Experiments in Digital Publishing: creating a digital compendium*

Matteo Romanello

Abstract

This chapter not only introduces the user and reader to the goals of the digitally provided index and the methods used for it. It also expands on the broader applicability of digital methods in view of electronic publishing, and to the problems involved. The chapter focuses on two aspects of the making of the *Structures of Epic Poetry* compendium, where digital tools played a central role: the creation of the *index locorum* and the development of a digital compendium to the printed volumes.

I. Introduction

The *index locorum* (or index of cited passages) is one of the essential tools for Classical scholars, as it allows them to find rapidly where a given text (or text passage) is discussed within a publication, without necessarily having to read it sequentially.[1] Yet, creating a traditional *index locorum* generally requires a substantial amount of mostly manual work, which is time-consuming and expensive to produce. Notwithstanding its costs, there are publications – like this copious compendium in three volumes – where the high number of references to ancient texts makes the creation of an *index locorum* virtually impracticable.

The editors of this compendium – Christiane Reitz and Simone Finkmann – were able to see not only that a digital publication nicely complements the printed volumes, but also that digital tools could do much more, such as help them produce an electronic index of the cited passages. They saw, in other words, that the extraction of cited passages from the publication chapters could help them in solving two problems at once: firstly, it would considerably speed up the process of producing the *index locorum* for the printed publication and, secondly, it would allow them to create a digital inventory of all passages cited, together with information about the poetic structure of Graeco-Roman epic from Homer to Neolatin Epic, which they are exploring in their research.

* This is a pre-print version; for the published one please refer to https://doi.org/10.1515/9783110492590-074.

1 Talking about the creation of an another genre of indices, the *index verborum*, Oldfather (1937, p. 1) notes "No competent scholar needs to be convinced of the utility of indices".

The making of the *Structures of Epic Poetry* compendium has been a digital publishing experiment in two ways. First, it is the first time, to the best of our knowledge, that an *index locourm* is created semi-automatically, by using a citation mining tool to extract references before alternatively correcting the remaining errors by hand and retraining and thus continually improving the programme's accuracy. This tool was originally created to recover cited passages from existing publications, but it can be integrated as well into the publishing workflow, as described in this chapter. The second experiment was the creation of a digital companion that allows readers to access and explore the publication contents better, while, at the same time, it can be used much like a database of incorporated text passages on the structures of epic poetry.[2]

The rest of this chapter is organized as follows: in section 2 I situate these two experiments within the broader context of work that strives to move digital publishing beyond the paradigm of PDF documents. Section 3 is dedicated to the first experiment, and describes the workflow we devised to semi-automate the creation of the *index locorum* for this compendium. Section 4 addresses the issue of the digital publication, in terms of both user and machine interface. Finally, in section 5 I reflect on how the digital medium is changing, and how it will change the way in which we conceive and consume publications.

II. Beyond the PDF

The title of this section is a provocative reference to the current status of digital publishing, at least or rather especially in the area of *Altertumswissenschaft*, where most of what is published in a digital format still holds on to the PDF as a document paradigm. *Semantic publishing* is an attempt to overcome this very situation, by promoting the use of semantic technologies so as to make publication contents more reusable, more interconnected and interoperable, and more easily discoverable.[3]

Work in this area – both within and beyond the realm of *Altertumswissenschaft* – has focussed on and emphasized various aspects of publications, namely: a) reproducibility, b) explicitness and machine readability, c) data reusability and interconnectivity.

a) Reproducibility

Reproducibility of published research is a concern especially in the Scientific, Technical and Medical (STM) sector, where there exists a tight connection between publication, experiments, and underlying data. Publications in this area contain, more often than not, visualizations produced by running programmatic analysis on primary data. A novel publishing paradigm is being put forward in this area, which deems the

2 The digital companion is openly available at http://epibau.ub.uni-rostock.de/app/, while all source code is published at https://github.com/CitedLoci/EpiBau-Digital-Companion.

3 Shotton (2009).

reproducibility of results described in a research paper as a key feature of digital publications. Technical solutions, like the *executable paper* proposed by Sato *et al.* (2011), need to address a wide range of technical issues like supporting the collaborative work of scientists, running the required computation in the background, and enabling access to primary data as defined by the license and depending on user affiliation.

b) Explicitness and **machine readability**

Explicitness and machine readability were the main goals of applying semantic technologies to publications. Semantic enhancements to publications include the provision of interactive figures, the explicit encoding in a machine-readable format (i.e. RDF) of elements of interest such as bibliographic references, and the linking of technical terms used in the publication with specialized thesauri.[4] While the immediate advantages of such enhanced publications are readily understood, the limited uptake of these technologies is due to the substantial amount of time it takes authors to encode their publications semantically. Current research to overcome this issue seeks, on the one hand, to exploit Natural Language Processing (NLP) techniques to automate the semantic encoding of publication contents (e.g. REF) and, on the other hand, to leverage purely structural and compositional features of publications to derive their corresponding semantic classifications.[5]

c) **Data reusability and interconnectivity**

When it comes to publications, data reusability can only be achieved by uncoupling (i.e. keeping separate and distinguished) data and interfaces. If a digital publication is designed following this simple pattern, it becomes then possible to reuse the data independently of any user interface and, at the same time, visualize the same data in a multiplicity of specialized user interfaces. From a technical point of view, an effective way of uncoupling data from interfaces is to expose the data to be displayed in an interface by means of a machine-friendly interface or Application Programming Interface (API). McGuire (2013), for instance, has argued that the job of "good publishers of the future" is to provide APIs for their publications and suggests that an API is the natural translation of a printed index in a digital environment. Witt (2018) has recently made a similar claim for a different type of texts, i.e. digital editions. He argues that in the current development of digital scholarly editions too much effort is wasted in creating editions whose data and user interface cannot exist separately from one another.

A notable example of the potentials opened up when publications are designed with a focus on APIs is provided by *A Homer Commentary in Progress*, a project of the Center for Hellenic Studies.[6] All the commentary data are exposed by means of an API and a shared set of unique identifiers – the so-called CTS URNs – is used to refer

[4] Cf. Shotton *et al.* 2009.
[5] Cf. Peroni (2017).
[6] See Elmer *et al.* (2011). The commentary is available online at https://ahcip.chs.harvard.edu/.

to the Homeric lines that are commented upon. This technical setting makes it possible to repurpose excerpts of the commentary outside of their original context; in fact, users of the newest front-end of the Perseus Digital Library (the Scaife viewer) have the possibility of visualizing the commentary for the range of Homeric lines in focus (see Fig. 1).

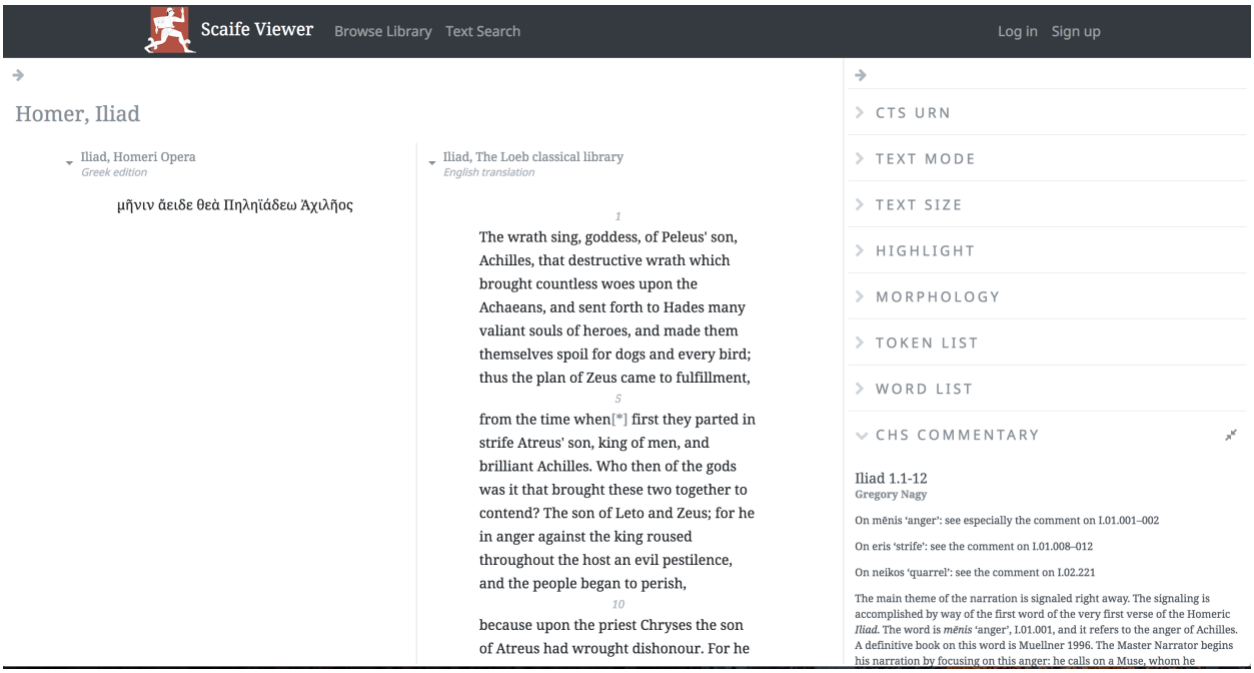

Fig. 1 Reading the *incipit* of the *Iliad* through Perseus Digital Library's Scaife viewer; commentaries on this passage, drawn from *A Homer Commentary in Progress*, are displayed in the bottom-right corner.

The work I describe in this chapter relates to current work in the area of semantic publishing described above in two ways. First, an NLP-based citation mining software is used to semi-automate the task of transforming canonical references into machine readable and actionable data, as I will describe in more detail in the next section of this chapter. Second, the design and implementation of a digital companion for the *Structures of epic poetry* compendium was profoundly informed by this logical separation of data and interface, as I will explain in section 3.

III. The semi-automatic creation of an *index locorum*

In this section I introduce the technology employed to produce the *index locorum* for the *Structures of Epic Poetry* compendium, and discuss the challenges related to its integration into an ongoing publishing workflow.

Mining Digitized Publications

The semi-automatic creation of the *index locorum* was made possible by a technology resulting from the Cited Loci[7] project, originally developed to index canonical references found in existing publications – be they born-digital or digitized.

This technology consists of four software components, working together to perform the extraction of references (see Fig. 2). The Citation Extractor (1) is responsible for identifying the citation components within the stream of text. Subsequently, the Citation Matcher (2) attempts to assign to each extracted reference a unique identifier, in the form of a CTS URN. To this end, it relies on a Knowledge Base (3), a database containing unique identifiers, abbreviations, and variant forms for classical authors and their works. Finally, the Citation Parser (4) takes care of transforming reference scopes into a normalized form, suitable to be embedded into a CTS URN.[8]

[7] On the project see http://citedloci.org, and Colavizza & Romanello (2019). For a more detailed description of the citation mining technology see Romanello (2015), pp. 110-166.

[8] For example, the scope "XII 10s." needs to become "12.10-12.11".

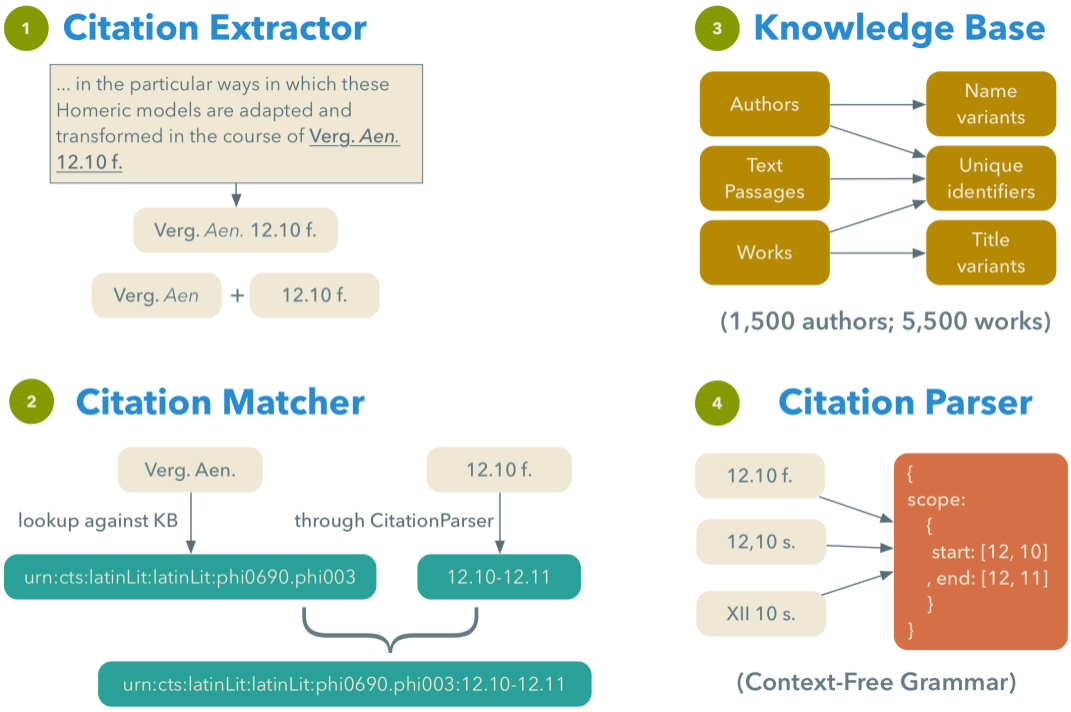

Fig. 2 The four software components used for the automatic extraction of canonical references.

In the context of Cited Loci, this technology was used to index all journal articles contained in JSTOR and classified as belonging to Classics, making it possible to develop new search interfaces that allow scholars to search through JSTOR publications by the references they contain. *Cited Loci of the Aeneid*[9] is a proof of concept of how such new interfaces could look like: it is a web application allowing users to find JSTOR articles containing quotations of or references to the Vergilian poem (see Fig. 3).

9 The tool is openly available at http://aeneid.citedloci.org. For a more detailed description of the interface design and functionalities see Romanello 2019, pp. 83-84, and also https://labs.jstor.org/blog/#!cited_loci_of_the_aeneid-searching_through_jstors_content_the_classicists_way.

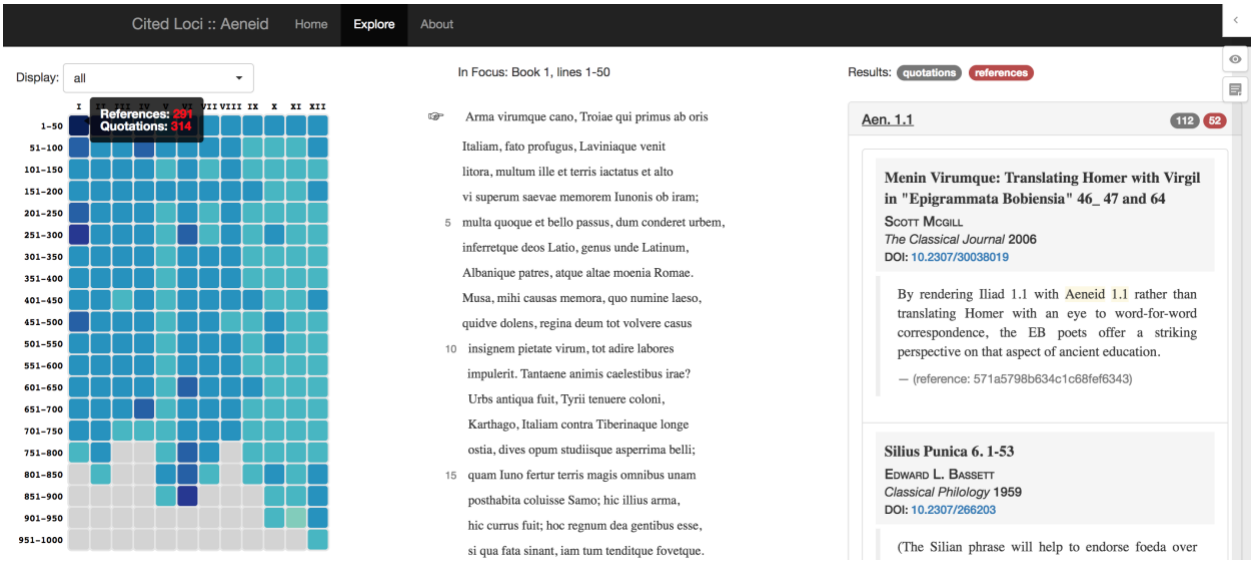

Fig. 3 The interface of *Cited Loci of the Aeneid*.

The starting point for the user is a visual index of the *Aeneid*, displayed on the left. This index uses a heat map to visualise the density of references and quotations for a given section of the poem: the darker a given chunk is, the higher is its density of references and quotations. In this sense, the visual index can be used to identify at a glance sections of the text characterised by an especially high (or low) density of references and quotations. One can already see, for example, how the first half of the poem seems to be more quoted (and referred to) than the second half. Upon selection of a single text chunk, the corresponding Latin text (middle panel) and the matching articles in JSTOR (right panel) are displayed. For each matching article, a snippet of the passage containing the quotation (or reference) is showed.

Publishing Scenarios

How can such a technology for the semi-automatic creation of *indexes locorum* enter the publishing workflows? I believe there are three possible scenarios:

1. **author-centric scenario**: authors directly insert canonical references in a standardised format as they prepare the manuscript.
2. **editor-centric scenario**: editors and their collaborators encode the semi-automatic (or computer-assisted) references, while authors follow a set of citation guidelines when preparing their manuscripts.
3. **publisher-centric scenario**: publisher staff encode references, while the incurred costs are covered by the publication fees.

While the publisher-centric scenario is certainly the most desirable, at least from the perspective of authors and editors, it seems unlikely to be realised in the near future. In fact, not only this scenario requires that publishers have in place the expertise and technical infrastructure needed, but it also implies that they see this as a profitable

endeavour. And, in case the publisher *does not* have already the expertise and infrastructure to deploy the necessary technology, the investment in terms of time and resources will have to be rather substantial.

On the longer run, the author-centric scenario seems the most sustainable option, as it makes (better) use of the time already spent by authors in inserting their references into the manuscript. Such a scenario, however, requires the availability of word processors plugins (similar to what Zotero and Mendeley already do for modern bibliographic references), which unfortunately do not exist yet.

What one is left with, at least for the time being, is the editor-centric scenario, which has the downside of putting an additional and considerable amount of work on the shoulders of (already very busy) editors and their collaborators. The only advantage of this scenario is that the editors can enforce the citation guidelines that are known to work best with the citation mining technology, thus minimizing the need for manual corrections.

Integrating Workflows

Since the ultimate goal of indexing canonical references is the preparation of an *index locorum*, the output of any automatic tool needs to be double-checked manually so as to guarantee the overall accuracy and reliability of the final index. In the workflow we have implemented for the compendium, automatic processing and manual checking go hand by hand and take place at each processing step.

A challenging aspect of such a workflow has been the synchronization of the various publication phases. The manual correction needs to be performed on the final chapter manuscripts, as it is unfeasible to map existing annotations onto documents that are different from those that were originally annotated. As a result, student assistants cannot start working until the final manuscripts have been handed in by the authors. Moreover, since the *index locorum* has to provide the exact page locations of cited text passages, the production of the index can necessarily happen only *after* the camera-ready manuscript of entire compendium has been prepared. Getting the exact page numbers of cited passages is in itself not an easy task given that documents used to typeset the final manuscript and those employed for the extraction of references have different formats (LaTeX for the former and an XML-based format for the latter). To reconcile this discrepancy, it was necessary to re-align the

indexed passages with their corresponding location within the PDF pages so as to be able to include their page numbers into the final *index locorum*, as schematically illustrated in Fig. 4.

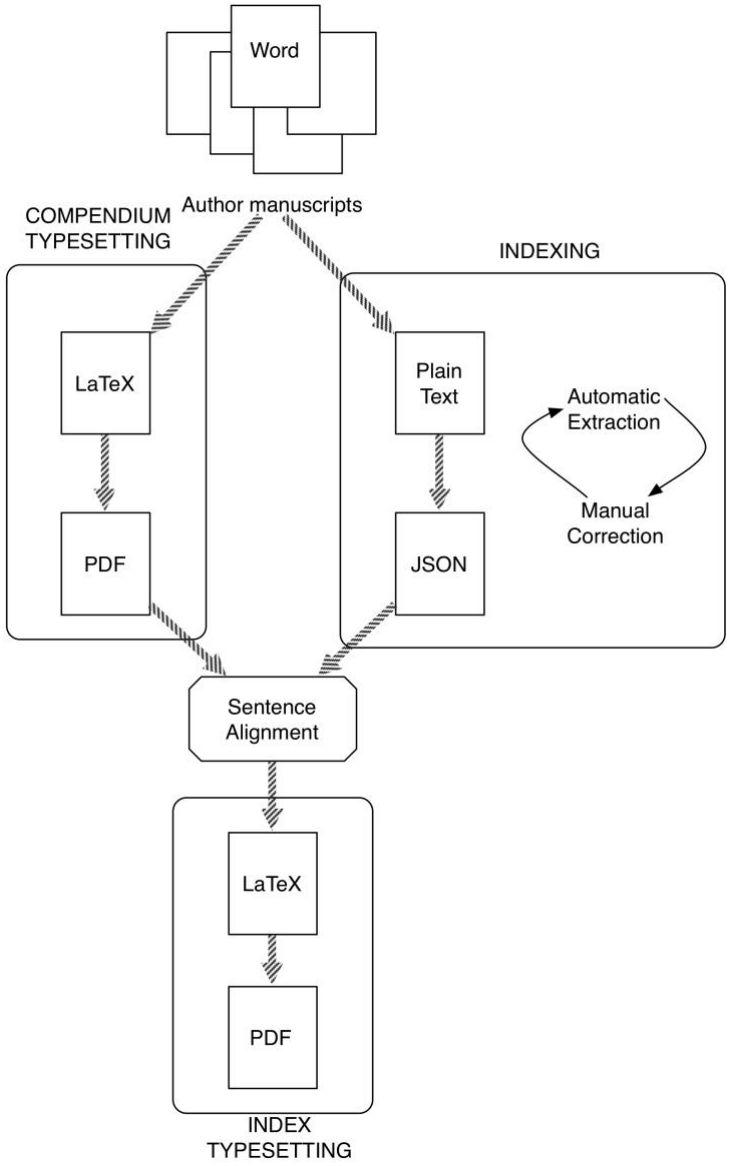

Fig. 4 Integration of semi-automatic indexing into the book production workflow.

The manual and iterative correction of automatically extracted references raised a number of technical issues concerning the choice of an annotation environment where this correction could take place. Such an environment needed to be as quick and reactive as possible, especially on long texts, in order to save precious time; it had to be fairly easy to learn so as to allow student assistants to perform this task; finally, it had to interact nicely with the reference extraction software and its various components.

During an initial phase of the project we used the annotation environment *brat* (Stenetorp et al. 2012) together with a shared spreadsheet to carry out the association of extracted references with identifiers from the knowledge base (in fact, *brat* does not provide support for external knowledge bases). However, *brat* proved to have some serious limitations when applied for our purposes: it becomes considerably slow when working on long texts (that is the case for many of the individual chapters in this publication); it does not provide any functionality to manage the annotation projects, such as monitoring the progress of annotators, calculating the inter-annotator agreement or reconciliating annotations created by several users on the same document.

Based on these considerations, we switched to INCePTION[10], an annotation environment based on WebAnno and partly on *brat*, which solves all the issues above, and most importantly provides seamless integration with external knowledge bases – an essential requirement in our case. This meant that student assistants could correct the extracted references from within one single tool, while their progress could easily be followed and monitored (see Fig. 5).

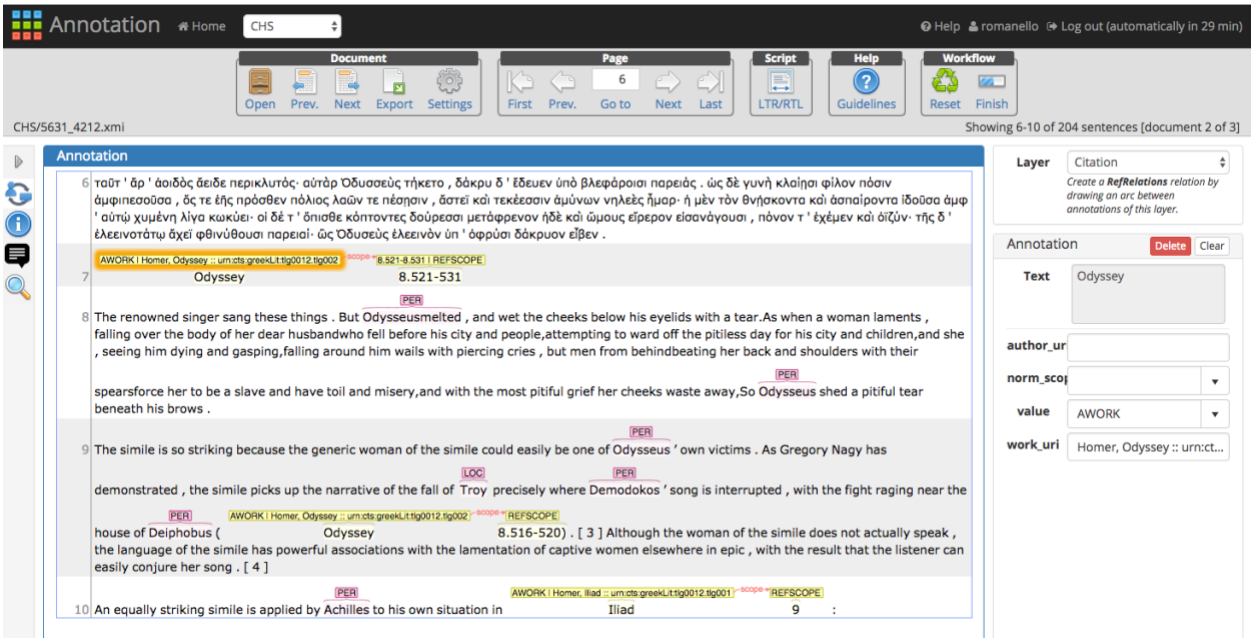

Fig. 5 Correction of extracted canonical references and their disambiguations in INCEpTION.

The main limitation of the citation mining process described above is that references to non-classical texts (e.g. late antique authors) and non-canonical texts (e.g. fragments) are currently not supported. The problem does not lie in the reference extraction phase, but rather in the disambiguation of references: in order for such references to be disambiguated we would need to have unique identifiers, possibly in the CTS URN format, for late antique texts and fragmentary texts.

[10] INCEpTION, https://inception-project.github.io. See also Boullosa *et al.* (2018).

IV. The design of the *Structures of Epic Poetry* digital companion

Once canonical references have been indexed and translated into machine-actionable identifiers, a whole range of new possibilities opens up. In this section I will discuss how such references were exploited in building the digital companion to the *Structures of Epic Poetry* compendium.

Design rationale

There were two main goals of the digital companion: first, to make the compendium's contents more readily and easily searchable by its readers and, second, to publish part of the raw data on which the individual chapters are based.

The principle of *loose coupling* between data and interface, discussed in section 2, deeply informed the design of a digital companion that could both provide a rich search interface and serve as a data publication platform.

Three pieces of information were considered of essential importance for readers when searching through the publication contents:

1. **project categories**: they derive from the taxonomy of epic structures as defined in the *Epische Bauformen* project, and they roughly correspond to the organization of the compendium's subject matter and chapters.
2. **cited passages**: they are the same passages listed in the *index locorum*; they are cited throughout the three volumes of the compendium and classified according to the taxonomy employed in the *Structures of Epic Poetry* compendium, as well as in the Epische Bauformen project[11].
3. **keywords**: they were extracted automatically from the individual chapters by means of a tool called Keyphrase Digger (Moretti, Sprugnoli, and Tonelli 2015), and then refined manually by the editors in cooperation with the authors.

The combination of these three search criteria allows users to identify chapters and single sections of their interest more easily within the entire compendium. A text-based export of search results is provided, so that users can store the result of a search session for later use.

11 See https://www.epische-bauformen.uni-rostock.de/.

User Interface

The design of the user interface was inspired by the work of Rodighiero, Halkia, and Gusmini (2009), who integrated the *indexes* of multiple information sources into a single search interface by means of a list-based layout. In particular, we structured the companion's user interface around the *Focus+Context* paradigm and the concept of *closure*, both at the core of their work.

The Focus+Context[12] paradigm consists in displaying on the screen all available information; upon selection of a specific element the displayed information is utterly reconfigured in order to illustrate the context of the selection. Based on this paradigm, we decided to keep the three indices always visible and available in the left half of the screen, in order to provide the user with both a search/filter function, as well as sufficient contextual information.

In our companion (Fig. 6), what Rodighiero, Halkia, and Gusmini (2009) define as closure, i.e. a way of "enabl[ing] information discovery by visualizing contextual relations between objects", is obtained by making the three *indexes* (categories, keywords and passages) dynamic and mutually dependant. For example, the action of selecting the category "departure scenes" from the categories index will trigger the following actions:

– the right half of the screen will be populated with chapter sections belonging to this specific category;
– the keywords index will be refreshed, so that only the keywords occurring within the "departure scene" sections will be shown;
– the passages index will also be refreshed, now displaying only a navigable tree of authors whose passages are cited in this subset of chapter sections.

In other words, the companion's user interface uses closure to allow users to explore the relations between categories, keywords and passages by means of three dynamic and interlinked list-based filters. Moreover, each index is equipped with a search function, thus enabling users to find search terms without having to scroll long lists.

12 Spence & Apperley (1982).

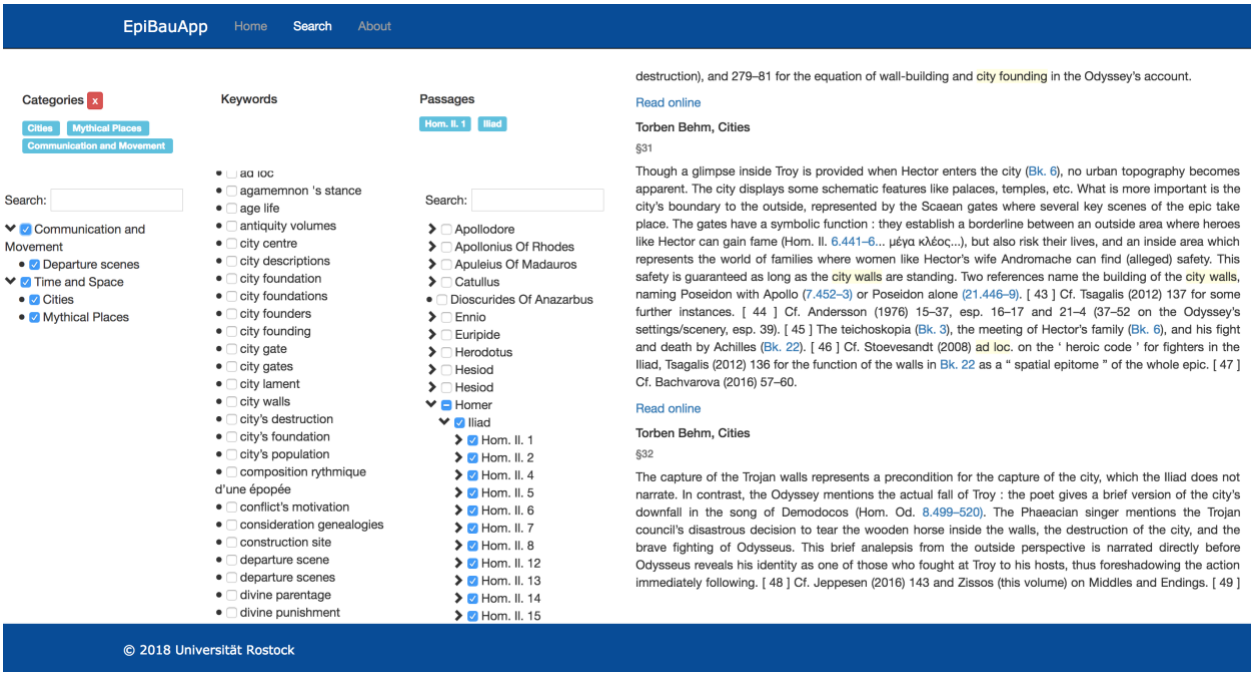

Fig. 6. Detail of the digital companion's search: the user can filter search results based on three filters, namely categories (i.e. different types of epic structures), extracted keywords, and cited passages.

Machine Interface (API)

Besides a user interface, the digital companion provides a machine interface (or API), which can be used to obtain programmatically (e.g. by means of scripts) some of the compendium's data. Thanks to this API, the compendium stops being a static publication to become a publication whose underlying data can be reused in research contexts different from the original ones.

The base URL for the API is http://epibau.ub.uni-rostock.de/api and it provides overall four endpoints[13]. The API's responses are encoded using the Javascript Object Notation (JSON) as a data exchange format.

The available endpoints are:

- **idxlocorum:** it returns the *index locorum* in the form of a hierarchical tree, where each hierarchical level is identified by a CTS URN (e.g. author/work/book/line);

- **keywords:** it returns the list of keywords that can be used as search filters, where each keyword is defined by a label and an identifier;

[13] An endpoint is the address at which a specific collection of resources can be queried. The URL of an endpoint is obtained by chaining together the endpoint's name with the API's base URL (e.g. http://epibau.ub.uni-rostock.de/api/idxlocorum, http://epibau.ub.uni-rostock.de/api/keywords, etc.).

- **categories:** similarly to the previous endpoint, it returns the list of project categories, where each category is represented by a label and an identifier;
- **search:** it allows for searching through the compendium's contents by using one or more category, keyword or passage as filters.

This is, for example, how one could get all extracted keywords via the API[14]:

```
curl -X GET "http://epibau.ub.uni-rostock.de/api/keywords/" -H "accept: application/json"
```

One could then further explore the compendium based on a keyword of interest, e.g. "city walls", designated in this case by the keyword identifier "5b0278833c630e4c9e770313":

```
curl -X GET " http://epibau.ub.uni-rostock.de/api/search/?kw=5b0278833c630e4c9e770313" -H
"accept: application/json"
```

Finally, the keyword identifier can be combined with a passage identifier, in the form of a CTS URN, to retrieve all publication sections containing a specific keyword (or set of keywords) and citing one or more text passages. For example, one could search for passages where the keyword "city walls" occurs *and* Statius' *Thebaid* is explicitly cited (urn:cts:latinLit:phi1020.phi001 is the CTS URN of the *Thebaid*):

```
curl -X GET "http://epibau.ub.uni-
rostock.de/api/search/?kw=5b0278833c630e4c9e770313&urn=urn:cts:latinLit:phi1020.phi001" -H
"accept: application/json"
```

To sum up, the *Structure of Epic Poetry* digital companion not only offers a web interface with a powerful mechanism to search within the compendium, but it also provides a machine interface (i.e. API) which allows for interacting programmatically with the contents of the compendium – especially project categories, extracted keywords and cited passages.

[14] The code examples below make use of the command line utility cURL in order to issue queries to the Digital Companion's API (API's responses are not displayed for the sake of readability).

V. Future Prospects

In this chapter I have described the technical work that has been happening behind the scenes in the production of the compendium *Structures of Epic Poetry* as well as of its digital companion. The latter, in particular, exemplifies the advanced user interfaces that can be conceived to explore and read publications whose contents have been richly annotated. The convenience of the digital companion, compared with traditional (printed) *indexes locorum*, is striking: the characteristic list-based structure of the index remains, but the reader is now able to draw search terms from several indices at once and combine them to form complex queries.

Publishing Workflows

With regards to the computer-assisted creation of the *index locorum*, we followed the editor-centric scenario discussed in section 3.1. The main limit of this scenario is that it puts an additional overhead on the shoulders of editors and their collaborators, and it does not leverage the time that authors already dedicate to inserting bibliographic references into their manuscripts.

In the longer run, we should aim to enable authors to insert directly such references in a semantic (or at least structured) format at manuscript preparation phase. To achieve this, one basic piece of technical infrastructure is still missing, namely the availability of word processor plugins similar to those existing for reference management software like Zotero or Mendeley. Another advantage of providing such a plugin for authors to manage their references of primary sources while writing would be the possibility of applying different formatting (i.e. citation styles) to the same document, based on the needs.

Interconnectivity and Discoverability

If we are to take a look into the future of digital publishing from an *Altertumswissenschaft* standpoint, providing publications with appropriate machine interfaces (or APIs) will be a very impactful technical advancement. Such APIs can exist either at the level of single publications – such is the case with the compendium – or can be developed for entire portals, publication series or even publisher's offers. Thanks to these APIs, the discoverability of relevant publications – a task greatly hindered by the current information overload – can be enhanced by implementing e.g. services that provide researchers with publication alerts based on specific sources being cited, or with links of links to publications on a specific passage, like the above mentioned Scaife Viewer is doing with respect to the CHS' commentaries (see section 2.1).

Ultimately, making available publications through this kind of APIs will have the effect of increasing the discoverability of Classics publications, which is currently hindered, among other things, by the limitations of general purpose citation indexes like Google Scholar. These indexes, in fact, do not support the retrieval of

documents based on the references to Classical texts they contain – which was instead the main goal and outcome of the *Cited Loci* project, on which the work described in this chapter has built upon. As a result, scholars in disciplines outside of Classics struggle to find relevant literature about classical works, which *does exist* but is somewhat hard to find via tools like Google Scholar.[15] While the available citation indexes render Classics scholarship essentially as an echo-chamber, whose outputs are hard to access for scholars from other disciplines, ad-hoc APIs could help us making what is published in our field more easily discoverable.

Nachleben of Data

Finally, the compendium's data may have a life beyond the actual publication. Since all data are available via the API, other scholars or projects could build upon them. It is not too hard to imagine for example scholars of intertextuality being interested in gathering all sections of the compendium where a given set of parallel passages are cited (Coffee 2018). Or to imagine scholars working on the computer-assisted detection of allusions and other text reuse phenomena, to leverage the thematic classification of passages discussed in the compendium to improve the performance of their systems (Nelis et al. 2018).

Acknowledgements

My deepest gratitude goes to Christiane Reitz and Simone, the editors of this compendium, for having fully supported and believed in the potentially risky enterprise of producing semi-automatically an *index locorum*. I am indebted to Chris Forstall, Damien Nelis and Lavinia Galli Milić, without whom this collaboration would have probably never happened. I am also grateful to Jeffrey Witt, Ethan Gruber, Emmanuelle Morlock and Thibault Clérice for fruitful online conversations about "digital publications as APIs". Finally, a thanks to Dario Rodighiero for having read earlier versions of this chapters and for the precious advices on the design of the digital companion.

References

Shotton, David. 2009. "Semantic Publishing: the Coming Revolution in Scientific Journal Publishing". *Learned Publishing* 22 (2): 85–94. doi:10.1087/2009202.

Sato, Mitsuhisa, Satoshi Matsuoka, Piotr Nowakowski, Eryk Ciepiela, Daniel Harężlak, Joanna Kocot, Marek Kasztelnik, et al. 2011. "The Collage Authoring Environment". In *Proceedings of the International Conference on*

[15] On this problem, see Gainsford (2018).

Computational Science, ICCS 2011, edited by Mitsuhisa Sato, Satoshi Matsuoka, Peter M. Sloot, G. Dick van Albada, and Jack Dongarra, 4:608–17. doi:10.1016/j.procs.2011.04.064.

Shotton, David, Katie Portwin, Graham Klyne, and Alistair Miles. 2009. "Adventures in Semantic Publishing: Exemplar Semantic Enhancements of a Research Article". *PLoS Computational Biology* 5 (4).

Peroni, Silvio. 2017. "Automating Semantic Publishing". Edited by Michel Dumontier. *Data Science* 1 (1-2). IOS Press: 1–19. doi:10.3233/DS-170012.

McGuire, Hugh. 2013. "A Publisher's Job Is to Provide a Good API for Books: You Can Start with Your Index". *The Indexer* 31 (1). Society of Indexers: 36–38. http://www.ingentaconnect.com/content/index/tiji/2013/00000031/00000001/art00008{#}.

Witt, Jeffrey C. 2018. "DSE's AND API Consuming Applications". In *Digital Editions as Interfaces*, edited by Stefan Dumont.

Elmer, David, Douglas Frame, Richard Martin, Leonard Muellner, and Gregory Nagy. 2011. "Introduction to A Homer Commentary in Progress". *Classics@* 8. https://chs.harvard.edu/CHS/article/display/6470.

Romanello, Matteo. 2015. "From Index Locorum to Citation Network: an Approach to the Automatic Extraction of Canonical References and Its Applications to the Study of Classical Texts". PhD thesis, King's College London. doi:11858/00-1780-0000-002A-4537-A.

Stenetorp, Pontus, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. 2012. "Brat: a Web-Based Tool for NLP-Assisted Text Annotation". In *EACL '12 Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 102–7. Avignon, France: Association for Computational Linguistics.

Moretti, Giovanni, Rachele Sprugnoli, and Sara Tonelli. 2015. "Digging in the Dirt: Extracting Keyphrases from Texts with KD". In *Proceedings of the Second Italian Conference on Computational Linguistics CLiC-It 2015*, 198–203. Accademia University Press. doi:10.4000/books.aaccademia.1518.

Rodighiero, Dario, Matina Halkia, and Massimiliano Gusmini. 2009. "Mapping for Multi-Source Visualization: Scientific Information Retrieval Service (SIRS)". In *Human-Computer Interaction*, edited by J.A. Jacko, 5613:597–605. https://link.springer.com/content/pdf/10.1007{%}2F978-3-642-02583-9{_}65.pdf.

Coffee, Neil. 2018. "An Agenda for the Study of Intertextuality". *TAPA* 148: 205–23. doi:10.1353/apa.2018.0008.

Nelis, Damien, Christopher Forstall, Lavinia Galli Milić, and Christopher W Forstall. 2018. "Intertextuality and Narrative Context: Digital Narratology?". *Journal of Data Mining and Digital Humanities*. https://hal.inria.fr/hal-01480773/document.

Gruber, Ethan. 2018. "Linked Open Data for Numismatic Library, Archive and Museum Integration". In *Proceedings of the 44th Conference on Computer Applications and Quantitative Methods in ArchaeologyCAA2016—Oceans of Data*, edited by M. Matsumoto and E. Uleberg, 33–40.

Clérice, Thibault, and Thibault. 2017. "Les Outils CapiTainS, l'éDition NumériQue Et l'Exploitation Des Textes". *MéDiéVales* 73 (73). Presses universitaires de Vincennes: 115–31. doi:10.4000/medievales.8211.

Romanello, Matteo, and Michele Pasin. 2017. "Using Linked Open Data to Bootstrap a Knowledge Base of Classical Texts". In *Proceedings of the Second Workshop on Humanities in the Semantic Web (WHiSe {II)} Co-Located with 16th International Semantic Web Conference {(ISWC} 2017), Vienna, Austria, October 22, 2017.*, edited by Alessandro Adamou, Enrico Daga, and Leif Isaksen, 2014:3–14. {CEUR} Workshop Proceedings. CEUR-WS.org. http://ceur-ws.org/Vol-2014/paper-01.pdf.

Elmer, David, Douglas Frame, Richard Martin, Leonard Muellner, and Gregory Nagy. 201. "Introduction to A Homer Commentary in Progress". *Classics@* 8.

Romanello, Matteo. 2019. "Large-Scale Extraction of Canonical References : Challenges and Prospects". In *Digital Humanities and Antiquity. Humanités Numériques et Antiquité. Actes du colloque international, Grenoble 2-4 septembre 2015.*, edited by Isabella Cogitore and Elena Pierazzo.

Colavizza, Giovanni, and Matteo Romanello. 2019. "Citation Mining of Humanities Journals: a Comparison of the Projects *Cited Loci* and *Linked Books*". In *Journal of European Periodical Studies,* xx-zz.

Boullosa, Beto, Richard Eckart de Castilho, N. Kumar, J.-C. Klie, Irina Gurevych. 2018. "Integrating Knowledge-Supported Search into the INCEpTION Annotation Platform". *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP),* Brussels, Belgium.

Spence, Robert, and Mark Apperley. 1982. "Data Base Navigation: An Office Environment for the Professional." *Behav. Inf. Technol.* 1 (1) (January 26): 43–54. doi:10.1080/01449298208914435.

Oldfather, W.A. 1937. "Suggestions for Guidance in the Preparation of a Critical Index Verborum for Latin and Greek Authors." *Trans. Proc. Am. Philol. Assoc.* 68: 1. doi:10.2307/283249.

Gainsford, Peter 2018. "The Citation Problem." *Kiwi Hell. Mod. Myth. about Anc. World.* Available at : http://kiwihellenist.blogspot.com/2018/09/the-citation-problem.html