Supplement to AETHIOPICA
International Journal of Ethiopian and Eritrean Studies
8

Pietro Maria Liuzzo

Digital Approaches to Ethiopian and Eritrean Studies

Harrassowitz Verlag

Pietro Maria Liuzzo

Digital Approaches to Ethiopian and Eritrean Studies

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Supplement to Aethiopica. International Journal of Ethiopian and Eritrean Studies
8

Edited in the Asien-Afrika-Institut
Abteilung für Afrikanistik und Äthiopistik
Hiob-Ludolf-Zentrum für Äthiopistik
der Universität Hamburg

Series Editor: Alessandro Bausi
in cooperation with Bairu Tafla, Ludwig Gerhardt
and Hilke Meyer-Bahlburg

2019
Harrassowitz Verlag · Wiesbaden

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Pietro Maria Liuzzo

Digital Approaches to Ethiopian and Eritrean Studies

2019

Harrassowitz Verlag · Wiesbaden

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

The publication of this volume was supported by the project TraCES: From Translation to Creation: Changes in Ethiopic Style and Lexicon from Late Antiquity to the Middle Ages, European Union Seventh Framework Programme IDEAS (FP7/2007–2013) / ERC Advanced Grant agreement no. 338756, and includes immediate Open Access.

On the front cover (left to right): Hermann Zotenberg, *Catalogue des manuscrits éthiopiens (gheez et amharique) de la Bibliothèque nationale*, Manuscrits Orientaux (Paris: Imprimerie nationale, 1877), 96b, MS Éthiopien 101 (from the Internet Archive, https://archive.org/details/manuscritsorient00bibl/page/96); MS Paris, Bibliothèque nationale de France, Éthiopien 101 in Beta maṣāḥəft encoded by Dorothea Reule (https://betamasaheft.eu/manuscripts/BNFet101/main); Henry Salt, *A Voyage to Abyssinia, and Travels into the Interior of That Country, Executed under the Orders of the British Government in the Years 1809 and 1810* (London: Printed for F. C. and J. Rivington by W. Bulmer and Co., 1814), 'Map of Abyssinia and the Adjacent Districts laid down partly from Original Observations taken in the Country and partly compiled from Information collected there by Henry Salt Esq. in 1809 & 1810' (from the Internet Archive, https://archive.org/details/voyageAbyssinia00Salt/page/n174); Sankey chart of textual units linked in first level <msItem>s of manuscript record grouped by period of the manuscript and keywords of the work excluding biblical material limited to 'poetry' (<https://pietroliuzzo.github.io/DHEth/MSS/EthioLitFlowWorksinMSSnoBiblPOETRY.html>).

On the spine: MS Bəḥerāwi Kəlləlāwi Mangəśti Təgrāy, Ḥarennat Gabazayti Qəddəst Māryām, MHG-010, f. 22v, Bishop Dexius writing the book of the *Miracles of Mary* (?) (photo courtesy of the Ethio-SPaRe project).

On the back cover (left to right): በለዕለ: in the Online *Lexicon Linguae Aethiopicae* (https://betamasaheft.eu/Dillmann/lemma/L1e891762428043be8130481dc731da73); Christian Friedrich August Dillmann, *Lexicon linguae aethiopicae, Cum indice latino. Adiectum est vocabularium tigre dialecti septentrionalis compilatum a W. Munziger* (Lipsiae: T. O. Weigel, 1865) (from the Internet Archive, https://archive.org/details/lexiconlinguaeae00dilluoft/page/n51), col. 56; encoding of MS Bəḥerāwi Kəlləlāwi Mangəśti Təgrāy, Ḥarennat Gabazayti Qəddəst Māryām, MHG-010, f. 22v, Bishop Dexius writing the book of the *Miracles of Mary* (?) by Abreham Adugna, Denis Nosnitsin, Pietro Maria Liuzzo, and Eugenia Sokolinski (https://betamasaheft.eu/manuscripts/ESmhg010.xml).

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.dnb.de abrufbar.

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the internet at http://dnb.dnb.de.

For further information about our publishing program consult our website http://www.harrassowitz-verlag.de

Printed on permanent/durable paper.
Printing and binding: Rosch-Buch Druckerei GmbH, Scheßlitz
Printed in Germany

ISSN 2196-7180
ISBN 978-3-447-11291-8
e-ISBN 978-3-447-19913-1

Table of Contents

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Notes for the Reader

Terms, especially acronyms, which are preceeded by a ↗ have a short definition in the Glossary at the end of the volume. In the PDF, this arrow will lead you to the correct point. If more occurrences of the same indexed word appear in the same paragraph, only the first will have an arrow. These terms do not appear also in the Index but backwards references to each occurrence are provided for each of them in the Glossary, and, in the PDF, these backwards page-references are clickable.

XML element names and attributes are in a different font from the rest of the text and, in the case of elements, use the standard notation `<myElement>`, in case of attributes `@myAttribute`. Unless otherwise stated, the XML namespace of the quoted elements is always TEI (http://www.tei-c.org/ns/1.0).

In the PDF, elements are linked to their specification in the TEI guidelines; attributes are linked to their page in the Beta maṣāḥǝft Guidelines; classes and properties are linked to a search in Linked Open Vocabularies (https://lov.linkeddata.es).

All entities in Beta maṣāḥǝft are accompained by their identifier. Appending this identifier to the home URL of Beta maṣāḥǝft https://betamasaheft.eu/ will redirect to the most up-to-date version of the information about the item. For literary works, the CAe number is also provided, linking to the main entry for the textual unit in the PDF.

Tagged place and person names which are in the Indexes, as well as cited passages, are linked in the text and in the Indexes to the URI used to identify them and are clickable in the PDF, as also all page and chapter cross-references.

The bibliography of this book is available in Zotero at https://www.zotero.org/pietroliuzzo/items/collectionKey/Z5MQAVPK. In the PDF all citations link to the correct reference and the page numbers in the Bibliography will lead back to each citation.

All code, as well as the online applications, are described in their status at the time of writing, the factors which can affect them and produce different results at the time of reading are not foreseable, especially for Beta maṣāḥǝft and the online *Lexicon linguae aethiopicae*, so there is no guarantee on what the reader will see to be exactly as described.

The code and the visualizations associated with this book are linked directly in the text where relevant. They are all available at the following link in a website which is bound to the present volume: https://pietroliuzzo.github.io/DHEth/. These will not be modified, unless for corrections, after the publication of this book.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

The data associated with the book is also stored by the Zentrum für nachhaltiges Forschungsdatenmanagement (https://www.fdr.uni-hamburg.de/) of Universität Hamburg and some other data repositories.

1) The application to produce this book, including scripts, website, and bibliography in Zotero RDF 10.5072/uhhfdm.128;
2) Beta maṣāḥəft data 10.5072/uhhfdm.132;
3) Beta maṣāḥəft Application 10.5072/uhhfdm.122;
4) Online *Lexicon linguae aethiopicae* Application 10.5072/uhhfdm.130;
5) Beta maṣāḥəft Guidelines 10.5072/uhhfdm.122;
6) Gəᶜəz Morphological Parser Application with TraCES data in TEI 10.5072/uhhfdm.124;
7) Trismegistos data https://www.trismegistos.org/tmcorpusdata/13/;
8) EAGLE Vocabularies https://github.com/EAGLE-BPN/epidocup-conversion/tree/master/edm%2Bvoc/vocabularies%20testing;
9) ISicily https://github.com/JonPrag/ISicily.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

List of Figures

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Acknowledgements

The current book, whose idea came from Alessandro Bausi in the first place, together with all its contents were produced thanks to the project Beta maṣāḥəft: Manuscripts of Ethiopia and Eritrea, Schriftkultur des christlichen Äthiopiens und Eritreas: eine multimediale Forschungsumgebung funded within the framework of the Academies' Programme (coordinated by the Union of the German Academies of Sciences and Humanities) under survey of the Akademie der Wissenschaften in Hamburg and thanks to the ERC Advanced Grant TraCES: From Translation to Creation: Changes in Ethiopic Style and Lexicon from Late Antiquity to the Middle Ages (Grant agreement no. 338756).

I would like to thank all those who have discussed with me the issues, the solutions, and the workflows presented in this book, providing me at every stage with invaluable suggestions, especially Bridget Almas, Patrick Andrist, Alessandro Bausi, Alice Bencivenni, Gabriel Bodard, Julian Bogdani, Antonella Brita, Yanne Broux, Hugh Cayless, Thibault Clérice, Eliana Dal Sasso, Steve Delamarter, Mark Depauw, Tom Elliott, Andreas Ellwardt, Achille Felicetti, Solomon Gebreyes Beyene, Jacopo Gnisci, Frank Grieshaber, Simone Maria Liuzzo, Marilena Maniaci, David A. Michelson, Matthew Munson, Francesca Murano, Silvia Orlandi, Jonathan Prag, Dorothea Reule, Matteo Romanello, Charlotte Roueché, Rainer Simon, Eugenia Sokolinski, Peter Stokes, Nafisa Valieva, Massimo Villa, and Anaïs Wion.

I would like to thank also all the other people which have made of the Hiob Ludolf Centre for Ethiopian Studies (HLCES) in the years in which I have been also part of it, the amazing research centre it is, Sophia Dege-Müller, Wolfgang Dickhut, Daria Elagina, Jonas Karlsson, Susanne Hummel, Magdalena Krzyżanowska, Denis Nosnitsin, Francesca Panini, Vitagrazia Pisani, Hewan Semon, Sisay Sahile, and Cristina Vertan.

I would also like to acknowledge the contribution to the development of this book by the Centre for the Study of Manuscript Cultures (CSMC), a real hub for discussion and research with continuous conferences and workshops at the highest level of discussion.

Without the continuous support of the members of the eXist-db community, none of this would have ever been possible for me. I would like to especially thank Wolfgang Meier, Leif-Jöran Olsson, Adam Retter, Winona Salesky, Claudius Teodorescu, Joe Wicentowski, and Lars Windauer.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

A special thank should be also given to the members of the TEI community, too many to be named one by one, and to the organizers of one of the most useful conferences I have ever attended, The Balisage Markup Conference (<https://www.balisage.net/>). A number of questions on the daily struggle of getting code to work are answered in Stack Overflow. I would like to thank all the generous contributors of that community as well. Any errors, mistakes, and imprecisions are my own responsibility and I will be grateful to those who will be so patient to point them out and let me know.

Pietro Maria Liuzzo
Universität Hamburg
pietro.liuzzo@uni-hamburg.de
July 2019

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Introduction

πάντα γὰρ ἡ χρεία διδάσκει τὴν φύσιν, οἰκείως τοῖς ὑποκειμένοις καιροῖς ἁρμοζομένην πρὸς τὴν ἐκ τῆς ἐλπίδος εὐχρηστίαν
(Necessity teaches Nature everything, adapting suitably with the present circumstances for the best of what can be hoped for)
(Diod. 3.15.7)

1 Aims and scope

Ten years ago, in a brilliant article, Sculley and Pasanek wrote, 'the collaborative field of digital humanities, still in its nascency, is characterized by new collaborations between scholars of language and literature on the one hand, and computer scientist on the other.'[1]

After ten years we hear voices already claiming that digital humanities are about to die. All scholars and researchers in the so called 'humanities' or 'Geisteswissenschaften' will use digital tools and there will be no more need for specialists in that particular field. In 2018 humanities scholars have enough digital competences and there are tools smart enough for them to engage with, so that the collaboration characterizing digital humanities *can* now be among scholars in the humanities without intermediation, in some cases.[2] The additional collaborations with computer scientist are thus more informed, eventually on both sides, and more fruitful. However, to lose the possibility of discussing the methodologies and directing the theory which is driving what is arguably the biggest revolution since the arrival of printing, would clearly be an undesirable side effect of decades of pioneering work which is now increasingly informing the decisions of policy makers. This discussion needs specialists in the collaboration between humanities scholars and computer scientists, not just members of two groups. The distinction of the three profiles itself should be blurred and questioned instead.

Digital humanities are no longer in their nascency and they are hopefully not dying. They can perhaps be said to have reached a maturity phase in which it is finally less difficult to pass on competences, which can thus spread more

1 Sculley and Pasanek 2008, 409–410.

2 Van Zundert 2016, 88–89 for a brief and clear discussion of the denomination 'digital humanities'.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

quickly and securely. A researcher, who is willing to learn, does not need to have a mathematical or classificatory mind, but sometimes just to be pointed to existing resources or tools and to keep in touch with other people. This is a luxury which previous generations did not have, and that we ought to treasure, so let me begin by acknowledging this fortunate situation.

However, digital methods are not good per se and can lead to errors and imprecisions as much as non-digital methods. They do not guarantee a better quality of contents or better research results but they should be especially preferred when we can actually get more from their use, where this 'more' is not simply an increased diversification of the types of access. The evaluation of the outputs of projects using digital methods as well as the digital methods themselves which lead to these outputs should follow the same principles which apply to any scientific product. It might well be that a laudable software methodology produces an output which is from the perspective of the users useless or wrong. While users and readers should have full critical awareness of the fact that no formalization of information is free from interpretation, the medium in which the results are presented has to be taken into consideration. In most cases of encoding I would say that 'objectiveness' is not even desired, while the strive to make clear and transparent an interpretation is part of the daily work. Hopefully the following chapters will be persuasive about the added value which web-based standards and methods developed by digital humanities researchers can bring, enabling us to 'see how to use programming as a means of inquiry—all without becoming full-blown computer scientists'.[3]

1.1 Code

There is a second point in the same article by Sculley and Pasanek from which I began which I would like to take into consideration: 'the No Free Lunch

3 Montfort 2016, 108. Resources to get introduced to digital humanities abound. A good companion is certainly Schreibman et al. 2016 and, among the numerous digitally available resources, see the peer reviewed <https://programminghistorian.org/>. See also Graham et al. 2015 and <http://www.themacroscope.org/2.0/>. The SunoikisisDC International Consortium of Digital Classics Programs (<https://www.dh.uni-leipzig.de/wo/sunoikisisdc/>) offers plenty of resources including references, videos, and other materials. There are also plenty of Frequently Asked Questions websites and projects which will help in finding answers, like the Stack Overflow (<https://stackoverflow.com/>) communities, Digital Humanities Questions and Answers (<http://digitalhumanities.org/answers/>), and the XML FAQ by Peter Flynn (<http://xml.silmaril.ie/>) which has been also translated into Amharic by Abass Alamnehe, <http://senamirmir.com/xml/faq/xml_faq_amh.html>. Needless to say, one can learn plenty on active mailing lists like the TEI List, the Markup List, the exist-open mailing list, and many others.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Theorem means that there is no single best learning algorithm, and we may have to employ a good deal of ingenuity to learn from difficult data.'[4]

This was meant specifically for the learning algorithms discussed, but let me extend the scope of the sentence a bit to simply say that there is no single best algorithm; thus, in particular, the scripts found in this book are not necessarily even among the plurality of the best algorithms. They are scripts which strive to translate parts of research questions in order to interrogate the data, but I do hope that some readers will find comprehensible what I am trying to say in a slightly more humanly readable gibberish.

The fact that I am not an engineer nor have I ever received formal training in programming or digital humanities might appear evident to the reader from my code examples as well as in the limited number of programming languages and functions which I use. However, what I learned from my mentors and colleagues, by trial and error, or reading good manuals,[5] was enough to produce working solutions for the needs of the projects I work for which might not be the latest technological trend, but are kept as close as possible to the needs of the team and users. No piece of code or data should be taken as exemplary as such and removed from its context, and there are certainly many issues of opinion or misinterpretation in each of the many lines of code written for this book and the resources it speaks about.

I have endeavoured to expose and explain all the code and the data that I am using, because I believe that this code is required in order to be as clear as possible and not to keep the magic formulas hidden. I also want to show that anyone can use them; it does not take a particular mindset to use that magic. It is easy enough for anyone to learn.

I am aiming for a descriptive and lightly didactic approach by example, which might persuade some of the interested practitioners of the usefulness of working with community-based standards and disenchant some of the blindly enthusiastic about the endless possibilities of the Web or of databases. There are things which have many more prerequisites than some tend to think, in order to work in the way they believe they ought to. I do not want to push for a confusion between methods and aims and to make of the technical standards employed the actual aim of the effort; that would be like writing a book just to demonstrate the use of an alphabet. I also do not want to evaluate the advantages of one or the other technique or workflow, nor to compare the age of paper against the digital world.

4 Sculley and Pasanek 2008, 412.

5 For example Kay 2008 for ↗XSLT, Walmsley 2015 for ↗XQuery and Siegel and Retter 2015 for ↗eXist-db.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

I would like instead to stress from the outset the utmost importance of the basic research in comparison to the web-based technologies and standards used. We can do great things with ↗XML, with ↗XQuery, ↗Python and ↗Docker, ↗Java and ↗JavaScript, and we can do them very quickly all in all. There are plenty of possibilities and plenty of occasions but we rarely learn anything which is thought-provoking or inspiring from the software stack alone. On the other side, after dedicating some time to programming, it is almost unavoidable that we look back at some of the ingenuities of the so-called traditional methodologies, and start thinking about them with fresh tools to formulate and answer questions. Scholars in the humanities have learned a lot from interactions with colleagues with computational background. One day perhaps (and perhaps that day is already here) classical philologists will try and answer simple and more complex questions in Stack Overflow leaving with generosity their answer to the evaluation of unknown peers; but perhaps this is not really the most important part of learning in this process, especially when compared with logic, separation of concerns, and documentation.

1.2 Readers

This book is primarily intended to share this experience-based know-how with scholars in Ethiopian and Eritrean studies, as well as of other disciplines, and especially practitioners and students interested in using some of the possibilities offered by the Web and by technologies related to ↗XML and ↗RDF. Readers will find here also synthetic evidence of the enormous amount of collaborative thinking and reasoning that has gone into developing all of the functions, from the least to the most complex, described in the frame of the Beta maṣāḥəft project. I hope that they might find some of these interesting even if in the future the technologies and techniques used will no longer be remembered. This is also intended to be part of a series of actions taken to make the open and collaborative approach of this project a reality and to keep it as transparent as possible, encourage as many interested people as possible to exploit it and the resources it provides for their own work, whatever their interest might be.

I will show some examples of the use of web-based technologies and standards to face common tasks and questions. I will do this mainly by taking specific examples of methodologies or basic requirements or questions and showing what can be achieved with these technologies either in support of the research process, to verify its parts, or to simply document it to make the results as useful and as easily reusable as possible with reproducible queries and workflows.

The book is intended primarily for scholars in Ethiopian and Eritrean studies without or with little digital training, but might be of interest to scholars in

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

related disciplines like Semitic studies, codicology and epigraphy, who might find here some examples or information useful for their studies. Also scholars in digital humanities and related disciplines may find it useful, insofar as they could find here information about the Beta maṣāḥəft project and some of the principles it tried to follow.

1.3 Scope

This book is written at the beginning of a long-term project, since Beta maṣāḥəft started in 2016 and is expected to continue its work until 2040. The book is based on the experience gained so far, and, hopefully, much more will have to be said in the future, about the project and about the use and reuse of the data it curates.

The scope of the research presented is thus that of the project in which it was hosted, Beta maṣāḥəft: Manuscripts of Ethiopia and Eritrea, Schriftkultur des christlichen Äthiopiens und Eritreas: eine multimediale Forschungsumgebung, and, therefore, the manuscript tradition and written culture of Ethiopia and Eritrea.[6] All examples are related to the Beta maṣāḥəft project which means, among other things, that none of the examples can be applied without caution to any other linguistic context in Ethiopia or Eritrea. The research presented here only deals with Classical Ethiopic or Gəʿəz. I do not deal with linguistics and corpus linguistics either.

The parts of the book related to lexicography and the morphological parser have been developed in collaboration with the project TraCES: From Translation to Creation: Changes in Ethiopic Style and Lexicon from Late Antiquity to the Middle Ages.[7]

The project Beta maṣāḥəft aims at serving a range of research fields with a common scope or type of interest wider than simply Ethiopian and Eritrean studies. Some of the cases presented, for example the one about inscriptions (Ch. 2) will offer insights into this, some others (e.g. Ch. 6) will show some of the possibilities to reuse the data and the code for other projects and interests.

However, Beta maṣāḥəft was not the first and is not the only project with interests in the application of digital technologies to the field of Ethiopian studies. In this Introduction I will as a first task draw a short 'State of the data' in which I will describe the resources and the projects which have been accomplished or are available. This is done without aiming at completeness as

6 Several contributions have already been published with details on some aspects of the project, e.g. Reule 2018; Solomon Gebreyes and Liuzzo 2018; Villa 2018; and Liuzzo 2017. The official website of the project is <https://www.aai.uni-hamburg.de/en/ethiostudies/research/betamasaheft.html>.

7 <https://www.traces.uni-hamburg.de/>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

I hope many more initiatives will appear in the time between the completion and the publishing of this book. This 'State of the data' should also serve to draw a picture of the sector-specific digital humanities tools available. I will not list many other digital projects and enterprises which are very relevant but not specific to Ethiopian studies. It is of course one of the best features of digital resources, to be reusable for different purposes, and the examples of projects and initiatives in other fields of research which have been taken as examples are many, but these will be discussed where needed in the various chapters.[8]

It is also convenient to state at the outset what is the set of web standards and technologies which are used for the Beta maṣāḥǝft project and for all the contents of the current book.[9] I will thus provide here also a short introduction to the project workflow and tools.

The last part of the Introduction, starting from the definition of the reference concept of 'written artefact',[10] will outline the contents of the book.[11]

2 State of the data

More than twenty years ago, Manfred Kropp gave a general overview of the development of Ethiopian studies including early computational projects. He stressed already at that time the need of these technologies for Semitic studies when he said, for example, 'the era of heroes like A. Dillmann has gone; not because we are so different or lazy, but because of the general conditions of modern life – thus collective work on well organized data bases might well be the melody of the future.'[12] We happen to be in that future, and to play that melody, but let us see what it sounds like.

At a conference in Verona in 2010, Alessandro Bausi listed the main lacunae and desiderata for Ethiopian studies, in a more precise way and with already some existing first attempts to look at.[13]

8 A good place to start from when looking for examples of digital projects is the Digital Classicist Wiki (<https://wiki.digitalclassicist.org/Main_Page>) which does have a lot of projects which are not necessarily linked to classics (Greek and Latin).

9 See p. xxxiv.

10 See p. xli.

11 See p. xlvii.

12 Kropp 1994, 118.

13 Alessandro Bausi kindly made available to me his notes for the contribution 'Etiopistica e filologia digitale' to a conference held in Verona in 2010, with title III Incontro di Filologia Digitale. The text and part of the description in the list in the following section were re-elaborated from that unpublished contribution.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

There are actually lacunae for every aspect of the digital work in digital philology for Ethiopic. As to the desiderata, one may list some 'ideal' necessary tools well beyond strict digital philology:

1) *Bibliotheca aethiopica*, a series of new critical editions:
   a) text editions should follow updated and adequate philological methodology;
   b) special care should be devoted to the text encoding process, which should be consistent with international standards;

2) *Clavis aethiopica*, a complete repertory of all the known Ethiopic texts (literary, either original or not, historical, liturgical, documentary, etc.), with extensive 'incipit' and 'desinit'; the entries should be identified with an alphanumeric string and furnished with complete lists of bibliographical references (editions, translations, secondary literature) and of all the known manuscripts which attest them. This all-embracing 'Clavis' should be done for all the texts and conceived since the beginning as an online repertory;
3) *Thesaurus linguae aethiopicae*, an as large as possible digital corpus ('thesaurus') of Ethiopic texts: this goes well beyond the obvious need of more concordances of texts; epigraphical texts should also be included in the corpus;
4) *Lexicon linguae aethiopicae*: an updated and enlarged Ethiopic dictionary: the *Lexicon linguae aethiopicae* by August Dillmann, dating from 1865, is still the only true Ethiopic dictionary for philological study, but hundreds if not thousands of new texts as well as new editions of already known texts (let's think for instance of the great enhancement of our knowledge of the Ethiopic New Testament) have been produced in the meanwhile, not to speak, again, of the epigraphical texts (obviously absent from Dillmann's *Lexicon* and little considered even in Wolf Leslau's *Comparative Dictionary of Geᶜez*);
5) *Catalogus manuscriptorum aethiopicorum*: besides preliminary working tools (such as updated inventories of catalogues, libraries, lists of manuscripts) it is crucial to keep a high standard in dealing with the texts: from this point of view, the KOHD-project (Katalogisierung der Orientalischen Handschriften in Deutschland) has attained remarkable standards of excellency, and what may be done is just to enhance the description of the material aspects, and, most of all, to conceive cataloguing since the beginning as destined to an online platform;

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

6) *Bibliographia aethiopica*: a systematic implementation of the already available bibliographies (German and English) to include all the contributions written in Italian, French, and Portuguese at least.

In the fifteen years between Kropp's article and Bausi's survey only some progress had been made which left a lot to be desired. But it was much more clear already, by then, what to wish for and some of the early enthusiasm had given way to a more achievable approach. We had learned by then with twenty years of World Wide Web that not all the magic imagined in the nineties was actually possible, that there was magic we had not yet imagined, and that, in case it was magic, was not that much faster than non-magic. In the first decades of this century the diversity of possibilities and solutions also widened so broadly that additional issues of choice of technologies and lack of skills arose.

It took almost another ten years to get to the point we are at now, where this music is starting to play and the desiderata begin to become reality. We have also learned that this kind of work would not make our efforts less heroic than those of August Dillmann, in that it takes a lot of work and effort to do this collective work and organize well these 'data bases'. It takes additional time to annotate the texts and again additional time to learn programming languages, develop scripts, or adapt existing ones to get results which are usable. There is also a major shift in focus which from computer-assisted research techniques moves to the aim of creating resources available online, that is, from a software-oriented to a resource-oriented approach.

This section gives an overview of the digital projects in Ethiopian studies known to me to date, and of some projects which produced data which is now available in databases or contributed to the definition of the requirements for digital resources for Ethiopian studies. Better descriptions of the projects can be found at the links provided.[14]

14 This is not a list of digitization projects but of projects with at least one web-related component, like a website or a web-based database, a catalogue, a digital edition, etc. Especially important and relevant to any work with Amharic, Gəʿəz, and other languages of Ethiopia are the efforts of Daniel Yacob and the Ethiopic Layout Task Force (<https://www.w3.org/International/groups/ethiopic-layout/>) towards a recommendation for Ethiopic layout (<https://w3c.github.io/elreq/>). A recent more general survey of online resources is also offered by Zaccaria 2018.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

2.1 Relevant projects

2.1.1 Accordance Ethiopic (Gəʿəz) Old Testament

This is a standard text of the Bible, based on the work by the team of the Capuchin Friary in ʾAddis ʾAbabā and the texts digitized by Ran HaCohen.[15] It can only be purchased with the Accordance software to compare it to other texts in that software. This resource will be certainly useful, although limited to customers of this software. The texts provided here are not new critical editions as far as I could gathered.

<https://www.accordancebible.com/store/details/?pid=ETHIOP>

2.1.2 Biblia Veteris Testamenti Aethiopica, Tell Aviv University, Israel

Project head: Ran HaCohen.

Since 2009, Prof. Ran HaCohen has made a number of texts in Unicode and PDFs of editions available online. There is a digitization of the Ethiopic Octateuch according to August Dillmann's edition, and several PDFs, including the *Lexicon linguae aethiopicae*.

<http://www.tau.ac.il/~hacohen/Biblia.html>

2.1.3 Comparative Oriental Manuscript Studies (COMSt)

Chair of the steering committee: Alessandro Bausi.

Comparative Oriental Manuscript Studies was a European Science Foundation Networking project which ran from 2009 to 2014. Its aim was to coordinate existing work in manuscript studies relative to different languages. It produced a handbook, *Comparative Oriental Manuscript Studies: An Introduction*, which is a guide for any future study of manuscripts and set a new benchmark line for digital resources.

The comparison of achievements and methodologies resulted in a unified approach to manuscript studies, on which the Beta maṣāḥəft views of manuscripts are largely based.

<https://www.aai.uni-hamburg.de/en/comst/>

2.1.4 Centre for the Study of Manuscript Cultures (SFB 950—Manuscript Cultures in Asia, Africa and Europe)

Antonella Brita organized missions to find and digitize manuscripts in Ethiopia for the last eight years. These have been partly entered in the database of the Centre for the Study of Manuscript Cultures. These will be inte-

15 This information was provided by Ralph Lee in an email exchange.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

grated in Beta maṣāḥəft as soon as possible. These activities were part of two sub-projects directed by Alessandro Bausi:

1) 2011–2015: Cross-Section Views of Evolving Knowledge: Canonico-Liturgical and Hagiographic Ethiopic Christian Manuscripts as Corpus-Organizers;
2) 2015–2019: 'Parchment Saints': The Making of Ethiopian Hagiographic Manuscripts—Matter and Devotion in Manuscript Practices in Medieval and Pre-modern Ethiopia.

<https://www.manuscript-cultures.uni-hamburg.de/Projekte_p2_e.html#C05>

2.1.5 Chojnacki Stanisław archive at the Digital Vatican Library

Jacopo Gnisci, as part of the Beta maṣāḥəft project, is digitizing the photographic archive of Stanisław Chojnacki.[16]

<https://digi.vatlib.it/all/search?k_f=0&k_v=Chojnacki>

2.1.6 Digital Archive for the Study of Pre-Islamic Arabian Inscriptions (DASI)

Director: Alessandra Avanzini.

The Digital Archive for the Study of Pre-Islamic Arabian Inscriptions is a collection of corpora of inscriptions with images and deep markup which includes also a work-in-progress Corpus of Early Sabaic Inscriptions from Ethiopia.[17]

<http://dasi.cnr.it/>

2.1.7 Database of digitized Ethiopic texts

Eugenia Sokolinski has produced a database of digitized text with an indication of where they can be obtained.

<https://creator.zoho.com/comstudies/ethiopictexts/>

2.1.8 Digital library for Tigrinya

Gezae Haile Weldemichael, working at Maqala University as lecturer and researcher in the Department of History and Heritage Management is preparing a digital library for Tigrinya language, translating manuscripts from Gəʿəz. The work, not yet published as a website, is still at the beginning and is car-

16 See p. xxxii.

17 Avanzini et al. 2014.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

ried out using Greenstone and Adobe Dreamweaver.[18] This project is not yet online.

2.1.9 *Encyclopaedia Aethiopica* (EAe)

General editors: Siegbert Uhlig, Alessandro Bausi.

The well-known reference work to access most of the primary topics about the 'Orbis Aethiopicus' is an invaluable source of information and its indexes have been curated and digitized, with the attribution of a unique identifier, by Eugenia Sokolinski. These indexes have formed the core data for the Beta maṣāḥəft project, and there is digital material which could be used for a future digital edition, connected to Wikipedia and other encyclopedic resources available online.

<https://www.aai.uni-hamburg.de/en/ethiostudies/research/encyclopaedia/eae.html>

2.1.10 Ethiomap

Project heads: Éloi Ficquet, Wolbert Smidt.

Ethiomap—Exploring Modern Maps of the Horn of Africa (18th–20th c.) combines online visualization and indexation tools to explore and study a collection of historical maps of north-eastern African territories. If places found within today's boundaries of Ethiopia are topographically central in the collection, hence the project's name, other places related to neighbouring countries are also part of the work. The scope of the research is limited to maps dated from 1790 to 1944, which are deemed to be important as these represent either scientific advances or historical turning points.[19]

Éloi Ficquet has also been awarded recently a ↗Pelagios Resource Development Grant for a project entitled 'Exposing 18th–20th geo-historical knowledge of the Horn of Africa into the Web of Data' to expose linked data and connect the ↗annotations in Ethiomap with Pelagios and Beta maṣāḥəft.

<https://ethiomap.huma-num.fr/ >

2.1.11 Ethiopic Psalter Project

Head of project: Steve Delamarter.

18 I was informed by Gezae Haile Weldemichael about this during the Summer School in Ethiopian and Eritrean Manuscript Studies held in Maq̱ala, Ethiopia, 24–29 September 2018 with the support of the Volkswagen Stiftung.

19 Éloi Ficquet has also an independent workflow to map the itineraries in the *Chronicle of ʾIyāsu I* (CAe 3950, ID: LIT3950ChronIyasu), which he analyses with the support of Google Earth.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

The Ethiopic Psalter Project is collecting features about several hundreds of Psalters to produce computable data for quantitative analyses.

During a period in which Steve Delamarter was doing research at the Hiob Ludolf Centre for Ethiopian Studies, we worked together towards a workflow producing some comparable ↗KML data to be used in the Dariah-DE Geobrowser, starting from the Google Spreadsheet in use for the data collection. This allowed parallel map and time analyses of the data at hand, however this project has now added an additional workflow and tools.

2.1.12 Ethiopian Manuscripts Archives (EMA)

Head of project: Anaïs Wion.

The Ethiopian Manuscripts Archives is a project focused on archival and historiographical documents contained in Ethiopic manuscripts. It is organized by corpus and has a team of editors for each group of documents which are annotated also according to diplomatic features.

<http://www.cn-telma.fr/publication/zekra-nagar-ema>

2.1.13 Ethiopic Manuscripts Imaging Project (EMIP)

Head of project: Steve Delamarter.

The Ethiopic Manuscripts Imaging Project (EMIP) was started in the spring of 2005 by Steve Delamarter with the mission to help preserve images of Ethiopia's manuscript heritage and make them available for scholarly study. The manuscripts were located in the possession of university libraries, dealers, and private owners. All of the owners have either loaned or rented their manuscripts to be digitized and made available for scholarly study.

The project has produced several catalogues of these manuscripts.

There are now more than 12.000 manuscripts and more than 1200 scrolls in this collection.

Part of the images of the collection was available via Vivarium, the Online Digital Collections of Saint John's University and the College of Saint Benedict hosted by Hill Museum & Manuscript Library (HMML). They will be available also in Beta maṣāḥǝft.

<http://cdm.csbsju.edu/digital/>

2.1.14 Ethiopian Manuscript Microfilm Library (EMML)

Director: Getatchew Haile, William F. Macomber.

The Ethiopian Manuscript Microfilm Library at Hill Museum & Manuscript Library is a collection of microfilmed manuscripts (more than 8000) which were collected in the framework of a conservation and preservation project. Started in the pre-digital era and conducted between 1973 and 1991,

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

with an interruption between 1987 and 1991, it is the result of the collaboration between the Saint John's University, the Ethiopian Orthodox Tawāḥədo Church, and the Ministry of Culture of Ethiopia. The manuscripts are now available via vHMML.

<http://hmml.org/manuscripts/>

2.1.15 Ethiopicist

This project by Augustin Dickinson and Sean M. Winslow is a nascent website aimed at those who are beginning their journey in Ethiopian studies. It will collect resources, introductory posts, and some selected materials.

Two further projects are part of this, on magico-religious texts (<https://asmat.ethiopicist.com>) and saints (<https://qeddus.ethiopicist.com>).

<https://ethiopicist.com/>

2.1.16 Ethio-SPaRe

Head of project: Denis Nosnitsin.

The project Ethio-SPaRe: Cultural Heritage of Christian Ethiopia: Salvation, Preservation and Research digitized more than 2000 manuscripts in the northern region of Ethiopia and entered in an online accessible database little more than 1000 of these new manuscripts (<https://www.aai.uni-hamburg.de/en/ethiostudies/research/ethiospare.html>). It was the most recent large-scale achievement in codicological field research in Ethiopia and the descriptions continue to be improved and updated in Beta maṣāḥəft. The descriptions, which follow largely up-to-date codicological standards (they include quire descriptions, formulas for the ruling and pricking, etc.), are available in a publicly accessible ↗MyCoRe database (<https://mycms-vs03.rrz.uni-hamburg.de/domlib/content/below/index.xml>) and so are the images upon request of permission.

All the descriptions are available in Beta maṣāḥəft, while only part of the images are.

2.1.17 Gunda Gunde Project

Directors: Michael Gervers, Ewa Balicka-Witakowska.

The Gunda Gunde Project offers to the end users about 120 (119 at the time of writing) manuscripts descriptions and images from the Gunda Gunde monastery.

<http://gundagunde.digitalscholarship.utsc.utoronto.ca>

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

2.1.18 Inventaire des bibliothèques et des catalogues des manuscrits éthiopiens

Heads of project: Anaïs Wion, Claire Bosc-Tiessé, Marie-Laure Derat.

This is an updated online version of a repertory of catalogues of Ethiopic manuscripts published in paper form in 1995.[20] These have been linked from Beta maṣāḥəft and, where possible, to the Wikidata entites available for each of the institutions in the list.

<http://www.menestrel.fr/spip.php?rubrique694>

2.1.19 Islam in the Horn of Africa (IslHornAfr)

Head of project: Alessandro Gori.

Islam in the Horn of Africa has done ground-breaking field research in the Islamic manuscript culture of Ethiopia, collecting images and catalogue descriptions in the last five years.

<http://www.islhornafr.eu/>

2.1.20 Mäzgäbä Səəlat

Directors: Michael Gervers, Ewa Balicka-Witakowska.

Mäzgäbä Səəlat: Treasury of Ethiopian Images is a database of images related to Ethiopia including some manuscripts. The database contains an extremely large number of documents and some simple search functionalities.

<http://ethiopia.deeds.utoronto.ca/>

2.1.21 Princeton Ethiopian Miracles of Mary Project (PEMM)

Principal Investigator: Wendy L. Belcher.

This newly launched project aims at the study of the *Miracles of Mary* in the Princeton collection.

<https://wendybelcher.com/african-literature/pemmproject/>

2.1.22 The Textual History of the Ethiopic Old Testament Project (THEOT)

Heads of project: Steve Delamarter, Curt Niccum.

The Textual History of the Ethiopic Old Testament Project (THEOT) aims to provide an initial assessment of the transmission history of the Bible in Ethiopia from extant manuscripts. The project works on one side on the development of computer scripts to generate, aggregate, and analyse data regarding textual variation in Ethiopic manuscripts and, on the other side, on practices to interpret and visualize the data.

20 Beylot and Rodinson 1995.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

<https://ethiopic-tool.firebaseapp.com/>

2.1.23 Thesaurus linguae aethiopicae (TLA)

Head of project: Manfred Kropp.

The Thesaurus linguae aethiopicae was a computer-based texts database for Ethiopic language and literature, which unfortunately is not any more available, and was started as early as 1991 (<https://www.rlp-forschung.de/public/people/Manfred_Kropp/research_projects/1315>). It was conceived and founded in the 1990s by Manfred Kropp, who has been one of the first and few to show interest towards the application of computer science to Ethiopian studies. It is the first attempt at an application to Ethiopic of computational methodologies which used web technologies (↗SGML). The site comprised some texts encoded according to pre-Unicode standards.

2.1.24 TraCES

Head of Project: Alessandro Bausi.

TraCES: From Translation to Creation: Changes in Ethiopic Style and Lexicon from Late Antiquity to the Middle Ages does deep morphological ↗annotation of a diachronic selection of edited texts using a newly developed Multilevel annotation tool, ↗GeTa (Gəʿəz Text Annotation).

The team has also produced a curated and updated version of Dillmann's *Lexicon linguae aethiopicae*.[21]

<https://www.traces.uni-hamburg.de/>

2.1.25 Virtual HMML (vHMML)

Head of project: Columba Stewart.

vHMML is a highly dynamic access point to learning and research materials offered by the HMML. Its reading room gives access to the collections of manuscripts of HMML, including the Gunda Gunde manuscripts and the EMML manuscripts with detailed descriptions.

vHMML also has a data portal (<https://www.vhmml.org/dataPortal>) and few samples of digital humanities research carried out on this data in vHMML DH (<https://www.vhmmldh.org/>).

<https://www.vhmml.org/>

21 See Ch. 5.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

2.1.26 Layout and Styling of Zaima Chant with Web Standards

Daniel Yacob, founder of Gəᶜəz Frontier Foundation,[22] has published online his work on the musical notation of hymns in Gəᶜəz. Using Web Standards and the Ruby notation, borrowed from Japanese, he is able to produce a viewer which deals with several layers of musical notation to display it in a web-based viewer using ↗CSS.

Related to this is also the work of the same author on a new font, which includes in the private use area some 300 extra signs for the musical notation.

<https://w3c.github.io/elreq/zaima/>

2.2 Available images of manuscripts

There are several other institutions that hold and make available images of Ethiopian manuscripts. The Beta maṣāḥəft project maintains a list of available images and descriptions of Ethiopic manuscripts online.[23] This list is curated especially by Dorothea Reule and Massimo Villa and is not to be confused with the list of repositories or with the list of catalogues of manuscripts which the project has as part of its main architecture.[24] These lists include all institutions which the project team knows of (and needs for the manuscript records), regardless of the availability of images from within the project or not. For those manuscripts for which the project has a record and the location of the images is known, if the images are publicly available via ↗IIIF, then the ↗manifest is added to the metadata; if the images are available in any other format a link is provided.

Images from the Chonjacki collection, which includes pages of manuscripts, have been digitized by Jacopo Gnisci within a collaboration between the Beta maṣāḥəft project and the Digital Vatican Library. Although these resources are disconnected from the others and are also physically stored at the Vatican Library, they are accessible in Beta maṣāḥəft within a dedicated section from the places to which they are related and as a full list.[25]

2.3 Cooperations

The list of projects and available resources presented shows quite well the amount of digital work which was carried out for Ethiopian manuscript studies. All this is a large contribution to the desiderata which have been listed

22 <http://www.geez.org/>.

23 <https://betamasaheft.eu/availableImages.html>.

24 Respectively available at <https://betamasaheft.eu/manuscripts/browse> and <https://betamasaheft.eu/catalogues/list>.

25 <https://betamasaheft.eu/chojnacki/viewer>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

above and which, however, largely remain desiderata, with the added frustration of having to traipse across websites figuring out what is where and what level of accuracy and completeness is provided. Among the projects listed only the *Bibliographia aethiopica*, on its way as a ↗Zotero group library (Ch. 8, p. 223), and the online *Lexicon linguae aethiopicae* (Ch. 5), produced by the TraCES project, are actually online resources up and running available to everyone.

The project Beta maṣāḥəft started from those needs and from the available resources and, thanks to its long-term funding, can tackle the other desiderata, especially the *Catalogus manuscriptorum aethiopicorum* (Ch. 1), the *Clavis aethiopica*, and the *Bibliotheca aethiopica* (Ch. 3), while adding to that list two further major tasks, that of a gazetteer of ancient places in Ethiopia (Ch. 4) and of a prosopography of Ethiopian persons.

Data from the above listed DASI, EMA, EMIP, EMML, Ethio-SPaRe, and IslHornAfr projects, who kindly contributed their efforts to this new collective enterprise, was converted to ↗TEI, indexed into Beta maṣāḥəft and linked to existing resources (like the manuscript reading room of vHMML, with the permalink provided). It can then be used and updated collaboratively also there.

This is the state of the data digitally available for Ethiopian studies to date to my knowledge. But I would like to at least name here also some other projects which have parallel aims and structures in cognate fields of research in Oriental studies and which have been important for the development of the Beta maṣāḥəft resources and with which the project has active cooperation in place, like the PAThs project,[26] the Corpus dei Manoscritti Copti Letterari (CMCL),[27] and Syriaca.org.[28]

Now that the context is clear, we can progress to briefly describe the project Beta maṣāḥəft and the core technologies it uses, which are at the basis of all the examples provided in this book.[29]

26 Tracking Papyrus and Parchment Paths: An Archaeological Atlas of Coptic Literature Literary Texts in their Geographical Context (<https://atlas.paths-erc.eu/>) is a ERC project directed by Paola Buzi at Sapienza Università di Roma.

27 The project, directed by Tito Orlandi, is the reference resource for Coptic texts (<http://cmcl.let.uniroma1.it>).

28 The Syriac reference portal comprising numerous resources for the study of Syriac (<http://syriaca.org>).

29 This is true also of the online *Lexicon linguae aethiopicae* (Ch. 5) which is also based on ↗TEI data and powered by ↗eXist-db, and the Gəʿəz Morphological Parser (Ch. 6), although they are not part of Beta maṣāḥəft as a project.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

3 Beta maṣāḥəft

Thirty years ago, there was possibly not even the thought of digitally shared resources quickly available via a thing called 'the Web', especially for Ethiopian studies. Now we have Beta maṣāḥəft: Manuscripts of Ethiopia and Eritrea, Schriftkultur des christlichen Äthiopiens und Eritreas: eine multimediale Forschungsumgebung.[30] The ↗XML data is the principal outcome of the project, and is curated with priority to the online representations of it.[31] The project aims to achieve a collection of manuscripts, literary and documentary texts, as well as reference authority files for historical persons and places.[32]

Beta maṣāḥəft is based on international standards for data and uses open software, following in as much as it is possible the ↗FAIR Guiding Principles for both data and software produced.[33] Let me try to self-assess the FAIRness of Beta maṣāḥəft based on these guiding principles to make clear at the outset how the project strives to 'facilitate and simplify the ongoing process of discovery, evaluation, and reuse in downstream studies'[34] of the data it curates. The resources in Beta maṣāḥəft are *Findable* in the FAIR meaning as they all get a unique ↗URI and are described with rich metadata in ↗TEI, including the URI in all standard formats with which the data is exposed. Beta maṣāḥəft data is *Accessible* because it can be retrieved via ↗HTTP, is open and free. Everything is encoded in ↗XML using TEI elements and guidelines,[35] and our ↗schema is a customization of the TEI using its modules.[36] The web applica-

30 The project is led by Alessandro Bausi and coordinated by Eugenia Sokolinski. The author of the present book is in charge of all technical and digital humanities related aspects of the project. A video presentation of the project is available at <https://youtu.be/bI950izCu2E>. Beta maṣāḥəft is a ↗TEI project. We use ↗GitHub to store and control versions of our data and we deploy our data with ↗eXist-db. None of these means that our TEI data is perfect, on the contrary it is subject of continuous revision as we find out inconsistencies or others point them out.

31 See Turska et al. 2016, § 4 'the encoded texts are the most important long-term outcome'.

32 Kropp 1994, 122.

33 Wilkinson et al. 2016.

34 Wilkinson et al. 2016, 1.

35 TEI Consortium et al. 2018. The ↗XML ↗TEI descriptions can all be seen both in the ↗GitHub repositories (<https://github.com/BetaMasaheft>), and appending `.xml` to any identifier, e.g. <https://betamasaheft.eu/BLadd24994.xml>, or clicking on the TEI /XML buttons in the web ↗application.

36 The current ↗ODD, which is at the base of the ↗schema in ↗Relax-NG used to validate the ↗TEI data, is here <https://github.com/BetaMasaheft/Schema/blob/master/tei-betamesaheft.xml>, and can be viewed also in a more readable fashion in the Beta maṣāḥəft Guidelines, <https://betamasaheft.eu/Guidelines/?id=schemaView>. Many

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

tions which Beta maṣāḥəft uses are powered by ↗eXist-db,[37] a native XML database which allows full data flexibility.[38] This, together with the parallel storage of the data as ↗RDF, makes the data in Beta maṣāḥəft *Interoperable* because it uses a 'formal, accessible, shared, and broadly applicable language for knowledge representation'[39] which allows *Interoperability* as 'the ability of data or tools from non-cooperating resources to integrate or work together with minimal effort.'[40]

misconceptions about TEI are based on the fact that the ODD customization mechanisms are not known, on which see Cummings 2018 and Burnard 2019, §§ 7–8.

37 Siegel and Retter 2015, <http://exist-db.org/>.

38 On ↗ODD see also Flanders and Jannidis 2016, 231 and the recent introduction of Pure ODD with the release of ↗TEI 3.0.0, for which see <https://wiki.tei-c.org/index.php/ODD>.

39 Wilkinson et al. 2016, 4.

40 Wilkinson et al. 2016, 2.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

```xml
<msItem xml:id="ms_i1">
  <locus from="14r" to="59r"/>
  <title type="complete" ref="LIT2246salotz"/>
  <textLang mainLang="gez"/>
  <incipit xml:lang="gez">
    <hi rend="rubric">
      በስመ፡ አብ፡ ወወልድ፡ ወመንፈስ፡ ቅዱስ፡ አሐዱ፡ አምላክ፡ ዛቲ፡
    </hi>
    ጸሎት፡ ዘእግዝእትነ፡
    <hi rend="rubric">ማርያም፡</hi>
    እሙ፡ ለብርሃን፡ ወላዲተ፡ አምላክ፡ ቅድስ
    <hi rend="rubric">
      ት፡ ድንግል፡ ንጽሕት፡ ዘጸሐፋ፡ አብሮኮሮስ፡ ረድእ፡
    </hi>
    ዮሐንስ። ጸሎተ፡ ወበረከተ፡ ወምሕረተ፡ ፍቁር፡ ወልዳ፡ የሃሉ፡ ምስለ፡ ገብራ፡
    <hi rend="rubric">
      <persName ref="PRS11744HabtaM">ሀብተ፡ ማርያም፡</persName>
      ለዓለመ፡
    </hi>
    ዓለም፡ አሜን። ጸሎት፡ ዘ
    <hi rend="rubric">ጸ</hi>
    ለ
    <hi rend="rubric">ዮ</hi>
    ት፡
    <hi rend="rubric">በ</hi>
    ቲ፡ እሙ፡
    <hi rend="rubric">ጿ</hi>
    ወ
    <hi rend="rubric">ፅ</hi>
    ለወርኀ፡ ሰኔ፡ በደብረ፡ ጎልጎታ፡ ዘውእቱ፡ መቃብረ፡ እግዚእነ፡ ወአምላክነ፡ ወመድኃኒነ፡ ኢየሱስ፡ ክርስቶስ፡ ሎቱ፡ ስብሐት፡ ለዓለመ፡ ዓለም፡ አሜን።
  </incipit>
</msItem>
<msItem xml:id="ms_i2">
  <locus from="101v" to="107r"/>
  <title type="complete" ref="LIT1113Anqasa"/>
  <textLang mainLang="gez"/>
</msItem>
<msItem xml:id="ms_i3">
  <locus from="107v" to="129v"/>
  <title type="complete" ref="LIT4663Eqabanni"/>
  <textLang mainLang="gez"/>
  <msItem xml:id="ms_i3.1">
    <locus from="107v" to="123r"/>
    <title type="complete" ref="LIT4663Eqabanni#Eqabanni"/>
    <textLang mainLang="gez"/>
  </msItem>
  <msItem xml:id="ms_i3.2">
    <locus from="123r" to="129v"/>
    <!-- check 129v -->
    <title type="complete" ref="LIT4663Eqabanni#Franks"/>
    <textLang mainLang="gez"/>
  </msItem>
</msItem>
```

Fig. 0.1 Example of a part of the ↗XML ↗TEI description of Frankfurt, Universitätsbibliothek Johann Christian Senckenberg, Ms. or. 17 (FSUor17) in Beta maṣāḥǝft. See <https://betamasaheft.eu/manuscripts/FSUor17.xml>.

Reusability in the ↗FAIR sense of it is achieved with a clear license stated everywhere, together with provenance and full attribution, as well as exposure of the data via community agreed and supported standard ↗APIs like ↗IIIF (International Image Interoperability Framework) and ↗DTS (Distributed Text Services). The project supports permanent ↗URLs (permalinks) to all revisions of each single file, which makes each version citable, either simply importing the embedded metadata relevant to each version or using the suggested citation.[41]

41 On the permanent links offered by the project, see the documentation at <https://betamasaheft.eu/pid.html>. The structure of these permanent links is based on the observations made in Arnold and Müller 2017. The project can offer this thanks to the fact that ↗GitHub does.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

The choice of ↗TEI was, at the moment of initiating the project, near to obvious, not just because this was the path indicated already more than two decades before and which had passed the proof of time, but because TEI has everything a project with these aims needs. Many projects existed already,[42] which used TEI to achieve similar goals, and tools, which could provide support for TEI, were already available, having been developed in diverse contexts and for different purposes. The ↗schema specification for the project was initially created starting from the Ethio-SPaRe data jointly with the process of ↗up-conversion of that data from the ↗MyCoRe ↗XML to TEI, making decisions for the techniques to use in encoding based on those already made by previous projects,[43] especially e-codices,[44] and FIHRIST.[45]

The ↗XML data is presented via ↗eXist-db and especially the ↗RESTXQ technology which allows one to very easily build ↗APIs, using ↗XQuery, thus making the information accessible in multiple different output formats for all the different present and future needs.[46] From ↗TEI encoded manuscript descriptions, for example, also ↗IIIF (International Image Interoperability Framework) ↗manifests are produced which are connected to the images stored on the local server.[47]

The database supports also several ↗RDF (Resource Description Framework) and ↗LOD (Linked Open Data) formats (Ch. 7), like ↗DTS (Distributed Text Services),[48] the ↗Pelagios Gazetteer Interconnection format and Pelagios ↗Annotations.[49] PDFs following in part the guidelines of *Aethiopica: International Journal of Ethiopian and Eritrean Studies* are also provided as a

42 Just to mention a few among those which have been more directly taken as examples, see e-codices (<http://www.e-codices.unifr.ch/>), FIHRIST (<https://www.fihrist.org.uk/>), EAGLE (<https://www.eagle-network.eu/>) which uses the ↗EpiDoc guidelines and ↗schema (<http://www.stoa.org/epidoc/gl/latest/>, Elliott et al. 2007) and the already mentioned Syriaca.org.

43 Reule 2018. The Beta maṣāḥǝft Guidelines (<https://betamasaheft.eu/Guidelines/>), contain a full description, integrated with the Beta maṣāḥǝft ↗schema, of the encoding decisions of the project. These guidelines are themselves a ↗TEI project maintained in ↗GitHub and powered by ↗eXist-db. See Liuzzo et al. 2018.

44 <https://www.e-codices.unifr.ch/it>.

45 <https://www.fihrist.org.uk/>.

46 Retter 2012.

47 See p. 16.

48 Almas et al. 2018.

49 <https://github.com/pelagios/pelagios-cookbook/wiki/Pelagios-Gazetteer-Interconnection-Format>, see now also the Linked Pasts Interconnection Format, <http://whgazetteer.org/2018/09/11/lp-format/>. See also Liuzzo 2017 for a general introduction to some encoding practices and Solomon Gebreyes and Liuzzo 2018.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

standard feature,[50] although not yet fully developed, and can be downloaded for every resource.

This was set from the very beginning of the project following ideas which Kropp already stated,

> But furthermore, let us not forget that machine readable data is extremely flexible and can be potentially linked and related to each other under varying sets of criteria freely formulated for changing problems. That means that human work of mental association, induction, is greatly facilitated and supported by immediate full record from the material, and, in consequence, new configurations and associations are created.[51]

All the project data (the ↗TEI for every resource, as well as the code that powers the applications) is maintained in a public ↗GitHub organization with several repositories where anyone can use and contribute to the data.[52]

The fact that the data is in ↗GitHub and is version-controlled from day one, not only is practical, but it also implies that the data is maintained independently from any ↗application, is not only software independent, and therefore the only things which should always be shipped with the data are the ↗schema and the guidelines. The schema, a file describing the content of the ↗XML files, is linked from each file and, in the schema, the guidelines are linked in several ways. On the other side the guidelines refer to and include the schema. Even where internet connection is not always available, the data can be used, validated to a local copy of the schema, and updated, then synced at the first occasion, being also very light weight. As time passes and software becomes obsolete, the data should always be available and reusable for new applications. This is a key for sustainability in the long term. The Beta maṣāḥǝft data is not tied to a specific software or interface and the website is just one possible presentation.

Some of the ↗GitHub repositories, the ones with the source data, are synced to our virtual server in Hamburg, where our ↗eXist-db instance with the Beta maṣāḥǝft ↗application is hosted. This allows users who enter data to benefit from a fully synced research environment and thus to see always the latest changes and updates to the data.[53] We can keep track of each other's progress and immediately visualize our work and navigate between entities,

50 <https://journals.sub.uni-hamburg.de/aethiopica>.

51 Kropp 1994, 132.

52 <https://github.com/BetaMasaheft>.

53 See Van Zundert 2016, 90–93 for a discussion of the eLaborate project and of some concrete examples of trading zones and linguistic negotiation between computer scientists and humanists. For a general introduction on collaborative projects and infrastructures, see Edmond 2016 and McCarty 2016.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

and not just for what concerns the team's work, but for any contribution, by anyone.[54]

Apart from significantly facilitating research, this workflow also means having a preview of the experience of end users, which can therefore be considered from the very first project stages, as the web ↗application is developed. Beta maṣāḥəft is not the only digital project of the centre and it is developed to integrate with other resources as the online *Lexicon linguae aethiopicae* and the morphological ↗annotations of the TraCES project (Fig. 0.2).[55]

The Beta maṣāḥəft research environment, which should manage the entry, presentation, and analysis of data related to the manuscript tradition of the Ethiopian and Eritrean highlands, is thus not one box or software providing all the required functionalities, but makes use of existing high quality and flexible resources like ↗GitHub and ↗eXist-db as well as well-established standards such as ↗TEI which are openly available, to empower a real and not a virtual research environment, based on digital resources. Instead of trying to be the one project to rule them all, Beta maṣāḥəft curates data which every project can use, hoping that this approach will support a cooperative culture of exchange and collaboration where each contribution to the common good impacts positively the entire community.[56]

In Beta maṣāḥəft, the flexibility and expandability of ↗XML are what allows us to build a research environment as a 'research ecosystem'. You can enter what you want to collect, marking up what is relevant for your own research. You can use the software you want to enter data and to use the data. You can tell us how you want it to look like in the applications we maintain. You can take the data and make it look like what you want in your own web or local ↗application.

A research environment and ecosystem means also taking full advantage of encoding practice as a research methodology which allows for multiple interpretations to coexist. Since any tag and any ↗attribute are interpretations and result of an inquiry which has led to a decision, there is room in the encoding for all possible and conflicting interpretations. It is possible to markup a role name with a type but also to add an element to say that another encoder thinks it should have not been marked up or that it should have been marked up differently. If the workflow tools help to reach an agreement for the sake of simplicity, that is welcome and possible, but where no agreement is possible, there is nothing to prevent the coexistence of several interpretations.

54 The example for this workflow has been taken from projects such as <http://syriaca.org> and <http://papyri.info>.

55 Finlay McCourt developed a working editor using ↗Python. This is linked to the organization repository and is still available (<https://github.com/BetaMasaheft/Editor>).

56 See Ch. 9.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Beta maṣāḥəft also has its own applications to access the data, and these are documented in many respects, so that there might be more chances of continuity in the future.[57]

The ↗XML data in ↗TEI allows us to

1) elaborate on the information structure and the contents;
2) compute frequency and relations between sub-strings of structured information;
3) structure and preserve meaning-carrying structures;
4) add several layers of meaning to a single source text (named entities, apparatus, citations, etc.);
5) control the quality of the core information;
6) check (validate) data and verify results with the highest possible precision;
7) keep updating consistently, while freely changing output methods;
8) let others reuse the data or contribute to it.

Beta maṣāḥəft does not have a data entry form,[58] collaborators write ↗XML using an editor of their choice, mostly ↗Oxygen XML Editor or ↗Atom. This leaves them complete freedom on what and how to encode within the boundaries of the ↗schema and guidelines. If the ↗ODD or the guidelines, which are also in XML, need to be updated, the team will discuss that and go ahead with the changes. This implies empowerment in the encoding process. Nothing is impossible because there is 'no box to put it in' and discussion is encouraged and leads to developments in the semantics of the data, instead of a boring conversation on the features of a form and where to best have a button. The team of collaborators is fully aware of what is happening and becomes aware of the importance of separation of concerns, as it concentrates on those about the contents, not those about how this is displayed. Compared to a workflow set up where people have to learn how an editor works, the learning curve might be a bit scarier, but it is not steeper, and it leads to benefits in the long term also for sustainability by enlarging the number of people who are actu-

57 Full version control via ↗GitHub for data and code (plus additional backups and a separate backup process for images) are in place. The project specific markup practices are fully documented from the ↗schema and explanatory guidelines, see Liuzzo et al. 2018. All functions are documented thanks to ↗eXist-db documentation app, which allows to expose details of all imported modules and modules specific to the ↗application, <https://betamasaheft.eu:8080/exist/apps/fundocs/index.html>. Data ↗API (the way data is exposed from the back-end) are also documented in a special page on the website, <https://betamasaheft.eu/api>. All the discussion about encoding, practice, schema, visualizations is archived in Issues on GitHub, <https://github.com/BetaMasaheft/Documentation/issues>, which are all publicly available and open.

58 There is a small form to create new records in the web ↗application, but this is only used in defined circumstances.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

ally aware of the data. Finally awareness of the data allows the team members to know what can be used to answer their questions, a fundamental step towards being able to make use of any digital technique.[59]

There are certainly many ways and many software and technologies which can be used to achieve the same results which are presented in the Beta maṣāḥǝft ↗application and in this book; some might certainly get better or more refined results, but it is not my aim to evaluate technologies or to compare methods. Instead the use of one set of technologies will hopefully facilitate the presentation of each chapter. This is done also with the intent of attracting contributions and users to the common effort of achieving a common base of information which can be used by anyone, and some of the desiderata listed above. Using few common standards and technologies should also make a first access to digital methodologies smoother, without discouraging the further exploration of the available possibilities.

Fig. 0.3 shows how the different entities (manuscript records, work records, etc.) are used to produce the output desired. No single view (e.g. a web page or the PDF) actually only reproduces the content of the ↗TEI file describing one entity, they all use the database to fetch information from related entities in some respect.

The ↗schema which controls the encoding of all files and serves also data entry facilitation purposes is maintained by the project and is part of continuous work.[60]

Now that I have outlined the main principle driving the Beta maṣāḥǝft project, let me briefly outline the reference concept of 'written artefacts', before moving on to outline the contents of the present book in the following section.

4 Written artefacts

The first two chapters (Ch. 1 and Ch. 2) deal with written artefacts,[61] and respectively manuscripts and inscriptions. This concept, borrowed from the

59 Bodard and Stoyanova 2016; Dee et al. 2016. On the benefits of learning to programme for humanists, Montfort 2016.

60 The Beta maṣāḥǝft ↗schema is available here <https://github.com/BetaMasaheft/Schema> and is part of the project guidelines (<https://betamasaheft.eu/Guidelines/?id=schemaView>).

61 This is an all inclusive denomination often used in the context of the activities of the SFB950 Manuskriptkulturen in Asien, Afrika und Europa and which turns out to be very useful and productive when discussing digital methods and their wider reach compared to both traditional disciplines (epigraphy, codicology, papyrology, etc.) and geographical areas of research. Many other general denominations have been used, like 'cultural heritage objects', for example, in the context of Europeana projects, which is

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

newly founded Cluster of Excellence at Universität Hamburg, 'Understanding Written Artefacts',[62] requires a little explanation before continuing in the description of the structure of the book.

The simple fact that the digital description of a manuscript or an inscription finds nicely its place in a description using ↗TEI means that we can encode both manuscripts and inscriptions specifically and according to their features but have them available in the same format to use their contents and description together, availing ourselves of the common encoding. There is no novelty in this. I would say that it is part of the aim of encoding in TEI itself. However, in my experience of teaching at trainings, this creates concerns that are neither few nor little for the specialists. Spending a lot of time to define an object of research for any kind of project, and then seeing it thrown into a larger context might turn out to be at least frustrating.

Between the all inclusive and interchanging ideas of linked data on one side, and the long struggle for the definition of an object of research which a researcher in the humanities faces, there is a probably unavoidable clash, and one which cannot be underestimated and thus requires some explanation. It is not self-evident that making data interchangeable and interoperable and linking it, not to mention making it openly available, is beneficial to research. For the data scientist these are prerequisites, but humanities researchers may have genuine concerns for their research data, which may be the outcome of years and years of work. It is a hard job to persuade either of the reasons of the other, and probably an unnecessary waste of time in most cases. The effort in this direction has been made especially by those interested in the data because it is for them a starting need, but it is also arguably because of the paradigm-shifting potential of the Web. New concepts and objects of research like 'written artefacts' come out of this productive process and provide new definitions which redefine the scope of research.

'dependent on the practices of libraries in terms of digitisation and cataloguing'<https://pro.europeana.eu/page/europeana-libraries-edm>; Sahle 2016 uses 'cultural artefacts' 'such as texts, images or physical objects. Usually they are kept in libraries, archives and museums and are thus not encountered as original material objects; rather, scholars work with surrogates of them created especially to make them more accessible and to facilitate research'; the ↗CIDOC-CRM has a ↗class called `E22_Man-Made_Object` (<http://www.cidoc-crm.org/Entity/e22-man-made-object/version-6.2>) which 'comprises physical objects purposely created by human activity'. The Handschriftenportal project uses instead the concept of '*Kulturobjekt*'. A fuller review of all this overarching concepts, which are taking a more and more relevant role in research, would make an interesting contribution. See also p. 93.

62 <https://www.written-artefacts.uni-hamburg.de/>. It is to be noted however that the cluster focuses on other research methods and explicitly not on those related to digital humanities.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

I would like to take the example of the definition of inscription and the way in which the argument is made,[63] to show how this process of redefinition is in continuity with the long-term reflection among specialists.

Silvio Panciera,[64] who gave the initial impulse to the digital epigraphy projects of the EAGLE consortium,[65] made the most recent attempt to define an inscription. This definition shows already the signs of a digital paradigm shift. Let me review the definition's principles he lists in his article and how each of them could be served by an encoded dataset which uses a wider concept as defining entity for its contents.[66]

The first observation of Panciera is that 'the words have taken their meaning from the discipline, according to its evolution, rather than the reverse, even if the words were introduced by the discipline itself.'[67] So scholars have defined an inscription and applied their definition to select their materials for each work. Other scholars might have referred to previous definitions and perhaps expanded or narrowed them. It would be probably interesting to be able to reapply each of the defining criteria used by each scholar though time for each publication, to see, for example, what would the results look like with a different selection. In a well defined dataset of written artefacts, where a very wide set of entities is available to which as little as possible of the defining criteria has been applied at the data entry point, one could select those entities which fit in a specific type of selection, for example simply filtering the data from the bibliography related to the discipline to take all those entities which are discussed in their publications by a given group of authors following a specific definition. Needless to say to do such thing one would need to know

63 Definitions of manuscript abound. An organized list with an attempt to a synthesis can be found in Lorusso et al. 2015. Unfortunately this paper does not comment each definition given. See also the comment to this paper made by Marilena Maniaci, <https://www.manuscript-cultures.uni-hamburg.de/papers.html#definition>.

64 Panciera 2012. The same journal published more recently also Grossi 2016, who proposes a new definition of inscription. However, this new attempt does not contribute much towards a definition, in fact, in my opinion, constitutes a regression from the steps made by Panciera and is very far from the experience which emerges from the pages of the article discussed here. Interesting are the collected references to inscriptions in papyri, which deserve more attention.

65 The Electronic Archive of Greek and Latin Epigraphy, <http://www.eagle-eagle.it/>, included in but not to be confused with the Europeana Network for Ancient Greek and Latin Epigraphy, <https://www.eagle-network.eu/>, coordinated by Silvia Orlandi, see Orlandi et al. 2014.

66 As there are indeed many. All ↗aggregators, like Europeana (<https://www.europeana.eu>) or the Text Grid Repository (<http://www.textgridrep.de>) collect materials which, from a traditional research perspective, could be said heterogeneous.

67 Panciera 2012, 2.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

much more about each publication and its relation to the entities: knowing that an article contains a reference to a document would not suffice, and one would need to know how is that relation qualified. Even more interesting would be to store each of the definitions proposed by scholars as parts of queries to the data to be able to restrict searches according to the principles of one or another scholar. In this way the meaning given to 'inscription', for example, by different scholars would be a selection principle rather than a defining one. Defining reusable queries would need a standardized data format and a database of definitions, linked to proper bibliographical records with stable URIs. This would be a useful tool for the future, but since both the authoritative bibliography and such database are not there for the moment,[68] the best and fastest available summary to access these developing meanings, at least for the concept of inscription, are publications like the one we are discussing.

Panciera initially groups the defining principles which have been used by scholars into 'material' and 'immaterial' principles to distinguish an inscriptions from other 'products of writing'. The concept of 'products of writing' is quite close to that of 'written artefacts' which interestingly emerges at the beginning of this article, and is a possible sign of a perspective which has already broadened its horizon. The concept of inscription needs in fact to be defined within a wider set which is the one of the 'products of writing/written artefacts', whose definition is not discussed in the article but is assumed as the necessary context within which to move.

The first of the material principles is that of 'durable materials'. Panciera criticizes this *criterium*. In an inclusive dataset of written artefacts one could filter its data selecting specific types of material. This implies that, if the researcher applying this definition thinks of parchment or papyrus as a durable material, he might include that in a corpus of 'inscriptions', or he could base his selection on a defined list of 'durability' parameters. There are stones of which we know, which did not last as long as the manuscripts carrying their descriptions, like *RIÉ* 276 and 277.[69] Given enough precision in the encoding of the data a researcher willing to apply this principle and definition could define his corpus for investigation based on a precise list of support materials and possibly also distinguish between an actual and an attested support.

The writing technique is the second of the 'material' criteria discussed by Panciera for their use in defining an 'inscription'. It is in fact a very suggestive way of distinguishing inscriptions from manuscripts that of saying that the first have been written by subtracting material (e.g. with a chisel) and the second by adding it (e.g. inks). The problematic implications are similar to the

68 See Ch. 2 and Liuzzo 2018 on the state of the epigraphic data available to date.

69 See Ch. 4, p. 133.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

previous kind of definition and, similarly, using standard encoding and vocabularies a researcher persuaded by this way of defining his object of study could get all the written artefacts which have been written using a specific technique of writing and include or not, for example, painted inscriptions or inscribed palm-leaf manuscripts according to his selection of defined methods.[70]

According to Panciera, the 'immaterial' criteria try to isolate inscriptions from other 'products of writing' (or written artefacts) by considering what the intent of an inscription would be, for example, 'the desire to perpetuate memory, the intention to communicate something publicly, and the aim of self-representation.'[71] These are features which are not easy to encode, unless parameters and values are established to define each of these intents, for instance the intent of public communication could be decided on the basis of the specific place of origin of a text. This *criterium* would exclude coins, manuscripts, and papyri, and perhaps also most of the *instrumentum domesticum*. A parameter defining the 'quest for eternity' of a written artefact could be sought on textual basis and evaluated only for a given list of materials which, in combination with such patterns of word's occurrence, might indicate this intention.

Potentially, given basic information is provided in the data in one of many possible standard ways, one could filter the data and apply any given definition and present clearly his selection principles. Stating those principles according to a standard offers the benefit of reproducibility. Let us imagine, for example, that one wants to verify an hypothesis made by an author on the basis of a certain corpus, in his own dataset. It will be of the utmost importance to make sure that the data is selected in the same way and thus, if the standard used is the same, this can be reproduced reliably.

Reflecting on his own definition of 1998, Panciera brings into the discussion the principle of 'intentional deviation from what may be said to be "normal" writing in the context in which it was produced.'[72] This could also be computable in as far as the variables are given, but it would *require* that these objects are in a wider context, the one of the 'products of writing' mentioned above, and can be isolated from it. While one could not apply this principle in a database of inscriptions, one could select the written artefacts in a dataset and compute the most frequent supports and writing techniques to isolate the deviant and call that group 'inscriptions'. Of course this negative principle applied to a corpus which, because of its history and constitution, contains

70 On this specific aspect see Evangelisti 2017 and p. 67.

71 Panciera 2012, 3.

72 Panciera 2012, 8.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

mainly inscriptions, would return the opposite of the result desired and can be applied only on datasets which are already known to represent well the proportional distribution of the different typologies.[73]

Panciera moves on from this defining principle to stress the one of 'address to a collectivity' as decisive to distinguish an inscription. There is no way I can think of in which this information could be retrieved by formal parameters from a generic encoded description, also because, as Panciera says, a 'document' might *become* an inscription when someone decides to address it to a collectivity. It is also clear that this principle alone does not suffice, as there are so many other 'products of writing' which would not be considered an inscription although they are intended for a collectivity.

If in fact 'neither epigraphists nor any one else has yet produced a generally accepted definition of "inscription"'[74] we might let everyone duly reason about this and apply its own definition on the basis of a generally accepted encoding standard instead. The task of the tools and methods is not that of 'understanding the specificity of epigraphic writing and the difference between it and other forms of expression of written culture or for defining the proper tasks and the "status" of the epigraphist'[75] as much as it is to let specialists of different provenances have proper ways to clearly do that. Some scholars in linguistics might want to see the attestations of the language without distinguishing on the basis of the type of writing or support. On the other side, some historians might want to find written artefacts of any type to base their research upon.

Panciera shows in this article an important step in the development of a critical definition, and demonstrates the need for clarity in principles and flexibility in their application and selection, which digital data can provide when using shared open standards. This by no means closes the discussion on such definitions, as the issues in the ontological definition discussed at the end of Ch. 2 shows.[76]

The meaning of this is that from the same set of written artefacts, given any of these principle is set on the desired value, a researcher could circumscribe its dataset of interest according to the chosen criteria, and, despite different approaches, researchers can work together and share data usefully, additionally being able to easily test their results against alternative options, by simply

73 This kind of definition has also the limit of evidence, which might make our knowledge of the normality distorted in as far as, for example, we have many more stamps than papyri from the imperial Rome and the papyri would appear to be deviant. But this limit could be easily overcome by setting declaratively the default value.

74 Panciera 2012, 1.

75 Panciera 2012, 1.

76 See p. 67.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

applying different criteria. In this way, encoding in ↗TEI does not only mean using the TEI guidelines, it means to refer to the same standard definitions to allow for more specific ones to exist all side by side. It also means for these definitions to be clearly and precisely applied each time by researchers both when they produce research on that data and when they study the work of others and want to reproduce it. In an ideal world where the standards are actually used, of course.

Basing the encoding and collaboration on ↗TEI and using the concept of written artefact as pivotal for the architecture of the data are the two major assumptions in Beta maṣāḥəft and for most of the contents of this book. Given these two elements as the ones defining the data, I can now conclude this Introduction with a brief outline of the contents of the book and their organization.

5 Contents outline

Forty years ago it was useful as it is today to provide a short overview of the contents of a book, which is what follows in this section.

Each chapter presents some data resources with some simple but specific questions and gives some concrete examples of encoded data and code used to help towards answering these questions or visualize the data in ways that can move in that direction. Each of the examples is kept as simple as possible and should be reproducible by any reader. Also, each chapter aims at being an independent module, so that interested readers might jump there and skip all the rest as in a handbook. The examples should show how old and new questions can be asked, what methodologies can be applied and what results can be achieved with that methodology and tools, distinguishing technical possibilities from research practice.

Fig. 0.2 gives a complete summary of the resources curated at the Hiob Ludolf Centre for Ethiopian Studies in Hamburg and helps also to place in the right relation with one another the different parts of this book, so that this figure and the following Fig. 0.3 can be considered as maps of the book as well as of these resources. Blue squares are the data entry tools (except for the ↗TEI editor which is part of the ↗application for the online *Lexicon linguae aethiopicae*) and the cylinders are where the data is stored. The top blocks are the front end applications, while the green blocks are different service components and software.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

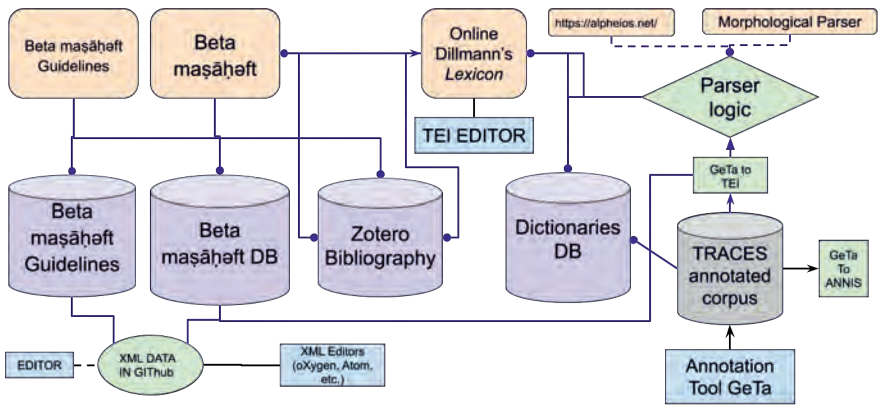


Fig. 0.2 Interactions of the Beta maṣāḥǝft data with other resources produced at the Hiob Ludolf Centre for Ethiopian Studies in Hamburg.

Ch. 1, Ch. 2, Ch. 3, and Ch. 4 deal with the data in Beta maṣāḥǝft, which in Fig. 0.2 is the second cylinder on the left side. The simple data architecture inside this block is represented in Fig. 0.3. Ch. 1 focuses on the upper part of what is represented in Fig. 0.3, written artefacts and the way they are modelled and fit into the architecture of Beta maṣāḥǝft, with a necessary narrowing of the scope to the most important type of written artefacts described in this phase of the project, namely manuscripts. Ch. 2 instead isolates inscriptions and uses the data to provide some analytical insights on the basis of comparison to other datasets.

Both manuscripts and inscriptions, as with any written artefact, carry content which can (but does not need to) be related with an abstract idea of a textual unit. Ch. 3 deals with these artefacts without their material support, and thus again with some of the data in Beta maṣāḥǝft, but, this time, the data in the lower part of Fig. 0.3, textual and narrative units,[77] as well as the central authority files, used for periodization and classification, and especially the ways in which textual fluidity is accounted for in the encoding and in the visualizations of the history of literature.

77 See definition on p. 79.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

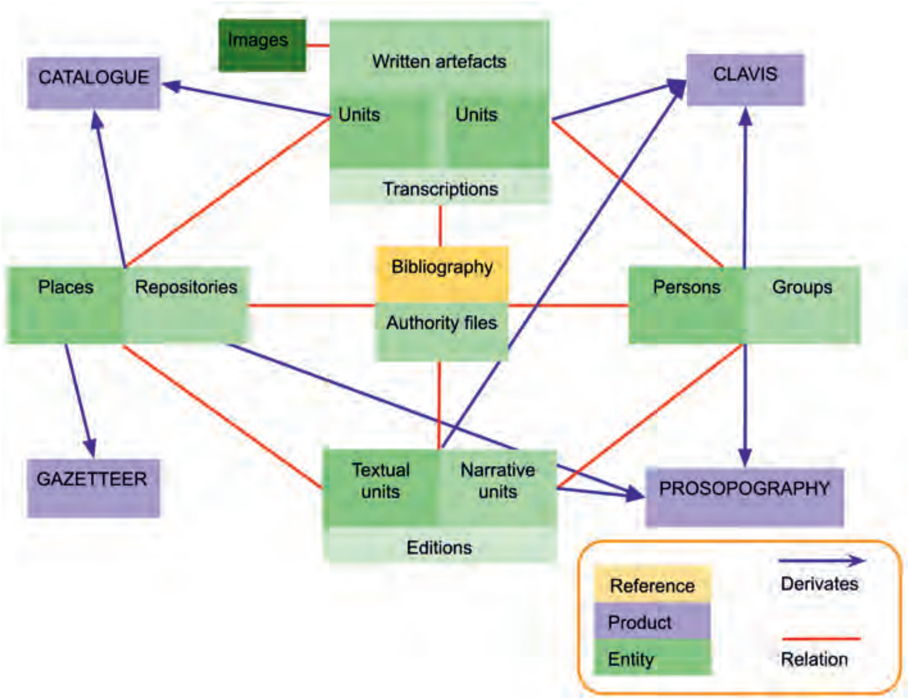


Fig. 0.3 Record types, their connections, and the outputs produced with each type of entity.

Identification and annotation of toponyms for the study of Historical Geography is the main topic of Ch. 4, considering data already seen in the previous chapters and focusing on the left side of Fig. 0.3, namely places. This chapter will take as examples Greek and Latin sources to highlight the great potential of encoding and Linked Data for this field of research. It will thus open up to more data than in Ch. 2 and see in more details some of the texts which are also seen from a distance in Ch. 3.

What has been missing up to this point is the data about persons on the right of Fig. 0.3, but in Ch. 7 some of the examples make use of this data as well. From this point on, we move a bit away from Beta maṣāḥəft.

The study of written artefacts and the abstract artefacts of human intellect they carry, written in Gəʿəz, cannot be done without the use of basic reference dictionaries. Among these the *Lexicon linguae aethiopicae* is the most important and its digitization will be described in Ch. 5, together with the functionalities provided by the web ↗application and its integration with oth-

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

er resources of the TraCES project. Here we will be moving towards the central part of Fig. 0.2, where the dictionaries database cylinder is placed.

However, having a good dictionary and a good text to study is not enough. Grammatical knowledge (*lato sensu*) of the language is needed and in Ch. 6 I will describe a possible way to build a morphological parser for Gəʿəz, which is represented in Fig. 0.2 at the right top corner.

From Ch. 7 I will be dealing with themes which cross each of the fields of interest described in the previous chapters, reaching out to other resources even more than in Ch. 2 and Ch. 3, beginning with the use of ↗RDF ↗triples and ontologies to represent the data in a way that provides several advantages for some tasks. This does not mean that you will not see this data, and queries to it, in the previous chapters: it only means that, at this point, you will get the full story about it.

To conclude the book by looking at the more pressing needs of its potential readers, who need to concretely publish research, often primarily in print, in Ch. 8 I will briefly deal with techniques which can be used to produce research output in print as PDF from ↗XML data, including the management of the bibliography using ↗Zotero. At this point the book should have covered most of the items in Fig. 0.2 which I have worked with.

The Conclusion in Ch. 9 tries to summarize the aim of this book and hints at key points of further development, and the most important ideas of collaboration and openness.

The book is completed by a Glossary, which contains short working definitions of many terms and frequently used acronyms, which should somehow help to disentangle the jargon.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Chapter 1. Manuscripts in Beta maṣāḥəft and out

καὶ πάλιν ἐκ τῶν ὀνομάτων καὶ ῥημάτων μέγα ἤδη τι καὶ καλὸν καὶ ὅλον συστήσομεν, ὥσπερ ἐκεῖ τὸ ζῷον τῇ γραφικῇ, ἐνταῦθα τὸν λόγον τῇ ὀνομαστικῇ ἢ ῥητορικῇ ἢ ἥτις ἐστὶν ἡ τέχνη
(And from nouns and verbs again we shall finally construct something great and fair and complete. Just as in our comparison we made the picture by the art of painting, so now we shall make language by the art of naming, or of rhetoric, or whatever it be)

(Plat. Crat. 425a)

Commenting on the Platonic dialogue where the passage in the epigraph occurs,[1] Sedley says that 'those words, it turns out, have not been attached in a merely arbitrary way to their objects, but are encoded *descriptions* of them.'[2] Encoding, as the act of describing other words with named tags, is also a τέχνη which allows one to make something 'fair and complete', if not great. I will try in this chapter to give some examples of the techniques used to make the artefacts produced with this τέχνη, into 'fair' visualizations.

Among the written artefacts described in Beta maṣāḥəft,[3] this chapter focuses only on manuscripts, and will call them thus throughout because also the web application, object of all the examples, does so. Although the chapter uses largely examples from Beta maṣāḥəft and its practices, there is no reason why, in the same way in which this project took ideas and methodologies from other projects, also others might find the way we have approached some issues to be relevant for their needs in manuscripts cataloguing. The description offered in this chapter is also intended to give the reader, who might not be familiar with concepts as that of 'separation of concerns', an overview of some encoding practices and the ways in which the encoded description is used to make visualizations, so that the reader without digital competences may find it easier to approach the following chapters knowing a little bit more about what the data looks like.

1 The text and the translation provided are those available in ↗Perseus.
2 Sedley 2018, § 4, emphasis in original.
3 Introduction p. xli. Eventually one could also easily argue that the encoded texts are as much a written artefact as a parchment manuscript or a printed book.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Beta maṣāḥəft is at the moment to my knowledge the only project doing XML encoding of manuscripts from Ethiopia and Eritrea, and it would be nice to see in the coming years more projects using this shared data for their own new projects or contributing to it for the benefit of the entire community.[4] This chapter would like to show how this interaction with the data can happen in terms of reuse of the data.[5] I will describe what can be done with encoded catalogue descriptions and give practical examples of some basic possibilities. After setting briefly the scope of a digital catalogue of Ethiopic manuscripts, in the first part I will show some examples of how the encoded ↗XML is used in our application, to show the separation of the encoding and the visualization concerns and how the second is based on the first as a separate task. The second part of the chapter will offer some example query on the described data accompanied by visualizations with tools entirely independent of Beta maṣāḥəft to demonstrate this added potential of openly accessible and reusable data. Finally I will discuss with an example using also data which is not in Beta maṣāḥəft what is actually possible in terms of interchange and interoperability.[6]

1 Digital catalogue of Ethiopic manuscripts

Only a fraction of the existing manuscripts in and from Ethiopia and Eritrea has been catalogued and is available to researchers. The quantity of still unknown documents is in itself a challenge.[7]

To accomplish the aims of the first phase of the project Beta maṣāḥəft, that is, the creation of a *Catalogus Manuscriptorum Aethiopicorum*, the work was based on two main input sources: existing historical catalogues and existing databases with descriptions of manuscripts.

4 This started to happen with various forms of collaboration, contribution to, and reuse of the data in Beta maṣāḥəft. Hopefully these practices will become more and more common.

5 To contribute to the data there are also several ways mentioned throughout this book. See <https://betamasaheft.eu/Guidelines/?id=howto>.

6 I hope in this way also to encourage reuse of the Beta maṣāḥəft data in any of the ways discussed and via any of the available access points to it, directly from ↗GitHub or from the channels offered by the ↗application like the data ↗API, the ↗DTS or ↗IIIF APIs, the ↗SPARQL endpoint, the linked data exposed via ↗VoID, the ↗RDFa shipped with the ↗HTML, one of the ↗TEI formats, and so on, most of which will be mentioned in this book.

7 See Bausi 2007, 89–92 for the quantification and explanation of the ground problems. The catalogues referenced in Beta maṣāḥəft can be found at <https://betamasaheft.eu/catalogues/list>, while the list of repositories holding manuscripts (without much precision for all the EMML and EMIP data) can be found here <https://betamasaheft.eu/institutions/list>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Core to the encoding and visualization of manuscripts in Beta maṣāḥəft are the five objectives put forward by Bausi for the census, preservation, and cataloguing of Ethiopic manuscripts,[8] although they are designed for printed catalogues. Let me recall them here briefly.

1) More space should be given to the physical description of manuscripts with attention to those features which are meaningful for comparison with other information.
2) Contents should be precisely identified and point to the essential repertories of biblical, apocryphal, hagiographic, and patristic literature, particularly the Arabic-Christian literature, and include all the specific bibliography or a reference to the *Encyclopaedia Aethiopica*.
3) Existing printed catalogues should be put online and new ones, whether online or in print, should have adequate space for textual data.
4) A catalogue of catalogues should be made, or at least a list of manuscripts, using adequate collaboration standards.
5) A bibliography for every manuscript, starting from the manuscripts in the historical European collections, should be available.

Beta maṣāḥəft started from the last two of these points and chose ↗TEI ↗XML to be able, among other things, to have both online and printed editions. The objective in the fourth point in the list is achieved primarily with several layers of attribution of each type of change which can occur, that is, by

1) the use of ↗GitHub for version control of each change to each file,
2) the use of <revisionDesc> for each meaningful change to the description to be properly attributed,
3) the use of <editionStmt> to detail the various roles and phases of elaboration of a given text before it made its way in the XML version.

We can thus accept and integrate by aligning to encoded description in ↗TEI, any sort of contribution, from the regular editorial work of the team to the occasional edit by a guest.

The differences in approach and choices of each source catalogue remain, but the information encoded in ↗TEI can be queried for all. With regard to the last point in the above list, each TEI description in Beta maṣāḥəft can have as much bibliography as needed (space not being a problem any more) as well as text quoted.[9] Additionally, the parallel task of assigning unique identifiers and creating records for each content of a manuscript being described and encoded means that the literary tradition is documented, starting from the man-

8 Bausi 2007, 108.

9 See the list of cited publications, which also lists for each title all entities where this is referred to, <https://betamasaheft.eu/bibliography>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

uscripts, in its entirety. Each of the main textual contents is recorded inside a <msItem> which is individually identified and can be linked directly to the identifier of a textual unit.[10] Since the publication of Bausi's article, however, a lot more has become possible and digital methodologies allow us to link each person named in a text, each date and toponym.[11]

Digital methodologies can offer also the possibility to directly compute the comparable data and make the implicit connections explicit, for example by gathering all manuscripts for a given literary work or comparing the structure of the contents.

2 Presenting Ethiopic manuscripts online

In this section I will describe some features of the online visualization of manuscript's catalogue records which are based on the ↗TEI descriptions. I will deal first with views which retrieve data from multiple records, like lists, and then with single manuscripts resources.

2.1 Groups of manuscripts, lists, and charts

The process of encoding manuscripts starting from historical catalogues results in an entirely new manuscript description for each item.[12] In Beta maṣāḥəft the features of the description are all checked against the photos whenever available and some will be checked on the manuscripts according to the project plan. However this does not mean at all that any given piece of information is perfect. On the contrary, it is in continuous evolution and is always in progress for many reasons: the lack of information or imprecision of the source, the connection between entities which determines changes in the view of any given item for each change, and so on. Therefore, the system needs to be capable of dealing with non-complete, partial, and even wrong data, to be flexible enough in its data structure and workflow to adapt to users of different kinds.

Each description encoded following the project guidelines becomes computable structured data which can be gathered and presented in many different ways,[13] instead of being only thought as a final product whose unique purpose is the delivery in one chosen format, often property of the author

10 See Ch. 3, p. 79.

11 The recognition of named entities is done in Beta maṣāḥəft without any assisted methodology, but only by researchers encoding the text. A named entities recognition service might be developed in the future, in relation with machine reading of manuscript images, as soon as the needed amount of training material will be available.

12 This has been described in Reule 2018.

13 Liuzzo et al. 2018, <https://betamasaheft.eu/Guidelines/>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

or editor. Fig. 1.1 shows in a list view a selection of manuscripts which have been marked with the keyword ‘Golden Gospel’, by adding a <term> in <keywords> with @key pointing to that value in the taxonomy.[14] In the list views we give the basic information needed at a glance: the name and shelf mark(s) of the manuscript (from <msIdentifier>); the number of parts (counting <msPart>s and <msFrag>) together with the information about the presence or not of a collation of the quires (encoded with <collation>); the number of content items (counting <msItem>); the number of hands (counting <handNote>); the script (distinct values of handNote/@script); and a series of available actions to get to the ↗XML source, print,[15] or compare with other manuscripts, as well as links to some of the views of the resource.

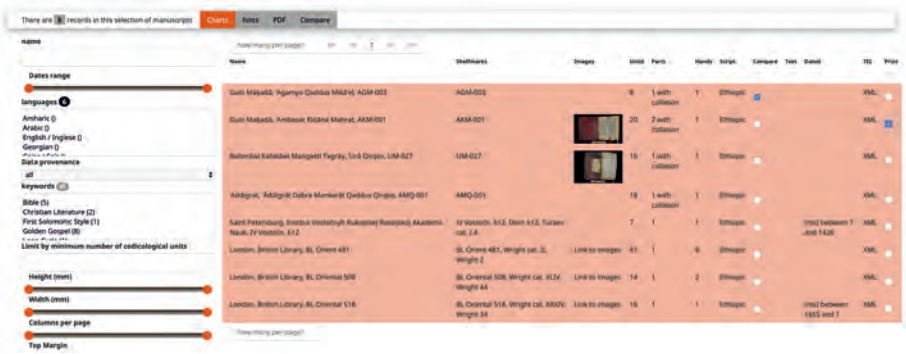


Fig. 1.1 List view of manuscripts, filtered to show only Golden Gospels. See <https://betamasaheft.eu/manuscripts/list?keyword=GoldenGospel>.

The list view for any given selection of manuscripts, by catalogue, repository, filtered by keyword, as in the example, or any of the other available parameters from the side bar, offers then a list of results which, because they are part of a group, might have a particular meaning to the researcher who has grouped them.[16]

14 <https://betamasaheft.eu/authority-files/GoldenGospel/main> is the main view about this keyword. The <taxonomy> is currently not included in the source files, as well as <prefixDef>.

15 The print function uses a script in all similar to that described in Ch. 8. Selecting more than one record, those will be printed together as one file.

16 Most of these examples are based on the Ethio-SPaRe data, which was converted from ↗MyCoRe to ↗TEI in 2016 and is since then in the process of being updated. This dataset offers the most up-to-date codicological descriptions, systematically including features which are often missing or scarcely present in other descriptions coming from different sources.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Already the faceted filters offer some information on the group as a whole, by telling the user, for example, that at a given repository (Fig. 1.2) there are a given number of manuscripts with the keyword ‘Canon Law’. In the case of repositories, the information about the monastery where the manuscripts are kept is also relevant to the manuscripts, and is therefore shown together with them and the filters.

We can show which scribes, donors, and patrons are linked to a repository or a selection of manuscripts, so that some studies on the collection can be done directly from this view and the filters it provides.[17]

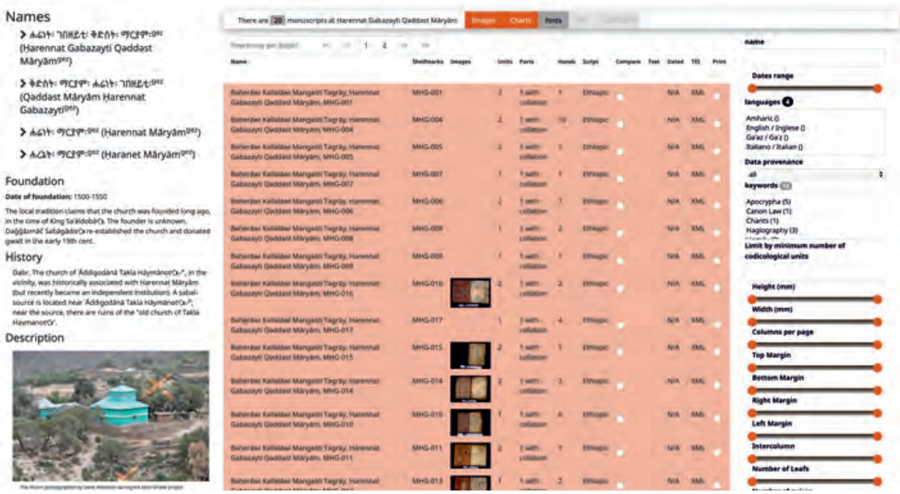


Fig. 1.2 Manuscripts at Ḥārennat Gabazayti Qəddəst Māryām, *waradā* Gāntā ʾAfašum, *ṭābiyā* Sāsun Bet Ḥawāryāt, *qušat* Sāsun (INS0144MHG). See <https://betamasaheft.eu/manuscripts/INS0144MHG/list>.

The data selected is also already offering the basis for relevant comparisons, and there are a number of representations which can be provided on the fly computing from the selection of results.

Fig. 1.3, Fig. 1.4 and Fig. 1.5 show some information from the ten manuscripts currently added to the collection by Eliana Dal Sasso starting from

17 Some repository’s descriptions include information which collides with the results of a search. For example, in Fig. 1.2 the description currently says twenty-two manuscripts photographed, but only twenty are listed. The reason for this is simply that it is true that twenty-two have been photographed but it is also true that only twenty of them have been described and are in the database, in this case as inheritance from the Ethio-SPaRe project.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

the information in the catalogue curated by Delio Vania Proverbio of the manuscripts at the Exarchic Greek Abbey of St Mary of Grottaferrata (INS0414Abbey_of_St_Mary_of_Grottaferrata) which has been extracted to produce tables which are sent to Google Charts, and returned from there as pie or column charts according to the requirements.[18]

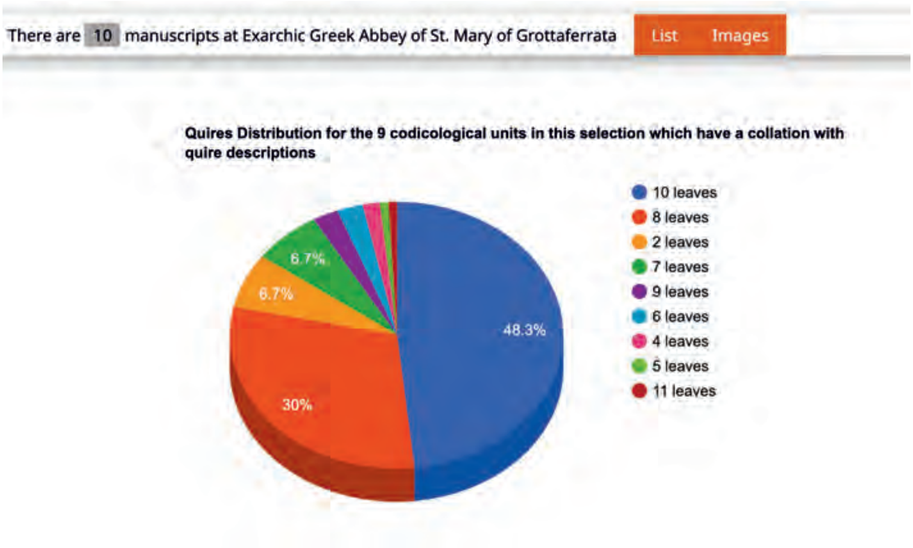


Fig. 1.3 Example pie chart with the quire distribution for the manuscripts from Exarchic Greek Abbey of St Mary of Grottaferrata. See <https://betamasaheft.eu/manuscripts/INS0414Abbey_of_St_Mary_of_Grottaferrata/listChart>.

18 <https://developers.google.com/chart/>. This is the same method used by <https://www.trismegistos.org/> and I would like to thank Mark Depauw who pointed me in this direction. The current view of the manuscripts from this repositories can be seen at <https://betamasaheft.eu/manuscripts/INS0414Abbey_of_St_Mary_of_Grottaferrata/list>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

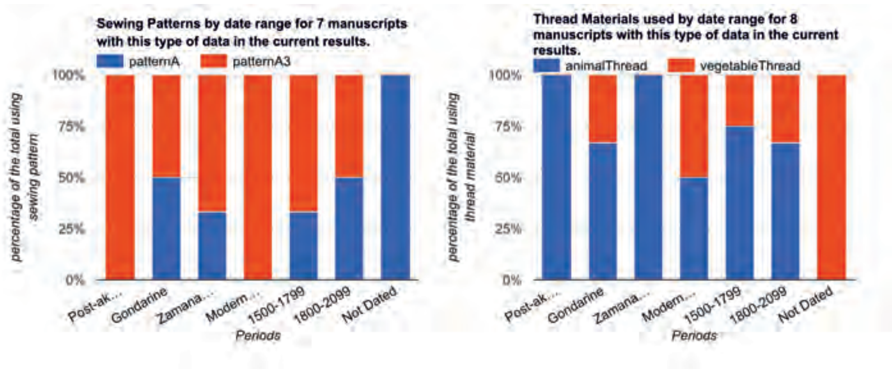


Fig. 1.4 Example charts related to binding and physical features in the manuscripts from Exarchic Greek Abbey of St Mary of Grottaferrata (part 1). See <https://betamasaheft.eu/manuscripts/INS0414Abbey_of_St_Mary_of_Grottaferrata/listChart>.

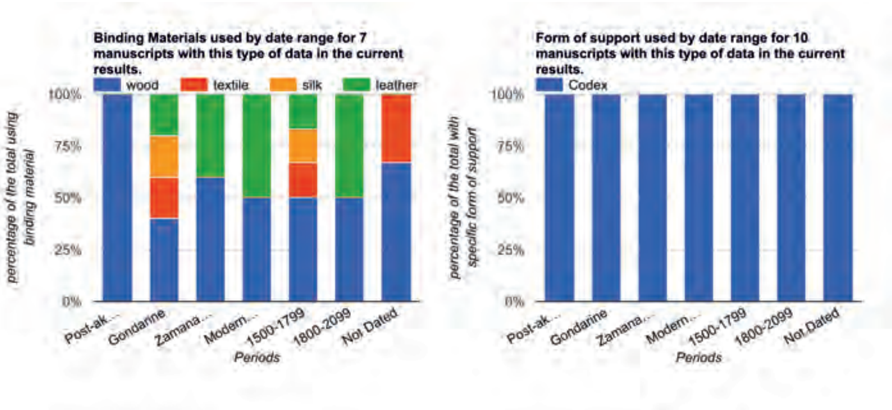


Fig. 1.5 Example charts related to binding and physical features in the manuscripts from Exarchic Greek Abbey of St Mary of Grottaferrata (part 2). See <https://betamasaheft.eu/manuscripts/INS0414Abbey_of_St_Mary_of_Grottaferrata/listChart>.

While the pie chart with the quire distribution is relatively easy to produce, given that each quire has a dimension and this can be simply counted (how many <item> contain a <dim> equal to 8?), the data for other charts is not as easy to produce.

In order to produce these charts we tried to impact as little as possible on the data entry process, and thus not to ask that cataloguers enter a specific type of information for this purpose. It seemed preferable to the project team that the cataloguers' effort would be put to record only what is strictly neces-

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

sary and as quickly as possible to get on with the enormous amount of work still to be done. Furthermore, the smaller the quantity of keystrokes the less opportunity there is for mistakes. Let me make an example of a chart which uses the information available.

The bar charts in the list views aim to show the distribution over different time periods of a given feature, for example the binding materials used.[19] This is encoded in the ↗TEI files using a typed <decoNote> in the <binding> element, where <material> is used to point to a value in the taxonomy which can be selected easily by looking at //t:decoNote[@type='bindingMaterial'][parent::t:binding]/t:material/@key. It is not certain that this information for a given selection of manuscripts is always given, so first, the manuscripts which have that information are selected. These manuscripts are then grouped by date (looking at <origDate>) and a column is produced for each of two types of periods: arbitrary periods of three centuries (e.g. 1500–1799) and meaningful periods taken from the canonical periodization of the project.[20] If a manuscript falls in both groups, it will be counted in both groups.[21] Then, for each of these groups, the values are computed and made into a percentile value so that they can be comparable regardless of the actual number of attestations of one piece of information.

A similar process is used to produce the chart which reproduces the one used by Maniaci,[22] and which shows with each line a given period (defined as above) on the x-axis ranges of sizes (height + width), and on the y-axis the percentage of manuscripts of that period with that size (Fig. 1.6 and Fig. 1.7). The measurements of height and width need to be given in the same unit of measure,[23] which is generally not a big problem except when the unit is not consistent with the data given, like a manuscript of 20 m.

While the data charted in Fig. 1.6 does not show interesting results for the limited number of items, Fig. 1.7 shows instead data for 183 manuscripts, which can be more significant. This is however selected on the basis of the

19 See <https://betamasaheft.eu/Guidelines/?id=bindingDescription> for a description of our encoding practice.

20 See the periods in ↗PeriodO, <http://n2t.net/ark:/99152/p03tcss4qvv>. On PeriodO see Rabinowitz forthcoming. In the <date> element we encode also dates relative to different calendars (<https://betamasaheft.eu/Guidelines/?id=date>). These ↗annotations are not yet available in the ↗RDF data.

21 The module is available with the code of the Beta maṣāḥǝft ↗application and is called charts.xqm, <https://github.com/BetaMasaheft/BetMas>.

22 Maniaci 2012, 486. The use of this chart was suggested by Eliana Dal Sasso.

23 This information is encoded in //extent/dimensions[@type='outer'] in our records, see <https://betamasaheft.eu/Guidelines/?q=dimensions&id=objectDescription>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

current repository, which might be not as relevant as grouping by area of provenance.[24]

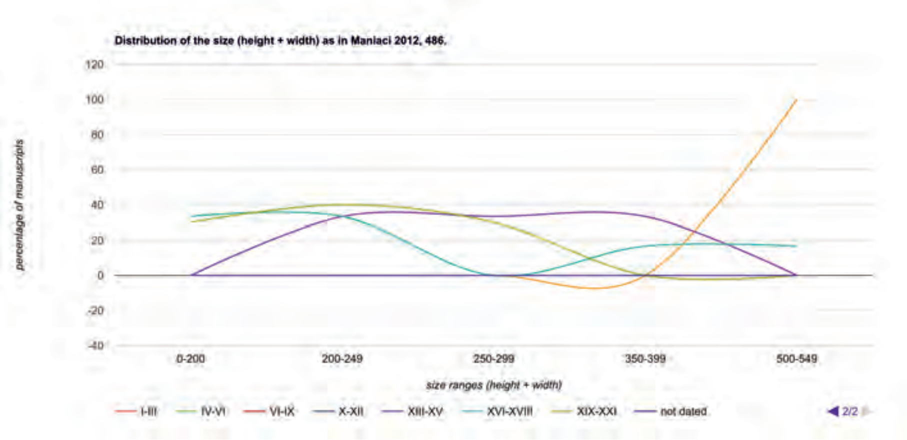


Fig. 1.6 Chart of the manuscript's tallies at the Monumento Nazionale Abbazia di Casamari. See <https://betamasaheft.eu/manuscripts/INS0452Mon_Naz_Casamari/list Chart>.

24 This can be done as well, pointing to a list based on a place, e.g. <https://betamasa heft.eu/manuscripts/place/listChart?place=LOC7175Sasi>. The link can be found on top of every place record and the resulting list will include any manuscript from a repository within that place.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

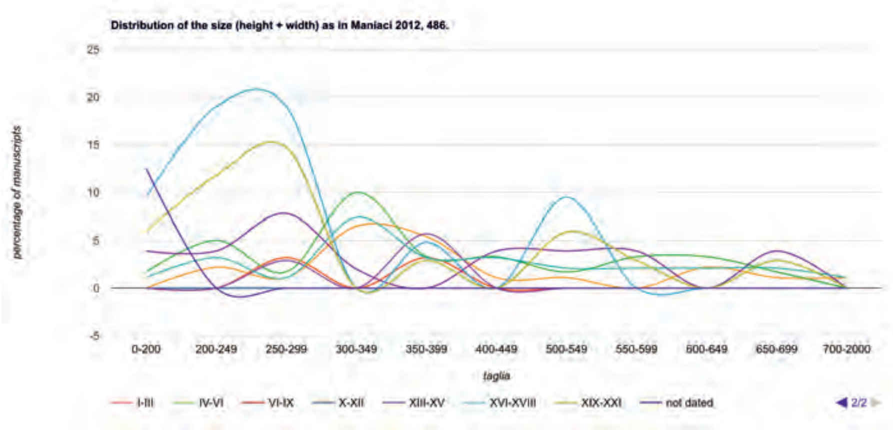


Fig. 1.7 Chart of the manuscript's tallies at the Bibliothèque nationale de France. See <https://betamasaheft.eu/manuscripts/INS0303BNF/listChart>.

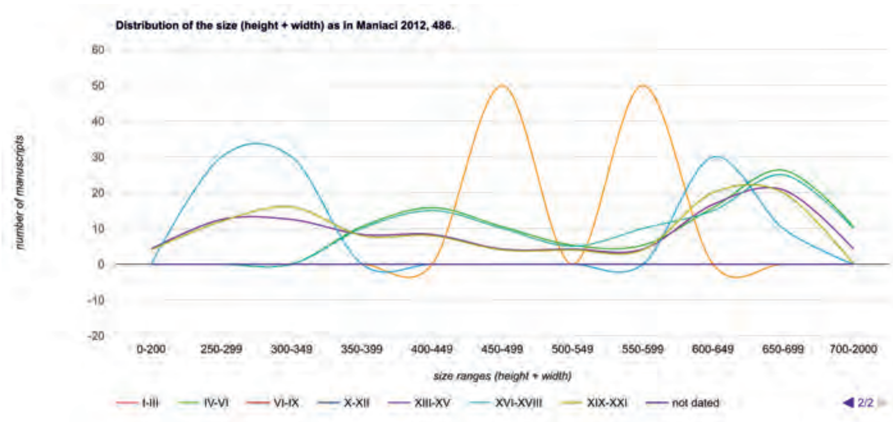


Fig. 1.8 Chart according to Maniaci 2012 of thirty-four manuscripts from the Ethio-SPaRe collection with the keyword 'Theology'. See <https://betamasaheft.eu/manuscripts/listChart?cp=ES&keyword=Theology>.

However, the chart looks more interesting if, for example, we take the thirty-four 'Theology' manuscripts in the Ethio-SPaRe collection as in Fig. 1.8. But I will let experts read it, or select the data they want and look at the automatically produced charts.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

In some manuscripts also information on ruling, pricking, and punctuation is recorded within an element <ab> using values for @type and @subtype which have been defined in the ↗schema:

```
<ab type="ruling" subtype="pattern"> Ruling pattern: 1A-1A/0-0/0-0/C. </ab>
<ab type="pricking"> Primary pricks are partly visible. </ab>
```

The scheme used for the ruling pattern is the one by Denis Muzerelle,[25] widely used in contemporary cataloguing. Although we actually do not need to enter formulas because of constraints of space, still formulas like this one help the cataloguer to get immediately a set of information which he can decode knowing Muzerelle's system. These formulas are preferable to a lengthy explanation, even where there is space for it, because they condense a large amount of information which would be otherwise tedious to describe as text and would inevitably be unclear. The ruling pattern formula already serves the end user very well. Additionally it is also easy to parse with ↗RegEx.[26] In this case (Fig. 1.9) while the comparison of the whole formula as string is already interesting as it gives an idea of the overall diversity, also more close-up comparisons of the single zones of the formula are useful.

25 <http://palaeographia.org/muzerelle/index.htm>. See also Andrist 2015b, 525.

26 It is also easily linked to the graphic representation nicely available at <http://palaeographia.org/muzerelle/grecs1.htm> with URLs which include the pattern, making of it a perfect candidate to become a useful ↗URI.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

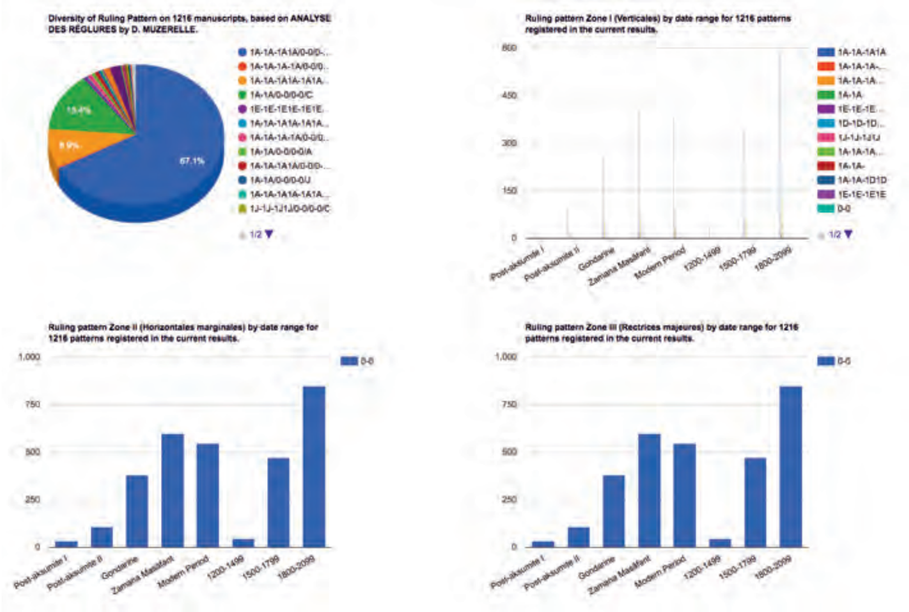


Fig. 1.9 Ruling patterns in Ethio-SPaRe manuscripts. See <https://betamasaheft.eu/manuscripts/listChart?cp=ES>.

The Ethio-SPaRe project had also recorded in the 1025 records available some information on pricking, which makes possible the distinction of four patterns, which can be dated,[27] but, at the moment, the encoding of the information has not been formalized enough to be able to reliably group the data according to these patterns.[28]

All these charts are available for any selection of manuscripts as stated previously. Also the comparison tool, which organizes the results in a slightly different way, prioritizing the contents instead of the physical features, is another list view and includes these charted data. Fig. 1.10 shows a comparison of the same manuscripts in Fig. 1.1. Fig. 1.11 shows instead a comparison of the manuscripts which contain a given work, which means they have a <msItem> with a <title> pointing with @ref to the identifier of a textual unit.[29] Actually, selecting the check-boxes on list views it is possible to compare in

27 Nosnitsin 2015.

28 Some of the decisive distinctions are based on the position of the first line of writing with respect to the main horizontal rulings and this information is simply given as text.

29 See <https://betamasaheft.eu/Guidelines/?id=manuscriptContents> and Ch. 3 here.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

this way arbitrarily a given set of manuscripts from a search result. Charts are provided for the manuscripts for which the information is available, declaring also the provenance of the information and the subset of the selection.

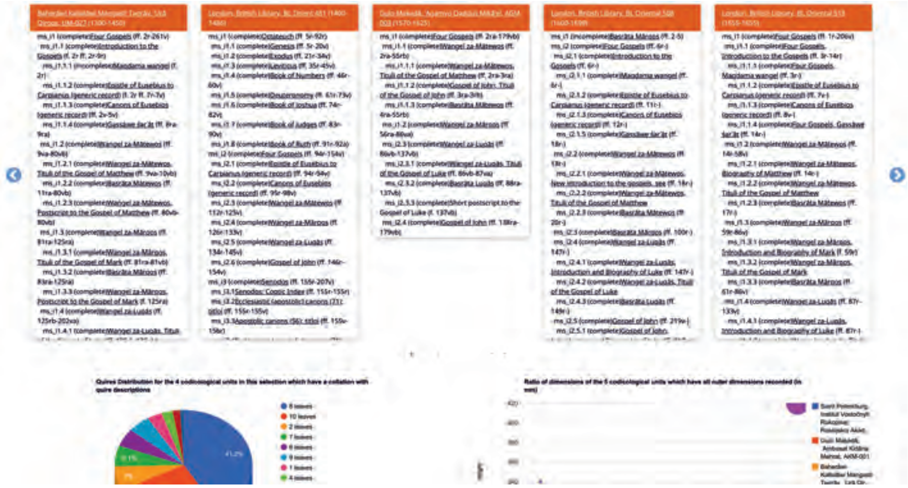


Fig. 1.10 Comparison visualization of manuscripts tagged as 'Golden Gospels'. See <https://betamasaheft.eu/compareSelected?mss=ESagm003,ESakm001,IVvostochn612,BLorient481,BLorient508,BLorient518>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

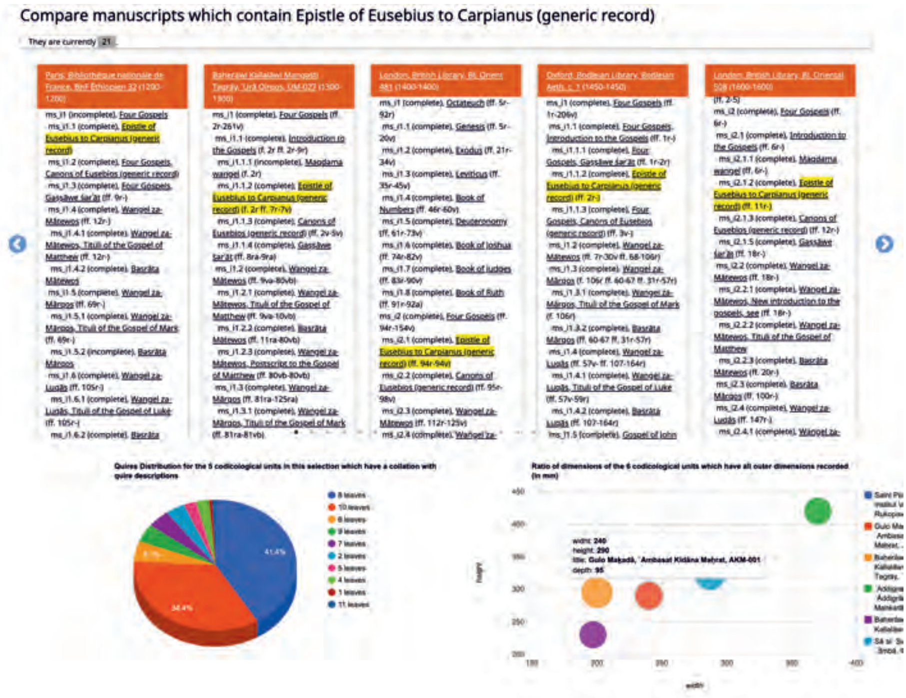


Fig. 1.11 Comparison of manuscripts containing the *Epistle of Eusebius to Carpianus* (CAe 1349, ID: LIT1349EpistlEusebius). See <https://betamasaheft.eu/compare?workid=LIT1349EpistlEusebius>.

The way in which these lists are structured leaves the specialist, which is not a codicologist in the first place, with a resource they have to navigate through. To serve other specialists, who want, for example, to search only the decorative elements, we offer dedicated search interfaces which give them direct access to their object of interest, without having to go through the manuscript's description organization.[30]

We can do this because of the data being stored and queried in ↗eXist-db, where we can, with a simple ↗XQuery, select only the <decoDesc> elements available in any description.

30 For decorations <https://betamasaheft.eu/decorations>, for additions <https://betamasaheft.eu/additions>, and for binding <https://betamasaheft.eu/bindings>. Also a list of cited publications is available, with links back to each of the files where a bibliographical reference is used, <https://betamasaheft.eu/bibliography>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

In addition to the list view and these thematic searches, Beta maṣāḥəft also offers a simple search on the homepage and an advanced search where several general filters can be set. Results are grouped by type and provided, if the search was for a string, with three matches with a ↗KWIC view.[31]

Now that I have described the main features of the list views and charts, I will continue to the single manuscript visualizations and their encoding.

2.2 The online catalogue record

The individual manuscript descriptions which are encoded in the current phase of the project come mainly from information contained in historical catalogues of Ethiopic manuscripts in European collections.[32] These catalogues have provided often the first access to texts which were otherwise unknown and were very generous with quotations and transcribed parts of texts,[33] while adding also bibliography which pertained to the contents of the manuscript. In a digital environment, it is easy and rewarding, as many projects have done and also Beta maṣāḥəft does,[34] to keep two distinct entities, that of written artefacts and that of textual units, as abstract works, and relate them accordingly.[35] This results in splitting the information provided by the catalogue in several places.

We can then present those new catalogue entries regardless of the way in which they were laid out in the original catalogue on which they are based. I will list only three features of this view: the way in which it is organized in general, the way in which the titles are produced, and the way in which images are linked.

2.2.1 Structure of the manuscript description

For each manuscript there might be several views, which represent in different ways (another example of multiple output from the same source input data) the data entered for the manuscript. For manuscripts there is the main entry view, the link to the images viewer or to an external image viewer, the ↗TEI export, a PDF print out, a relations view, and a syntax view.[36] Fig. 1.12 is the main view of a manuscript entry.

31 <https://betamasaheft.eu/as.html>.

32 A full list of the bibliographic entries used in the catalogue bibliography of manuscript records is available here <https://betamasaheft.eu/catalogues/list>.

33 Bausi 2007, 94–95.

34 e.g. <http://pinakes.irht.cnrs.fr/>.

35 See p. 79.

36 Each of these views is prompted by changing the last part of the ↗URL, either clicking the optional buttons on the top or by hand. `/main` is the main entry, and `/viewer` will

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

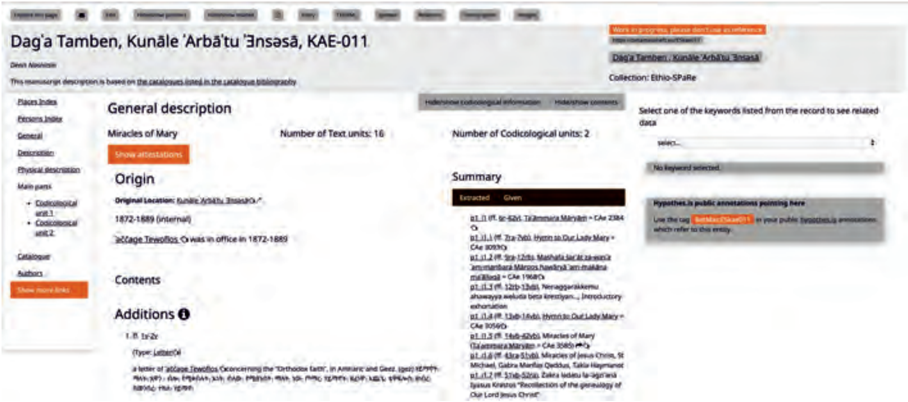


Fig. 1.12 Part of the main view of KAE-011 (ESkae011).

Using the classification proposed by Andrist in the COMSt handbook, our main view should be of the type B (reproduced in the following table) as it takes every codicological unit (equated in our guidelines to a <msPart>)[37] and prints physical description (<physDesc>) and contents separately.[38] This is easily done based on the ↗TEI encoding, which allows the manuscript description (<msDesc>) to have a <physDesc> and many <msPart>s or <msFrag>s (manuscript fragment), each of which can have its own <physDesc> and can be nested in other <msPart>s, which can therefore be limited to those aspects which are not in the general one. This is very useful because, given the latest circulation status, there will be features which we do not want to repeat. The Beta maṣāḥəft description is not originally structural, but a principle is applied according to which, as a result of the research carried out on the manuscripts, parts are encoded with <msPart> and correspond to production units (Unité de Production).[39]

show the image viewer, if available. Similarly, adding .xml after the identifier, or .pdf will prompt respectively the source file or a PDF rendering of the same. This is true for each version of a given file, which is always permanently accessible.

37 This praxis was inherited from e-codices, <https://www.e-codices.unifr.ch/it>.

38 Andrist 2015a, 517; see also Andrist 2014.

39 As defined in Andrist et al. 2013, 59–62. See Liuzzo forthcoming a for a full explanation of how the methodology for a structural description is integrated in the Beta maṣāḥəft workflow.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Table 1.1 **Skeleton descriptions of manuscripts according to Andrist 2015a, 517–518.**

Type B	Type C	Beta maṣāḥəft
Codex N (first part)	**Codex N**	**Codex N**
Content	Heading	Heading
Physical features	Content	Common features
etc.	1. … of stratum 1	Bibliographies
Codex N (second part)	2. … of stratum 2	**Codex N (first part)**
Content	etc.	Content
Physical features	Physical feature (a)	Specific features
etc.	1. … of stratum 1	etc.
Information about the grouping	2. … of stratum 2	**Codex N (second part)**
Bibliography	etc.	Content
	Physical feature (b)	Specific features
	1. … of stratum 1	etc.
	2. … of stratum 2	
	etc.	
	Common features (Binding, History…)	
	Bibliography	

Although this view aims at the model of the type B among those classified by Andrist,[40] still, it is a hybrid, because we also have a number for the different parts as in type C and because the division inherited from the ↗TEI affects also shelf marks, contents, and all other information which can be nested in one another as needed. A further virtue, in this respect, of the way in which TEI guides the encoding of a manuscript description is that the counting is not broken, as these parts, although independent, still belong with the manuscript being described. The layout in ↗HTML also allows us to show contents and features side by side, so that the specific features of a given part are directly on the side of the contents of that part.

Contents, but also other parts of the description, will include canonical names of resources, let them be places, people, textual units, or manuscripts. Because these are all results of a query, let me describe briefly how one such title is built.

2.2.2 Titles

Among all 'titles' or canonical labels, which are present in a manuscript description, the view of each entry reports at the top the full denomination. This title attribution is made by a script which runs each time a title is needed and

40 Andrist 2015a, 517.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

works slightly differently according to which type of entity is involved. In the case of a manuscript, this title is actually composed of information stored in several of the ↗XML files in the database. For example '*waradā* Dagʿa Tamben, *ṭābiyā* Salām, *qušat* Kunāle, Kunāle ʾArbāʿtu ʾƎnsəsā, KAE-011' is never entered by the cataloguer, but is instead composed of

1) The place in which the manuscript is at the time of the last recording: '*waradā* Dagʿa Tamben, *ṭābiyā* Salām, *qušat* Kunāle', which in this case reproduced the Ethiopian administrative divisions. This information comes in most cases from a place record linked with a @ref inside a <settlement> element inside the repository record. This means that the information is stored once for all the repositories which happen to be in that place.
2) The canonical name of the repository holding the manuscript: 'Kunāle ʾArbāʿtu ʾƎnsəsā'. This information comes from the repository record which is linked with a @ref inside the <repository> element of the manuscript record. This means that the information is stored once for all manuscripts in the same repository.
3) The authoritative shelf mark of the main unit: 'KAE-011'. This information is in the <msIdentifier> child of <msDesc>, inside <idno>.

The cataloguer, given that the repository record exists already and is correct, will only have to know the identifier of the repository and the shelf mark to complete the basic information necessary to print the full correct title for the manuscript in every place in the ↗application where this is needed. The <msIdentifier> of this manuscript looks like this:

```
1 <msIdentifier>
2   <repository ref="INS0234KAE"/>
3     <collection>Ethio-SPaRe</collection>
4     <idno facs="KAE/011/KAE-011_" n="66">KAE-011</idno>
5 </msIdentifier>
```

It can be seen that the cataloguer has only to enter the information inside the <idno> element (Line 4 in the example) and the @ref of <repository> (Line 2 in the example).

This condensed ↗XML and non-canonical ↗TEI which the project team edits is not very readable, and an external user of this code, for example a humanist with some digital competence who is interested in this data or a programmer who needs to connect the data in one of his projects to the one in Beta maṣāḥəft, would need to know quite a lot to figure out, for example, what the identifiers in the @ref attributes point to even with the ↗schema and

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

guidelines.[41] For this reason a different XML is offered for download, which makes the data a bit more explicit. The example above will be expanded as follows:[42]

```
<msIdentifier>
   <repository ref="https://betamasaheft.eu/INS0234KAE">
            Kunāle ʾArbāʿtu ʾƎnsəsā
   </repository>
   <collection>Ethio-SPaRe</collection>
   <idno
            facs="https://betamasaheft.eu/api/iiif/ESkae011/manifest"
            n="66">
            KAE-011
  </idno>
</msIdentifier>
```

Also, in this version of the ↗TEI which we try to make more conform, the name of the repository is added, and the @ref is expanded to contain the ↗URI for the repository.[43]

The additional information inside the ↗attributes @facs and @n of <idno> is used to produce the ↗IIIF ↗manifest which will allow the display of the images everywhere they are needed in the ↗application, providing the location of the images relative to the base IIIF image server and the total number of images available. Let me then move on to describe these images and the way they are linked in the ↗TEI.

2.2.3 Images

To expose and make accessible the data about the images of manuscripts which we use, we choose to adopt a further standard, ↗IIIF, which is a specification for the way in which the information to access images should be presented. We produce this presentation API for the information about images

41 On the importance of documentation and customization of TEI, see now.

42 This can be obtained clicking the TEI/XML button on each view or pointing the browser to <https://betamasaheft.eu/tei/ESkae011.xml>, while the source TEI can be seen either clicking on the edit button, or as raw data, <https://betamasaheft.eu/ESkae011.xml>. The identifiers are the actual file names and the latter request simply retrieves them from the file system, as well as the query for the corresponding ↗RDF, <https://betamasaheft.eu/rdf/manuscripts/ESkae011.rdf>, see also Ch. 7, p. 187.

43 See Ch. 7 for the further machine readable clarification added to the ↗HTML page in form of ↗RDFa and other machine operable formats provided.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

following this specification, which is shared and used by many other institutions, and benefit from it ourselves. To gather the information needed to present the images, we start from the ↗TEI, always with a script as in the previous example, but this time to produce a series of alternative full or partial representations according to the IIIF presentation API specification. We will say what images are available on the basis of the encoded data, and specifically the presence or not of an attribute @facs in <idno> as seen in the previous section, and according to the amount of encoded data in TEI we will present more or less information in the IIIF. So, if there are images, this will be said in the TEI and will trigger the possibility to request a IIIF presentation following the API specification. We are thus serving data in an additional international standard starting from the same initial data, adding with a layer of interoperability also a further layer of access.

In Fig. 1.13 the ↗Mirador viewer is used to show the information about the images organized in a ↗IIIF ↗manifest. The script producing this manifest from the ↗TEI-encoded data also produces ranges for any meaningful set of information,[44] in this case anything which has a <locus> element that nails it to a precise image,[45] so that in the viewer the user can navigate the images starting from the content and, for example, jump to a given quire or to the image corresponding to the beginning of a text.

44 The module is available with the code of the Beta maṣāḥǝft ↗application and is called iiif.xqm, <https://github.com/BetaMasaheft/BetMas>.

45 The images served by Beta maṣāḥǝft, mainly those of the Ethio-SPaRe project, EMIP, and the libraries with which we have an agreement to publish the images, like the Biblioteca Medicea Laurenziana (INS0339BML), are stored at Universität Hamburg thanks to the services and support offered by the Regionales Rechenzentrum. They are then served from the Beta maṣāḥǝft web server, where the disk is mounted, using the IIPI Image server <http://iipimage.sourceforge.net/> by Ruven Pillay.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Fig. 1.13 The ↗Mirador ↗IIIF viewer used to display the images of KAE-011 (ESkae011). See <https://betamasaheft.eu/manuscripts/ESkae011/viewer>.

Having images exposed online according to ↗IIIF involves another comparison possibility, across resources for images.[46] In the ↗Mirador viewer multiple manifests can be loaded and seen in parallel. In the following example I have loaded from the ↗manifest ↗URL also Vatican City, Biblioteca Apostolica Vaticana, Aeth. 24, provided via IIIF by the DigiVatLib.[47]

46 What can be done with well digitized images of manuscripts is out of the scope of the present book, but see for example Busch and Chandna 2017 on analysis of layout and measurements of manuscripts and the work being carried out by Daniel Stökl Ben Ezra and Hayim Lapin for <http://www.erabbinica.org> which is not yet available but I could see during a presentation at the Centre for the Study of Manuscript Cultures in 2018.

47 <https://digi.vatlib.it/>. Also the Bibliothèque nationale de France and some other institutions provide ↗IIIF access to images. You can reproduce this for example starting from <https://betamasaheft.eu/manuscripts/ESkae011/viewer> and loading the manifest ↗URL from DigiVatLib, e.g. <https://digi.vatlib.it/iiif/MSS_Vat.et.24/manifest.json> in there. In any other IIIF viewer which allows this, you could load the ↗manifests from Beta maṣāḥəft as well as those of other IIIF providers.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

The same technology is used to call all available images of manuscripts of a given text (Ch. 3).[48]

Fig. 1.14 The ↗Mirador ↗IIIF viewer used to display the images of KAE-011 (ESkae011) and Vat. Aeth. 24 (BAVet24), both containing the *Miracles of Mary* (CAe 2384, ID: LIT2384Taamme).

We have now seen all the main features of visualization for a single manuscript either from the structured ↗TEI or the ↗IIIF presentation ↗API derived from it and linking to the images. This last case with the Mirador viewer, although placed here to give a more complete idea of the visualization of one manuscript in Beta maṣāḥǝft is however already an example of what I will discuss in the following section, which are additional basic visualizations obtained also with the support of tools which have been developed by other projects.

3 Features of the visualization of manuscripts

Encoding is interpretation, but also each visualization is interpretation. I have discussed some of the possibilities offered by encoded ↗XML descriptions of manuscripts and by the use of ↗IIIF to present images and visualize them.[49] I will now introduce some of the additional visualizations in Beta maṣāḥǝft,

48 The manifests ↗URLs available in vHMML are accessible only to their instance of the viewer, which is said to be in line with their partner institutions' requests. It is not possible thus to load in the ↗Mirador viewer one of those manifests. For reasons which are unknown to me, also the possibility to load external accessible manifests into the Mirador instance in vHMML are currently prevented.

49 More on this at p. 191.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

which complement the general entry and image viewer for a single manuscript.[50]

3.1 Additional visualizations for manuscript descriptions

On the example of other projects,[51] where images are available via a ↗manifest, the ↗TEI element <locus> used throughout the description of the manuscript and containing the placement information, is made in the website into a clickable link to allow the direct connection of the description with the images. In Beta maṣāḥəft this produces an OpenSeadragon viewer in a modal block,[52] which will give either the specific image linked or the range of images requested (Fig. 1.15).

50 Ch. 7 gives also some further features of these manuscript descriptions based on the ↗RDF representation of the data, including a short overview of the workflow used to implement the methodology of Andrist et al. 2013. For the full account of this, see Liuzzo forthcoming a and here p. 211.

51 e.g. e-codices <http://www.e-codices.unifr.ch/en>.

52 <https://openseadragon.github.io/>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

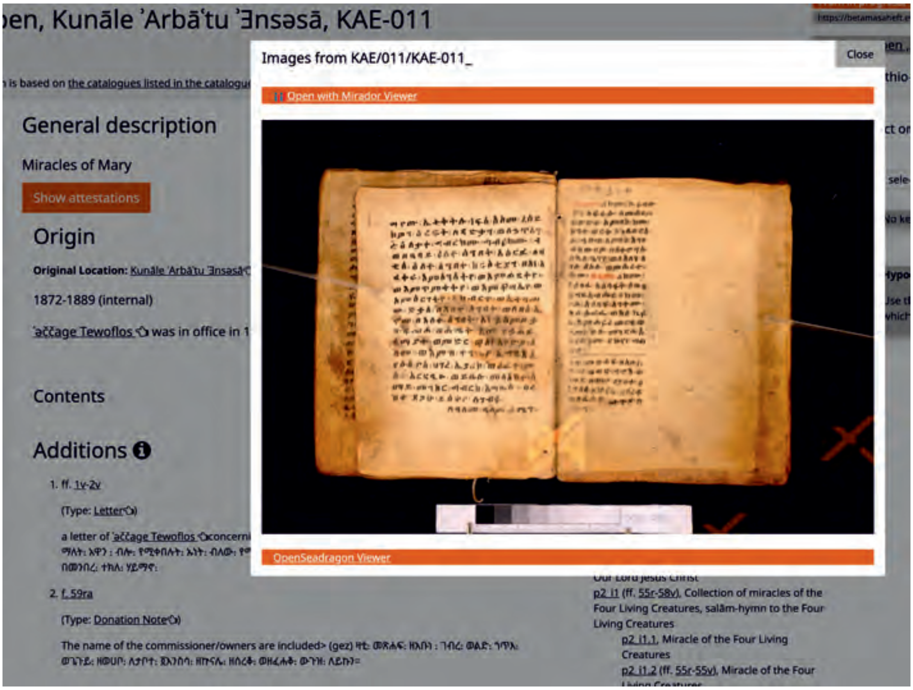


Fig. 1.15 Image from KAE-011 from ↗IIIF ↗manifest on click on *locus* indication for f. 59ra. See <https://betamasaheft.eu/manuscripts/ESkae011/main>.

This is useful to navigate the correct images of the manuscript involved in a given piece of description. Beside description and images, in Beta maṣāḥəft there might also be a transcription of the manuscript, in the same way in which there is an edition for a textual unit. The eventually available transcription of the manuscript's contents includes the visualization of editorial interventions and features of the written page like rubrication (Fig. 1.16).[53]

53 Transcriptions are in all cases partial, and more parts of transcription of manuscripts would be a nice contribution to the project data, although there are plans to extract the bulk of these transcriptions with automated methodologies, using for example Transkribus (<https://transkribus.eu/Transkribus/>).

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Fig. 1.16 Text transcription view of London, British Library, Oriental 718 (BLorient718). See <https://betamasaheft.eu/manuscripts/BLorient718/text>.

To come back to the main description and its parts, as well as giving a prominent position to the declared date of the manuscript as encoded by the cataloguers in the <origDate>, Beta maṣāḥəft collects all dates available in a manuscript description and prints them in a time line (Fig. 1.17) available in the relations view.[54] This allows us to show a wider range of dated units for a manuscript and even competing dates of the same unit, with their attribution.[55]

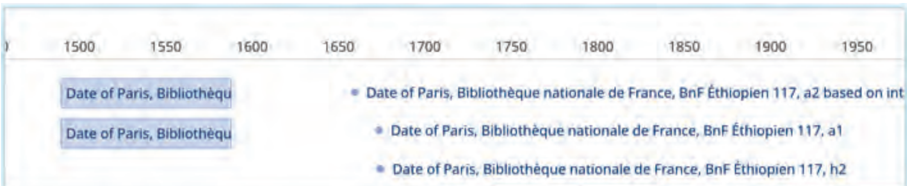


Fig. 1.17 Timeline with the dates in Paris, Bibliothèque nationale de France, Éthiopien 117 (BNFet117). See <https://betamasaheft.eu/manuscripts/BNFet117/analytic>.

To make a further example of an additional extraction of data viewed along with the standard views, we can look instead at the persons involved in the manuscript description. There might be a number of individuals and places

54 Produced with <vis.js>.

55 This should be at least close to 'all the relevant available data, and only those' as foreseen by Andrist 2015a, 517, since each date is attached to its context in the ↗TEI description.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

named in the description or in colophons and other additions to the text. These are collected and grouped by the role they are eventually assigned in the description of the manuscript. Other relevant information about them is also gathered from the XML database, and specifically any other occurrence in manuscript records of a <persName> with the same identifier in @ref and with a specific @role. This will collect, for each relevant actor in the life of the manuscript, a list of other manuscripts which he or she might have interacted with in different roles. In this way, a simple list of the activity of the persons involved is immediately visible in any of the manuscripts they have dealt with.

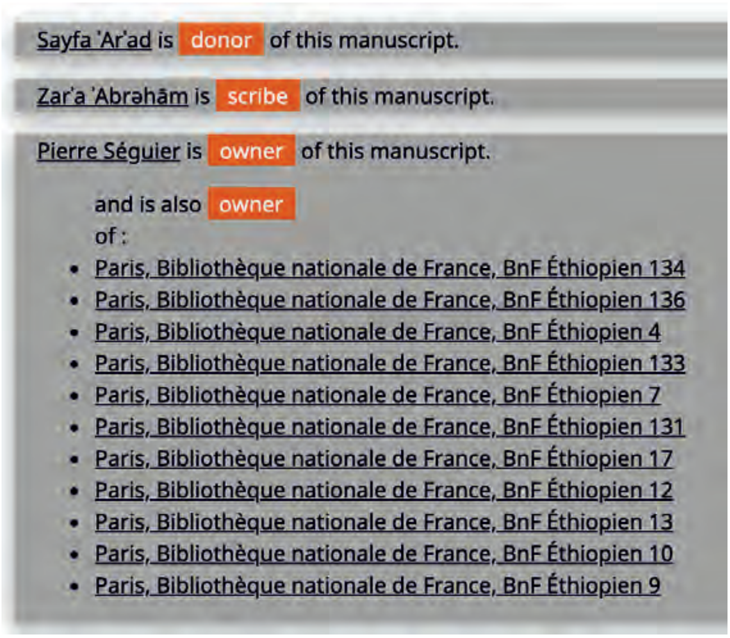


Fig. 1.18 Persons who had a role in the history of Paris, Bibliothèque nationale de France, Éthiopien 32 (BNFet32) related to other manuscripts with which they had something to do. See <https://betamasaheft.eu/manuscripts/BNFet32/analytic>.

While these few examples show some piece of the general view dedicated to a specific type of information, we will look at a more complex case in the next example, which discusses the way in which in Beta maṣāḥəft we reuse software to visualize the collation of the manuscript.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

3.2 Visualization of quires with VisColl and Mirador

Quire structure is indicated in a <collation> element in the description of a manuscript. In a ↗TEI <list> element inside <collation>,[56] each quire is encoded in a <item> element containing information on its dimension (<dim>) and <locus>, with an @xml:id and possibly a @n as in the following example:

```
<item xml:id="q2" n="1">
  <dim unit="leaf">8</dim>
  <locus from="3r" to="10v"/>
</item>
```

The visualization of the quire structure is obtained with ↗VisColl starting from the data in the list which is transformed to the quire and leaf model needed.[57] The initial data model required by VisColl is produced by an ↗XSLT script on the fly and the VisColl XSLT is applied to get the ↗SVG for the ↗browser view.[58] But let me unpack this statement to show what the stages of the transformation are.

The information in each <item> is parsed by the ↗XSLT to check for statements about stubs, added or missing leaves. The parsed information is used, as in the ↗VisColl pipeline,[59] to produce the quire model. The text nodes inside <item> are not entirely free text, as they require for the parsing that the information given for special leaves adheres to a light syntax, which has been inherited as much as possible from what was already the praxis in the Ethio-SPaRe project, namely using phrases like ‘1 added’ or ‘1, stub after 7’, as in the example below.

```
<item xml:id="q1">
  <dim unit="leaf">7</dim>
  <locus from="1" to="7"/>
```

56 Porter et al. 2017, 85 recalls that there is no further specification in TEI Consortium et al. 2018 so Beta maṣāḥəft inherited the Ethio-SPaRe conventions for recording collation, and these are the ones which have been parsed to produce the model which we now use with the mentioned elements. For collation formulas see also Andrist 2015b, 525.

57 <https://github.com/leoba/VisColl>, see Porter et al. 2017.

58 <https://github.com/BetaMasaheft/BetMas/tree/master/BetMas/xslt> which includes a modified version of the original scripts from ↗GitHub. A tool to generate the visualization is also available, see Porter et al. 2017, 91–92.

59 Porter et al. 2017, 86.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

```
   1 added.
   c.1 was originally conjoint to c.8. Today c.1 is a
   single leaf stitched to a stripe of parchment and linked to c.7.
</item>
<item xml:id="q2">
   <dim unit="leaf">10</dim>
   <locus from="8" to="17"/>
   1 added,
   2 added,
   9 added,
   10 added.
   c.8 was originally conjoint to c.1 and c.17 was originally
   conjoint to c.24. Today c.8 and c.17 are single leaves
   linked to the quire by means of stitching.
</item>
```

These two quires described above in ↗TEI XML will be transformed into the XML which can be seen in the following example, and which is in the form required by the version in use of ↗VisColl.

```
<quire n="1">
   <leaf n="1" folio_number="1" mode="added" single="true" conjoin="" position="1"/>
   <leaf n="2"  folio_number="2" mode="original" single="false" position="2" conjoin="6"/>
   <leaf n="3" folio_number="3" mode="original" single="false" position="3" conjoin="5"/>
   <leaf n="4" folio_number="4" mode="original" single="false" position="4" conjoin="4"/>
   <leaf n="5" folio_number="5" mode="original" single="false" position="5" conjoin="3"/>
   <leaf n="6" folio_number="6" mode="original" single="false" position="6" conjoin="2"/>
   <leaf n="7" folio_number="7" mode="original" single="false" position="7" conjoin="1"/>
   <leaf conjoin="1" position="8"/>
</quire>
<quire n="2">
   <leaf n="1" folio_number="8" mode="added" single="true" conjoin="" position="1"/>
   <leaf conjoin="10" position="2"/>
   <leaf n="2" folio_number="9" mode="added" single="true" conjoin="" position="3"/>
   <leaf  conjoin="9" position="4"/>
   <leaf n="3" folio_number="10" mode="original" single="false" position="5" conjoin="6"/>
   <leaf n="4" folio_number="11" mode="original" single="false" position="6" conjoin="5"/>
   <leaf n="5" folio_number="12" mode="original" single="false" position="7" conjoin="4"/>
   <leaf n="6" folio_number="13" mode="original" single="false" position="8" conjoin="3"/>
   <leaf n="7" folio_number="14" mode="original" single="false"  position="9" conjoin="2"/>
   <leaf n="8" folio_number="15" mode="original" single="false" position="10" conjoin="1"/>
   <leaf n="9" folio_number="16" mode="added" single="true" conjoin="" position="11"/>
```


ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

```
  <leaf conjoin="2" position="12"/>
  <leaf n="10" folio_number="17" mode="added" single="true" conjoin="" position="13"/>
  <leaf conjoin="1" position="14"/>
</quire>
```

Starting from this format where all the needed information about each leaf are spelled out, ↗VisColl produces the ↗SVG diagrams which can be seen in Fig. 1.19.

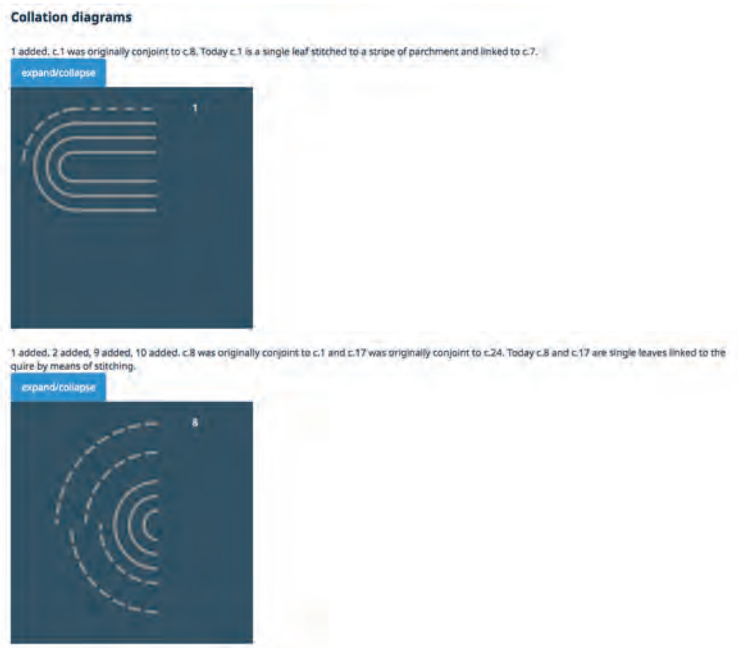


Fig. 1.19 VisColl used to display the collation of Grottaferrata, Exarchic Greek Abbey of St Mary of Grottaferrata, Crypt. Aet. 7 (GAet7). See <https://betamasaheft.eu/manuscripts/GAet7/main>.

We do not use the possibility offered by ↗VisColl to associate images to these diagrams, but we use the same XML of the quire description in our source data to produce a range in the ↗IIIF ↗manifest so that quires can be navigated in the viewer (Fig. 1.13).[60]

60 Beside quires, ranges are produced for content items, decorations, and additions. For a detailed description of how manuscripts and other entities are encoded, please see the Beta maṣāḥǝft Guidelines, <https://betamasaheft.eu/Guidelines/?id=manuscri

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

In this section I have presented a selection of visualizations in Beta maṣāḥǝft, and others will be presented later in the book, but I hope it will be enough to have given a clear idea of the way in which the underlying ↗XML encoding allows anyone to produce multiple visualizations for different purposes.

3.3 Encoding VS data entry

Now that I have discussed some of the features of the visualization of written artefacts in Beta maṣāḥǝft, let me consider some of the points raised by Gippert about online resources for manuscripts.[61] Because Beta maṣāḥǝft is a research environment which needs to host diverse contributions,[62] there is no way the project will ever be able to claim completeness. It would actually be a denial of its definition as research environment. Some of the manuscript records in it will remain stubs, some will be huge analytical entries.

The evaluation of the advantages of working with ↗XML and exploiting it in a native XML database, in comparison to storing data in a database whose architecture is constructed for that purpose, is not to be based on the completeness or quantity of information. For example, a record with a shelf mark and a repository identifier, is already valid according to our ↗schema.

What makes a ↗TEI encoded catalogue description preferable is the freedom it comes with to extend it (or not). And that is also the reason why data entry forms for a TEI encoded description are often meaningless, forcing into a database-like structure something which is, in its nature, different and much more flexible once the encoder has learned it.[63]

Gippert points out clearly the problem of users' expectations from an online catalogue.[64] Very often users expect things from the Web, and not only from a specific ↗application, which are simply not possible.[65] But, while

pts>, and Chapter 10 of the TEI Guidelines, <https://tei-c.org/release/doc/tei-p5-doc/en/html/MS.html>.

61 Gippert 2015a and Gippert 2015b.

62 See p. xxxiv.

63 The Beta maṣāḥǝft team is an example of how this is not at all difficult and is instead rewarding, especially for the researchers doing the encoding. That is also the reason why ↗TEI cannot be too prescriptive, it would void the whole point of being able to say what you need, and to the level of detail needed. One cannot in this respect compare a ↗SQL database to an XML encoded TEI file. It could perhaps be compared to a ↗XQuery capable database like ↗eXist-db.

64 Gippert 2015b, 533.

65 I had once the occasion to defend in public the very clear and well documented statement made in Europeana's calculation of the extent of digitized cultural heritage,

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

a project should make enough effort to meet the requests of selected users' needs, anyone who has ever developed anything for someone else knows how 'flexible' these requirements can be, meaning that from one day to the other they can be reversed ten times. And a good eight out of these ten times, the decisions are reversed with good arguments.

Since this process of encoding is based on a series of layers of decision-taking and investigation, as I have tried to show in the previous examples, and it is thus radically different from a data entry model in which the researcher has to adapt to a form and the requirements of such form, once more it is impossible to consider encoding as simple data entry. Each encoded catalogue description, even where it is largely based on a specific historical catalogue as in the current phase of Beta maṣāḥǝft, is a new cataloguing act. Authorship attribution of the new catalogue descriptions which are collaboratively edited is a problem which still awaits a definitive solution. At the moment what the Beta maṣāḥǝft project does is to be as explicit as possible about each contribution to the encoding, and providing a bibliography which is as complete as possible.[66]

To conclude, there are several good reasons to encode written artefacts in ↗TEI as a base format, which make of it, if not an obvious, a very sensible choice for any new project. Ways of using that structured data and related resources to produce a high quality online resource abound and only very few have been presented. While until now, the focus has been on visualization of the data in the Beta maṣāḥǝft ↗application, in the following section I will give an example of how to query and visualize this same data using tools which are not already in Beta maṣāḥǝft.

4 Moving manuscripts and text circulation

Encoding has not as its aim any visualization, although it allows to build them rather easily as shown above. An encoded manuscript description is a product of research which can be easily reused for further research.[67] We have until now seen how encoded text is used to get some specific output. I would like now to give an example of a type of query, that tries to collect specific information in the encoded descriptions to answer a question, using visualizations to explore the extracted information and support research and arguments.

<http://strategy2020.europeana.eu/>, against perpetrators of the misconception that 'everything is online', without eventually a lot of success despite the evidence.

66 See p. 2.

67 That is also the reason why it is necessary that it is citable in each of its versions, a feature offered via permanent links, see <https://betamasaheft.eu/pid.html>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

I will start from this example question: how did manuscripts and texts circulate? This is not a question that a database or any encoded description can answer alone. It requires years of research and a deep understanding of each case and, if using that source, also a deep understanding of the database hosting the information queried to gather data in support of the research to answer that question. We can use the same data encoded in ↗TEI which is used to offer the views in the website in order to gather analysable information which might help organize various bits of information in ways which are more easily digestible and analysable, so that we may have visualizations of the data used to answer a research question, which are not part of a standard visualization offered by a specific website. The open access availability of this specific type of data, which are encoded descriptions, gives this additional freedom to reuse the data with other tools, in the way the users prefer, without forcing the use of the interface offered by a specific project.

To try to gather data which could help towards answering the example question about manuscript's circulation, we are going to look for occurrences of the <placeName> element inside <colophon>, <incipit>, and <explicit> elements in manuscript descriptions. To date there are fifty-five such elements in <colophon>s, thirteen in <incipit>, and seven in <explicit>.[68] In this example we take these as possible contexts where information on places relevant to the life of the manuscript could be found, together with the current repository. The place of origin of each relevant manuscript could be easily added and incipits and explicits omitted. The number of occurrences is not at all big, and actually these <placeName>s might be different things: places of provenance of a scribe or donor, for example, or any other toponym in the text which is entirely unrelated to the production of the manuscript. This demonstrates that the markup needs to be more precise and grounded on research on these topics,[69] to distinguish different pieces of information, but, at the same time, that it is already possible to get possibly relevant data to reuse. It is not here my aim to provide answers, rather to simply show one of many possible ways to collect and reuse data, and how this can be manipulated to analyse it.

68 The ↗Xpath search functionality of Beta maṣāḥəft (<https://betamasaheft.eu/xpath>) can provide up-to-date figures quickly on this, selecting in the manuscript collection //t:incipit//t:placeName for example.

69 Antonella Brita carries out research related to these questions at the Centre for the Study of Manuscript Cultures in Hamburg, and the present example was briefly discussed with her in Maq̱ala in October 2018.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

4.1 Using SPARQL to get all annotations

These <placeName>s could be queried in Beta maṣāḥəft either from the XML with ↗XQuery or from the ↗RDF with ↗SPARQL.[70] We want to know which manuscripts have these <placeName>s and where they are kept now. Additionally, for each place we want, where they are available, some representative coordinates to be able to plot the results on a map. However, the RDF which contains the ↗Pelagios ↗annotations is not precise enough for this purpose,[71] as the annotations for any mentioned place are related to the main record, not to the exact position in the description, so we cannot know if the annotated occurrence saying, for example, that 'manuscript 1 has target place 2' refers to a place in the colophon, incipit or explicit. The following query has been produced with the query editor in ↗Palladio pointing to the Beta maṣāḥəft SPARQL endpoint returning results in ↗JSON and should be easily reproducible there.[72]

```
PREFIX bm: <https://betamasaheft.eu/>
PREFIX crm: <http://www.cidoc-crm.org/cidoc-crm/>
PREFIX oa: <http://www.w3.org/ns/oa#>
PREFIX lawd: <http://lawd.info/ontology/>
PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
SELECT ?repositoryID ?repositoryPlaceName
                    ?coordinates ?manuscriptID ?manuscriptName ?type
WHERE {
      {?manuscriptID crm:P55_has_current_location ?repositoryID .
         BIND ('current location' as ?type)}
            UNION
       {?annotation oa:hasBody ?repositoryID ;
                    oa:hasTarget ?manuscriptID .
            BIND ('place named' as ?type)}
    ?repositoryID a lawd:Place ;
                        a bm:ins ;
```

70 See Ch. 7 on how the two representations are kept in sync to be used alternatively to query the same information.

71 See Ch. 4.

72 <https://betamasaheft.eu/api/SPARQL/json>, see Seaborne 2013 for the specification of this results format. For other selected visualization tools, see <http://dhd2016.digitale-akademie.de/workshop/tools/>. My use, here and in other places, of this tool is minimal, and more hints on its use can be found at <http://hdlab.stanford.edu/palladio/help/>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

```
16                         lawd:hasName/lawd:primaryForm ?repositoryPlaceName .
17      ?manuscriptID a lawd:AssembledWork ;
18      crm:P48_has_preferred_identifier ?manuscriptName .
19 OPTIONAL {
20      ?repositoryID geo:location ?coord .
21        ?coord geo:lat ?lat ;
22                  geo:long ?long .
23  BIND (concat(STR(?lat),',',STR(?long)) as ?coordinates) }}
```

We are querying here first for manuscript repositories (Lines 8–9) then for triples in the ↗Pelagios format for ↗annotations (Lines 10–13). After declaring the prefixes which are involved in the query, at Line 6 we declare six variables, which are the information we want to obtain and will be rendered as a table with a column for each of these. At Lines 11–12 we get every statement matching the format of a oa:annotation. At Line 14 for each place matching the statements above in the variable ?repositoryID we make sure that they are a places and are also repositories in the bm:ins ↗class (Line 15) before we select their primary name (Line 16). While for the statements with crm:P55_has_current_location we expect the subject of the statement to be a manuscript, this cannot be assumed for the annotations. At Line 17 we need then to restrict the total number of annotations by specifying that the target of the annotation can only be a manuscript. We then retrieve the name of the manuscript (Line 18). This UNION statement (Line 10) puts together annotations and the current repository statements, and for all of these we check with OPTIONAL in Line 19 for any coordinates that are present (Lines 20–22), binding them to one variable as string (Line 23), if available.

With this query a simpler dataset can be obtained which ↗Palladio can allow us to explore and visualize nicely, as in the tabular example in Fig. 1.20. We expect at least one row for each manuscript, because the XML schema rules that there must be a repository, and therefore there must be in the ↗RDF a statement in that format for each manuscript. The ?type variable allows us to filter the two in Palladio simply adding a facet filter.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Fig. 1.20 Table view in ↗Palladio of the results of the ↗SPARQL query in the previous example.

4.2 Using XQuery to focus on specific elements

In the previous query, we do not know if the place ↗annotations are in colophons or subscriptions because, although this would be perfectly possible, the current state of the ↗RDF data in Beta maṣāḥǝft does not allow us to retrieve that information.[73] Querying the XML instead we can make the ↗XQuery ask for this explicitly availing us of the ↗XML tree structure.[74]

```
let $incipitPlaces := $config:collection-rootMS//t:incipit//t:placeName
let $explicitPlaces := $config:collection-rootMS//t:explicit//t:placeName
let $colophonPlaces := $config:collection-rootMS//t:colophon//t:placeName
let $all := ($incipitPlaces, $explicitPlaces, $colophonPlaces)
let $repositories := for $mss in $all let $root := root($mss)
                          group by $R := $root
                           return $R//t:repository
for $place in ($all, $repositories)
```

73 Subscriptions, supplications, and other similar things are encoded in Beta maṣāḥǝft as types of <incipit>, <explicit>, or <colophon> at the time of writing. See <https://betamasaheft.eu/Guidelines/?id=manuscriptContents>.

74 In this example some variables defined in the application are used, starting with $config: and they replace the path to the collection where the manuscripts are stored.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

```
9    let $manuscriptName := string(root($place)/t:TEI/@xml:id)
10     let $repositoryPlaceName := replace(titles:printTitleMainID($place/@ref), ',', '')
11     let $getcoordinates := coord:getCoords($place/@ref)
12  let $coordinates := if(contains($getcoordinates, 'no coor')) then () else $getcoordinates
13  let $type := if($place/name()='repository')
14                    then 'current location' else
15                    'place named in ' || $place/name()
16  return
17     concat($manuscriptName,';', $repositoryPlaceName,';', $coordinates,';',$type)
```

In Lines 1 to 3 we select from <incipit>, <explicit> and <colophon> their descendants <placeName>s and in Line 4 we join them all in the same sequence (i.e. if there are doubles they will stay). At Lines 5–7 we select repositories of the manuscripts where the <placeName> selected before occurs by checking the root element and its descendant <repository>, grouping by the root, so that we get the repositories only once.[75] At this point we can look through a sequence which has all <placeName> elements in a manuscript and all repositories in which these manuscripts are kept currently. While in the previous example we have added the requirement that these named places are repositories, here we can avoid this, as we have restricted to specific parts of the description where the annotation occurs, so that we know that the toponym occurs in a relevant context and we do not care if that named place is also a repository. At Line 8 we can iterate through this joined sequence. With a local function, not reproduce here, in Line 9 we store the canonical title of each of the resources pointed to by the <placeName>s and <repository> collected, and, with other locally defined functions at Lines 10–11, we try to extract coordinates for each place in the list.[76] This coord:getCoords() function determines also how to get the coordinates based on the identifier provided, meaning if from Wikidata, or from the Beta maṣāḥəft data, or from ↗Pelagios. At Line 17 we are finally ready to let the query return for each place a semicolon separated row, with the identifier of the manuscript (from Line 9), the name of the place (from Line 10), its coordinates (from Line 12) and type (from Line 13), meaning if it is a named place or a repository, so that we have produced on the fly a table with the information we want to look at.

75 There are other elements in the encoded text which could be relevant, especially those in <history>, but they are not used enough at the moment in Beta maṣāḥəft to be interesting for this example. See <https://betamasaheft.eu/Guidelines/?id=History>.

76 The modules which output titles and coordinates are available in the ↗GitHub repository of the application and are called respectively titles.xqm and coordinates.xqm.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

The result will look like the following example, where I reproduce only a few lines.

MasKa003;Beta Māryām;12.033657,39.043390;place named in colophon/incipit/explicit
FSUor41;Məsr;27.0,29.0;place named in colophon/incipit/explicit
FSUor134;Məsr;27.0,29.0;place named in colophon/incipit/explicit
FSUor134;Məsr;27.0,29.0;place named in colophon/incipit/explicit
Parm3852;Beta Māryām;12.033657,39.043390;place named in colophon/incipit/explicit
BAVcerulli223;Beta Māryām;12.033657,39.043390;place named in colophon/incipit/explicit
EMML6964;Beta Māryām;12.033657,39.043390;place named in colophon/incipit/explicit
IVEf78;Golgotha;;place named in colophon/incipit/explicit
BNFet149;ᵓAdāl;10.45222,41.177806;place named in colophon/incipit/explicit

Pasting the result of this query, which is a simple tabular format (↗csv) with a semicolon as a separator, in ↗Palladio already allows us to see some more interesting things,[77] like the relation to common places of different manuscripts and the networks among them.

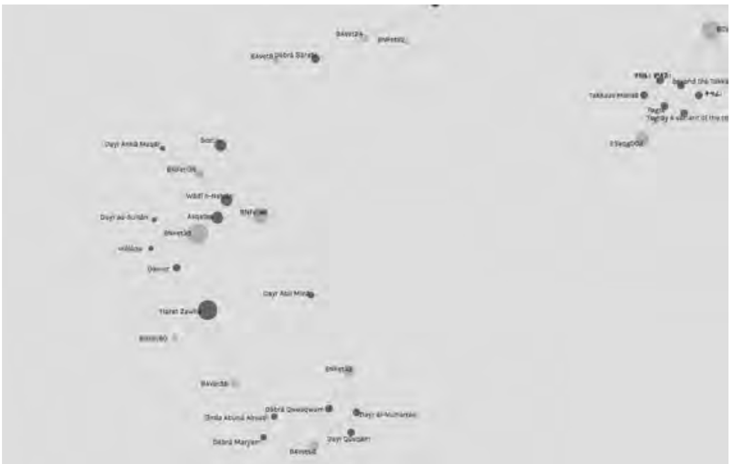


Fig. 1.21 ↗Palladio graph view of the results of the ↗XQuery from the previous example.

77 The ↗XQuery module placesIncipExplicColop.xql, to be run within the Beta maṣāḥəft ↗application, and the results (placesIncipExplicColop.csv) are available in the project website at <https://github.com/PietroLiuzzo/DHEth/tree/master/1%20Manuscripts>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Fig. 1.22 ↗Palladio faceted map view of the results of the ↗XQuery in the previous example.

4.3 Focus on a specific textual unit

A query similar to the one in the previous example can be made with a specific textual unit for its subject. In this case in the query we will only focus on the repositories and will add the dates related to the manuscripts.

```
let $work := 'LIT1146Argano'
let $mss := collection($config:data-rootMS)//t:title[contains(@ref , $work)]
for $ms in $mss
   let $repo := root($ms)//t:repository
   let $manuscriptName := string(root($ms)/t:TEI/@xml:id)
   let $date := root($ms)//t:origDate
   let $stringDate := for $d in $date
   let $atts := for $att in ($d/@notBefore, $d/@notAfter, $d/@when) return string($att)
                return min($atts)
   let $coordinates := coord:getCoords($repo/@ref)
   let $repositoryPlaceName := titles:printTitleMainID($repo/@ref)
   return
      $repositoryPlaceName || ';' || $coordinates || ';' || $manuscriptName || ';' || min($stringDate)
```


ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

With the ↗XQuery in this example we ask (with the selector in Line 2) for manuscripts which contain a given textual unit, *ʾArgānona Māryām* (CAe 1146, ID: LIT1146Argano), with their date at Lines 6 to 9, precisely the earliest date reduced to the year part only,[78] and coordinates of the current repository at Line 10. The same could be done for the original place of conservation of the manuscript (<origPlace>) and the places named in <colophon>s, <incipit>s, and <explicit>s or any other part of the description deemed potentially relevant.[79]

In Fig. 1.23 you can see how you can get filters, timeline, and map to explore the results.[80]

Fig. 1.23 Map of the current repositories of the *ʾArgānona Māryām* (CAe 1146, ID: LIT1146Argano) in ↗Palladio.

78 This is due to ↗Palladio's date format requirement, which, as far as I could see, does not take ranges.
79 The ↗XQuery module WhereIsLIT1146Argano.xql, to be run within the Beta maṣāḥəft ↗application, and the results (WhereIsLIT1146Argano.csv) are available in the project website at <https://github.com/PietroLiuzzo/DHEth/tree/master/1%20Manuscripts>.
80 A basic visualization of this kind, with much less flexibility on the input data, is also offered in Beta maṣāḥəft, <https://betamasaheft.eu/workmap>, where repositories and places of origin can be shown.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

5 Looking outside of Beta maṣāḥəft

All is nice and easy until we work with data in the same system. If in the previous section I have shown data from Beta maṣāḥəft with other tools, now I want to conclude this chapter by going even further away and giving an example of joining the Beta maṣāḥəft data to other resources.[81] It would have been nice, for example, to expand the query in the last example of the previous section and look at textual units which are represented in more than one of the literatures of the Christian Orient. A starting point for this is the fact that Syriaca.org and the PAThs project share indication of the corresponding repertories numbers (clavis), as well as Pinakes,[82] which would have been a perfect port of call for retrieving information about manuscripts with contents related to relevant textual units.[83]. Ideally one could have looked at these in a programmatic way and extract the same information above, which is mostly present also in these databases, to be able to look at the geographical and chronological distribution of the works which are attested in at least two of these databases. Unfortunately it is not yet possible to get a list of these identifiers from reference repertories from PAThs or Syriaca.org.[84] We should be able, with not that much effort of programming and implementation, to know about all witnesses with one query from any of the access points provided. For example, given that CPG 3939 (Ephraem Graecus,[85] *In Transfigurationem Domini*) is the same as CC 0215 and CAe 4410,[86] there are eighty-five manuscripts containing CPG 3939 in Pinakes, in PAThs there is for CC 0215 one manuscript pertaining to the Coptic tradition, in Beta maṣāḥəft there are four and a series of others which attest textual units containing the one textual unit in question, and in Syriaca.org there are others with the same

81 There are many more ways to access and reuse in other resources the data produced and exposed by Beta maṣāḥəft, some of which will be detailed in other chapters.

82 <https://pinakes.irht.cnrs.fr/recherche-generale.html>.

83 The clavis offered by <https://biblia-arabica.com/> is not yet online as a database at the time of writing.

84 I could search the identifiers one by one to the best, and while for Pinakes there is no public ↗API or ↗SPARQL Endpoint, PAThs API only offers a search based on one of these identifiers, and I could not find this information in the Syriaca.org ↗RDF to get to know how to query it. However the Syriaca.org SPARQL endpoint can be queried for related data, as shown in the examples of the section of Ch. 7 at p. 194. It is to be hoped that some agreement about standards of access is reached so that similar queries will not be a mirage any more in few years time from now. See p. 79.

85 <https://pinakes.irht.cnrs.fr/notices/oeuvre/8707/>.

86 For CC 0215 cf. <https://atlas.paths-erc.eu/works/215>; for CAe 4410 cf. <https://betamasaheft.eu/LIT4410Transfiguration>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

identifiers, related to CPG 3939.[87] We should also easily get to know if there is an authorship attribution and other available pieces of information. All these repositories share references to these repertories, and a user can still simply search in each and get its data, and will be lucky enough to find the identifiers to make his own matchings.[88]

An equally interesting test would have been to be able to search more generic ↗SPARQL endpoints for manuscripts of a given work, and simply join those results to the ones of Beta maṣāḥəft and eventually the repositories mentioned above.[89] Such access points fortunately exist and offer Linked Open Data, and the biggest one, to my knowledge, is Europeana.[90] The attempts I made with <https://sparql.europeana.eu>, however, looking at manuscripts which contained in their title part of the title of a work like the *Miracles of Mary* (CAe 2384, ID: LIT2384Taamme), which have a tradition in medieval Europe, almost succeeded, in as far as it is rather easy to get a list from the user interface simply searching for the title.[91] The exercise highlighted, however, the amount of prerequisite knowledge needed to query a model even if documentation is provided. The Europeana Data Model provides title, provenance location, and date, although it does not match contents precisely in most cases, but it would have been enough. The Europeana ↗API offers instead results which we can search and navigate as in the portal interface,[92] and which in ↗XQuery we can parse to join them with those of Beta maṣāḥəft.[93] In this last example I will query the Europeana API from a XQuery script to retrieve results as in the previous example, producing a tabular format to

87 I could not find this work among those list under <http://syriaca.org/person/13> at the time of writing.

88 One possibility to achieve this is now offered by the ↗DTS Specification. If all four databases in this example provide an ↗API following it, it would be quite easy for each of them, and anyone else, to gather the information consistently.

89 Some libraries offer this kind of service, but it is off course not that convenient to query a library endpoint, as it restricts the query to the holdings of that institution and will have to know before hand that there is there something interesting. In this case a query to an ↗aggregator is much better.

90 <https://www.europeana.eu/>. Biblissima (<http://beta.biblissima.fr/>) is another example where a lot of promising work has been put into producing ↗RDF. See <https://demos.biblissima.fr/snorql/>.

91 <https://www.europeana.eu/portal/it/collections/manuscripts?q=Miracles+Mary>.

92 <https://pro.europeana.eu/resources/apis/search#metadata-sets>.

93 The ↗XQuery module `MiraclesOfMary.xql`, and the results (`miraclesofmary.csv`) are available in the project website at <https://github.com/PietroLiuzzo/DHEth/tree/master/1%20Manuscripts>. An additional parameter is required for the query, called `wskey` and containing the API key obtained with a simple form on the Europeana website.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

display in ↗Palladio. I will drop the coordinates functionality, however, and focus only on the time.[94]

```
let $apiroot := 'https://www.europeana.eu/api/v2/search.json'
let $parameters :=  '?query=Miracles+Mary&rows=100'
let $apirequest := $apiroot || $parameters
let $apiresponse := httpclient:get(xs:anyURI($apirequest), true(), <Headers/>)
let $response := util:base64-decode($apiresponse)
    let $parse-response := parse-json($response)
    for $item in $parse-response?items?*
    let $manuscriptName : = normalize-space(replace($item?title, ';', ', '))
    let $type := $item?type
     let $link := $item?link
     let $apiresponseRecord := httpclient:get(xs:anyURI($link), true(), <Headers/>)
     let $responseRecord := util:base64-decode($apiresponseRecord)
     let $parse-responseRecord := parse-json($responseRecord)
     let $dataProvider := $item?dataProvider
     let $repoPlaceName := $parse-responseRecord?object?proxies?*?dctermsProvenance?*
     let $repositoryPN := if(count($repoPlaceName) gt 0)
                                            then $repoPlaceName else $dataProvider
```

94 An additional ↗SPARQL query to Wikidata (like <https://query.wikidata.org/#SELECT%20DISTINCT%20%3Fitem%20%3FitemLabel%20%3Fcoord%0AWHERE%0A%7B%7B%0A%20%20%23Consortium%20of%20European%20Research%20Libraries%20%0A%20%20%3Fitem%20wdt%3AP463%20wd%3AQ1127581.%7D%0A%20UNION%0A%20%7B%23IIIF%20consortium%0A%20%20%20%3Fitem%20wdt%3AP463%20wd%3AQ35677307%7D%0A%20%20%3Fitem%20wdt%3AP625%20%3Fcoord%20%3B%0A%20%20%20%20%20%20%20%20rdfs%3Alabel%20%3FitemLabel%20%3B%0A%20%20SERVICE%20wikibase%3Alabel%20%7B%20%0A%20%20%20%20bd%3AserviceParam%20wikibase%3Alanguage%20%22en%22%20.%0A%20%20%20%7D%0A%7D%20LIMIT%20100>) could have been added to get coordinates for the institutions, for example, by looking at coordinates of institutions member of the IIIF Consortium or the Consortium of European Research Libraries and filtering the label of the entity, but this would have not been entirely reliable, also given the information mapped to the repository name variable in this example, and the fact that it would have been the result of a search, with a series of possible matches, and not an exact match. Taking the time to check the Wikidata item for each controlled repository name would allow to add a simple request to the Wikidata endpoint like `SELECT ?coord WHERE {wd:Q383931 wdt:P625 ?coord .}`. Wikidata also has introductory pages with loads of information on how to learn and use SPARQL, see <https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/Wikidata_Query_Help>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

```
17 let $repositoryPlaceName := normalize-space(replace($repositoryPN, ';', ', '))
18 let $dcTermsCreated := $parse-responseRecord?object?proxies?*?dctermsCreated?def?*
19 let $dcDate := $parse-responseRecord?object?proxies?*?dcDate?def?*
20 let $stringDate := ($dcDate,$dcTermsCreated)
21 return
22 $repositoryPlaceName || ';' || $manuscriptName|| ';' ||
                  string-join($stringDate, ' ') || '; Europeana ' || $type
```

Lines 1–3 are where the text of the ↗HTTP request to the Europeana ↗API is set up. At Line 4 the request is sent and it is parsed so that it can be navigated using the support for ↗JSON provided by ↗XQuery in ↗eXist-db.[95] For each item in the response to the search API we can immediately get the title (Line 8) which needs to be formatted a bit, to avoid the semicolon separator and to normalize the spaces to avoid new lines. We can also already store important information from Europeana, which is the item type. Europeana being an ↗aggregator of many items of different origin and scope this will help us to filter. The link, which we store in a variable at Line 10, is important because it allows a further request to be sent, to retrieve information about each item which would have not been available in the search response directly. Lines 11–13 send this further request for each item in the list of search results obtained from the previous HTTP request. It is not that simple, at least for me, to decide between the data provider and the provenance, so I store in two variables each of these. If the `dctermsProvenance` is available, I assume it is more precise than `dataProvider` which I use as a fallback (Line 16). The value is then formatted to be used as place name for the repository of the item. The fields in the JSON response, which may contain a date for the object in question, are saved and joined in Lines 18–20 so that the same row as in the previous example can be built, as a row in the tabular output. Adding this list of rows to the rows obtained running the query in the previous example for the *Miracles of Mary* (CAe 2384, ID: LIT2384Taamme) gives a unified list of objects which can be filtered and searched like in Fig. 1.24.

95 <https://exist-db.org/exist/apps/wiki/blogs/eXist/XQuery31>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

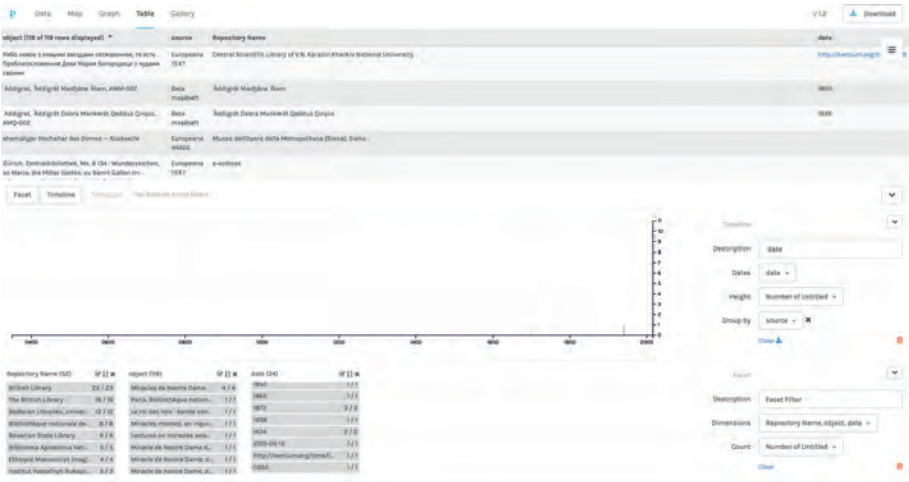


Fig. 1.24 Table and filters for a joint list of items related to the *Miracles of Mary* (CAe 2384, ID: LIT2384Taamme) from Beta maṣāḥǝft and Europeana viewed in ↗Palladio.

In conclusion to this section, I hope I have shown that there are several ways to query data to answer interesting questions about manuscripts, and by no means the ones presented are even a fraction of the possibilities, but that the availability of this data in standard, documented, and accessible ways is crucial to be able to use them for any specific individual question. If the data in Beta maṣāḥǝft had not been openly accessible via its ↗SPARQL endpoint, there would have been no using of ↗Palladio in the way shown in the first example. The same is true for the last example, where it is thanks to the fact that Europeana exposes an accessible and documented ↗API that we can query it and reuse the information using the tools we want. Offering a very nice interface is not enough to allow people to use a resource, and, while producing meaningful and dynamic visualizations is certainly important, letting users do their own nice visualization is even more useful, especially in the long term. Curating reliable data is thus a separate and far more important task which should be kept well distinct from the needs of one specific final product, let it be a database or a printed book. A research environment and the community involved in it should be built around the data, not around one access point or interface of any kind. Additionally, this nice data should be available and accessible in standard ways. It would have been nice to be able to query different SPARQL endpoints for data following the same model or at least sharing some properties or classes. It would have been nice to access data without having to learn specific API responses and requests but relying

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

on shared standard specifications like ↗IIIF and ↗DTS. But this is still hardly achievable, and we are back at a level of interchange of data quite far from interoperability.

With these examples I have also introduced the way in which all queries in this book are made, not to teach anyone how to write ↗SPARQL or ↗XQuery, but to show how it is nothing which cannot be read and understood. I have shown a SPARQL query to ↗RDF data, a XQuery to ↗XML data and an other XQuery requesting information from a web ↗API and parsing a ↗JSON response, all of these producing a simple tabular output which has been visualized using ↗Palladio instead of the Beta maṣāḥəft application. This chapter should have given enough evidence both of what Beta maṣāḥəft does with the ↗TEI data for manuscripts and at least some examples of what else can be done with that to reuse the data also outside the project. In the next chapter I will develop on this aspect, by selecting another group of written artefacts, namely Aksumite inscriptions, to put them in context by comparing them to other datasets of inscriptions from the same period, relying on their availability as ↗EpiDoc, a customization of TEI for ancient documents.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Chapter 2. Comparing Inscriptions

> Als Kübelreiter, die Hand oben am Griff, dem einfachsten Zaumzeug, drehe ich mich beschwerlich die Treppe hinab; unten aber steigt mein Kübel auf, prächtig, prächtig...
> (As bucket rider, hand on the handle, the simplest bridle, I turn down the stairs, painfully; but below my bucket rises on, glorious, glorious...)
>
> (F. Kafka, *Der Kübelreiter*, 1917)

'Papyri and ostraca from Roman Egypt survive in sufficient numbers to invite statistical analysis and thus to teach us something out of the numbers themselves that is not evident in the body of any single text.'[1] What MacMullen so neatly formulates for papyri and ostraka is certainly true of inscriptions as well, as shown by its consideration later in the same article on Latin epigraphy, and not just from the Roman Egypt,[2] but especially if we take *all* the epigraphic evidence, meaning all what is openly available without time and space limits. We can learn even more from the numbers (and this is one of the driving ideas of producing and sharing structured data), if we have multiple datasets which we can compare in order to learn something from comparing the numbers that is not evident from a single set of numbers. Since this article 'scholars have been acutely aware of just how characteristic of the ancient world was the carving of inscriptions, at least in certain phases like the Hellenistic Greek and the Roman imperial periods'[3] and, in this chapter, I would like to show with a limited example a possible way in which the perspective can be fruitfully enlarged to analyse in more details, from a detached point of view and ignoring entirely the contents, the *how* in 'how characteristic' of the previously quoted passage. It needs to be anticipated that the result will be rather disappointing.

1 MacMullen 1982, 234.

2 For quantitative studies of inscriptions of Sicily, see Prag 2002 and more recently Prag 2018. Another example of quantitative research, which is especially interesting for its focus on the transitions, can be found in Di Segni 2017.

3 Bolle et al. 2017, 17.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

We could also, theoretically, increment this statement even more.[4] There are certainly numbers which can be mashed up from papyri, inscriptions, ostraka, manuscripts, and other written artefacts which can teach us something we could not learn from inscriptions only or manuscripts only. But this is a matter of future investigation which will require, to begin with, that proper questions are formulated; secondly, that for these questions data is openly available and easily accessible; thirdly, that access methods and data follow such standards that make them understandable with as little effort as possible. These requirements are typically not met, especially when multiplied by the different types of relevant written artefacts, so I will leave also this for now.

In this chapter I will focus on inscriptions and attempt a data comparison exercise trying to contextualize inscriptions from Ethiopia of the Aksumite period in the wider epigraphic context of the period from 300 to 700 CE.[5] This is not meant to provide researchers with results from these comparisons, but to demonstrate what researchers could use and how they could use it for their own research. We will be looking at a different selection of written artefacts in comparison to the previous chapter, and expand the data according to that selection to other data sources. It will be a view on the data from above, as also in the following Ch. 3, but a humble one, like the flight of Kafka's man in the epigraph to this chapter, who is without coal almost as we lack data.

1 Inscriptions from Ethiopia in context

To be able to compare several datasets there are a few basic requirements, especially if one wants his analyses to be reproducible and verifiable by others. First of all, the data needs to be available and to remain so for some time at least.

Secondly, the available data needs to be in a format which is reusable because it is understandable to its potential users. It almost does not matter if it is in ↗CSV or ↗RDF or ↗XML, as far as the data is described properly. It is instead important that some data is provided which is not simply the visualization of it via a web interface. Having data visible online does not automatically mean that it is available for research, it is only available to users of the application. When the data is provided also as structured data, then it reaches all those people who can and would like to reuse the data in their own research, free of any interface imposed on it.[6] It is unfortunately very often

4 There are several ↗aggregators, portals, and online repositories which include several types of data, and, among them, the most important to mention is certainly Europeana, <http://europeana.eu/>.

5 The Aksumite period in Beta maṣāḥəft.

6 See also Ch. 7, p. 187 and the basic Linked Open Data principles.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

the case that projects do not understand their data as the real primary scientific output, but instead concentrate on a product such as a book or a website.

Thirdly, and very importantly, if the data is in a controlled and structured format documented collaboratively by a community of interested stakeholders who agree on common vocabularies, then it becomes possible to formalize the software to a level to which it is really possible to use data from a number of different sources together and make useful comparisons without the burden of multiple, insecure, data-loss prone transformation passages.

In this chapter I will compare ↗EpiDoc datasets which are available online, and that therefore meet the three requirements listed above. Additionally the chosen datasets make use of or have been aligned by myself to the EAGLE Vocabularies.[7] These are ↗SKOS vocabularies in which the terms used in archives of inscriptions to describe a specific category of information have been listed, linked to one another, and grouped, assigning to each of them a unique ↗URI. Their purpose is to align via these URIs the diverse terms used across languages and epigraphic traditions, without flattening all to one language or one list, but exploiting the potential of a non-hierarchical ↗RDF representation. The EAGLE Vocabularies are used to align some core information about the inscription and its support: material, type of inscription, object type, and execution technique especially.[8]

7 Liuzzo 2014, 194–196. These resources were produced by myself and a large team of contributors from the EAGLE Network, especially Silvia Evangelisti, Anita Rocco, Donato Fasolini, and Francesco Mambrini, during the EAGLE Best Practice Network project: Grant Agreement CIP 325122 (2013–2016). After the end of the project they continued to be occasionally updated in the ↗GitHub repository where they reside, which is linked to the EAGLE website (<https://www.eagle-network.eu/resources/vocabularies/>). The Ancient Graffiti Project (<http://ancientgraffiti.org/Graffiti/>), ISicily project (<http://sicily.classics.ox.ac.uk/>), Epigraphic Database Heidelberg (<https://edh-www.adw.uni-heidelberg.de/home>), and Supplementum Epigraphicum Graecum through Nicolò Bettegazzi have contributed to the vocabularies after the end of the project. In Prag and Chartrand 2018, 248, the authors rightly comment on the lack of adoption of the EAGLE Vocabularies. These are actually adopted by a number of large projects and it is possible to show what can be done, although the amount of work yet to be done on these vocabularies as well as on the data is without doubts still enormous.

8 The latter is the only vocabulary based on a specific study, now published in Evangelisti 2017 which should guide its usage. In this article both the complexity of the matter and the status of the current encoding, which rarely meets the terminological requirements listed, are exemplified. In none of the datasets in use in the present chapter is a distinction made with regard to the execution technique between the tool used and the shape of the letters, only one value is added. This article actually defines a specific 'ontology', in the semantic web terminology (see Ch. 7), for the description of the execution techniques, which could be translated for example to a ↗CIDOC-CRM

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Because there is this 'shared authority list' and because for epigraphy there is a standard ↗TEI customization, namely ↗EpiDoc, which is used and shared by several projects, we have the necessary data to make simple comparisons and see if they can lead to interesting observations. Let me now present the data used for this comparison.

1.1 Inscriptions from Ethiopia and Eritrea

Inscriptions from Ethiopia and Eritrea are not many,[9] but are of core importance for the study of the history and language of the region and especially of the Aksumite kingdom with regard to the example selected in this chapter.[10] In Beta maṣāḥəft these inscriptions are encoded following the same ↗schema as the manuscripts:[11] they are always ↗TEI encoded descriptions, but they additionally validate to the ↗EpiDoc schema although they follow primarily the project guidelines.[12] Only the Aksumite inscriptions have a stub record at the moment, but here I have added alignment to the EAGLE Vocabularies for object type, material and execution technique, basing the information on the data available in *RIÉ* only.[13] Thanks to a cooperation with the project Digital Archive for the Study of pre-islamic Arabian Inscriptions (DASI)[14] there are in Beta maṣāḥəft also some earlier inscriptions, which have been transformed from the DASI EpiDoc to validate to the Beta maṣāḥəft schema, but these will not be part of the data used in the current chapter. I will only deal with the inscriptions of the Aksumite period, including the Greek inscriptions,[15] with the addition of a funerary stele,[16] found in 2005, and two new bronze inscriptions from this period,[17] for a total of 113 written artefacts.[18]

extension. Translating this kind of studies, based on understanding both the needs of the discipline and those of computational resources, is in my opinion the kind of intellectual exercise necessary for the future of data in the Web, as discussed in Oldman et al. 2016, 255 and 257. I have attempted this kind of exercise for Andrist et al. 2013, see Liuzzo forthcoming a.

9 See Bausi and Liuzzo 2018 for a brief overview on the inscriptions and how they are encoded in Beta maṣāḥəft.

10 See Ch. 4.

11 See Ch. 1.

12 Elliott et al. 2007. ↗EpiDoc Guidelines: <http://www.stoa.org/epidoc/gl/latest/>. The bibliography page there provides also a selection of publications.

13 Bernand et al. 1991.

14 <http://dasi.humnet.unipi.it>, Avanzini et al. 2014.

15 Which means that *RIÉ* 276 and *RIÉ* 277 are not part of this corpus, as they might be Hellenistic. See Ch. 4.

16 Fiaccadori 2007.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

It is to be noted that this dataset has some peculiarities.

1) Most of the inscriptions in *RIÉ* would have not been even recorded in other database contexts (e.g. *RIÉ* 207–211) as they do not offer enough understandable text.
2) *RIÉ* 212 is a record collecting sixty-four monogram inscriptions, all engraved before being placed in the wall of the tomb of King Gabra Masqal (PRS4327GabraMa).
3) The quality of the information in the collection regarding the selected grounds for comparison is not homogeneous: materials are not always given, nor are details about the execution techniques.

Nevertheless this small set of data has the characteristics listed above which allow the extraction and quantitative analysis of the data.

1.2 Comparison datasets

The other datasets taken into consideration are described in the following sections, considering only the subset of inscriptions dated to the Aksumite date range, meaning 300 CE to 700 CE as registered by the Beta maṣāḥǝft project with ↗PeriodO.[19] The date range was set strictly, so that inscriptions with a date range like 250 to 400, 340 to 701 or similar should *not* be included in the selections from other datasets.[20]

17 Discussed in Nebes 2017, and Bausi 2018 with colour photos in Yohannes Gebreselassie 2017.

18 The static dataset used for the present chapter is available here <https://github.com/PietroLiuzzo/DHEth/tree/master/2%20Inscriptions/AksumiteInscriptions>. The latest data about the inscriptions can be downloaded from the website or obtained from the official repository directly (<https://github.com/BetaMasaheft/Manuscripts/tree/master/RIE>).

19 <https://chronontology.dainst.org/period/bKEIhoB9UmVV> provides a view of the period in parallel with other period definitions from the Getty Thesaurus of Art and Architecture and PeriodO. See Rabinowitz forthcoming.

20 The selection is done in the script with a simple ↗XPath on the XML requesting inscriptions with an element <origDate> which has @notBefore-custom greater than or equal to (ge) 0300 and @notAfter-custom less than or equal to (le) 0700, i.e. //tei:origDate[@notBefore-custom ge 0300][@notAfter-custom le 0700]. It was not possible, at the moment of writing this chapter, to select the relevant inscriptions from the South Arabian peninsula, collected by the DASI project. This dataset offers a much wider set of questions, which would go well beyond the scope of this chapter and might be instead object of a future contribution.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

1.2.1 ISicily

The ↗EpiDoc data enriched with EAGLE Vocabularies ↗URIs is available online and can be easily downloaded, with a total of 327 inscriptions, downloaded as ↗CSV and ↗XML from the website (<http://sicily.classics.ox.ac.uk/>) or the GitHub repository (<https://github.com/JonPrag/ISicily>). This is a geographically well defined, multilingual corpus of the inscriptions from Sicily digitally available online, where data can be easily accessed, downloaded, and verified. This dataset provides a good comparison as there are very few possible direct connections with the data from Ethiopia and Eritrea.

1.2.2 Trismegistos

This dataset is a selection of texts from Trismegistos, kindly provided by Mark Depauw, extracting everything which was not on papyrus or parchment or on wood in the form of a tablet between 300 and 700 CE from Egypt.[21] Trismegistos data is famously multilingual and is the first port of call for any scholar working with Egypt. The data was aligned to the EAGLE Vocabularies by myself and includes a very wide variety of languages, a unique feature of Trismegistos. The selection of 1466 inscriptions on a variety of supports and materials, includes also inscriptions with provenance from *Aethiopia* (see Ch. 4) and Sudan. Trismegistos offers also dynamic charts produced via Google Charts in several views of the data. This dataset is the closest geographically to the Ethiopian inscriptions and can provide hints on continuities. The Epigraphic Database Heidelberg (EDH) includes for the selected time frame eleven inscriptions from Egypt only and is not multilingual. The data was cleaned and transformed to XML to extract the relevant values for alignment, then transformed once more injecting the aligned values into the XML.

1.2.3 Epigraphic Database Heidelberg (EDH)

The ↗EpiDoc export of the data, which contains alignment to the EAGLE Vocabularies can be obtained online, as well as an export of search results in the form of an Excel table. EDH is also the only online archive dedicated exclusively to inscriptions which provides dynamic chart visualizations of the data in the statistical data view of the search results. The coverage of EDH is very wide although not complete and constitutes a non-multilingual (only bilingual inscriptions are recorded systematically for the completed provinces) comparison base for space distribution. There are 4040 dated inscriptions in EDH to date, which fall in the time span from 300 to 700 CE. This

21 The dataset is available in TM Corpusdata at the page <https://www.trismegistos.org/tmcorpusdata/13> thanks to Yanne Broux and Mark Depauw.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

dataset is also included, in an older version, in the following dataset, which would be otherwise representing basically only Rome. The data is available in several formats from the website, <https://edh-www.adw.uni-heidelberg.de/home>.[22]

1.2.4 The Europeana network for Ancient Greek and Latin Epigraphy (EAGLE)

This is the largest accessible collection of inscriptions for which ↗EpiDoc files with alignment to the EAGLE Vocabularies can be obtained. The collection is multilingual to the extent to which the contributing partners are and contains also non-inscribed objects. The data is available also from other websites, like Europeana and Trismegistos,[23] but I could avail myself of a copy of all the EpiDoc data in EAGLE already accessible to me, which is available to anyone for reuse on demand, while waiting for further development of the website to facilitate its direct exportability, which is now possible only for single items. This dataset is interesting for comparison because of its breadth, although the scope is hardly larger than that of EDH. I have kept from this data only the datasets from content providers which had aligned their authority lists to the EAGLE Vocabularies. In this EAGLE dataset, comprising in total 32,320 inscriptions dated to the selected time frame, are included (in their state in EAGLE as of November 2017):[24]

1) Epigraphic Database Bari, <http://www.edb.uniba.it/> (EDB), with 21,990 inscriptions in the time frame;
2) Epigraphic Database Roma, <http://www.edr-edr.it> (EDR), with 4689 inscriptions in the time frame;
3) Epigraphic Database Heidelberg, <https://edh-www.adw.uni-heidelberg.de/home> (EDH), with 3363 inscriptions in the time frame (see above);
4) Inscriptions of Roman Tripolitania, <http://inslib.kcl.ac.uk/irt2009/> (IRT), with 129 inscriptions in the time frame;
5) Ashmolean Museum Latin Inscriptions, <http://latininscriptions.ashmus.ox.ac.uk/> (AshLi), with 31 inscriptions in the time frame;
6) Ausonius, <http://petrae.huma-num.fr/fr/> (PETRAE), with 129 inscriptions in the time frame;

22 At the time of writing this article in the summer of 2018 EDH had developed the set of resources now available at <https://edh-www.adw.uni-heidelberg.de/data> which include a ↗DTS ↗API, a ↗SPARQL endpoint and several other data access options which I did not use here. See Grieshaber 2019.

23 Respectively <https://www.europeana.eu> and <https://www.trismegistos.org>.

24 The EAGLE datasets were either exported in ↗EpiDoc from databases or transformed from existing EpiDoc formats: details of this process are given in Liuzzo et al. 2014 and Liuzzo et al. 2015.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

7) Roman Inscriptions of Britain, <https://romaninscriptionsofbritain.org> (RIB), with 20 inscriptions in the time frame, from the collection of stone inscriptions from Britain;
8) Last Statues of Antiquity, <http://laststatues.classics.ox.ac.uk/> (LSA), with 2000 inscriptions in the time frame, related to statue bases.

All these datasets (and many others) are searchable via the website of the contributing partners as well as via the EAGLE portal,[25] where each partner is also acknowledged and the available editions of inscriptions are grouped by Trismegistos identifier. Some overlap might be present, and one inscription could occur twice in the corpus, as the disambiguation possible in the website is not reproducible in a small sample. However, it is possible to disambiguate now via the Trismegistos ↗API, <https://www.trismegistos.org/dataservices/texrelations/documentation>, developed by Frederic Pietowski.

1.3 Comparison

The comparisons made do not take into consideration location: each of the datasets does have that type of information, but with different types of identifiers that mostly are not in ↗Pelagios.[26] In an ideal world, an ↗EpiDoc file should contain at least identifications of places which make use of the stable URIs like those provided by ↗Pleiades or Trismegistos which are part of the Pelagios Commons, but this is not always the case. So the aspect of location has been left for a future study.[27] On the other side, here the focus is really on the typologies and what can be seen by looking at their distribution. Removing the location element lets us focus on those typologies instead of simply counting inscriptions by language.[28]

One aim of quantitative comparisons of multilingual datasets, in general, can certainly be to set the epigraphic evidence from Ethiopia and Eritrea in a wider context,[29] to the extent to which this is possible. With all the caveats of the changing data and quality of data.

25 <https://eagle-network.eu>.
26 For details on this, see Liuzzo 2018. The project Latin Now, led by Alex Mullen, is working on the EAGLE data to improve it also in this respect.
27 Trying to find out relevant information online on a findspot called Edit (*RIÉ* 228–231) was for me an interesting experiment.
28 Prag exposes the limitations of such studies well where he says ‘such quantification of Latin in isolation can be problematic; […] simply counting the proportions of Greek and Latin inscriptions in a period and/or location should no longer be considered a viable way of determining the principal spoken language’ (Prag 2002, 15–16).
29 A method supported, especially for this period of epigraphic history, also by Bolle et al. 2017, 17, although the kind of contextualization offered here is rather different, as the context is defined by the data and is not the historical context.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

For the same reason stated above for space, the examples here also do not break down by date. Any inscription within the defined time span for the Aksumite period is counted without further specification. In this case, this is due to the fact that there is not enough data and not precise enough to see diachronic change at this level for the Ethiopic dataset. The diachronic analysis is one of the most important ones and the promising perspective of work for the observation of historical changes and, hopefully, more precise dating of the texts will become available and will be encoded, which will enable this kind of comparisons in the future.

Some general features of the epigraphical evidence from Aksumite Ethiopia can be already outlined, and show a discrete continuity with the general features of the epigraphic practices in Late Antiquity.[30] These are the following: the impact of Christianity as a propulsor of epigraphic production, the tendency to reuse epigraphic material, and the varied forms of execution,[31] although not studied in enough detail to date to allow a proper classification.

30 Bolle et al. 2017, 19–21.

31 e.g. *RIÉ* 200, 201, 214, 220, etc.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

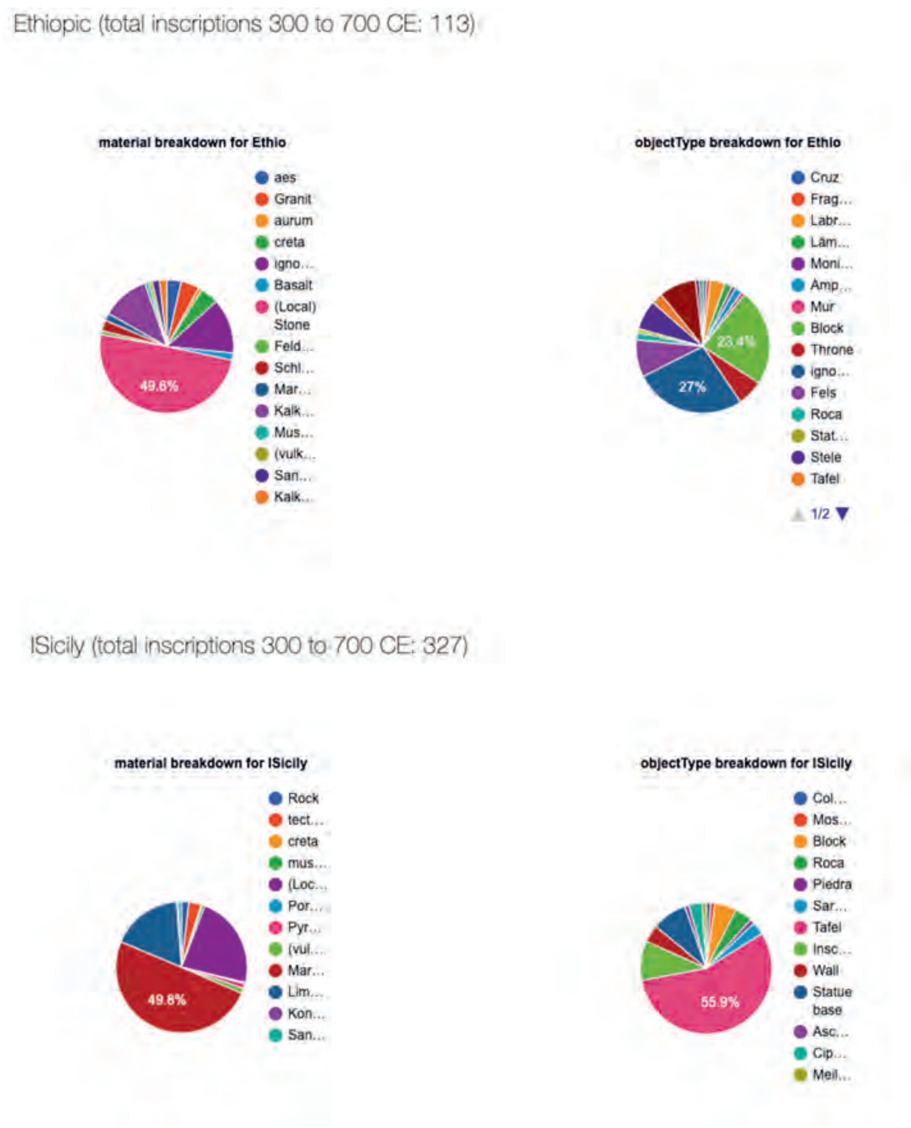


Fig. 2.1 Partial view of a set of pie charts generated to compare quantitative data. See <https://pietroliuzzo.github.io/DHEth/charts/charts.html>.

The charts in Fig. 2.1 are only few of the ones which are discussed here, and which are available in the GitHub repository connected to this book as simple ↗HTML pages. The images provided here are only meant to visualize

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

the charts under discussion which are in the website.[32] The labels in the charts here as in the following are the skos:prefLabel of the ↗URI of the reference concept. EAGLE applied criteria in establishing this label among the many possible for each concept, based on availability of definitions rather than on a choice about the main language of the vocabulary, so, it should come as no surprise that some labels are in German, some in French, and others in Spanish or Latin.

Now that the premises are set, we shall move on to try and see if something can be observed in these charts. There are up to four charts for each main dataset, and for each dataset in the EAGLE group there is a more specific breakdown to allow better consideration of what ends up in the big numbers of this ↗aggregator dataset, building forty-eight charts in total, minus the ones where no data is available.

Let us start from the following consideration: 'la peculiarità dell'epigrafia etiopica […] è quella di produrre iscrizioni solo in occasione di grandi eventi militari (ben diversamente da quella sudarabica, dove si commemorano spesso costruzioni di edifici o scavo di canali e cisterne)' ('the peculiarity of Ethiopian epigraphy [...] is that of producing inscriptions only in occasion of major military events (quite differently from South Arabian epigraphy, where often the construction of buildings and the excavation of canals are commemorated)')[33]. The typology distribution of the corpus is clearly marked even without quantitative confirmation, as Marrassini says. In the effort to compare this set to others, the first observation to be made is also an obvious one, and is not actually based on the data visualized: funerary inscriptions are by far the majority in all other corpora,[34] whilst in Ethiopia there are only very few,[35] although the existing ones document an alignment to the Christian epigraphy in the Mediterranean for the period.[36] Secondly, the distribution of object typologies with eighteen different types, with the generic 'block' nearing a quar-

32 <https://pietroliuzzo.github.io/DHEth/charts/charts.html>. The script producing this output is here <https://github.com/PietroLiuzzo/DHEth/tree/master/2%20Inscriptions/modules/piecharts.xql> and assumes data being stored in a ↗eXist-db ↗application in a specific collection where, together with the three main collections of data for comparison, also a copy of the EAGLE Vocabularies as ↗RDF-XML is stored. The result of the module is the ↗HTML page with several <script> elements with Google Charts ↗JavaScript code using data in arrays produced from the ↗XML.

33 Marrassini 2014, 10.

34 With the exception of the Inscriptions of Roman Tripolitania and of Last Statues of Antiquity. In this last case the totality of inscriptions are honorific.

35 *RIÉ* 274 (Fiaccadori 2003), and Fiaccadori 2007 (Fiaccadori 2007) for the Greek inscriptions, with few more examples like *RIÉ* 232 for the documentation in other languages.

36 Fiaccadori 2007, 71.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

ter of the total, is very high compared to ISicily for the same period, which has much less diversity with only thirteen types attested, where slabs (<https://www.eagle-network.eu/voc/objtyp/lod/257>) dominate without doubt. Interesting might also be the very similar percentage of inscribed panels (<https://www.eagle-network.eu/voc/objtyp/lod/259>) in both datasets. Thrones (<https://www.eagle-network.eu/voc/objtyp/lod/1996>) are a specific feature of the Ethiopian epigraphic production and they do not occur anywhere else. Actually the value for this type of object had to be added to the EAGLE Vocabulary. The EDH as well as the EAGLE data confirm the image of predominance of some objects given by ISicily, although in EDH 'sarchophagi' have a very relevant position which goes lost in the general EAGLE data, where the high number of EDB, with over 75 per cent of object types being *tabulae*, does not leave space for anything else. Here then the comparison with EDH data could tell us more about the contextualization of the inscriptions from Ethiopia, which are more similar to the picture given of the provinces of the empire, but the language focus and the presence of a considerable number of inscriptions from Rome in the dataset do not allow for such statement. Limiting the search to the provinces closer to Ethiopia (Aegyptus, Arabia, Cyrene, Iudaea, and Syria, see Fig. 2.2) the result are only a handful of inscriptions, thirty-nine at the date of writing, of which only twenty-four with information on the object type, where we note a very probably random similarity to the Ethiopian situation, in which 'block' has a share second only to milestones, the latter being also the dominant object in the nineteen texts from RIB when isolated from the EAGLE data.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

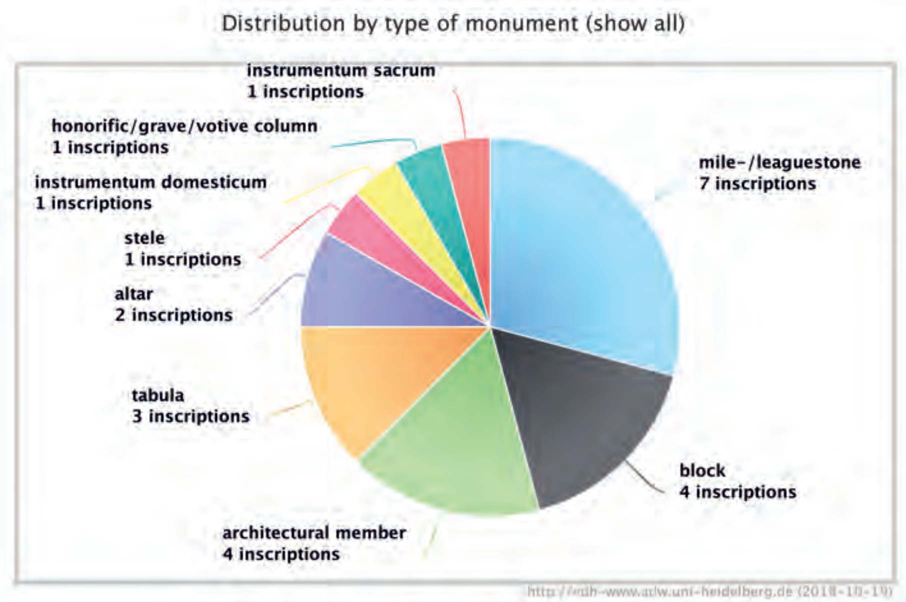


Fig. 2.2 Pie chart of the distribution by type of monument of inscriptions limited to the time frame 300 to 700 CE, for the provinces of Aegyptus, Arabia, Cyrene, Iudaea, and Syria. Source: <https://edh-www.adw.uni-heidelberg.de>.

The comparison with the Trismegistos data from Egypt is more complex and much less clear. The much higher diversity of the data meant also a high level of normalization in the alignment of the information about typologies, which is available only for a subset of the texts. *Stelai* (<https://www.eagle-network.eu/voc/objtyp/lod/250>) are predominant as an object type while *tabulae* only have a secondary attestation.

The diversity of materials (number of different materials over total) for the inscriptions of Ethiopia might be deceiving, as information in the source is often vague and have led me to indulge in the use of ‘local stone’ (<https://www.eagle-network.eu/voc/material/lod/2>) as the most generic value to be picked where no better precision was given. The local provenance is unsure, as much as the name ‘stone’ is generic. Nevertheless, the same value, although possibly with a more precise choice behind it, can be seen as having a heavy role also in ISicily, second only to marble (<https://www.eagle-network.eu/voc/material/lod/48>). Marble, in turn, when looking at the EAGLE data, is the most used material due to the heavy impact of both EDR and EDB datasets, where it is the prominent material. Local stone is, however, the predominant material encoded in the data, when comparing to the Egypt-

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

ian data from Trismegistos. Interesting is to observe instead the similar percentage of unknown materials (<https://www.eagle-network.eu/voc/material/lod/138>) between the *RIÉ* inscriptions and EAGLE (13 per cent in the *RIÉ* inscriptions, 15 per cent in EAGLE).

Here the only hope is specialized autopsy with proper instruments, allowing to iterate with better data these charts. Even better would be to make this dynamically linked to the incoming data and configurable, in the context of the Epigraphy.info developments.[37]

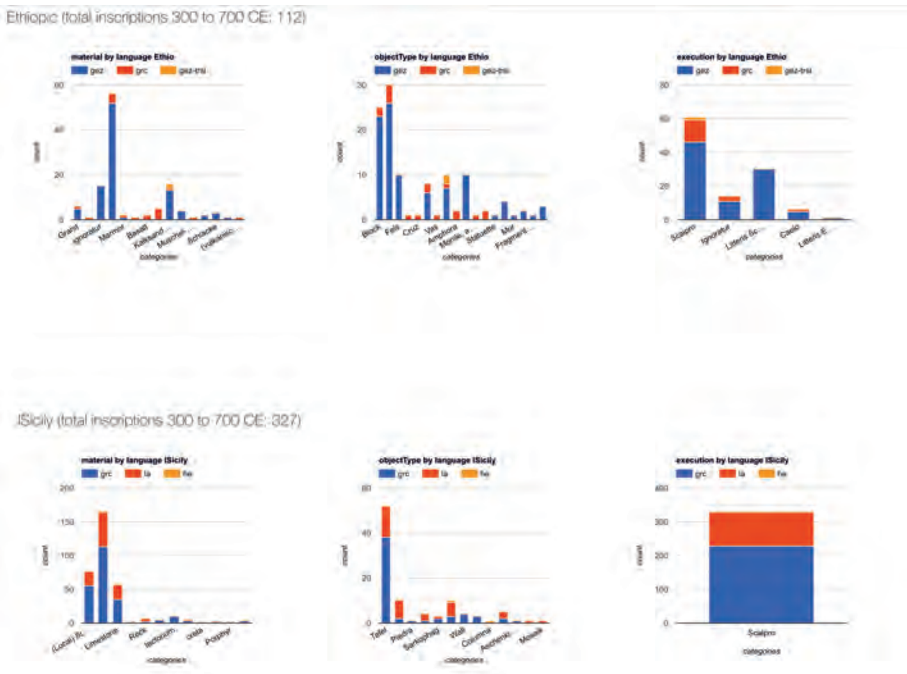


Fig. 2.3 Language breakdown for each category, partial view of series of graphs for comparison. See <https://pietroliuzzo.github.io/DHEth/charts/bylanguage.html>.

The charts in Fig. 2.3 add one more variable to each chart: language.[38] Each column is broken down into the number of inscriptions associated to any of the languages documented in the corpus in question. Also in this case the im-

37 See Feraudi-Gruénais and Grieshaber 2016; Feraudi-Gruénais and Grieshaber 2018 and Feraudi-Gruénais et al. 2019.

38 Charts types have been taken from Prag 2002. Here no account is taken of the interesting relation between script and language of the bilingual inscriptions, for example, Marrassini 2014, 37, nor of the absent, partial, or complete vocalization.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

age shows just a few of the charts discussed which are available in the GitHub repository connected to this book as simple ↗HTML pages.[39] There are types and materials which for this period are specific to a certain language. Greek inscriptions are on a distinct set of materials compared to Gəʿəz inscriptions. Gold, basalt, shell-limestone, and pottery are materials which only see Greek inscriptions.[40] Equally, the object types made of these materials confirm a specific relation with the language of the inscription they carry: vase, lamp, jewellery, altar, and cross only see Greek inscriptions. On the other side, the wider breakdown of the object types also shows that rock-hewn inscriptions are only in Gəʿəz, and so are inscribed plaques, basins, and several other types. We might consider this casual, but the comparison with the results of the same query on the data from ISicily confirms the impression that there is in this period at regional level a preference in the connection between language and object types. In Sicily, unsurprisingly perhaps, inscriptions on milestones, statue bases, sarchophagi, and rock-hewn inscriptions are only in Latin in this period,[41] whereas inscriptions on walls are only in Greek.

In the data from Egypt we can observe this situation, with several object types carrying only Greek inscriptions but there are also cases, like statue bases and socles, which only present Latin inscriptions, as well as cases in which a fair share for Coptic and Greek is to be observed: fifty-seven and fifty-three respectively for tomb (<https://www.eagle-network.eu/voc/objtyp/lod/447>), twenty-four and twenty-eight for temple (<https://www.eagle-network.eu/voc/objtyp/lod/5>), eight and eight for ostrakon (<https://www.eagle-network.eu/voc/objtyp/lod/124>). However, all of these could be due to double assignation to a language where this is not entirely clear.

This distribution of languages on different object types might have different explanations in the different contexts, of course, and the numbers are in all cases small, so that it would really be enough to find one more inscription

39 <https://pietroliuzzo.github.io/DHEth/charts/bylanguage.html>. The script producing this output is here <https://github.com/PietroLiuzzo/DHEth/tree/master/2%20Inscriptions/modules/bylangcharts.xql>. The same considerations made for the previous charts are valid.

40 Marble is an interesting case here, as the two inscriptions on this material are one in Greek, from ʾAdulis, where the impact of Hellenistic society was strong, the other is an inscription in Gəʿəz found in Ẓafār (capital city of the Ḥimyarite kingdom now in Yemen) which might witness another kind of phenomenon.

41 Here the table uses *roca* (skos:prefLabel of <https://www.eagle-network.eu/voc/objtyp/lod/211>) instead of *fels* (skos:prefLabel of <https://www.eagle-network.eu/voc/objtyp/lod/210.html>), apparently because of an ambiguity in the EAGLE Vocabularies which led to the choice of different terms for what is probably the same thing. However there is in the vocabulary an explicit relation between the two which allows their assimilation.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

to break the results. This would not be a big problem if we could quickly and dynamically update the conclusions together with the data and the visualization which we produce from them. At the moment, this is what the data shows and it might be interesting in the case of the epigraphic evidence from Eritrea and Ethiopia to highlight two distinct epigraphic practices.

The bilingual inscriptions are on three different materials: granit (*RIÉ* 185+270), porphyr (*RIÉ* 185 bis+270 bis), and calcareous sandstone (*RIÉ* 190+271). In the first and last cases there is a predominant use of this material for Gəᶜəz inscriptions, which might lead to confirm that the Greek language is 'guest' of an inscription of the Gəᶜəz production context, without necessarily denying that the most important text was the Greek and the Gəᶜəz only the local translation.[42]

To be able to compare also the EAGLE data, I tried to remove the columns with an overwhelming majority of entries, like marble (<https://www.eagle-network.eu/voc/material/lod/48>) and slab (<https://www.eagle-network.eu/voc/objtyp/lod/257>), but still the lack of language diversity in the datasets does not allow much to be said: it simply confirms the under-representation of any other language but Latin in the EAGLE data, and the unavailability of accessible and reusable resources for Greek epigraphy.[43]

Other charts are easily computable to breakdown each of the figures in the pie charts by each of the other typologies.[44] I am not going to discuss these here, however, because beside the obvious relations (i.e. some materials are bound to some object type, like pottery to vases) any other observation would need a much deeper knowledge of the documentation to be able to draw relevant observations.

More interesting is perhaps to see if the combinations of material and object type attested in the Ethiopian corpus are attested in other datasets. The following table summarizes the attested couples of material and object type in the inscriptions from Ethiopia and then lists each of the inscriptions which have these features in the corpus and counts if there is any other attestation in three comparison datasets, EAGLE, ISicily, and Trismegistos, taking in-

42 See the quote from Marrassini at the end of this section.

43 This is a very well-known problem, although the owners of the Greek inscriptions corpora have been cooperating with larger efforts, as, for example, the Integrating Digital Epigraphies Project at Duke Univeristy. See Cayless 2015 where also a video of the presentation is available. Although the major databases are not easily available for reuse as data, the number of excellent exceptions abound and include some of the most famous and pioneering projects in digital epigraphy, which are listed among the epigraphy projects in the Digital Classicist Wiki, <https://wiki.digitalclassicist.org/Category:Epigraphy>.

44 They are available in the website, <https://pietroliuzzo.github.io/DHEth/>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

to account twenty-seven pairs (the ones with an unknown value have been omitted).[45]

Table 2.1 **Comparison of the combinations of material and object type attested**

Material/Object	Ethiopic	EAGLE	ISicily	Trismeg.
(Local) Stone Block	20: *RIÉ* 201, 203, 204, 208, 212, 213, 214, 216, 222, 224, 225, 226, 227, 232, 234, 261, 262, 264, 268, 272	12	2	12
(Local) Stone Fels	10: *RIÉ* 196, 197, 228, 229, 230, 231, 233, 250, 251, 252	0	0	57
Kalksandstein Throne	6: *RIÉ* 186, 187, 188, 189, 190+271, 193	0	0	0
(Local) Stone Labrum	4: *RIÉ* 198, 199, 215, 219	0	0	0
Kalksandstein Stele	4: *RIÉ* 184 A, 184 B, 184 C, 184 D	0	0	0
Granit Block	4: *RIÉ* 200, 210, 211, 217	0	0	0
Schlacke Inscribed plaque	3: *RIÉ* 181, 182, 183	0	0	0
(Local) Stone Inscribed plaque	2: *RIÉ* 194, 207	0	0	0
Creta Lámpara	2: *RIÉ* 282, 283	0	0	22
Sandstein Block	2: *RIÉ* 287, 269	3	0	13
(Local) Stone Tafel	2: *RIÉ* 266, 267	179	2	2
(Local) Stone Roca	2: *RIÉ* 220 221	21	0	0
Aes Inscribed plaque	2: Gebresellasie 2017, 37 lower; Gebresellasie 2017, 37 upper	0	0	0
(Local) Stone Stele	2: *RIÉ* 218, 223	19	0	61
Creta Amphora	2: *RIÉ* 281, 285	9	0	55
Kalkstein/Alabaster Inscribed plaque	2: *RIÉ* 195, 265	0	0	0
Aes Statuette	1: *RIÉ* 184	0	0	0
Muschel-Kalkstein Monile, anulus	1: *RIÉ* 280	0	0	0

45 The materials as well as the object types can be in languages different from English, according to what is the preferred label in the ↗SKOS vocabulary.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Material/Object	Ethiopic	EAGLE	ISicily	Trismeg.
Kalksandstein Inscribed plaque	1: *RIÉ* 205	0	0	0
Marmor Tafel	1: *RIÉ* 263	18947	26	9
(Local) Stone Fragmentum	1: *RIÉ* 206	8	0	1
Aurum Cruz	1: *RIÉ* 279	0	0	0
Creta Vas	1: (*RIÉ* 284)	3	0	64
Basalt Stele	1: *RIÉ* 286	0	0	0
(Local) Stone Mur	1: *RIÉ* 209	0	0	0
Marmor Altar	1: *RIÉ* 278	23	0	0
(Local) Stone Throne	1: *RIÉ* 202	0	0	0

Fifteen out of twenty-seven combinations attested in the Ethiopic corpus are not found in the other datasets, marking a neat specificity. This table, where the individual texts are also listed, can be reorganized and cleaned up to contain only those cases for which there are features to be compared, namely the twelve combinations which are attested also outside of the Ethiopic corpus, and to translate in English the labels; but it is already quite clear that there is not much to compare, even if instead of the absolute number, we take into account the percentage of representation (Fig. 2.4).

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Table 2.2 **Inscriptions on same object and same material**

Material/Object	Ethiopia	EAGLE	ISicily	Trismegistos
(Local) Stone Block	20 (18%)	12 (0%)	2 (1%)	12 (1%)
(Local) Stone Cliff	10 (9%)	0 (0%)	0 (0%)	57 (4%)
(Local) Stone Fragment	1 (1%)	8 (0%)	0 (0%)	1 (0%)
(Local) Stone Rock	2 (2%)	21 (0%)	0 (0%)	0 (0%)
(Local) Stone Stele	2 (2%)	19 (0%)	0 (0%)	61 (4%)
(Local) Stone Slab	2 (2%)	179 (1%)	2 (1%)	2 (0%)
Pottery Amphora	2 (2%)	9 (0%)	0 (0%)	55 (4%)
Pottery Lamp	2 (2%)	0 (0%)	0 (0%)	22 (2%)
Pottery Vase	1 (1%)	3 (0%)	0 (0%)	64 (4%)
Marble Altar	1 (1%)	23 (0%)	0 (0%)	0 (0%)
Marble Slab	1 (1%)	18947 (59%)	26 (8%)	9 (1%)
Sandstone Block	2 (2%)	3 (0%)	0 (0%)	13 (1%)

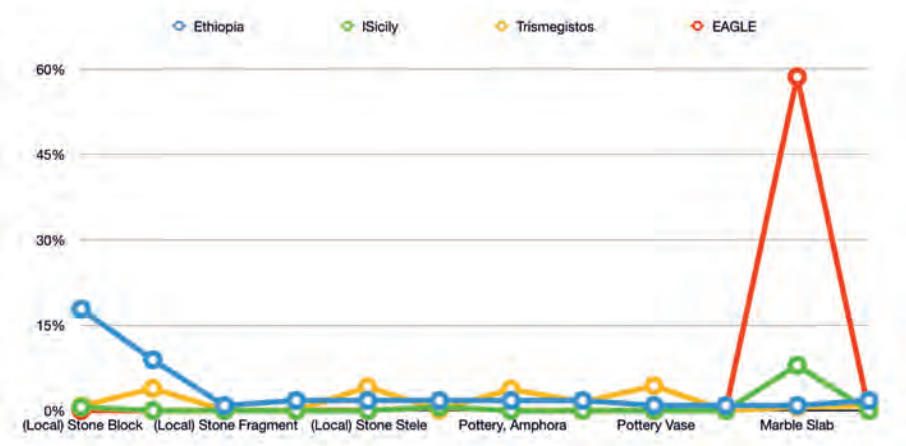


Fig. 2.4 Comparison of inscriptions with relevant combinations of material and object type.

We obtain instead something slightly more interesting when restricting to the inscriptions in Greek, the only language attested in all the datasets, as it can be seen in Fig. 2.5.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Table 2.3 **Greek inscriptions on same object and material**

Material/Object	Ethiopia	EAGLE	Trismegistos
Pottery Amphora	2 (2%)	5 (0%)	54 (4%)
Pottery Lamp	2 (2%)	0 (0%)	22 (2%)
Pottery Vase	1 (1%)	2 (0%)	64 (4%)
(Local) Stone Block	1 (1%)	2 (0%)	12 (1%)
Sandstone Block	1 (1%)	0 (0%)	13 (1%)

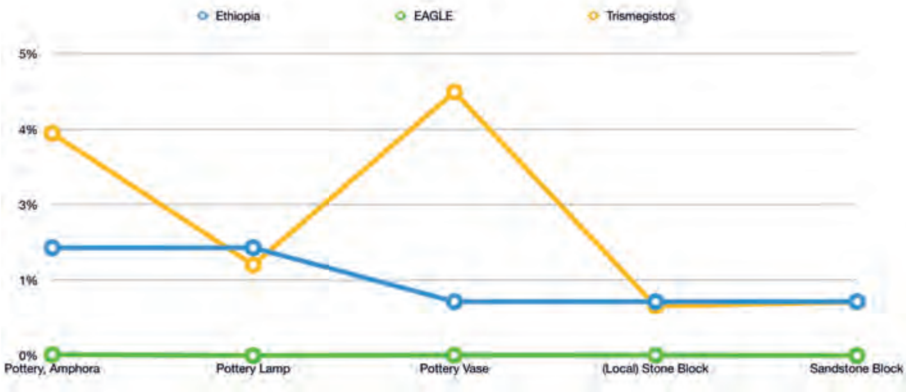


Fig. 2.5 Comparison of inscriptions in Greek with relevant combinations of material and object type.

No comparable items are present in ISicily, as expected, and the numbers of Greek inscriptions are too little in EAGLE to make the result of any relevance, but for three out of five of the types of inscriptions selected the percentage of the Greek inscriptions from Ethiopia is similar to those of the inscriptions in Greek from Egypt. This in a way might confirm what already Marrassini noted,

> Dunque, queste iscrizioni, in lingua geez ma in scrittura sudarabica, potrebbero costituire una categoria, e diremmo un 'genere epigrafico', a parte, che ha una sua linea evolutiva specifica dal III secolo […] al VI e che non ha niente a che vedere, dal punto di vista tipologico e funzionale, con quelle in etiopico vocalizzato: quelle in grafia sudarabica e in greco erano quelle 'vere', ufficiali; quelle in etiopico, le stesse—potremmo dire—spiegate al popolo.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

> (Thus, these inscriptions, in geez language but in South Arabian script, could constitute a separate category, we could say an 'epigraphic genre', a separate one, with its own evolution from the III century to the IV and which has nothing to do, from the point of view of typology and function, with that of the vocalized Ethiopic: those in South Arabian script and in Greek were the 'true' ones, the officials; the Ethiopic ones, were the same—we could say—explained to the people.)[46]

This is perhaps a meagre booty, as it could have been easily expected even without counting. But, considered the state of the data, I hope it is a good test, which will be hopefully repeated when information on the type of inscriptions will be available for the data about Ethiopia and the data about the execution technique will be available for the inscriptions recorded in Trismegistos. So, nothing could be more preliminary, even without mentioning the fact that the vocabularies also need a lot of work in order to be able to perform more intelligent associations. These associations could, for example, exploit the relations between the terms to group the values which are related to one another.

However, both the quality of the data and vocabulary, and the formats of the data are irrelevant when we face the major problem of data availability, which is evident when looking at the Greek inscriptions in EAGLE. It would have been nice to be able to observe this distribution also for types of inscription, or looking at structured patterns in the text of the inscriptions, but this information is not available in the data used at the moment. Also no mention has been made here of the contents. It is worth noting that, for example, also isolated official documents in manuscripts could be compared with inscriptions, since the text of the inscriptions is not stored only as transcription in the inscription record: formulas and, in general, diplomatic studies across documents on any material,[47] including parchment, could naturally be carried out.

2 Linking inscriptions

The curator of the Datenbank zur jüdischen Grabsteinepigraphik epidat (EPIDAT)[48] recently stated that 'the stable quality of data and sufficient quantity of temporal, spatial and gender related information contributed by epigraphy, is suitable for a wide area of inter- and trans-disciplinary research ques-

46 Marrassini 2014, 76.

47 As, for example, the one object of the study of the Ethiopic Manuscript Archives project, see Wion 2018.

48 EPIDAT is available at <http://www.steinheim-institut.de/cgi-bin/epidat>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

tions.'[49] We have seen that this statement could be extended to typology. Also in this sentence, the adjective 'stable' for the quality of the data is rather appropriate to describe a situation where we would like to have better data, but what we do have is already a good starting point. One could think that, when we lack so much data altogether, and of the published data the part which also has an online counterpart is so imperfect (although stable) one should focus on that, and not on yet another format for the digital data. But the push towards more digital resources, which will be able to support always better scientific results, which are reproducible and dynamically updated, is so urgent that discussing, although briefly, the work towards as much standardization and interoperation as possible seems to me necessary to complete this chapter.

As much as a standard encoding in XML helps the interchange of data, as shown in the previous section, so the use of stable URIs helps to connect resources in a stable and precise way. To navigate the connections between ↗URI-defined information which are part of a standard description with certainty, we need an ontology which formalizes the way in which these connections can happen and allows a discrete amount of reasoning to be possible also for the machine which knows what connections to expect and what type of information it will find in the node at the other side of that connecting edge.[50]

URIs for texts are provided by a good number of digital projects online, and services like those offered by Trismegistos allow us to disambiguate identifiers and provide a stable Trismegistos Text Identifier, in the form of a stable ↗URI. Places can be identified by pointing to the stable URIs of ↗Pleiades or other relevant gazetteers such as Beta maṣāḥəft for ancient places in Ethiopia. We have also seen the EAGLE Vocabularies which provide stable URIs for a series of concepts used in the classification of inscriptions and GODOT (Graph of Dated Objects and Texts) will provide stable URIs for ancient calendar dates.[51] Persons are also often identified by a URI in general or specific authorities and Beta maṣāḥəft maintains a list of URIs to identify people related to Ethiopian history and Ethiopian studies, for example.

However, there is no shared epigraphic ontology which would allow one to interoperably serialize ↗RDF triples which make use of these ↗URIs. There are tools to produce different kinds of RDF and many projects which have produced RDF and Linked Open Data (↗LOD) for epigraphic records.

49 Kollatz 2018, 235.

50 More on ↗RDF and Linked Open Data in Ch. 7.

51 <https://godot.date/home> is a project led by Frank Grieshaber in Heidelberg which will become the reference for URIs 'of calendar dates in different calendar systems used in the Greek and Roman antiquity all across the Mediterranean sea'. See Rabinowitz forthcoming.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

That of the EPIDAT and Deutsche Inschriften Online (DIO)[52] projects is one of the most recent efforts in this direction,[53] mapping the information in a script to produce triples on the fly. Also the CIDOC-CRM special interest groups and the EAGLE project have done some work about this and are certainly not the only ones to have done that.[54] I will briefly describe what is available to date, to contribute with this collection to the possible development of a shared epigraphic ontology in the near future.[55] This ontology will need to start from taking into consideration existing cookbooks and specifications and formalizing those first to be application-oriented and not just a theoretic exercise.

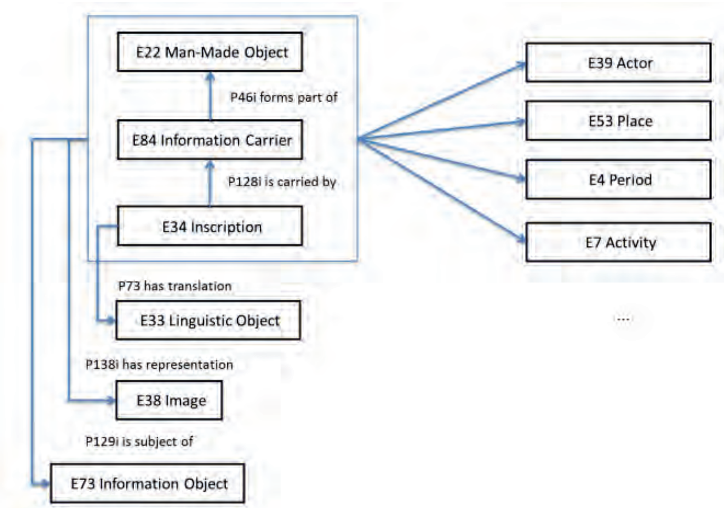


Fig. 2.6 The EAGLE high-level model for inscriptions (artefacts).

The EAGLE project had theorized a structure for the ↗RDF data which was generated by the project but never made it into a public and shared ontology, to my knowledge.[56] This structure had a high-level model (Fig. 2.6)

52 <http://www.inschriften.net/>.

53 Kollatz 2018, 237. See <https://digicademy.github.io/DARIAH-GT/> and <https://digicademy.github.io/dariah-workshop2016-xtriples-cidoc/>.

54 For example, see the Epigraphische Datenbank Erlangen-Nürnberg, <http://wisski.cs.fau.de/eden/>, developed with <http://wiss-ki.eu/>, a scientific communication infrastructure based on ↗CIDOC-CRM.

55 Gabriel Bodard has been working on an effort in this direction for some time, and there is a dedicated mailing list available.

56 Manghi et al. 2015.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

and two sub-models for the 'object' or 'artefact' (Fig. 2.7) and the 'document' (Fig. 2.8). In the figures ↗CIDOC-CRM entities (prefixed by E and a number) and properties (prefixed by P and a number) are used.

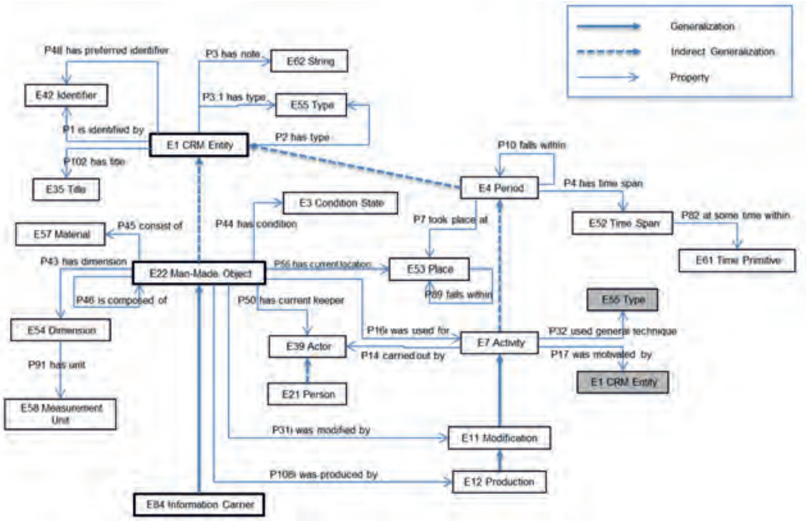


Fig. 2.7 The EAGLE model for objects.

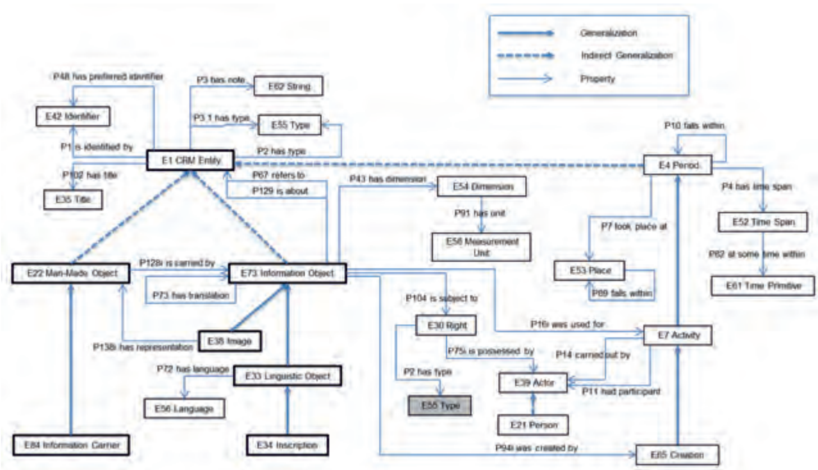


Fig. 2.8 The EAGLE model for documents.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

The categories used in the previous section (material, object type, execution technique) are mapped, according to the ↗CIDOC-CRM use suggested by EAGLE as follows:

1) Object Type: E55_Type is related to E22_Man-Made_Object in the high-level model via P2_has_type;
2) Inscription Type: E55_Type is related to E22_Inscription in the high-level model via P103_was_intended_for;
3) Execution Tecnique: E55_Type is related to E12_Production in the high-level model via P32_used_general_technique;
4) Material: E57_Material is related to E22_Man-Made_Object in the high-level model via P45_consist_of.

The mapping of EAGLE is complete but, although it was discussed by a large network of partners, it was never used for any concrete application beside EAGLE and especially it never made its way into a proper ontology remaining only a model in theory.[57]

The lack of a common alignment can be observed also in the DIO and EPIDAT mappings to ↗CIDOC-CRM done with XTriples conversions, which use different properties for the same kind of information, and the same purposes of data interoperation. The inscription is put in a ↗class E19_Physical_Object and then linked to the material of the support directly with the property P9_consists_of.[58]

Another independent move towards modelling triples to describe inscriptions has been made by Achille Felicetti and Francesca Murano,[59] who have pointed out how in ↗CIDOC-CRM, E34_Inscription is not adequate and proposed a series of new classes and properties (Fig. 2.9) like EPI1_Epigraph defined as 'subclass of E25_Man-Made_Feature intended to describe a particular feature created by humans in various ways and on various kinds of support with the declared purpose of conveying a specific message towards a given recipient or group of recipients'.

57 On data modelling see Flanders and Jannidis 2016.

58 See <https://digicademy.github.io/DARIAH-GT/data/xtriples_config_dio.xml> and <https://digicademy.github.io/DARIAH-GT/data/xtriples_config_epi.xml> for example configurations with which the EPIDAT and DIO datasets have been mapped and converted to ↗RDF using XTriples (<http://xtriples.spatialhumanities.de/index.html>).

59 See Felicetti et al. 2015a; Felicetti and Murano 2016; and Felicetti and Murano 2017.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

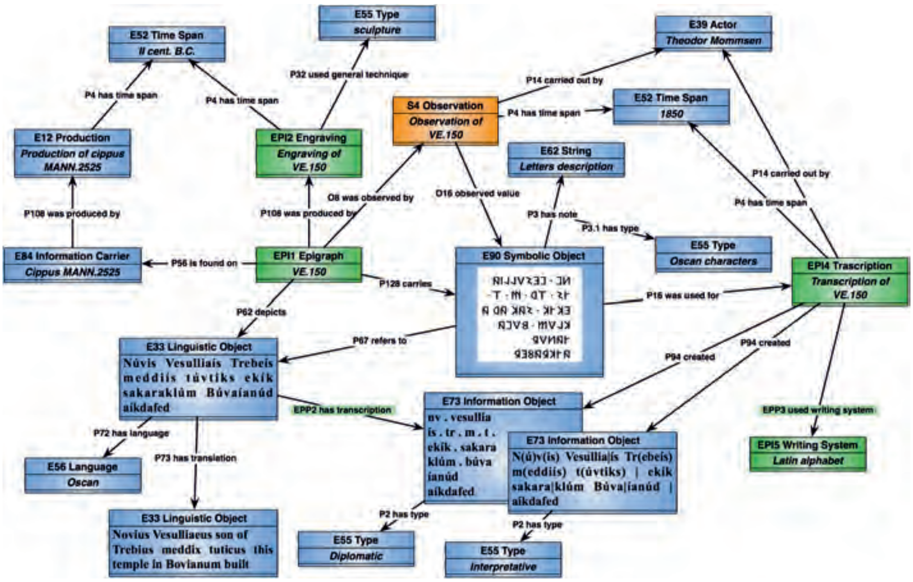


Fig. 2.9 Model of CRMepi from Felicetti et al. 2015b, slide 22.

A complex but understandable modelling of the information, in this new set of classes and properties of CRMepi would lead to classes like EPI2_Engraving, which could be further expanded with the proposal by Evangelisti on the classification and description of execution techniques.[60] The same authors have also proposed an extension of the ↗CIDOC-CRM for texts, CRMTex,[61] which further generalizes the concept EPI1_Epigraph into TX1_Written_Text,[62] in very much the same way in which in this volume I have made use of the concept of written artefacts to subsume inscriptions and manuscripts.[63] This model was based also on discussion in some EAGLE workshops and could be considered as a real successor of the EAGLE model aimed at a larger interest group, but to date it has no application known to me.

60 Evangelisti 2017.

61 <http://www.cidoc-crm.org/crmtex/>.

62 Felicetti and Murano 2017, 4–6.

63 Although having a type 'Epigraph' would make the life of many much easier, it is perhaps not strictly necessary. In the new updates, the authors also model textual features like gaps and abbreviations.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

The approach of the Ontology Based Data Integration (OBDI)[64] starts from the same premises but the ontology used to map the concepts is, as in the case of the XTriples example above, specific to the task and not a shared effort from the community. As a methodology it is a wonder of efficiency, as it does not need any update and works independently of the data and its format so that the users see only the benefit and have no stress.[65] This is the most promising methodology applied to date to epigraphic data integration, and it is to be hoped that it will not remain confined to the data of the Economic & Political Network (EPNet) project.[66] Having been developed for the amphoric inscriptions of the CEIPAC database,[67] the model is very specific to this dataset. The EPNet project implementation uses `E34_Inscription` disliked by Felicetti and Murano, linking it to `E84_Information_Carrier` and `E33_Linguistic_Object` as in the EAGLE model, but then expands on it with a set of new classes and properties specific to the EPNet CRM, in a similar way as the previous proposal does in the CRMepi. This, however, largely reflects the need of this very specific kind of written artefacts.

These four examples, which by no means cover all efforts gone in this direction, and, to a large extent, do not overlap with each other, attest for how different the approaches to modelling the nodes and edges needed to describe an inscription can be.

It needs to be said that, where there is a common node or edge, this might still be reused, but the knowledge necessary to query the different bases would be as much as the one needed to reuse different ↗APIs or different data formats all together, with the only advantage of using one query language.

One main problem, when looking at the use of such models, is their actual application. Not many projects have adopted any of these propositions, probably because the purpose of such efforts was not very clear or too specific to the project application needs, thus avoiding the lengthy and troublesome process of agreeing on a common model. Comparisons and computation can be easily done on the XML so a case needs to be made for ↗RDF, and, since there is already a standard source format, ontologies can run wild and free without too much to worry about.

There are, however, other advantages to producing ↗RDF which conforms to a shared ontology, as I will try to show in Ch. 7, which make it desirable even if people do not like the idea of sharing that data openly. One main reason is the possibility of seamlessly joining sets of information. For example,

64 Discussed in Calvanese et al. 2016.

65 See <http://www.roman-ep.net/wb/home/> and <http://www.romanopendata.eu/>.

66 <http://www.roman-ep.net/wb/>.

67 <http://ceipac.ub.edu/index_en.html>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

if a researcher has collected a list of identifiers and only those, and another one has only metadata or only texts, but all of them produce a serialization of their data according to the specification of the shared ontology, they will not even need to exchange the datasets and integrate them, they will be already implicitly linked and would just need to be queried together.

It has never been possible to directly enrich a specific dataset with data from other datasets, but, now that it is an option on the table, this would enable the suppression of conversions, transformations, and all the steps which break the continuity of information. A comparative approach as the one in the previous section could be really offered dynamically and update itself on the fly, eventually sparing the researcher who does not want to do that, the time to set up all the data and the scripts. Linguistic, chronological, and spatial boundaries would be superseded opening up an entirely new range of possibilities. But this is not yet there, although it looks one step from happening, and hopes reside with the group of people around IDEA, Epigraphy.info, and the epigraphic ontology groups to make this happen.[68]

3 Conclusion

'It is desirable, especially in regional studies, to collect the epigraphic evidence in a comprehensive way, including all different types of inscriptions and not preferring certain groups like the Christian *tituli*'[69]. This is a very important statement, which has hopefully found support in the consideration made in this chapter when trying to show methods to contextualize quantitatively the epigraphic evidence from *RIÉ*. However, it is also evident that nobody can limit interests, focus, and scope of research and there will always be specific and selective corpora, even when it is clear how the selective criteria fail to be such.[70] This is, however, the kind of issue which standards and protocols, at least in the digital realm, can solve. Digital corpora of inscriptions with diverse scope and aim, diverse structures and interest *can* and *have been* used in an interoperable way so that the specific and the general needs are both served. To be able to do this, the research interests need not be changed, and

68 An initial mission statement from the Epigraphy.info group can be found in the report from the meeting held in Zadar in December 2018, Feraudi-Gruénais et al. 2019.

69 Bolle et al. 2017, 18.

70 Roueché and Sotinel 2017, 508–511 especially, although the entire article argues very clearly and convincingly about the need of putting all inscriptions 'to be plunged again into the sea of all the inscriptions of the epoch in question'. Ogereau 2018 in his review points out that dismissing these selective corpora 'downplays the actual distinctiveness of such inscriptions' and, although in light of the previously quoted article it is exactly this distinctiveness which is problematic, his argument shows how not even clear historical evidence can impede the application of fallacious selection criteria.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

instead some attention needs to be put to the standards in play, data curation with ↗EpiDoc, disambiguation via Trismegistos, serialization as Linked Open Data according to a shared ontology and common vocabularies, and distribution via ↗DTS and ↗IIIF ↗API specifications (respectively for texts and images) being the main points which cannot be omitted any more by any digital epigraphic project.

It is quite disappointing, and a lot more needs to be done, more data and of better quality in terms of its distribution and encoding, more tools and support for protocols. In this chapter there are probably not very striking results, but hopefully somebody will iterate this, or even better, it will be possible to make this dynamically linked to the incoming data, in the context of the Epigraphy.info developments which will enable diachronic evaluation in the 'sea of all inscriptions'. In the next chapter I will look at another type of entities in Beta maṣāḥəft, textual units, this time in order to read them as a whole with the support of encoded relations with keywords and periods.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Chapter 3. Textual Units Trees and Charts

> La perfezione, dice un antico proverbio orientale, è bella ma è stupida: bisogna conoscerla ma romperla. Adesso che, come penso, vi sarà chiaro come disegnare un albero, non dovete seguire pedestremente quello che vi ho mostrato […].
> (Perfection, says an ancient oriental saying, is beautiful but stupid: you need to know it and break it. Now that, as I believe, it is clear to you how to draw a tree, you do not have to follow pedestrianly what I have shown you.)
> (B. Munari, *Disegnare un albero*, 1978)

Computational methods offer a great deal of possibilities, but it is without a doubt that the main issue of understanding a text and putting it into its context is not something which can be done with a program. Understanding a text to be able to present it with a commentary and a translation requires deep reflection on the text and a wide range of knowledge as well as a high number of painful decisions to be taken which no machine learning will ever be able to teach a computer to replicate. It is also a non-finite process by virtue of which new interpretations are always possible. Imprecision is sometimes all an author can offer and, on the other side, allusions, analogies, and associations build towards arguments which are sometimes much more productive and thought-provoking than precise computations to be compared. What computational method can do is to help out in many stages of such journeys.

In the previous chapters I have described and used for analysis XML trees encoded in ↗TEI, namely hierarchical descriptions of written artefacts which can be reused for presentation and study purposes. Taking these trees for granted and moving in this chapter to an analysis of the texts, these trees will be broken and restructured in other kinds of trees in an attempt to follow the advice of Munari, reproduced in the epigraph to this chapter.

Being able to formulate queries with creativity is an additional skill which can support original and creative thinking, without being a necessary component of this latter. This skill needs to be based on good knowledge of the data, which expert researchers have been able to encode using their expertise of the subject. A scholar has typically both these skills, namely good knowledge of

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

the subject and the ability to question it. A digital scholar perhaps has the additional skill of making these queries in formal languages on the basis of formally defined data.

Producing data which can reflect the complexity of a literary tradition and give it an organization, without failing it, is vital to this process. To begin with, the data architecture representing a literary tradition should be able to identify unequivocally the units which constitute it.[1] There are several stages of this process which, from an empirical point of view, require different decisions to be made when encoding which need to be supported by the architecture of the data. Both the encoding and the architecture should be as simple as possible, but also as precise as possible.

In this chapter I will first offer a description of the way in which textual units are identified and related in Beta maṣāḥəft and I will then propose an overview of the current data which takes into consideration the architecture we have opted for and the information currently stored in it.

1 Defining units

McGann wrote that, 'As we lay foundations for translating our inherited archive of cultural materials, including vast corpora of paper-based material, into digital repositories and forms, we are called to a clarity of thought about textuality that most people, even most scholars, rarely undertake'[2] which for us here means that it is not enough, although very important, to distinguish among written artefacts and works, meaning with the former the description which one would find in a catalogue of the object and its contents, and with the latter what would end up in a critical edition of an abstract entity which has several manuscripts as witnesses. However, this generic distinction can host much more refined ones following a fairly easy and understandable set of principles to apply to record them. In Beta maṣāḥəft, for example, these two types of resources, namely manuscripts and textual units, host a variety of other semantic distinctions which the editors encode following the criteria outlined in the Beta maṣāḥəft Guidelines.[3] In Ch. 1 and Ch. 2 I have discussed written artefacts and in the Introduction I have given also a brief overview of how to define them,[4] let me now give some more details on the working definition of textual units in this context.

1 This problem is the subject of Orlandi 2018.

2 McGann 2007, § 2.

3 <https://betamasaheft.eu/Guidelines/>.

4 Cf. p. xli.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

1.1 Written artefacts and textual units

The traditional distinction between 'text-as-witness' and 'text-as-opus'[5] is a common and practical architecture solution for online databases. It is also the most important one, because it reflects the two main research interests targeted by a research environment like Beta maṣāḥəft, that is, on one side, the philologist, linguist, historian, and so on more interested in the contents; and, on the other side, the codicologist, book restorer, art historian, among others, more focused on the artefact. This is a generalization, but I think it is a useful one in this case. Because of these different research interests, the characteristics of the ↗TEI files, all of which validate to the same Beta maṣāḥəft ↗schema, are slightly different and contain the information relevant for either of the general ideas. The manuscript TEI files will contain physical description, transcription, shelf mark, and any other information which would typically be found in a manuscript catalogue. On the other side, the TEI file of the 'text-as-opus', which we call 'work' for the sake of simplicity in our guidelines and documentation, will contain statements about the composition of the work, the authorship, the witnesses, and the critically edited text with its own encoded apparatus.[6]

To facilitate the distinction between these two types of information in the XML we use both different IDs (prefixed with 'LIT' for abstract works) and a `@type` at the root element. Also in the parallel ↗RDF representation these two concepts are distinguished by assigning them to different classes,[7] for example `lawd:AssembledWork` for manuscripts and `lawd:ConceptualWork` for works.

The identification of a manuscript is in general an easier task in the context of the project Beta maṣāḥəft compared to the identification of a literary work. This is due, among other reasons, to the fact that the project also aims at producing a *Clavis Aethiopica*.[8] It is therefore crucial to have an idea as clear as possible about how to identify what constitutes an abstract work.[9]

The necessary definition,[10] which is intended as a working definition, is based on the core distinction made by Orlandi between textual units and nar-

5 Andrist 2018, 135–136.

6 Kropp 1994, 125 already said that 'the use of a mark-up language for a free text data base is obvious', although in his remark Kropp was pointing to the earlier standard used by ↗TEI, i.e. ↗SGML.

7 See Ch. 7 for the ↗RDF data and Liuzzo forthcoming a for the identification of Units of Production and Units of Circulation according to Andrist et al. 2013.

8 See p. xxii.

9 Cf. p. 207 for the ways in which textual fluidity is represented in the ↗RDF.

10 <https://betamasaheft.eu/Guidelines/?id=definitionWorks>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

rative units,[11] and on the simple principle stated by Bausi that a work 'only refers to fixed contents and a precise sequence'[12] (see below). However, if it can be more intuitive to make the distinction between narrative and textual units as suggested by Orlandi, the application of both these principles in practice, as observed during the project, is not enough for the encoder who has to choose the correct encoding option to represent the information at hand. I reproduce here the principles we use in our guidelines.

1) Texts provided with an independent circulation get an individual record with a first-level ID (such as CAe 1931, ID: LIT1931Mashaf), all other are text parts.
2) Texts which typically form part of other texts are equally assigned an independent first-level ID/record if they also have an independent circulation. For example, the anaphoras contained in *Maṣḥafa qəddāse* (CAe 1960, ID: LIT1960Mashaf) all have their independent ID and a derived ID in LIT1960Mashaf: the independent abstract work *ʾAkkʷateta qʷərbān za-qəddus Həryāqos za-hagara Bəhnəsā* (CAe 1099, ID: LIT1099Anapho) is what inside *Maṣḥafa qəddāse* (CAe 1960, ID: LIT1960Mashaf) is identified with the @xml:id AnaphoraMary.
3) For these texts (both circulating as independent units and part of another unit), a <div> should also be created, with a @corresp giving the individual ID, in order to show the full content of a collection in correct order.

```
<div type="textpart" subtype="chapter"
          corresp="LIT1099Anapho"
          xml:id="AnaphoraMary">
     <label>Anaphora of Mary</label>
</div>
```

The text is then identified as textual unit but also as part of the text where it is contained with reference to that. This corresponds to making a statement as

```
<relation active="LIT1960Mashaf"
               name="saws:contains" passive="LIT1099Anapho"/>
```

11 Orlandi 2013.

12 Bausi 2010, 34, n. 3.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

You can then in a manuscript record decide what is more appropriate to point to, if it is better to point to the independent unit directly or if this is in a manuscript as part of another unit. See for example CAe 1960, ID: LIT1960Mashaf#AnaphoraMary in Frankfurt, Universitätsbibliothek Johann Christian Senckenberg, Frankfurt Ms. or. 15 and CAe 1099, ID: LIT1099Anapho in Vatican City, Biblioteca Apostolica Vaticana, Aeth. 15.

4) If a textpart has an individual circulation in traditions other than the Ethiopic, all information on this text should be given in the <div>: multiple titles can be inserted, with bibliography linking to the IDs of these titles, relations for the authors of the titles, etc.
5) For text portions (paragraphs, chapters, miracles, etc.) which are extant as different versions in multiple recensions of the same work or even in different works, a <Narrative Unit> is created.

[…] Often the first problem to face is which records and how many to make. We accept as a principle to create *the maximum number of records that is sensible* and to state *the minimum number of relations needed among them* with a view on expanding them potentially to all that is not essential, in order to clarify their relation with each other.[13]

Each of these 'works' is associated, when possible, to the corresponding identifiers in other repertories,[14] as in the following example from the *Hymn for the Transfiguration of Christ on Mount Tabor* (CAe 4410, ID: LIT4410Transfiguration):

```
<listBibl type="clavis">
    <bibl type="CC">
        <ptr target="bm:CC"/>
        <citedRange unit="item">0215</citedRange>
    </bibl>
    <bibl type="CPG">
        <ptr target="bm:CPG1"/>
        <citedRange unit="item">3939</citedRange>
```

13 Emphasis in original.

14 The aligned repertories are at the moment the *Clavis Coptica* (CC), the *Clavis Patrum Graecorum* (CPG), the *Bibliotheca Hagiographica Graeca* (BHG), the *Bibliotheca Hagiographica Orientalis* (BHO), the *Clavis Apocryphorum Veteris Testamenti* (CAVT), the *Clavis Apocryphorum Novi Testamenti* (CANT), Kinefe-Rigb Zelleke 1975 (KRZ) and Hammerschmidt 1987 (H). Cf. p. 41.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

```
    </bibl>
    <bibl type="BHG">
        <ptr target="bm:BHG"/>
        <citedRange unit="item">1984</citedRange>
     </bibl>
</listBibl>
```

This information, encoded as bibliography because these repertories are available only as printed books, contains a further layer of abstraction of a 'work'. This level is not yet modelled in the ↗RDF produced by the project in such a way that would allow the exploitation of it,[15] although the ↗SAWS Ontology serves us well in that direction,[16] describing links between textual units which we can use to specify what the actual relation is to the degree of precision we know. We register which are the correspondences between the identified texts, but not the exact relation between the concept defined by each of these *clavis* numbers.

In this case the data has been inherited from that exposed by the PAThs project,[17] with a simple interface which, given a string from a title or direct input, queries for it the PAThs ↗API and returns possible matches which the user can or can not add to Beta maṣāḥəft. The information related to these identifiers is also available via a dedicated API query to the Beta maṣāḥəft data API and in the ↗DTS API as part of dts:extentions.[18]

```
"dts:extensions" : {
   "crm:P1_is_identified_by" : [ "CC 0215", "BHG 1984", "CPG 3939" ],
   "saws:formsPartOf" : "https://betamasaheft.eu/LIT3178Deggwa"
     }
```

With these identifications and relations to other entities we can then define a network of textual units which I will detail in the following sections.[19]

15 In Beta maṣāḥəft all these are stored in the same type of ↗triple, related to the main ↗URI with a crm:P1_is_identified_by.

16 Roueché and Lawrence 2014.

17 <http://paths.uniroma1.it/>.

18 See <https://betamasaheft.eu/apidoc.html#getclavis> and the full ↗DTS collection endpoint view for this work at <https://betamasaheft.eu/api/dts/collections?id=urn:dts:betmas:LIT4410Transfiguration>.

19 See Witt 2018 on the importance of clearly separating data and applications presenting it. Beta maṣāḥəft being based on the capabilities of ↗eXist-db to deliver the data in a front-end web ↗application, is not a pure ↗API consuming application, but tries to use

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

1.2 Textual units and narrative units

Orlandi introduces a distinction between textual and narrative units, which we adopted in Beta maṣāḥəft, although with less practical use at this stage of the project. Orlandi starts from the consideration that 'the *textual units* correspond to what is generally called "the works" (*opus, oeuvre, Werk, opera*). They are identified in modern scholarship by means of author and title […], but also specifying the *literary genre*'[20]. He then argues on how this is not fitted for the Coptic tradition, where neither authors, nor titles or genre can be taken into account to identify such units. It is in this sense that in Beta maṣāḥəft we use the term textual units to refer to what is more generally named as 'work', to say that they are not so defined and can have multiple titles and multiple authors, known or attributed.

Even more innovative and important is the concept of 'narrative unit' proposed by Orlandi and inherited in Beta maṣāḥəft as a core reference concept. These are 'independent from the *textual units* of which they are part'[21].

In Beta maṣāḥəft however we extend the practical use of these narrative units, and define a unit which is only about form and general content and not one exact sequence. In this way a given type of miracle or a typical event in a life of a saint can be assigned a narrative unit.[22] If textual units must have, by definition, a fixed content and a precise sequence, narrative units do not have these requirements and allow for a constructive fuzziness which enables researchers to point to an ideal text which does not have fixed contents in a fixed sequence.

One practical use of this distinction is, for example, to associate narrative portions in different recensions of a text.[23] It is also helpful whenever in the identification of a fragment it is not possible to decide to which textual unit it might have belonged, for example in the absence of any name.

Together with the manuscripts and their structure and the person files in Beta maṣāḥəft, these two types of entities defined in this way let us build an architecture which, rather than forcing the literary tradition of Ethiopia and Eritrea into other more common schemes, strives to represent its actual

the concept and push the limits of the medium by opening up all its data and code in as many non-idiosyncratic ways as possible and encouraging the reuse and alternative display of the data.

20 Orlandi 2013, 91, emphasis in original.

21 Orlandi 2013, 93, emphasis in original.

22 There are no real world examples yet of this use due to the project priorities and time line.

23 See Villa 2018.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

features. The way in which this is reflected into the data is by using relations which are encoded in the ↗TEI files and have a specific semantic value. The identification of a content in a manuscript is realized by assigning an identifier which is specific to the part of the manuscript and the textual unit linked to it, and the textual units are related to other textual units with further statements with defined names.

In the last example of the previous section you can see one of these relations, which is about the fact that the abstract work with identifier CAe 4410, ID: LIT4410Transfiguration (*Hymn for the Transfiguration of Christ on Mount Tabor*) is part of (saws:formsPartOf) another abstract work with identifier CAe 3178, ID: LIT3178Deggwa (*Dəggʷā*).

I believe, however, that it is easier to see and understand these working concepts in practice, so, to have an idea of the complexity of these relations only, I have uploaded to ↗Palladio the data obtained with the following ↗SPARQL query from the Beta maṣāḥəft SPARQL endpoint.[24]

```
1  PREFIX bm: <https://betamasaheft.eu/>
2  PREFIX ecrm: <http://erlangen-crm.org/current/>
3  PREFIX saws: <http://purl.org/saws/ontology#>
4  SELECT DISTINCT ?sTitle ?pTitle
5  WHERE {
6     {?s saws:contains ?p}
7        UNION
8     {?s saws:formsPartOf ?p }
9        UNION
10    {?s ecrm:CLP46i_may_form_part_of ?p }
11       UNION
12     {?s saws:follows ?p }
13    ?s dc:title ?sTitle .
14    ?p dc:title ?pTitle .
```

Since in Beta maṣāḥəft there can be textual units for which a relation has been stated which says that a textual unit contains another one (saws:contains) and there are cases in which, instead, there is a relation stating that a textual unit is part of another one (saws:formsPartOf), and I cannot assume that they are both there, because this depends on the progress of the work of the project, in Lines 6–8 of the query I am selecting statements of either kind. To

24 <https://betamasaheft.eu/api/SPARQL/json>. Any query to the ↗RDF data in Beta maṣāḥəft can be visualized like this. See Ch. 7.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

increase the number of relations I have added to these also units related by ecrm:CLP46i_may_form_part_of and by saws:follows. The variables ?s and ?p will be the ↗URI of the units related in this way and in Lines 13–14 I am asking for their English title (dc:title) instead of the URI, so that the table will have that instead of the URI, as requested by the variables in Line 4.[25]

Fig. 3.1 Part of the graph obtained querying texts related with saws:contains, saws:formsPartOf, ecrm:CLP46i_may_form_part_of, and saws:follows as visualized in ↗Palladio.

The unreadable cloud (Fig. 3.1), generated by setting the graph view of ↗Palladio to link ?sTitle and ?pTitle, that is, the titles of the textual units which are respectively subject and predicate in the relations selected, is interesting especially where the clusters of texts are in contact with each other.

This cloud could be filtered by adding some more data in the query in order to use it in the facets of ↗Palladio. For example, we could save the type of relations as a variable in the results. In the images below you can see a zoom

25 This will have one row for each of multiple titles. One could filter this in the query, for example, by restricting to the English titles FILTER (lang(?sTitle) = 'en') or by using the ↗URI instead of any title. Another major exception is the identification of parts of a textual unit with other textual units. These in the ↗RDF are related to the part of text, so that it is the part of textual unit to have a relation with the independent textual unit, and not directly the main textual unit. This can be added to the above query, with the following ?sPart dcterms:hasPart ?sSubPart ; dcterms:isPartOf ?s . ?sSubPart dc:relation ?p . ?p a lawd:ConceptualWork . taking care to add the necessary prefixes to the query as well. See for instance <https://betamasaheft.eu/rdf/works/LIT5673AkkonuColl.rdf>. A basic request for textual units related with these properties is also possible starting from <https://betamasaheft,eu/litcomp> which presents a table with information as requested by Wendy Belcher for the PEMM project.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

on a cluster filtering only the units which *might* form part of other units, namely ecrm:CLP46i_may_form_part_of (Fig. 3.2), and the same cluster looking only at certain statements, namely saws:formsPartOf (Fig. 3.3).

Fig. 3.2 Part of the graph obtained querying texts related with ecrm:CLP46i_may_form_part_of as visualized in ↗Palladio.

Fig. 3.3 Part of the graph obtained querying texts related with saws:formsPartOf as visualized in ↗Palladio.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Here it is easy to see how a hierarchical definition cannot be applied, whereas the graph view supports the multiple relations among textual units which have been encoded as independent statements. The same is true for authorship. Rather than entering this information in a specific field, we simply make a statement by means of a relation named dcterms:creator or saws:isAttributedToAuthor.

```
SELECT DISTINCT ?author ?TU ?type
WHERE {
{?TU saws:isAttributedToAuthor ?author}
   UNION
 {?TU dcterms:creator ?author}
 ?TU a lawd:ConceptualWork ;
       a ?class .
 ?class a bm:auth ;
          crm:P48_has_preferred_identifier ?type }
```

By running this query including the keywords associated to each work in Beta maṣāḥəft one can explore this information as shown in Fig. 3.4.[26]

26 This can be obtained easily also in the clavis search of the project, <https://betamasaheft.eu/works/list> where the keywords are a primary filter and authors can be loaded as additional one.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

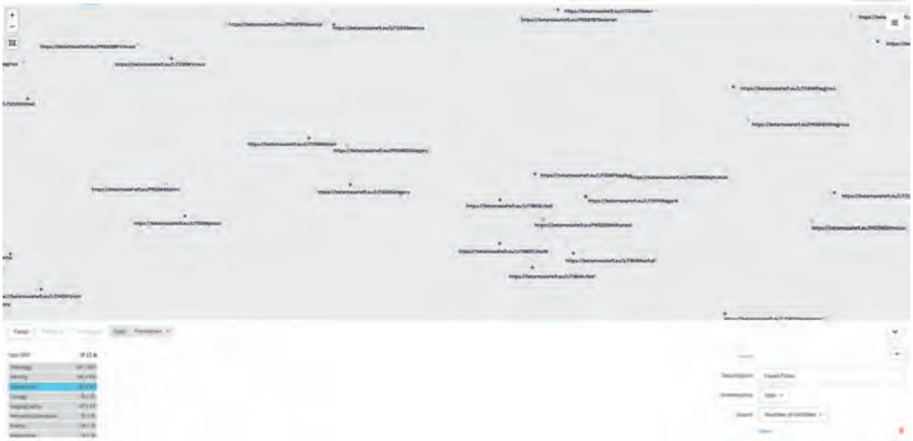


Fig. 3.4 Textual units and their attributed (saws:isAttributedToAuthor) or certain (dcterms:creator) authors viewed in ↗Palladio and filtered to translations only.

As we have seen in the previous chapter (Ch. 2), it would be even more useful to be able to run these queries over multiple compatible datasets from the different literary traditions which might be related, such as the Syriac, the Coptic, and so on.

These authorship statements can be made about the entire ↗TEI file, but, because inside the TEI description of a textual unit we find the structure of the work in its abstract form, encoded in the <text> element, with <div>s, <l>s, and so on, representing its asserted structure, we can look at a similar example making it a bit more specific.[27] For example, in the TEI file about *Wəddāse ᵓAmlāk* (CAe 2505, ID: LIT2505Weddas) there are several parts in the ideal order in which they occur in the abstract work. This needs not to be the same actual order in which the parts appear in the manuscripts, so that, when we compare the witnesses and their content, we can check what the actual order can be.[28]

In this work, which is a unit as it does circulate as such in manuscripts, each part is attributed to an author, and there are thus several authorship state-

27 Details on this aspect can be found in the guidelines, <https://betamasaheft.eu/Guidelines/?id=edition>.

28 <https://betamasaheft.eu/compare?workid=LIT2505Weddas>, see some examples also in Ch. 1.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

ments,[29] one for each part, without a need for each of these to be an abstract work record until its independent circulation in the Ethiopic tradition is attested and recorded in the database.

```
<div type="textpart" subtype="chapter" xml:id="Monday">
    <listRelation>
        <relation
                name="saws:isAttributedToAuthor"
                active="LIT2505Weddas#Monday"
                passive="PRS2546Basilth">
            <desc>Attributed to
             <persName ref="PRS2546Basilth">Bāsəlyos</persName>
            </desc>
        </relation>
    </listRelation>
</div>
```

In this way potentially any grouping of textual units which is attested in a manuscript can be a textual unit, for example, an abstract work, and maintain a clear relation to other such entities, to its witnesses and to the related information. This also allows us to study the distribution of texts in manuscripts which serve as 'corpus organizers', where 'neither the occurrence nor the sequence of the items are clearly fixed'[30]. In this way we can distinguish the grouping of units as concretely attested in a manuscript from the abstract groupings of texts which might have been reconstructed. By organizing the data in this way we try in Beta maṣāḥəft to represent the 'relational function that links a specific manuscript to others, thus collectively representing and attesting to a corpus of written knowledge and to tools used in a concrete praxis'[31], and to remain loyal to the fact that, as Bausi continues in his article, 'far from being conceived as an autonomous and well-defined witness of texts (as it would appear from purely philological perspective), each of these "corpus-organizers" acquires its full significance only in mutual relationship to the others. Each manuscript organizes an implicit but nevertheless also quite material and concrete, evolving knowledge.'

Being able to look at the literary tradition both from the text and from the manuscript is not enough. We also need to be able to look at the texts

29 The use of the element <relation> for such statements is described in Ch. 4 and Ch. 7 as well, where the mapping of the ID with anchor to a ↗URI is also detailed. The use of <listRelation> everywhere in a ↗TEI document is specified in the Beta maṣāḥəft ↗schema.

30 Bausi 2010, 34.

31 Bausi 2010, 34–35.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

in the manuscripts individually identified, and this multiplicity of possible perspectives is what the structure of the data in Beta maṣāḥəft strives to offer, as I will discuss in the following section.

1.3 Core content and additions

Given that we are now satisfactorily able to distinguish and relate the main textual units, some problems of application remain, especially when looking at the evidence provided by the manuscripts. We need here to distinguish and preserve two types of information, the one related to the status of a text in the manuscript and the one about the textual unit.

The observation decides the strata and units to be used but the cataloguer can still decide and add more or less precision. Andrist has proposed some principles to distinguish the main strata in a manuscript which I try here to summarize in a table.[32]

Table 3.1 **Andrist classification of strata in a manuscript**

Stratum	Example	Project	Self-Standing Content	Self-Standing Material
Primary	Two different writing projects bound together	Yes	Yes	Yes
Secondary	Extra texts copied on new quires and bound together	No	Yes	Yes
Tertiary	Added table of contents	No	No	Yes
Quaternary	Added written elements, small margin drawings	Yes	Yes	No

This principle, although it is not to be used for a detailed syntactic description according to Andrist,[33] is very useful to distinguish a so-called 'addition' from the core content, a task which cataloguers have to deal with not infrequently.

Beside more normal additions, it can happen that

1) a textual unit is added as a guest text in a quaternary stratum of a manuscript;
2) a text originally in a quaternary stratum of a manuscript becomes the core content of another manuscript.

32 Andrist 2015a, 511–512. See also Ch. 7 and Liuzzo forthcoming a.

33 Andrist 2015a.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

An addition is always quaternary stratum, yet, when its content is copied, it can become the primary stratum of the new manuscript. This is, for instance, the case of Rome, Biblioteca dell'Accademia Nazionale dei Lincei e Corsiniana, Conti Rossini 27 (ANLcr27).[34] In this manuscript the donation note is encoded as <msItem>, but an example of the same type of document in Paris, Bibliothèque nationale de France, Éthiopien 32 (BNFet32) is an addition encoded as an <item> in a <list> inside <additions>.[35]

The same document can thus be attested as addition or as core content and the encoding (as well as the visualization) needs to be able both to make that distinction and to group all the instances of such text. In Beta maṣāḥəft this is done by a link to the ID of that textual unit, that is, its identifier in the *Clavis Aethiopica*.

Conceptually, there is no difference in the model. Whether a text is core content or added in one of the manuscript's later strata, it can be identified by means of a textual or narrative unit as necessary, or by associating it with both wherever this helps. Also, there can be generic keywords from a taxonomy associated to each individual manifestation, as well as to the abstract entities they are linked to.

Another example is that of documents written in Golden Gospels as additions,[36] and then collected in the so-called archive manuscripts like Paris, Bibliothèque nationale de France, Éthiopien d'Abbadie 152 (BNFabb152) where they are the core content.

34 <https://betamasaheft.eu/manuscripts/ANLcr27/main>.

35 See <https://betamasaheft.eu/manuscripts/BNFet32/main#a2> and click on the folium indication to view the image.

36 See <https://betamasaheft.eu/authority-files/GoldenGospel/main>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Additions

1. (Type: Supplication)

   (ar) اذكر يا رب عبدك الخاطي المسكين قسيس غبريال المصري قاصد البطرك الاسكندراني وراهب دير العذرى مريم جبل الطير واذكر يا رب كاللذين يذكرنا من الاباء الحبش

   (en) Remember, o my Lord, your servant, the poor sinner priest Ġabriyāl the Egyptian, delegate of the Alexandrine patriarch and monk of the Monastery of the Virgin Mary of Ǧabal al-Ṭayr; and remember, o Lord, those who will commemorate us amongst the Ethiopian fathers

   The note is written in Arabic on the front board on its inner side. It was drawn up by the Egyptian priest Ġabriyāl.

2. f. 217v

   (Type: Record)

   Text in Gəʿəz

   Historical note on the monastic community of Santo Stefano de' Mori.

3. f. 219r

   (Type: Calendaric Note)

   (gez) [illegible]

   The note refers to the year 6616, year of John. The date can not be assessed with certainty; it might correspond to 1624 AD or be incorrect.

4. f. 219v

   (Type: Unclear)

   (gez) እምገበ፡ ለእከ፡ ... ማሮሮርእም፡ (?) ዘሮማ፡ ትብጸሕ፡ (!) ኢ<የ>ሩሰሌ<ም>

   The note is hardly legible because of the faded ink. It mentions Rome and Jerusalem.

Extras

1. f. 1r

   (la) Missale; n° 1. Aduerte Lector hunc Codicem esse Vitiosissime scriptum ejusq autoritatem exigui valoris.

   (en) Missal; nr. 1. Be careful o reader, that this codex is very defective and its reliability is of low value.

   Note written by a later librarian and by Johann-Michael Wansleben concerning the reliability of the manuscript. The note was probably written in 1665.

2. f. 1r

   (it) Etiopico.

Fig. 3.5 Additions and extras in Vatican City, Biblioteca Apostolica Vaticana, Aeth. 1 (BAVet1).

One of these documents can have in Beta maṣāḥəft an identifier as any other textual unit and be defined by the appropriate keywords. Additionally, due to the interaction with the Ethiopic Manuscript Archives project, it is possible to have a corpus of documents,[37] which collects the entities associated to it using the same saws:formsPartOf. In this case the 'corpus' is neither a literary work nor a manuscript and thus it is a concept on its own separated from the others, which can be used without affecting either the encoding of the manuscripts or the organization of the data about textual units.

It is also to be noted that there is no necessity for abstraction, if this is not needed, because each <msItem> and each <item> descendant of <additions>

37 <https://betamasaheft.eu/documentcorpora.html>. 'These pieces, of real value only if analyzed as statistical documents in a series, not as individual ones, should be stored in a full text data base and encoded according to several categories', Kropp 1994, 131.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

has its own and unique ID in the ↗TEI file, which then becomes a ↗URI in the ↗RDF representation. The abstraction is needed only when grouping is required and can happen with reference to a textual or narrative unit.

A number of other concepts, like the one nicely and clearly proposed by Andrist to give a better definition of paratexts and side-contents have not yet found their place in Beta maṣāḥəft.[38] However, there is no reason why they could not be incorporated in the future, given the evolving nature of the ↗schema and the guidelines associated with the data, which are maintained by a community of users which makes it as much up to date as possible with the current research.

Now, with a better idea of how the data is organized, we can move on to see some possible ways to analyse it.

2 Trees and charts

The tree metaphor, which is at the basis of the hierarchical structure of XML and serves well the encoding of a text for describing and transcribing all types of written artefacts, is also a good metaphor for literature and its historical development.[39] Let me quote a large passage from a recent book by Franco Moretti, where literary genres as 'artifactual types' are described.

> Different biological species usually do not interbreed, and on the rare occasions when they do their offspring are infertile. Artifactual types, on the other hand, are routinely combined to produce new and fruitful entities … The internal combustion engine branch was joined with that of the bicycle and horse-drawn carriage to create the automobile branch, which in turn merged with the dray wagon to produce the motor truck. (George Basalla, *The Evolution of Technology*, Cambridge 1988, p. 137–8)
>
> Artifactual species combined in new and fruitful entities: in support of his thesis, Basalla reproduces Alfred Kroeber's ingenious 'tree of culture' [here Fig. 3.6], whose Alice-in-Wonderland quality makes the reality of convergence unforgettably clear.

38 Andrist 2018, 142–147.

39 See Moretti 2013, 59–61 on this metaphor and that of waves.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

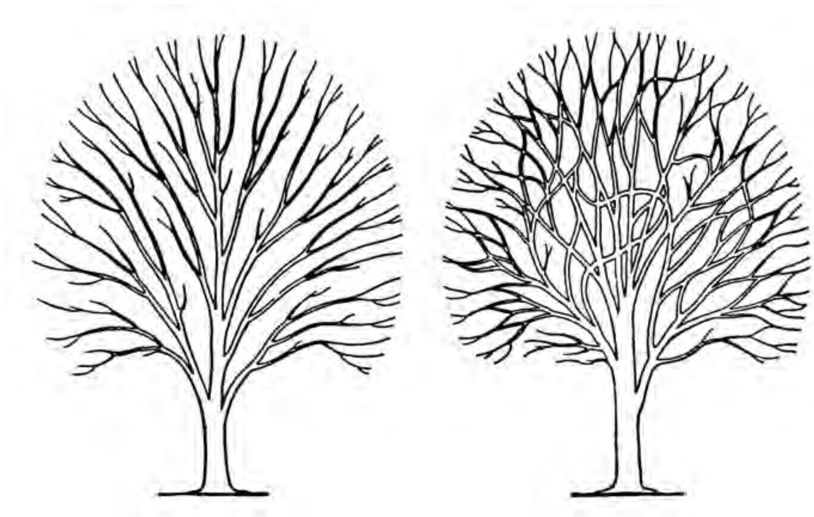


Fig. 3.6 The tree of life and the tree of the knowledge of Good and Evil—that is, of human culture (from Kroeber's *Anthropology* 1923) .

As it should be, because convergence is indeed a major factor of cultural evolution. But is it *the only one*? '*Culture diverges*, but it syncretizes and anastomoses too', runs Kroeber's comment on the tree of culture; and Basalla: 'the oldest surviving made things … stand at the beginning of the *interconnected, branching*, continuous series of artifacts shaped by deliberate human effort'. Interconnected *and* branching; syncretism *and* divergence: rather than irreconcilable 'differences in deep principle' between convergence and divergence, passages like these (which could be easily multiplied) suggest a sort of division of labour between them; or perhaps, better, a cycle to which they both contribute in turn. Convergence, I mean, only arises *on the basis of previous divergence*, and its power tends in fact to be directly proportional to the distance between the original branches (bicycles, and internal combustion engines). Conversely, a successful convergence usually produces *a powerful new burst of divergence*: like the 'new evolutionary series [which] began almost immediately after Whitney's [cotton gin] was put to work', and which quickly became, concludes Basalla, 'the point of origin for an entirely new set of artifacts'.[40]

40 Moretti 2005, 78–81. Emphasis in original.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Selecting some data from the XML trees, we could try to reproduce the convergences and divergences in the literature from Ethiopia and Eritrea, calculating the proportion of the branches to see where powerful divergences and convergences are visible. There are of course a few ground differences from the image discussed by Moretti. First of all, the fact that there cannot be one single trunk. Secondly, it cannot be taken for granted, as the tree image in this case suggests, that knowledge always grows, especially if the particular kind of artefact we look at is a literary artefact.

The most useful graph in my opinion to represent these divergences and convergences in the history of literature is a sankey diagram.[41] This chart is often used to represent flows, and, in this case, I will use it to represent the flow of literature through the main periods of its development.[42] While the basic functionality of this type of chart will return the connections between nodes in a flow, the significance of the chart lies in the definition of its nodes and edges.[43] These will give an idea of the 'power' of a given divergence or convergence.

The nodes in the charts which follow here are defined by grouping textual units on the basis of two parameters: the period to which they are associated and the subject keywords. The keywords in our ↗schema which define the assignment of a work to a period are only six and they refer to concepts in ↗PeriodO.[44] The keywords relating to subjects are, at the time of writing, more than forty and include also some grouping keywords as the taxonomy serves in Beta maṣāḥəft also a simple data retrieval function. Both types of keywords are stored in a <term> element both in the descriptions of written artefacts and in the descriptions of textual units.

The nodes in the sankey diagrams will be groups of textual units primarily arranged by period and secondarily on the basis of the keywords associated to them, so that each node is a group of one or more textual units which share the same set of keywords in a given period.

41 These are produced here using default setting examples from Google Charts.

42 All visualizations are available in the accompanying website <https://pietroliuzzo.github.io/DHEth/> and can be browsed there more comfortably. For charts with not too many nodes, also a breakdown of the groups is offered with the links to the entities they include in Beta maṣāḥəft. For an up-to-date version these charts can be obtained selecting keywords in <https://betamasaheft.eu/LitFlow> or for each subject keyword on the graph view, e.g. <https://betamasaheft.eu/authority-files/Apocrypha/graph>.

43 I would like to thank here especially my brother Simone Maria Liuzzo, and my colleagues Solomon Gebreyes Beyene and Sisay Sahile, who helped me figure out what was important and how to cleanly calculate it.

44 A list can be obtained also at <https://betamasaheft.eu/authority-files/list>. Pre-Aksumite and Early Aksumite are not taken into consideration.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

It is important to note that, in an ongoing project like Beta maṣāḥəft, the data changes daily and that some of the keywords have been inherited from previous projects, so that the examples provided here should not be taken in any way to describe the current state of knowledge about the Ethiopian literature, but only to demonstrate the possibility of the application of distant reading methodologies to the available data,[45] which may be improved and repeated in the future in order to update the knowledge we can gather from them.

The grouping process excludes the possibility that the same textual unit occurs in two groups of the same period, but, if more than one period keyword has been assigned, then the textual unit will be grouped in both periods. Also, it is obvious that textual units to which no other keyword than the one for the period has been assigned will not end up in any group.[46] Nodes could also be created for each keyword, grouping works one time for each of them, but I think the proposed grouping represents better the way in which few products of literature fit only and entirely in one classification.

If the nodes so defined represent points of convergences, it remains to be calculated which kind of divergence produces them, a measure which is reflected in the lines joining each node, the edges in the network. Edges in the sankey chart are calculated in order to represent with their weight the relation between two given groups as defined above in successive periods which share at least one subject keyword.[47] Since the same measure and same calcu-

45 For a discussion of 'distant reading' see Moretti 2013, 48–49.

46 For example, at the time of writing most of the inscriptions discussed in Ch. 2 have only a period keyword, and no subject, while the *Chronicle of ʿAmda Ṣəyon* (CAe 4275, ID: LIT4275ChronAmdS) has subject keywords but no period keywords. All these will not appear in the data here.

47 Aksumite groups are only compared to Post-Aksumite I, Post-Aksumite I only to Post-Aksumite II, etc. If a group is new to a period which is not the first of the six used, this will be lined up at the beginning of the flow. The weight, which must be a positive integer is derived from two main factors: (1) the difference in the number of textual units attested in the group (D), which can be a positive or negative integer; (2) the difference between the keywords sets (K) calculated by looking at the total number of shared keywords between the two groups and using that as a factor to multiply the total number of keywords in each group and get the difference between the two, which results in a positive or negative integer. Both measures tell us both of an increase or decrease in the feature represented and could be used to plot each group in a chart which shows the distance from the group for which the comparison is made. Keeping D on the axis and K on the ordinates, for example, the group for which we make the comparison will be at the zero and a group with a higher D and higher K will be placed at the point defined by these coordinates. A smaller D or K will result in a position in the negative quadrants. In order to make this always a positive factor,

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

lation are made for each group, also each group of relations to other groups is weighted.[48]

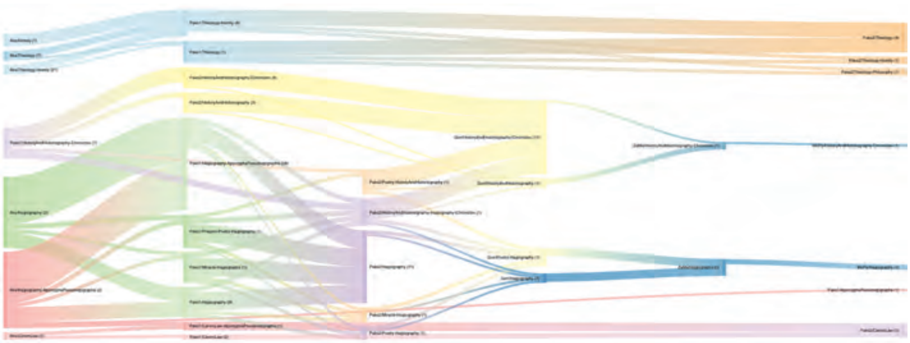


Fig. 3.7 Sankey chart of textual units grouped by period and keywords. See <https://pietroliuzzo.github.io/DHEth/Works/EthioLitFlowWorks.html>.

Fig. 3.7 looks at the textual units in Beta maṣāḥǝft, grouping by keywords which are directly added to these ↗TEI files, as defined above. Here all periods are attested, and the most striking feature is probably the distinction which is highlighted between the works related to 'theology' and 'homily', which seem to be rather distinct from the rest. It is also possible to quickly observe the importance of hagiography at least in the picture provided by the current data. Another striking feature of the chart is the diversity in the groups which can be seen increasing in the Post-Aksumite periods and then strongly decreasing in the Gondarine and subsequent periods.

However, the plot thickens as soon as we look at the literature from the manuscripts and not directly looking at the level of abstraction of the textual units. In Fig. 3.8 the selection is made starting from manuscripts which are associated to a specific period with a keyword. Then the textual units referred to from the first level only of nested <msItem>s in the <msContents> element of the manuscript record are selected and the keywords computed from there.

as required by the weight of the lines in the sankey diagram, all points in the plot are translated to a vector corresponding to the positive value of the minimal coordinates. The distance from zero is then calculated, which represents the relevance of the relation proportionally inside the group of relations.

48 The ↗XQuery functions used to produce the charts are available at <https://github.com/PietroLiuzzo/DHEth/tree/master/3%20textual%20units/modules> and need to be used in the Beta maṣāḥǝft ↗application, as they rely on other modules in there, and on the indexes. The modules contain annotations explaining some of the steps outlined here in their details.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

So the number of instances is much higher as each work is counted as many times as it occurs in the manuscripts and it does not need to have a period keyword, which instead is much more often present in manuscripts, in the current state of the data. The diversity in textual units is thus much higher and consequently the number of groups and relations (nodes and edges) is also higher.

Fig. 3.8 Sankey chart of textual units linked in first level <msItem>s of manuscript record grouped by period of the manuscript and keywords of the work excluding biblical material. See <https://pietroliuzzo.github.io/DHEth/MSS/EthioLitFlowWorksinMSSnoBibl.html>.

I have also filtered out some very frequent biblical keywords from this chart to make it slightly more readable, but the complexity it delivers would require a much deeper analysis. In comparison to the previous chart we see that the Aksumite period is lost, as we could expect, since almost no manuscript survives from that period. Also the high decrease in diversity, visible already for the Gondarine period in Fig. 3.7, here disappears. Given the very early state of encoding of the data this is no surprise.

In Fig. 3.9 the same selection starting from the manuscripts is performed taking into consideration only the textual units which include a specific keyword, which in the example is ‘poetry’. Because this selection forces the presence of a keyword, it is clear that there will be always edges between all nodes at each level, which was not a given in the previous charts.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Fig. 3.9 Sankey chart of textual units linked in first level <msItem>s of manuscript record grouped by period of the manuscript and keywords of the work excluding biblical material limited to 'poetry'. See <https://pietroliuzzo.github.io/DHEth/MSS/EthioLitFlowWorksinMSSnoBiblPOETRY.html>.

More interesting is perhaps to look at two or more selected keywords to see more clearly the relations which occur between two genres. In Fig. 3.10 and Fig. 3.11, I have selected both the keyword 'hagiography' and 'history and historiography', in the first case looking directly at the textual unit and in the second at the actual presence of textual units in the manuscripts.

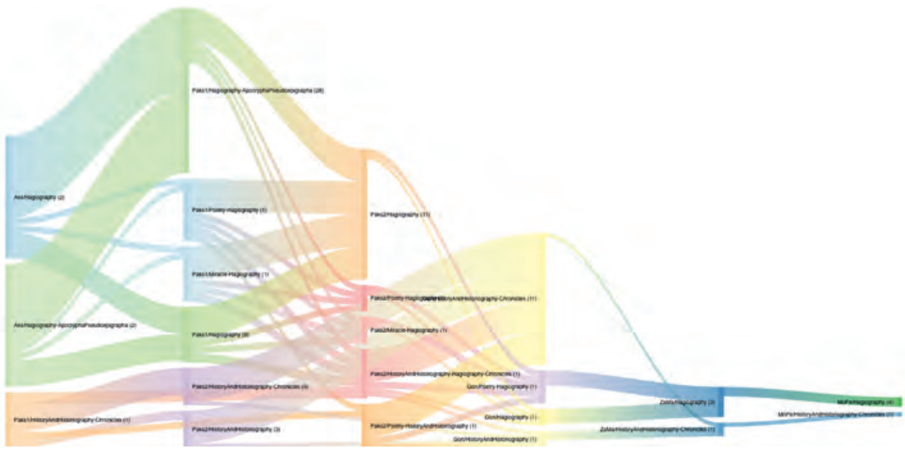


Fig. 3.10 Sankey chart of textual units grouped by period and keywords excluding biblical material limited to 'hagiography' and 'history and historiography'. See <https://pietroliuzzo.github.io/DHEth/Works/EthioLitFlowWorksnoBiblHagiographyHistoriography.html>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

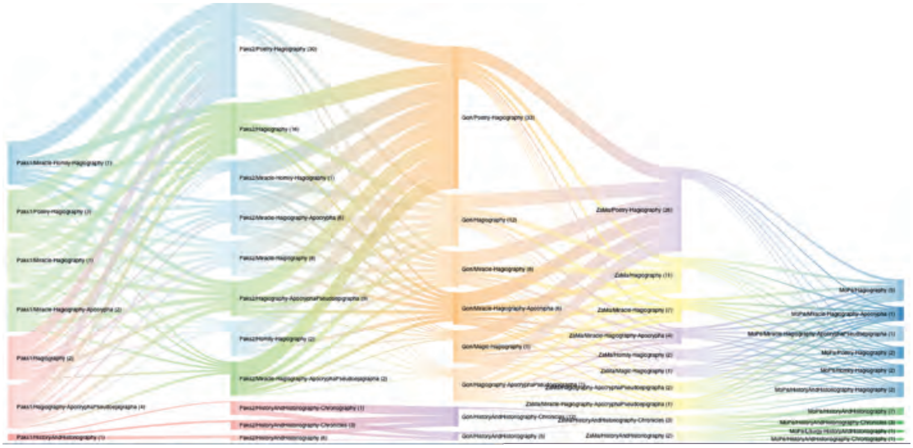


Fig. 3.11 Sankey chart of textual units linked in first level <msItem>s of manuscript record grouped by period of the manuscript and keywords of the work excluding biblical material limited to 'hagiography' and 'history and historiography'. See <https://pietroliuzzo.github.io/DHEth/Works/EthioLitFlowWorksinMSSnoBiblHagiographyHistoriography.html>.

The first chart (Fig. 3.10) shows the deep roots of hagiography in the Aksumite period, and the later birth of the historical genres, as well as a possible higher relevance that historiographical work obtained in the Gondarine period, and respective decrease of importance of hagiography, detailed by the thick edges for historiography and the very tiny ones for hagiography.

From Fig. 3.11 we can observe something rather different. The supremacy of 'hagiography' in comparison to the almost annihilated 'history and historiography' holds on in the Gondarine period, and especially in connection with poetry. Moreover, the groups of works which contain 'hagiography' as a keyword never collide with those who have 'history and historiography' until the Modern Period where two textual units have been classified under both terms.

All the above visualization additionally seems to confirm the general observation made by Cerulli, who stated the following:

> Facile ad accettare e ad assimilare adattando, la letteratura etiopica è pure, non raramente, facile a coagulare, a sclerotizzare queste ispirazioni ed espressioni nuovamente accettate, quando esse non si rinnovino con successivi elementi; sì che si arriva, qui come in altre letter-

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

ature orientali, a canoni rigidi che regolano l'espressione artistica sui modelli prestabiliti e che divengono tanto assolutamente obbligatori da giungere a rallentare e poi fermare lo slancio iniziale.

(Ethiopic literature easily accepts and assimilates adapting. It also, not rarely, easily coagulates and sclerotizes these inspirations and expressions which it has newly accepted, when they do not renew themselves with successive elements. In this way it comes, as in other Oriental literatures, to rigid canons which regulate the artistic expression on preset models and become so absolutely obligatory to relent and finally stop the initial élan.)[49]

These could of course be highly deformed by the changing state of the data, and indeed one keyword of difference does influence the results (as it should) but I hope to have provided a test of how, even only working with keywords, the data can be used to visualize interesting patterns which would otherwise remain hidden by the quantity of the data itself. More visualizations of these data could be attempted in the future, which could speak about the characteristic assimilation through translation of the Ethiopic literature.

Kropp said that a corpus could even help 'to decide on the style and nature of texts: for example to resolve the problem of whether a given text is a translation from the Arabic or not'[50]. However, for this kind of question an extensive annotated corpus, like the one curated by the TraCES project,[51] would be necessary in conjunction to the metadata which can be gathered from Beta maṣāḥəft.[52]

49 Cerulli 1968, 12–13.

50 Kropp 1994, 126.

51 See Ch. 5.

52 See on translations to Gəʿəz Marrassini 2001. I have extracted some data selecting those `<relation>` elements with `@name` equal to `saws:isVersionInAnotherLanguageOf` or `saws:isVersionOf` and made a network graph of the resulting information, which is available here <https://pietroliuzzo.github.io/DHEth/translations.html>. Cf. also p. 207. There is one node for each period, one for each language, and one for each ID in the `<relation>` element. Edges are instead created among entity and period, entity and language, and entity and entity. The <http://visjs.org> network graph also offers some interaction taken from the examples of the software. I have also stored some tables which can be used, e.g. in ↗Palladio (<http://hdlab.stanford.edu/palladio-app/#/visualization>) and which are here: <https://github.com/PietroLiuzzo/DHEth/tree/master/3%20textual%20units>. The data is however not very telling in this case.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

3 Editions in context

Sahle gave a very practical definition of digital edition saying that 'A digital edition cannot be given in print without significant loss of content and functionality'[53]. This definition helps to sum up this chapter about textual units and the ways in which they can be structured and used for analysis. In Beta maṣāḥəft we would like for each textual unit to be a digital edition of the kind defined by Sahle, and we would like for each manuscript description to also be of the same kind, that is printable but, at the very same time, is much more meaningful in relation with all other digital editions, manuscript descriptions, places and persons files. A digital edition in a research environment should have its value and defining characteristic in being connected to the other editions and resources of different kinds.[54] In addition to this, an edition 'in' the Web, not simply online, should be as connected as possible to other available resources and this is possible only if the data is freely available and is produced and served following community-driven international standards.

One way to connect with other resources will be discussed in the next chapter and in Ch. 7 as well as in Ch. 5 where it is detailed how text and lexical resources are connected and a simple link is provided to give access to analysis services in Voyant Tools (Fig. 3.12):[55] from a selection of texts, transcriptions, or single textual units which can be investigated with those computational and visualization tools.

53 Sahle 2016, 27.

54 It is indispensable also that available texts, editions, or transcriptions are easily retrievable and citable. While Beta maṣāḥəft supports the standardization of edition-independent citation structures, by encoding text in a structured way, it also supports the multiple existence of defined citation methodologies and the referencing of texts, on which see p. 155.

55 <https://voyant-tools.org/>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Fig. 3.12 All the twenty-one available texts from the list selection of textual units with the keyword 'history and historiography' passed to Voyant Tools.

While describing textual and narrative units in this chapter I have tried to provide some insights from a distant reading perspective. In the next chapter we will look at single groups of annotations of toponyms, enlarging as in Ch. 2 the perspective to other datasets, this time using web ↗annotations made with freely available tools.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Chapter 4. Places in Ancient Ethiopia

> 'In the two lectures God occurs twenty-seven times […]. But there is a complication', said Murke, 'which is the following: if we replace the word God, at least that's how it is in your lecture, apart from the genitive, the casual connection will not be clear with "that higher being, which we worship", but it must be made clear […].'
>
> Bur-Malottke had actually not thought of these complications; he started to sweat, the casual shift made him worry.
>
> (H. Böll, *Doktor Murkes gesammeltes Schweigen*, 1953)

As in the text quoted in the epigraph to this chapter, where the issue is the replacement of a word with a longer construct for a radio programme, there is sometimes an invisible clash between what the 'technical people' can quite normally utter and the reactions these statements can generate in people which have a more 'aesthetic' mind. This is even more true for historical geography, especially when done with digital methodologies, as the most common concepts, as 'place', become slippery and technicalities quickly and necessarily take the centre of the stage.

Both because of its conceptualization and of the ways in which these are represented, the study of ancient places done with digital methodologies can make people sweat like Bur-Malottke in the text quoted as epigraph, but it is totally worth the fatigue. The gains in precision and accuracy in the study of space allow the enforcement of the connecting nature of spatial concepts.

Not being myself an expert in either digital methodologies for the encoding of information relevant to historical geography or of the historical geography of Ethiopia as such, I will give in this chapter a practitioner's description of the main issues and concepts with a few examples of how in the context of Beta maṣāḥəft we have put them into practice.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

1 The Pleiades conceptual overview

I will start by briefly commenting on the definition given by the ↗Pleiades project of the concept of 'place'.[1] In its third paragraph this definition says,

> Places are entirely abstract, conceptual entities. They are objects of thought, speech, or writing, not tangible, mappable points on the earth's surface. They have no spatial or temporal attributes of their own. A place can exist in name only in an ancient source, without any material correlate; conversely, an archaeological site can exist as a place without an ancient name. The spatial aspects of ↗Pleiades places (i.e. latitude and longitude coordinates in space), as well as their ancient and modern names, are addressed through two other conceptual entities: locations and names. Temporal characteristics are also recorded at the name and location levels.

What is especially important here is the core separation of the 'place' concept from its 'locations' and 'names', meaning the ability of this conceptual model, with its simplicity, to maintain clarity in the collection of the data and produce knowledge by ensuring a very solid basis.

Especially when dealing with historical geography, it is of vital importance to be able to model the available information in a way that allows us to be precise about what we know and what we do not know. With this core separation of the three concepts of 'place', 'location', and 'name', where the temporal component is attached to the second and third of these concepts, we can even store information about 'place' concepts about which we do not know any name or location (relative or precise), although in most cases we would know at least one of these without having to provide any point for them on the map. Very often the desire of seeing on a clear map the places of interest for a given historical period or a given source has led to drawings which serve little purpose. These have for sure their value in the context of the studies in which they were included, but it is here that the digital turn allows us to do more, because we can say all we know without necessarily having to draw that point or to decide for a name.

1 <https://pleiades.stoa.org/>, see Elliott and Gillies 2009, §§ 40–46 and Elliott and Gillies 2011 where the core point about ↗Pleiades is briefly and effectively made. The definition, at the date of writing, was created by Sean Gillies, Jeffrey Becker, Elizabeth Robinson, Adam Rabinowitz, Tom Elliott, Sarah Bond, Brian Turner, Stuart Dunn, Noah Kaye, and Ryan Horne, see <https://pleiades.stoa.org/help/conceptual-overview> for the full text. See also Palladino 2016, who discusses more spatial models and some important concepts for ancient geography.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

In this chapter I will present a series of examples,[2] starting from that of the 'place' concept 'Ethiopia', to show how digital methods do not make our life easier, and instead force us to more precision thus leading to a better overall quality of the data produced and made available for future generations to be reused in ways we cannot yet foresee.[3]

I will offer some examples of place identification and ↗annotation,[4] building on the toponyms marked up in some of the Greek and Latin sources about Ethiopia and the Red Sea region,[5] to give a few examples of tools which can be used to annotate, and of the problems which annotation can force to face. I will also provide some examples of tools which can be used to study ancient maps of Ethiopia,[6] especially the Georeferencer.[7]

What this chapter aims to highlight with its practical examples is the improved quality of research data production and, consequently, of the results which can be obtained from it for our understanding of ancient sources, using digital and web-based standards and technologies. One could store (if they were available) all relevant texts in a full text database and search for a place name. This is already a positive result,[8] and I do not want to discourage anyone from doing so, but I would like to add to the many already existing, some more examples from the Ethiopian studies domain to show how it is possible to do much more than that already, thanks to many good projects which have led the way in the past decades.[9] Among these, it is important to name at the outset, beside the already mentioned ↗Pleiades, also the ↗Pelagios Commons project,[10] which has become over the years the most important international

2 Details on the basic ↗annotation procedures can be found in Solomon Gebreyes and Liuzzo 2018. See also Buzi et al. 2018, 42–50 for the approach of the PAThs project.

3 See the quote at the beginning of the section at p. 78.

4 On the concept of ↗annotation see Simon et al. 2016.

5 These concepts are voluntarily vague at this stage, please see below for a fuller breakdown and definition.

6 For a detailed overview of the cartography about Ethiopia see 'Cartography', *EAe*, I (2003), 686b–692b (R. Voigt, W. Smidt, and Mekete Belachew).

7 <http://www.georeferencer.com/> is a Klokan Technologies GmbH product.

8 There is still a major need for texts in whichever format they can be contributed, especially for Oriental studies where a wealth of resources are informally shared and available but inaccessible to those who do not know the actual owner of a given transcription or text.

9 We are probably not quite yet at the point envisaged by Elliott and Gillies 2009, §§ 1–5 in the section 'The View From 2017', but the years of work have shown both the potential and the effectiveness of the methodology pushed forward, together with the enormous size of the enterprise undertaken.

10 <http://commons.pelagios.org/>, see Simon et al. 2012 for an early description of the core concepts of this project.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

hub for any project related to historical geography and ↗annotation of geographical data. Beside the tools it has produced, its greatest achievement is possibly the community of interested people around it. Another project also named 'pleiades',[11] but in Arabic, is al-Ṯurayyā which includes a gazetteer and a very nice map-based search facility for Arabic sources, which are already very well-structured and systematic, thus offering the encoder a reliable data structure to begin with.[12]

The sources dealt with in this chapter are necessarily not Gəʿəz sources, but rather Greek and Latin sources which I have more competence in handling. They are not the texts which contain more, or more relevant, or more interesting place names and place descriptions, as the chronicles could be for the later period of the Ethiopian Christian kingdom, nor are they geographical texts in Gəʿəz, like the *Tārika Walda ʿAmid* (CAe 4723, ID: LIT4723TarikaWaldaAmid), which are being edited and might be the objects of a digital project in the future.[13] The chronicles and some other documents are being marked up by Solomon Gebreyes Beyene and other team members of the Beta maṣāḥəft project, but are not taken into consideration in this chapter, although they constitute the biggest bulk of ↗annotations about ancient places in Ethiopia.[14]

Let us then take the very simple example of running a search on a text for a place name in a dataset which has more than one language. It would simply fail to return all results or have to be run several times with each of the languages used.

One step further which would ideally spare the above to future users is to mark up in such a dataset all the occurrences of the place name and perhaps distinguish the ethnonyms from the toponyms. This allows us to pick up directly any piece of text which is potentially useful to a geographical research aim.

Given a dataset where place names are identified as such, for example by tagging them with the ↗TEI element <placeName>, already allows to do much

11 I would like to thank my colleague Dorothea Reule who told me about this.

12 <https://althurayya.github.io/>. See also Masoumeh Seydi et al. 2017, section 2.1 where the encoding process of the spatial description is detailed, with examples from Arabic sources as well as Roman *itineraria*.

13 See Hoffmann 2013.

14 In this case Ethiopia needs to be understood as Orbis Aethiopicus, 'Orbis Aethiopicus', *EAe*, IV (2010), 42a–42b (S. Uhlig). All the ↗annotations are available via the Beta maṣāḥəft website, or in Peripleo. The ↗turtle format dumps can be obtained starting from the ↗VoID file here <https://betamasaheft.eu/api/placeNames/void> or directly from <https://github.com/BetaMasaheft/Documentation>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

more than a simple string search,[15] but this step provides an even greater measure of semantic enrichment when it is associated with the crucial scientific activity of critical identification of the abstract place intended by the text or source under scrutiny ('georesolution') when using the given marked up toponym. The identification of toponyms can be automatized with Named Entity Recognition techniques,[16] but the result, in any case, is a text enriched with scientific information so that the toponyms are connected where possible to one or more place concepts as identified in gazetteers. The encoded text contains thus a series of scientific statements which can be used within their context.

While the first example focuses on the problem of identification of a place concept, the following section will look at how the series of place names georesolved in a text can be used to compare sequences and co-occurrences. My second example of ↗annotations of some texts regarding the Red Sea will try to show this (p. 121). My third example with the encoding of variously identified places in the *Monumentum Adulitanum* (p. 133) will try to show how the challenges of encoding lead to a richer text which can be used in several new ways.

The last brief examples (p. 138) will show how the encoding transformed in ↗annotations and reused via the ↗Pelagios hub can expand even more the potential value of the identifications, adding even further scientific value to the data. The last section will offer briefly some more examples of the potentials of the study of georeferenced old maps.

2 Identifying a place: Ethiopia

When dealing with the *Chronicle of Galāwdewos* (CAe 3122, ID: LIT3122Galaw) encoding both **እልኅቡስ ፡** (*ʾəlḫəbus*)[17] and **ኢትዮጵያ ፡** (*ʾityopyā*) with the same concept identification was unproblematic. At the time of the sources in question, the concept had already developed and was used to identify the Christian kingdom without doubt. The process by which the Greek word Αἰθιοπία came to be used to identify a nation was very complex. Shiferaw Bekele, starting from the observations made by Bausi and Gori in 'Tradizioni orientali del "Martirio di Areta": la prima recensione araba e la versione etiopica',[18] states that 'the Ethiopians of the high middle ages did

15 Wion 2018, 35–36 on reducing the labelling and preferring the tagging of the vocabulary used.

16 The collective name given to the initial steps of 'geoparsing', Elliott and Gillies 2009, §§ 11–18.

17 On this term, see Solomon Gebreyes 2019, 6, n. 16.

18 Bausi and Gori 2006.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

not want their country to be called Bilad al-Habaš; they preferred very clearly to be known by the name of Ethiopia.'[19] According to Shiferaw Bekele, the *manbara tabot*s provide evidence to state that, already between the seventh and ninth century, a national epic was formed which preferred the name Ethiopia. So, the concept of Ethiopia as state or kingdom is actually late. Before this time, both the concept of Ethiopia and the toponyms used for it are fluid. Munro-Hay provides a rather clear statement about the 'confusion' which can be faced by the scholar in the sources regarding the metropolitan bishops of Ethiopia in the period between the fourth and the thirteenth century.[20] He says, 'There is considerable confusion in the terminology employed by various authors, ancient and modern, for the Aksumite kingdom, present-day Ethiopia (Abyssinia), Nubia, Meroitic Ethiopia in the modern republic of Sudan, and India.'[21]

This is indeed the case as the survey of associations of toponyms to actual locations and place concepts proves. Undeniably one always feels the need to be more precise than the term he uses in such contexts: each attestation of a toponym can be made into the object of speculation on its identification and thus each association of a toponym in a source with a place concept (located or not) is the product of a research effort.

In the most ancient documents we have from the Ethiopian and Eritrean highlands, the inscriptions,[22] we have the same two terms which in the *Chronicle of Galāwdewos* are associated with certainty to the same concept of Ethiopia as the Christian kingdom, but here their identification may be different, not just the location of a static concept. The most important evidence is here given by inscriptions in multiple languages, where it can be observed which term is used in each language. In *RIÉ* 185 (and *RIÉ* 185 bis) the Greek βασιλεὺς Αἰθιόπων is parallaled by *ngś Ḥbśt* in Ethiopic.[23] Additionally, as Voigt points out,[24] if the Κάσου of the inscriptions can be identified with Kush, then it can be safely stated that Αἰθιοπία/Ḥbśt as a concept cannot to be identified with the Nubian kingdom. In fact the term Αἰθιοπία was used to refer to Kush, and terms used to refer to a place concept centred in Meroe included often also what we would now refer to as Ethiopia.

19 Shiferaw Bekele 2018, 12–13.

20 Munro-Hay 1997, 11–14.

21 Munro-Hay 1997, 11.

22 See Ch. 2.

23 'Ḥabašāt', *EAe*, II (2005), 948b–949a (W. W. Müller) offers an overview of the possible identification of the people connected to this toponym. See also Marrassini 2014, 18–19.

24 'Aithiopía', *EAe*, I (2003), 162b–165a (R. Voigt).

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Voigt traced a brief outline of the variation in meaning of the concept behind the word Αἰθιοπία in ancient sources and highlighted the important role of the biblical texts in the fortune of the Greek word compared to other names used for the same concept.[25] At the origin of the use in the Bible of the term Αἰθιοπία,[26] there is the Greek and Roman tradition and this presents a variety of meanings for the term, which I am now going to describe. I will focus on the earliest part of the development of the place concepts in Greek and Roman sources, occasionally bound to a toponym which can be associated with 'Ethiopia' to show how the perspective and the sources of each text influence the possible identification of a toponym and its association as a name to a place concept.[27]

2.1 Ancient concepts of Ethiopia

The place concept used for the *Chronicle of Galāwdewos* does not suffice: to have one concept linked to a variety of toponyms is not enough. This can be demonstrated with the example of the number of concepts related to the term Αἰθιοπία and few other toponyms.

The starting point of this enquiry has been a set of selected texts collected by Felix Jacoby in the *Anhang* dedicated to Ethiopia to his *Die Fragmente der griechischen Historiker*, as recently edited by Burstein for Brill's New Jacoby Online.[28] I will discuss only some selected passages which highlight the different concepts in play.

2.1.1 The most ancient concepts of Ethiopia

According to Burstein,[29] the most productive line for the construction of conceptions of Ethiopia in antiquity was the Homeric passage in the *Odyssey* 'Αἰθίοπας τοὶ διχθὰ δεδαίαται, ἔσχατοι ἀνδρῶν, οἱ μὲν δυσομένου Ὑπερίονος, οἱ δ' ἀνιόντος' ('the Ethiopians, who are divided into two, the most remote of men, some where Hyperion sets and some where he rises')[30], which generated

25 See also Marrassini 2014, 19–21 who gives the narrow interpretation of the place concept.

26 The most famous passage being Acts 8:27. See Marrassini 2014, 39.

27 For the terms used in Gəʿəz texts to refer to Europeans, see Kropp 1983.

28 Burstein 2016a.

29 Burstein 2016a. See especially the 'Biographical Essay' at the end of the entry <https://referenceworks.brillonline.com/entries/brill-s-new-jacoby/aithiopia-appendix-673-a673>, which is unfortunately not freely accessible.

30 Hom. Od. 1.22–23, trace of which may have also remained in the inscription *RIÉ* 277 which I propose to be a Ptolemaic text with other references to the Greek literary tradition. See Liuzzo forthcoming b.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

discussions about the two different groups of Ethiopians (Αἰθίοπας) and the location of these peoples in the known world.

In his description of the army of Xerxes,[31] Herodotus divides them between the Indian-Ethiopians and the African-Ethiopians and in various other passages documents the beginning of an attempt to locate the places from where these people came. In fact a narrower definition of the broad term used for any dark-skinned people is already used to introduce the passage:[32] 'τῶν μὲν δὴ ὑπὲρ Αἰγύπτου Αἰθιόπων καὶ Ἀραβίων ἦρχε Ἀρσάμης οἱ δὲ ἀπὸ ἡλίου ἀνατολέων Αἰθίοπες διξοὶ γὰρ δὴ ἐστρατεύοντο προσετετάχατο τοῖσι Ἰνδοῖσι' ('Arsames led the Aithiopians above Egypt and the Arabians. The Aithiopians from the east, however, for the two groups took part in the campaign, were brigaded with the Indians')[33]. In this passage, we have in fact two place concepts associated with the single generic term Αἰθίοπες. The first is the place inhabited by dark-skinned people above Egypt (i.e. south of Egypt), the second is the place inhabited by dark-skinned people coming from the east (the sunrise) and closer to India.[34]

Since we cannot use either of these concepts to annotate the occurrence of Αἰθίοπας in the *Odyssey* mentioned above, we actually already have three concepts for the place inhabited by the Αἰθίοπες. Each of these remains attested in the sources independently and can have several locations and names. Among these concepts the one in Homer is the broader in scope and might be said to contain both others as an additional feature. We have also some core associations to be recorded, because both the specifications given by Herodotus are provided in relation to another place, Egypt for the first group and India for the second, and they apply only to one or the other of the two additional place concepts.

Each of these concepts and each of these relations would need to be encoded. If each place concept is an entity, then we would need three identifiers for the three concepts and we would need each of them to be a resource we can say something about, meaning that it is defined with reference to another place, more or less definitely identified. We need a ↗URI for each of these.[35]

Strabo summarizes three ways of interpreting the Homeric passage in order to describe the part of the world inhabited by the Αἰθίοπες and charac-

31 Hdt 7.70.

32 Text and translation are from Brill's New Jacoby Appendix 673, Burstein 2016a. See also Sergew Hable Selassie 1972, 45–48.

33 Hdt 7.70.1.

34 On the confusion of Ethiopia with India, and in general on the problem of the geographical identification of Ethiopia, see Schneider 2004. Cf. Marrassini 2014, 20–21 on the confusion extending also to Himyar since Philostorgius.

35 On the importance of URIs for places see also Gibson et al. 2016, 3–4.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

terized by its division in two parts.[36] The first hypothesis discussed, that of Krates, says that the Αἰθίοπες are divided by the ocean.[37]

> ὁ μὲν γὰρ ἀκολουθῶν τοῖς μαθηματικῶς λέγεσθαι δοκοῦσι τὴν διακεκαυμένην ζώνην κατέχεσθαί φησιν ὑπὸ τοῦ ὠκεανοῦ: παρ᾽ ἑκάτερον δὲ ταύτης εἶναι τὴν εὔκρατον, τήν τε καθ᾽ ἡμᾶς καὶ τὴν ἐπὶ θάτερον μέρος. ὥσπερ οὖν οἱ παρ᾽ ἡμῖν Αἰθίοπες οὗτοι λέγονται οἱ πρὸς μεσημβρίαν κεκλιμένοι παρ᾽ ὅλην τὴν οἰκουμένην ἔσχατοι τῶν ἄλλων παροικοῦντες τὸν ὠκεανόν, οὕτως οἴεται δεῖν καὶ πέραν τοῦ ὠκεανοῦ νοεῖσθαί τινας Αἰθίοπας ἐσχάτους τῶν ἄλλων τῶν ἐν τῇ ἑτέρᾳ εὐκράτῳ, παροικοῦντας τὸν αὐτὸν τοῦτον ὠκεανόν: διττοὺς δὲ εἶναι καὶ διχθὰ δεδάσθαι ὑπὸ τοῦ ὠκεανοῦ.
>
> (One of them, in fact, takes what he considers the mathematical view of the case, and says that the torrid zone is occupied by the ocean, and that on each side of this there is a temperate zone, one inhabited by us and another opposite thereto. And as we call the Ethiopians, who are situated to the south, and dwell along the shores of the ocean, the most distant on the face of the inhabited globe; so he supposed that, on the other side of the ocean, there were certain Ethiopians dwelling along the shores, who would in like manner be considered the most distant by the inhabitants of the other temperate zone; and thus that the Ethiopians were double, separated into two divisions by the ocean.)[38]

According to this thesis, the interpretation of the Homeric passage to formulate a place concept uses two spatial references: the land inhabited by the Ethiopians is south of the Greek inhabited areas and of Egypt,[39] and is the last on the shores of the Ocean.

In the second hypothesis, the one of Aristarchus, the Homeric division of the Ethiopians is rooted in the division of τῆς καθ᾽ ἡμᾶς Αἰθιοπίας, 'our Aithiopia', by the Nile, rather than by the Ocean. By the first century BCE, in fact, the Nubian kingdom of Meroe is identified with the land inhabited by the dark-skinned man, and is rather well-known, as Strabo himself states.[40] A further relation is thus obtained in this interpretation and this associates the two groups of dark-skinned people to either of the continents, Libya and Asia, divided by the Nile.

36 Strabo 1.2.24–26. Burstein 2016a for commentary on the sources of these theories.

37 Text and translation from ↗Perseus, <http://www.perseus.tufts.edu/hopper/text?doc=Strab.+1.2.24&fromdoc=Perseus%3Atext%3A1999.01.0197>.

38 Strabo 1.2.24.

39 See also the continuation of the passage, not quoted.

40 Strabo 1.2.25.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

The last two theses exposed by Strabo, the one of the isthmus and the one attributed to Ephorus of Cuma, although not as neatly understandable, provide a further reference point for the definition, the Red Sea (ἀπὸ τῆς Ἐρυθρᾶς).[41]

While the Herodotean passage associates the two military contingents with the two groups of Αἰθίοπες and provides a relative location by means of his description, starting from this passage of Strabo, we could venture in the imaginative task of marking up the passage in the *Odyssey* according to the first two theories. Krates's thesis would require two place concepts, the Ocean place concept as well as an additional place concept for the land of the people living on the other side of the ocean in the temperate zones. Krates's astronomical justification of the spatial references to the east and west would have to be provided as well. Encoding the verses of the *Odyssey* according to the thesis of Strabo would require one place concept only related to Egypt. Eventually the definition of this place with reference to the Nile and the two continents would need to be modelled.

How do we then encode the passages of Homer and especially Strabo to reflect the development of these different place concepts? A generic and vague concept would probably suffice to encode the Homeric passage, but the occurrences of the toponym in Strabo would probably require a more diversified identification. The same identifier for a vague place, as used in Homer, could be used in all places in a text where reference is made to the term in question, but the occurrences which are specific to Strabo and reflect his place concept, the one identified with the Meroitic kingdom, would need to have a different identifier, one which points to that place concept while each concept of those exposed by the sources discussed by Strabo would also need a separate place concept. The following chapter of the *Geography* further defines what the author believed to be the place concept for the land inhabited by the Αἰθίοπες of the ancient Greeks, that is, the whole southern regions of the world on the Eritrean sea, basing his argument on Aeschylus and Euripides.[42]

Strabo 1.2.28 finally reports the opinion about Aithiopia (περὶ τῆς Αἰθιοπίας δόξαν) of Ephorus of Cuma which provides a further specification of the broad concept with reference to a spatial representation of the entire world.

What is more interesting here is that the same process of identification of the place concepts behind attestations of toponyms, or more generally references to places, are discussed by the authors. On the other side, it is evident

41 Strabo 1.2.26.

42 Strabo 1.2.27.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

how these different place concepts, and indeed many others, like the one of the long lived Ethiopians,[43] kept existing together throughout antiquity.[44]

The discussion of Strabo is concluded with a remark which is very important:[45]

> And if the moderns have confined the term of Ethiopians to those only who dwell near to Egypt, and have also restricted the Pygmies in like manner, this must not be allowed to interfere with the meaning of the ancients. We do not speak of all the people who fought against Troy as merely Achæans and Argives, though Homer describes the whole under those two names. Similar to this is my remark concerning the separation of the Ethiopians into two divisions, that under that designation we should understand the whole of the nations inhabiting the sea-board from east to west. The Ethiopians taken in this sense are naturally separated into two parts by the Arabian Gulf, which occupies a considerable portion of a meridian circle, and resembles a river, being in length nearly 15,000 stadia, and in breadth not above 1,000 at the widest point. In addition to the length, the recess of the Gulf is distant from the sea at Pelusium only three or four days' journey across the isthmus. On this account those who are most felicitous in their division of Asia and Africa, prefer the Gulf as a better boundary line for the two continents than the Nile, since it extends almost entirely from sea to sea, whereas the Nile is so remote from the ocean that it does not by any means divide the whole of Asia from Africa. On this account I believe it was the Gulf which the poet looked upon as dividing into two portions the whole southern regions of the inhabited earth. Is it possible, then, that he was unacquainted with the isthmus which separates this Gulf from the Egyptian Sea?[46]

In this passage the first thing to be noted is the concern with the separation and identification of place concepts, which must have been a reality as it kept and keeps being a problem. In the process of ↗annotation of place concepts this is crucial. In the quoted passage from Strabo, as well as in Herodotus, but indeed in any ancient text, the same toponym is associated to several place concepts and the identification needs to be done (and redone) after the core concepts used have been identified and registered.

43 Hdt 3.97.2–3.

44 See Burstein 2016a, 'Biographical Essay'.

45 Translation from ↗Perseus.

46 Strabo 1.2.28.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Additionally we note that the opinion of Strabo is that the places inhabited by the Αἰθίοπες are divided by the Red Sea, not by the Nile. Because in this passage the Greek says τῷ Ἀραβίῳ κόλπῳ for what we call today 'the Red Sea', we would then need to confirm also this place concept. It will suffice here to maintain the generic and indistinct place concept which extends the identification of this place with the whole Indian Ocean and keep this as reference of the definition of Ethiopia as a place south divided in two by the Red Sea.

2.1.2 The Ptolemaic explorations

In between Ephorus and Strabo, a major event took place which led to the certainty about the location and identification of the place concept of Ethiopia to which Strabo refers to as 'τῆς καθ' ἡμᾶς Αἰθιοπίας' ('our Aithiopia'), and these are the great expeditions to the south of Egypt sent by Alexander the Great, but especially Ptolemy II Philadelphus (309–246 BCE) and his direct successors, Ptolemy III Euergetes (284–222 BCE) and Ptolemy IV Philopator (244–204 BCE).[47]

As Diodorus Siculus explicitly tells us,[48] from the times of Ptolemy II's campaigns in Αἰθιοπία with a Greek army, facts about the lands south of Egypt become more accurately known.[49] Although this is the same place concept which Herodotus used for his western dark-skinned people, after these expeditions it takes the boundaries and the precision it will retain for centuries: it includes Meroe, excludes the regions of the Trogodytice from the Nile valley to the Red Sea, but does not have a southern boundary itself, thus yet still expanding in scope to anywhere to the south and to the west of Egypt. This place concept which got its clearer boundaries during the Ptolemaic reign is most probably also the one which is identified as Kush and used in the Greek translation of the Hebrew Bible also commissioned by Ptolemy II Philadelphus.

47 Sergew Hable Selassie 1972, 45–48 provides a long period overview where he stresses this important period. More specifically on the Ptolemaic expeditions see Burstein 1996; Burstein 2000; Burstein 2002; and Casson 1993. Discussing the relation between Agatharchides of Knidus (FGrHist 86, Burstein 2016b), Diodorus Siculus, and the *hypomnemata* of the Ptolemaic court, Peremans 1967 provides a discussion of all explorers who might have left such documents. The text of P. Oxy. 4458 (whose author Fowler 2000 identified as Poseidonius of Apamea) is an attestation in an earlier work than that of Agatharchides of Knidus used by Diodorus Siculus. On the *hypomnemata* see also Burstein 1989, 30–32.

48 Diod. 1.37.5.

49 The date of the expedition to which Ptolemy II took part is known from this passage and dated by reference in Theocritus Idylls 17.87 written before 270 BCE. On the interest of the Philadelphus for this region, see Burstein 1989, 4.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

If the place concept identifying the region centred in Meroe continues to be used alongside the newly refined concept for Aithiopia, there is also at least one other place concept which we need to take into account and which emerges in the first centuries of our era and is the one which associates the Aksumite kingdom and its extension with a place concept of Ethiopia. By this time the explorations had gone also beyond Cape Guardafui and we know from Pliny of the itinerary of the Romans sent by Nero in Nubia all the way to Meroe,[50] with distances, used to complement earlier accounts. As Burstein says,[51] after the conquest of Meroe in the mid-fourth century CE the kingdom of Aksum is identified with the place concept of Ethiopia and the discussion over the Homeric Αἰθίοπες starts to die off.

We reach here again the time of the oldest inscriptions from the Aksumite kingdom, where the term Αἰθιοπία also occurs and is associated with Ḥbśt and distinct from Kush.

In the following table I have summarized six different place concepts which might have different locations and names, which coexisted at many points in history. The table should not be read as a development of a general concept of Ethiopia, as much as a series of coexisting relevant place concepts to which several different toponyms have been associated in Greek and Roman sources.[52]

50 Pliny 6.181–186.

51 Burstein 2016a.

52 None of these can actually be equated or put into a direct relation with the present-day Ethiopia (<http://www.geonames.org/337996>) for example. For the reader's convenience, the often used ↗Pleiades place concepts have been all written as for instance pleiades:39274 which is an abbreviation for <https://pleiades.stoa.org/places/39274>; the Beta maṣāḥəft place concepts use only the identifier, like, for example, LOC7053Aithiopia, which stands for <https://betamasaheft.eu/places/LOC7053Aithiopia>. Bibliographical references for all Pleiades place concepts can be found in the dedicated part of the general bibliography. Bibliographical references for the Beta maṣāḥəft place concepts are to be found at the end of this book in the bibliography dedicated to Beta maṣāḥəft entities. A similar discussion of the place name and relative concepts can be found in Fiaccadori 2004, 125–127.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Table 4.1 **Six different coexisting concepts related to Ethiopia**

Main Label	Scope	Example	Perspective	Identifier
Αἰθιοπές	Large undelimited region south of the Greek inhabited countries	Homer, Herodotus	Greco-Roman	pleiades:334481
Αἰθιοπία	South of Egypt	Diodorus, Strabo	Greco-Roman	pleiades:39274
Αἰθιοπία	West of ʾAdulis, close to Sasou	*RIÉ* 185, *RIÉ* 277	Aksumite	LOC7053Aithiopia
Ḥbśt		*RIÉ* 185	Aksumite	LOC3843HBS
Aksum	The Aksumite kingdom		Greco-Roman and Aksumite	LOC1310Aksum
ኢትዮጵያ	The Christian kingdom of the highlands	Chronicles and sources from circa thirteenth century	Christian kingdom	LOC3010Ethiop

2.2 Relations between place concepts

Once into Peripleo, these concepts and their relations become part of an ecosystem where also many other concepts and definitions are available. The visualization of the relations of one of these concepts in Peripleo (Fig. 4.1) like the one which is found searching for Aithiopia and labelled Nubia, provides good insights on the extension of this network of place concepts.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

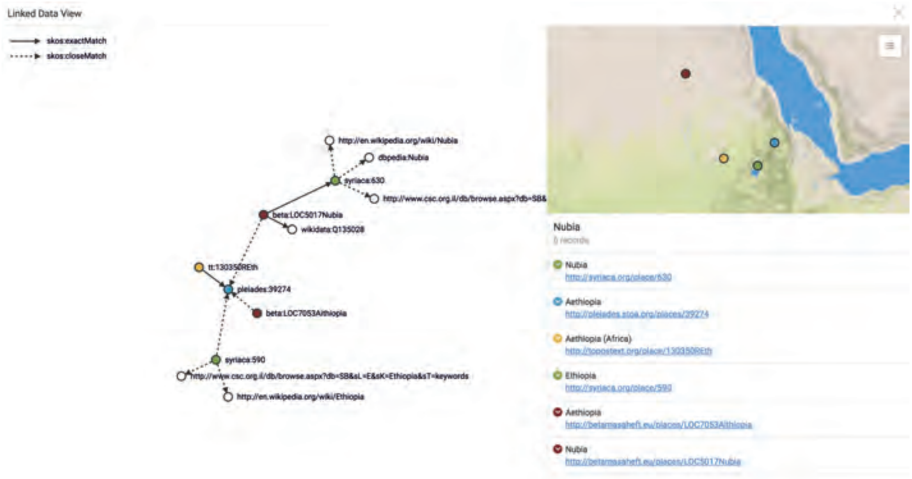


Fig. 4.1 The linked data view in Peripleo for the concepts related to Aithiopia.

We cannot control what the exact identification of the concept <http ://topostext.org/place/130350REth> with pleiades:39274 means, neither we know why that concept is labelled ‘Aethiopia (Africa)’, or why it has a representative point in the southern part of Sudan, but we get a connection to that concept via the ↗Pleiades place concept, which is closely matched to Beta maṣāḥəft places LOC7053Aithiopia and LOC5017Nubia.[53]

Additionally, each of the resources connected to one of those entities in the graph is linked to the others and this related information can be queried and retrieved. One could eventually also expand the graph of related concepts to include those which are actually important for the definition of a given concept, like Egypt and the Red Sea (in its wide and narrow extent) in this case.

This is of course not an exhaustive list of all possible place concepts associated to the Greek word Αἰθιοπία or to any other related toponym, but is a set of distinctions based on the sources discussed in the previous pages of this chapter, which allows us firstly to point to a network of place concepts as

53 ‘Il termine Αἰθιοπία è passato a designare estensivamente quella porzione dell’altipiano, a s. di Meroe e dell’Etiopia classica, che l’emergente regno di Aksum veniva annettendo alla propria signoria territoriale’ (‘The term Αἰθιοπία come to mean extensively that part of the highlands to the south of Meroe and of the classical Ethiopia, which the emerging kingdom of Aksum was slowly bringing under its control’), Fiaccadori 2004, 125.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

an entitiy rather than reducing it, and secondly allows us to demonstrate the coexistence of different place concepts which need to be researched before beginning the ↗annotation process. The encoder will need to know what is available and what the distinctive features of a concept are to be able to annotate the occurrences of toponyms in a text and associate them to one of these concepts. On the other side, each of these concepts is not simply defined by its scope, point of view, and possible set of labels, but by the relation it entertains with the other place concepts, as we have seen. The encoder needs thus to be able to create new concepts, and to create statements about the relations which exist among them across the gazetteers which provide such identifiers and stable ↗URIs.

For the Greco-Roman domain, ↗Pleiades allows one to do this collaboratively as does Beta maṣāḥəft for the domain of Ethiopian studies. The place concepts relevant in each domain can be used together given that the same conceptual model is in use so that there is no need to duplicate identifiers.

Each of these resources which identify a place concept can be pointed to and is a citable resource authored as such by one or more contributors. This allows one to keep distinguished, to some extent, the scientific effort of identifying a place attestation and associating it to a place concept from the scientific activity of definition of these place concepts as such.[54]

However, there are cases in which one might need to have an identifier which already exists with the same definition, in which case also a relation between the two is needed. This means that part of the definition of my concept Aksum (LOC1310Aksum) will be the statement that says that it is the same as ↗Pleiades place concept <https://pleiades.stoa.org/places/39303>. Similarly one can create self-standing statements connecting the place concepts to formalize the relation between each place. If we wanted to say that LOC7053Aithiopia is a place concept very close to <https://pleiades.stoa.org/places/39274>, we could do so with a self-standing ↗annotation stating just that. The software used to do this is a choice depending on taste and what matters is that the triples are openly accessible. In Beta maṣāḥəft this is done by adding a `<relation>` element in the ↗TEI file,[55] which is then transformed into a statement in the dataset which is provided to ↗Pelagios Commons for Peripleo and the Pelagios ↗API.

```
<relation name="skos:closeMatch" active="LOC7053Aithiopia" passive="pleiades:39274"/>
```

54 See in the bibliography the section about ↗Pleiades resources and Beta maṣāḥəft entities.

55 See <https://betamasaheft.eu/Guidelines/?id=places> in Liuzzo et al. 2018.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

To partly describe the different relations between the concepts in the above table, we have added a statement that says, for example, that pleiades:39274 has a close match to LOC5017Nubia (Nubia)[56] in Beta maṣāḥəft, and we could add that pleiades:39274 includes LOC1310Aksum but the latter had a different relation to LOC5017Nubia, because this is also generally true, but we could still use them independently to identify places named in one text, as for example in *RIÉ* 185. A concept like LOC3843HBS could even be associated, if that was the case in the sources, with the Bilād al-Ḥabašat in al-Ṯurayyā. LOC3010Ethiop is also to be related to the ethnonym **አግዓዚ** (ʾAgʿāzi), which is a name, not a place concept, identified by ETH1084Agazi, used in the Gəʿəz language to designate the Ethiopians.[57]

The process of definition of the place concepts to be used is only a first step. Let us now look at some examples of how the encoding can offer further ways to be specific and clear about associations of concepts to attestations.

3 Encoding place names

We will look in this section at some features of the encoding of place names and association of them to place concepts. Each attestation counts as a name given to that place concept at a given time but, especially with toponyms which are difficult to identify, it might be not that straightforward to find a fitting place concept.

Annotated <placeName>s can be usefully grouped in various ways and I will try to give some examples of what can be done.

The corpus of annotated sources includes some of the most famous descriptions of the Red Sea coasts and some of the most famous ancient maps of the area.

1) The Pithom stele dated 264 BCE;[58]
2) Diodorus Siculus's *Library*, Book 3, Chapters 37 to 41, from the first century BCE, which uses Agatarchides of Knidus's works from the second century BCE;[59]

56 Burstein 2016a commentary to Josephus's Jews. Ant. 1.131 (exegesis of Gen. 10:5 in light of Greek geography) says that the Greeks knew the name Kush used to refer to the Nubian kingdom, but preferred Αἰθιοπία for the Homeric identification.

57 Fauvelle-Aymar 2009, 143.

58 <http://www.attalus.org/docs/other/inscr_16.html>. Catalogue général des antiquités égyptiennes du Musée du Caire (CGC) 22183 (Kamal 1905, 171–177. See image at <https://archive.org/details/KamalStelesPlanches1905/page/n57>.

59 <http://www.perseus.tufts.edu/hopper/text?doc=Diod.+3.37&fromdoc=Perseus%3Atext%3A2008.01.0540>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

3) Strabo's *Geography*, Book 16, Chapter 4, from the first century CE;[60]
4) Pliny the Elder's *Natural History*, Book 6 Chapters 33 and 34, from the first century CE;[61]
5) The *Periplus of the Erythraean Sea* (CAe 2170, ID: LIT2170Peripl), Chapters 1 to 12 and 19, from the first century CE;[62]
6) The ostrakon published in SB 24 16187, dated *c.*150 CE;[63]
7) The inscription *RIÉ* 270, Greek version of *RIÉ* 185 (CAe 4851, ID: LIT4851greekRoyal);[64]
8) The text of the two inscriptions of the *Monumentum Adulitanum*, *RIÉ* 276 (CAe 5012, ID: LIT5012MonumentumAdulitanum1)[65] and *RIÉ* 277 (CAe 5019, ID: LIT5019MonumentumAdulitanum2)[66] as recorded by Cosmas Incopleustes in the sixth century BCE;
9) The map drawn in the manuscript Vat. Gr. 699, f. 15v of Cosmas, dated to the ninth century CE;[67]
10) Fra Mauro's map dated around 1450;[68]
11) Hiob Ludolf's map of 1683 (part).[69]

In the need to gather information from diverse sources, nothing has changed and there is still, if not more than before, when all was to be found

60 <http://www.perseus.tufts.edu/hopper/text?doc=Strab.+16.4&fromdoc=Perseus%3Atext%3A1999.01.0239>.

61 <http://www.perseus.tufts.edu/hopper/text?doc=Perseus%3Atext%3A1999.02.0137%3Abook%3D6%3Achapter%3D34>.

62 <https://betamasaheft.eu/works/LIT2170Peripl/main>.

63 <http://papyri.info/ddbdp/sb;24;16187>.

64 <https://betamasaheft.eu/works/LIT4851greekRoyal/main>.

65 <https://betamasaheft.eu/works/LIT5012MonumentumAdulitanum1/main>.

66 <https://betamasaheft.eu/works/LIT5019MonumentumAdulitanum2/main>.

67 <https://recogito.pelagios.org/document/vkliauohujyndq/part/1/edit>. The image uploaded to ↗Recogito for the ↗annotation was downloaded from <https://digi.vatlib.it/view/MSS_Vat.gr.699>.

68 <https://recogito.pelagios.org/document/xlidop0zvdwwwt>. The image uploaded to ↗Recogito for the ↗annotation was downloaded from <https://commons.wikimedia.org/wiki/File:FraMauroDetailedMap.jpg>.

69 <https://recogito.pelagios.org/document/t7e6q1bznco148>. The image uploaded to ↗Recogito for the ↗annotation was found via <http://www.georeferencer.com/> downloaded from <http://cuni.georeferencer.com/maps/540129335857/>. This map is also fully indexed and searchable in the Ethiomap project, <https://ethiomap.huma-num.fr/public/?id_article=22>, however it is not possible to link one of those names on the map to other resources, and thus the Recogito method has been preferred. This will be possible once the work for the resource development grant allocated to Ethiomap by the ↗Pelagios Network will be accomplished.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

in books, the need to deal with different types of data. The above examples present a variety of sources types:

1) The Pithom stele is an inscription in Hieroglyphic, of which the only version accessible to me was an English translation, available online;
2) The *RIÉ* inscriptions and the *Periplus* are relevant for the Beta maṣāḥǝft project and have been thus stored there;
3) The Greek and Latin texts come from ↗Perseus;
4) The maps have been loaded to ↗Recogito.

The ↗annotation methods are also different.

1) For the Beta maṣāḥǝft entities I could annotate directly in the XML;
2) The ↗Perseus texts have been annotated with <https://hypothes.is>;
3) ↗Recogito is also an ↗annotation editor and allowed me to select on the image a portion of text to transcribe and align to an identifier.

To be used together, these ↗annotations needed to be aligned to the same format. The ↗Recogito annotations were downloaded as ↗RDF from the software. One could have also carried out the whole annotation process there of course, uploading the texts and images in the same way, but I wanted to show this process moving as little as possible and focusing instead on the easiest way to make reusable annotations.

The ↗annotations on the Greek and Latin texts were (1) initially made as public annotations with <https://hypothes.is>; (2) then downloaded as ↗JSON from the ↗Hypothes.is ↗API;[70] (3) then transformed to ↗XML with <http://convertjson.com/json-to-xml.htm> and then transformed to ↗RDF-XML with a ↗XSLT.[71] I then (4) substituted the ↗URLs used by Hypothes.is with the canonical citation ↗URN provided by ↗Perseus,[72] where possible, and (5) used < http://www.easyrdf.org/converter> to make the ↗turtle files which I (6) validated with <http://ttl.summerofcode.be/> and <https://peripleo.pelagios.org/validator/annotations> to (7) store and (8) link from the ↗VoID file, describing the dataset contributed to the ↗Pelagios Commons project from Beta maṣāḥǝft.[73] The RDF-XML versions were stored instead in the local repository which indexes them as ↗RDF with the ↗SPARQL

70 <https://h.readthedocs.io/en/latest/api-reference/#section/Hypothesis-API-Reference>.

71 The transformation is available here: <https://github.com/BetaMasaheft/RDF/blob/master/transformations/HypothesisXML2PelagioOA.xsl>.

72 <http://www.perseus.tufts.edu/>.

73 <https://betamasaheft.eu/api/placeNames/void>. This method for publishing the Beta maṣāḥǝft dataset for ↗Pelagios, was kindly suggested by Rainer Simon who has generously assisted throughout.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

module for ↗eXist-db.[74] The annotations on the text in Beta maṣāḥəft, done directly in the XML, are also stored in the same ↗triple store by a routine transformation described in Ch. 7.

At this stage we then have all the ↗annotations associating portions of text to place concepts identifiers in the form of URIs, inside ↗RDF stores openly accessible, following the guidelines for the annotations offered by the ↗Pelagios Commons Cookbook.[75] They are also available in Pelagios already, for anyone to reuse them in Peripleo or via the Pelagios ↗API.[76]

3.1 Annotating sources related to the Red Sea

Each of the ↗annotation processes carries its own challenges. Let me start from the XML encoding in Beta maṣāḥəft.

Encoding is the way in which more information can be added directly, free of interfaces and software capabilities, and ↗TEI has established methodologies to encode properly, for example, multiple alternative identifications. The second inscription copied by Cosmas Indicopleustes, *RIÉ* 277,[77] has a whole list of places whose identification is dubious if not impossible, as we shall see, but even toponyms which, at first sight, would give no problem have been rightly put in question. In the following example, I have encoded the doubt regarding the identification of the word Νείλου in the text.

```
<placeName ref="LOC6850Takkaze" cert="low">
  <certainty locus="value"
                 match="../@ref"
                 assertedValue="pleiades:727172"
                 resp="#LIT5019MonumentumAdulitanum2"/>
  <certainty locus="value"
                 match="../@ref"
                 assertedValue="LOC4560Marab"
                 resp="#Fauvelle-Aymar"/>
            Νείλου
</placeName>
```

74 <https://github.com/ljo/exist-sparql>.

75 <https://github.com/pelagios/pelagios-cookbook>.

76 See respectively <http://peripleo.pelagios.org/> and <https://github.com/pelagios/peripleo/blob/master/README.md>.

77 In the Beta maṣāḥəft record RIE277 further bibliography can be found, <https://betamasaheft.eu/RIE277>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

While the most common and accepted identification is with the river Takkaze, between Aksum and the Səmen mountains, the text of the inscription refers to the Nile directly, and thus we can provide this information, using the standard methods of the <certainty> element.[78] However, recently Fauvelle-Aymar has argued in favour of an identification with the Marab river,[79] which is recorded in the same way, with reference to an identifier in the @resp which points to the @xml:id of the <bibl> element containing the reference to this publication.

This encoding allows us also to visualize synthetically important information directly with the text. In this example, the Beta maṣāḥəft text view (Fig. 4.2) offers, beside the direct link to the entity where attestations can also be gathered with their co-occurring named entities (Fig. 4.3), also the possibility to fetch the related entities from the ↗Pelagios ↗API and, in addition, the two alternatives to the identification offered as the preferred one, with the same options available. Beside the word Νείλου the arrow prompts to the Pelagios entities related to the main identification (the Takkaze river); the hand pointing to the word will produce a popup with the list of entities in Beta maṣāḥəft which have a reference to the main identification; the exclamation marks will present a popup with the alternative identifications.[80] Someone else reusing this data could infer some sort of relation between the three rivers and use it for other purposes.

78 <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CE.html> and the generic identifier for the Nile found in ↗Pleiades.

79 Fauvelle-Aymar 2009, 152.

80 See <https://betamasaheft.eu/works/LIT5019MonumentumAdulitanum2/text>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Fig. 4.2 The text of *RIÉ* 277 in Beta maṣāḥǝft.

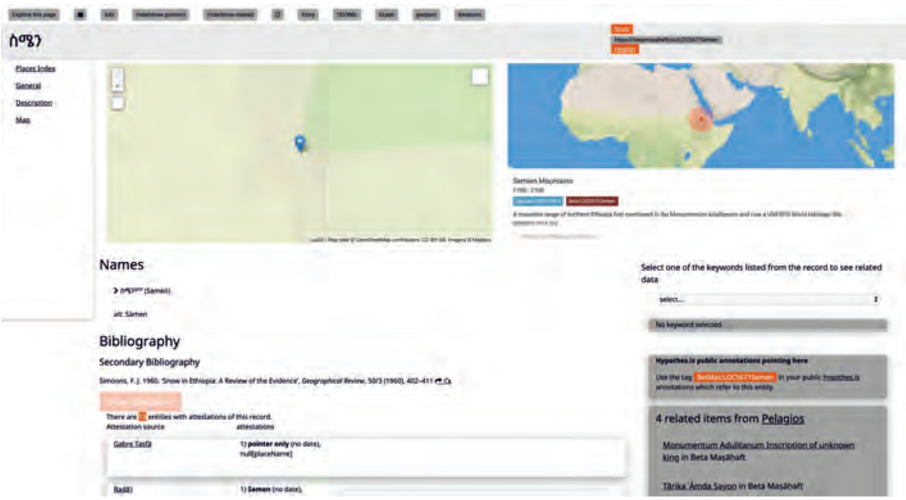


Fig. 4.3 The Beta maṣāḥǝft attestations view (beginning of a very long list) showing the related items from ↗Pelagios Commons of the place Sǝmen (LOC5671Semen). See <https://betamasaheft.eu/places/LOC5671Semen/main>.

↗Annotation with open access software like the ↗Hypothes.is service offers great flexibility and ease of access to such procedures and workflows, but

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

still needs decisions to be taken.[81] I had decided for this chapter to keep the ↗annotation to a trivial simplicity, adding simply in the annotation the single most appropriate place identifier available in Peripleo (Fig. 4.4). But there is still the problem of how to deal with place names unidentified or without an available identifier. I opted for a standard phrase ‘a place name’ or ‘ethnic’. For doubtful identification I added ‘perhaps’ before the identifier.[82] On the other side, as in ↗Recogito, each single annotation is identified with a ↗URL, which makes it possible to establish direct relations between exactly two such annotations.

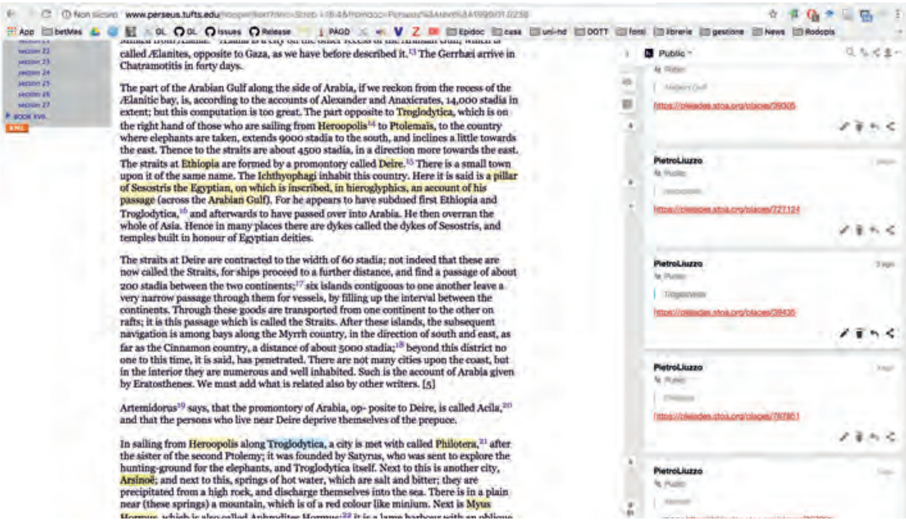


Fig. 4.4 ↗Hypothes.is ↗annotations on Strabo.

The ↗annotations offer already, taken one by one, the possibility, as shown, to make immediately visible connections between sources which share a reference to a specific place, like the Səmen mountains, which are in the Fra Mauro’s map as well as in *RIÉ* 277 and in many other later texts, or the Kinaidokolpitai, mentioned in an ostrakon and in *RIÉ* 277, or Leuke Kome, named in the *Periplus* and by Strabo.[83]

81 See for example the project Journey of the Hero which based its work on structured ↗Hypothes.is ↗annotations: <http://joth.perseids.org/joth.html#index>.
82 See <https://hypothes.is/users/PietroLiuzzo?q=place+name> and <https://hypothes.is/users/PietroLiuzzo?q=ethnic>.
83 ‘Periplus of the Erythreaen Sea’, *EAe*, IV (2010), 133a–134b (G. Fiaccadori).

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

At the same time it is clear, also to the machine, that the αἰθιοπικῶν in *RIÉ* 276 (on page 58 of the edition of the *Topographia Christiana*)[84] has little to do with the Αἰθιοπίας in *RIÉ* 277 (on page 63 of the edition of the *Topographia Christiana*, LOC7053Aithiopia, see below) or, at least, the second is under discussion, while the first is certainly the place concept which identified the land south of Egypt (pleiades:39274).

With these identifications, we also need not to pinpoint on a map the things we are not sure about. If to a place entity, like Gabazā, no location is associated, it does not need a place on the map.[85]

Now that we have all this information available, we want to get a general look at it. To produce the following maps with the Dariah-DE Geo-Browser a further transformation of the data to the ↗KML used by the web ↗application was made,[86] using ↗XQuery, starting from the XML for the text in Beta maṣāḥəft and from the ↗XML transformation of the ↗JSON obtained from ↗Hypothes.is for the Greek and Roman texts, in order in both cases to preserve the order in which the ↗annotations occur in the text and the context in which they occur.[87]

This basic format allows great analysis potential. While perhaps seeing all ↗annotations which can be placed on a map (Fig. 4.5) only provides for the extent of the general knowledge about the area and the quantity of annotations might hint at how well a place was known, looking at split sets of anno-

84 Wolska-Conus 1968, 58.

85 For the methodology used to add locations and names in ↗TEI inside the place entities of Beta maṣāḥəft, we follow the guidelines developed by Syriaca.org for the syriaca.org Gazetteer, Carlson and Michelson forthcoming. Huntingford 1989, 44 already treated the image in Vat. Gr. 699, f. 15v as a map, but the legend says more: it is a map put there to show a road to Aksum, so we could expect that the places marked are stops on the road to that city, i.e. you have to go through those places before reaching Aksum. Let us split the information. The identification of the two places north and south of ʾAdulis (Gabazā and Samidi) makes of Gabazā the harbour of ʾAdulis, where ʾAdulis is the port of Aksum. We know about Gabazā otherwise, from *RIÉ* 186, 13–14*RIÉ* 275, inscription of King Ousanas and the *Greek Martyrdom of Arethas of Nagran* VII.29. There is no reason to identify the *Telwnion Gabazas* with the port of ʾAdulis. The scale could be totally different from one place to the other, they are just relatively located as in an itinerary.

86 <https://geobrowser.de.dariah.eu/>.

87 It is not possible to reproduce the dynamic feature of the Geo-Browser in any screenshot, but you can load the ↗KML files which are provided here <https://github.com/BetaMasaheft/Documentation/tree/master/KML>, to obtain the same results presented here.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

tations together, relying on the colouring of the different sources done by the software, allows us to see immediately and precisely the overlaps.[88]

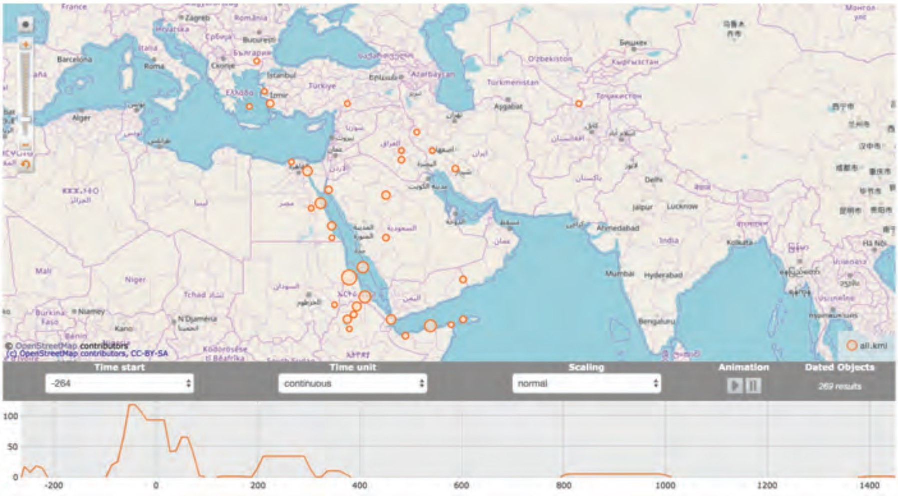


Fig. 4.5 All places in the annotated texts with coordinates.

In Fig. 4.6, which includes only Diodorus, Strabo, Pliny, and the *Periplus*, it can be immediately seen which places in the Red Sea were known by all or only some of the authors, and partially also the progressive increase of knowledge in time. Knowing that both Diodorus and Strabo worked on Agatarchides, and having also ↗annotations on the text summarized by Photius, it could also be used to study this lost author.

88 The same is possible in a dynamic way also in ↗Recogito.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

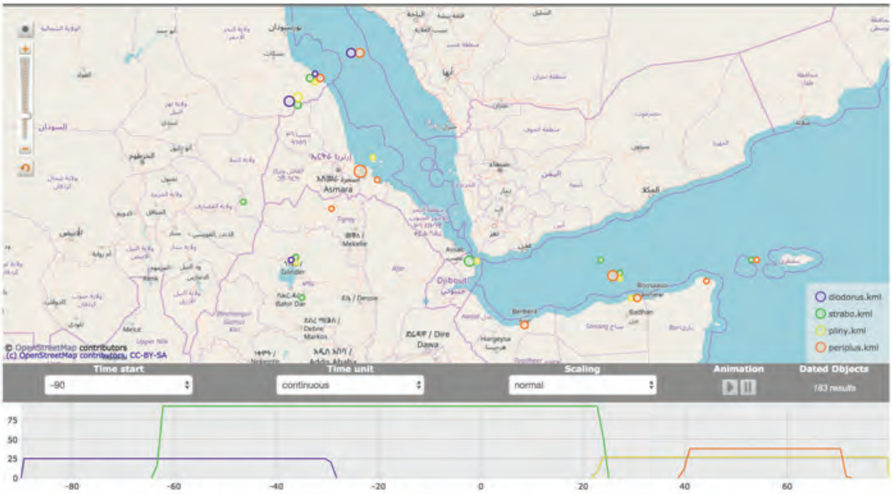


Fig. 4.6 All places annotated in texts from the literary Greco-Roman tradition.

Comparing by selecting only one type of written artefacts, for example inscriptions (Fig. 4.7), or defining a precise time range, for example only the first and second centuries (Fig. 4.8), might also lead to interesting observations. For example, in Fig. 4.7 it can be observed that the georesolved toponyms attested in the two inscriptions copied by Cosmas never overlap,[89] while there is an overlap with the Pithom stele for *RIÉ* 276 (Persia), while for *RIÉ* 277 there is one overlap with *RIÉ* 185+270 (Aksum) and one with the Pithom stele (Red Sea). Fig. 4.8, which excludes Diodorus, shows a consistent knowledge about the area for the different sources, shown by the clusters which often have three out of four (the ostrakon has only one toponym) sources attesting the same places.

89 On Cosmas, see 'Cosmas Indicopleustes', *EAe*, I (2003), 806a–807a (A. Sima).

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

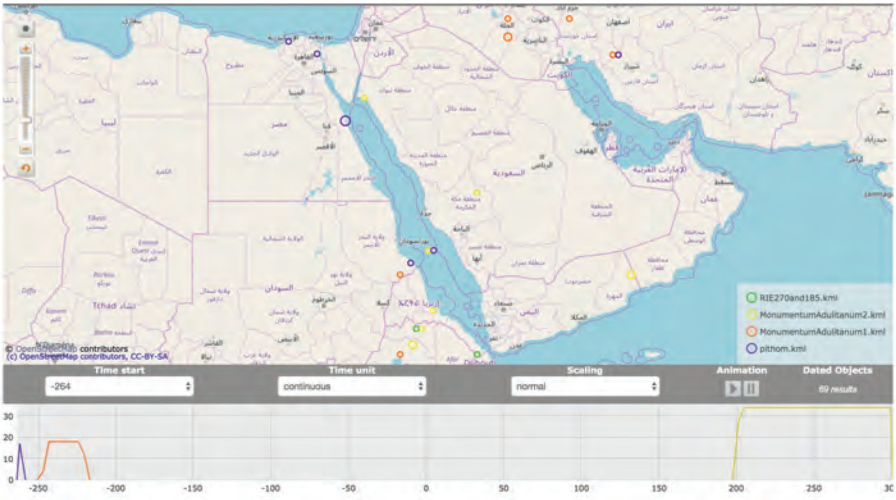


Fig. 4.7 All places annotated in inscriptions.

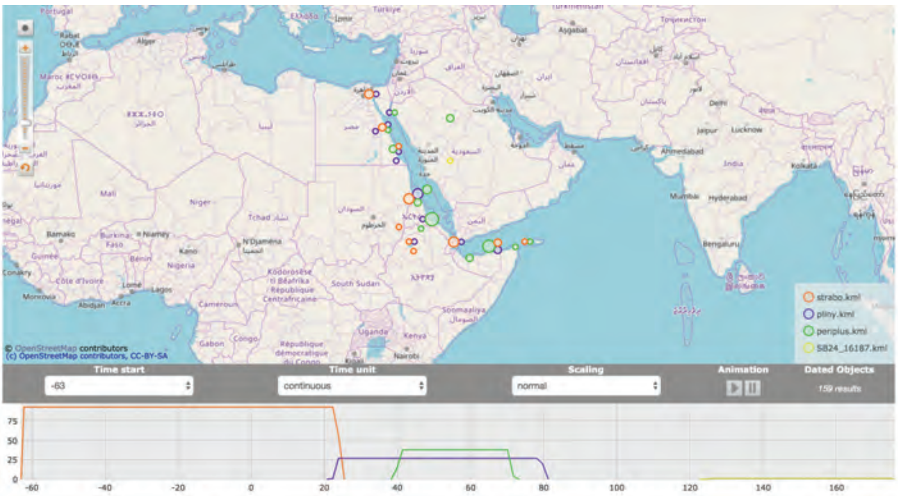


Fig. 4.8 All places annotated in sources from the first century CE to the second century CE.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Table 4.2 **Comparison of the sequence of places along the Red Sea mentioned by the sources**

Pleiades	**Pithom**	**Diodorus**	**Strabo**	**Pliny**	***Periplus***	**Cl. Ptolemy**	***Itineraria***
727124	Heroon-polis						Heroon-polis
727101	Kemuer = Arsinoe?	Arsinoe		Arsinoe		Arsinoe	
767851			Philotera				
							Serapion
786069		λιμὴν σκολιὸν ἔχων τὸν εἴσπλουν, ἐπώνυμος Ἀφροδίτης	Myus Hormus	Myus Hormus	Myus Hormus	Myus Hormus	
767851						Philotera	
						Leukos Limen	
						Nechesia	
785986		Berenike	Berenike	Berenike	Berenike		
		Ὀφιώδης	Ophiodes				
		λιμένος τοῦ προσαγορε-υθέντος σωτηρίας	harbour of the goddess Soteira				
39409	a great city to the king with the illustrious name of the king, the lord of Egypt, Ptolemy	Πτολεμαίου τὴν τῶν ἐλεφάντων θήραν	Ptolemaïs	Ptolemaïs	Ptolemaïs		

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Pleiades	Pithom	Diodorus	Strabo	Pliny	*Periplus*	Cl. Ptolemy	*Itineraria*
39271				ʾAdulis[90]	ʾAdulis		
39295		Arsinoe					
39335		Deire	Epidires				

But maps are not all that can be done with place ↗annotations, clearly enough. In Table 4.2 I expanded the comparison offered already by Desanges in which the sequence of places mentioned by each of the sources is put side by side.[91] This highlights clearly the continuity of some of the settlements in the sources (Myus Hormus, Berenike, Ptolemaïs) compared to others which appear only in one or the other source or which are not identified.[92]

The list would be much longer, including all those places which are without an identification, among which the several ports for the hunting of elephants mentioned by Strabo are very interesting.[93]

3.2 *RIÉ* 277

While the previous example looked at place attestations in different sources, this second example takes into consideration some of the diverse interpretations of *RIÉ* 277, one of the texts used in the previous example, which depend on the centre of the geographical representation of the account of the deeds of the ruler, whose name was probably mentioned in the initial part of the inscription, that was already lost at the time of Cosmas. This central point of the perspective is, in turn, the result of considerations about the other toponyms present in the text of the inscription.[94] The text is interesting also because it is a case where the study of written artefacts of the two kinds which I have discussed in Ch. 1 and Ch. 2, namely manuscripts and inscriptions, cannot be separated.

This inscription is known only via manuscript tradition, as its text is preserved in the *Topographia Christiana* of Cosmas Indicopleustes together with

90 Pliny NH 6, 172–173.

91 Desanges 1978, 263–279. In the table, the information about the *Itineraria Antonini* is taken from Huntingford 1980, 79.

92 See Burstein 1989, 6–8, for a description of the Ptolemaic settlements, canals, and roads along the Red Sea.

93 In the Dariah-DE Geo-Browser, these place names are listed as they appear in the ↗annotations in the table view below the map.

94 The hypothesis of a South Arabian king was also formulated and is discussed in Fauvelle-Aymar 2009, 143–144.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

RIÉ 276, an inscription of Ptolemy III Euergetes.[95] The text we have, which is complemented by inline scholia and notes, has to be studied and philologically confirmed before it can be used as 'epigraphic' evidence.[96] This is not an uncommon situation, as we have many lost inscriptions which we know about only from the reports of authors which have been transmitted to us in manuscripts. Herodotus already used inscriptions in his *Histories* and so did most historians after him. This text has also a long history of places identification which makes it an even more interesting example. The following is the list of hypotheses encoded in the XML:

1) The identifications in the scholia (flattened to a unique time after the composition and before the manuscript completion);
2) Salt's identifications in *Voyages and travels to India*,[97] based on Ludolf's map of 1683 annotated at <https://recogito.pelagios.org/document/t7e6q1bznco148>;
3) Huntingford's identifications with the integrations of Pankurst, which include the scholia and are based on the Aksumite kingdom hypothesis;[98]
4) The identifications discussed in the commentary to the inscription by Bernand. The author says that the point of departure of the campaigns might have been Aksum or ʾAdulis. These identifications are a collection and summary of some previous hypotheses like those of Dillmann, Littmann, Vivien de Saint Martin, and Kobishchanov;[99]
5) The identifications proposed by Fauvelle-Aymar, who argues for an otherwise unknown ʾAdulis kingdom/Trogodyte kingdom 'situé dans les basses terres côtières de l'actuelle Érythrée.' ('situated in the low lands on the cost of today's Eritrea.');[100]
6) The assumption attributed to Cosmas that this text was a Ptolemaic inscription like *RIÉ* 276.

The table (Fig. 4.10) discussed in the following paragraphs,[101] which aims at demonstrating the methodology only, would be much more useful if it included all the hypotheses made. With the data remaining openly available

95 'Monumentum Adulitanum', *EAe*, III (2007), 1010a–1012b (G. Fiaccadori).

96 The text is that of *RIÉ 277* (Bernand et al. 1991, 378–382) which is based on Wolska-Conus 1968 as documented in the Beta maṣāḥəft record, CAe 5019, ID: LIT5019MonumentumAdulitanum2.

97 Salt in Viscount Valentia 1809, 179–201.

98 Huntingford 1989, 40–47.

99 Bernand 2000, 32–45.

100 Fauvelle-Aymar 2009, 142–146.

101 Full view is available at <https://pietroliuzzo.github.io/DHEth/rie277table.html>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

in its latest updated form, as well as the script to produce the table,[102] the reader will be able to update the data with missing hypotheses or their own and reproduce the same table in an updated version or do whatever else they want with this data.

Additionally with respect to the encoding described above for <certainty> I have added some empty <seg> elements to mark the end of meaningful groupings made by the various scholars. These groups are important for the interpretation and have been named in various ways according to that. While Huntingford and Fauvelle-Aymar simply speak of 'groups', Salt and Bernand have a more interpretive denomination and speak of them as 'expeditions', which in Salt are also vaguely grouped in 'seasons'.

While I have used the <app> element to record the text of the scholia present in the witnesses of the text of Cosmas Indicopleustes, I have recorded as bibliographic entries the other works above mentioned and assigned to each an @xml:id to be able to refer to them in <placeName>, in the value of @resp inside <certainty> and in the <seg> elements used for the end of the groups.

The <desc> element allowed inside <certainty> provided space to complete the identifications. In most cases the identification process did not work from a list of identified places in a gazetteer, but from a search for similar place names or from a map at the disposal of the researcher.

The most interesting case is that of Salt's identifications, which are made on Ludolf's map of 1683, which is available in digital format online and happens to help much more to follow the argument than the 'corrected' map printed along his notes in the volume. I could then annotate the place names he associated with those in the text of the inscription directly with the names as written on the map (Fig. 4.9).

102 This ↗XSLT can be found here: <https://github.com/PietroLiuzzo/DHEth/blob/master/4%20Historical%20Geography/table.xsl>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Fig. 4.9 ↗Annotations of place names on the map of Ethiopia by Hiob Ludolf of 1683, made with ↗Recogito.

On the other side, especially in texts like this, it is more often the case that no precise identification is possible, but having the information organized allows us to have data which takes the available observations into consideration, even if these observations are never systematically made for each toponym attested.

The alternative visualizations of the XML source file help to clearly show the options at hand. Each of the sets of identifications could be mapped to a single map, like the one offered by Fauvelle-Aymar.[103]

Two examples are particularly relevant. The identification of the concept behind Αἰθιοπίας, as it is attested in this text in 'ἀπὸ δὲ δύσεως μέχρι τῶν τῆς Αἰθιοπίας καὶ Σάσου' ('to the west until the lands of Ethiopia and Sasou'), is crucial to the whole understanding of the text of the inscription. In my encoding, I have kept as valid the most specific possible meaning, meaning a concept based on this occurrence of the word (LOC7053Aithiopia). However, one could interpret this toponym as (1) embracing the Kushite kingdom

103 Fauvelle-Aymar 2009, 160.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

(LOC5017Nubia);[104] or (2) as the Ptolemaic concept of Ethiopia, which is distinguished by Trogodytice and is south of Egypt (pleiades:39274), given that, in the inscription, it is explicitly said to be to the west. This, however, means locating the centre of the geographic representation. Equally to be noted is the variance of the identification of the Nile (see above), which is associated with two different rivers (neither of which is the Nile). The process of ↗annotation thus allows one to precisely record the specific hypotheses made for each single fragment of information and revise them for new evaluations.

A table (Fig. 4.10), generated from the encoding of the above sources in the ↗TEI, might help to avoid the circularities by which the 'geographic order' of the text determines the centre of the geographic representation, but, at the same time, this centre determines the geographic order of the text and its meaning.

104 See for example Speidel 2016.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Fauvelle-Aymar **B**, Bernand **1**, salt **2**, Huntingford **2** — Καλαὰ

ID	desc	resp	related
pleiades:40169			Peripleo
	"Halaat or Salait" Salait on Ludolf map (1683) west of Axum	salt	
LOC5232Qohayt		OGIS	Peripleo
	la peuplade Kelau	dillmann	
	Kalaoué, district de la rive gauche du Takazzé, au-dessous des montagnes du Semen	vivien	

Fauvelle-Aymar **B**, Bernand **1**, salt **2**, Huntingford **2** — Σαμῆνε

ID	desc	resp	related
LOC5671Semen		salt	Peripleo
	"Sembre or Samen" Samen R. on Ludolf map (1683) west of Axum	salt	

Fauvelle-Aymar **B**, Bernand **1**, salt **2**, Huntingford **2** — Νείλου

ID	desc	resp	related
LOC6850Takkaze		salt	Peripleo
pleiades:727172		LIT5019MonumentumAdulitanum2	Peripleo
LOC4560Marab		Fauvelle-Aymar	Peripleo

Bernand **1**, salt **2**, Fauvelle-Aymar **C**, Huntingford **2** — δυσβάτοις καὶ χιονώδεσι ὄρεσιν

ID	desc	resp	related
LOC5671Semen			Peripleo

salt **2**, Fauvelle-Aymar **C**, Huntingford **2**, Bernand **2** — Λασινὲ

ID	desc	resp	related
pleiades:40192			Peripleo
LOC4259Lasta	"Lasta" Lasta M. hinc Lastenses on Ludolf map (1683) west of Axum	salt	Peripleo
	Nord d'Axoum	Bernand	

Fauvelle-Aymar **C**, salt **3**, Huntingford **2**, Bernand **2** — Ζαὰ

ID	desc	resp	related
pleiades:40347			Peripleo
LOC5597Sawa	"Shawa or Zaa" Sewa or Shawada P. on Ludolf map (1683) south-west of Axum	salt	Peripleo
	Nord d'Axoum	Bernand	
	Mount Zaad	Fauvelle-Aymar	

Fig. 4.10 Selected place identifications in *RIÉ* 277 plotted to a simple table from the XML-encoded text.

4 Georeferencing maps for research

Although in most cases with early maps georeferencing would not bring any interesting results, replotting ancient representations to a modern map can help to identify places and their whereabouts. In this last section, I will illustrate some simple examples using a specific tool, <https://www.georeferencer.com>. This is also another good way to visualize data and support annotations of maps or texts.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

If in most cases for later maps the matching of the places allows a fair amount of reliable overlap, for maps like Ludolf's, used by Salt, the comparison in parallel is a better way to approach the data (Fig. 4.11).

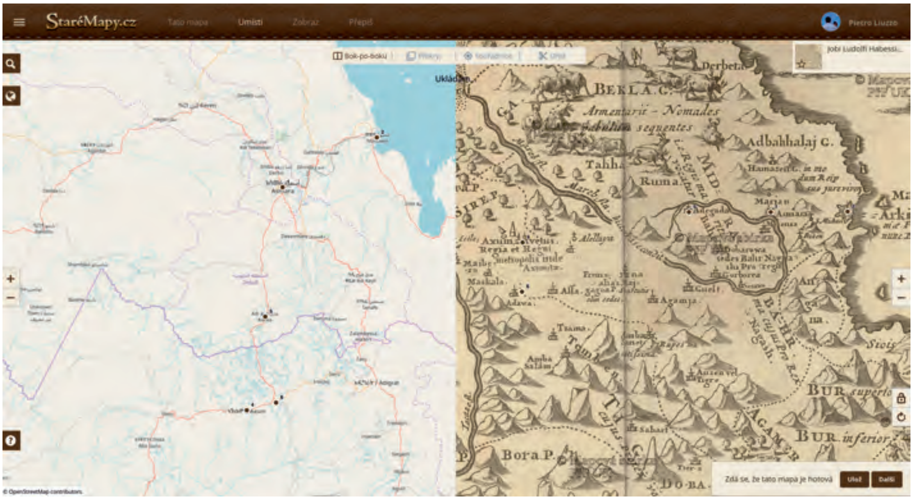


Fig. 4.11 Hiob Ludolf's map of Ethiopia compared with OpenStreetMap.

The encoding and the table in the previous example, or the georeferencing, certainly do not tell us who is the king, but we can see some of the considerations which have been made a bit more clearly.

This resource integrates large amounts of freely available sources, that is, digitized collections of old maps and easy interface-driven functionalities to overlay and compare these maps.[105] I have already shown the comparison function (Fig. 4.11), which allows one not only to put a modern and an old map side by side, but also to nail points of the old maps on the modern one.

For some maps, especially those drawn by hand, like the ones at the end of Huntingford's *Historical Geography of Ethiopia*,[106] this can help locate better some of the places plotted (Fig. 4.12 and Fig. 4.13).

105 For an overview and introduction to the spatial turn in digital humanities see Presner and Shepard 2016, 205–207.
106 Huntingford 1989.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Fig. 4.12 Overlay of Vat. Gr. 699, f. 15v, Egyptus Novelo map (1454), Fra Mauro's map (1460) replotted on modern basis, as well as the names in Zorzi's *Itineria* by Crawford 1958, figures 2, 3, and 20.

Fig. 4.13 Close-up view of the overlay of Huntingford's map 3.

An integration of this with ↗Recogito for the ↗annotations would make a very good step forward, allowing to do both the georeferencing and the place names transcriptions and georesolution in one place.

Tools like this, which are intuitive to use and accessible, can be of great help in the avoidance of common mistakes, like the ones made with map orientation, which led to famous works of literature pointing onto a reference map places whose location is actually unknown and exchanging their relative position, like in the case of Gabazā and Samidi of the Vat. Gr. 699, f. 15v in the

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

map printed by Conti Rossini.[107] On the other side, it should not be thought that the tools are the panacea and solution to all issues. Instead I hope I have shown how the issues multiply when the precision increases in the first section of this chapter.

The important work of indexing ancient maps done by the Ethiomap project is a very good example of the ground work needed to be able to bring together the information in texts with that on maps at a textual level.[108]

5 Conclusion

The study of ancient places in Ethiopia still needs major attention. After the work of Huntingford, no other attempt has been made to systematize the scattered but considerable knowledge we have of ancient Ethiopia at different stages. The systematic markup of place names and the organization of a gazetteer for ancient places in Ethiopia started only two years ago, with the Beta maṣāḥəft project, and the indexing of cartographic sources is for now limited to the efforts of the Ethiomap project.[109]

We would need annotated, critically edited chronicles (a task started by Solomon Gebreyes Beyene) and itineraries as well as ↗annotations of the accounts of the explorers in the fifteenth and sixteenth centuries. Additionally, a good interoperation with the annotation of sources in Arabic would be very important and the prospective research results are enormous, especially opening up to new forms of research done with maps, for instance 'critical practices of geo-temporal narration, forms of counter-mapping, and notions of "deep mapping" or "thick mapping" which privilege experimental navigation, time-based approaches, participatory mapping, and alternative rhetoric of visualization.'[110]

Relative locations and space definitions could be used for place identification and the structuring of the knowledge about ancient places, continuity and discontinuities, itineraries and roads, development and movements could be studied. This is a field of research requiring a collaborative effort and the

107 Conti Rossini 1928, Tavola XXVI.

108 At the 20th International Conference of Ethiopian Studies held in Maq̱ala in October 2018, a promising cooperation between the Beta maṣāḥəft Gazetteer and the Ethiomap project was initiated, which will lead at the geo-resolution of the place names in the maps to the URIs of Beta maṣāḥəft and to the production of ↗Pelagios ↗annotations also for this dataset, so that it will be possible from the maps to see texts related to the same place, and from the texts also the maps, collecting toponyms attestations from both types of written sources.

109 <https://ethiomap.huma-num.fr/>.

110 Presner and Shepard 2016, 207.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

use of resources which are open to contributions, accessible to collaborative efforts, and flexibly interoperable, so that, perhaps, if we did not manage by 2017,[111] by 2027 we could start being able to see the fruits of this work.

111 Elliott and Gillies 2009, §§ 1–5, see above.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Chapter 5. Dillmann's *Lexicon* as Online Resource

> [Theodor Holm Nelson] coined more obscure but extremely provocative terms, including 'transclusion' (a principle of his hypertext concept), 'intertwingled' (the complex interconnected condition of human knowledge), and 'teledildonics' (the technological facility for having sex at a distance).
>
> (N. Montfort, *The Future*, 2017)

The productivity of a language in its lexicon is a feature which does not escape markup languages and web applications. Programming languages might be a good example in their continuous proliferation. Nobody would doubt the core importance of a lexicon as printed work and also here, as for editions of texts,[1] a digital version should be able to offer even more and interact with the texts which use this language and, why not, also others.

In this chapter, I will describe the process which led to the establishment of the online *Lexicon linguae aethiopicae* as an 'intertwingled' resource in the World Wide Web.[2]

1 Introduction

'When does a certain word, or idiomatic expression first appear in literature? Where does it come from? When does it fall in disuse? How does its meaning change?'[3] Among the lexicographic resources available for Gəᶜəz,[4] the monumental work by August Dillmann is still the most important one which can be used to answer the above question,[5] especially in connection to sets of annotated texts, like the ones which the TraCES project is working on. However its usage was limited by the fact that it is written in Latin and accessible only in print.

1 See Ch. 3.

2 <https://betamasaheft.eu/Dillmann>. The work presented in this chapter was carried out for the ERC Advanced Grant TraCES: From Translation to Creation: Changes in Ethiopic Style and Lexicon from Late Antiquity to the Middle Ages (Grant agreement no. 338756).

3 Kropp 1994, 132.

4 See Kropp 2016, 202–204 for a state of the art until 2015.

5 Dillmann 1865.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

I will describe here the steps taken to transform the text, which had already been reproduced as a digital file, into an accessible online ↗application,[6] which, in turn, was connected to the resources in Beta maṣāḥəft described in the previous chapters. This is also offered here as an example of the many ↗up-conversion processes which have been involved in the setting up of Beta maṣāḥəft, especially to prepare the initial data and which are often involved in any task to produce a digital resource.[7]

2 Preprocessing

The ↗up-conversion process which took the *Lexicon linguae aethiopicae* from a book to a web ↗application serving this resource dynamically was not aimed at a simple digitization of this work but at gaining a starting point to expand and renew it into an updated lexicon of Gəʿəz texts, based on synergies with other resources. For example one aim was to be able to gather attestations of a word from a corpus directly, in order to expand the examples provided.

Heavy preprocessing of data has been carried out with ↗RegEx in ↗Oxygen XML Editor Editor to encode the text starting from a .txt version that was the outcome of many years of work and was started by Alessandro Bausi, with the help of Andreas Ellwardt, Irene Roticiani, and others.[8] This dillmann.txt file was developed in the TraCES project by Cristina Vertan and was enriched,[9] while it continued to be curated under the main responsibility of Andreas Ellwardt and more recently Magdalena Krzyżanowska. The data has been processed within this new ↗application project to improve the structure, the accessibility, readability, and the reusability of the large amount of information provided by the *Lexicon*. All parts of speech, abbreviations, citations, translations, notes, nesting of meaning, column breaks, internal references, and so on, have been marked up in ↗TEI ↗XML and the app relies on that markup as data structure.

Let me describe in a bit more detail the steps involved, starting from a minimal example extracted from the PDF of the *Lexicon* as reproduced in Fig. 5.1.[10]

6 The code of the ↗application is available at <https://github.com/BetaMasaheft/Dillmann/tree/master/gez-en-2.5> with some sample data.

7 A version of this chapter was presented at the 20th International Conference of Ethiopian Studies in Maq̲ala, in October 2018.

8 See Bausi 2016a and Bausi 2016b for more details.

9 This is just a conventional name for various files in .txt with parts or all of this text which circulated.

10 The PDF has long been available online in <https://archive.org> and later on at <https://www.tau.ac.il/%7Ehacohen/Lexicon.html> thanks to the work of Prof. Ran Ha-Cohen.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

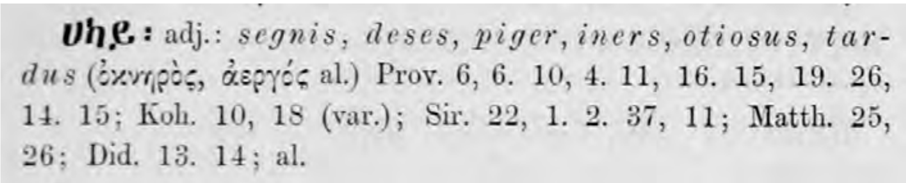

ሀካይ : adj.: *segnis, deses, piger, iners, otiosus, tardus* (ὀκνηρὸς, ἀεργός al.) Prov. 6, 6. 10, 4. 11, 16. 15, 19. 26, 14. 15; Koh. 10, 18 (var.); Sir. 22, 1. 2. 37, 11; Matth. 25, 26; Did. 13. 14; al.

Fig. 5.1 The entry ሀካይ in Dillmann's *Lexicon*.

Alessandro Bausi led the project For a New Lexicon of Ethiopic Texts, jointly funded by the Istituto Italiano per l'Africa e l'Oriente and the Istituto Universitario Orientale di Napoli (now Università degli Studi di Napoli "L'Orientale") as early as between 2000 and 2007. This project was conducted thanks to the personal efforts of the curator and some students, mainly from the University of Naples "L'Orientale". The project digitally recovered August Dillmann's *Lexicon linguae aethiopicae* according to Unicode standard. The *Lexicon* was digitized mainly by retyping it and was rearranged so that references and quotations were normalized and made searchable in the text file. For example, citations like 'Prov. 6, 6. 10, 4. 11, 16' (Fig. 5.1) were split up into 'Prov. 6,6; Prov. 10,4; Prov. 11,16' so that a pattern could be searched for as string or one could search all the occurrences of 'Prov.' to find all entries with some quotation from that work. Another important aspect was the encoding of page breaks and the integration in the main text of Dillmann's *Addenda et Emendanda*.[11]

This step was without a doubt the most important in terms of length and effort spent. As soon as the need arose for something more than searching into a file, a second step in the life of this resource was carried out. Cristina Vertan and Andreas Ellwardt, while continuing to hand fix the text, produced a .txt version, enriched with unique identifiers, and the example would then look like the following example.

11 See especially Ellwardt 2016, 94–96 for details on the *Lexicon* peculiarities.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

```
Lab84275dc11347ce818680a6949d3db1$ሀካይ$ adj.: <i>segnis</i>,
<i>deses</i>, <i>piger</i>, <i>iners</i>,
<i>otiosus</i>, <i>tardus</i>(ὀκνηρός, ἀεργός al.)
Prov. 6,6; Prov. 10,4; Prov. 11,16; Prov. 15,19; Prov. 26,14; Prov. 26,15; Koh.
10,18 (var.); Sir. 22,1; Sir. 22,2; Sir. 37,11; Matth. 25,26; Did. 13; Did.14.
```

This .txt file could be used in ↗GeTA to link morphological ↗annotations to a lemma and thus identify each annotated token properly. At the same time, the content was now in a non-proprietary format which could be read without the specific software used for the previous version. A third step took place with the beginning of Beta maṣāḥəft when the use of ↗TEI and ↗eXist-db made it possible to produce a text database fully linked with the *Lexicon*. The .txt version was taken to start from and produce TEI using elements from the dictionary module, like the one in the following example.[12]

```
<entry xml:id="Lab84275dc11347ce818680a6949d3db1" n="57">
  <form>
     <foreign xml:lang="gez">ሀካይ</foreign>
  </form>
  <sense xml:lang="la" source="#dillmann">
   <gramGrp><pos expand="adjectivum">adj.</pos></gramGrp>:
   <cit type="translation" xml:lang="la"><quote> segnis </quote></cit>,
   <cit type="translation" xml:lang="la"><quote> deses </quote></cit>,
   <cit type="translation" xml:lang="la"><quote> piger </quote></cit>,
   <cit type="translation" xml:lang="la"><quote> iners </quote></cit>,
   <cit type="translation" xml:lang="la"><quote> otiosus </quote></cit>,
   <cit type="translation" xml:lang="la"><quote> tardus </quote></cit>
   (<foreign xml:lang="grc">ὀκνηρός</foreign>,
   <foreign xml:lang="grc">ἀεργός</foreign> al.) <ref cRef="Prov." loc="6,6"
   >Prov.6,6</ref> ; <ref cRef="Prov." loc="10,4">Prov. 10,4</ref> ;
   <ref cRef="Prov." loc="11,16">Prov. 11,16</ref> ;
   <ref cRef="Prov. " loc="15,19">Prov. 15,19</ref> ;
   <ref cRef="Prov." loc="26,14">Prov. 26,14</ref> ;
   <ref cRef="Prov." loc="26,15">Prov. 26,15</ref> ;
   <ref cRef="Koh." loc="10,18">Koh. 10,18</ref> (var.);
   <ref cRef="Sir." loc="22,1">Sir. 22,1</ref> ;
   <ref cRef="Sir." loc="22,2">Sir. 22,2</ref> ;
   <ref cRef="Sir." loc="37,11">Sir. 37,11</ref> ;
   <ref cRef="Matth." loc="25,26">Matth. 25,26</ref> ;
```

12 <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

```
    <ref cRef="Did." loc="13">Did. 13</ref> ; Did.14; al.
  </sense>
</entry>
```

This stage of transformation, which I carried out between December 2016 and April 2017, provided additionally the markup of features, which could be isolated in the .txt using ↗RegEx:

1) Abbreviations, which are separated into
   1) Parts of Speech;
   2) Grammatical information labels;
2) Other abbreviations and bibliography;
3) Texts citations;
4) The different languages used for examples and parallels.

This was not perfect, as it can be seen also in the not matched 'Did.14' in the last line, for example. The identifiers were preserved in their original form to grant interoperation, and from this XML encoded text a range of new outputs could be generated, including the .txt from which it came.

↗Oxygen XML Editor has a very powerful search and replace functionality which allowed searching patterns with ↗RegEx and replacing with markup.[13] Searching for (L[0-9A-Za-z]+)(\$)([\p{InEthiopic}\s\w\.,]+)(\$), for example, matching in our example 'Lab84275dc11347ce818680a6949d3db1$ሀከየ$', isolated parts of the string corresponding to the identifier and the main lemma associated to it. The first group ((L[0-9A-Za-z]+)) could then be used in the @xml:id of <entry> and the third (([\p{InEthiopic}\s\w\.,]+)) as content of <form>.

The possibility to search codepoints belonging to different Unicode Charts (e.g. \p{InEthiopic}) allowed us to easily isolate the many languages used by Dillmann in his Latin text: in particular Arabic, Coptic, Gəᶜəz, Greek, Hebrew, and Syriac.[14]

The text upconverted in this way was then run through an ↗XSLT transformation to add the ↗TEI header and a few other features, such as the progressive sequence of the entries. Instead of one huge file, one file for each lemma was created, thus about 12,000 files, named after the unique ID of the entry.

13 This could have been done in many other ways with a script, but by the time I had figured out the patterns and tested the replacements, the task was accomplished and no need arose to put it in a script, given it was not to be iterated.

14 See Weninger 2016 on the use of Arabic in Gəᶜəz lexicography. Lists of these can be found at <https://betamasaheft.eu/Dillmann/reverse?start=1&mode=foreign>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

This was still far from being completely clean, and additionally it inherited some conversion errors and mismatches. However a set of possibilities became available on the basis of the structured data, most of which could be served via a web ↗application (Fig. 5.2), whose primary aim was to deliver this resource to a greater number of potential users from a simple database.[15]

Fig. 5.2 The entry ሀካይ in the online Dillmann's *Lexicon linguae aethiopicae*.

Not all features were marked at this stage and there was some notable semantics of the original text which escaped this rudimentary step of mass encoding.[16] For example, the classification of verbs was not encoded nor was the grouping of the lemmas belonging to a same stem (Fig. 5.5), done by Dillmann with the use of lines between groups of entries, which were already omitted at the stage of conversion to a text file. Leonard Bahr added back this information in the XML data at the beginning of 2018, so that the information is now restored.

In the following pages, I will discuss the added features which could be gained with the structured text to improve accessibility and readability of this core research resource, which was largely inaccessible due to its main language and lack of availability.

15 This is also an ↗eXist-db ↗application.

16 Not all of these have yet been corrected.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

3 Features

This web ↗application was built with ↗eXist-db, which uses Lucene as the standard search engine. This meant that the simple fact of loading the ↗XML data and configuring the indexes in the XML database made a countless series of additional things possible, like the ones of the Lucene indexes themselves which allow to search in the indexed text with wildcards or adding a degree of 'fuzziness'.[17] Thanks to another module available in eXist-db,[18] results of the searches can be easily displayed as ↗KWIC highlighting the match of the query.

17 <https://lucene.apache.org/core/2_9_4/queryparsersyntax.html#Fuzzy%20Searches>.

18 <http://exist-db.org/exist/apps/fundocs/view.html?uri=http://exist-db.org/xquery/kwic&location=&details=true>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

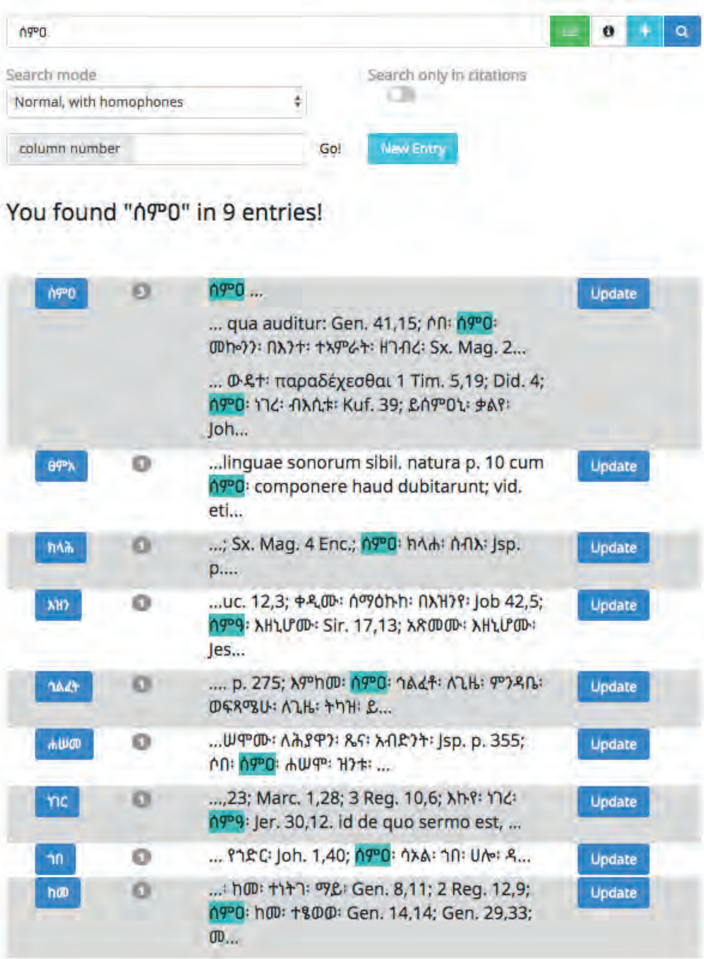


Fig. 5.3 KWIC search results.

However, there were other things which had to be improved for a search to be more powerful than that run on the Word document. A series of functions was added before passing the search query to the indexes which would cope with orthographic variants of Gəʿəz, that is, letters which are in practice

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

interchangeable and would force the user to try all possible variants to find out the one chosen to be recorded by Dillmann.[19]

1) ሀ, ሐ, ኀ, ሃ, ሓ, ኃ;
2) ሀ, ሐ, ኀ;
3) ሁ, ሑ, ኁ;
4) ሂ, ሒ, ኂ;
5) ሄ, ሔ, ኄ;
6) ህ, ሕ, ኅ;
7) ሆ, ሖ, ኆ;
8) ሠ, ሰ;
9) ሡ, ሱ;
10) ሢ, ሲ;
11) ሣ, ሳ;
12) ሥ, ስ;
13) ሦ, ሶ;
14) ሤ, ሴ;
15) ጸ, ፀ;
16) ጹ, ፁ;
17) ጺ, ፂ;
18) ጻ, ፃ;
19) ጼ, ፄ;
20) ጽ, ፅ;
21) ጾ, ፆ;
22) አ, ዐ, ኣ, ዓ;
23) ኡ, ዑ;
24) ኢ, ዒ;
25) ኤ, ዔ;
26) እ, ዕ;
27) ኦ, ዖ.

By searching one or the other of the letters in any of these groups also all other options are considered, building to hundreds of optional strings to be searched at times.

The potential of the marked up text is also leveraged with some of the search options available with the + button on light blue background on the side of the search box (see Fig. 5.3 top right corner), which loads several search options. Here the user can search all those entries which have a certain value for the <pos> element, for example, meaning find all marked nouns in Dill-

19 This feature has been inherited also from the Beta maṣāḥəft ↗application, where it is the standard behaviour of all searches.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

mann, or all conjunctions or all those entries which do not have an Arabic string in them.

The XML source allows one to easily reproduce a PDF and print it out on demand with editorial and layout decisions which can improve the readability of this resource as well.[20] In case of doubt, it is always possible to check the PDF of the original in Ran HaCohen's website, which is linked at the top of the record. The link is computed on the basis of the <cb> elements in the XML and relies on the very tidy organization of the PDF provided online.[21]

Additionally, a web visualization of structured data can offer a clearer presentation of the nesting of different meanings as in the example below (Fig. 5.4), and can reproduce features of basic search functionality in word processors like highlighting the searched text in the record.

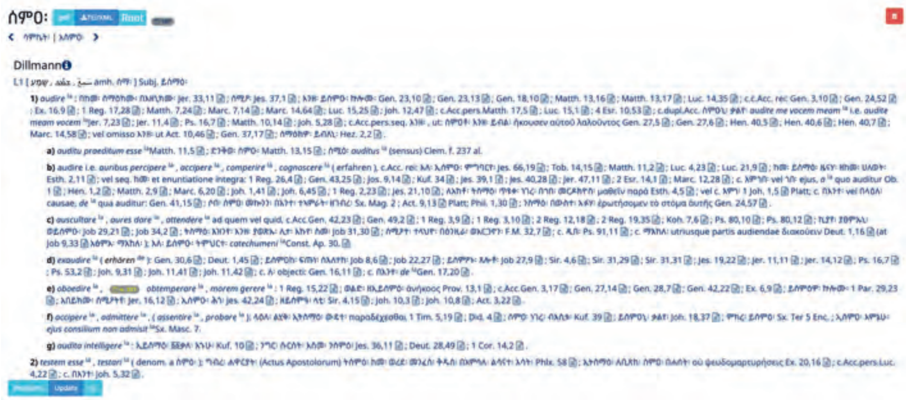


Fig. 5.4 Nesting of the different meanings of **ሰምዐ**.

Additional information can be gained when needed, if, for example, the user wants to know what kind of actions have been carried out (Fig. 5.5) or in which group of lemmas Dillmann put the current one.[22]

20 A reprint of the updated version is also possible and is foreseen for a point in time when the cleaning of the data and the integration of other resources will be completed.

21 <https://www.tau.ac.il/~hacohen/Lexicon.html>.

22 For the other search feature of the ↗application, please see the help box online.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

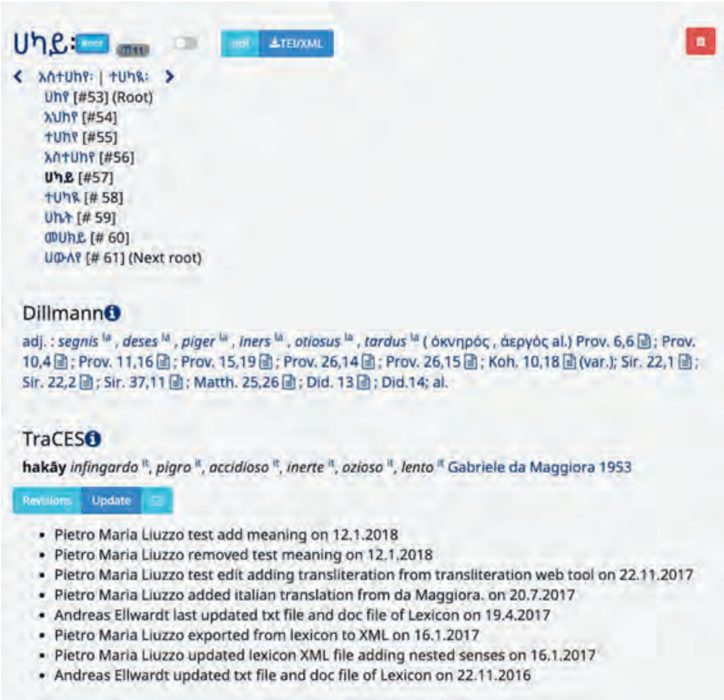


Fig. 5.5 Semantic grouping by Dillmann.

4 Edit dictionary records

The newly structured XML text can be edited with any XML editor directly, but the initial project requirements prompted the development of a further feature, an online editor which would allow the correction and normalization effort to continue as well as the addition of new lexical sources and of new lemmas or new meanings for a known lemma. The editing workflow implemented uses only an ↗HTML form (Fig. 5.6), enhanced by some ↗JavaScript for a few functionalities and enriched with data entry guidelines in the same page.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

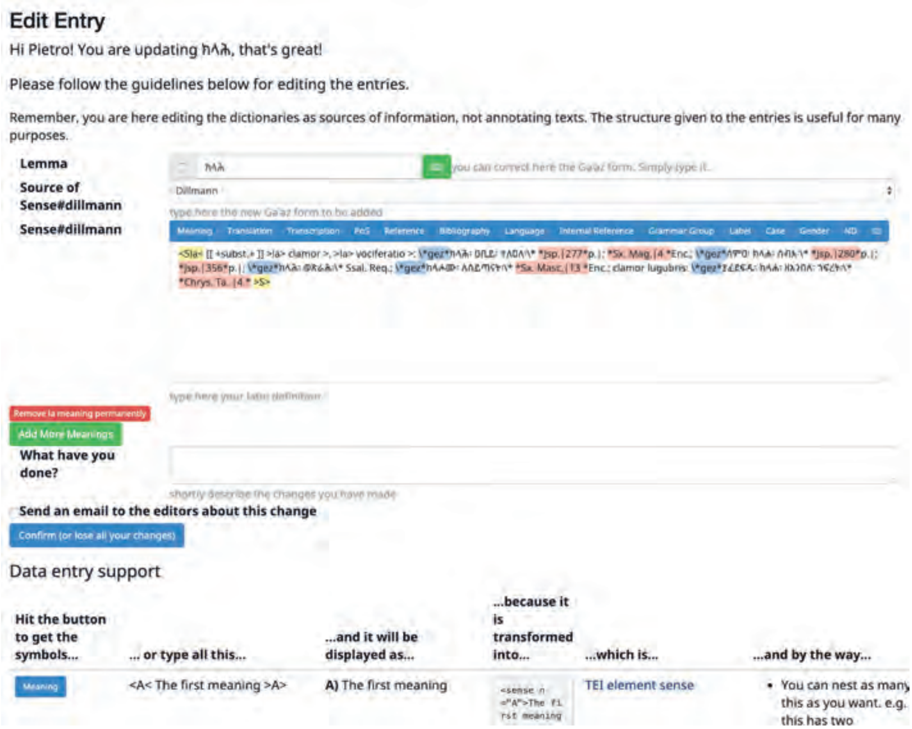


Fig. 5.6 The editor of the online *Lexicon linguae aethiopicae*.

The XML is made available for editing in the form transformed with a ↗XSLT script into a string which uses some conventions for marking up the text, which can be keyed in more quickly, and highlights some of these to facilitate readability. The text conversion of our example looks like in the following example.

<Sla<
[[+adj.+]]:
>la> segnis >, >la> deses >, >la> piger >, >la> iners >, >la> otiosus >, >la> tardus >
(*grc*ὀκνηρός*, *grc*ἀεργός* al.)
Prov.|6,6; *Prov.|10,4*; *Prov.|11,16*; *Prov. |15,19*; *Prov.|26,14*; *Prov.|26,15*;
Koh.|10,18 (var.);
Sir.|22,1; *Sir.|22,2*; *Sir.|37,11*;
Matth.|25,26;
Did.|13; Did.14; al.
>S>

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

It can be easily observed how the >la> … > string pattern corresponds in the ↗TEI to //cit[@type='translation']/quote. On saving this string the text is newly upconverted with a simple ↗XSLT to the ↗XML entry in the ↗eXist-db database and is validated against the ↗schema before being stored in order to be able to report any error before it can make its way to the file system. The one-to-one correspondence of matchable patterns and elements in the TEI makes the bidirectionality of the transformation safe.[23] The string patterns to be used are listed just below the editor with examples and values.

New meanings can be added from other sources, like the vocabulary by Da Maggiora in our example (Fig. 5.2), which has been added with a bibliographic reference to the source as an additional meaning entered by the TraCES project, rather than from the *Lexicon*.[24] In this case both the provenance of the data entry and the provenance of the actual information are provided, the first in the section header, where it is stated that this is an addition made by the TraCES project, the second as a bibliographical reference. In fact a new meaning added by members of the TraCES project or other contributors could be potentially based on many different sources, lexicographical or not.

The same form is used to create new entries, with some additional features to check for the existence of a lemma. On accomplishment of any of these basic operations of creating, updating, or deleting, the group of editors can be informed with an email notification, so that they are made aware of more relevant changes if the author of the edit thinks that this is necessary.

5 Linking the online *Lexicon linguae aethiopicae* to Beta maṣāḥəft

The indexed texts in Beta maṣāḥəft offer a corpus of texts increasing daily which can be used to give to the user a list of attestations of any given lemma in the online *Lexicon*. Pending the full deployment of the Gəʿəz Morphological Parser (Ch. 6), the online *Lexicon linguae aethiopicae* currently sends a request to Beta maṣāḥəft for the exact string used as main lemma and the results of this query are displayed as attestations below the entry with ↗KWIC.

These attestations have a different value from the selected passages in the source which can point to passages containing examples with any of the possible morphological forms. These are instead directly linked to Beta maṣāḥəft

23 The concept was developed on the example of the Leiden+ conventions used in <https://papyri.info> to edit ↗EpiDoc of inscriptions. <https://papyri.info> uses a different technology to do this.

24 Da Maggiora 1953. In this way extracting only the data originally from Dillmann's *Lexicon* is still possible, as it is to isolate what is new.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

when possible. The format of the encoded reference, as we have seen, looks like in the following example after being marked up.

```
<ref cRef="Gen." loc="31,12">Gen. 31,12</ref>
```

The information is not sufficient to resolve the citation to a passage in a text. First of all, the abbreviation must be paired with the available text to which it refers. In this case, the texts are in Beta maṣāḥǝft and we thus need to have a table where each standardized abbreviation is associated with the identifier in Beta maṣāḥǝft of the cited text, textual unit, or manuscript transcription, where Dillmann uses that directly.

Once we know that 'Gen.' corresponds to CAe 1546, ID: LIT1546Genesi with an association in a table, which is, in our case, another XML file, we can look up that resource which is available online to fetch the piece of text in the citation. For this purpose the text in question needs to be marked up in a way that it can answer to that request. Our Genesis text available online will thus have to have some markup for the chapters and verses in it, which are used by the citation to point to a specific passage. Thanks to the generous contribution of Ran HaCohen, this is the case.[25]

However this text structure is by no means unique or unequivocally reliable and thus a step is necessary in between to present this text structure, whichever it is, to resolve the format of the given citation. This intermediate data model lets us know that CAe 1546, ID: LIT1546Genesi is organized in chapters and verses and that the citation's first digit refers to the chapter and the second to the verse, but when presented with a citation like 'Kid. f. 9'[26] will look at the relevant resource (a manuscript in this case) for the page break corresponding to the digit after 'f.'. Beta maṣāḥǝft serves this kind of information using the ↗DTS ↗API specification,[27] for all and any text present in the written artefacts (mainly manuscripts) or textual units (works) collections.[28]

This means that I can use the correspondence to 'Gen.', namely CAe 1546, ID: LIT1546Genesi, and the information in the value of `@loc` inside my `<ref>` to send a request to the Beta maṣāḥǝft ↗API and retrieve the passage cit-

25 See <https://betamasaheft.eu/LIT1546Genesi>.

26 Ellwardt and Dickhut 2018.

27 <http://w3id.org/dts>. Although we implement the latest draft to date, the version of the script actually used by the online *Lexicon linguae aethiopicae* is an older, not fully conformant one, which remains in parallel with the newer implementation in the system. The update is due soon. See Ch. 7, p. 187.

28 See Ch. 1 and Ch. 2.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

ed. The ↗JavaScript performing this will then build a request ↗URL as defined by the API specification, with the pattern /api/dts/document?id=urn:dts:betmas:LIT1546Genesi:31.12, and will get the defined response,[29] which is a fragment of the ↗TEI file selected according to the structure of the information and the parameters passed with the request. The logic in the ↗DTS implementation will, given the identifier, look at the file and know that the request is for Chapter 31 and Verse 12 and will fetch that from the TEI. It will then take the specified text and return it to the online *Lexicon linguae aethiopicae* as in Fig. 5.7.

The same ↗DTS ↗API, receiving a request for a text of which a record exists in Beta maṣāḥǝft but perhaps without any text, will at least link to the main page about that textual unit. If the reference is to a manuscript, the API will perform the appropriate reasoning, so that, if a transcription is present, it will return the range indicated looking at <pb> and <cb> milestone elements in the encoded transcription.

29 The current implementation still uses the pattern /api/dts/{$id}/{$level1}/{$level2} and the response is in a bespoke format. The code of the ↗DTS module is available with the Beta maṣāḥǝft ↗application (<https://github.com/BetaMasaheft/BetMas>), in the dts.xqm module, although, as any ↗API implementation, is very specific to the data presented.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Fig. 5.7 The text of Gen. 30:35 fetched in the online *Lexicon linguae aethiopicae* from Beta maṣāḥəft via ↗DTS.

6 Indexes and navigation

The ↗application in ↗eXist-db, using ↗XQuery can also offer a number of other ways to navigate the data. The online *Lexicon linguae aethiopicae* offers a browse function to see all lemmas as cards with the translation and a citations index which gives access to the data starting from the texts referenced and gives explicit evidence of the association of the citation with a text in Beta maṣāḥəft. Also a reverse index is available (Fig. 5.8) where the user can browse attestations of words in other languages and see entries where those are in an ordered list instead of simply using the advanced search.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

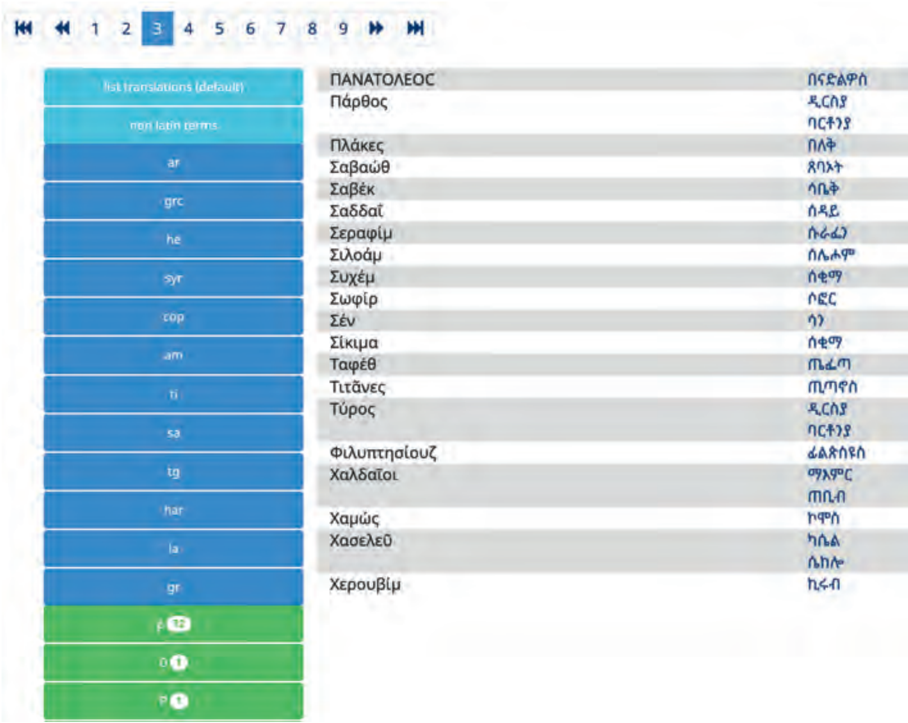


Fig. 5.8 Reverse indexes in the online *Lexicon linguae aethiopicae*.

Also the bibliography is navigable with backwards references at the exact places of use. This is of course true only for those bibliographic references which have been marked in the text as such, so that, for example, where Dillmann had a reference as 'Lud. Comm. hist.', this cannot be picked up but, as soon as it is updated with the proper ↗Zotero tag bm:Ludolf1691Commentarius, we can have a link to the full record and it will be counted in the bibliographic index.[30]

7 External connections

Whilst all these remain uses of the data within the *Lexicon*, there are many other ways the data can be used by external resources. The TraCES tool ↗GeTa, for example, uses the data ↗API to access the information and add it to the

30 <https://betamasaheft.eu/Dillmann/bibl.html>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

↗annotations.[31] For example, every text in Beta maṣāḥǝft is linked word by word to the Dillmann's *Lexicon* by adding during the transformation phase to each space-separated word a link to a fuzzy search in the *Lexicon*, together with other optional actions starting from the selected word.[32] Due to the way in which the language is structured and consequently the dictionary, it is not possible to search for the exact lemma in the *Lexicon* which is relevant for a given word in a text, but the fuzzy search provides a decent landing.

The more precise identification of the lemma to be linked in the *Lexicon* depends on the morphological analysis of a given string and there might be indeed several relevant entries to link to. In the next chapter (Ch. 6) we will look at a tool written in ↗XQuery, which tries to close this gap and offer morphological analysis. The data of the *Lexicon* is used there in the first place to support the process of morphological analysis as the next chapter will show, and then to validate the hypotheses made by the parser. For each acceptable hypothesis the verified value becomes directly a link to the relevant lemma in the *Lexicon*, thus serving a second function directly.[33]

31 <https://betamasaheft.eu/Dillmann/apidoc.html>.

32 See Fig. 6.5.

33 Other reuse of the data is highly encouraged and supported by the data ↗API.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Chapter 6. Gəᶜəz Morphological Parser

(Wurzel. / Wurzel Abrahams. Wurzel Jesse. Niemandes / Wurzel – o / unser.)
(P. Celan, *Radix, Matrix*, 1963)

A well-trained eye will be able to parse a Gəᶜəz word and understand its morphological structure in few milliseconds using intuition, analogy, and experience. There are cases yet when the attested form under scrutiny will be problematic and require the reader to formulate hypotheses to understand it. A non-well-trained eye would take some time to parse a word, and a untrained eye has no chance, even knowing the script.

However, even if a computer cannot be trained as a human and cannot learn a language in the same way in which we do, it can apply the same rules and methods to formulate and verify hypotheses and thus help any kind of users with different competence levels. In this chapter I will present a small prototype ↗application (one ↗XQuery module and ten ↗XML files)[1] which parses a Gəᶜəz word to return plausible hypotheses for a morphological description of that word.[2]

The parsing of Gəᶜəz poses specific problems, which will be discussed, and some which are instead common to Semitic languages.[3] At the same time, the

1 These files, as it will be detailed later in the chapter, contain only morphological information in the form of tables, they do not include the morphological ↗annotations themselves. The first file is the one called letters, containing a model for each sign and its transliteration. The second most important one contains the list of patterns, modelled according to the structure of the verb. A third file is for conjugation and contains a self-made model of the verb conjugation containing the distinctive features. A similar file has this information for pronouns. Two files contain nominal forms and suffixes, and the remaining contain lists of invariables such as particles, numbers, and proclitics. For the validation, one last file contains an extract of all lemmas in the *Lexicon* which is indexed separately to speed up lookup. <https://github.com/TraCES-Lexicon/lexicon/tree/master/geezParser>.

2 There are several other word analysers or parsers, for Latin, Greek, Arabic, and other languages. See for example Ruzicka 2018 and SomMorph for Somali, Jama Musse Jama n.d.

3 Other examples are for Arabic, <http://elvira.lllf.uam.es/jabalin/analizarForma.php>, or ElixirFM, <http://quest.ms.mff.cuni.cz/cgi-bin/elixir/index.fcgi>; and for Syriac, <https://sedra.bethmardutho.org/lexeme/paradigm>. The main reference examples

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

systematic vocalization of Gəʿəz helps more than a little in the parsing, which is thus easier than for other languages.[4] I will present here some of the problems I encountered while building a parser for Gəʿəz and some of the solutions adopted. I tried to replicate human processes of hypothesis production and validation and only a grammar-based knowledge of the language,[5] which I never practised myself.

1 Principles and aims

To let the machine understand the morphological aspects of a word, we can either use a database of annotated forms or reconstructed possible forms or give the parser engine the same information we would use and let it calculate the way we would do that. While the morphological information useful to the parser is in the ten files mentioned above, for the Gəʿəz parser I was able to use also a set of existing morphological ↗annotations on a selected corpus of texts, produced by the TraCES project and exported from the ↗GeTa tool as ↗TEI feature structures.[6] In this dataset the transliteration was manually corrected from a first automatic run and then annotated using the tool which kept it aligned to the *fidal*, that is the Ethiopian syllabic script.[7] The export looks like the following example:

```
<fs type="graphunit" xml:id="Wf2e5da2f-322e-4eab-94df-19c2487220c3">
<f name="fidäl">በስመ</f>
<f name="translit">ba-səma</f>
<f name="analysis">
    <fs type="tokens">
        <f name="lit" fVal="ba">
            <fs type="morpho">
                <f name="pos">Preposition</f>
                <f name="state">Nominal state</f>
                <f name="lex">L79528bf91b00402f9cb3505bd6ea3010--በ</f>
            </fs>
```

are, however, the Greek Word Study Tool and the Latin Word Study Tool of the ↗Perseus project, <http://www.perseus.tufts.edu/hopper/morph>.

4 The parser presented in this chapter is available at <https://betamasaheft.eu/morpho> and was developed for the ERC Advanced grant TraCES, Grant Agreement no. 338756.

5 Grammars used are mainly Procházka 2005; Tropper 2002; and Dillmann 1907.

6 <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/FS.html>. The mapping and transformation was done by Cristina Vertan.

7 See Vertan 2016 and Hummel and Dickhut 2016. *Fidal* is the script used in Gəʿəz, ‘Script, Ethiopic’, *EAe*, IV (2010), 580b–585a (S. Frantsouzoff).

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

```
        </f>
        <f name="lit" fVal="səma">
            <fs type="morpho">
                <f name="pos">Common Noun</f>
                <f name="gender">unmarked</f>
                <f name="number">SingularP UnmarkedS</f>
                <f name="case">Nominative</f>
                <f name="state">Construct state</f>
                <f name="lex">L9ad0e25e4767494686ee0bb77e9571e5--ስም</f>
            </fs>
        </f>
    </fs>
  </f>
</fs>
```

This can be easily indexed to be able to quickly run a query for a string in either transcription or *fidal*, or to match a given morphological form and find attestations of it, or to look up all ↗annotations of a lexeme, in whichever form using the content of <f name="lex">, which contains an ID identifying an entry in the Dillmann's *Lexicon linguae aethiopicae* online.[8]

The TraCES project team also provided tables of the most frequent schemas of the verb and noun formation in Gəᶜəz including *fidal*, transliteration, and affixes, which have been modelled in the ten ↗XML files used by the script.[9] This information can instead be used to analyse the words, and this is the way we use tables in any grammar book for any language.

The fundamental difference between using the annotated texts and looking up information in reference tables is that, while the first looks up a list of individually annotated occurrences of a word and returns the ↗annotation of that precise word in that context and helps by providing possibly analogous cases, the second approach applies rules to give absolute options, which are not grounded on evidence but only on rules. The first approach allows us to make analogies while the second gives us options which we can test in the given context, eventually also with the aim of producing annotations which will be then fed to the previous method. Both methods have benefits and, in the best of possible worlds, one wants to make a valid hypothesis, and know of other options, but then also have the possibility to verify its actual existence, because language is not only rules.

8 See Ch. 5.

9 These schemas are in no way to be confused with schemas for ↗XML validation, often referred to in other chapters of this book.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

A morphological parser like this, in principle, takes a string as input and analyses it to get as much information as possible. In this case, I have tried to gain this information in both of these two ways, for each separable bit of information. Each word is searched in the TraCES annotated texts and an attempt is made to analyse it with rules only. If a result is found and has a matching entry in the *Lexicon linguae aethiopicae*, then the ID is used also to link to ↗annotations in the TraCES corpus which have that identifier, as an additional validation for the user.

Using this system, the software should be able to make many of the mistakes we can make, both annotating or writing code to analyse, but can provide options which we did not think of and becomes, in this respect, potentially useful especially to people which already have some knowledge of the language.

2 *Fidal*, transliteration, schemes, and affixes

I will first describe the types of information available. We have first of all the *fidal* forms. There are then transliterations of these forms and the schemas or patterns. The schemas present, in a simplified format, all the information needed to convert a verb root into each of its paradigmatic forms. For example, let us take as example a form **ወለደ**, which itself is, in this example, in the schema form **1a2a3a** (or more precisely Wa2a3a), that is the schema of the perfect. If another schema is given, for example **yəWa22ə3**, which identifies the imperfect, we know that there is a prefix (*yə*), the first position which is a **ወ** in the first order as the *a* tells us, and then the second consonant in the root is geminated and in the sixth order (*ə*), while the third is also in the sixth order but the vowel is not expressed in the pattern. We can thus figure out, replacing the figures with the correct consonants and writing it up in *fidal*, the impossible form **ይወልልድ**, which is actually **ይወልድ** because *fidal* does not show the geminated consonant sound. Also to be noted already in this process is that the distinction which the schema records between **2ə** and **3**, namely the presence or not of the vowel, is also not reflected in the *fidal*.

These last observations are a first item of interest for our parser, because that means that it is not true that, given a form **ይወልድ**, one can automatically derive its transliteration or its schema.

A mechanical conversion of the string to a schema format would bring us to **1ə2a3ə4ə**, or, given we know that the prefix is a prefix, it would give us **yə1a2ə3ə**. We could use this mechanical result to build the form but we would not know its distinctive features, the ones which carry the morphological meaning, like the gemination, and it would thus be not much more useful than a mechanical transliteration.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

The second feature of the data to be taken into consideration is that there are several competing transliterations methods.[10] Although I will always use here the Beta maṣāḥəft transliteration conventions, a parser should be able to provide the information and to allow requests to be sent in each of these alternative transliterations, not only a given one. This is important for the data entry and request but it needs to be noted that the transliteration itself depends on the understanding of the schema, not directly on the *fidal* string.[11] Therefore, to transliterate the two examples above, one additionally needs to know what letter of the Latin alphabet is used to transliterate a given letter in one of its seven orders in *fidal*. Transliteration is thus useful for a grammar that aims at beginners, who want to learn the language and are not yet familiar with the script as they learn it. The transliterations of the examples above would be *walada* and *yəwalləd* respectively, based on their schema.

What we are left with in terms of conversion, given knowledge of the root, is the following:

1) We cannot produce a correct transliteration from *fidal* directly without knowing the schema;
2) We can produce a correct transliteration from the schema;
3) We cannot always produce a correct schema from a *fidal* string;
4) We can produce a correct schema from a transliteration string given we recognize affixes;
5) We can produce *fidal* from a transliteration, knowing a few rules (e.g. ignore gemination);
6) We can produce *fidal* from a schema, and a few rules (e.g. ignore gemination).

Language does not need too much precision to work, it is the grammatical and didactic need to formalize the description of the language and its morphology that create the need for a transliteration. Additionally, the schema is coded in a way that it is not directly comparable to the transliteration string, because it carries information to match the consonants part of the root.

We are then back at the beginning, that is, the only things that matter are root and schema. However, in the real world what one has in front of his eyes will always be a *fidal* or transliterated form, and the same is true for a web ↗application or any other script presented with this task. The parser, which receives a string as input and wants to return that as part of its output, needs thus to be able to retrieve root and schema from this provided string. Once the root is hypothesised, some transformations become possible and allow further ones.

10 Bulakh 2016.

11 Hummel et al. 2018, 101.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

There is also an additional point concerning the affixes which provide the means to distinguish the different forms. These are given in transliteration for a grammatical reason, which is that only in this way can they be isolated and described. An affix *-a* is not going to increment the *fidal*, but only going to change the order of the last sign in the string. To be able to use these affixes in transcription in order to modify forms of a word we need to know what letter it is and which sign corresponds to the given order. The way I have chosen to make these comparisons, transformations and conjugations is to map the information to a 'matrix', a minimal data model describing the word.[12] Let me take the word **ተወሰንክሙ** as example.

Table 6.1

ተ	**ወ**		**ሰ**	**ን**	**ክ**	**ሙ**
ta	1a	2	2a	3a		
					kə	*mu*
ta	*wa*	*s*	*sa*	*n*	*kə*	*mu*

Always holding the assumption that we know how to transfer a single letter to its equivalent Latin script and that we have already recognized schema and root, in the above table, even without breaking it at each letter of the transliteration but staying with the signs sequence in *fidal*, it is easy to notice that there are differences between each two consecutive lines, which have been put in a logical order to produce a transliteration from *fidal*. When I have recognized the pattern, a first distinction is generated between the first and the second line in that there is one more position for the geminated consonant in the schema. The schema is different from the affix in the third line in that it is a transliteration and not a schema. The transliteration is not a plain transfer of the consonants of the root in the schema, because the affix causes a change in the order of the third position of the schema.[13]

It can also be observed how the tokens in each cell are different. To get the relevant token I simply make the string into a sequence of characters for the *fidal*, but I need to parse with a ↗RegEx both the schema notation and the transliteration, possibly after performing some preliminary normalization on the string to resolve ambiguities, for instance adding a zero between geminated consonants. The RegEx parsing the schema notation will also have to take

12 The Gəʿəz Morphological Parser ↗application source code is available at <https://github.com/TraCES-Lexicon/lexicon/tree/master/geezParser>.

13 The form without the affix would be in the transliteration *-du*.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

into special consideration digits and the capital letters used to identify special letters like **W** (waw), **Y** (yod), **L** (any laryngeal), **S** (any sibilant), and **D** (any dental).

There are some features of each of the layers which can be mapped to a small data format containing

1) The distinction between affixes and root signs;
2) The order of the letter;
3) The position of the letter.

Optionally, some basic knowledge which is handy to have in this model directly for the processing, so that the code does not have to look it up again, is the following:

1) The actual realization in the input string (any of the rows in the table);
2) The transcription value of the letter;
3) The first order of that letter.

These last optional values can be obtained starting from the input string, by looking up in a file with modelled letters the actual part of the string we are interested in.[14] For this parser, I have structured each letter as in the example below:

```
<letter type="laryngeal">
    <transcription>h</transcription>
    <realizations>
        <realization>ሀ</realization>
        <realization>ሀ</realization>
        <realization>ሁ</realization>
        <realization>ሂ</realization>
        <realization>ሃ</realization>
        <realization>ሄ</realization>
        <realization>ህ</realization>
        <realization>ሆ</realization>
    </realizations>
</letter>
```

Where for a number of transcription systems, another mini-model is given like this:

14 For all the XML data used by the analyser, I use the ↗namespace 'http://fidal.parser' which is in no way a registered ↗URL, but just a conventional namespace used only locally and unofficial. The letters file can be viewed in XML at <https://raw.githubusercontent.com/TraCES-Lexicon/lexicon/master/geezParser/morpho/letters.xml> or as a ↗HTML page at <https://betamasaheft.eu/morpho/letters>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

```
<transcription type="BM">
    <vowel/>
    <vowel>a</vowel>
    <vowel>u</vowel>
    <vowel>i</vowel>
    <vowel>ā</vowel>
    <vowel>e</vowel>
    <vowel>ə</vowel>
    <vowel>o</vowel>
</transcription>
```

It is immediately evident that I have added to the model a ‘ghost’ order, which is used to represent those positions in the transcription which do not have a vowel. Giving the above six pieces of information, partially obtained from the analyses of the string and partially looking at the letters file, I can build this model for the *fidal* string in our example:

```
<chars xmlns="http://fidal.parser">
<prefix>
  <char>ተ</char>
  <firstOrder>ተ</firstOrder>
  <order>1</order>
  <transcription>t</transcription>
</prefix>
<syllab>
  <char>ወ</char>
  <firstOrder>ወ</firstOrder>
  <position>2</position>
  <order>1</order>
  <transcription>w</transcription>
</syllab>
<syllab>
  <char>ሰ</char>
  <firstOrder>ሰ</firstOrder>
  <position>3</position>
  <order>1</order>
  <transcription>s</transcription>
</syllab>
<syllab>
  <char>ን</char>
  <firstOrder>ነ</firstOrder>
```


ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

```
      <position>4</position>
      <order>6</order>
      <transcription>n</transcription>
   </syllab>
   <syllab>
      <char>ክ</char>
      <firstOrder>ከ</firstOrder>
      <position>5</position>
      <order>6</order>
      <transcription>k</transcription>
   </syllab>
   <syllab>
      <char>ሙ</char>
      <firstOrder>መ</firstOrder>
      <position>6</position>
      <order>2</order>
      <transcription>m</transcription>
   </syllab>
</chars>
```

Fortunately there is a definite number of possible occurring affixes to forms, so it is possible to test if a string starts with one of them and hypothesize its identification as a prefix. The positions give the numeric part of the schema information (which will have to deduct the number of prefix tokens) and the order allows for the correct vowel to be identified,[15] without relying on a transliteration. I could make from this model a hypothesis of the perfect of the I,1 form by simply picking the first, second, and third <syllab> elements child nodes <firstOrder>.

Given the same matrix of information for another schema, I can match the two and select the values for another form. Taking the previous example, if I want to build the corresponding form of this word in the schema **yə1a22ə3**, I will produce this model as well, taking into account the meaning carrying digit positions of the schema notation:

```
<chars xmlns="http://fidal.parser">
   <prefix>
      <order>6</order>
      <transcription>y</transcription>
```

15 The position of the 0 vowel in the letters model requires a bit of attention because, to retrieve the vowel at a given position index, which is based 1, the value needs to be incremented.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

```
  </prefix>
  <syllab>
    <position>1</position>
    <order>1</order>
  </syllab>
  <syllab>
    <position>2</position>
    <order>6</order>
  </syllab>
  <syllab>
    <position>2</position>
    <order>6</order>
  </syllab>
  <syllab>
    <position>3</position>
    <order>0</order>
  </syllab>
</chars>
```

I can then copy over for each <syllab> the <firstOrder> or <transcription> values from the <syllab> with the same position in my starting form.

```
<chars xmlns="http://fidal.parser">
  <prefix>
    <order>6</order>
    <transcription>y</transcription>
  </prefix>
  <syllab>
    <firstOrder>ወ</firstOrder>
    <position>1</position>
    <order>1</order>
    <transcription>w</transcription>
  </syllab>
  <syllab>
    <firstOrder>ሰ</firstOrder>
    <position>2</position>
    <order>6</order>
    <transcription>s</transcription>
  </syllab>
  <syllab>
    <firstOrder>ሰ</firstOrder>
    <position>2</position>
    <order>6</order>
```


ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

```
      <transcription>s</transcription>
    </syllab>
    <syllab>
      <firstOrder>ን</firstOrder>
      <position>3</position>
      <order>0</order>
      <transcription>n</transcription>
    </syllab>
  </chars>
```

Looking up the vowel to be used in the letters file for each order, I can already get from this a correct transliteration *yəwassən* and I can get the correct *fidal* form **ይወስን**, by looking up the same letters file for the realization of a letter at that order.

Secondly, I have enough information to join to a form the affixes and apply the rules which alter the schema in case the affix starts with a vowel or a consonant (in transliteration form), for example. The base form of the perfect III,2 of the lemma in our example would have been as follows:

```
<chars xmlns="http://fidal.parser">
  <prefix>
    <char>ተ</char>
    <firstOrder>ተ</firstOrder>
    <order>1</order>
    <transcription>t</transcription>
  </prefix>
  <syllab>
    <char>ወ</char>
    <firstOrder>ወ</firstOrder>
    <position>2</position>
    <order>1</order>
    <transcription>w</transcription>
  </syllab>
  <syllab>
    <char>ሰ</char>
    <firstOrder>ሰ</firstOrder>
    <position>3</position>
    <order>1</order>
    <transcription>s</transcription>
  </syllab>
  <syllab>
    <char>ነ</char>
    <firstOrder>ነ</firstOrder>
```


ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

```
      <position>4</position>
      <order>1</order>
      <transcription>n</transcription>
    </syllab>
  </chars>
```

When we add to it the affix *-u* for the third person masculine plural, we actually only need to modify the corresponding <order> in the last <syllab> to be able to derive ኑ as new value for <char>.

```
<syllab>
  <firstOrder>ነ</firstOrder>
  <position>4</position>
  <order>2</order>
  <transcription>n</transcription>
</syllab>
```

If we have to add *-kəmu* we will modify the order of the last <syllab> and add to the model the affix and we will have the same as the first code example in this section.

3 Hypotheses making process

Now that we have an interchange data format allowing us to use the three different kinds of information at our disposal, we can design a process which will try to reproduce the human behaviour in analysing a string to parse it and return the morphological analysis. As already said, we want to look up the input string as is in the TraCES ↗annotations. This already gives the full solution in most cases, but we want also to apply mechanical rules to independently validate these results, or even simply to get other possibilities.

We need to identify the form and its affixes, as well as hypothesize a schema and a root to be able to gather information from the schema and the affix. But let me follow the steps I would do without a computer to parse the word ዘኢወለደተኒ (*zaʾiwaladattanni*).

This form is annotated in Chron. Am. 95 (CAe 4275, ID: LIT4275ChronAmdS) and is made of the following parts:

1) Relative Pronoun Masculine Singular ዘ;
2) Negative Particle ኢ;
3) Verb Perfect Third Feminine Singular of ወለደ;
4) Pronominal Suffix First Communis Singular ኒ.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

First of all, I would discern proclitic particles so that they can be parsed independently. This is a preliminary step which attempts to isolate units of meaning. (1) and (2) in the list above are then parsed on their own. Looking up the presence of such proclitic particles from a list, each value *and* the original one are parsed.[16]

Removing the relative pronoun and the negative particle, the remaining string is then only **ወለደትኒ**. In this substring we look at the next most evident feature, the suffix. To identify this from the list of transcription values (Fig. 6.1), it is actually enough to produce a mechanical transliteration and match the ending to a list of affixes. This list of affixes, which is always an XML file, contains information on all the possible changes to the base root.[17]

```
<conjugations xmlns="http://fidal.parser">
  <type name="Perfect">
    <num type="Singular">
      <person type="Third" n="3">
        <gender type="Masculine"> [72 lines]
        <gender type="Feminine">
          <affixes>
            <affix>at</affix>
          </affixes>
          <pronouns>
            <num type="Singular">
              <person type="Third" n="3"> [11 lines]
              <person type="Second" n="2"> [11 lines]
              <person type="First" n="1">
                <gender type="Common">
                  <affixes>
                    <affix>attanni</affix>
                  </affixes>
                </gender>
              </person>
            </num>
            <num type="Plural">
              <person type="Third" n="3">
                <gender type="Masculine">
                  <affixes>
```

Fig. 6.1 The affix -*attanni* in the conjugation.xml file.

The transcription *waladattanni* with a few mechanical replacements for the limited cases of geminations occurring in the suffixes (nn, tt, kk) can be com-

16 <https://raw.githubusercontent.com/TraCES-Lexicon/lexicon/master/geezParser/morpho/proclitics.xml>.

17 <https://raw.githubusercontent.com/TraCES-Lexicon/lexicon/master/geezParser/morpho/conjugation.xml>. The data for these tables was provided by Magdalena Krzyżanowska and Vitagrazia Pisani.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

pared to each of the values in the affixes files simply with a ↗RegEx checking the end of the transliteration. This will, in most cases, return several candidate solutions which can be then narrowed down based on the pertinence to the identified schema (an affix of the perfect is not valid for a schema of the imperative) and the length of the string.[18]

We can now move on to extract a possible schema from our model of the string as described above. We will obtain a schema **1a2a3a4a5i**. This is now what we want to match to a list of patterns. This list is also encoded in XML and has the same structure as the affixes file.[19] Here, however, also the type of schema patterns is annotated to distinguish patterns which are possible for a given type of root.

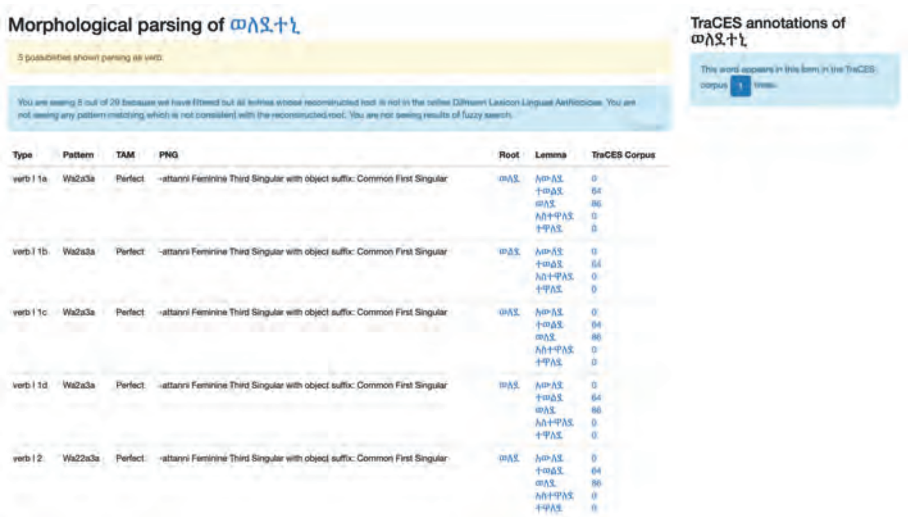
```
<pos name="verb">
  <group name="I">
    <type name="1a">
      <pattern name="Perfect">
        <formula>1a2a3a</formula>
        <formula type="w1">Wa2a3a</formula>
        <formula type="w2">1aWa3a</formula>
        <formula type="w2">1aW3a</formula>
        <formula type="w3">1a2aWa</formula>
        <formula type="y1">Ya23a</formula>
        <formula type="y2">1aYa3a</formula>
        <formula type="y3">1a2aYa</formula>
        <formula type="l1">La2a3a</formula>
        <formula type="l2">1aLa3a</formula>
        <formula type="l3">1a2La</formula>
        <solution>I,1a Perf. 3 m. sg.</solution>
      </pattern>
      ....
    </type>
  </group>
</pos>
```

18 A string ending in *kəmu* will match *kəmu* but also *u*. The latter is however impossible if the query string length minus the hypothetical affix does not equal the identified base schema length.

19 <https://raw.githubusercontent.com/TraCES-Lexicon/lexicon/master/geezParser/morpho/patterns.xml> based on tables provided by Vitagrazia Pisani and Magdalena Krzyżanowska to the TraCES project.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

To look into the patterns for three radicals, I can then omit the positions 4 and 5 from my schema hypothesis, and look for **1a2a3a**. And I find it easily, however it is not the correct one, as I need to match **Wa2a3a**. The schema hypotheses need to be multiplied to many alternatives (with W, Y, L, etc.) and then tested for their existence in the schema file. Let us evaluate only the two options just mentioned, although many more are produced by the script. I can make both hypothesis and test both of them against the list and both will result in a valid matching schema for my form, from which I can hypothesize a root. I need to decide which type of root this is, if a regular one or a *primae w*. For each of my schema hypotheses I attempt a reconstruction of the stem of the perfect I,1, which in both cases will return **ወለደ**. I can parse this and check if, at any given position, a *w* occurs. If it does, in the first position the hypothetical root is *primae w*, or *w1* in the short XML notation I used. I have then a regular schema with a hypothetical root *w1* and a *primae w* schema with a hypothetical root *primae w*. The second wins and, of the two hypotheses I made, I can carry on with the second, **Wa2a3a**. By the way, I also have gained a hypothetical root with a first level of conformance to the schema. I can then produce a tentative response (Fig. 6.2) using the information associated to the matched schema and that associated to matching affixes pertinent to that schema.



Type	Pattern	TAM	PNG	Root	Lemma	TraCES Corpus
verb I 1a	Wa2a3a	Perfect	-attanni Feminine Third Singular with object suffix: Common First Singular	ወለደ	ኣውለደ ተወልደ ወለደ አስተዋለደ ተዋለደ	0 64 86 0 0
verb I 1b	Wa2a3a	Perfect	-attanni Feminine Third Singular with object suffix: Common First Singular	ወልደ	ኣውለደ ተወልደ አስተዋለደ ተዋለደ	0 64 0 0
verb I 1c	Wa2a3a	Perfect	-attanni Feminine Third Singular with object suffix: Common First Singular	ወለደ	ኣውለደ ተወልደ ወለደ አስተዋለደ ተዋለደ	0 64 86 0 0
verb I 1d	Wa2a3a	Perfect	-attanni Feminine Third Singular with object suffix: Common First Singular	ወለደ	ኣውለደ ተወልደ ወለደ አስተዋለደ ተዋለደ	0 64 86 0 0
verb I 2	Wa22a3a	Perfect	-attanni Feminine Third Singular with object suffix: Common First Singular	ወለደ	ኣውለደ ተወልደ ወለደ አስተዋለደ ተዋለደ	0 64 86 0 0

Fig. 6.2 The result of morphological parsing of **ወለደተኒ**.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

4 Validation

In the previous example we had only two options, and we could have been fine with that. The current number of hypotheses for the schema, which need to be made to be sure to match the right one among the results, is larger, due to the variation described above. I need to hypothesize geminations and the presence or absence of the sixth order vowel, but I need also to combine these cases and, for example, take into consideration the prefixes, which are part of the paradigmatic schemas provided in the list. I will need to add another step to the analysis of the results and I can do that by validating them against an external authoritative resource like the Dillmann's online *Lexicon linguae aethiopicae*: if the hypothetical perfect stem is among the lemmas of Dillmann's lexicon, it has a higher probability to be a solution.

However, it is not true that all forms of the paradigm are always lexically attested, so, it could be that Dillmann's *Lexicon* has a relevant entry which is not in the form of the perfect I,1 and it might be there in another form (I,2 or II,1, etc.) in addition to the perfect third masculine singular, which Dillmann always gives. For each possible schema and root of a given type, I have to use the model described above to construct the other perfect forms and check all of them, not only the perfect I,1, and validate the result if any or more of these is actually there.[20]

There is a further problem, which is that of roots which are not represented in the model schemas, e.g. because they have a laryngeal *and* a *w*. Some of these will be retrieved, but the pattern of the root will not match that of the schema. It is also possible that the root will be wrongly hypothesized, invalidating the final results and returning no match. For these reasons, the mismatching results need to be kept, and optionally a fuzzy string search can be run, instead of an identity matching on the schema list, producing many more false results, but possibly the right one for a verb of a mixed pattern or a type not represented in the schema present in the tables or with more than three radicals.

5 Paradigms and conjugation

From a valid hypothesis for which we now know a root, we can apply the known schemas for that type of verb and the affixes structures and generate the forms of the paradigm (Fig. 6.3) and conjugation (Fig. 6.4) of each form.

20 There are many other issues in this respect which would make this chapter too long and I leave for a further stage of development of this parser.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Even those which are not attested. This gives some potential to discover new possible solutions or identify yet unattested forms.

fidal	translit	pattern	form	conjugation	links
አስተነገረ	'astangara	'asta12a3a	IV,1 Perfect	[illegible] with pronouns	[illegible]
ያስተናግር	yāstanaggər	yāsta1a22ə3	IV,1 Imperfect	[illegible] with pronouns	[illegible]
ያስተንግር	yāstangər	yāsta12ə3	IV,1 Subjunctive	[illegible] with pronouns	[illegible]
አስተንግር	'astangər	'asta12ə3	IV,1 Imperative	[illegible] with pronouns	[illegible]
አስተንጊሮ	'astangiro	'asta12i3o	IV,1 Gerundive	[illegible] with pronouns	[illegible]
አስተንግሮ	'astangəro	'asta12ə3o	IV,1 Infinitive	[illegible] with pronouns	[illegible]
አስተንግሮት	'astangərot	'asta12ə3ot	IV,1 Infinitive	[illegible] with pronouns	[illegible]

Fig. 6.3 Paradigm of **ነገረ** IV,1a.

number	person	gender	form
Singular	Third	Masculine	ነገረ
Singular	Third	Feminine	ነገረት
Singular	Second	Masculine	ነገርከ
Singular	Second	Feminine	ነገርኪ
Singular	First	Common	ነገርኩ
Plural	Third	Masculine	ነገሩ
Plural	Third	Feminine	ነገራ
Plural	Second	Masculine	ነገርክሙ
Plural	Second	Feminine	ነገርክን
Plural	First	Common	ነገርነ

Fig. 6.4 Conjugation of **ነገረ** perfect I,1a.

6 Integration

Knowing the root and the Dillmann's online *Lexicon linguae aethiopicae* identifier, one can expand further the possibilities to match relevant ↗annotations in the TraCES corpus. While the first matching of annotation is by necessity string based, given the identifier a search can be made to return annotations of that lemma, in any of its forms.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

But perhaps even more interesting is the possibility to integrate the analysis directly on the texts with a tool like Alpheios,[21] to provide access, for any word on a page, to three resources: the TraCES corpus, the Dillmann's *Lexicon* and the Gəᶜəz Morphological Parser. Alpheios can either be embedded by a project as a ↗JavaScript module directly in the specific ↗HTML pages published by that project or installed as ↗browser add-on and used with any HTML page on the Web.

The format of the output can be returned in a way that it is consumable by Alpheios and, in that case, both the options of solutions in TraCES and those offered by the parser can be provided along one another. Below is an example of response by the morphological parser constructed according to the Alpheios schema.[22] Fig. 6.5 shows the results for a query to the Morphological Parser as displayed via Alpheios.


```
<words>
  <phrase>ተወለድክሙ</phrase>
  <word>
    <form>ተወለድክሙ</form>
    <entry>
      <dict>
        <hdwd>ወለደ</hdwd>
        <src>
          https://betamasaheft.eu/Dillmann/lemma/
          L4b1fd3ec8adc45bcb9a98a3a9f12c44a
        </src>
      </dict>
      <infl>
        <term>
          <stem>taWa22a3a</stem>
          <suff>kəmu</suff>
        </term>
        <pofs>verb</pofs>
        <note>III</note>
        <note>2</note>
        <mood>perfect</mood>
        <gend>masculine</gend>
```


21 <https://alpheios.net/>. A test page with two texts, one annotated in TraCES and the other not, is available in the project website at <https://pietroliuzzo.github.io/DHEth/morphoparserTest.html>. Any web page with Gəᶜəz text can be used in conjunction with the Alpheios plug-in.

22 <https://raw.githubusercontent.com/alpheios-project/xml_ctl_files/master/schemas/trunk/lexicon.xsd>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

```
                <num>plural</num>
                <pers>second</pers>
            </infl>
        </entry>
    </word>
</words>
```

Fig. 6.5 Results of a query made by Alpheios set to use Gəʿəz as language.

This has the further advantage of making available in one way to a user a series of tools to help with the reading of several languages, and, for example, the Dillmann's *Lexicon* itself as a resource would benefit from the reading support for most of the languages there used (Latin, Gəʿəz, Arabic, and Greek, among those currently covered by Alpheios).

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Chapter 7. Linked Data

> and the maitre gave the signal and let glasses of white wine of the Moselle be poured, and my time came because, as I saw, they had forgotten to pour the wine to the emperor himself, I took a bottle with a serviette and I didn't even know what was on my mind, I got closer to the emperor, knelt on one knee like a church attendant, I bowed down, but when I got up, everyone was looking at me, and on my forehead, or rather in my forehead, the emperor printed a cross on me, he blessed me like that with his finger, and I poured
>
> (B. Hrabal, *I Served The King of England*, 1953)

Let us imagine that the passage quoted in the epigraph was alone in a web page with its citation. How could any crawler find out that the emperor is Ḫāyla Śəllāse? Or how would a human reader, who does not have access to the rest of the text, know? If the machine were able to do some very clever work on the text itself, it could, maybe, figure out an hypothesis of an English 'emperor' based on similarity with the title. Maybe. If we want a piece of information to be machine-retrievable, we have to tell the machines what that is, and possibly in an unambiguous way. In the passage, the relation between the citation and the quote from the text would have to be clarified and structured and this can be done by structuring the information, for example in ↗XML. Additionally, one can identify this distinction and identify the relation which the quote entertains with the emperor named in it, as well as the emperor himself, by pointing to concepts via links which the machine can see and use to figure out that they are different things. The machine could know that our page is not only a page on the Web, but that it is a page which contains a quote (and this is a specified concept) and that the quote is associated with a named relation to a citation (another defined concept) which says it is taken from a book authored by Bohumil Hrabal. Eventually it might know that this is actually a translation in English of the original Czech work. And eventually it might know that the word 'emperor' in the quote is actually a character in the novel and this character is Ḫāyla Śəllāse. We need, as always, to spell things out if we want any software to know them.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

If we do this work of identification and unpacking of the information for the machine, using the Resource Description Framework (↗RDF) with Unique Resource Identifiers (↗URIs) for each bit of information, using shared vocabularies for all values, the text will be *in* the Web, and not any more just *on* the Web,[1] as a linked resource, pointing to others and (possibly) pointed to by others. In this chapter I will describe how we can do this and which steps Beta maṣāḥǝft has taken for its data to be not only available and structured but also open and linked.[2] Although Linked Open Data (↗LOD) does not have the power of a blessing by the emperor (at least in the life of Dítě, the protagonist of the novel by Bohumil Hrabal), it is a very useful thing, and I will try to show that.

I will start from a few core concepts which need a brief introduction, to then move on to provide more precise details on the aims of ↗RDF and ↗LOD. I will then describe the way our data architecture is done to conclude the chapter with some examples.[3]

1 Core concepts

In a brief theoretical presentation, which I use when teaching ↗XML, I explain the core concepts of hierarchy and the basic rules of the language following the book *Disegnare un albero* ('How to draw a tree') by the designer Bruno Munari.[4] The parallel between the XML tree and the real world trees, or their drawings, is a very powerful metaphor, and a productive one,[5] which helps to understand the relationship between the encoded data and the ↗schema as well as the difference between a tree and a graph representation, not to speak of the fact that it allows one to understand how different can two trees be, even when they follow the same schema. If a text encoded in XML is a tree, a graph (and especially a linked data graph) might be thought of as a tree seen from the top, as part of a forest, where the branches touch and intersect with each other.[6] If a tree, real or in XML, has its beauty, so does the forest seen from below or from above, and indeed it is just a matter of focus: one could argue the same for a single square centimetre of cork or for a leaf.

1 woddiscovery 2010.

2 The online *Lexicon linguae aethiopicae* data is not yet serialized also as ↗RDF.

3 Some of the contents of this chapter have been presented in Tokyo at the ↗TEI conference in September 2018.

4 Munari 1978.

5 See also Ch. 3.

6 Similarly, Moretti 2013, 18 proposes bush as a metaphor for literary development. See Ch. 3.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

What matters here is that a tree and a graph are not necessarily alternative representations, they serve different purposes and naturally coexist as it is made clear in the following passage by Oldman, Doerr, and Gradmann.[7]

> The same principles and mindset established in more discrete digital research activities should be applied to large repositories of Linked Data, and we should not be distracted by quantity. This requires the removal of a 'two cultures' history that implies that memory institution database systems have less value than, for example, crafted TEI type representations. Linked Data resources become richer the more they integrate and can provide independent or complementary contexts. They should not be seen as being in opposition or competing.[8]

How to translate this idea in implementation to be able to benefit from the ↗TEI as well as the ↗LOD is a further problem and requires that we determine how to contextualize and enrich by creating links which break hierarchies.[9] As in a real forest, not all plants can adapt to all climates and conditions, nor can they grow anywhere in their climate, so a project needs to find its place and interact properly in its ecosystem.

For a ↗TEI-based project like Beta maṣāḥəft producing ↗RDF to be served as ↗LOD means, first of all, to be able to produce and maintain the triples to be stored and knowing what can be usefully stored there to begin with. Before moving to what sort of information is in the RDF and for which aims, I will give some more general implementation details.

1.1 RDF and SPARQL

It is necessary to start from few short but important definitions. ↗RDF is a model native to the Web made of triples.[10] ↗Triples are made of a subject, a predicate, and an object.[11] ↗SPARQL is a language used to query the data stored in this model as triples.[12]

A series of simple statements in this format allows for a non-hierarchical or poly-hierarchical structure of data, where each ↗triple can share with another

7 See also Oldman et al. 2016, 255–258 discusses also the relation between XML, ↗RDF, ↗LOD, and ↗OWL, with a special focus on ↗RDFS and OWL.

8 Oldman et al. 2016, 260.

9 This is also something which has been done in different ways by several projects, and the usefulness of this approach was already underlined by Ciula et al. 2008.

10 <https://www.w3.org/standards/techs/rdfa#w3c_all>.

11 Oldman et al. 2016, 252–253.

12 Some examples in this book were already given using this language in Ch. 1, p. 32 and Ch. 3, p. 83.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

one any of the three parts of the statement and statements can be made about a triple as a whole.

It is then clear that it is nothing which has directly to do with ↗XML or ↗TEI, although asking what is the difference between the two is a legitimate question in a world of extreme and sometimes confusing acronyms of all sorts.

Conceptually, triples in ↗RDF are the standard model for data interchange on the Web.[13] RDF data can be serialized as ↗turtle, XML, ↗json-LD, and so on. The core feature of this model is that it is minimal being all built of statements made of subject, predicate, and object. Additionally, all three parts of the ↗triple can be expressed as ↗URIs, which is the important part to produce Linked Data. These triples build graphs whose nodes are the subjects and objects of each statement and predicates can be seen as the edges between the nodes.

Because it is just a matter of formalization of the data, it follows that, to produce ↗RDF-XML, that is, ↗RDF serialized in XML, from ↗TEI data in XML, it might just be the matter of a ↗XSLT script, which is the most obvious way to transform XML.[14] To produce the RDF data you need to know how you want your graph to be like, to define the structure of your graph and thus the architecture of the knowledge you are storing in the data.

The definition of this architecture of the knowledge is done also with the help of ↗RDFS and ↗OWL, which means producing a web ontology which defines what subject is linked to what object and how.[15] Once the ontology is there, then the ↗XSLT transformation can be done, knowing what the resulting graph should be like and thus which information stored in the ↗TEI should be transformed to which ↗triple in the ↗RDF representation.[16] The result of the transformation of a TEI file will be a set of triples in RDF which fit into their ontology and represent some of the information which can be used as Linked Data.

13 <https://www.w3.org/RDF/>.

14 If one wishes to avoid ↗XSLT, there is also XTriples, <http://xtriples.spatialhumanities.de/documentation.html>, a configurable service developed by Torsten Schrade at the Academy of Sciences and Literature of Mainz, which, given a simple configuration, can produce triples in several formats from any XML. This is also an ↗eXist-db package which can be deployed locally, <https://github.com/spatialhumanities/xtriples>.

15 It is entirely out of the scope of this book to delve into what this involves, but the Web is very generous in instructive materials and formal specifications on these, which will certainly serve better the interested readers.

16 This is possible, and quite practical, also starting from examples and models.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

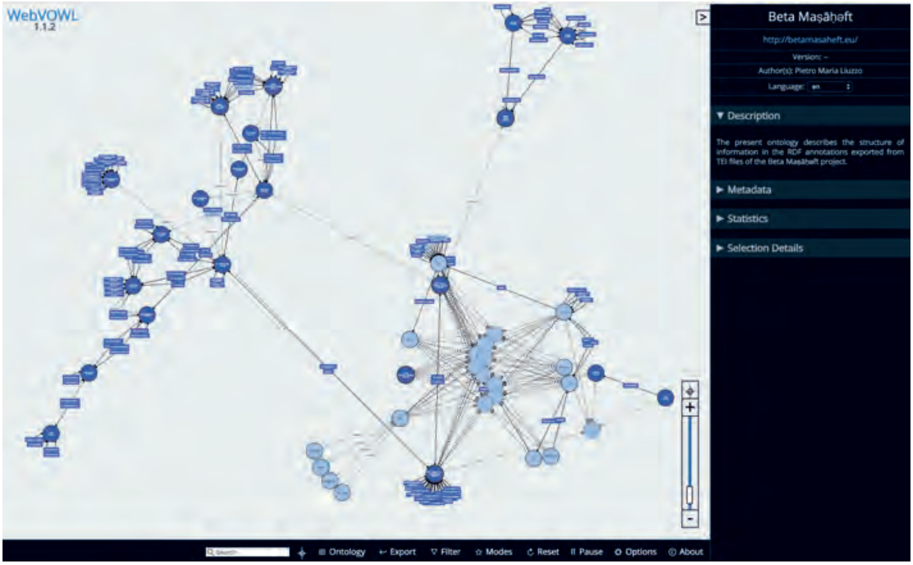


Fig. 7.1 WebVOWL visualization of the Beta maṣāḥəft Ontology. See <https://betamasaheft.eu/Guidelines/?id=ontologyView>.

This transformation is done in Beta maṣāḥəft at the time of storing a XML resource in the database.[17] Beta maṣāḥəft uses ↗GitHub for the versioning of the data and contributors push their changes to the shared repositories. In GitHub a ↗webhook fires a ↗POST request to the Beta maṣāḥəft production server,[18] where a script acts upon the content of the payload sent from GitHub which contains information about who has changed what and where exactly. At this stage the script stores the new or edited XML file and additionally transforms it in a series of ↗RDF triples which are also stored. While the XML resource is stored and indexed as such, in our ↗eXist-db instance, the ↗RDF-XML triples are stored in a parallel collection and indexed using the eXist-sparql module by Leif-Jöran Olsson.[19] The ↗TEI XML and RDF-XML are thus kept in sync, with the RDF being a derivative read-only version

17 The ↗XSLT is available here <https://github.com/BetaMasaheft/RDF/blob/master/transformations/data2rdf.xsl>.
18 The code for this, available in the repository of the Beta maṣāḥəft ↗application, is a modified version of the fantastic script <https://gist.github.com/wsalesky/bf26507f-f593f0c99a35> by Winona Salesky used also by Syriaca.org. This is now available also as a library module, <https://github.com/eXist-db/github-xq>.
19 <https://github.com/ljo/exist-sparql> available also via the public repository from ↗eXist-db dashboard.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

(i.e. the triples are never modified creating the need to update the TEI consequently), and we have two parallel sets of indexes which can be queried.[20]

In our ↗eXist-db instance we have, for example, a ↗TEI fragment like the following:

```
<title xml:id="t1">Periplus of the Erythraean Sea</title>
```

At the same time, a ↗triple of this kind in ↗RDF-XML is stored alongside.[21]

```
<rdf:Description rdf:about="https://betamasaheft.eu/LIT2170Peripl">
    <dc:title>Periplus of the Erythraean Sea</dc:title>
</rdf:Description>
```

This same statement could be equally expressed in ↗turtle as in the following example.

```
1 <https://betamasaheft.eu/LIT2170Peripl>
2 <http://purl.org/dc/elements/1.1/title>
3 "Periplus of the Erythraean Sea" .
```

In this case the subject is the ↗URI of the work (Line 1), the predicate is a ↗Dublin Core property (Line 2) called 'title' and the object in the ↗triple is the literal string in Line 3.[22] The equation between <title> in ↗TEI and <http://purl.org/dc/elements/1.1/title> is done in the ↗XSLT transformation, and does not have to be considered canonical. It is simply the way we have mapped it so far.[23]

20 Documentation of the ↗RDF is available at <https://betamasaheft.eu/lod.html>. The ontology was written in ↗OWL using Protégé (<https://protege.stanford.edu/>) and ↗Atom (Fig. 7.1). The visualization available at <https://betamasaheft.eu/Guidelines/?id=ontologyView> is made with WebVOWL (<http://www.visualdataweb.de/webvowl/>).

21 I am omitting the root <rdf:RDF> node with namespaces. dc: here refers to the ↗namespace for <http://purl.org/dc/elements/1.1/>.

22 It could have been a ↗URI and the URI would have had a <rdf:label> property with the literal value.

23 A special interest group of the ↗TEI consortium has been active for some time and has seen recently a revival, see Eide and Ore 2006; Ore and Eide 2009; Eide 2014; Ciotti and Tomasi 2016; and Chiarcos and Gracia forthcoming. Especially for inscriptions

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

As we shall see below in this chapter, other types of ↗LOD are also produced by Beta maṣāḥəft, but they are not materialized and indexed anywhere, for example the ↗IIIF presentation ↗API and the ↗DTS APIs,[24] both ↗json-LD which are produced on the fly with a call to a specific module following the ↗RESTXQ XQuery protocol by Adam Retter.[25] These scripts serving triples on demand query according to convenience either the index created on the stored ↗RDF or one of the indexes created on the source ↗XML data.

1.2 Linked Open Data

Having ↗RDF with URIs does not mean automatically having ↗LOD.[26] The ↗TEI-XML available online is already open data, is structured, and has a CC-BY-SA licence on it. Adding more explicit links into it makes it also ‘linked’. Also the ↗HTML main view in the website of Beta maṣāḥəft contains ↗RDFa and is accessible online and structured as a further way of exposing and sharing LOD.[27] This is to say that the accessibility of the data as RDF is an additional possibility and can be provided in several ways.

Beta maṣāḥəft strives to make the ties which keep its data in the Web in as many ways as possible because also the way in which the access to the ↗LOD is provided depends on the users. This means that there are several representations and, because they are there, it is important to clarify in which way they relate to one another and what the available options are. In fact, on a first level, this is what *content negotiation* is about. Content negotiation is a core part of implementing LOD, and it involves letting the client specify which format of information it requires (the ↗HTML view or the ↗RDF for example) and returning the desired response starting from one ↗URI.[28] The following is a ↗schema from the Cool URIs document which explains, much

(Ch. 2, p. 67) some work has started also from the ↗CIDOC-CRM group, see Felicetti and Murano 2017 and Felicetti and Murano 2016.

24 <https://iiif.io>. See also Ch. 1, p. 16. The ↗namespace for ↗DTS is <http://w3id.org/dts>. <https://betamasaheft.eu/api/dts> is the starting endpoint for Beta maṣāḥəft. For an example use of this, see Ch. 5 and my presentation at the Texts and APIs Workshop held in Hamburg 15–16 July 2019, <https://distributed-text-services.github.io/workshops/events/2019-hamburg/>.

25 Retter 2012.

26 See <https://5stardata.info/en/>, a proposition by Sir Tim Berners-Lee. See Van Zundert 2016, 96–99 on the hypertext and its theoretical origins in the ideas of Vannevar Bush and Theodor Holm Nelson.

27 See also other examples in Chiarcos and Gracia forthcoming.

28 In Beta maṣāḥəft this is achieved in the `controller.xql` module of the standard packaged ↗application.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

better than I could try to do here, why this is important and how it can be achieved.

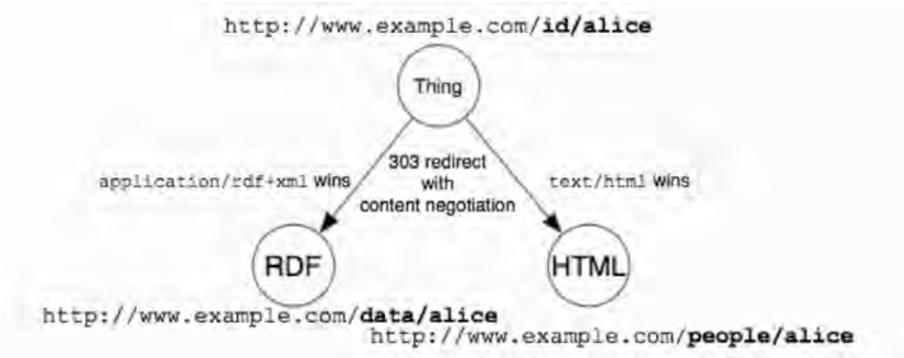


Fig. 7.2 Content negotiation schema from Sauermann and Cyganiak 2008.

To give a concrete example, in Beta maṣāḥəft this means that for an ID <https://betamasaheft.eu/BNFet32>, which is guaranteed by the project and its funding institution to be persistent, the ↗HTML representation and human readable resource is available at <https://betamasaheft.eu/manuscript/BNFet32/main> and this is where a browser will point when requested the above ↗URI, unless the ↗RDF location <https://betamasaheft.eu/rdf/manuscripts/BNFet32.rdf> is specified or the request for the ID is explicitly asking for application/rdf+xml in the request header Accept, in which case a graph centred at the requested URI will be produced.

This is not the same set of ↗RDF triples as in the file at the RDF location, because <https://betamasaheft.eu/rdf/manuscripts/BNFet32.rdf> is a representation of all the information in the XML file which is present at the ↗HTML page, directly derived from the same XML. The RDF graph centred at <https://betamasaheft.eu/BNFet32> does not contain the triples specific to parts and instead contains statements made in other XML files which are indexed as RDF in ↗eXist-db as described above.[29]

While this step of content negotiation ties the different representations to the ↗URI, also these need to be linked among them explicitly, so that there is always a way to navigate from one to the other. The ↗HTML has a <link rel="alternate" type="application/rdf+xml" title="RDF Representation" href="https://betamasaheft.eu/rdf/works/LIT2170Peripl.rdf"> pointing to the corresponding transformation to ↗RDF-XML. The ↗RDF file resulting from the transformation, on

29 See Section 4.3 of Sauermann and Cyganiak 2008.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

the other side, has a property foaf:homepage with the link to the main HTML page and a property dcterms:source pointing to the source XML.

Because also a ↗TEI-XML with links can be produced, in case the request header Accept specifies the response type application/tei+xml, then ↗eXist-db will redirect to the TEI-XML with links also produced by another ↗XSLT transformation as in Fig. 7.3.

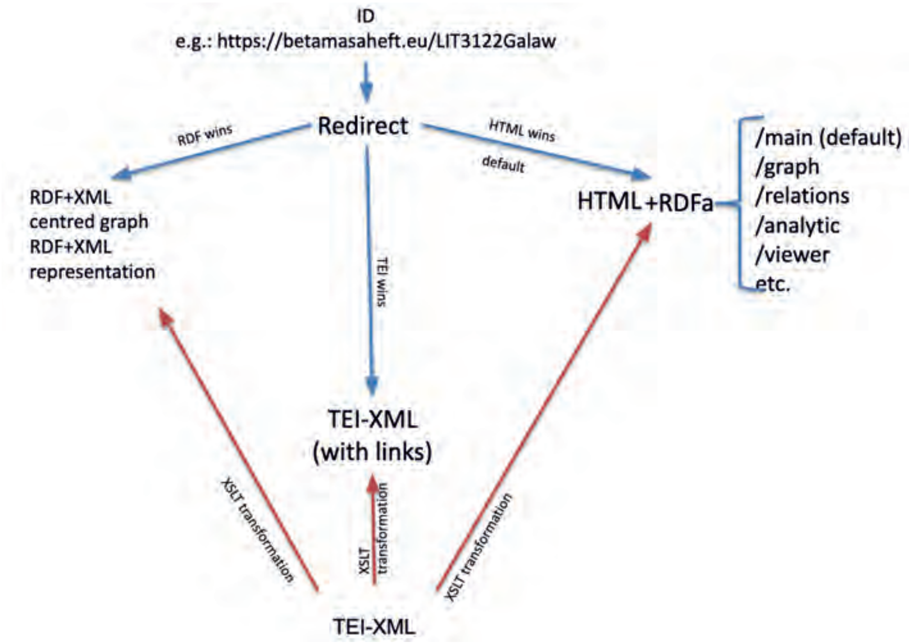


Fig. 7.3 XML workflow and content negotiation.

As briefly mentioned above, the ↗RDF stored in ↗eXist-db is not the only type of ↗LOD produced by the project. There are, according to the different types of resources (textual units, places, written artefacts, etc.), different representations available.

Using a specific vocabulary, ↗VoID,[30] some additional nodes are made openly available (in this case in ↗turtle) which can lead to the alternative representations as outlined in Fig. 7.4.[31]

30 See also Meadows and Gruber forthcoming for the use of this vocabulary to share datasets.

31 e.g. <https://betamasaheft.eu/api/void/BNFet32>. Each ↗HTML page contains a visible link to these nodes.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

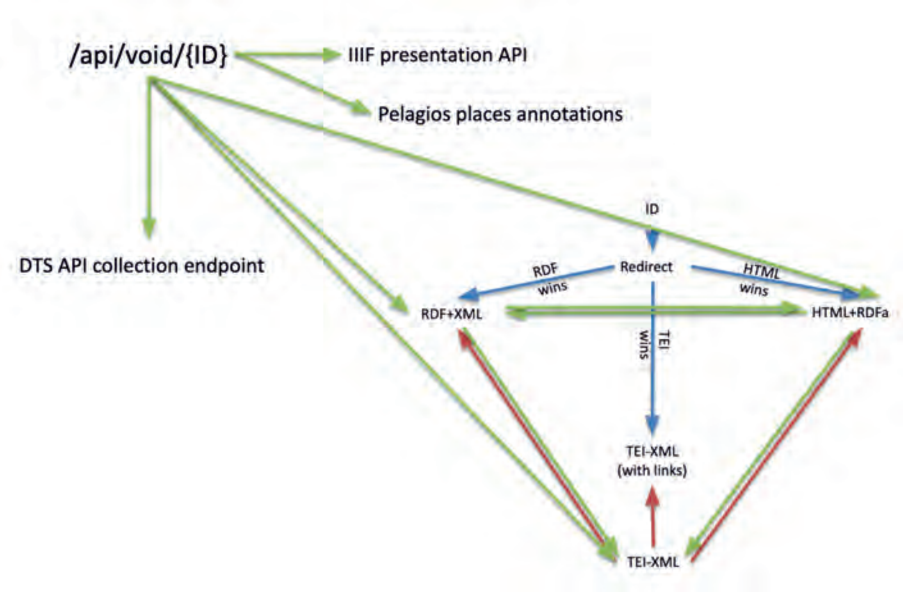


Fig. 7.4 Interlinking data with ↗VoID.

Users can then gain knowledge about the existence or not of a ↗DTS representation for a given resource even if they did not start their navigation from the DTS endpoint. Given this general implementation infrastructure, which is more easily done than explained in writing, the so-called technical barriers are removed and we can look at what can be done which is useful with this kind of data.

2 Aims of RDF and LOD

Let me start from another passage from the article quoted above, which summarizes very well what the usefulness of ↗LOD is.

> In the semantics of the structural elements, the relationships which can be expressed explicitly within Linked Data become critical to the ↗application, much more than the world describes. Even the smallest piece of information, placed in context, may provide the missing link needed to unlock a chain of relationships in data sourced from diverse locations. The discovery of potentially related facts through the use of a particular pattern of context allows us to debate similarities and differ-

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

ences which we can reuse to further infer and assert various arguments and apply other evidence.[32]

One shared link is enough to do a lot, if it is meaningful and used well. Actually, those few cases in which one specific external entity is linked are more useful than the mass of triples which can be serialized in a transformation and those are the ones which make the whole game worthwhile and enrich the data with unexpected possibilities.

There is no need for ↗RDF or ↗LOD to get fancy drawings, an ↗XQuery or ↗XSLT can serialize data in the format required by ↗JavaScript libraries to produce nice graphical representations or charts starting from the XML.[33] Although often, at least in presentations, networks, and clouds are shown to illustrate 'what can be done', let me stress once more here that this is not the aim, and the chosen visualization does not demonstrate the usefulness of LOD.[34]

There are instead situations in which the ↗RDF data model allows one to represent information in the data more correctly (without undue multiplications of entities for example), and, on the other side, to visualize dynamically the actual state of a complex situation, like, for example, a text tradition with multiple versions, by taking into consideration the state of the data, and not forcing the information into structures.

In other cases we can use ↗LOD, in ↗RDF or not, to retrieve more context and related items for a given piece of information, which allows us to offer verification, for example, or to explicitly represent a complex set of relations, like in the example of the place concept Ethiopia in Ch. 4.[35]

In the following sections I will provide a few examples of this, starting from a case of interaction between freely available resources and moving towards more complex ways of using this model to apply advanced methodologies in research.

2.1 LOD in practice

Publishing data with an open licence allows an enriching exchange, by virtue of which resources can be correctly and precisely related and one can have the correct page inside a manuscript load in a viewer directly from the institu-

32 Oldman et al. 2016, 259–260.

33 See for example Ch. 3.

34 See Ch. 4, Fig. 4.1 for an example of a useful network of place ↗URIs offered by Pelagios, and Ch. 3, Fig. 3.1 for a self-made example of an unreadable cloud. A very nice and navigable network visualization is the one offered for Kindred Britain, <http://kindred.stanford.edu/#>.

35 See p. 109.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

tion which has the responsibility of hosting that artefact.[36] Providing ↗LOD for a memory institution means that they can pursue their aims of preservation and access to the information in a way which allows us, when we encode our data and do our part to produce correct scientific data about a text in its edition, to complement their work. In the same way in which our ↗application fetches, via LOD representations, the links to the images in order to display them, the repository of that given manuscript could fetch our edition because it is also exposed as LOD. Let me give an example here, and namely that of the set of chapters from the *Periplus of the Erythrean Sea* (CAe 2170, ID: LIT2170Peripl) which I have encoded in Beta maṣāḥəft to mark up the toponyms discussed in Ch. 4.

The Universitätsbibliothek Heidelberg made available under CC-BY-SA images of the manuscript Heidelberg, Universitätsbibliothek Heidelberg, Cod. Pal. graec. 398 where the *Periplus* is preserved, using the ↗IIIF presentation ↗API. Together with a link to the page where Universitätsbibliothek Heidelberg has a presentation of this manuscript, we have entered in our record in Beta maṣāḥəft the link to the ↗manifest.

```
<witness type="external"
            xml:id="P"
            corresp="Cod. Pal. graec. 398"
            facs="http://digi.ub.uni-heidelberg.de/diglit/cpgraec398/0569">
   <ptr target="https://digi.ub.uni-heidelberg.de/diglit/iiif/cpgraec398/manifest.json"/>
   Sammelhandschrift — Konstantinopel, letztes Viertel 9. Jh.
</witness>
```

By doing so, we can use the ↗manifest ↗URI and pass it on to the ↗Mirador viewer to view the images of the manuscript. This is possible because Mirador, which is freely available, knows and works with the ↗IIIF presentation ↗API, and the Universitätsbibliothek Heidelberg publishes information using this API openly.

It is of course possible, as some institutions do, to exploit this only for their benefit and use their ↗IIIF presentation, without making them available for others to use, although this misses a bit the whole point of interoperability of the framework.

However the whole manuscript is not relevant to the text we are publishing in Beta maṣāḥəft, only ff. 40v to 54v contain this text. If the ↗Mirador viewer was just showing to the user the first image of the manuscript, then one

36 See p. 16 for the way in which these ↗manifests are produced from the encoded catalogue descriptions in ↗TEI.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

would have to navigate again inside the contents of the manuscript displayed in Mirador thanks to the data in the ↗manifest published by the Universitätsbibliothek Heidelberg, and this would not be so convenient.

By adding to our <pb> element the exact links to the canvases relevant to our text, ↗Mirador can start displaying from the correct point, and, exploiting the same information in the ↗XQuery and ↗XSLT which produce the view, can also follow up the browsing of the text, so that, if Chapter 5 of the edition is on f. 41r, the viewer will focus on that page when loaded (see Fig. 7.5).

Fig. 7.5 *Periplus of the Erythrean Sea* (CAe 2170, ID: LIT2170Peripl) in the text view on Beta maṣāḥǝft using ↗IIIF images of the manuscript Cod. Pal. graec. 398 made available under CC-BY-SA by Universitätsbibliothek Heidelberg. See <https://betamasaheft.eu/works/LIT2170Peripl/text>.

The manuscript images are thus accessible where relevant and also the text needs to be so. This is among the aims of the Distributed Text Services (↗DTS) ↗API specifications. Given a third-party text browser or even an ↗annotation tool which understands DTS, which still does not exist, Universitätsbibliothek Heidelberg or any other user could use it to display the edition of the text side by side with the images also there, with minimal development time. Indeed a third or fourth interested party could use the ↗IIIF ↗manifest and DTS APIs data to produce yet another view of the core text and image resources using available viewers or building its own without duplicating the reference resources, or storing them somewhere else.

This is a reality in which the data curators, the repositories of the manuscripts, and the curators of the editions of the text serve the data they produce

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

and let the developers work on the software, so that everyone can use their favourite tools to interact with the same curated data to simply read, or to edit or mix it with other sources.

This works when the data is linked and is openly accessible in a structured format which follows shared standards both for its ground encoding (in our case for the text ↗TEI) and the presentation of this available data, namely via ↗IIIF and ↗DTS. In the same Fig. 7.5, further elements of connection can be seen, like the related items from ↗Pelagios and the bibliography,[37] called from the ↗Zotero public ↗API. This should be sufficient to persuade anyone, user or provider, of the benefits of open data which follows agreed shared standards for its presentation, but let me show some more use cases.

2.2 Querying LOD

The first priority in undertaking work to produce ↗LOD should be defining the aims of the representation. This is especially true for projects in the humanities, where the utility of the openness of the data is perhaps less obvious than it is for state agencies and memory institutions. For sure one reason could be all the unexpected research questions people may ask from this data when it is made available to them. On the other side, all the unforeseeable implementations which users might be able to put together are definitely a justification worth the effort.

But even where 'the digital' is not any more in question there are still many considerations to be made on how things are digitally published. Producing and working with ↗LOD is a way of publishing data online which allows to expand the possibilities. Van Zundert recently said, discussing the potential of digital scholarly editions to express the relations between the witnesses, that:

> The point here is that a list [of the witnesses of a text, ordered by date from a database] so represented is not a hypertext representation of the chronological 'linkedness' of the witnesses, it is a mere list of individuated metadata. This is different from the idea of hypertext that all information is expressed as machine negotiable nodes and links, so that an expressive network of knowledge is created. […] but if the result of that effort only allows user-level navigation of relational metadata represented as a graphical interface, then the digital scholarly edition is not an effective hypertext knowledge space.[38]

A digital edition, although with many links and navigation possibilities, is not an 'hypertext knowledge space'. Having machine negotiable nodes and

37 See Ch. 4 and Ch. 8.

38 Van Zundert 2016, 102.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

links helps along that path. This statement leads, in my understanding of it, quite clearly and directly to the need of producing and exposing as many forms as possible of ↗LOD which are meaningful and can allow resources online to form a 'hypertext knowledge space'.

One practical way to try and achieve this, for example, is the ability to do simple federated queries to selected endpoints.[39] If two or more databases expose some ↗RDF triples with links, they could also implement expansions of their queries which interrogate data made available by others in the same way, without changing the query. Having a common ontology and ↗annotation model would strongly facilitate any implementation because one would not have to go and check what the model looks like for each dataset.

This calls for a shared graph structure, so that the same query can be applied. For Beta maṣāḥəft we have started by looking at the triples produced by Syriaca.org.[40] Both projects implement an endpoint for ↗SPARQL queries and this means that, from the Beta maṣāḥəft endpoint, one could query the Syriaca.org ↗RDF as a service and complete information mixing it from both sources in one query.

```
SELECT DISTINCT ?variant
WHERE {
  bm:LOC5017Nubia lawd:hasName ?name ;
                          skos:exactMatch ?match .
  SERVICE <http://wwwb.library.vanderbilt.edu/api/sparql>
      {?match lawd:hasName ?name ;
                 rdfs:label ?label .}
   {?name lawd:variantForm ?variant .}
              UNION
   {?name  lawd:primaryForm ?variant .}
             UNION
    {BIND(?label as ?variant)}    }
```

39 This is done, for example, using specific technologies and workflows by EAGLE (<https://www.eagle-network.eu/>) and the Roman Open Data (<http://www.romanopendata.eu/#!/main>). See also Calvanese et al. 2016.

40 <https://github.com/srophe/srophe-data-rdf>. See Michelson 2016a and Michelson 2016b.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Because both projects use the ↗LAWD Ontology,[41] I can expect the ↗property <http://lawd.info/ontology/hasName> (Lines 3 and 6) to be there and used in the same way (Lines 8–10) meaning the same thing in both datasets, so I can join the results to get all possible variant names, including those used as label. This is just a basic working example but shows a case of real world interoperability, which can be exploited to gather related contents or to complete information, like in the following example, also very basic, where we query, for all statements which are related with skos:exactMatch and whose ↗URI as a string matches that of a place in Syriaca.org, to retrieve from the Syriaca.org SPARQL endpoint the description provided there (Lines 5–6).[42]

```
1 SELECT *
2 WHERE {
3       ?bm skos:exactMatch ?match .
4        FILTER(STRSTARTS(STR(?match), 'http://syriaca.org/place/'))
5        SERVICE <http://wwwb.library.vanderbilt.edu/api/sparql>
6                { ?match dcterms:description ?description}  }
```

Table 7.1 **Example results of the previous query: fetching descriptions of a place concept based on matching identifiers.**

bm	match	description
https://betamasaheft.eu/LOC5611Saydna	http://syriaca.org/place/617	A monastery outside Ṣaydnāyā, north of Damascus.
https://betamasaheft.eu/LOC5017Nubia	http://syriaca.org/place/630	A region of north-eastern Africa, south of Egypt.

In Ch. 2 I have used the data produced as ↗RDF with links to the EAGLE Vocabularies to link locally and compare epigraphical data from different traditions for a comparative aim. This could be achieved also simply from the ↗TEI without links, of course, given the data is open, accessible and has been advertised to interested parties. A second use case is that which I have dis-

41 <https://github.com/lawdi/LAWD> is the core interoperability vocabulary for Linked Ancient World Data. It includes ↗CIDOC-CRM and is used by ↗Pelagios and ↗SNAP:DRGN.

42 Because there are also matches to Wikidata, we could have also fetched information from there, like SERVICE <https://query.wikidata.org/> { ?match <http://www.wikidata.org/prop/direct/P625> ?coordinates}, but it was not possible at the time of writing. See also the SPARQL query examples library at <https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/queries/examples>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

cussed in Ch. 4, where the publication of ↗LOD made it possible to join the ↗Pelagios effort and gather data and visualizations of contextualizing information as well as of related items from other domains.

However, in a ↗LOD world I would like to be able to do even more. I would like to visualize distributions of the same type of inscription in time and space on a map (which I could do in Peripleo); I would want to visualize a force graph which is centred on a person and shows the inscriptions he is named in regardless of language and provenance; I would like to see continuity in the use of specific patterns and, thus, I would like to be able to easily get to ↗TEI encoded texts from the ↗RDF. I would like to be able to see a graph which shows me the modifications undergone by a given artefact, for example, that it was a stele in the forum, then a part of a fountain some hundreds years later, then part of a collapse in an excavation, then stored in museum X, and subsequently moved to lapidarium Y. I would love if the available LOD could join for me the attestations marked up in the text and available via an epigraphic database in TEI, with the second part of the information, probably more easily available via a museum database.

To achieve this is not enough to be able to produce ↗LOD. It is indispensable to use shared vocabularies and communities of practice, which for Beta maṣāḥəft are those of the ancient world data, where ↗LAWD is the core interoperability vocabulary.[43]

This is even more important where core concepts (like that of 'text' or 'author') are fluid, like in the Ethiopian tradition and, in general, in the literature of the Christian Orient, where also the language is not defining, due to the frequent and productive processes of translation from one to another language.[44]

Hoping to have argued enough for the necessary standards and sharing culture required by ↗LOD to be fruitful, I will now move on to give some concrete examples of what we use our ↗RDF data for and how LOD is benefiting Beta maṣāḥəft.

2.3 Contextualizing place names with Pelagios

The only application of ↗LOD that Beta maṣāḥəft can display is currently the connection to the ↗Pelagios Commons network. Because it takes more than one player to play this game and LOD cannot be exploited alone.

The ↗Pelagios Network, in all the Pelagios projects iterations, has been a pioneering effort in leading digital projects related to the ancient world in the

43 See Isaksen et al. 2014 for a short and clear description of the meaning of Linked Open Data for the humanities.

44 See Ch. 3.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

new world of ↗LOD.[45] It used a very simple and low-barrier methodology to allow anyone to easily join the network.

The idea is that there are gazetteers of place concepts and there are ↗annotations of resources which refer to those gazetteers. All a project needs to do to link its annotated resources to ↗Pelagios is to publish those annotations as ↗LOD. The annotations have very basic requirements but can be expanded as much as wanted. The following is an example of an annotation made on the *Confessio Claudii* (CAe 1252, ID: LIT1252Confes).

```
<placeName ref="pleiades:50004">ህንድ ።</placeName>
```

This is transformed by a ↗XQuery into an ↗annotation,[46] serialized as ↗turtle.

```
<https://betamasaheft.eu/LIT1252Confes/place/annotation/9>
    a oa:Annotation ;
    oa:hasTarget <https://betamasaheft.eu/LIT1252Confes> ;
    oa:hasBody <https://pleiades.stoa.org/places/50004> ;
    oa:annotatedAt "2018-09-24T11:18:10.686+02:00"^^xsd:date .
```

The ↗annotation itself receives a ↗URI, and it is defined as being part of the ↗class Annotation in the Open Annotation vocabulary.[47] Three other statements, part of the Open Annotation Data Model are then used to describe the annotation and say that the annotation has a target, the work annotated, a body which is the georesolution of the toponym in the text, and the date at which this statement was produced. The following graphic (Fig. 7.6) represents these four triples about the annotation.

45 See Ch. 4.

46 This script is available in the application repository on ↗GitHub and is called places.xqm, <https://github.com/BetaMasaheft/BetMas>.

47 <http://www.openannotation.org/spec/core/>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

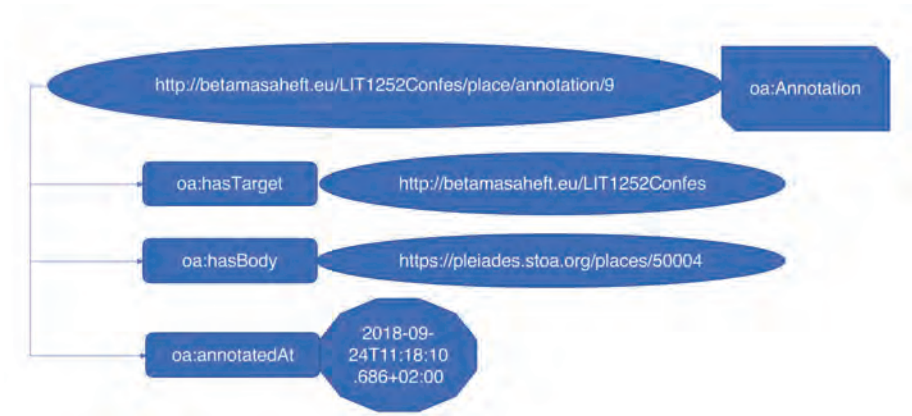


Fig. 7.6 Graphic view of one ↗annotation for ↗Pelagios.

Although these ↗annotations are presented separately, they are also part of the ↗RDF stored in Beta maṣāḥǝft, so that they can be queried as in the following example, which returns all the annotations about pleiades:50004.

```
SELECT DISTINCT ?target
  WHERE {
    ?x a oa:Annotation ;
    oa:hasBody <https://pleiades.stoa.org/places/50004> ;
    oa:hasTarget ?target .
      }
```

One finds out in this way that also the *Chronicle of Galāwdewos* (CAe 3122, ID: LIT3122Galaw) has a reference to India (pleaides:50004). This is not much of an improvement, as I could have found out the same with an even simpler ↗XPath on the collection, for instance `//placeName[@ref='pleiades:50004']`. To expand this query outside my collection, I could have used a federated query, and also include the Syriaca.org data, for example, like in the previous section, but this would have returned me only the places annotated with the ↗Pleiades place concept identifier in these two datasets, none instead linked, for example, to a Wikidata identifier for a concept related to that. But there is a third way one can use to exploit the advantages of discovery and aggregation of ↗LOD. Given that Syriaca.org, Beta maṣāḥǝft, and many others share the above ↗annotations with ↗Pelagios, I can instead query the Pelagios ↗API and get all the annotations which refer to any of the

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

concepts related to any of the identifiers associated with India.[48] Additionally, since this aggregated dataset is served via an API, I can display these results directly in my ↗application, as in the example above (Fig. 7.5) or in Peripleo (Fig. 7.7).[49]

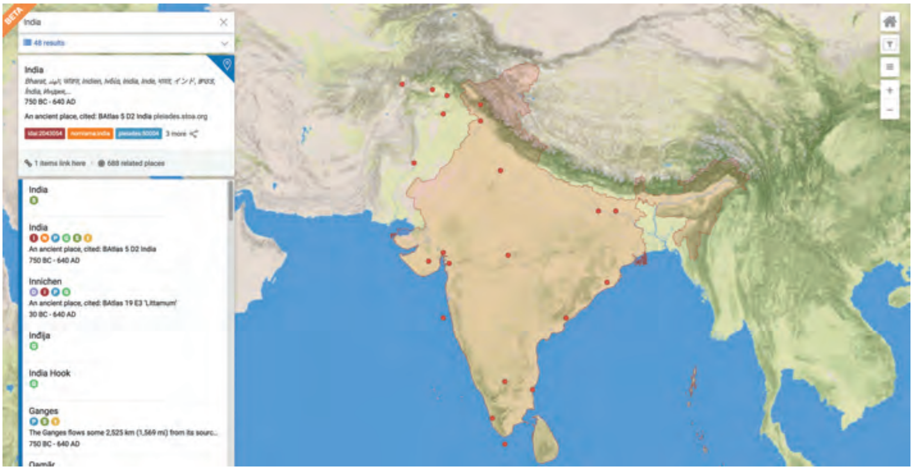


Fig. 7.7 Searching India in Peripleo.

We get thus more identifications, fully related, all accessible in one go, in multiple places, without updates and without much effort, just publishing a few links: this makes the contextualization of resources very easy and clear and it would be very nice if much more, which is relevant to ancient world data, could be done in this way.

2.4 Encoding networks of personal relationships

More or less the same kind of service provided to the community by the ↗Pelagios Network for places has been started for personal names and people networks by the ↗SNAP:DRGN (Standards for Networking Ancient Prosopographies: Data and Relations in Greco-Roman Names) project.[50] Although the project primarily aimed at Greco-Roman names, the ontology produced

48 See p. 41 for an example query to a third-party ↗API. The queries to the Pelagios API in Beta maṣāḥəft are done in various places in ↗XQuery or in ↗JavaScript. Even nicer would be to be able to point to a registry and navigate available datasets registered by partners, but this is not yet available.

49 <http://peripleo.pelagios.org>.

50 <https://snapdrgn.net/>, see also Bodard forthcoming. The version of the ontology used here is 1.0.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

has a much larger impact and usability. Therefore, looking forward to the continuation of this project, Beta maṣāḥǝft has made use of it to start modelling information about people in historical Ethiopia.

The core authority list for people has been formed from the analytic indexes of the *Encyclopaedia Aethiopica*, so that Beta maṣāḥǝft had one unique identifier for each person which is in that index.[51] Additionally, always thanks to the indexes of the *Encyclopaedia Aethiopica*, Beta maṣāḥǝft could start from an existing authority list of ethnic groups. Both these lists, as all the others which formed the starting data for Beta maṣāḥǝft were curated by Eugenia Sokolinski, who also assigned the identifiers according to the pattern still used by the project.

Each of the person records represents a person which existed at some point in some respect and can contain a series of pieces of information, starting from various names which are known as in the following example from the person record of ʾ*Aṣe* Yāʿqob (PRS10198Yaeqob), emperor of Ethiopia between 1597 and 1607.[52]

```
<persName xml:lang="gez" xml:id="n1"
                type="birth">
    ያዕቆብ ፡
</persName>
<persName xml:lang="gez"
                 type="normalized" corresp="#n1">
    <roleName type="title">ʾaṣe</roleName>
     Yāʿqob
</persName>
<persName xml:lang="en"
                corresp="#n1">
    Jacob
</persName>
<persName xml:lang="gez"
                type="regnal" xml:id="n2">
                መልክአ ፡ ሰገድ ፡
</persName>
<persName xml:lang="gez"
                type="normalized" corresp="#n2">
    Malkǝʾa Sagad
</persName>
```

51 <https://betamasaheft.eu/persons/list> allows one to search these records and all those subsequently created.

52 See Liuzzo et al. 2018 and especially <https://betamasaheft.eu/Guidelines/?id=persons> about the way names are encoded.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

The types of standard names registered here do not represent the collection of all the attestations of a given person in the corpus, which are given, to the extent to which they are annotated,[53] in the graph view and in the index,[54] but are rather a representation of the actual knowledge we have about the names of that person and are encoded there mainly with the purpose of disambiguation.

Also the <birth>, <death>, <floruit>, <occupation>, and <faith> elements are used, in order to provide a minimum set of disambiguating elements, together with bibliography about the person in question, when available.

Following the ↗SNAP:DRGN cookbook for producing ↗RDF data for a prosopography,[55] the triples produced from these XML records assign the identified person to the ↗class lawd:Person and record the basic information required as in the shortened example below.

```
<rdf:Description rdf:about="https://betamasaheft.eu/PRS10198Yaeqob">
    <dcterms:isPartOf rdf:resource="https://betamasaheft.eu"/>
    <rdf:type rdf:resource="http://lawd.info/ontology/Person"/>
    <foaf:name xml:lang="gez">ያዕቆብ ፡</foaf:name>
    <foaf:name xml:lang="gez">ʾaṣe Yāʿqob</foaf:name>
    <foaf:name xml:lang="en">Jacob</foaf:name>
    <foaf:name xml:lang="gez">መልክአ ፡ ሰገድ ፡</foaf:name>
    <foaf:name xml:lang="gez">Malkəʾa Sagad</foaf:name>
</rdf:Description>
```

Beta maṣāḥəft uses the same ↗annotation model which is used for places also for persons, assigning a different ↗URI structure to the annotation. Beside the utility of serving users with a stable URI for each person, for each name and attestation, using the ↗SNAP:DRGN Ontology helps Beta maṣāḥəft to model information about persons and their relationships.[56] For example snap: ontology properties are used, such as snap:occupation to represent in the ↗RDF the information encoded in the source ↗TEI. Also in this case, querying in

53 e.g. <https://betamasaheft.eu/persons/PRS10198Yaeqob/graph>.

54 <https://betamasaheft.eu/IndexPersons?pointer=PRS10198Yaeqob>.

55 <http://snapdrgn.net/cookbook>.

56 On ↗RDF for prosopographical knowledge basis see Chiarcos and Gracia forthcoming.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

↗SPARQL for `SELECT ?x WHERE {?x snap:occupation 'dabtarā' .}` or in ↗XQuery for `//tei:occupation[.='dabtarā']` does not return different results.[57]

Much more interesting and purposeful is instead the use of this to model relationships. If, for example, we want to say that *ʾAṣe* Yāʿqob is son of Māryām Śənā (PRS6829Maryams), this is an eminently relational piece of information, which we do not want to simply record as a sentence which could not be actioned by a machine or discovered in a graph. As for other pieces of information of this kind, we exploit the `<relation>` element to make this statement like in the example below.

```
<relation
   active="PRS10198Yaeqob"
  name="snap:SonOf"
  passive="PRS6829Maryams"/>
```

However, ↗SNAP:DRGN is more precise in the way it structures this information, and requires a node with ↗class `snap:Bond` and one in the class `snap:SonOf`. One statement will only say that the first person has a bond (`snap:hasBond`) with the second and will identify this bond with a ↗URI and a second statement will qualify this bond by giving it a type (`snap:SonOf`). A third and final statement will allow adding to this identified bond a further statement about the second person in the relationship (`snap:bond-with`). This might seem a lot to say something simple, but it unpacks the information enough to be able to query it in interesting ways by looking at the bonds instead of the persons, for example.

The ↗XSLT transformation producing the indexed ↗RDF will then elaborate the information in the `<relation>` element and its ↗attributes to produce the following triples.

```
<rdf:Description rdf:about="https://betamasaheft.eu/PRS10198Yaeqob">
   <snap:hasBond
      rdf:resource="https://betamasaheft.eu/bond/snap:SonOf-PRS6829Maryams"/>
</rdf:Description>
<rdf:Description
   rdf:about="https://betamasaheft.eu/bond/snap:SonOf-PRS6829Maryams">
  <rdf:type
      rdf:resource="http://data.snapdrgn.net/ontology/snap#SonOf"/>
```

57 In fact, for example, the attestations function of the Beta maṣāḥəft project uses a ↗XQuery function. This is fine until there are other ↗annotations around which Beta maṣāḥəft wants to query as well as its own, as in the case of places above.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

```
    <snap:bond-with
        rdf:resource="https://betamasaheft.eu/PRS6829Maryams"/>
</rdf:Description>
```

Having this set of nodes allows us to run ↗SPARQL queries to the ↗RDF data in order to build a graph of relations (Fig. 7.8), which can be then directly visualized as such. The query with ↗XQuery would in this case have to be serialized as nodes and edges and would be possible but less convenient compared to the SPARQL results formats.[58]

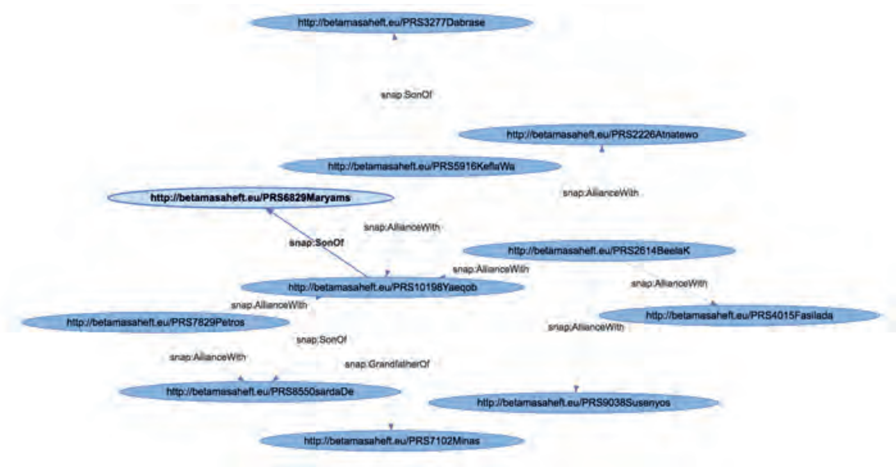


Fig. 7.8 Graph of the personal relationships of ʾ*Aṣe* Yāʿqob up to the second level in Beta maṣāḥəft.

Given this ↗RDF and the ↗annotations, which include a role ↗class derived from the @role in the <persName>, defined in the Beta maṣāḥəft Ontology (Fig. 7.1), we could do a generic query for 'the manuscripts with a patron of the imperial family'. A part of the query referring to the data described above would look like the following:

58 For the ↗SPARQL results serialization see Beckett and Broekstra 2013 for the specification in XML, and Seaborne 2013 for the specification in ↗JSON. Since the eXist-sparql module returns results in XML, Beta maṣāḥəft has an ↗XQuery function which transforms that XML to the corresponding JSON representation for use with libraries which support that, <https://github.com/BetaMasaheft/BetMas/blob/f9e0f9b01cd2f-f56ed332f67af68ce6a021b928e/BetMas/modules/sparqlRest.xqm#L169>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

```
1 ?annotation
2         a bm:patron ;
3         oa:hasTarget ?manuscript ;
4         oa:hasBody ?patron .
5 ?manuscript a bm:mss .
6 ?patron snap:hasBond ?bondName .
7 ?bondName rdf:type ?relation ;
8                 snap:bond-with ?ruler .
9 ?ruler snap:occupation 'Emperor' .
```

In this ↗SPARQL query we are looking first at the ↗annotation to isolate those which are about a patron of a manuscript (Lines 1–2), relying on the fact that the ↗RDF will have a statement linking a person to a manuscript and that, if that person is a patron, there will be a statement about that derived from the ↗attribute @role in a <persName> element. To make sure that the target of the annotation is a manuscript, at Line 5 we specify that the target of the annotation needs to be in the ↗class bm:mss.[59] The annotation body will be a person entity for sure, given that this is already controlled in the ↗schema which rules the XML so we do not need to specify anything else and we can instead ask that this person should have a bond of any type (the ?relation variable) with another person entity, which we expect there because of the ↗SNAP:DRGN Ontology. As a last step we specify that the person with which the patron should have a relation needs to be an emperor, and, therefore, we add that to the query by saying at Line 9 that any entity in the variable ruler should have a statement snap:occupation which links to a literal value which is equal to the word 'emperor'. This is so, again because in the Beta maṣāḥǝft RDF data that information is stored in this way, with a literal value for this property coming from a list of controlled values in the taxonomy used for the value of the attribute @type in <occupation>.

A full query, which organizes also slightly the results, could return for Beta maṣāḥǝft, in October 2018, the following:[60]

59 Here we are using two classes of the Beta maṣāḥǝft Ontology, but we could have limited also using the ↗LAWD ↗class lawd:assembledWork.

60 This can be reproduced, for example, using the Beta maṣāḥǝft ↗SPARQL endpoint, <https://betamasaheft.eu/sparql>. See above. The same query with a few lines of ↗JavaScript for rendering is used on this page in the book's website, and can be used to check the current answer: <https://pietroliuzzo.github.io/DHEth/table/mssSPARQL.html>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Table 7.2 **Manuscripts with a patron of the imperial family**

Manuscript	Patron	Relation	Ruler
London, British Library, Oriental 650 (BLorient650)	Zarʾa Yāʿqob	1) father of 2) son of 3) brother of	1) Baʾəda Māryām I 2) Dāwit I 3) Yəsḥāq
Oxford, Bodleian Library, Bruce 86 (BDLbruce86)	Sabla Wangel	1) wife of 2) grandmother of 3) mother of	1) Ləbna Dəngəl 2) Śarḍa Dəngəl 3) Minās
Oxford, Bodleian Library, Bruce 93 (BDLbruce93)	Mikāʾel Səḥul	enemy of	ʾIyoʾas I

Needless to say, this is just one example of a query on generic terms which can return precise results on the basis of the ↗RDF data with a simple query. But let me stress, once more, that this does not mean that just any question can be asked and magic will happen. This specific generic question can be asked because the information is available. One could not ask, for example, for scribes who worked in Egypt, because, although scribes could be listed and relations to Egypt or places in Egypt could also be highlighted for those of them who have them, there is no part of the ontology and no statement in the RDF to which one could rely to make sure that what they did in Egypt was 'working', in any understanding of what 'working' means. Given appropriate encoding and representation in the RDF of relevant information, this might become a possible question and obtain a processable response.

One last example can be made for role names attestations (Fig. 7.9), where the presence of ↗annotations allows one to call for each attestation of a `<roleName>` in association with the persons which carried that role name.[61]

61 It goes without saying that the assignment of a given role to a typology is an encoding decision which reflects the interpretation of the encoder of the text.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Fig. 7.9 The ↗SPARQL query on the <roleName> in *Edict of King Galāwdewos* (CAe 4969, ID: LIT4969EdictGala), tagged in the text returns other persons associated with the same role. See <https://betamasaheft.eu/works/LIT4969EdictGala/text?start=4>.

Given that data is updated constantly, the increase in available information will make this kind of query return more interesting results, but what is more important here is that, although it would be possible for somebody to contribute only the individual statements in Beta maṣāḥəft, it would be also possible for an independent project producing triples with the ↗SNAP:DRGN Ontology to query this information, which would be interesting, given that many patrons (and donors, sponsors, etc.)[62] of manuscripts are not Ethiopian and might fall in the scope of other projects.

Hopefully this data, as with the ↗Pelagios data, will make it into a larger network of interlinked datasets and allow the exploration of the many relationships of Ethiopians with other people throughout history.

2.5 Representing textual fluidity

It is a known fact that categories traditionally used for the Western modern book culture, such as 'work' or 'author', hardly apply to the literature of the

62 This can be checked in the advanced search of the Beta maṣāḥəft project, by selecting the specific filter for persons, then loading the role option and selecting the desired role, for example 'patron'. This will query the ↗eXist-db database to return all the patrons known and, additionally, the list of manuscripts they are patrons of, on a further click.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Christian Orient and Islam.[63] Texts are copied and varied, they can be in several recensions with each an independent manuscript transmission; they can be variously attributed to different persons for reasons which are also not always easy to determine, can be abridged or expanded, and can use narrative material from other works. This is not a simple thing to model in the data with any technology if one wants to stay as close as possible to the few known facts.

In Beta maṣāḥǝft we use the ↗SAWS (Sharing Ancient WisdomS)[64] Ontology to make these statements and rely on it and few other properties from other ontologies, in order to flexibly link textual units, narrative units,[65] persons, manuscripts, and other textual units.[66] Also in this case, like in the previous, Beta maṣāḥǝft uses the element <relation> to hold the triples in the ↗TEI data entry phase of the workflow.

Villa has already presented the way we use this ontology to help in modelling the tradition of the *Physiologus* (CAe 1401, ID: LIT1401Physio, see Fig. 7.10),[67] using the property saws:isVersionOf to record information about two abstract textual units.[68] To this end it is essential that there are ↗URIs which allow us to identify a text in different ways, for example as part of the contents of a manuscript or as abstract work or part of another work. In one example presented by Villa there are four textual units and one additional narrative unit for each animal discussed, but there is much more which is done with the ↗SAWS relations and requires making statements about a single manuscript part or a single part in a work.[69]

63 See the recent contribution Buzi et al. 2018, 50–56 and Orlandi 2018 on textual fluidity and the issues of identifying a ‘work’. See also Ch. 3.

64 <http://www.ancientwisdoms.ac.uk/method/ontology/>, see Tupman and Jordanous 2014.

65 These definitions are borrowed from Orlandi 2013 and the use Beta maṣāḥǝft makes of them is explained in the Beta maṣāḥǝft Guidelines, <https://betamasaheft.eu/Guidelines/?id=definitionWorks>, where, however, the term ‘work’ is used for textual unit.

66 See the Beta maṣāḥǝft Guidelines and, especially, <https://betamasaheft.eu/Guidelines/?id=definitionWorks> and <https://betamasaheft.eu/Guidelines/?id=relations>.

67 Villa 2018.

68 See also <https://betamasaheft.eu/Guidelines/?id=work-teiHeader>.

69 Only two examples have been created so far as a test-case.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

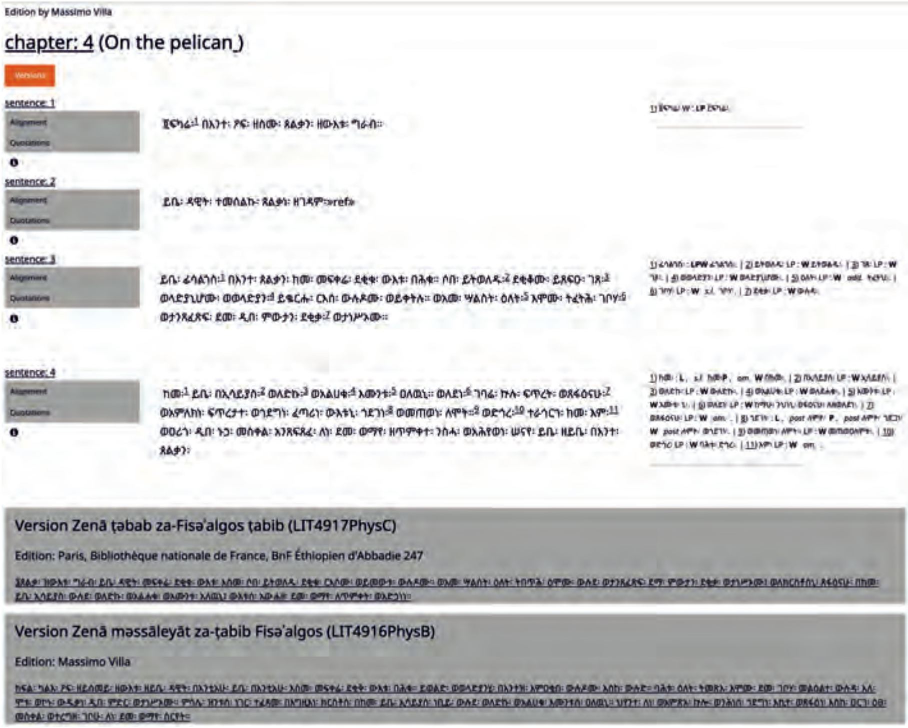


Fig. 7.10 Multiple editions and parallel versions of the *Physiologus* (CAe 4915, ID: LIT4915PhysA) as described in Villa 2018. See <https://betamasaheft.eu/works/LIT4915PhysA/text?start=4>.

In the following example we are linking directly from a specific content item in a manuscript to another manuscript, saying that, according to the sources used by the encoder, the content item described in the <msItem> with @xml:id ms_i1 in Paris, Bibliothèque nationale de France, Éthiopien 2 (BNFet2) is a direct copy of Cambridge, Bible Society's Library, BFBS Mss 169 (BSLet169).

```
<relation
    name="saws:isDirectCopyOf"
    active="BNFet2#ms_i1"
    passive="BSLet169"/>
```


ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Another example of direct use of this ontology are the authorship statements.[70] Instead of saying that a person X is author of the work Y, Beta maṣāḥəft uses saws:isAttributedToAuthor, which is a lighter statement than that of saying directly 'is author' in any form this might take, as an element <author> or a dc:creator. For example, in *Dərsān za-qəddus Yoḥannəs ʾafa warq baʾənta ṭəmqatu la-ʾəgziʾəna ʾIyasus Krəstos* (CAe 4866, ID: LIT4866OntheB) there are two such statements because the Greek version of this text is attributed to Gregory of Antioch while the Gəʿəz is attributed to John Chrysostom:

```
<relation
    name="saws:isAttributedToAuthor"
    active="LIT4866OntheB#t1" passive="PRS5720JohnChr"/>
<relation
    name="saws:isAttributedToAuthor"
    active="LIT4866OntheB#t2greek" passive="PRS4821Gregory"/>
```

We have to point to the titles of versions of the same text in other traditions, and this is in many respects wrong, but no better solution could be found at the moment to keep this information and produce reusable ↗annotations. The best way to go would certainly be to join in a ↗Pelagios-like network similar sets of triples from projects with a different scope (e.g. Arabic, Coptic, Syriac, and Hebrew literature) also using the ↗SAWS Ontology. This would require a new project and consortium to be put into the context of Linked Ancient World Data.

It is in this case very important to stress how any visualization or study of these triples without the assertions made by other specialists in other contexts is of little use, but, on the other side, this is exactly the point of producing these ↗annotations as ↗LOD, as their only usefulness is in connection with one another. No data repositories, virtual reading rooms, or research environments will ever contain all that is needed for any study aiming at some relevance.

We could also look at the properties saw:isVersionInAnotherLanguageOf, saws:isVariantTranslationOf and saws:isCloseTranslationOf to have a glimpse of the relationships between translations to Gəʿəz in different periods, but a representation of this information at this stage would not bring any interesting re-

70 See p. 83 where some other saws: properties are queried.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

sult.[71] Beta maṣāḥəft at the moment only shows the individual relations established (Fig. 7.11).

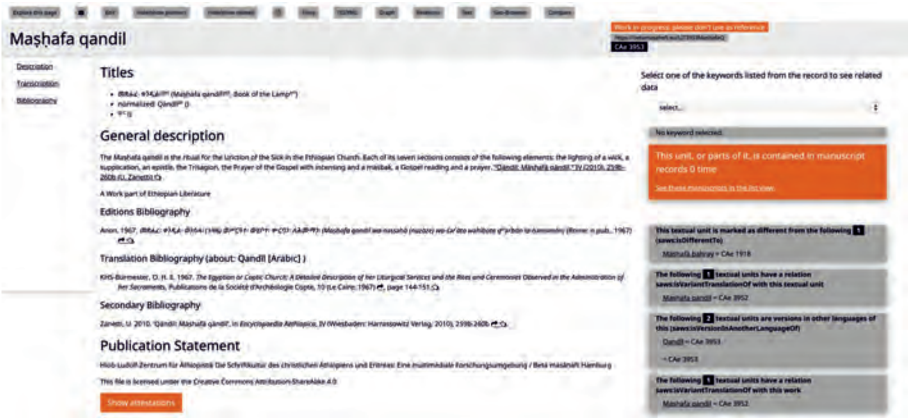


Fig. 7.11 ↗SAWS relations in *Maṣḥafa qandil* (CAe 3952, ID: LIT3952MashafaQ). See <https://betamasaheft.eu/works/LIT3952MashafaQ/main>.

2.6 Supporting syntactic manuscript descriptions

'It is best practice to use or extend an existing vocabulary before creating a new vocabulary'[72] but this is not always possible. And actually it is an opportunity offered by the Web of Data, that of being able to expand it. Whilst all previous examples are about the use of existing ontologies for specific tasks and areas of interest, no ontology existed that would allow the integration of the methodology described in *La syntaxe du codex*,[73] a ground-breaking methodological essay which allows for a full syntactical description of complex manuscripts.[74]

There is no plan at the moment in the project to redo all the descriptions of the manuscripts following this methodology, but there are too many things which a normal description would not be able to hold and we have then opted for the development of a new ontology. The book is, in fact, already an on-

71 See also p. 93.
72 <https://www.w3.org/TR/ld-bp/#ANNOUNCE>.
73 Andrist et al. 2013.
74 In Liuzzo forthcoming a I discuss in detail the development and implementation of this methodology in Beta maṣāḥəft.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

tology and needed only to be declined into a formal ↗OWL file.[75] Also in this case, as for previous examples, Beta maṣāḥəft uses the <relation> element of ↗TEI, with values for @name, @active, and @passive, to encode the statements necessary for these descriptions. The values used are either anchors in the file, identifiers, or classes in *La syntaxe du codex* Ontology. The use of these values strictly requires knowledge of the method described in the book, which cannot be summarized here.

The specific phenomena which can be described in addition to the formal description of a manuscript in the <teiHeader> are related to the history of the manuscript and can be encoded as a result of the application of the methodology, while supporting the application of it with the possibility to visualize intermediate stages. This is a main advantage of this flexible model, the ability to make very partial statements, and I will give an example of application to one such case, in which, although we are not currently in the position to perform a complete description following the methodology, the information offered by the previous cataloguers can be encoded and preserved with enough precision.

While cataloguing in Beta maṣāḥəft the manuscript Saint Petersburg, Institut Vostočnyh Rukopisej Rossijskoj Akademii Nauk, Orlov20,[76] Daria Elagina faced this situation, which she nicely reported in an issue on ↗GitHub as part of the project workflow:[77]

> In IVorlov20, I have two cases, where, according to Turaev's description, some text was presumably added later in another hand. The first one is the list of abbots of Dabra Libanos. Four last names are added in another hand. The second case is the 'chronology in cycles' (Turaev) from Adam till Iyasu I. 'The end of the text, starting from Yekuno Amlak' (Turaev) is written in another hand. I have no idea, how these 'additions' are laid out, but as next texts in both cases start on next folia, there is a possibility that these 'additions' are written at the bottom. And there is no hint whether they were written in one hand or in two different hands. In my opinion, these added portions should be regarded as additions, because they are written in another hand, presumably later. And these texts might also be regarded as complete without

75 The ontology prefix used in the example is sdc: which resolves to the ↗namespace of the ontology, <https://w3id.org/sdc/ontology>, available at <https://raw.githubusercontent.com/BetaMasaheft/SyntaxeDuCodex/master/SyntaxeDuCodex.owl>.

76 <https://betamasaheft.eu/manuscripts/IVorlov20/main>.

77 <https://github.com/BetaMasaheft/Documentation/issues/927>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

> this added information, so it was not about the completion of the text, rather about adding new information.

In the following code snippet I reproduce the encoding I have made of this.

```
<msItem xml:id="p1_i14">
    <locus target="#9v"></locus>
    <title ref="LIT4952History" cert="medium"></title>
    <textLang mainLang="gez"/>
    <msItem  xml:id="p1_i14.1" >
        <title type="incomplete" ref="LIT4952History"></title>
        <note>Extends till <persName ref="PRS5616IyasuI"></persName></note>
    </msItem>
    <msItem  xml:id="p1_i14.2" corresp="#h5">
        <title type="incomplete" ref="LIT4952History"></title>
     </msItem>
     <note>Starting from <persName ref="PRS10303Yekunno"></persName>
       till <persName ref="PRS5616IyasuI"></persName> the text is written in another hand.
      </note>
      <listRelation>
          <relation active="IVorlov20#UniProd1"
                 name="sdc:undergoesTransformation"
                 passive="IVorlov20#tr2"/>
           <relation active="IVorlov20#tr2"
                 name="sdc:produces"
                 passive="IVorlov20#UniProd3"/>
           <relation active="IVorlov20#tr2"
                 name="sdc:resultsIn" passive="IVorlov20#UniCirc2"/>
           <relation active="IVorlov20#tr2"
                 name="sdc:hasTransformationModel"
                 passive="sdc:A2"/>
           <relation active="IVorlov20#UniProd3"
                 name="skos:exactMatch"
                 passive="IVorlov20#p1_i14.2" />
           <relation active="IVorlov20#UniProd3"
                 name="sdc:constituteUnit"
                 passive="IVorlov20#UniCont" />
            <relation active="IVorlov20#UniProd3"
                 name="sdc:constituteUnit"
                 passive="IVorlov20#UniMain4" />
      </listRelation>
</msItem>
```


ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

The encoder had already structured the information about the content of the manuscript in a series of nested <msItem>s, each with a @xml:id (Lines 1, 5, and 9). She had also clarified the information given by the cataloguer in a <note>. This would be already sufficient and serve well enough the users, but I have then added a representation of this common phenomenon using the properties of the ontology to formalize this information. This, as any other element or statement has only the value of the hypothesis and tries to reflect the information available. I have added a <listRelation> element (Lines 15–36), which in the Beta maṣāḥəft ↗schema is allowed everywhere, containing a series of <relation> elements. In this example, since we do not know the order in which the two 'additions of contents without material support',[78] of which the cataloguer and the encoder speak, occurred, I have not stated an explicit equation with the main identifier of the manuscript, as we would normally do, to say that the final UniCirc (circulation unit), the one we have in Saint Petersburg, corresponds to that final stage. This is to say that we know that the manuscript being described is the latest available unit of circulation, but we do not know in which relation it is with the units of circulation resulting from each transformation. We also do not know if the two additions in question are part of the same action, so we have to consider them distinct for the moment.

I made instead two statements about the resulting UniCirc with each addition, one of which can be seen in the example above at Line 23. The ↗URI of the manuscript, as well as the UniCirc2 and UniCirc1 URIs defined in these sets of triples will be all in the ↗class of circulation units, but we are not stating which one comes first and which one comes after, since we do not yet know. The numeric part of the UniCirc identifier has no sequential value and is totally arbitrary, although it should be unique. In the same way, the UniProd anchors are set there and will become URIs in the ↗RDF, but are here defined for the first time.

To read the series of statements in the example from the beginning, what is being stated is that a unit of production which has been assigned the arbitrary numeric value 1 (Line 16) has undergone a transformation, whose name is 'tr2' (Line 18). The fact that the property sdc:undergoesTranformation relates an object in the ↗class UniProd or UniCirc to an object in the class of transformations is dictated by the ontology and inherited from the book directly. The transformation produces (Line 20) a new unit of production (UniProd), to which I have assigned the arbitrary numeric part '3'. The transformation is then said at Line 23 to result in a new unit of circulation (UniCirc2). The transformation is then qualified with the property sdc:hasTranformationModel

78 Type A2 in Andrist et al. 2013, 64.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

(Line 25) to say that it has one of a series of examples which are given in the book and can help to begin a classification of common phenomena (A2, addition of content without material support). The last three statements are about the newly produced UniProd3. The first states to which <msItem> it corresponds (Lines 27–29), the second states it is a unit of content (Lines 30–32), and the third (Lines 33–35) assigns to this a unit of hand (UniMain4).

The structure of the description in ↗TEI is thus not altered and an additional layer is added to it which can be exploited from the ↗RDF serialization as well. The additional use of the ontology based on *La syntaxe du codex* enriches the description where needed and allows for this further layer of encoding and further possibility of enquiry without requiring architectural modifications in the data structure. There will be no graph visualization nor table, although these are supported by Beta maṣāḥəft, because we are only trying to record a partial observation present in the catalogue.[79]

Being able to record these transformations, although separately and disjointed from the full account of the stratigraphy of the codex in question, still has its point when this basic statement become computable and we know in a modelled way that we have here cases of transformation A2 even if we cannot say what the sequence in which they occurred is in order to get a clear idea of the history of the manuscript.

Once again, adding ↗annotations to the ↗TEI which remain there and are serialized as ↗RDF and published as ↗LOD allows for an additional layer of distinctively relational queries to be made. For example, given sets of annotations which make use of this ontology, about manuscripts in any given tradition, quantitative analysis could be carried out on the transformations undergone by manuscripts with a simple and very generic query like the following, where all instances of the ↗class 'Transformation' are selected (Line 3), taking into consideration their products (Line 4) and model (Line 5). The results of the query, which can be run in Beta maṣāḥəft endpoint at any time, are exemplified in Fig. 7.12.

79 Examples can be seen for any manuscript with a completed encoding. See for instance Vatican City, Biblioteca Apostolica Vaticana, Cerulli 37 (BAVcerulli37), <https://betamasaheft.eu/manuscripts/BAVcerulli37/graph>; and Oxford, Bodleian Library, Aeth. e. 8 (BDLaethe8), <https://betamasaheft.eu/manuscripts/BDLaethe8/graph>. These visualizations are produced with d3graph.js and <vis.js> according to convenience, building on the results of queries to the ↗SPARQL endpoint, <https://betamasaheft.eu/api/SPARQL>, run from ↗JavaScript. The data can also be loaded in smarter and flexible tools like ↗Palladio, <http://hdlab.stanford.edu/palladio/>, where data can be loaded directly from the endpoint, although the ↗JSON variant needs to be used, <https://betamasaheft.eu/api/SPARQL/json>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

```
SELECT *
  WHERE {
        ?transformation a sdc:Transformation;
                        sdc:produces ?UnitProduct;
                        sdc:hasTransformationModel ?model .
        }
```

Transformation	UnitProduct	Model
https://betamasaheft.eu/BAVcerulli37/transformation/tr1	https://betamasaheft.eu/BAVcerulli37/UniProd/UniProd2	https://w3id.org/sdc/ontology#A1
https://betamasaheft.eu/BDLaethe8/transformation/tr1	https://betamasaheft.eu/BDLaethe8/mspart/p2	https://w3id.org/sdc/ontology#A1
https://betamasaheft.eu/BDLaethe8/transformation/tr1	https://betamasaheft.eu/BDLaethe8/mspart/p1	https://w3id.org/sdc/ontology#A1
https://betamasaheft.eu/BLorient452/transformation/tr1	https://betamasaheft.eu/BLorient452/mspart/p2	https://w3id.org/sdc/ontology#A1
https://betamasaheft.eu/BLorient452/transformation/tr1	https://betamasaheft.eu/BLorient452/mspart/p1	https://w3id.org/sdc/ontology#A1
https://betamasaheft.eu/BLorient655/transformation/tr1	https://betamasaheft.eu/BLorient655/UniProd/UniProd2	https://w3id.org/sdc/ontology#A2
https://betamasaheft.eu/BNFet45/transformation/tr1	https://betamasaheft.eu/BNFet45/UniProd/UniProd1i	https://w3id.org/sdc/ontology#D2
https://betamasaheft.eu/BNFet45/transformation/tr1	https://betamasaheft.eu/BNFet45/UniProd/UniProd1ii	https://w3id.org/sdc/ontology#D2
https://betamasaheft.eu/IVorlov20/transformation/tr2	https://betamasaheft.eu/IVorlov20/UniProd/UniProd3	https://w3id.org/sdc/ontology#A2
https://betamasaheft.eu/IVorlov20/transformation/tr1	https://betamasaheft.eu/IVorlov20/UniProd/UniProd2	https://w3id.org/sdc/ontology#A2

Fig. 7.12 Results of the ↗SPARQL query in the previous example.

From this representation of the data, the ↗TEI manuscript description gains a depth, by virtue of which I can query information about a specific UniCirc or UniProd in the history of the manuscript, without having multiple representations but only linking the evidence of the description of the current manuscript to the abstract entities defined in its syntactic description.

Pointing the browser to these URIs for UniProd and UniCirc or other units, the controller.xql will redirect to the relevant part of the description, so that reduplication of entity and consequent possible ambiguity are limited.

I hope I have shown enough cases for the need to use non-hierarchical statements together with the ↗TEI encoded data, as well as the importance of having this data shared openly for a new stage in the development of knowledge availability on the Web.

3 Conclusion

In this chapter I have described several ways in which Beta maṣāḥǝft produces and uses ↗LOD, but let me conclude quoting some words by Moretti which nicely summarize some key concept about data, cooperation, and sharing: 'Quantitative work is truly *cooperation*: not only in the pragmatic sense that it takes forever to gather the data, but because such data are ideally independent from any individual researcher, and can thus be shared by others, and combined in more than one way.'[80]

80 Moretti 2005, 5, emphasis in original.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

This is true not only of quantitative work, and it can be safely expanded to embrace more largely all research. ↗LOD offers a good way of sharing and a perfect way to combine and recombine. The availability of several forms of LOD (the ↗RDF, the ↗RDFa, the ↗json-LD according to ↗IIIF and ↗DTS specifications) is as vital to the future of digital resources for Eritrean and Ethiopian studies, as it is the openness of the source data in ↗TEI and its availability together with the Beta maṣāḥəft ↗schema,[81] because it is a core way to share data generously and foster collaboration, without which the collection of the data and the development of meaningful resources would be impossible. A culture of sharing and open access is fundamental to foster the future of these disciplines. It is not an additional value, but the basis for future development, paving the way with the establishment of URIs for many more projects to produce linked data and enrich one another in a network.

↗LOD allows us to cross the boundaries of the scope of the project by connecting to resources potentially unknown, and allows us to precisely qualify the connections established. In this respect, again, together with the ↗TEI encoding, it gives us tools which allow us to broaden the definitions, increase the precision, and allow multiple opinions and interpretations to coexist and flourish.

81 See Ch. 9, where these two concepts are considered again.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Chapter 8. Using XML Data for Research Products

> Die Sachen, die zu Tausend, müssen noch getan werden, und dazu braucht man etwas mehr Unbedenklichkeit, als ich habe. Etwas mehr Husarenhaftigkeit.
> (The things that in thousands still must be done, and for that one needs a little more unobjectionability than I have. To be a little more like a Hussar.)
>
> (E. Barlach, *Brief an Charitas Lindemann, 14. Juni 1910*)

This last chapter, which would like to continue a tradition of publications about technical possibilities for scholars in Ethiopian studies,[1] aims at giving only some basic information about the use of ↗XML to encode data, in order to produce so called 'traditional' research output, like articles or books. In previous chapters I have given examples of other traditional research activities, like cataloguing, studying literature, identifying places, and stating relationships. Here I will focus on the last stage of research activities, and the most important, namely publication. It will deal with how XML encoding, with a bit of programming, can be consistent with traditional research activities,[2] and how it can be useful even if one does not want to go all the way into the complexity of modelling of information or does not need reproducible visualizations and computations for his research.

It would be untrue to say that working on a publication in ↗XML instead of using a more traditional software makes things easier. It makes things better, however, even if then an author still has a quantity of tasks to take care of for publication. The main advantage of this is to be able to keep the work on the content as much separated as possible from that on the layout. An advantage which does not look like one until one tries that out, and which comes with many additional tasks to obtain a final product.

After a brief overview of the process of the production of output from ↗XML to PDF, I will give some details on the use of ↗Zotero for the bibliography management and discuss a few tools and basic options for getting XML out in print. This is not intended to give any more relevance to a print medi-

1 See Berhanu Beyene et al. 2000 for an article on encoding Ethiopic for ↗LaTeX.

2 Montfort 2016, 107.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

um over a digital one, as much as it should instead clarify how a digital-first approach to any research output is possible and beneficial.

1 Using XML to write a book

Today everything seems to be written in ↗JavaScript, from Node.js to React and the world revolving about these very nice, fast, and modular programming languages. So why don't we just use ↗JSON? Why XML and why ↗TEI? One could answer that we encode in ↗XML because we want to encode in TEI, and it would be already a good answer, but in the same way in which a two-cultures world for XML and ↗LOD is not productive, so is also this dichotomy.

This book is written in ↗XML and uses the data from Beta maṣāḥǝft in the same framework used for all other things discussed in it, namely ↗eXist-db. I am typing these words in ↗Oxygen XML Editor or ↗Atom, and using a ↗schema derived from ↗TEI with the Roma tool by simply using a few modules.[3] There are other more specialized XML vocabularies for encoding book content,[4] and one could have used ↗HTML directly: but why change really? In this book I have tried to use TEI everywhere, hopefully in a non-abusive way. There is no actual need to change to another standard to produce this type of output, when TEI has everything it takes.

The ↗TEI files are organized in a collection and one ↗XQuery module transforms the data in ↗XSL-FO,[5] which is then processed to become a PDF file with Apache FOP,[6] which comes as extension with ↗eXist-db.

For example the first part of the above paragraph looks like this in the ↗XML.

3 <http://roma.tei-c.org/>. The schema validating the files forming this book is available at <https://github.com/PietroLiuzzo/DHEth/tree/master/schema> as ↗ODD and compiled RELAX-NG. The modules included are core, tei, header, textstructure, tagdocs, namesdates, and transcr.

4 Publishing from ↗XML, is certainly not an innovation as many publishers and software for producing output for print use XML as a format at some stage.

5 This is available from <https://pietroliuzzo.github.io/DHEth/> or directly here <https://github.com/PietroLiuzzo/DHEth/blob/master/8%20Encoding%20Research/dheth2fo.xql>, although it is not directly reusable and needs to be tailored to the project in question. The module assumes a certain collection structure. It also assumes a particular configuration for the location of the fonts, which are stored in a separate library.

6 <https://xmlgraphics.apache.org/fop/> with the great reference by Antenna House, <https://www.antennahouse.com/antenna1/comprehensive-xsl-fo-tutorials-and-samples-collection/>. It is important to obtain hyphenation to add fop-hyph.jar and edit the settings, as described here: <https://stackoverflow.com/questions/34901011/exist-db-xsl-fo-hyphenation-and-settings>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

```
<p>The <term>TEI</term> files are organized in a collection and one
   <term>XQuery</term> module transforms
     the data in  <term>XSL-FO</term>,<note>This is available from
     <ref target="https://pietroliuzzo.github.io/DHEth/"/>. ... </note>
</p>
```

It is then transformed by the ↗XQuery module into the following bit of ↗XSL-FO:

```
<fo:block hyphenate="true" text-indent="0.43cm">
     The TEI files are organized in a
     collection and one
     <fo:inline font-size="smaller">↗</fo:inline><fo:inline
               id="N1.5.2.2.11.2">XQuery</fo:inline>
               module transforms
                 the data in  <fo:inline font-size="smaller">↗</fo:inline><fo:inline
                           id="N1.5.2.2.11.3">XSL-FO</fo:inline>,<fo:footnote>
                 <fo:inline font-size="7pt" vertical-align="text-top">2</fo:inline>
                 <fo:footnote-body text-align="justify" margin-left="0pt" text-indent="0">
                   <fo:list-block>
                     <fo:list-item>
                       <fo:list-item-label>
                         <fo:block><fo:inline vertical-align="text-top"
                           font-size="9pt">2</fo:inline></fo:block>
                       </fo:list-item-label>
                       <fo:list-item-body>
                         <fo:block space-before="0.45cm" font-size="9pt"
                           line-height="11pt" margin-left="0.45cm">This is
                           available from   <fo:basic-link
                           external-destination="https://pietroliuzzo.github.io/DHEth/">
                           <https://pietroliuzzo.github.io/DHEth/>
                           <fo:basic-link>  ... .</fo:block>
                       </fo:list-item-body>
                     </fo:list-item>
                   </fo:list-block>
                 </fo:footnote-body>
               </fo:footnote>
               ...
</fo:block>
```


ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

The ↗XSL-FO will contain then, because they are specified in the ↗XQuery, also a series of information which describes the layout of the PDF, which, in this case, follows the style of the Supplements to Aethiopica, and will produce the first line of the paragraph and the footnote in the example as you see it in the previous page.

The ↗XQuery relies on only one extra script,[7] which compiles a reference bibliography.xml from ↗Zotero and handles, for example, the letters at the end of a citation like 'Bausi 2018a'.[8]

The current version of this script would not support the production of another supplement or of a different publication, it would of course need to be tailored to the actual use of elements and ↗schema of the project, as well as to the structure of the book. For example, a critical edition could be produced by handling in the ↗XQuery functions also the elements which are part of the apparatus criticus in ↗TEI.[9]

This is not a better method to produce output compared to text processors like Word or any other publication-oriented software, because it requires some additional knowledge, but it offers complete control from source to output and typographic quality together.

To write up a chapter in ↗XML needs certainly knowledge of the basic rules of XML and of the chosen type of encoding, but, if done with an editor like ↗Oxygen XML Editor, for example, in author mode, it might be not that much of a different experience from that of writing with a more common text processor. Usually templating is also available, so that in Oxygen XML Editor for example it is possible to associate entire chunks of code which tend to be repeated often with a short-cut on the keyboard.[10] I have set a series of those, for example, which help me type 'Gəʿəz' and 'Beta maṣāḥəft', or one to enter elements for figures in this book or to make links.

7 Available from the book code base <https://pietroliuzzo.github.io/DHEth/> or directly at <https://github.com/PietroLiuzzo/DHEth/blob/master/8%20Encoding%20Research/generateBibliography.xql>. This ↗XQuery module makes a series of calls to the ↗Zotero ↗API for references formatted according to the HLCES Citation Style. The output needs post-processing because the style includes some italicization which gets escaped in the API responses.

8 This is done automatically by the ↗Zotero ↗API requesting a bibliography, but only for up to 150 titles.

9 On the representation of variants in ↗TEI, see Cayless 2018 discussing examples from the Digital Latin Library, <http://digitallatin.org/>, and comparing the printed and digital interfaces to the knowledge about the text we have.

10 See <https://www.oxygenxml.com/doc/versions/20.1/ug-author/topics/code-templates.html>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

The process is thus quite straightforward, but let me give a few more details on the only additional preprocessing step needed, which is necessary for the production of the citations and the bibliography.

1.1 Bibliography

↗Zotero has been mentioned a number of times but the way in which it is actually used for Beta maṣāḥəft or in this book was not clarified. Zotero is used to manage bibliographical references. It has several pieces of software, like a stand-alone and a plug-in version, and is able to import bibliographic references from web pages to create and maintain a structured bibliographic database which is kept in sync online. Additionally it offers a very rich ↗API,[11] which can be used to interact with the database.

To print the selected citations or references from the database in the desired format, ↗Zotero uses the ↗CSL (Citation Style Language)[12] so that, by associating a given style from the public repository to the selected data, the citations and references are formatted accordingly.

It does a lot of magic, especially when used openly, to share bibliographies and collaborate on them, saving a lot of time in data management and serving also as a discovery tool. It does not do everything, however, and the user must pay attention to his practice in entering data in the database.[13]

The generateBibliography.xml script, used in the production process of this book, for example, collects from the ↗XML files of all chapters of this book each and every distinct value of the @target inside the <ptr> element child of <bibl>, and queries the ↗Zotero ↗API for each of these values, which are tags in the Zotero bibliography pointed to by the script, to get the formatted bibliographical reference using the ↗CSL of the Hiob Ludolf Centre for Ethiopian Studies.[14] After querying all the references, the script builds the citation handle by counting the citations that are the same and adding to them a consecutive letter. The result is stored as an XML file bibliography.xml containing ↗HTML <div>s. This can be edited for those small things which are not possible to do within Zotero or because of further specification by the editor. For example here the entries of the *Encyclopaedia Aethiopica* are treated differently, and so the citation handle is what is printed in the citations, but only

11 <https://www.zotero.org/support/dev/web_api/v3/start>.

12 <http://citationstyles.org/>.

13 The following page <https://betamasaheft.github.io/bibliography/> details the ways in which we recommend to use the Ethio-Studies ↗Zotero Library, <https://www.zotero.org/groups/358366/ethiostudies/items>.

14 <https://www.zotero.org/styles?q=HLC>, see also <https://betamasaheft.github.io/bibliography/>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

one general reference is given in the bibliography, and this needs to be edited in the output XML file. The ↗XQuery which transforms the ↗TEI to ↗XSL-FO looks at this bibliography.xml file both for the citations and for the final bibliography, and prints the handle in the text and the full citation in the Bibliography assigning IDs and linking to them, so that it is possible to click on the one citation and get to the full reference, but also to get from the full reference back to the individual citation where one was, at least when looking at the PDF in a viewer, not at the paper book.

It is simple. But not all publications are this easy in their requirements and, in the following section, I will discuss a few options which are available for more complex needs.

1.2 Printed critical editions

In Ch. 3, I have detailed a few of the features which make a digital edition richer and more desirable than a print edition. However, one of the benefits about working with ↗XML is that one can produce different outputs from the same source (a concept referred to as 'plurality of representations'), and the printed book is still the main output of research.[15]

This process of professionally typesetting the encoded text is not one that, in most cases, will be actually performed by the author, and depends on the type of output, on the style of the series and publication, and so on, so that the choice of the workflow to use will not always be entirely free. This is, of course, all to the advantage of quality, since the author can care about the content mainly and the typesetter about the ways in which to correctly and nicely lay that out on pages.

The range of possibilities is enormous and I am not going to give a complete review of all what is available to researchers in order to lay out an edition autonomously for print starting from their ↗XML source, I will only describe some of those possibilities which are freely available and have been produced for researchers.

There is no way to provide a fully generic 'TEI to book' software,[16] because there are many 'TEI' and even more 'book' formats, so readers should not expect to find here the solution to their problems, nor should they consequently walk away from any encoding just because it does not directly give a printed output (which is not what it is supposed to do). It remains important

15 On the lively debate on digital scholarly editions see the recently collected contributions in Driscoll and Pierazzo 2016 and Bleier et al. 2018.

16 However, the standard ↗TEI ↗XSLTs do provide a transformation to PDF and other formats which can be customized if sufficient.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

to encode texts and make them available in some format, as described in the previous chapters, regardless of the final intended output of a project.

Let me take as example, to demonstrate some more complex requirements of a print product, one section of the *Historia Episcopatus Alexandriae* (CAe 5064, ID: LIT5064HEpA), as edited by Bausi and Camplani (Fig. 8.1).[17] This edition has some interesting features:

1) The parallel translations on the right page are synced to the parallel texts on the left;
2) The two parallel texts in Gəʿəz and Latin have a different numeration;
3) There are multiple apparatus linked to any of the four texts;
4) The contemporary presence of elements of the transcription in the edition;
5) Synced marginal texts.

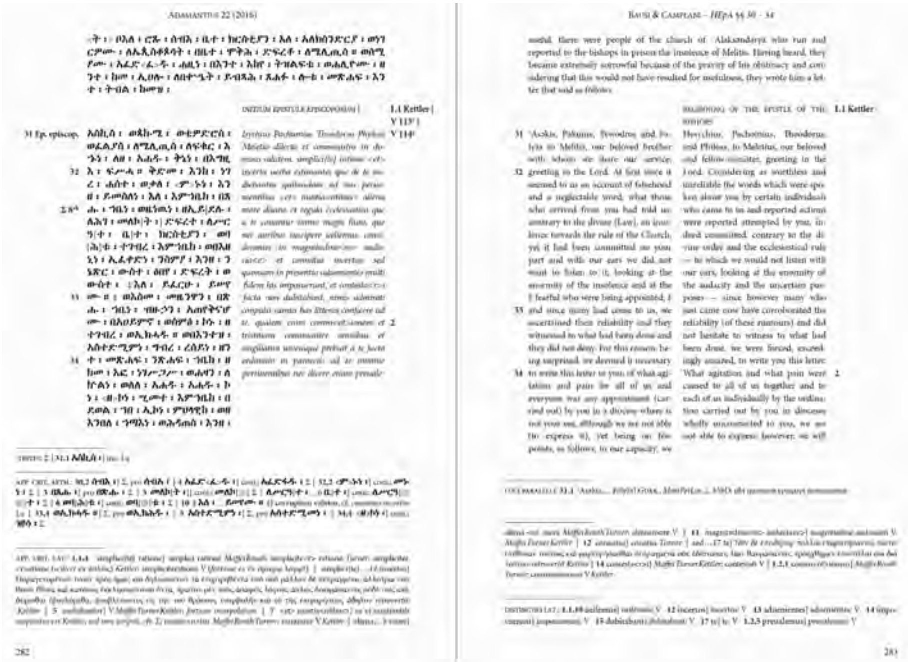


Fig. 8.1 Parallel versions of the *Historia Episcopatus Alexandriae* with critical apparatus and facing parallel translations.

17 Bausi and Camplani 2016, 283–284.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

To be noted here are also some features of the TEI which are not immediately evident from the PDF and which Alessandro Bausi kindly discussed with me.

1) < > indicate critical emendation in general. For example, in the apparatus note 'magnitudine<m> audacia<e>] magnitudine audaciam V' the first angle brackets, as it can be deduced from the witness V reported, indicate an omission, while, the second ones indicate a correction. The encoded text in ↗TEI looks like the following.

```
magnitudine<supplied reason="omitted">m</supplied>
audacia<choice><corr>e</corr><sic>m</sic></choice>
```

2) Lacunae in the Gəʿəz manuscript (the unique witness) are represented in the text of the edition with square brackets, whereas in the apparatus a special character is used to show the missing material support.
3) The *fidal* text breaks at any point inside a word, without hyphenation, except before punctuation.
4) There is a separate apparatus for the punctuation (*distinctio*).
5) There is a separate apparatus for the parallel texts.

I have encoded these chapters (only) in a ↗TEI edition for testing the different options available, and the ↗XML encoded text is available in Beta maṣāḥəft. However I am not going to be able to show you how close I could get to this output with other methods, because I could not get very close in a reasonable amount of time. On the other side, I would not want to give the impression that this is an easy task and that it can be done by anyone: it does instead require professional care and it can be achieved with one or the other of the techniques and tools listed below according to the needs of each project.

1.2.1 Web publication

In the best of possible worlds, the two texts, Latin and Gəʿəz, would need to be in two different repositories and should be optionally visualized in parallel with the use of a ↗DTS viewer, for example, collecting texts from different APIs which have, for instance, a ↗SAWS relation `saws:isVersionInAnotherLanguageOf` stated,[18] but we are not yet at this stage in the availability of resources online, so the Latin parallel version has been also encoded with the Gəʿəz one, although it is not available in Beta maṣāḥəft.[19]

18 See Ch. 3 and Ch. 5.

19 According to the Beta maṣāḥəft ↗schema and Guidelines, the Latin version would have to be linked with a ↗SAWS relation.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

An arguably better web-based digital edition can be generated from a ↗TEI encoded text using several freely available tools, even without looking at database software like ↗eXist-db.[20] The number of options for web publication, in terms of applications, frameworks, and so on, does not allow a brief review: any can be used. I will focus only on tools which have support for TEI, and, also here, by far just a small selection of them.[21]

CETEIcean,[22] by using ↗Web Components, allows one to 'use slightly-modified ↗TEI directly in the ↗browser with ↗CSS and JavaScript enhancements. We can leverage the TEI data model directly in the reading environment.'[23] It is not difficult to create several <div>s for different elements and similarly to define the way the apparatus should look like as the examples show very nicely.

The software EVT2 (Edition Visualization Technology)[24] provides a very flexible and lightweight way to publish a critical edition with several witnesses, simply generating a configuration file with an editor.[25] This is an easy and fully featured solution for those who want to produce a website in no time from their encoded texts.

Textual Communities provides a collaboration environment for TEI editions, including support for images via ↗IIIF,[26] collation with ↗CollateX and a collation editor and is, thus, a very good option for projects looking at a lot of encoding with a goal of web publication.

Other more generic software, like OMEKA, can also be used successfully for web publication of critical editions,[27] and the traditional simple transformation with ↗XSLT from ↗XML to ↗HTML is still a good option for static websites.

20 An issue of RIDE (<https://ride.i-d-e.de>) is in preparation, about tools for digital scholarly editing, which will provide a lot more details about a number of useful tools, see <https://ride.i-d-e.de/reviewers/call-for-reviews/tools-and-environments-for-digital-scholarly-editing/>.

21 Also the Beta maṣāḥəft presented in Ch. 1 is an example of visualization using a generic framework.

22 <https://github.com/TEIC/CETEIcean>, Cayless and Viglianti 2018.

23 Cayless 2018, 257.

24 <https://visualizationtechnology.wordpress.com/2017/10/31/evt-2-beta-1-available-for-download/>, see Rosselli Del Turco et al. 2014; and Di Pietro and Rosselli Del Turco 2018.

25 <http://evt.labcd.unipi.it/evt2-config/>.

26 <https://textualcommunities.org>.

27 <https://omeka.org/classic/showcase/>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

A further set of tools which has been used for critical editions and aimed at being a solution for both web and print publication is ediarum,[28] developed and maintained by the Telota group at the Berlin-Brandenburgische Akademie der Wissenschaften.[29] This set of resources, based on ↗Oxygen XML Editor, ↗eXist-db and ConTeXt offers Web output, print output, collaborative work and tools to check the contents produced, and it is based on TEI while striving for user friendliness. Its only disadvantage is that it needs somebody to be able to install it and maintain it.

A new and promising way of publishing texts, only online, which is also a suite of resources which a developer has to implement and which uses CTS (Canonical Texts Services)[30] or ↗DTS is the CapiTainS suite, 'an informal open-source organization which aims at providing a suite of tools and guidelines for Citable Text APIs standards'[31] with its different components and also a template app, Nemo. This includes also a continuous integration testing feature which makes sure that the ↗XML is not only valid according to the ↗schema but also sensible for the ↗API representation, and thus able to be properly displayed in the ↗application.[32] This is by now, in my opinion, the most complete and well-documented software suite available oriented to text publication which fully supports DTS.

A web-based output is also preferable due to the nature of a digital edition, which remains open for quick updates, however, 'while the open-endedness of digital scholarship is often held up as an advantage over print publication because it allows its authors to incorporate new findings and documents as they are discovered, this very open-endedness of the digital can also be interpreted as one of its greatest weaknesses.'[33] This is especially true since researchers are still asked to present printed books as their main output.

None of the above software is intended for a print publication and it is easily arguable that a web-based edition offers much more and in a better way than a printed edition. An openly accessible edition is even better in as far as

28 <http://www.bbaw.de/telota/software/ediarum>, see Dumont and Fechner 2014; Dumont and Fechner 2019.

29 <http://www.bbaw.de/telota/telota>.

30 Blackwell and Smith 2014; Almas and Clérice 2018.

31 See <http://capitains.org/> and <https://github.com/Capitains>, Clérice et al. 2017. More bibliography about CapiTainS can be found at <http://capitains.org/pages/publications>.

32 The project Formulae - Litterae - Chartae (<https://werkstatt.formulae.uni-hamburg.de/>) in Hamburg uses this suite of tools successfully based on CTS data and ↗XML according to the Digital Latin Library Guidelines (<https://digitallatin.github.io/guidelines/>).

33 Edmond 2016, 61.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

it breaks the barrier set by price on a book or the accessibility of a library. However, it is also arguable that the sustainability of the digital resource is not comparable to that of a book, and having both web and printed results is almost a must for any serious project to which a ↗TEI-based workflow offers a solid semantic basis. Finally we cannot take for granted that the audience of any research result will like to navigate it in a website. A very large portion of the scientific community still prefers a PDF to download and print.

Hence, I will give a few details in the following sections regarding the tools which can be used to produce a print edition from ↗TEI,[34] with the notable exclusion of TUSTEP.[35]

1.2.2 Example of workflows for print publication

1.2.2.1 Using XSL-FO

This is the method used for this book. I could not however do all that the example requires, especially the syncing on the two pages of two synced texts, in conjunction with the apparatus.[36]

It is not that easy and accessible to everyone to work with ↗XSL-FO or with ↗LaTeX and, if the LaTeX package reledpar does not currently support columns inside parallel texts, in XSL-FO one could get parallel texts in a table, for example, but parallel flows in facing pages are not supported.

1.2.2.1.1 TEI Publisher

The TEI Publisher,[37] which is an ↗eXist-db ↗application, is the only tool which, to my knowledge, actually allows for a 'digital first' workflow which

34 See also <https://wiki.tei-c.org/index.php/Publishing_printed_critical_editions_from_TEI> for a community maintained list of options.

35 <http://www.tustep.uni-tuebingen.de/tustep_eng.html>. I am also not going to discuss generic conversion tools like OxGarage (<http://oxgarage.tei-c.org/#>) or Pandoc (<http://pandoc.org/>), although they are very useful at many stages of any workflow.

36 I was not able to implement the solution proposed on Stack Overflow by Tony Gram, <https://stackoverflow.com/questions/52775123/original-text-on-left-side-with-facing-translation-on-the-right>. nor the one kindly suggested to me by Wolfgang Meier, starting from neither of the ↗XML files at hand (the one produced for Beta maṣāḥəft, the one tested in EVT2, and the one from the ↗TEI Critical Edition Toolbox). That is probably because of lack of skills and time on my side, so, this remains the top option to be considered from my point of view, for a resaercher doing everything on his own.

37 <https://teipublisher.com>. Wicentowski and Meier 2015; Turska et al. 2016.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

out of the box delivers the online version of the data, but also PDF, either via ↗LaTeX or ↗XSL-FO,[38] and ↗ePub as well.

TEI Publisher does this also in a very nice way, that is, allowing the user to define this in the ↗ODD, where the definition of the ↗schema, which provides explicit evidence of the ↗XML structure, resides. The ODD editing is also facilitated by a specific interface, and, although this still requires knowledge of ↗XPath, ↗CSS, and of the models and behaviours used by this ↗application, this is with no doubt to date the best available resource, which is not only providing what is needed but is doing it in a neat and meaningful way.

For some simple encodings, it might well be possible that this ↗application can already produce the desired result, or a usable intermediate format to edit for the final book product.

1.2.2.2 Using LaTeX

A conversion with ↗XSLT or ↗XQuery from the ↗TEI to ↗LaTeX is a rather straightforward task, if it is clear which is the source code and which the desired output.

Users have especially two packages which are relevant to the production of critical editions, reledmac and reledpar by Maïeul Rouquette.[39]

However, the parallel texts in facing pages are not supported yet, which means that it is not possible to date to obtain with ↗LaTeX the result in our example.[40] However, this is an extreme case and all the other features of the critical edition in the example can be typeset using LaTeX to obtain the desired result.

But as there are few people ready to encode their text in ↗TEI, probably not many more like to work with ↗LaTeX, and this has prompted the development of several tools which produce this encoding for typesetting on the basis of more or less defined TEI encoding, so that the user can potentially see the PDF doing only the TEI.

1.2.2.2.1 Lombard press

Jeffrey C. Witt maintains in relation to his project SCTA (<https://scta.info/>) a nice library called lombardpress-print which, once installed, allows one to

38 A user can also get the source file in ↗LaTeX and ↗XSL-FO, to edit it and compile it locally.

39 Both available from CTAN.

40 This was kindly confirmed to me via email by Maïeul Rouquette on the 10 October 2018.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

directly view a PDF output running one script in the command line.[41] The conversion to ↗LaTeX is driven by ↗XSLT, so that, for any specific pagination need, these can be changed or more modules can be added.

1.2.2.2.2 TEI Critical Edition Toolbox

The TEI Critical Edition Toolbox has an interface to customize directly online a transformation which will produce a ↗LaTeX file and the PDF output.[42] This is useful in case you want to include the LaTeX file in another LaTeX project or run it again after modifying it.

In case you do not have TeX installed, you can use online software like Overleaf to compile the ↗LaTeX file again.[43] Also here an example is provided which can inform the choices for the encoding, if one decides to use this tool in his workflow.

1.2.2.3 Using the Classical Text Editor (CTE)

The example from which we started is a PDF produced with the Classical Text Editor (CTE) by Stefan Hagel (version 9). This is a very good visual editor for a critical edition,[44] the only one to my knowledge that can achieve that. CTE is certainly to be recommended if the primary aim of the work is to produce a book and there is no primary interest in sharing data in reusable formats for any of the other possible purposes presented in this book.

CTE provides functionality to export some ↗TEI, but of course this is only representing the current edition and does not validate to any ↗schema.[45] This laudable effort requires thus additional work to have also a digital resource shareable for collaboration in the same way in which primarily web-oriented resources require extra work to produce a print version.

41 <http://lombardpress.org/seamlessly-converting-a-tei-critical-edition-to-camera-ready-print-proofs/>.

42 Burghart 2016, <http://teicat.huma-num.fr/check.php>.

43 <https://www.overleaf.com/>.

44 <http://cte.oeaw.ac.at/>. To use CTE on Mac see also <https://andrewdunning.ca/cte-on-the-mac>.

45 Some projects rely on this export from CTE to produce the ↗TEI data which they then process and present in websites, for example Formulae - Litterae - Chartae and the HUNAYNNET project, for which see Kessel et al. 2018.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

1.2.2.4 Using Word or InDesign

You can convert a ↗TEI file to Word or one of many other formats you could use in your favourite text editor by simply using the TEI stylesheets.[46] Some Beta maṣāḥəft users actually convert to Word with online tools the PDF produced (via ↗XSL-FO) in the website, as it already does a lot of formatting.

However it is unlikely that any out-of-the-box conversion will be exactly what you need or do everything without the need of some final fixes. The Faust Edition project, for example, has a workflow via an ↗XSLT which transforms to InDesign Markup Language (IDML).[47]

2 Conclusion

So, in the end, there is no winner, just a lot of valid choices to be made early in each project, on top of many options, but one thing remains the only one which is really important, that is to curate interchangeable data in ↗TEI and share it widely and generously. In this sense, any of the workflows which start from TEI, maintain and publish it online with an open licence are preferable and can be used also for additional products. Delegating somebody else the encoding as a secondary task duplicates or triplicates the effort: all work to get a book, then again to the get the TEI from it, and again to get a product from the TEI, print or web based. So, let me conclude with the following quote:

> If scholarly communications in the digital age are to overcome a publication culture based on printed texts delivered in electronic format, then the reader will also engage differently with the arguments and evidence contained within, with sources and with the overt (as in thematic) or hidden (as in a database structure) organization of knowledge.[48]

This citation brings us to the conclusion of this book, in the next chapter, which summarizes these two as the core concepts for a research in Ethiopian and Eritrean studies.

46 <https://github.com/TEIC/Stylesheets/releases>. However, this is not as easy as it sounds, but the ↗TEI List has occasional exchanges where the latest methods used and implemented are shared.

47 <http://faustedition.net/>. Thanks to Gerrit Brüning who has kindly informed me about this workflow, which might become available in the near future with documentation to make it reproducible. The pipeline is already available, <https://subversion.le-tex.de/customers/faustedition/transpect/>.

48 Edmond 2016, 57–58.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Chapter 9. Collaboration and Openness

> Delineare il futuro è un atto politico. Senza questo, che fine fa l'idea universalistica della 'fratellanza umana'?
> (To sketch the future is a political act. Without this, where will the universal idea of 'human brotherhood' end up?)
> (Stefano Levi Della Torre, 'Qualche considerazione sulla storia in corso', *Una Città*, 252, 2018)

Beta maṣāḥəft is for everyone, and especially the source data it produces (with its ↗schema and guidelines) and the diverse data formats it offers for the study of written artefacts and related entities. It might, as all digital projects, never be accomplished, never be complete. As a research environment, it needs to remain open to all the little bits that anyone will want to contribute to the common effort to become bigger and bigger for everyone to enjoy 'facilitating each others' work, and validating each others' results.'[1] What is to be hoped is that, at some point, it might reach and surpass the knowledge that is available today to become a reference resource not for undeniable truths but for numerous interpretations and stable data structures.

The collection and encoding of data, in ways which allow its reusability and continuous repurposing in new resources and products of different kinds and the production of a research environment, do change the way in which some parts of research can be done, but they do not change the way in which questions need to be formulated. A research environment can support methodologies and be part of them, but does not remove the complexities of research, as Vattuone formulated in the final words of his introductory lessons on Ancient Greek historiography:

> è lecito dubitare, tuttavia, che la rapidità di selezione e raccolta di un materiale immenso, con risparmio di tempo impressionante, muti in qualche modo la sostanza dell'inchiesta: dal linguaggio serrato, asciutto del *net* escono *notizie, argomenti, testi,* che sarebbe imprudente considerare *oltre* la dimensione documentale. L'epoca degli 'ipertesti' non

1 Edmond 2016, 55.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

ci esimerà purtroppo dalla fatica di *trovare* i fatti storici e di costruirne un intreccio.

(one can doubt, however, that the speed in the selection and collection of a huge quantity of materials, with an impressing time saving, will change in some way the substance of the inquiry. From the short, dry language of the *net* emerge *news, topics, texts* which would be imprudent to take into consideration *beyond* the documental dimension. The age of 'hypertexts' will not exempt us from the labour of *finding* the historical facts and reconstruct their interweaving.)[2]

While in general it is vital that students and scholars are aware and in full command of the ways in which information can be, and is, collected in the Web and as digital resources,[3] this should not replace the scientific interests of codicologists, papyrologists, philologists, and historians, but rather enhance their capability to investigate their research questions. This is the first point I would like to make concluding this book.

A second final consideration is that I believe that an open approach to digital resources is one of the few ways in which collaboration in the humanities can be fostered for the progress of knowledge about written artefacts from Eritrea and Ethiopia. In this respect, the importance of the summer school organized in September 2018 by Maq̠ala University and the Beta maṣāḥəft project, with the support of the Volkswagen Foundation, cannot be understated, also for the fact that it included a training in digitization and a short introduction to digital methodologies. More such events should follow in Ethiopia and Eritrea and continue to spread access to common knowledge bases,[4] encouraging the reuse and production of open data to benefit every researcher and user of any level.[5] The possibility to work digitally should benefit Eritrean and Ethiopian scholars in the first place.

One can type angle brackets in ↗TEI on his or her laptop also in a monastery in Ethiopia or Eritrea, and it is no news that researchers can work with their laptops in the field. Light and easily portable machines with more than ten hours batteries enhance these possibilities and it becomes just really a matter of choices.

The usability of the standard Beta maṣāḥəft tools in the field was tested during the 20th Internal Conference of Ethiopian Studies in Maq̠ala, on a vis-

2 Vattuone 2006, 96, emphasis in original.

3 See on this the concerns expressed in Bagnall and Heath 2018, 187.

4 This will actually continue with a summer school in ᵓAddis ᵓAbabā in 2019.

5 On the critical point of access to the Internet, see also Zaccaria 2018.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

it to the monastery of Koholo Yoḥannəs Dabra Betel (INS0171KY)[6] with Antonella Brita, Alessandro Bausi, Gianfranco Agosti, Solomon Gebreyes Beyene, and myself.

Very generously the monks brought us some manuscripts, among which a *Gadla samāʿətat* (*Acts of Saints and Martyrs*, CAe 1493, ID: LIT1493Gadlas), already known to the scientific community having been digitized in one of the early missions of the Ethio-SPaRe project by Denis Nosnitsin and his team, and since Antonella Brita had already studied it during her research and Alessandro Bausi also used it for a recent edition of the *Gadla ʾAzqir*.[7]

We proceeded at the analysis of the quire structure (collation), mainly carried out by Antonella Brita, while I registered her observations directly in ↗XML on my laptop using ↗Oxygen XML Editor and validating to the locally synced copy of the Beta maṣāḥəft ↗schema. Using then ↗eXist-db as installed on my computer, we could run the application locally and check the visualization with ↗VisColl. Only few imprecisions in my encoding came up and we could fix them there and share with the monks the visualization of the quire structure in the application while eating together bananas and bread.

Later, upon return to the hotel with an internet connection, the file could be pushed to GitHub and go live in ESky001.[8] This is how quick and easy it can be. The choices of a free and open software and of a free and open repository, we hope, will encourage this sort of contributions to the data by many other scholars who are able to reuse the entirety of the data produced in the research environment.

This was not a planned test, but it helps to prove that working with ↗XML in a monastery or a church in Eritrea or Ethiopia does not present more issues than bringing a camera and a tripod and can thus be integrated in digitization plans and research activities. It would be actually nice to have a programme to bring computers and cameras to the monasteries, train young people locally to do the encoding on the spot, and contribute to their heritage preservation, while passing on skills useful also to be spent in the digital technologies market. Taking back these tasks to the place where they belong would have multiple benefits and, although needing time, it could lead to a much more thorough digitization and documentation project, which could include paintings, objects, painted inscriptions, and many other features. But, even more importantly, it would mean more involvement of the owners of the heritage in its preservation and study and, perhaps, encourage more people to follow on in their study with the research on these topics.

6 Nosnitsin 2013, 259–263.

7 Bausi 2017.

8 <https://betamasaheft.eu/manuscripts/ESky001/main>.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

It is important to photograph and digitize manuscripts, but, without a proper description, openly shared, and open access to these images, the result can build on more frustration especially for young researchers. It should become a norm that, together with the digitization of the manuscripts, also metadata in ↗TEI following a common ↗schema should be produced. I hope that, having been developed in the public and with open methodologies, Beta maṣāḥǝft will be the first port of call for any future project embracing this quality standard for the documentation of the manuscript tradition of Ethiopia and Eritrea, and that future projects will add as a non-optional tool to their equipment, together with cameras and tripods, also some of these very light angle brackets, contributing to the definition of a common schema. The numerous collaborations which the project has already seen, the generosity and engagement of other digital projects in the field and the numerous participants to the several workshop run in addition to the Summer schools and direct contributors to the online resources, are a good sign of a shared will to proceed in that direction. These two first conclusive remarks hint at collaboration especially. Let me add one final thought to this Conclusion, on openness.

Openness was one of the possible titles of the sixth of the memos written by Italo Calvino for the Charles Eliot Norton Poetry Lectures to be held at Harvard in 1985–1986. Part of this theme is what probably became *consistency*.[9] It is uncertain if only the title was changed, or, together with many others, the topic changed and what we know about *consistency* does not apply at all to what *openness* would have been.[10] But without indulging in imagining what Italo Calvino would have noted down for the next millennium about *openness*, which is so important to all the topics of this book and of digital humanities and to the Web as a whole, there are other interesting points of reflection that digital methodologies for the study of written artefacts can take

9 There is no available digital edition of the manuscript notebooks on which these lectures have been prepared, nor photos as far as I know. It would be very useful indeed to have such edition with photos. However, Calvino 1995, 2957–2964 gives a synthetic account where it looks like *openness* is linked to *lightness* to begin with, then becomes a separate theme related to 'sense of space', the relation between inside and outside (separated from the 'sense of connection'), then becomes the concluding lecture where 'sarebbero state recuperate idee relative al rapporto tra l'io e gli altri' ('ideas relative to the relationship between the I and the others would have been newly addressed'). Beside the many editions of this classic, Italian-speaking readers might also listen to the wonderful radio series <https://www.raiplayradio.it/programmi/pantheon/archivio/puntate/Lezioni-americane-1988-2018--01cb1170-5e23-4968-9c17-6887b1eb5c57>.

10 Scarpa 2005, 91–93 makes me think this is the case.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

from this work of Calvino, and I would like to conclude with two of them, from *Exactitude* and *Multiplicity*.

The points made by Calvino speak also of the Web in some way, and should be important for those encoding and publishing resources online as much as they are for those writing a good novel. There is a shared responsibility between the digital curator and the writer in as far as they deal with the most precious of the things human kind has, ideas.

In *Exactitude*, the first and third of the three meanings, which he sets out to begin with, all speak also to the researcher as encoder, who makes the largest part of his work defining everything precisely and so that it might be computed correctly, but, at the same time, is concerned with the correct encoding of each piece of information, so that it is semantically correct and precisely encoded. In a digital project often we do a work as the one that Calvino exemplifies with his *Città Invisibili*, 'una struttura sfaccettata in cui ogni breve testo sta vicino agli altri in una successione che non implica una consequenzialità o una gerarchia ma una rete entro la quale si possono tracciare molteplici percorsi e ricavare conclusioni plurime e ramificate' ('a multifaceted structure in which each short text is close to the others in a succession which does not imply consequentiality or hierarchy but a network in which is it possible to trace multiple paths and extract multiple and ramified conclusions')[11] The making of digitally encoded texts is in its essence a struggle towards the kind of exactitude which does not close arguments but actually opens up more questions, more encodings, more interpretations. The encoded text has a machine operable structure which can be queried in many ways and leads to many different interpretations.

In the fifth lecture, *Multiplicity*, Calvino wrote of the novel as a network of facts, people, and things in the world. I really wondered what he would have thought of Linked Open Data, but I think he would have seen the way in which, at least theoretically, it encompasses the complexity of multiplicity, and groups it into a consistent and coherent whole. At the same time, however, the semantic Web is no novel and cannot be read as such, it requires authors of research and narrative to use it. And in this is not different from history which needs an historian to be told and become such.

A research environment dealing with written artefacts is certainly not enough without the researchers involved in it and does not by itself determine a decisive innovation in research. Perhaps, however, the resources, made available and easy to interact with, will enable scholars of Eritrean and Ethiopian studies to jump into a future of online collaboration and *openness*?

11 Calvino 2016, 72.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Glossary

Aggregator A service which harvests data from content providers and harmonizes them for others, <https://en.wikipedia.org/wiki/Data_aggregation>. The EAGLE network, for example, is a data aggregator for Europeana. [Introd.: xliii; Ch. 1: 42, 44; Ch. 2: 48, 57]

Annotation In this book the term is used loosely to refer to triples in a less technical way. An annotation is often intended as a statement or a series of statements about something, <https://www.w3.org/TR/annotation-vocab/>. [Introd.: xxxvii, xxvii, xxxi, xxxix; Ch. 1: 9, 34, 34, 35, 35, 35, 35, 36; Ch. 3: 103; Ch. 4: 126, 136, 107, 107, 107, 107, 108, 108, 108, 109, 109, 115, 120, 120, 122, 122, 122, 123, 123, 123, 123, 123, 123, 123, 124, 124, 124, 124, 127, 127, 127, 127, 127, 127, 127, 128, 128, 128, 128, 129, 133, 133, 137, 140, 141, 141, 141; Ch. 5: 146, 160; Ch. 6: 161, 162, 163, 163, 164, 172, 177, 177, 177; Ch. 7: 193, 195, 198, 198, 198, 198, 198, 198, 198, 198, 198, 199, 199, 199, 199, 199, 202, 202, 203, 204, 205, 205, 205, 205, 206, 210, 210, 215, 215]

API Application Processing Interface, <https://it.wikipedia.org/wiki/Application_programming_interface>. [Introd.: xxxvi, xxxvii, xl; Ch. 1: 2, 2, 23, 41, 41, 42, 42, 42, 44, 44, 45, 45, 46; Ch. 2: 53, 54, 73, 75; Ch. 3: 82, 82, 82, 82, 82; Ch. 4: 120, 123, 124, 125; Ch. 5: 156, 156, 157, 157, 157, 157, 159, 160; Ch. 7: 187, 192, 192, 192, 193, 194, 199, 200, 200; Ch. 8: 222, 222, 222, 223, 223, 228]

Application A web application is software running in a browser. Beta maṣāḥəft, as well as the online *Lexicon linguae aethiopicae*, the Beta maṣāḥəft Guidelines and the Gəʿəz Morphological Parser presented in this book are all web applications powered by eXist-db. [Introd.: xxxiv, xxxviii, xxxviii, xxxix, xxxix, xl, xl, xli, xlvii, xlix; Ch. 1: 2, 9, 19, 20, 21, 31, 32, 38, 40; Ch. 2: 57; Ch. 3: 82, 82, 97; Ch. 4: 128; Ch. 5: 144, 144, 144, 144, 148, 148, 149, 151, 152, 157, 158; Ch. 6: 161, 165, 166; Ch. 7: 185, 187, 190, 192, 200; Ch. 8: 228, 229, 230, 230]

Atom <https://atom.io/> is a free and open source text editor, very useful for code. It has a large number of packages for the needs of almost anyone, especially developers. It can be used to ed-

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

it and validate XML, and is integrated with GitHub, so that it can be used for the entire workflow of the project, from editing XML and validating it (in Beta maṣāḥəft we use <https://atom.io/packages/linter-autocomplete-jing>) to push to GitHub and manage the database (with the dedicated package, <https://atom.io/packages/existdb>). [Introd.: xl; Ch. 7: 186; Ch. 8: 220]

Attribute In XML, inside the opening tag, attributes can be specified separated with a space from the element name. [Introd.: xxxix; Ch. 1: 20; Ch. 7: 203, 205, 205]

Browser A software to navigate the Web, e.g. Chrome, Firefox, Opera, <https://en.wikipedia.org/wiki/Web_browser>. [Ch. 1: 28; Ch. 6: 178; Ch. 8: 227]

CIDOC-CRM Conceptual Reference Model of the International Committee for Documentation, <http://www.cidoc-crm.org/>. [Introd.: xlii; Ch. 2: 49, 69, 70, 71, 71, 71, 72; Ch. 7: 187, 196]

Class In a web ontology, corresponds to `rdfs:Class`, <https://en.wikipedia.org/wiki/RDF_Schema>. [Introd.: xlii; Ch. 1: 35; Ch. 2: 71; Ch. 7: 198, 202, 203, 203, 204, 205, 205, 214, 214, 214, 215]

Collatex It is a software used for textual criticism and especially to collate manuscript transcriptions, <https://collatex.net/>. [Ch. 8: 227]

CSL Citation Style Language, <https://citationstyles.org/>. This is the format of the bibliography style used by Zotero and other bibliography management systems. [Ch. 8: 223, 223]

CSS Cascading Style Sheets, <https://it.wikipedia.org/wiki/CSS>. [Introd.: xxxii; Ch. 8: 227, 230]

CSV Comma-Separated Values, <https://it.wikipedia.org/wiki/Comma-separated_values>. [Ch. 1: 38; Ch. 2: 48, 52]

Docker <https://www.docker.com/>. [Introd.: xx]

DTS Distributed Text Services, <https://w3id.org/dts>, see Almas et al. 2018. [Introd.: xxxvi, xxxvii; Ch. 1: 2, 42, 46; Ch. 2: 53, 75; Ch. 3: 82, 82; Ch. 5: 156, 157, 157, 157, 158; Ch. 7: 187, 187, 190, 190, 193, 193, 193, 194, 217; Ch. 8: 226, 228, 228]

Dublin Core Dublin Core, <http://dublincore.org/specifications/>. [Ch. 7: 186]

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

EpiDoc Schema, guidelines and example stylesheets for ancient documents in TEI, < http://www.stoa.org/epidoc/gl/latest/>. [Introd.: xxxvii; Ch. 1: 46; Ch. 2: 49, 50, 50, 50, 50, 52, 52, 53, 53, 53, 53, 54, 75; Ch. 5: 155]

ePub An open standard for the publication of digital books, <https://en.wikipedia.org/wiki/EPUB>. [Ch. 8: 230]

eXist-db <http://eXist-db.org/> shipped with several core applications like the Integrated Development Environment app called eXide which can be a good alternative to either Oxygen XML Editor or Atom for editing XML files. [Introd.: xix, xxxiii, xxxiv, xxxv, xxxvii, xxxvii, xxxviii, xxxix, xl; Ch. 1: 15, 31, 44; Ch. 2: 57; Ch. 3: 82; Ch. 4: 124; Ch. 5: 146, 148, 149, 149, 155, 158; Ch. 7: 184, 185, 185, 186, 188, 189, 189, 207; Ch. 8: 220, 220, 227, 228, 229; Ch. 9: 235]

FAIR Findable, Accessible, Interoperable, Reusable, <http://www.datafairport.org/>. [Introd.: xxxiv, xxxiv, xxxvi]

GeTA A software for multilevel linguistic annotation developed by Cristina Vertan for the TraCES project. [Introd.: xxxi; Ch. 5: 146, 159; Ch. 6: 162]

GitHub A hosting service primarily for software with plenty of useful features for deploying code to the Web, <https://github.com/>. [Introd.: xxxiv, xxxiv, xxxvi, xxxvii, xxxviii, xxxviii, xxxviii, xxxix, xl, xl; Ch. 1: 2, 3, 28, 37; Ch. 2: 49; Ch. 7: 185, 185, 185, 198, 212]

HTML HyperText Markup Language is the core technology used to present content over the World Wide Web, for example in a browser, <https://en.wikipedia.org/wiki/HTML>. [Ch. 1: 2, 18, 20; Ch. 2: 56, 57, 61; Ch. 5: 153; Ch. 6: 167, 178, 178; Ch. 7: 187, 187, 188, 188, 188, 189, 189; Ch. 8: 220, 223, 227]

HTTP HyperText Transfer Protocol allows the exchange of information over the Web, <https://en.wikipedia.org/wiki/Hypertext_Transfer_Protocol> (and HTTPS, same with Secure, <https://en.wikipedia.org/wiki/HTTPS>). [Introd.: xxxiv; Ch. 1: 44, 44]

Hypothes.is A tool for web annotations, <https://web.hypothes.is/>. [Ch. 4: 123, 123, 126, 127, 127, 128]

IIIF International Image Interoperability Framework defines standard APIs to interact and present images. To serve im-

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

ages according to IIIF, an image server like IIPI Image Server (<http://iipimage.sourceforge.net/>) is needed. [Introd.: xxxii, xxxvi, xxxvii; Ch. 1: 2, 20, 20, 20, 21, 21, 21, 21, 22, 22, 22, 22, 22, 22, 23, 23, 23, 25, 30, 46; Ch. 2: 75; Ch. 7: 187, 192, 192, 192, 193, 193, 194, 217; Ch. 8: 227]

JavaScript A programming language core to the Web which is used to enable interaction between user and a web page. Together with HTML and CSS is one of the core technologies enabling our experience of navigation of the World Wide Web, <https://en.wikipedia.org/wiki/JavaScript>. [Introd.: xx; Ch. 2: 57; Ch. 5: 153, 157; Ch. 6: 178; Ch. 7: 191, 200, 205, 215; Ch. 8: 220]

JSON JavaScript Object Notation is a standard data format, <https://en.wikipedia.org/wiki/JSON>. [Ch. 1: 34, 44, 44, 46; Ch. 4: 123, 128; Ch. 7: 204, 204, 215; Ch. 8: 220]

json-LD JSON Linked Data is a further JSON specification to have Linked Data in it, <https://en.wikipedia.org/wiki/JSON-LD>. [Ch. 7: 184, 187, 217]

KML Keyhole Markup Language is an XML data structure which is used for geographic information, <https://en.wikipedia.org/wiki/Keyhole_Markup_Language>. [Introd.: xxviii; Ch. 4: 128, 128]

KWIC Key Words In Context is a way to present results of a search for text as concordance where the result is presented with the string preceding and following it, <https://en.wikipedia.org/wiki/Key_Word_in_Context>. [Ch. 1: 16; Ch. 5: 149, 155]

LaTeX It is a descriptive markup language which is used for high quality typography, <https://en.wikipedia.org/wiki/LaTeX>. [Ch. 8: 219, 229, 229, 230, 230, 230, 230, 230, 230, 231, 231, 231, 231, 231]

LAWD Linking Ancient World Data is an ontology which functions to close the gaps between the use of different vocabularies for the representation in RDF of data about the ancient world, <http://lawd.info/>. [Ch. 7: 196, 197, 205]

LOD Linked Open Data is a format for structured data which is openly available through the Web allowing semantic queries, <https://en.wikipedia.org/wiki/Linked_data>. [Introd.: xxxvii; Ch. 2: 68; Ch. 7: 182, 182, 183, 183, 183, 187, 187,

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

187, 187, 187, 189, 190, 191, 191, 191, 192, 192, 192, 194, 194, 195, 197, 197, 197, 197, 197, 197, 197, 197, 198, 198, 199, 210, 215, 216, 217, 217, 217; Ch. 8: 220]

Manifest It is the first of the primary resources in a IIIF presentation API, containing the description of the digital representation of an object, <https://iiif.io/api/presentation/2.0/#primary-resource-types>. [Introd.: xxxii, xxxvii; Ch. 1: 20, 21, 21, 22, 22, 24, 25, 30; Ch. 7: 192, 192, 192, 193, 193]

Mirador It is a great IIIF data viewer, which allows one to easily interact with the presented images and compare them, <http://projectmirador.org/>. [Ch. 1: 21, 22, 22, 23, 23, 23; Ch. 7: 192, 192, 192, 193, 193]

MyCoRe It is a content management system, <http://www.mycore.de/index.html>. [Introd.: xxix, xxxvii; Ch. 1: 5]

Namespace It defines to what set of definitions your data refers. If you say that 't:' is the abbreviation for the TEI namespace, then saying in your data and code t:p means that you are referring to a paragraph (<p>) in TEI, <https://en.wikipedia.org/wiki/Namespace>. [Ch. 6: 167, 167; Ch. 7: 186, 187, 212]

ODD One Document Does it all is the abbreviation for the most magic of the parts of TEI: a TEI customization. This is the file where a local TEI schema is defined, where you can add your project documentation and also rendering instructions, <http://www.tei-c.org/guidelines/customization/getting-started-with-p5-odds/>. [Introd.: xxxiv, xxxv, xxxv, xxxv, xl; Ch. 8: 220, 230, 230]

OWL Web Ontology Language, <https://it.wikipedia.org/wiki/Web_Ontology_Language>. [Ch. 7: 183, 183, 184, 186, 212]

Oxygen XML Editor It is the standard software used to edit and work with XML <https://www.oxygenxml.com/>. It does really all you need to do. Alternatives might be the eXide, the editor application shipped with eXist-db, and Atom with specific packages. This book is written with Oxygen XML Editor. [Introd.: xl; Ch. 5: 144, 147; Ch. 8: 220, 222, 228; Ch. 9: 235]

Palladio It is a set of 'data-driven tools for analyzing relationships across time' by the Humanities + Design Research Lab, Stanford University, <http://hdlab.stanford.edu/palladio/>. [Ch.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

1: 34, 35, 35, 36, 38, 38, 39, 40, 40, 43, 45, 45, 46; Ch. 3: 84, 85, 85, 85, 86, 86, 88, 101; Ch. 7: 215]

Pelagios The Pelagios Commons is a network of projects and people which share annotations about ancient places as LOD, <http ://commons.pelagios.org/>. [Introd.: xxvii, xxvii, xxxvii, xxxvii; Ch. 1: 34, 35, 37; Ch. 2: 54, 54; Ch. 4: 107, 109, 120, 120, 122, 123, 123, 124, 124, 124, 125, 125, 126, 141; Ch. 7: 194, 196, 197, 197, 197, 198, 199, 199, 199, 200, 207, 210]

PeriodO It is a gazetteer of period definitions, <http://perio.do/en/>. Beta maṣāḥəft defined the main periods of Ethiopian history here to be able to associate them to the data and LOD produced, especially for Pelagios. [Ch. 1: 9; Ch. 2: 51; Ch. 3: 95]

Perseus It is the digital library for classical texts which has led the way for years as open project serving huge amounts of free and TEI encoded ancient texts, together with a series of very useful tools, <http://www.perseus.tufts.edu/hopper/>. [Ch. 1: 1; Ch. 4: 113, 115, 123, 123, 123; Ch. 6: 162]

Pleiades It is the gazetteer of ancient places. It is community-built and serves as a graph of ancient places and their relations. It was originally based on the work done for the *Barrington Atlas of the Greek and Roman World* , <http://pleiades.stoa.org/>. [Ch. 2: 54, 68; Ch. 4: 106, 106, 106, 107, 117, 117, 119, 120, 120, 120, 125; Ch. 7: 199]

POST An HTTP request method used to send data to be stored, <https://en.wikipedia.org/wiki/POST_(HTTP)>. [Ch. 7: 185]

Python It is a very quick and powerful modular programming language, <https://www.python.org/>. [Introd.: xx, xxxix]

RDF Resource Description Framework is the standard for data interchange in the Web, <https://www.w3.org/RDF/>. [Introd.: xx, xxxv, xxxvii, l; Ch. 1: 9, 20, 24, 34, 34, 35, 36, 41, 42, 46; Ch. 2: 48, 49, 68, 68, 68, 68, 69, 71, 73, 73; Ch. 3: 79, 79, 79, 82, 84, 85, 93; Ch. 4: 123, 123, 124; Ch. 7: 182, 182, 182, 183, 183, 183, 183, 184, 184, 184, 184, 184, 184, 185, 185, 186, 187, 187, 187, 187, 188, 188, 188, 188, 188, 188, 189, 191, 191, 191, 195, 195, 196, 197, 197, 199, 202, 202, 202, 203, 204, 204, 205, 205, 206, 206, 206, 214, 215, 215, 217]

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

RDFa RDF in the attributes is a standard defining how to include RDF in the HTML page served, by adding URIs in the attributes of HTML elements, <https://rdfa.info/>. [Ch. 1: 2, 20; Ch. 7: 187, 217]

RDF-XML It is a way of serializing RDF in XML, <https://en.wikipedia.org/wiki/RDF/XML>. [Ch. 2: 57; Ch. 4: 123, 123; Ch. 7: 184, 185, 185, 186, 188]

Recogito It is a tool for the annotation of texts and images, especially dedicated to geo-resolving toponyms by associating them to a Pleiades place URI, <https://recogito.pelagios.org/>. [Ch. 4: 122, 122, 122, 122, 123, 123, 123, 127, 129, 136, 140]

RegEx Regular Expressions is a way of searching patterns in text instead of strings and is used in most programming languages, in different flavours. It is extremely useful in any search context, <https://en.wikipedia.org/wiki/Regular_expression>. [Ch. 1: 12; Ch. 5: 144, 147, 147; Ch. 6: 166, 166, 174]

RESTXQ RESTful XQuery, developed by Adam Retter, see Retter 2012. [Introd.: xxxvii; Ch. 7: 187]

SAWS Sharing Ancient WisdomS is the name of an ontology produced during a project by the same name for structuring the fluid relations 'between or within' texts, <http://www.ancientwisdoms.ac.uk/>. [Ch. 3: 82; Ch. 7: 208, 208, 210, 211; Ch. 8: 226, 226]

Schema It is the document against which the correctness of a XML document is validated. It can be written in several schema languages, <https://en.wikipedia.org/wiki/XML_Schema_(W3C)>. Beta maṣāḥəft uses RELAX-NG (REgular LAnguage for XML Next Generation, <http://www.relaxng.org/>). The schema for a TEI project can be produced from the ODD. [Introd.: xxxiv, xxxiv, xxxvii, xxxvii, xxxvii, xxxviii, xxxviii, xxxviii, xxxviii, xxxviii, xl, xl, xl, xli, xli; Ch. 1: 12, 19, 31; Ch. 2: 50, 50, 50; Ch. 3: 79, 89, 93, 95; Ch. 5: 155; Ch. 7: 182, 187, 205, 214, 217; Ch. 8: 220, 222, 226, 228, 230, 231; Ch. 9: 233, 235, 236, 236]

SGML Standard Generalized Markup Language, <https://en.wikipedia.org/wiki/Standard_Generalized_Markup_Language>. [Introd.: xxxi; Ch. 3: 79]

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

SKOS Simple Knowledge Organization System is used to publish thesauri and vocabularies for use in the Semantic Web, <https://www.w3.org/2004/02/skos/>. [Ch. 2: 49, 63]

SNAP:DRGN Standards for Networking Ancient Prosopographies defines an ontology and model for publishing information about the relationships among ancient people, <http://snapdrgn.net>. [Ch. 7: 196, 200, 202, 202, 203, 205, 207]

SPARQL SPARQL is the query language for RDF, <https://www.w3.org/TR/sparql11-query/>. [Ch. 1: 2, 34, 34, 36, 41, 41, 42, 43, 43, 45, 45, 46, 46; Ch. 2: 53; Ch. 3: 84, 84; Ch. 4: 123; Ch. 7: 183, 195, 203, 204, 204, 204, 205, 205, 207, 215, 216]

SVG It is the standard used to present vector images in the Web, <https://www.w3.org/TR/sparql11-query/>. [Ch. 1: 28, 30]

TEI Text Encoding Initiative is a consortium and a community of members of the consortium and users which maintain the TEI Guidelines, and a series of other resources core to the TEI, <www.tei-c.org>. [Introd.: xxxiii, xxxiii, xxxiv, xxxiv, xxxiv, xxxiv, xxxiv, xxxiv, xxxiv, xxxiv, xxxv, xxxv, xxxvi, xxxvii, xxxvii, xxxvii, xxxvii, xxxvii, xxxvii, xxxvii, xxxviii, xxxix, xl, xli, xlii, xlii, xlvii, xlvii, xlvii, xlvii; Ch. 1: 2, 3, 3, 3, 3, 4, 5, 9, 16, 17, 18, 18, 19, 20, 20, 21, 21, 21, 21, 23, 24, 26, 28, 29, 31, 31, 31, 31, 32, 33, 46, 46; Ch. 2: 50, 50; Ch. 3: 77, 79, 79, 79, 79, 84, 88, 88, 88, 89, 93, 97; Ch. 4: 108, 120, 124, 128, 137; Ch. 5: 144, 146, 146, 147, 155, 155, 157, 157; Ch. 6: 162; Ch. 7: 182, 183, 183, 184, 184, 184, 184, 185, 186, 186, 186, 186, 187, 189, 189, 192, 194, 196, 197, 197, 202, 208, 212, 215, 215, 216, 216, 217, 217; Ch. 8: 220, 220, 220, 220, 220, 220, 222, 222, 224, 224, 226, 226, 227, 227, 227, 229, 229, 229, 230, 230, 230, 230, 231, 231, 232, 232, 232, 232, 232, 232; Ch. 9: 234, 236]

Triple It is a statement made of three parts: subject, predicate, and object. It is the core data model of RDF, <https://en.wikipedia.org/wiki/Semantic_triple>. [Introd.: l; Ch. 3: 82; Ch. 4: 124; Ch. 7: 183, 183, 184, 184, 184, 186, 186]

Turtle Terse RDF Triple Language is one of many possible ways of serializing RDF triples, <https://en.wikipedia.org/wiki/Turtle_(syntax)>. [Ch. 4: 108, 123; Ch. 7: 184, 186, 189, 198]

Up-conversion The process of taking a string text to structured data, e.g. XML; Kay 2008, 906. [Introd.: xxxvii; Ch. 5: 144, 144]

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

URI	Unique Resource Identifier, <https://www.w3.org/TR/cooluris/>. [Introd.: xxxiv, xxxiv; Ch. 1: 12, 20; Ch. 2: 49, 49, 52, 57, 68, 68, 68, 68; Ch. 3: 82, 85, 85, 85, 85, 89, 93; Ch. 4: 112, 120; Ch. 7: 182, 184, 186, 186, 186, 187, 187, 188, 188, 188, 191, 192, 196, 198, 202, 202, 203, 208, 214, 214]
URL	Unique Resource Locator, <https://en.wikipedia.org/wiki/URL>. [Introd.: xxxvi; Ch. 1: 16, 22, 22, 23; Ch. 4: 123, 127; Ch. 5: 157; Ch. 6: 167]
URN	Unique Resource Name, <https://en.wikipedia.org/wiki/Uniform_Resource_Name>. [Ch. 4: 123]
VisColl	It is a set of scripts in XSLT to display the collation of a manuscript starting from a TEI encoded description and a set of images, <https://github.com/leoba/VisColl>. [Ch. 1: 28, 28, 28, 28, 29, 30, 30; Ch. 9: 235]
VoID	Vocabulary of Interlinked Datasets, <https://www.w3.org/TR/void/>. [Ch. 1: 2; Ch. 4: 108, 123; Ch. 7: 189, 190]
Web Component	See <https://www.w3.org/TR/custom-elements/> and <https://developer.mozilla.org/en-US/docs/Web/Web_Components>. [Ch. 8: 227]
webhook	<https://developer.github.com/webhooks/>. [Ch. 7: 185]
XML	eXtensible Markup Language, <https://en.wikipedia.org/wiki/XML>. [Introd.: xx, xx, xxxiv, xxxiv, xxxv, xxxiv, xxxiv, xxxvi, xxxvii, xxxvii, xxxviii, xxxix, xl, xl, xl, l; Ch. 1: 2, 3, 5, 19, 19, 23, 31, 31, 36, 46; Ch. 2: 48, 52, 57; Ch. 4: 123, 128; Ch. 5: 144, 149, 155; Ch. 6: 161, 163, 163; Ch. 7: 181, 182, 182, 184, 187; Ch. 8: 219, 219, 219, 219, 219, 220, 220, 220, 220, 220, 220, 220, 222, 222, 223, 223, 224, 224, 224, 226, 227, 228, 228, 229, 230; Ch. 9: 235, 235]
XPath	It is a query language used to select nodes in an XML tree, used, for example, by XSLT and XQuery, <https://en.wikipedia.org/wiki/XPath>. [Ch. 1: 33; Ch. 2: 51; Ch. 7: 199; Ch. 8: 230]
XQuery	XML Query is a language for programming functionally specific to XML, <https://www.w3.org/TR/xquery-31/>. [Introd.: xix, xx, xxxvii; Ch. 1: 15, 31, 34, 36, 38, 38, 39, 40, 40, 42, 42, 42, 44, 46, 46, 46; Ch. 3: 97; Ch. 4: 128; Ch. 5: 158, 160;

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Ch. 6: 161; Ch. 7: 191, 193, 198, 200, 203, 203, 204, 204; Ch. 8: 220, 221, 222, 222, 222, 222, 224, 230]

XSL-FO XSL Formatting Objects is an XML markup language for formatting documents used to generate with a processor a PDF file, <https://en.wikipedia.org/wiki/XSL_Formatting_Objects>. [Ch. 8: 220, 221, 222, 224, 229, 229, 230, 230, 232]

XSLT eXtensible Stylesheet Language Transformations is the main programming language used to transform XML into other formats or other XML, <https://en.wikipedia.org/wiki/XSLT>. [Introd.: xix; Ch. 1: 28, 28, 28; Ch. 4: 123, 135; Ch. 5: 147, 154, 155; Ch. 7: 184, 184, 184, 185, 186, 189, 191, 193, 203; Ch. 8: 224, 227, 230, 231, 232]

Zotero It is a bibliography management software which allows the user to extract and import data from online resources to a bibliography and organize research materials. With a CSL style it can print the bibliography following those rules, <https://www.zotero.org/>. The bibliography of this book is managed with Zotero and both citations in the text and bibliographies are obtained from the Zotero API. [Introd.: xxxiii, l; Ch. 5: 159; Ch. 7: 194; Ch. 8: 219, 222, 222, 222, 223, 223, 223, 223, 223, 223, 223]

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Bibliography

Almas and Clérice 2018 | Almas, B. and T. Clérice 2018. 'Continuous Integration and Unit Testing of Digital Editions', *Digital Humanities Quarterly*, 11/4 (2018). Cited in Ch. 8: 228.

Almas et al. 2018 | Almas, B., T. Clérice, J. Robie, H. A. Cayless, Z. Fletcher, V. Jolivet, P. M. Liuzzo, M. Romanello, J. Tauber, and J. C. Witt 2018. *Distributed Text Services API Specification* (2018), <https://w3id.org/dts>. Cited in Introd.: xxxvii; Gloss.: 240.

Andrist 2014 | Andrist, P. 2014. 'Going Online is not Enough! Electronic Descriptions of Ancient Manuscripts, and the Needs of Manuscript Studies', in T. Andrews and C. Macé, eds, *Analysis of Ancient and Medieval Texts and Manuscripts: Digital Approaches*, Lectio, Studies in the Transmission of Texts & Ideas, 1 (Turnhout: Brepols, 2014), 309–334. Cited in Ch. 1: 17.

Andrist 2015a | Andrist, P. 2015. 'Syntactical Description of Manuscripts', in A. Bausi, P. G. Borbone, F. Briquel-Chatonnet, P. Buzi, J. Gippert, C. Macé, M. Maniaci, Z. Melissakis, L. E. Parodi, W. Witakowski, and E. Sokolinski, eds, *Comparative Oriental Manuscript Studies: An Introduction* (Hamburg: COMSt, 2015), 511–520. Cited in Ch. 1: 17, 17, 18, 26; Ch. 3: 90, 90.

Andrist 2015b | Andrist, P. 2015. 'The Physical description', in A. Bausi, P. G. Borbone, F. Briquel-Chatonnet, P. Buzi, J. Gippert, C. Macé, M. Maniaci, Z. Melissakis, L. E. Parodi, W. Witakowski, and E. Sokolinski, eds, *Comparative Oriental Manuscript Studies: An Introduction* (Hamburg: COMSt, 2015), 521–530. Cited in Ch. 1: 12, 28.

Andrist 2018 | Andrist, P. 2018. 'Toward a definition of paratexts and paratextuality: The case of ancient Greek manuscripts', in M. Maniaci and L. I. Lied, eds, *Bible as NotepadTracing Annotations and Annotation Practices in Late Antique and Medieval Biblical Manuscripts* (Berlin–Boston, MA: De Gruyter, 2018), 130–150 (DOI: 10.1515/9783110603477-010). Cited in Ch. 3: 79, 93.

Andrist et al. 2013 | Andrist, P., P. Canart, and M. Maniaci 2013. *La syntaxe du codex: Essai de codicologie structurale*, Bibliologia, 34 (Turnhout: Brepols, 2013). Cited in Ch. 1: 17, 24; Ch. 2: 50; Ch. 3: 79; Ch. 7: 211, 214.

Arnold and Müller 2017 | Arnold, E. and S. Müller 2017. 'Wie permanent sind Permalinks?', *Informationspraxis*, 3/1 (2017) (DOI: 10.11588/ip.2016.2.33483). Cited in Introd.: xxxvi.

Avanzini et al. 2014 | Avanzini, A., A. De Santis, D. Marotta, and I. Rossi 2014. 'Between Harmonization and Peculiarities of Scientific Domains: Digitizing the epigraphic heritage of pre-Islamic Arabia in the project DASI', in S. Orlandi, R. Santucci, V. Casarosa, and P. M. Liuzzo, eds, *Information Technologies for Epigraphy and Cultural Heritage: Proceedings of the First EAGLE International Conference*, Convegni, 26, Studi umanistici, Serie Antichistica, (Roma: Sapienza Università Editrice, 2014), 69–93. Cited in Introd.: xxvi; Ch. 2: 50.

Bagnall and Heath 2018 | Bagnall, R. S. and S. Heath 2018. 'Roman Studies and Digital Resources', *The Journal of Roman Studies*, 108 (2018), 171–189 (DOI: 10.1017/S0075435818000874). Cited in Gloss.: 234.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Bausi 2007 | Bausi, A. 2007. 'La catalogazione come base della ricerca. Il caso dell'Etiopia', in B. Cenni and C. M. F. Lalli, eds, *Zenit e Nadir II. I manoscritti dell'area del Mediterraneo: la catalogazione come base della ricerca. Atti del Seminario internazionale. Montepulciano, 6–8 luglio 2007*, Medieval Writing, Settimane poliziane di studi superiori sulla cultura scritta in età medievale e moderna (Montepulciano: Thesan&Turan, 2007), 87–108. Cited in Ch. 1: 2, 3, 16.

Bausi 2010 | Bausi, A. 2010. 'A Case for Multiple Text Manuscripts being "Corpus-Organizers"', *Manuscript Cultures Newsletter*, 3 (2010), 34–36. Cited in Ch. 3: 80, 89, 89.

Bausi 2016a | Bausi, A., 2016. *150 Years after Dillmann's* Lexicon*: Perspectives and Challenges of Gəʿəz Studies*, ed. A. Bausi, Supplement to Aethiopica, 5 (Wiesbaden: Harrassowitz Verlag, 2016). Cited in Ch. 5: 144.

Bausi 2016b | Bausi, A. 2016. 'Introduction. 150 Years After Dillmann's *Lexicon*', in A. Bausi, ed., *150 Years after Dillmann's* Lexicon*: Perspectives and Challenges of Gəʿəz Studies*, Supplement to Aethiopica, 5 (Wiesbaden: Harrassowitz Verlag, 2016), 3–10. Cited in Ch. 5: 144.

Bausi 2017 | Bausi, A. 2017. 'Il *Gadla ʾAzqir*', *Adamantius*, 23 (2017), 341–380. Cited in Gloss.: 235.

Bausi 2018 | Bausi, A. 2018. 'The recently published Ethiopic Inscriptions of King Ḥafilā (ΑΦΙΛΑΣ): a few Remarks', *Bibliotheca Orientalis*, 75/3–4 (2018), 285–295. Cited in Ch. 2: 51.

Bausi and Camplani 2016 | Bausi, A. and A. Camplani 2016. 'The History of the Episcopate of Alexandria (HEpA): Editio minor of the fragments preserved in the Aksumite Collection and in the Codex Veronensis LX (58)', *Adamantius*, 22 (2016), 249–302. Cited in Ch. 8: 225.

Bausi and Gori 2006 | Bausi, A. and A. Gori 2006. *Tradizioni orientali del «Martirio di Areta»: La prima recensione araba e la versione etiopica: edizione critica e traduzione*, Quaderni di Semitistica, 27 (Firenze: Dipartimento di Linguistica, Università di Firenze, 2006). Cited in Ch. 4: 109.

Bausi and Liuzzo 2018 | Bausi, A. and P. M. Liuzzo 2018. 'Inscriptions from Ethiopia', in A. De Santis and I. Rossi, eds, *Crossing Experiences in Digital Epigraphy: from Practice to Discipline* (Warsaw–Berlin: De Gruyter Open publisher, 2018), 84–92. Cited in Ch. 2: 50.

Beckett and Broekstra 2013 | Beckett, D. and J. Broekstra 2013. *SPARQL Query Results XML Format*, 2nd edn (2013) <https://www.w3.org/TR/rdf-sparql-XMLres/>, accessed on 28 March 2018. Cited in Ch. 7: 204.

Berhanu Beyene et al. 2000 | Berhanu Beyene, M. Kudlek, and O. Kummer 2000. 'Notes on Encoding Ethiopic for LATEX', *Aethiopica*, 3 (2000), 132–152 (DOI: 10.15460/aethiopica.3.1.575). Cited in Ch. 8: 219.

Bernand 2000 | Bernand, É., 2000. *Recueil des Inscriptions de l'Éthiopie des périodes pré-axoumite et axoumite*, ed. É. Bernand, III:*Traductions et commentaires*, A: *Les inscriptions grecques* (Paris: Diffusion de Boccard, 2000). Cited in Ch. 4: 134.

Bernand et al. 1991 | Bernand, É., A. J. Drewes, and R. Schneider 1991. *Recueil des inscriptions de l'Éthiopie des périodes pré-axoumite et axoumite, Introduction de Fr. Anfray*, I: *Les documents* (Paris: Diffusion de Boccard, 1991). Cited in Ch. 2: 50; Ch. 4: 134.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Beylot and Rodinson 1995 | Beylot, R. and M. Rodinson 1995. *Répertoire des bibliothèques et des catalogues de manuscrits éthiopiens*, Documents, études et répertoires publiés par l'Institut de Recherches et d'Histoire des Textes, 46 (Paris: Centre national de la recherche scientifique, Turnhout: Brepols, 1995). Cited in Introd.: xxx.

Blackwell and Smith 2014 | Blackwell, C. and N. Smith 2014. *The CITE architecture* (2014), <http://www.homermultitext.org/hmt-docs/cite/>, accessed on 26 November 2018. Cited in Ch. 8: 228.

Bleier et al. 2018 | Bleier, R., H. W. Klug, T. L. Andrews, J. van Zundert, W. Dillen, S. McGarry, G. Ferraro, A.-M. Sichani, S. Dumont, C. Di Pietro, R. Rosselli Del Turco, J. Schäuble, H. W. Gabler, E. Bleeker, A. Kelly, J. C. Witt, H. A. Cayless, F. Caria, B. Mathiak, and E. Leblanc 2018. *Digital Scholarly Editions as Interfaces*, eds R. Bleier, M. Bürgermeister, H. W. Klug, F. Neuber, and G. Schneider, Schriften des Instituts für Dokumentologie und Editorik, 12 (Norderstedt: Books on Demand, 2018). Cited in Ch. 8: 224.

Bodard and Stoyanova 2016 | Bodard, G. and S. Stoyanova 2016. 'Epigraphers and Encoders: Strategies for Teaching and Learning Digital Epigraphy', in G. Bodard and M. Romanello, eds, *Digital Classics Outside the Echo-Chamber. Teaching, Knowledge Exchange & Public Engagement* (London: Ubiquity Press, 2016), 51–68 (DOI: 10.5334/bat). Cited in Introd.: xli.

Bodard forthcoming | Bodard, G. forthcoming. 'Linked Open Data for Ancient Names and People', in S. Bond, P. Dilley, and R. Horne, eds, *Linked Ancient World Data: An Open Access Cookbook*, ISAW Papers (New York: forthcoming). Cited in Ch. 7: 200.

Bolle et al. 2017 | Bolle, K., C. Machado, and C. Witschel 2017. 'Introduction: Defining the Field–The Epigraphic Cultures of Late Antiquity', in K. Bolle, C. Machado, and C. Witschel, eds, *The Epigraphic Cultures of Late Antiquity*, Heidelberger Althistorische Beiträge, 60 (Stuttgart: Franz Steiner Verlag, 2017), 15–30. Cited in Ch. 2: 47, 54, 55, 74.

Bulakh 2016 | Bulakh, M. 2016. 'Some problems of transcribing Geez', in A. Bausi, ed., *150 Years after Dillmann's* Lexicon*: Perspectives and Challenges of Gəʿəz Studies*, Supplement to Aethiopica, 5 (Wiesbaden: Harrassowitz Verlag, 2016), 103–137. Cited in Ch. 6: 165.

Burghart 2016 | Burghart, M. 2016. 'The TEI Critical Apparatus Toolbox: Empowering Textual Scholars through Display, Control, and Comparison Features', *Journal of the Text Encoding Initiative*, 10 (2016), DOI: 10.4000/jtei.1520. Cited in Ch. 8: 231.

Burnard 2019 | Burnard, L. 2019. 'What is TEI Conformance, and Why Should You Care?', *Journal of the Text Encoding Initiative*, 12 (2019) (DOI: 10.4000/jtei.1777). Cited in Introd.: xxxv.

Burstein 1989 | Burstein, S. M., 1989. *Agatharchides of Cnidus. On the Erythreaan Sea*, ed., tr. S. M. Burstein (London: Hakluyt Society, 1989). Cited in Ch. 4: 116, 116, 133.

Burstein 1996 | Burstein, S. M. 1996. 'Ivory and Ptolemaic Exploration of the Red Sea. The Missing Factor', *Topoi*, 6/2 (1996), 799–807 (DOI: 10.3406/topoi.1996.1696). Cited in Ch. 4: 116.

Burstein 2000 | Burstein, S. M. 2000. 'Exploration and Ethnography in Ptolemaic Egypt', *The Ancient World*, 31 (2000), 31–37. Cited in Ch. 4: 116.

Burstein 2002 | Burstein, S. M. 2002. 'Kush, Axum and the Ancient Indian Ocean Trade', in T. A. Bács, ed., *A Tribute to Excellence: Studies Offered in Honor of* Ernö Gaál,

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Ulrich Luft, *and* László Török, Studia Aegyptiaca, 17 (Budapest: University, 2002), 127–137. Cited in Ch. 4: 116.

Burstein 2016a | Burstein, S. M. 2016. 'Aithiopia: Appendix (673)', in I. Worthington, E. M. Carawan, K. Dowden, J. Engels, A. Erskine, R. L. Fowler, and N. Jones, eds, *Brill's New Jacoby* (2016), DOI: 10.1163/1873-5363_bnj_a673. Cited in Ch. 4: 111, 111, 112, 113, 115, 117, 121.

Burstein 2016b | Burstein, S. M. 2016. 'Agatharchides of Knidos (86)', in I. Worthington, E. M. Carawan, K. Dowden, J. Engels, A. Erskine, R. L. Fowler, and N. Jones, eds, *Brill's New Jacoby* (2016), DOI: 10.1163/1873-5363_bnj_a86. Cited in Ch. 4: 116.

Busch and Chandna 2017 | Busch, H. and S. Chandna 2017. 'eCodicology: The Computer and the Mediaeval Library', in H. Busch, F. Fischer, and P. Sahle, eds, *Kodikologie & Paläographie im digitalen Zeitalter 4. Codicology & Palaeography in the Digital Age 4*, Schriften des Instituts für Dokumentologie und Editorik, 11 (Norderstedt: Books on Demand, 2017), 3–23. Cited in Ch. 1: 22.

Buzi et al. 2018 | Buzi, P., J. Bogdani, and F. Berno 2018. 'The "PAThs" Project: an Effort to Represent the Physical Dimension of Coptic Literary Production (Third–Eleventh centuries)', in A. Bausi, P. Buzi, P. M. Liuzzo, and E. Sokolinski, eds, *Proceedings of the Workshop* Linking Manuscripts from the Coptic, Ethiopian, and Syriac Domain: Present and Future Synergy Strategies, *Hamburg, 23 and 24 February 2018* (Hamburg: COMSt, 2018 = *COMSt Bulletin*, 4/1 (2018)), 39–58. Cited in Ch. 4: 107; Ch. 7: 208.

Calvanese et al. 2016 | Calvanese, D., P. Liuzzo, A. Mosca, J. Remesal, M. Rezk, and G. Rull 2016. 'Ontology-based data integration in EPNet: Production and distribution of food during the Roman Empire', *Engineering Applications of Artificial Intelligence*, 51 (2016 = *Mining the Humanities: Technologies and Applications*), 212–229 (DOI: 10.1016/j.engappai.2016.01.005). Cited in Ch. 2: 73; Ch. 7: 195.

Calvino 1995 | Calvino, I. 1995. *Saggi (1945–1985)*,II, ed. M. Barenghi, I Meridiani (Milano: Mondadori Bruno, 1995). Cited in Gloss.: 236.

Calvino 2016 | Calvino, I. 2016. *Lezioni americane: Sei proposte per il prossimo millennio* (Milano: Mondadori, 2016). Cited in Gloss.: 237.

Carlson and Michelson forthcoming | Carlson, T. A. and D. A. Michelson forthcoming. *The Syriac Gazetteer*, *Syriaca.org: The Syriac Reference Portal* (forthcoming), <http://syriaca.org/geo>, accessed on 19 May 2019. Cited in Ch. 4: 128.

Casson 1993 | Casson, L. 1993. 'Ptolemy II and the Hunting of African Elephants', *Transactions of the American Philological Association (1974–)*, 123 (1993), 247–260 (DOI: 10.2307/284331). Cited in Ch. 4: 116.

Cayless 2015 | Cayless, H. 2015.*Integrating Digital Epigraphies (IDEs)* (2015), <http://www.digitalclassicist.org/wip/wip2015-07hc.html>, accessed on 19 October 2018. Cited in Ch. 2: 62.

Cayless 2018 | Cayless, H. A. 2018. 'Critical Editions and the Data Model as Interface', in R. Bleier, M. Bürgermeister, H. W. Klug, F. Neuber, and G. Schneider, eds, *Digital Scholarly Editions as Interfaces*, Schriften des Instituts für Dokumentologie und Editorik, 12 (Norderstedt: Books on Demand, 2018), 249–263. Cited in Ch. 8: 222, 227.

Cayless and Viglianti 2018 | Cayless, H. and R. Viglianti 2018. 'CETEIcean: TEI in the Browser', in *Proceedings of Balisage: The Markup Conference 2018*, Balisage Series

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

on Markup Technologies, 21, (2018), DOI: 10.4242/balisagevol21.cayless01. Cited in Ch. 8: 227.

Cerulli 1968 | Cerulli, E. 1968. *La letteratura etiopica. Con un saggio sull'Oriente Cristiano: L'oriente cristiano nell'unità delle sue tradizioni*, Le letterature del mondo, 30, 3rd edn (Firenze: Sansoni, Milano: Accademia, 1968). Cited in Ch. 3: 101.

Chiarcos and Gracia forthcoming | Chiarcos, C. and J. Gracia forthcoming. 'Applications in Digital Humanities', in P. Cimiano, C. Chiarcos, J. P. McCrae, and J. Gracia, eds, *Linguistic Linked Data. Representation, Generation and Applications* (Springer Nature, forthcoming). Cited in Ch. 7: 186, 187, 202.

Ciotti and Tomasi 2016 | Ciotti, F. and F. Tomasi 2016. 'Formal Ontologies, Linked Data, and TEI Semantics', *Journal of the Text Encoding Initiative*, 9 (2016) (DOI: 10.4000/jtei.1480). Cited in Ch. 7: 186.

Ciula et al. 2008 | Ciula, A., P. Spence, and J. M. Vieira 2008. 'Expressing complex associations in medieval historical documents: the Henry III Fine Rolls Project', *Literary and Linguistic Computing*, 23/3 (2008), 311–325. Cited in Ch. 7: 183.

Clérice et al. 2017 | Clérice, T., M. Munson, and B. Almas 2017. 'Capitains/Capitains.Github.Io: 2.0.0', DOI: 10.5281/zenodo.570516. Cited in Ch. 8: 228.

Conti Rossini 1928 | Conti Rossini, C. 1928. *Storia d'Etiopia. Parte prima: Dalle origini all'avvento della dinastia salomonide*, Africa Italiana, 3 (Milano: Officina d'Arte Grafica A. Lucini & C., 1928). Cited in Ch. 4: 141.

Crawford 1958 | Crawford, O. G. S. 1958. *Ethiopian Itineraries circa 1400–1524: Including those Collected by Alessandro Zorzi at Venice in the Years 1519–24*, ed. O. G. S. Crawford, (Cambridge: University Press, 1958). Cited in Ch. 4: 140.

Cummings 2018 | Cummings, J. 2018. 'A world of difference: Myths and misconceptions about the TEI', *Digital Scholarship in the Humanities* (2018), DOI: 10.1093/llc/fqy071. Cited in Introd.: xxxv.

Da Maggiora 1953 | Da Maggiora, G. 1953. *Vocabolario etiopico–italiano–latino ad uso dei principianti* (Asmara: Scuola tipografica francescana, 1953). Cited in Ch. 5: 155.

Dee et al. 2016 | Dee, S., M. Foradi, and F. Šarić 2016. 'Learning by Doing: Learning to Implement the TEI Guidelines Through Digital Classics Publication', in G. Bodard and M. Romanello, eds, *Digital Classics Outside the Echo-Chamber. Teaching, Knowledge Exchange & Public Engagement* (London: Ubiquity Press, 2016), 15–32 (DOI: 10.5334/bat). Cited in Introd.: xli.

Desanges 1978 | Desanges, J. 1978. *Recherches sur l'activité des Méditerranéens aux confins de l'Afrique (VIe siècle avant J.-C.–IVe siècle après J.-C.)*, Collection de l'École Française de Rome, 38 (Rome: École Française de Rome, 1978). Cited in Ch. 4: 133.

Di Pietro and Rosselli Del Turco 2018 | Di Pietro, C. and R. Rosselli Del Turco 2018. 'Between Innovation and Conservation: The Narrow Path of User Interface Design for Digital Scholarly Editions', in R. Bleier, M. Bürgermeister, H. W. Klug, F. Neuber, and G. Schneider, eds, *Digital Scholarly Editions as Interfaces*, Schriften des Instituts für Dokumentologie und Editorik, 12 (Norderstedt: Books on Demand, 2018), 133–163. Cited in Ch. 8: 227.

Di Segni 2017 | Di Segni, L. 2017. 'Late Antique Inscriptions in the provinces of *Palestina* and *Arabia*: Realities and Change', in K. Bolle, C. Machado, and C. Witschel, eds,

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

The Epigraphic Cultures of Late Antiquity, Heidelberger Althistorische Beiträge, 60 (Stuttgart: Franz Steiner Verlag, 2017), 287–320, 609–615 (plates). Cited in Ch. 2: 47.

Dillmann 1865 | Dillmann, C. F. A. 1865. *Lexicon linguae aethiopicae, Cum indice latino. Adiectum est vocabularium tigre dialecti septentrionalis compilatum a W. Munziger* (Lipsiae: T. O. Weigel, 1865). Cited in Ch. 5: 143.

Dillmann 1907 | Dillmann, C. F. A. 1907. *Ethiopic Grammar. Second Edition Enlarged and Improved (1899)*, ed. C. Bezold, tr. J. A. Crichton (London: Williams & Norgate, 1907). Cited in Ch. 6: 162.

Driscoll and Pierazzo 2016 | Driscoll, M. J. and E. Pierazzo 2016. *Digital Scholarly Editing: Theories and Practices*, eds M. J. Driscoll and E. Pierazzo (Cambridge: Open Book Publishers, 2016). Cited in Ch. 8: 224.

Dumont and Fechner 2014 | Dumont, S. and M. Fechner 2014. 'Bridging the Gap: Greater Usability for TEI encoding', Journal of the Text Encoding Initiative, Issue 8 (2014) (DOI: 10.4000/jtei.1242). Cited in Ch. 8: 228.

Dumont and Fechner 2019 | Dumont, S. and M. Fechner 2019. *ediarum – Arbeits- und Publikationsumgebung für digitale Editionsvorhaben* (2 April 2019), DOI: 10.5281/zenodo.2621062. Cited in Ch. 8: 228.

EAe | S. Uhlig, ed., *Encyclopaedia Aethiopica*, I: *A–C*; II: *D–Ha*; III: *He–N*; in cooperation with A. Bausi, ed., IV: *O–X*; A. Bausi in cooperation with S. Uhlig, ed., V: *Y–Z. Supplementa. Addenda et Corrigenda. Maps. Index* (Wiesbaden: Harrassowitz Verlag, 2003, 2005, 2007, 2010, 2014).

Edmond 2016 | Edmond, J. 2016. 'Collaboration and Infrastructure', in S. Schreibman, R. Siemens, and J. Unsworth, eds, *A New Companion to Digital Humanities*, Blackwell Companions to Literature and Culture, 93 (Malden, MA–Oxford–Chichester: Wiley-Blackwell, 2016), 54–65. Cited in Introd.: xxxviii; Ch. 8: 228, 232; Gloss.: 233.

Eide 2014 | Eide, Ø. 2014. 'Ontologies, Data Modeling, and TEI', *Journal of the Text Encoding Initiative*, 8 (2014), DOI: 10.4000/jtei.1191. Cited in Ch. 7: 186.

Eide and Ore 2006 | Eide, Ø. and C.-E. Ore 2006. 'TEI, CIDOC-CRM and a Possible Interface between the Two', *Digital Humanities*, (2006), 5–9. Cited in Ch. 7: 186.

Elliott and Gillies 2009 | Elliott, T. and S. Gillies 2009. 'Digital Geography and Classics', *Digital Humanities Quarterly*, 3/1 (2009), <http://www.digitalhumanities.org/dhq/vol/3/1/000031/000031.html>, accessed on 1 August 2018. Cited in Ch. 4: 106, 107, 109, 142.

Elliott and Gillies 2011 | Elliott, T. and S. Gillies 2011. *'Poster': Pleiades: an un-GIS for Ancient Geography Digital Humanities 2011: June 19–22* (2011), <http://dh2011abstracts.stanford.edu/xtf/view?docId=tei/ab-192.xml;query=;brand=default>, accessed on 1 August 2018. Cited in Ch. 4: 106.

Elliott et al. 2007 | Elliott, T., G. Bodard, E. Milonas, S. Stoyanova, C. Tupman, and S. Vanderbilt 2017. *EpiDoc Guidelines: Ancient documents in TEI XML* (2007) <http://www.stoa.org/epidoc/gl/latest/>, accessed on 19 May 2019. Cited in Introd.: xxxvii; Ch. 2: 50.

Ellwardt 2016 | Ellwardt, A. 2016. 'Beyond Dillmann's *Lexicon*. Towards digital lexicography: Lessons from Syriac', in A. Bausi, ed., *150 Years after Dillmann's* Lexicon*: Per-*

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

spectives and Challenges of Gəʿəz Studies, Supplement to Aethiopica, 5 (Wiesbaden: Harrassowitz Verlag, 2016), 185–199. Cited in Ch. 5: 145.

Ellwardt and Dickhut 2018 | Ellwardt, A. and W. Dickhut 2018. 'Die Quellen von August Dillmanns *Lexicon linguae Aethiopicae*: Anmerkungen zu den Prolegomena und den verwendeten Sigla', in A. Bausi, P. Buzi, P. M. Liuzzo, and E. Sokolinski, eds, *Proceedings of the Workshop* Linking Manuscripts from the Coptic, Ethiopian, and Syriac Domain: Present and Future Synergy Strategies, *Hamburg, 23 and 24 February 2018* (Hamburg: COMSt, 2018 = *COMSt Bulletin*, 4/1 (2018)), 79–88. Cited in Ch. 5: 156.

Evangelisti 2017 | Evangelisti, S. 2017. 'Scrittura Epigrafica, alcune Riflessioni', in S. Antolini, S. M. Marengo, and G. Paci, eds, *Colonie e Municipi nell'Era digitale: Documentazione Epigrafica per La Conoscenza delle Città antiche, Atti del Convegno di studi (Macerata, 10–12 Dicembre 2015)*, Ichnia 14, (Tivoli: Edizioni Tored, 2017), 163–178. Cited in Introd.: xlv; Ch. 2: 49, 72.

Fauvelle-Aymar 2009 | Fauvelle-Aymar, F.-X. 2009. 'Les inscriptions d'Adoulis (Erythrée): Fragments d'un royaume d'influence hellénistique et gréco-romaine sur la côte africaine de la mer Rouge', *Bulletin de l'Institut Francais d'Archéologie Orientale*, 109 (2009), 135–160. Cited in Ch. 4: 121, 125, 133, 134, 136.

Felicetti and Murano 2016 | Felicetti, A. and F. Murano 2016. *CRMepi and CRMtex Updates* (2016), <https://www.academia.edu/27889385/CRMepi_and_CRMtex_Updates>, accessed on 7 September 2017. Cited in Ch. 2: 71; Ch. 7: 187.

Felicetti and Murano 2017 | Felicetti, A. and F. Murano 2017. 'Scripta manent: a CIDOC CRM semiotic reading of ancient texts', *International Journal on Digital Libraries*, 18/4 (2017), DOI: 10.1007/s00799-016-0189-z. Cited in Ch. 2: 71, 72; Ch. 7: 187.

Felicetti et al. 2015a | Felicetti, A., F. Murano, P. Ronzino, and F. Niccolucci 2015. 'CIDOC CRM and Epigraphy: a Hermeneutic Challenge', in P. Ronzino, ed., *Extending, Mapping and Focusing the CRM 2015. Proceedings Workshop EMF-CRM2015, Poznań, Poland, September 17, 2015* (2015), 55–68. Cited in Ch. 2: 71.

Felicetti et al. 2015b | Felicetti, A., F. Murano, P. Ronzino, and F. Niccolucci 2015. *CIDOC CRM and Epigraphy: a Hermeneutic Challenge* (2015), <https://www.academia.edu/19633265/CIDOC_CRM_and_Epigraphy_a_Hermeneutic_Challenge>, accessed on 18 October 2018. Cited in Ch. 2: 72.

Feraudi-Gruénais and Grieshaber 2016 | Feraudi-Gruénais, F. and F. Grieshaber 2016. 'Digital Epigraphy am Scheideweg? / Digital Epigraphy at a crossroads?', presented at the *Nachnutzung und Nachnutzbarkeit der Forschung im Akademienprogramm Workshop der Nordrhein-Westfälischen Akademie der Wissenschaften und der Künste und der Union der deutschen Akademien der Wissenschaften AG „eHumanities"* (Düsseldorf: 2016), DOI: 10.11588/heidok.00022141. Cited in Ch. 2: 60.

Feraudi-Gruénais and Grieshaber 2018 | Feraudi-Gruénais, F. and F. Grieshaber 2018. 'Empfehlungen für eine offene kollaborative Plattform für die antike Epigraphik—epigraphy.info / Recommendation for an open collaborative platform for ancient epigraphy—epigraphy.info' (2018), DOI: 10.11588/heidok.00024674. Cited in Ch. 2: 60.

Feraudi-Gruénais et al. 2019 | Feraudi-Gruénais, F., J. Cowey, T. Gheldof, F. Grieshaber, A. Kurilić, and P. Liuzzo 2019. 'Report on the second *Epigraphy.info* workshop held in Zadar, December 14–16, 2018 (Department of History, University of Zadar)',

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

presented at the *Second Epigraphy.info workshop* (Heidelberg: 2019), DOI: 10.11588/heidok.00026330. Cited in Ch. 2: 60, 74.

Fiaccadori 2003 | Fiaccadori, G. 2003. 'Un'epigrafe greca aksumita (*RIÉth* 274)', in V. Ruggeri and L. Pieralli, eds, *Eukosmia. Studi miscellannei per il 75° di Vincenzo Poggi S.J.* (Soveria Mannelli: Rubbettino Editore, 2003), 243–255. Cited in Ch. 2: 57.

Fiaccadori 2004 | Fiaccadori, G. 2004. 'Sembrouthes 'Gran Re' (*DAE* IV 3 = *RIÉth* 275): Per la storia del primo ellenismo aksumita', *La Parola del Passato*, 59/2 (335) (2004), 103–157. Cited in Ch. 4: 117, 119.

Fiaccadori 2007 | Fiaccadori, G. 2007. 'Nuova iscrizione greca da Aksum', *La Parola del Passato*, 62/1 (352) (2007), 70–76. Cited in Ch. 2: 50, 57, 57.

Flanders and Jannidis 2016 | Flanders, J. and F. Jannidis 2016. 'Data Modeling', in S. Schreibman, R. Siemens, and J. Unsworth, eds, *A New Companion to Digital Humanities*, Blackwell Companions to Literature and Culture, 93 (Malden, MA–Oxford–Chichester: Wiley-Blackwell, 2016), 229–237. Cited in Introd.: xxxv; Ch. 2: 71.

Fowler 2000 | Fowler, R. L. H. 2000. 'P. Oxy. 4458: Poseidonios', *Zeitschrift für Papyrologie und Epigraphik*, 132 (2000), 133–142. Cited in Ch. 4: 116.

Gibson et al. 2016 | Gibson, N. P., D. A. Michelson, and D. L. Schwartz 2016. 'From manuscript catalogues to a handbook of Syriac literature: Modeling an infrastructure for Syriaca.org', *Journal of Data Mining & Digital Humanities* (2016 = Computer-Aided Processing of Intertextuality in Ancient Languages). Cited in Ch. 4: 112.

Gippert 2015a | Gippert, J. 2015. 'Digital approaches to oriental manuscript studies', in A. Bausi, P. G. Borbone, F. Briquel-Chatonnet, P. Buzi, J. Gippert, C. Macé, M. Maniaci, Z. Melissakis, L. E. Parodi, W. Witakowski, and E. Sokolinski, eds, *Comparative Oriental Manuscript Studies: An Introduction* (Hamburg: COMSt, 2015), 12–26. Cited in Ch. 1: 31.

Gippert 2015b | Gippert, J. 2015. 'Catalogues and cataloguing of oriental manuscripts in the digital age', in A. Bausi, P. G. Borbone, F. Briquel-Chatonnet, P. Buzi, J. Gippert, C. Macé, M. Maniaci, Z. Melissakis, L. E. Parodi, W. Witakowski, and E. Sokolinski, eds, *Comparative Oriental Manuscript Studies: An Introduction* (Hamburg: COMSt, 2015), 531–538. Cited in Ch. 1: 31, 31.

Graham et al. 2015 | Graham, S., I. Milligan, and S. Weingart 2015. *Exploring Big Historical Data: The Historian's Macroscope* (London: Imperial College Press, 2015). Cited in Introd.: xviii.

Grieshaber 2019 | Grieshaber, F. 2019. 'Epigraphic Database Heidelberg – Data Reuse Options', (2019), 1–16 (DOI: 10.11588/heidok.00026599). Cited in Ch. 2: 53.

Grossi 2016 | Grossi, M. 2016. 'ΕΓΡΑΨΕΝ ΔΕ ΚΑΙ ΤΙΤΛΟΝ Ο ΠΙΛΑΤΟΣ (GV 19,19). Verso una nuova definizione di iscrizione', *Zeitschrift für Papyrologie und Epigraphik*, 197 (2016), 85–95. Cited in Introd.: xliii.

Hammerschmidt 1987 | Hammerschmidt, E. 1987. *Studies in the Ethiopic Anaphoras*, Äthiopistische Forschungen, 25 (Stuttgart: Franz Steiner Verlag Wiesbaden GmbH, 1987). Cited in Ch. 3: 81.

Hoffmann 2013 | Hoffmann, C. 2013. *Die Handschrift éth. d'Abbadie 20. Text, Übersetzung, Kommentar*, Masterarbeit, Freie Universität Berlin (2013). Cited in Ch. 4: 108.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Hummel and Dickhut 2016 | Hummel, S. and W. Dickhut 2016. 'A part of speech tag set for Ancient Ethiopic', in A. Bausi, ed., *150 Years after Dillmann's* Lexicon*: Perspectives and Challenges of Gəʿəz Studies*, Supplement to Aethiopica, 5 (Wiesbaden: Harrassowitz Verlag, 2016), 17–29. Cited in Ch. 6: 162.

Hummel et al. 2018 | Hummel, S., V. Pisani, and C. Vertan 2018. 'Multi-level Digital Annotation of Ethiopic Texts', in A. Bausi, P. Buzi, P. M. Liuzzo, and E. Sokolinski, eds, *Proceedings of the Workshop* Linking Manuscripts from the Coptic, Ethiopian, and Syriac Domain: Present and Future Synergy Strategies, *Hamburg, 23 and 24 February 2018* (Hamburg: COMSt, 2018 = *COMSt Bulletin*, 4/1 (2018)), 97–105. Cited in Ch. 6: 165.

Huntingford 1980 | Huntingford, G. W. B. 1980. *The Periplus of the Erythraean Sea, by an unknown author. With some Extracts from Agatharkhidēs 'On the Erythraean Sea'*, ed., tr. G. W. B. Huntingford (London: The Hakluyt Society, 1980). Cited in Ch. 4: 133.

Huntingford 1989 | Huntingford, G. W. B. 1989. *The Historical Geography of Ethiopia from the First Century AD to 1704*, ed. R. Pankhurst, Fontes historiae Africanae, Series varia, 4 (Oxford: Oxford University Press, 1989). Cited in Ch. 4: 128, 134, 139.

Isaksen et al. 2014 | Isaksen, L., R. Simon, E. T. E. Barker, and P. de Soto Cañamares 2014. 'Pelagios and the Emerging Graph of Ancient World Data', in *WebSci '14: Proceedings of the 2014 ACM conference on Web science, Bloomington, Indiana, USA June 23–26, 2014* (New York, NY: ACM, 2014), 197–201. Cited in Ch. 7: 197.

Jama Musse Jama n.d. | Jama Musse Jama n.d. *Somali Corpus: state of the art, and tools for linguistic analysis* (n.d.), <https://www.academia.edu/26504727/Somali_Corpus_state_of_the_art_and_tools_for_linguistic_analysis>, accessed on 6 October 2018. Cited in Ch. 6: 161.

Kamal 1905 | Ahmed Bey Kamal 1905. *Catalogue général des antiquités égyptiennes du Musée du Caire, Nos 22001–22208 Stèles ptolémaïques et romaines*, I, Texte (Le Caire: Imprimerie de l'Institut français d'archéologie orientale, 1905). Cited in Ch. 4: 121.

Kay 2008 | Kay, M. 2008. *XSLT 2.0 and XPath 2.0 Programmer's Reference*, Wrox Programmer to Programmer, 4th edn (Indianapolis, IN: Wiley Publishing, Inc., 2008). Cited in Introd.: xix; Gloss.: 246.

Kessel et al. 2018 | Kessel, G., R. Arnzen, Y. Arzhanov, N. Bamballi, and S. Čéplö 2018. 'Coding and Encoding: Towards a New Approach to the Study of Syriac and Arabic Translations of Greek Scientific and Philosophical Texts', *COMSt Bulletin*, 4/2 (2018), 205–213. Cited in Ch. 8: 231.

Kinefe-Rigb Zelleke 1975 | Kinefe-Rigb Zelleke 1975. 'Bibliography of the Ethiopic Hagiographical Traditions', *Journal of Ethiopian Studies*, 13/2 (1975), 57–102. Cited in Ch. 3: 81.

Kollatz 2018 | Kollatz, T. 2018. 'EPIDAT—Research Platform for Jewish Epigraphy', in A. De Santis and I. Rossi, eds, *Crossing Experiences in Digital Epigraphy: from Practice to Discipline* (Warsaw–Berlin: De Gruyter Open publisher, 2018), 231–239. Cited in Ch. 2: 68, 69.

Kropp 1983 | Kropp, M. 1983. '«Grecs» et Européens: Note de lexicographie éthiopienne', *Semitica*, 33 (1983), 125–131. Cited in Ch. 4: 111.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Kropp 1994 | Kropp, M. 1994. 'From Manuscripts to the Computer: Ethiopic Studies in the last 150 years', in K. J. Cathcart, ed., *The Edward Hincks Bicentenary Lectures* (Dublin: Department of Near Eastern Languages, University College Dublin, 1994), 114–137. Cited in Introd.: xxii, xxxiv, xxxviii; Ch. 3: 79, 92, 101; Ch. 5: 143.

Kropp 2016 | Kropp, M. 2016. 'Sergew Hable Selassies Fragment eines Gəʿəz-Belegstellenlexikons und Abraham Johannes Drewes' Glossare zum *Recueil des inscriptions de l'Éthiopie*. Zwei unveröffentlichte Beiträge zur äthiopischen Lexikographie und deren Bewertung und Lehren für die heutige informationstechnisch aufgerüstete Äthiopistik', in A. Bausi, ed., *150 Years after Dillmann's* Lexicon*: Perspectives and Challenges of Gəʿəz Studies*, Supplement to Aethiopica, 5 (Wiesbaden: Harrassowitz Verlag, 2016), 201–217. Cited in Ch. 5: 143.

Liuzzo 2014 | Liuzzo, P. M. 2014. 'Mapping Epigraphic Databases to EpiDoc', in S. Orlandi, R. Santucci, V. Casarosa, and P. M. Liuzzo, eds, *Information Technologies for Epigraphy and Cultural Heritage: Proceedings of the First EAGLE International Conference*, Convegni, 26, Studi umanistici, Serie Antichistica (Roma: Sapienza Università Editrice, 2014), 187–199. Cited in Ch. 2: 49.

Liuzzo 2017 | Liuzzo, P. M. 2017. 'Encoding the Ethiopic Manuscript Tradition: Encoding and representation challenges of the project Beta maṣāḥəft: Manuscripts of Ethiopia and Eritrea', in *Proceedings of Balisage: The Markup Conference 2017*, Balisage Series on Markup Technologies, 19 (2017), DOI: 10.4242/balisagevol19.liuzzo01. Cited in Introd.: xxi, xxxvii.

Liuzzo 2018 | Liuzzo, P. M. 2018. 'EAGLE continued: IDEA. The International Digital Epigraphy Association', in A. De Santis and I. Rossi, eds, *Crossing Experiences in Digital Epigraphy: from Practice to Discipline* (Warsaw–Berlin: De Gruyter Open publisher, 2018), 216–230. Cited in Introd.: xliv; Ch. 2: 54.

Liuzzo et al. 2014 | Liuzzo, P. M., S. Evangelisti, and H. Verreth 2014. *Content harmonization guidelines, including GIS and terminologies* (2014), <https://www.eagle-network.eu/wp-content/uploads/2013/06/EAGLE_D2.2.1_Content-harmonisation-guidelines-including-GIS-and-terminologies.pdf>, accessed on 19 May 2019. Cited in Ch. 2: 53.

Liuzzo et al. 2015 | Liuzzo, P. M., D. Fasolini, and A. Rocco 2015.*Content harmonization guidelines, including GIS and terminologies—Second Release* (2015), <https://www.eagle-network.eu/wp-content/uploads/2013/06/EAGLE_D2.2.2_Content-harmonisation-guidelines-including-GIS-and-terminologies-Second-Release.pdf>, accessed on 19 May 2019. Cited in Ch. 2: 53.

Liuzzo et al. 2018 | Liuzzo, P., D. Reule, E. Sokolinski, Solomon Gebreyes, D. Elagina, D. Nosnistin, E. Dal Sasso, and J. Gnisci 2018. *Beta maṣāḥəft Guidelines* (2018), <http://betamasaheft.eu/Guidelines/>, accessed on 30 April 2018. Cited in Introd.: xxxvii, xl; Ch. 1: 4; Ch. 4: 120; Ch. 7: 201.

Liuzzo forthcoming a | Liuzzo, P. M. forthcoming. 'Linked Open Data based on *La syntaxe du codex* for Manuscripts in Beta maṣāḥəft', in S. Bond, P. Dilley, and R. Horne, eds, *Linked Ancient World Data: An Open Access Cookbook*, ISAW Papers (New York NY: forthcoming). Cited in Ch. 1: 17, 24; Ch. 2: 50; Ch. 3: 79, 90; Ch. 7: 211.

Liuzzo forthcoming b | Liuzzo, P. M. forthcoming. 'RIÉ 277: an inscription of the time of Ptolemy II?', *Aethiopica,* 22 (forthcoming). Cited in Ch. 4: 111.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Lorusso et al. 2015 | Lorusso, V., A. Bausi, D. Bondarev, A. Brita, C. Brockmann, G. Ciotti, M. Friedrich, H. Isaacson, J. Karolewski, J. B. Quenzer, R. Samsom, S. Valente, E. Wilden, and H. Wimmer 2015. 'Searching for a definition of "manuscript"', *CSMC—Occasional Papers*, 1 (2015), <http://www.manuscript-cultures.uni-hamburg.de/papers_e.html>. Cited in Introd.: xliii.

MacMullen 1982 | MacMullen, R. 1982. 'The Epigraphic Habit in the Roman Empire', *The American Journal of Philology*, 103/3 (1982), 233–246 (DOI: 10.2307/294470). Cited in Ch. 2: 47.

Manghi et al. 2015 | Manghi, P., A. Mannocci, M. A. Sicilia, J. Gomez Pantoja, J. Rubiro Fuentes, E. Rivero Ruiz, and F. Zoppi 2015. *EAGLE metadata model specification—Second Release* (2015), <https://www.eagle-network.eu/wp-content/uploads/2013/06/EAGLE_D3.1_EAGLE-metadata-model-specification_v1.1.pdf>, accessed on 19 May 2019. Cited in Ch. 2: 69.

Maniaci 2012 | Maniaci, M. 2012. 'Costruzione e gestione dello spazio scritto fra oriente e occidente: principi generali e soluzioni specifiche', in Fondazione Centro Italiano di Studi sull'Alto Medioevo, *Scrivere e Leggere nell'Alto Medioevo, Spoleto, 28 aprile–4 maggio 2011*, I, Settimane di Studio del Centro italiano di studi sull'alto medioevo, 59 (Spoleto: Fondazione Centro italiano di studi sull'alto medioevo, 2012), 473–514. Cited in Ch. 1: 9, 11.

Marrassini 2001 | Marrassini, P. 2001. 'Traduzioni e citazioni dal greco in copto ed etiopico', in S. Settis, ed., *I Greci. Storia cultura arte società*, III: *I Greci oltre la Grecia*, Grandi Opere (Torino: Giulio Einaudi Editore, 2001), 985–1008. Cited in Ch. 3: 101.

Marrassini 2014 | Marrassini, P. 2014. *Storia e leggenda dell'Etiopia tardoantica: Le iscrizioni reali aksumite*, ed. P. Marrassini, Testi del Vicino Oriente antico, 9, Letteratura etiopica, 1 (Brescia: Paideia Editrice, 2014). Cited in Ch. 2: 57, 60, 67; Ch. 4: 110, 111, 111, 112.

Masoumeh Seydi et al. 2017 | Masoumeh Seydi, M. Romanov, and C. Palladino 2017. 'Premodern Geographical Description: Data Retrieval and Identification', in C.B. Jones and R.S. Purves, *GIR'17: Proceedings of the 11th Workshop on Geographic Information Retrieval*, (New York, NY: ACM, 2017), 1–10 (DOI: 10.1145/3155902.3155911). Cited in Ch. 4: 108.

McCarty 2016 | McCarty, W. 2016. 'Becoming Interdisciplinary', in S. Schreibman, R. Siemens, and J. Unsworth, eds, *A New Companion to Digital Humanities*, Blackwell Companions to Literature and Culture, 93 (Malden, MA–Oxford–Chichester: Wiley-Blackwell, 2016), 69–83. Cited in Introd.: xxxviii.

McGann 2007 | McGann, J. 2007. 'Marking Texts of Many Dimensions', in S. Schreibman, R. Siemens, and J. Unsworth, eds, *A Companion to Digital Humanities*, Blackwell Companions to Literature and Culture, 26 (Malden, MA - Oxford - Carlton: Wiley-Blackwell, 2007), 198–217 (DOI: 10.1002/9780470999875.ch16). Cited in Ch. 3: 78.

Meadows and Gruber forthcoming | Meadows, A. and E. Gruber forthcoming. 'Numismatics and Linked Open Data', in S. Bond, P. Dilley, and R. Horne, eds, *Linked Ancient World Data: An Open Access Cookbook*, ISAW Papers (New York, NY: forthcoming). Cited in Ch. 7: 189.

Michelson 2016a | Michelson, D. A. 2016. 'Mixed Up by Time and Chance? Using Digital Methods to "Re-Orient" the Syriac Religious Literature of Late Antiquity', *The Journal of Religion, Media and Digital Culture*, 5/1 (2016), 136–182. Cited in Ch. 7: 195.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Michelson 2016b | Michelson, D. A. *Syriaca.org Developing Linked Open Dataset: Project News from Syriaca.org* (2016), <http://syriaca.org/blog/2016/04/4192016-syriaca-org-developing-linked-open-dataset/>, accessed on 3 April 2018. Cited in Ch. 7: 195.

Montfort 2016 | Montfort, N. 2016. 'Exploratory Programming in Digital Humanities Pedagogy and Research', in S. Schreibman, R. Siemens, and J. Unsworth, eds, *A New Companion to Digital Humanities*, Blackwell Companions to Literature and Culture, 93 (Malden, MA–Oxford–Chichester: Wiley-Blackwell, 2016), 98–109. Cited in Introd.: xviii, xli; Ch. 8: 219.

Moretti 2005 | Moretti, F. 2005. *Graphs, Maps, Trees: Abstract Models for a Literary History* (London–New York, NY: Verso, 2005). Cited in Ch. 3: 94; Ch. 7: 216.

Moretti 2013 | Moretti, F. 2013. *Distant Reading* (London–New York, NY: Verso, 2013). Cited in Ch. 3: 93, 96; Ch. 7: 182.

Munari 1978 | Munari, B. 1978. *Disegnare un albero* (Mantova: Corraini, 1978). Cited in Ch. 7: 182.

Munro-Hay 1997 | Munro-Hay, S. C. 1997. *Ethiopia and Alexandria: The Metropolitan Episcopacy of Ethiopia*, Bibliotheca nubica et aethiopica (Warszawa: Zaś Pan, 1997). Cited in Ch. 4: 110, 110.

Nebes 2017 | Nebes, N. 2017. 'The Inscriptions of the Aksumite King Ḥafil and their Reference to Ethio-Sabaean Sources', *Zeitschrift für Orient-Archäologie*, 10 (2017), 356–369. Cited in Ch. 2: 51.

Nosnitsin 2013 | Nosnitsin, D. 2013. *Churches and Monasteries of Təgray: A Survey of Manuscript Collections*, Supplement to Aethiopica, 1 (Wiesbaden: Harrassowitz Verlag, 2013). Cited in Gloss.: 235.

Nosnitsin 2015 | Nosnitsin, D. 2015. 'Pricking and Ruling in Ethiopic Manuscripts: an Aid for Dating?', *Comparative Oriental Manuscript Studies Bulletin*, 1/2 (2015), 94–108. Cited in Ch. 1: 13.

Ogereau 2018 | Ogereau, J. M. 2018. 'Review of K. Bolle, C. Machado, and C. Witschel, eds, *The Epigraphic Cultures of Late Antiquity*, Heidelberger Althistorische Beiträge, 60 (Stuttgart: Fraz Steiner Verlag 2017)', *Zeitschrift für Antikes Christentum*, 22/2 (2018), 342–346. Cited in Ch. 2: 74.

Oldman et al. 2016 | Oldman, D., M. Doerr, and S. Gradmann 2016. 'Zen and the Art of Linked Data: New Strategies for a Semantic Web of Humanist Knowledge', in S. Schreibman, R. Siemens, and J. Unsworth, eds, *A New Companion to Digital Humanities*, Blackwell Companions to Literature and Culture, 93 (Malden, MA–Oxford–Chichester: Wiley-Blackwell, 2016), 251–273. Cited in Ch. 2: 50; Ch. 7: 183, 183, 183, 191.

Ore and Eide 2009 | Ore, C.-E. and Ø. Eide 2009. 'TEI and cultural heritage ontologies: Exchange of information?', *Literary and Linguistic Computing*, 24/2 (2009), 161–172. Cited in Ch. 7: 186.

Orlandi et al. 2014 | Orlandi, S., L. M. C. Giberti, and R. Santucci 2014. 'EAGLE: Europeana Network of Ancient Greek and Latin Epigraphy. Making the Ancient Inscriptions Accessible', *Lexicon Philosophicum International Journal for the History of Texts and Ideas*, 2 (2014), 315–326. Cited in Introd.: xliii.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Orlandi 2013 | Orlandi, T. 2013. 'A Terminology for the Identification of Coptic Literary Documents', *Journal of Coptic Studies*, 15 (2013), 87–94. Cited in Ch. 3: 80, 83, 83; Ch. 7: 208.

Orlandi 2018 | Orlandi, T. 2018. 'The CMCL *Clavis Coptica*. On Producing a Standardized List of (Coptic) Works and Manuscripts', in A. Bausi, P. Buzi, P. M. Liuzzo, and E. Sokolinski, eds, *Proceedings of the Workshop* Linking Manuscripts from the Coptic, Ethiopian, and Syriac Domain: Present and Future Synergy Strategies, *Hamburg, 23 and 24 February 2018* (Hamburg: COMSt, 2018 = *COMSt Bulletin*, 4/1 (2018)), 107–114. Cited in Ch. 3: 78; Ch. 7: 208.

Palladino 2016 | Palladino, C. 2016. 'New Approaches to Ancient Spatial Models: Digital Humanities and Classical Geography', *Bulletin of the Institute of Classical Studies*, 59/2 (2016), 56–70 (DOI: 10.1111/j.2041-5370.2016.12038.x). Cited in Ch. 4: 106.

Panciera 2012 | Panciera, S. 2012. 'What Is an Inscription? Problems of Definition and Identity of an Historical Source', *Zeitschrift für Papyrologie und Epigraphik*, 183 (2012), 1–10. Cited in Introd.: xliii, xliii, xlv, xlv, xlvi, xlvi.

Peremans 1967 | Peremans, W. 1967. 'Diodore de Sicile et Agatharchide de Cnide', *Historia*, 16/4 (1967), 432–455. Cited in Ch. 4: 116.

Porter et al. 2017 | Porter, D., A. Campagnolo, and E. Connelly 2017. 'VisColl: A New Collation Tool for Manuscript Studies', in H. Busch, F. Fischer, and P. Sahle, eds, *Kodikologie & Paläographie im digitalen Zeitalter 4. Codicology & Palaeography in the Digital Age 4*, Schriften des Instituts für Dokumentologie und Editorik, 11 (Norderstedt: Books on Demand, 2017), 81–100. Cited in Ch. 1: 28, 28, 28, 28.

Prag 2002 | Prag, J. R. W. 2002. 'Epigraphy by numbers: Latin and the epigraphic culture in Sicily', in A. E. Cooley, ed., *Becoming Roman, Writing Latin? Literacy and Epigraphy in the Roman West*, Journal of Roman Archaeology, Supplementary Series, 48, International Roman Archaeology Conference Series (Portsmouth, RI: Journal of Roman Archaeology, 2002), 15–31. Cited in Ch. 2: 47, 54, 60.

Prag 2018 | Prag, J. R. W. 2018. 'The Epigraphy of Agrigento in Context', in V. Caminneci, M. C. Parello, M. S. Rizzo, and C. Soraci, eds, *Agrigento ellenistico-romana: Coscienza Identitaria e Margini di Autonomia, Atti della Giornata di studi (Agrigento, 30 Giugno 2016)*, Bibliotheca archaeologica, 49 (Bari: Edipuglia, 2018), 27–35. Cited in Ch. 2: 47.

Prag and Chartrand 2018 | Prag, J. R. W. and J. Chartrand 2018. 'I.Sicily: Building a Digital Corpus of the Inscriptions of Ancient Sicily', in A. De Santis and I. Rossi, eds, *Crossing Experiences in Digital Epigraphy: from Practice to Discipline* (Warsaw–Berlin: De Gruyter Open publisher, 2018), 240–252. Cited in Ch. 2: 49.

Presner and Shepard 2016 | Presner, T. and D. Shepard 2016. 'Mapping the Geospatial Turn', in S. Schreibman, R. Siemens, and J. Unsworth, eds, *A New Companion to Digital Humanities*, Blackwell Companions to Literature and Culture, 93 (Malden, MA–Oxford–Chichester: Wiley-Blackwell, 2016), 201–212. Cited in Ch. 4: 139, 141.

Procházka 2005 | Procházka, S. 2005. *Altäthiopische Studiengrammatik*, Orbis Biblicus et Orientalis, Subsidia Linguistica, 2 (Göttingen: Vandenhoeck & Ruprecht, 2005). Cited in Ch. 6: 162.

Rabinowitz forthcoming | Rabinowitz, A. forthcoming. 'Time for Linked Open Data', in S. Bond, P. Dilley, and R. Horne, eds, *Linked Ancient World Data: An Open Access*

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Cookbook, ISAW Papers (NewYork, NY: n.pub., forthcoming). Cited in Ch. 1: 9; Ch. 2: 51, 68.

Retter 2012 | Retter, A. 2012. 'RESTful XQuery: Standardized XQuery 3.0 Annotations for REST', in *XML Prague 2012: Conference Proceedings, University of Economics, Prague, Czech Republic, February 10–12, 2012* (2012), 91–123. Cited in Introd.: xxxvii; Ch. 7: 187; Gloss.: 245.

Reule 2018 | Reule, D. 2018. '*Beta maṣāḥəft*: Manuscripts of Ethiopia and Eritrea', in A. Bausi, P. Buzi, P. M. Liuzzo, and E. Sokolinski, eds, *Proceedings of the Workshop* Linking Manuscripts from the Coptic, Ethiopian, and Syriac Domain: Present and Future Synergy Strategies, *Hamburg, 23 and 24 February 2018* (Hamburg: COMSt, 2018 = *COMSt Bulletin*, 4/1 (2018)), 13–27. Cited in Introd.: xxi, xxxvii; Ch. 1: 4.

Rosselli Del Turco et al. 2014 | Rosselli Del Turco, R., G. Buomprisco, C. Di Pietro, J. Kenny, R. Masotti, and J. Pugliese 2014. 'Edition Visualization Technology: A Simple Tool to Visualize TEI-based Digital Editions', *Journal of the Text Encoding Initiative*, 8 (2014), DOI: 10.4000/jtei.1077. Cited in Ch. 8: 227.

Roueché and Lawrence 2014 | Roueché, C. and F. K. Lawrence 2014. 'Linked Data and Ancient Wisdom', *ISAW Papers*, 7 (2014), <http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/roueche-lawrence-lawrence/>, accessed on 1 August 2018. Cited in Ch. 3: 82.

Roueché and Sotinel 2017 | Roueché, C. and C. Sotinel 2017. 'Christian and Late Antique Epigraphies', in K. Bolle, C. Machado, and C. Witschel, eds, *The Epigraphic Cultures of Late Antiquity*, Heidelberger Althistorische Beiträge, 60 (Stuttgart: Franz Steiner Verlag, 2017), 503–514. Cited in Ch. 2: 74.

Ruzicka 2018 | Ruzicka, R. 2018. 'KALAM: A Word Analyzer for Sabaic', in A. De Santis and I. Rossi, eds, *Crossing Experiences in Digital Epigraphy: from Practice to Discipline* (Warsaw–Berlin: De Gruyter Open publisher, 2018), 133–140. Cited in Ch. 6: 161.

Sahle 2016 | Sahle, P. 2016. 'What is a Scholarly Digital Edition?', in M. J. Driscoll and E. Pierazzo, eds, *Digital Scholarly Editing: Theories and Practices* (Cambridge: Open Book Publishers, 2016), 19–40 (DOI: 10.11647/OBP.0095.02). Cited in Introd.: xlii; Ch. 3: 102.

Sauermann and Cyganiak 2008 | Sauermann, L. and R. Cyganiak 2008. *Cool URIs for the Semantic Web W3C Interest Group Note* (2008), <https://www.w3.org/TR/cooluris/>, accessed on 27 August 2018. Cited in Ch. 7: 188, 188.

Scarpa 2005 | Scarpa, D. 2005. *Italo Calvino* (Milano: Mondadori Bruno, 2005). Cited in Gloss.: 236.

Schneider 2004 | Schneider, P. 2004. *L'Éthiopie et l'Inde: interférences et confusions aux extrémités du monde antique (VIII siècle avant J.C.–VI siècle après J.C.)*, Collection de l'École française de Rome, 335 (Rome: École française de Rome, 2004). Cited in Ch. 4: 112.

Schreibman et al. 2016 | Schreibman, S., R. Siemens, and J. Unsworth, eds, 2016. *A New Companion to Digital Humanities*, Blackwell Companions to Literature and Culture, 93 (Malden, MA–Oxford–Chichester: Wiley-Blackwell, 2016). Cited in Introd.: xviii.

Sculley and Pasanek 2008 | Sculley, D. and B. M. Pasanek 2008. 'Meaning and mining: the impact of implicit assumptions in data mining for the humanities', *Literary and Linguistic Computing*, 23/4 (2008), 409–424 (DOI: 10.1093/llc/fqn019). Cited in Introd.: xvii, xix.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Seaborne 2013 | Seaborne, A. 2013. *SPARQL 1.1 Query Results JSON Format W3C Recommendation* (2013), <https://www.w3.org/TR/sparql11-results-json/>, accessed on 28 March 2018. Cited in Ch. 1: 34; Ch. 7: 204.

Sedley 2018 | Sedley, D. 2018. 'Plato's *Cratylus*', in *The Stanford Encyclopedia of Philosophy*, (Fall 2018 Edition), ed. Edward N. Zalta (Stanford University: Metaphysics Research Lab, 2018), <https://plato.stanford.edu/archives/fall2018/entries/plato-cratylus/>. Cited in Ch. 1: 1.

Sergew Hable Selassie 1972 | Sergew Hable Selassie 1972. *Ancient and Medieval Ethiopian History to 1270* (Addis Ababa: United Printers, 1972). Cited in Ch. 4: 112, 116.

Shiferaw Bekele 2018 | Shiferaw Bekele 2018. 'The Genesis of Ethiopian Nationalism in Late Antique and Early Medieval Ethiopia in Light of Recent Historical and Philological Research', in M. Lafkioui and V. Brugnatelli, eds, *Written Sources about Africa and Their Study. Le fonti scritte sull'Africa e i loro studi*, Africana Ambrosiana, 3 (Milano: Centro Ambrosiano, Biblioteca Ambrosiana–ITLSH, 2018), 3–18. Cited in Ch. 4: 110.

Siegel and Retter 2015 | Siegel, E. and A. Retter 2015. *eXist: A NoSQL Document Database and Application Platform* (Sebastopol, CA: O'Reilly, 2015). Cited in Introd.: xix, xxxv.

Simon et al. 2012 | Simon, R., E. T. E. Barker, and L. Isaksen 2012. 'Exploring Pelagios: a visual browser for geo-tagged datasets.', in E. Agirre, ed., *International Workshop on Supporting Users' Exploration of Digital Libraries* (Cyprus: n.pub., 2012). Cited in Ch. 4: 107.

Simon et al. 2016 | Simon, R., L. Isaksen, E. Barker, and P. de Soto Cañamares 2016. 'Peripleo: a Tool for Exploring Heterogeneous Data through the Dimensions of Space and Time', *The Code4Lib Journal*, 31 (2016), <https://journal.code4lib.org/articles/11144>, accessed on 15 August 2018. Cited in Ch. 4: 107.

Solomon Gebreyes 2019 | Solomon Gebreyes 2019. *Chronicle of King Gälawdewos (1540–1559)*, ed., tr. Solomon Gebreyes, Corpus Scriptorum Christianorum Orientalium, 667, 668, Scriptores Aethiopici, 116, 117 (Lovanii: In aedibus Peeters, 2019). Cited in Ch. 4: 109.

Solomon Gebreyes and Liuzzo 2018 | Solomon Gebreyes and P. M. Liuzzo 2018. 'Encoding and Annotation of Ancient Places in Ethiopia', in A. Bausi, P. Buzi, P. M. Liuzzo, and E. Sokolinski, eds, *Proceedings of the Workshop* Linking Manuscripts from the Coptic, Ethiopian, and Syriac Domain: Present and Future Synergy Strategies, *Hamburg, 23 and 24 February 2018* (Hamburg: COMSt, 2018 = *COMSt Bulletin*, 4/1 (2018)), 121–141. Cited in Introd.: xxi, xxxvii; Ch. 4: 107.

Speidel 2016 | Speidel, M. A. 2016. 'Die Throninschrift von Adulis und das Römische Reich am Roten Meer zu Beginn des dritten Jahrhunderts n. Chr.', *Zeitschrift für Papyrologie und Epigraphik*, 200 (2016), 287–300. Cited in Ch. 4: 137.

TEI Consortium et al. 2018 | TEI Consortium, C. M. Sperberg-McQueen, and L. Burnard 2018. *TEI P5: Guidelines for Electronic Text Encoding and Interchange, Version 3.3.0. Last updated on 31st January 2018, revision f4d8439* (n.p.: TEI Consortium, 2018). Cited in Introd.: xxxiv; Ch. 1: 28.

Tropper 2002 | Tropper, J. 2002. *Altäthiopisch: Grammatik des Geʿez mit Übungstexten und Glossar*, Elementa Linguarum Orientis, 2 (Münster: Ugarit-Verlag, 2002). Cited in Ch. 6: 162.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Tupman and Jordanous 2014 | Tupman, C. and A. Jordanous 2014. 'Sharing Ancient Wisdoms across the Semantic Web using TEI and Ontologies', in T. Andrews and C. Macé, eds, *Analysis of Ancient and Medieval Texts and Manuscripts: Digital Approaches*, Lectio, Studies in the Transmission of Texts & Ideas, 1 (Turnhout: Brepols, 2014), 213–228. Cited in Ch. 7: 208.

Turska et al. 2016 | Turska, M., J. Cummings, and S. Rahtz 2016. 'Challenging the Myth of Presentation in Digital Editions', *Journal of the Text Encoding Initiative*, 9 (2016), DOI: 10.4000/jtei.1453. Cited in Introd.: xxxiv; Ch. 8: 229.

Van Zundert 2016 | Van Zundert, J. 2016. 'Barely Beyond the Book?', in M. J. Driscoll and E. Pierazzo, eds, *Digital Scholarly Editing: Theories and Practices* (Cambdridge: Open Book Publishers, 2016), 83–106 (DOI: 10.11647/OBP.0095.05). Cited in Introd.: xvii, xxxviii; Ch. 7: 187, 194.

Vattuone 2006 | Vattuone, R. 2006. Historíe*: Lezioni introduttive al corso di Storia greca* (Bologna: Pàtron Editore, 2006). Cited in Gloss.: 234.

Vertan 2016 | Vertan, C. 2016. 'Bringing Gəʿəz into the digital era: computational tools for processing Classical Ethiopic', in A. Bausi, ed., *150 Years after Dillmann's* Lexicon*: Perspectives and Challenges of Gəʿəz Studies*, Supplement to Aethiopica, 5 (Wiesbaden: Harrassowitz Verlag, 2016), 31–41. Cited in Ch. 6: 162.

Villa 2018 | Villa, M. 2018. 'Encoding Strategies and the Ethiopic Literary Heritage: The *Physiologus* as a Case Study', in A. Bausi, P. Buzi, P. M. Liuzzo, and E. Sokolinski, eds, *Proceedings of the Workshop* Linking Manuscripts from the Coptic, Ethiopian, and Syriac Domain: Present and Future Synergy Strategies, *Hamburg, 23 and 24 February 2018* (Hamburg: COMSt, 2018 = *COMSt Bulletin*, 4/1 (2018)), 143–149. Cited in Introd.: xxi; Ch. 3: 83; Ch. 7: 208, 209.

Viscount Valentia 1809 | Viscount Valentia, G. A. 1809. *Voyages and Travels to India, Ceylon, the Red Sea, Abyssinia, and Egypt, in the Years 1802, 1803, 1804, 1805, and 1806*, III (London: Printed for William Miller by W. Bulmer and Co., 1809). Cited in Ch. 4: 134.

Walmsley 2015 | Walmsley, P. 2015. *XQuery: Search Across a Variety of XML Data*, 2nd edn (Sebastopol, CA: O'Reilly, 2015). Cited in Introd.: xix.

Weninger 2016 | Weninger, S. 2016. 'The use of Arabic in Gəʿəz lexicography: from Dillmann to Leslau and beyond', in A. Bausi, ed., *150 Years after Dillmann's* Lexicon*: Perspectives and Challenges of Gəʿəz Studies*, Supplement to Aethiopica, 5 (Wiesbaden: Harrassowitz Verlag, 2016), 219–231. Cited in Ch. 5: 147.

Wicentowski and Meier 2015 | Wicentowski, J. C. and W. Meier 2015. 'Publishing TEI documents with TEI Simple: A Case Study at the U.S. Department of States's Office of the Historian', in *Proceedings of Balisage: The Markup Conference 2015*, Balisage Series on Markup Technologies, 15 (2015), DOI: 10.4242/balisagevol15.wicentowski01. Cited in Ch. 8: 229.

Wilkinson et al. 2016 | Wilkinson, M. D., M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons 2016. 'The FAIR Guiding Principles for scientific data management and stewardship', *Scientific Data*, 3 (2016), DOI: 10.1038/sdata.2016.18. Cited in Introd.: xxxiv, xxxiv, xxxv, xxxv.

Wion 2018 | Wion, A. 2018. 'The XML-TEI Architecture of *Ethiopian Manuscript Archives*: Respecting the Integrity of Primary Sources and Asserting Editorial Choices', in A. Bausi, P. Buzi, P. M. Liuzzo, and E. Sokolinski, eds, *Proceedings of the Workshop* Linking Manuscripts from the Coptic, Ethiopian, and Syriac Domain: Present and Future Synergy Strategies, *Hamburg, 23 and 24 February 2018* (Hamburg: COMSt, 2018 = *COMSt Bulletin*, 4/1 (2018)), 33–38. Cited in Ch. 2: 67; Ch. 4: 109.

Witt 2018 | Witt, J. C. 2018. 'Digital Scholarly Editions and API Consuming Applications', in R. Bleier, M. Bürgermeister, H. W. Klug, F. Neuber, and G. Schneider, eds, *Digital Scholarly Editions as Interfaces*, Schriften des Instituts für Dokumentologie und Editorik, 12 (Norderstedt: Books on Demand, 2018), 219–247. Cited in Ch. 3: 82.

woddiscovery 2010 | woddiscovery 2010. *Data and the Web—a great many of choices Web of Data* (2010), <https://webofdata.wordpress.com/2010/03/01/data-and-the-web-choices/>, accessed on 25 August 2018. Cited in Ch. 7: 182.

Wolska-Conus 1968 | Wolska-Conus, W. 1968. *Cosmas Indicopleustès: Topographie Chrétienne*, I, ed. W. Wolska-Conus (Paris: Cerf, 1968). Cited in Ch. 4: 128, 134.

Yohannes Gebreselassie 2017 | Yohannes Gebreselassie 2017. 'L'alphabet éthiopien. Une origine discutée', *Dossiers d'Archéologie*, 379 (2017), 36–39. Cited in Ch. 2: 51.

Zaccaria 2018 | Zaccaria, M. 2018. 'Digital Sources for the History of the Horn of Africa', *Oxford Research Encyclopedia of African History* (2018), DOI: 10.1093/acrefore/9780190277734.013.341. Cited in Introd.: xxiv; Gloss.: 234.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Beta maṣāḥǝft Entities

ANLcr27 | Elagina, D., 'Rome, Biblioteca dell'Accademia Nazionale dei Lincei, ANL Conti Rossini 27—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥǝft* (Last Modified: 13.2.2019), <https://betamasaheft.eu/ANLcr27>, accessed on 29 August 2019.

BAVcerulli37 | Valieva, N., 'Vatican City, Biblioteca Apostolica Vaticana, Biblioteca Apostolica Vaticana Cerulli 37—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥǝft* (Last Modified: 15.2.2019), <https://betamasaheft.eu/BAVcerulli37>, accessed on 29 August 2019.

BAVet1 | Villa, M. and D. Reule, 'Vatican City, Biblioteca Apostolica Vaticana, Aeth. 1—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥǝft* (Last Modified: 13.9.2016), <https://betamasaheft.eu/BAVet1>, accessed on 29 August 2019.

BAVet24 | Villa, M. and D. Reule, 'Vatican City, Biblioteca Apostolica Vaticana, Aeth. 24—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥǝft* (Last Modified: 4.7.2018), <https://betamasaheft.eu/BAVet24>, accessed on 29 August 2019.

BDLaethe8 | Reule, D., 'Oxford, Bodleian Library, Bodleian Aeth. e. 8—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥǝft* (Last Modified: 13.6.2017), <https://betamasaheft.eu/BDLaethe8>, accessed on 29 August 2019.

BDLbruce86 | Reule, D., 'Oxford, Bodleian Library, Bodleian Bruce 86—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥǝft* (Last Modified: 29.6.2017), <https://betamasaheft.eu/BDLbruce86>, accessed on 29 August 2019.

BDLbruce93 | Reule, D., 'Oxford, Bodleian Library, Bodleian Bruce 93—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥǝft* (Last Modified: 26.7.2017), <https://betamasaheft.eu/BDLbruce93>, accessed on 29 August 2019.

BLorient650 | Reule, D., 'London, British Library, BL Oriental 650—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥǝft* (Last Modified: 3.8.2018), <https://betamasaheft.eu/BLorient650>, accessed on 29 August 2019.

BLorient718 | Valieva, N., 'London, British Library, BL Oriental 718—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥǝft* (Last Modified: 15.1.2019), <https://betamasaheft.eu/BLorient718>, accessed on 29 August 2019.

BNFabb152 | Wion, A. and P. M. Liuzzo, 'Paris, Bibliothèque nationale de France, BnF Éthiopien d'Abbadie 152—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Be-*

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

ta maṣāḥəft (Last Modified: 17.11.2016), <https://betamasaheft.eu/BNFabb152>, accessed on 29 August 2019.

BNFet117 | Reule, D., 'Paris, Bibliothèque nationale de France, BnF Éthiopien 117—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 3.12.2018), <https://betamasaheft.eu/BNFet117>, accessed on 29 August 2019.

BNFet2 | Reule, D. and Solomon Gebreyes, 'Paris, Bibliothèque nationale de France, BnF Éthiopien 2—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 6.3.2017), <https://betamasaheft.eu/BNFet2>, accessed on 29 August 2019.

BNFet32 | Reule, D. and J. Gnisci, 'Paris, Bibliothèque nationale de France, BnF Éthiopien 32—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 22.11.2017), <https://betamasaheft.eu/BNFet32>, accessed on 29 August 2019.

BSLet169 | Reule, D., 'Cambridge, Bible Society's Library, BFBS Mss 169—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 20.4.2017), <https://betamasaheft.eu/BSLet169>, accessed on 29 August 2019.

ESkae011 | Nosnitsin, D. and E. Sokolinski, 'wäräda Dägʿa Tämben, ṭabiya Sälam, qušät Kunale, Kunale ʾArbaʿtu ʾƎnsəsa, KAE-011—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 28.9.2017), <https://betamasaheft.eu/ESkae011>, accessed on 29 August 2019.

ESky001 | Brita, A. and P. M. Liuzzo, 'Ǝndärta, Koholo Yoḥannəs Dabra Betel, KY-001—stub', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 12.4.2019), <https://betamasaheft.eu/ESky001>, accessed on 29 August 2019.

GAet7 | Dal Sasso, E., 'Grottaferrata, Exarchic Greek Abbey of St. Mary of Grottaferrata, Crypt. Aet. 7—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 10.1.2018), <https://betamasaheft.eu/GAet7>, accessed on 29 August 2019.

IVorlov20 | Elagina, D., 'Saint Petersburg, Institut Vostočnyh Rukopisej Rossijskoj Akademii Nauk, IV Orlov20–draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 31.1.2019), <https://betamasaheft.eu/IVorlov20>, accessed on 29 August 2019.

IVsuhtelen255 | Nosnitsin, D., 'Saint Petersburg, Institut Vostočnyh Rukopisej Rossijskoj Akademii Nauk, IV Suhtelen 255—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 10.12.2018), <https://betamasaheft.eu/IVsuhtelen255>, accessed on 29 August 2019.

LIT1146Argano | Reule, D. and E. Sokolinski, 'ʾArgānona Māryām—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale*

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Forschungsumgebung / Beta maṣāḥəft (Last Modified: 10.2.2017), <https://betamasaheft.eu/LIT1146Argano>, accessed on 29 August 2019.

LIT1252Confes | Solomon Gebreyes and E. Sokolinski, 'Confessio Claudii—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 8.5.2018), <https://betamasaheft.eu/LIT1252Confes>, accessed on 29 August 2019.

LIT1349EpistlEusebius | Bausi, A., M. Villa, and E. Sokolinski, 'Epistle of Eusebius to Carpianus (generic record)—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 27.11.2018), <https://betamasaheft.eu/LIT1349EpistlEusebius>, accessed on 29 August 2019.

LIT1493Gadlas | Reule, D. and E. Sokolinski, 'Acts of Saints and Martyrs—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 9.3.2017), <https://betamasaheft.eu/LIT1493Gadlas>, accessed on 29 August 2019.

LIT1546Genesi | HaCohen, R., 'Genesis', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 5.7.2017), <https://betamasaheft.eu/LIT1546Genesi>, accessed on 29 August 2019.

LIT2170Peripl | Liuzzo, P. M., 'Periplus of the Erythraean Sea—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 27.6.2018), <https://betamasaheft.eu/LIT2170Peripl>, accessed on 29 August 2019.

LIT2505Weddas | Reule, D., M. Villa, and E. Sokolinski, 'Wəddāse ʾAmlāk—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 26.5.2017), <https://betamasaheft.eu/LIT2505Weddas>, accessed on 29 August 2019.

LIT3122Galaw | Solomon Gebreyes and E. Sokolinski, 'Chronicle of Galāwdewos—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 8.9.2017), <https://betamasaheft.eu/LIT3122Galaw>, accessed on 29 August 2019.

LIT3950ChronIyasu | Sokolinski, E., Solomon Gebreyes, and D. Reule, 'Chronicle of ʾIyāsu I—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 8.8.2017), <https://betamasaheft.eu/LIT3950ChronIyasu>, accessed on 29 August 2019.

LIT3951ChronSusenyos | Solomon Gebreyes, D. Reule, and E. Sokolinski, 'Chronicle of Susənyos–draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 8.8.2017), <http://betamasaheft.eu/works/LIT3951ChronSusenyos/main>, accessed on 29 August 2019.

LIT3952MashafaQ | Reule, D. and E. Sokolinski, 'Maṣḥafa qandil—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 29.6.2017), <https://betamasaheft.eu/LIT3952MashafaQ>, accessed on 29 August 2019.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

LIT4275ChronAmdS | Pisani, V., E. Sokolinski, and D. Reule, 'Tārika ʿĀmda Ṣəyon—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 2.8.2018), <https://betamasaheft.eu/LIT4275ChronAmdS>, accessed on 29 August 2019.

LIT4723TarikaWaldaAmid | Reule, D., 'Tārika Walda ʿAmid—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 25.2.2019), <https://betamasaheft.eu/LIT4723TarikaWaldaAmid>, accessed on 29 August 2019.

LIT4851greekRoyal | Bausi, A. and P. M. Liuzzo, 'Pseudo-trilingual royal inscriptions, RIÉ 270—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 27.11.2017), <https://betamasaheft.eu/LIT4851greekRoyal>, accessed on 29 August 2019.

LIT4866OntheB | Villa, M., 'Dərsān za-qəddus Yoḥannəs ʾafa warq baʾənta ṭəmqatu la-ʾəgziʾəna ʾIyasus Krəstos—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 20.12.2017), <https://betamasaheft.eu/LIT4866OntheB>, accessed on 29 August 2019.

LIT5012MonumentumAdulitanum1 | Liuzzo, P. M., 'Monumentum Adulitanum—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 19.7.2018), <https://betamasaheft.eu/LIT5012MonumentumAdulitanum1>, accessed on 29 August 2019.

LIT5019MonumentumAdulitanum2 | Liuzzo, P. M., 'Monumentum Adulitanum Inscription of unknown king—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 21.7.2018), <https://betamasaheft.eu/LIT5019MonumentumAdulitanum2>, accessed on 29 August 2019.

LIT5064HEpA | Bausi, A., A. Camplani, and P. M. Liuzzo, 'History of the Episcopate of Alexandria—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 27.9.2018), <https://betamasaheft.eu/LIT5064HEpA>, accessed on 29 August 2019.

LOC1261Adulis | Panini, F. and E. Sokolinski, 'ʾAdulis', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 30.6.2017), <https://betamasaheft.eu/LOC1261Adulis>, accessed on 29 August 2019.

LOC1310Aksum | Sokolinski, E., 'ʾAksum', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥəft* (Last Modified: 27.4.2016), <https://betamasaheft.eu/LOC1310Aksum>, accessed on 29 August 2019.

LOC3010Ethiop | Sokolinski, E., 'Ethiopia', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta*

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

maṣāḥǝft (Last Modified: 27.4.2016), <https://betamasaheft.eu/LOC3010Ethiop>, accessed on 29 August 2019.

LOC3149Gabaz | Sokolinski, E. and P. M. Liuzzo, 'Gabazā', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥǝft* (Last Modified: 19.7.2018), <https://betamasaheft.eu/LOC3149Gabaz>, accessed on 29 August 2019.

LOC3843HBS | Sokolinski, E., 'Ḥ̣BŠ', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥǝft* (Last Modified: 27.4.2016), <https://betamasaheft.eu/LOC3843HBS>, accessed on 29 August 2019.

LOC4560Marab | Solomon Gebreyes and E. Sokolinski, 'Marab', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥǝft* (Last Modified: 19.7.2018), <https://betamasaheft.eu/LOC4560Marab>, accessed on 29 August 2019.

LOC5017Nubia | Sokolinski, E., 'Nubia', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥǝft* (Last Modified: 28.4.2016), <https://betamasaheft.eu/LOC5017Nubia>, accessed on 29 August 2019.

LOC6850Takkaze | Solomon Gebreyes, 'Takkaze', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥǝft* (Last Modified: 28.6.2017), <https://betamasaheft.eu/LOC6850Takkaze>, accessed on 29 August 2019.

LOC7053Aithiopia | Liuzzo, P. M., 'Aithiopia', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥǝft* (Last Modified: 22.7.2018), <https://betamasaheft.eu/LOC7053Aithiopia>, accessed on 29 August 2019.

PRS10198Yaeqob | Sokolinski, E. and D. Reule, 'ʾAṣe Yāʿqob—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥǝft* (Last Modified: 8.1.2018), <https://betamasaheft.eu/PRS10198Yaeqob>, accessed on 29 August 2019.

PRS4327GabraMa | Sokolinski, E. and Solomon Gebreyes, '**ገብረ ፡ መስቀል ፡**', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥǝft* (Last Modified: 24.1.2017), https://betamasaheft.eu/PRS4327GabraMa>, accessed on 29 August 2019.

PRS6829Maryams | Solomon Gebreyes and E. Sokolinski, 'Māryām Śǝnā—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥǝft* (Last Modified: 15.5.2017), <https://betamasaheft.eu/PRS6829Maryams>, accessed on 29 August 2019.

RIE276 | Liuzzo, P. M., 'RIE 276—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Beta maṣāḥǝft* (Last Modified: 19.7.2018), <https://betamasaheft.eu/RIE276>, accessed on 29 August 2019.

RIE277 | Liuzzo, P. M., 'RIE 277—draft', in Alessandro Bausi, ed., *Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung / Be-*

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

ta maṣāḥəft (Last Modified: 19.7.2018), <http://betamasaheft.eu/manuscripts/RIE277/main>, accessed on 29 August 2019.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Pleiades Resources

29475 | Roaf, M., 'Arabia: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2018), <https://pleiades.stoa.org/places/29475>, accessed on 15 August 2018.

30224 | Roaf, M., 'Chelonophagoi: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2013), <https://pleiades.stoa.org/places/30224>, accessed on 15 August 2018.

334481 | Fentress, E. W. B., 'Aethiopia: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2013), <https://pleiades.stoa.org/places/334481>, accessed on 15 August 2018.

39274 | Potts, D. T., 'Aethiopia: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2012), <https://pleiades.stoa.org/places/39274>, accessed on 15 August 2018.

39281 | Potts, D. T., 'Alalaiou (islands): a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2016), <https://pleiades.stoa.org/places/39281>, accessed on 15 August 2018.

39290 | Potts, D. T., 'Arabicus Sinus/Erythr(ae)um/Rubrum Mare: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2017), <https://pleiades.stoa.org/places/39290>, accessed on 15 August 2018.

39295 | Potts, D. T., 'Arsinoe: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2012), <https://pleiades.stoa.org/places/39295>, accessed on 15 August 2018.

39298 | Potts, D. T., 'Astapous (river): a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2017), <https://pleiades.stoa.org/places/39298>, accessed on 15 August 2018.

39303 | Potts, D. T., 'A(u)xoume: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2018), <https://pleiades.stoa.org/places/39303>, accessed on 15 August 2018.

39305 | Potts, D. T., 'Aualites Sinus: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2013), <https://pleiades.stoa.org/places/39305>, accessed on 15 August 2018.

39335 | Potts, D. T., 'Deire: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2014), <https://pleiades.stoa.org/places/39335>, accessed on 15 August 2018.

39340 | Potts, D. T., 'Elephas Pr.: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2013), <https://pleiades.stoa.org/places/39340>, accessed on 15 August 2018.

39360 | Potts, D. T., 'Kinaidokolpitai: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2012), <https://pleiades.stoa.org/places/39360>, accessed on 15 August 2018.

39361 | Potts, D. T., 'Kinnamomophoros Chora: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2014), <https://pleiades.stoa.org/places/39361>, accessed on 15 August 2018.

39387 | Potts, D. T., 'Mosylon Emporion: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2012), <https://pleiades.stoa.org/places/39387>, accessed on 15 August 2018.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

39409 | Potts, D. T, 'Ptolemais Theron: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2013), <https://pleiades.stoa.org/places/39409>, accessed on 15 August 2018.

39417 | Potts, D. T., 'Saba: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2012), <https://pleiades.stoa.org/places/39417>, accessed on 15 August 2018.

39435 | Potts, D. T., 'Trogodytice: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2013), <https://pleiades.stoa.org/places/39435>, accessed on 15 August 2018.

40028 | Potts, D. T., 'Aithiopes Ichthyophagoi: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2012), <https://pleiades.stoa.org/places/40028>, accessed on 15 August 2018.

40184 | Potts, D. T., 'Kreophagoi: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2012), <https://pleiades.stoa.org/places/40184>, accessed on 15 August 2018.

40324 | Potts, D. T., 'Strythophagoi: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2012), <https://pleiades.stoa.org/places/40324>, accessed on 15 August 2018.

50004 | Erdosy, M. U., 'India: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2017), <https://pleiades.stoa.org/places/50004>, accessed on 15 August 2018.

727101 | Bernand, A., 'Clysma: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2017), <https://pleiades.stoa.org/places/727101>, accessed on 15 August 2018.

727124 | Bernand, A., 'Heroonpolis: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2018), <https://pleiades.stoa.org/places/727124>, accessed on 15 August 2018.

727172 | Bernand, A., 'Nilus (river): a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2016), <https://pleiades.stoa.org/places/727172>, accessed on 16 August 2018.

746701 | Graf, D. F., 'Aelaniticus Sinus: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2015), <https://pleiades.stoa.org/places/746701>, accessed on 15 August 2018.

746710 | Graf, D. F., 'Arabia Eudaemon: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2014), <https://pleiades.stoa.org/places/746710>, accessed on 15 August 2018.

766 | Downs, M. E., B. Z. Lund, R. Talbert, and M. J. McDaniel, 'Aegyptus (Roman imperial province): a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2019), <https://pleiades.stoa.org/places/766>, accessed on 23 August 2019.

767851 | Sidebotham, S. E., 'Philoteras: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2012), <https://pleiades.stoa.org/places/767851>, accessed on 15 August 2018.

767858 | Sidebotham, S. E., 'Scytala: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2012), <https://pleiades.stoa.org/places/767858>, accessed on 15 August 2018.

785986 | Wilfong, T., S. E. Sidebotham, and J. Keenan, 'Berenice/Berenike: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2018), <https://pleiades.s-toa.org/places/785986>, accessed on 15 August 2018.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

786010 | Wilfong, T., S. E. Sidebotham, and J. Keenan, 'Coptos: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2018), <https://pleiades.stoa.org/places/786010>, accessed on 15 August 2018.

786069 | Wilfong, T., S. E. Sidebotham, and J. Keenan, 'Myos Hormos?: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2018), <https://pleiades.stoa.org/places/786069>, accessed on 15 August 2018.

795842 | Burstein, S. M., 'Pentedactylos M.: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2012), <https://pleiades.stoa.org/places/795842>, accessed on 15 August 2018.

805593 | Millet, N. B., 'Astaboras (river): a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2016), <https://pleiades.stoa.org/places/805593>, accessed on 15 August 2018.

805630 | Millet, N. B., 'Meroe: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2018), <https://pleiades.stoa.org/places/805630>, accessed on 15 August 2018.

845 | Talbert, R., B. Z. Lund, M. E. Downs, and M. J. McDaniel, 'Atlanticus Oceanus: a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2016), <https://pleiades.stoa.org/places/845>, accessed on 15 August 2018.

981506 | Wilkes, J. J., S. T. Parker, R. S. Bagnall, W. V. Harris, A. S. Esmonde-Cleary, C. M. Wells, J. F. Drinkwater, R. C. Knapp, and S. Mitchell, 'Arabia (province): a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2019), <https://pleiades.stoa.org/places/981506>, accessed on 23 August 2019.

981517 | Wilkes, J. J., S. T. Parker, R. S. Bagnall, W. V. Harris, A. S. Esmonde-Cleary, C. M. Wells, J. F. Drinkwater, R. C. Knapp, and S. Mitchell, 'Iudaea (province): a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2019), <https://pleiades.stoa.org/places/981517>, accessed on 23 August 2019.

981517 | Wilkes, J. J., S. T. Parker, R. S. Bagnall, W. V. Harris, A. S. Esmonde-Cleary, C. M. Wells, J. F. Drinkwater, R. C. Knapp, and S. Mitchell, 'Creta et Cyrene (Roman province): a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2019), <https://pleiades.stoa.org/places/981517>, accessed on 23 August 2019.

981550 | Wilkes, J. J., S. T. Parker, R. S. Bagnall, W. V. Harris, A. S. Esmonde-Cleary, C. M. Wells, J. F. Drinkwater, R. C. Knapp, and S. Mitchell, 'Syria (Roman imperial province): a Pleiades place resource', in *Pleiades: A Gazetteer of Past Places* (2018), <https://pleiades.stoa.org/places/981550>, accessed on 23 August 2019.

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Cited Passages

Inscriptions

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Persons

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

Places

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1

ISBN Print: 978-3-44711291-8 - ISBN E-Book: 978-3-447-19913-1