

Monica Berti

Historical Fragmentary Texts in the Digital Age

Abstract: This paper describes how the digital revolution is changing the way scholars access, analyze, and represent historical fragmentary texts, with a focus on traces of quotations and text reuses of ancient Greek and Latin sources. The contribution presents two different projects: 1) the Digital Fragmenta Historicorum Graecorum (DFHG), which is a digital collection of ancient Greek fragmentary historians enriched with functionalities for accessing and analyzing their texts; 2) the Digital Athenaeus, which provides experimental tools for reading the text of the *Deipnosophists* of Athenaeus of Naucratis and getting information about citations of authors and works that are preserved in it.

Introduction

In the last two centuries generations of scholars have been publishing many critical editions of historical fragmentary texts of Greek and Latin sources. These publications are the result of an intense work for individuating and assembling traces of quotations and text reuses of authors whose works are now mostly lost. Classical scholarship has adopted the word *fragmenta* to name these traces and describe their transmission in our textual heritage.¹ In this case the term doesn't refer to broken off pieces of material objects bearing textual evidence, but to the output of philological analyses of researchers who have to dig into the context of literary texts to individuate references to authors and works.² The goal of this paper is to describe how the digital revolution is changing the way scholars evaluate and represent fragmentary texts, while preserving the lesson of a long established editorial and philological tradition.³

1 (Most 1997).

2 (Berti 2012; 2103).

3 A detailed description of this topic is forthcoming in a monograph by Monica Berti entitled *Digital Editions of Historical Fragmentary Texts*.

Monica Berti, Universität Leipzig

Classical scholarship and fragmentary texts

Glenn Most individuates two main phases in the history of modern scholarship on collecting fragmentary texts:⁴ 1) the humanist and early modern phase that began in the second half of the sixteenth century and was interested more in publishing the very best fragments of the most important authors than in producing complete, critical, and exhaustive collections, and 2) the romantic and contemporary phase that began in the second half of the eighteenth century and brought a new attempt to understand the totality of the past beyond the few surviving canonical works. The second phase was fundamental for developing a new scholarship on ancient literary fragments that took off in the middle and the second half of the nineteenth century, when scholars began to establish rigorous philological methods for producing big collections of fragmentary texts belonging to many different genres, as for instance epic poetry, comedy, tragedy, philosophy and historiography.

These important efforts in collecting fragmentary texts depend not only on an interest in looking for every possible trace of the past, but also on the fact that fragmentary literature covers a significant percentage of what has been preserved from our tradition.⁵ Given the fragmentary state of ancient evidence and its complexity, counting the amount of textual fragments and calculate its proportion in relation to what has survived from the past is a difficult task that can't produce complete and definitive results, first of all because it is not possible to establish with precision and uniquely what is a fragmentary text. In spite of that, digital libraries of Greek and Latin sources allow us to undertake this task at least in a provisional way.

According to statistics performed on the online Thesaurus Linguae Graecae (TLG) for the period of time between the eighth century BC and the sixth century CE, about 50% of authors is represented by fragmentary authors (Figure 1). Within this group, more than 80% is represented by authors who are completely lost (e.g., Hellanicus), and about 18% by authors who have both fragmentary and still extant works (e.g., Sophocles) (Figure 2).⁶

⁴ (Most 2009, 15–17).

⁵ As far as it concerns ancient Greek historiography, Strasburger (1977, 9–15) tried to quantify the “land of ruins” of this genre and came to the conclusion that the tradition has preserved only about 2.5% of what was originally written, with a *ratio* of 1 to 40 between what is still extant and what is lost.

⁶ These percentages are based on TLG data as of early 2018: <http://stephanus.tlg.uci.edu> (last access 2019.01.31). For more information on this data and other digital collections,

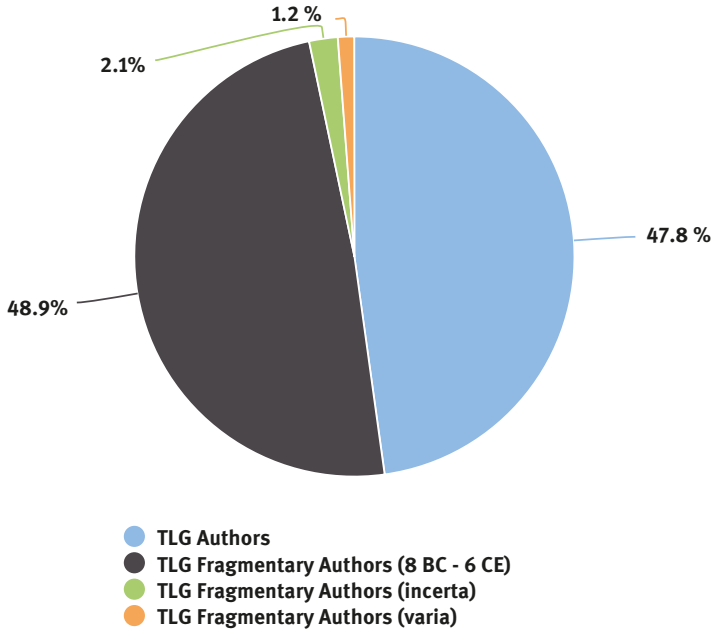


Figure 1: Fragmentary authors in the Thesaurus Linguae Graecae (TLG).

These percentages reflect the situation of Greek literature and the amount of fragmentary authors, showing the importance of working on this kind of evidence for improving our knowledge of Classical sources. As mentioned before, traditional scholarship has given an extraordinary contribution to the critical reconstruction of the intellectual personality of many lost authors by establishing philological criteria to study and edit them with the technology of the printed book. Today digital tools offer a new environment that requires us to rethink the way we analyze and represent this kind of evidence. In the following paragraphs we will describe concrete opportunities and challenges of the digital revolution by presenting two projects focused on ancient Greek fragmentary texts.

see <http://www.dfhg-project.org/Fragmentary-Texts> (last access 2019.01.31). A detailed description of these resources will be available in the monograph mentioned at note 3.

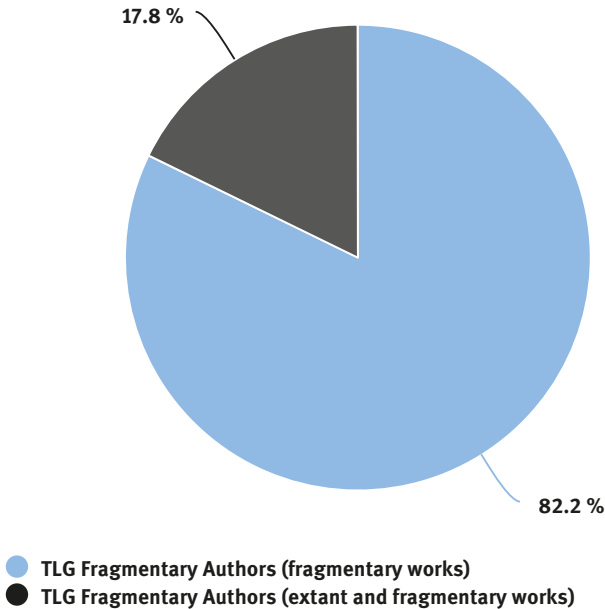


Figure 2: Fragmentary authors in the Thesaurus Linguae Graecae (TLG).

Fragmentary texts and the digital revolution

Digital Classical philology is working on two main challenges. The first one is the conversion of printed editions of Greek and Latin sources into a machine readable format that preserves their textual and editorial heritage.⁷ The second challenge is the publication of new digital critical editions that make use of computational technologies and help define standards and scholarly models that are different from those developed through the technology of traditional books.⁸ The first generation of digital libraries has digitized the reconstructed

⁷ In this regard a lot of work is currently developed by the Open Greek and Latin (OGL) project at the University of Leipzig. See also the contributions by Leonard Muellner and Samuel J. Huskey in this volume.

⁸ On this aspect see the paper by Franz Fischer in this volume. On the technology of the printed book, see Borsuk (2018).

text of single editions of Classical works.⁹ The goal of the second generation of digital libraries is to publish multiple editions of the same work, reproduce the critical apparatus and all other paratextual elements (prefaces, introductions, indexes, bibliographies, notes, etc.), and generate collaborative environments for critical editing of Greek and Latin sources.¹⁰

Fragmentary works are directly involved in this process because they consist of quotations and text reuses preserved by still extant sources. This means that also in this case efforts are focused both on the digitization of printed critical editions of fragmentary authors and on the implementation of a new model for representing fragments inside digital contexts. The Digital Fragmenta Historicorum Graecorum (DFHG) and the Digital Athenaeus are two projects focused on these aspects for dealing with fragmentary authors and works in a digital environment.

The Digital Fragmenta Historicorum Graecorum (DFHG)

The Digital Fragmenta Historicorum Graecorum (DFHG) is the digital version of the five volumes of the *Fragmenta Historicorum Graecorum* (FHG), which is the first big collection of ancient Greek historical fragments published by Karl Müller.¹¹ The FHG is a collection of quotations and text reuses (*fragmenta*) extracted from many different sources pertaining to 636 ancient Greek fragmentary historians. Except for the first volume, authors are chronologically distributed and date from the sixth century BC through the seventh century CE. Fragments are numbered sequentially and arranged by works and book numbers with Latin translations, commentaries, and critical notes. A separate appendix at the end of the first volume includes the *Marmor Parium* and the Greek text of the *Marmor Rosettanum* with translations and commentaries.¹²

⁹ Examples are the TLG, the Perseus Digital Library, and the PHI Latin Texts.

¹⁰ For reasons of space, we only refer to two generations of digital libraries, but see Babeu (2011, 2–3) on “several generations of digital corpora in Classics”.

¹¹ (Müller 1841–1873). The project is available online at <http://www.dfhg-project.org> (last access 2019.01.31). See also Berti (2019).

¹² Two digital projects are currently developed on the *Marmor Parium* and the *Marmor Rosettanum*: see Berti and Stoyanova (2014) and Berti et al. (2016c).

The fifth volume collects Greek and Syriac historical fragments preserved in Armenian texts.¹³

The DFHG is not a new edition of ancient Greek fragmentary historians, but a digital experiment to provide textual, philological, and computational methods for representing fragmentary authors and works in a digital environment. The reason for choosing the collection of the FHG depends on different factors: 1) an interest in Greek fragmentary historiography, which offers many examples of reuse of prose texts whose complexities are shared by other genres of fragmentary literature;¹⁴ 2) the necessity of digitizing printed editions and preserving them not only as image files but also as structured machine readable collections, that can be accessed for experimenting with text mining of historical languages; 3) the importance of the FHG for understanding more recent editions of Greek historical fragments and in particular *Die Fragmente der griechischen Historiker* (FGrHist) by Felix Jacoby, who spent his life to change and improve the collection created by Karl Müller;¹⁵ 4) the fact that the corpus of the FHG is open (i.e., free of copyright) and big enough to perform computational experiments and obtain results.¹⁶

The DFHG is an ongoing project that has been developing many tools and services not only for accessing the entire collection of the FHG, but also for providing a new digital and philological model that can be applied to other collections of fragmentary authors. The complete text of the five volumes of the FHG has been converted into a machine readable format with Optical Character Recognition (OCR) systems as part of the Open Greek and Latin (OGL) project at the University of Leipzig.¹⁷ The digital version of the FHG has been produced starting from the OCR output by creating an SQL database for delivering web services and tools. Web pages are generated using the Ajax technique to retrieve data from the database and increase the usability of the huge amount of FHG contents. Functionalities of the DFHG are

13 On the collection of the FHG, see Petitmengin (1983) and Grafton (1997).

14 (Berti 2012; 2103).

15 (Jacoby 1909; 2015).

16 The FHG is a corpus of more than 2 million words (in Greek, Latin, and French) with more than 600,000 Greek tokens.

17 On OCR for ancient Greek and Latin see the contribution by Bruce Robertson in this volume. Even if nowadays it is possible to obtain good results when OCRing nineteenth century editions of ancient Greek and Latin sources, errors are still present in OCRed texts. The DFHG project is working on OCR post-correction and also includes an experimental editing environment for manual corrections.

presented below and descriptions are grouped according to tools and add-ons developed by the project.¹⁸

1. Visualization of DFHG Contents. Contents of the DFHG can be browsed by selecting the entire collection or one single volume in the homepage of the project. The slide in/out navigation menu represents the whole structure of volumes, books, authors, works and fragments collected in the printed edition, and it is available for the entire collection and for each volume.¹⁹ The “Expand All” and “Collapse All” functions allow scholars to navigate the FHG with a comprehensive view of the structure of the whole collection by expanding and collapsing every volume, book, author and work down to the fragment level. This structure is very helpful because the printed version of the FHG does not contain detailed tables of contents of its volumes.²⁰ Following each navigation menu element, users are able to jump to the relevant section of the FHG without reloading the page. The navigation menu gives access to the following contents as they are arranged in the FHG: *volumina* (FHG I-V), *praefationes* (FHG I, II, IV and V), *libri* and other volume divisions (FHG I-V), list of authors, books and fragments (FHG I-V), *Index Nominum et Rerum* (FHG I), *Index Marmoris Rosettani* (*Table de mots grecs, et des principaux faits expliqués*) (FHG I) and *addenda et corrigenda* (FHG I-V).²¹ The DFHG main page of the entire collection and of each volume allows to visualize and navigate the following contents: a) introductions to FHG authors with notes;²² b) five-item rows for each fragment with the following data: (1) the number of the fragment with links to the relevant page of the printed edition of the FHG, to the *Index Nominum et Rerum* and the *Index Marmoris Rosettani*, and to the OpenNLP POSTagger for Ancient Greek, (2) a reference to the source text of the fragment (sometimes with

18 Tools and add-ons are available through the homepage of the project with detailed descriptions and instructions.

19 The menu faithfully represents the arrangement of authors and fragments in the FHG.

20 The FHG only provides an *index auctorum* and an *index titulorum* at the end volume IV.

21 FHG III doesn't have a *praefatio*. Still missing in the DFHG are the *index auctorum*, the *index titulorum*, and the *index nominum et rerum* of volume II-IV that are published at the end of FHG IV, and the *indices* of the two sections of FHG V. Also, *addenda et corrigenda* in the DFHG are represented as separate web pages at the end of each volume due to the fact that their integration in the relevant passages of the collection would have required too much manual work.

22 FHG I has a unique introduction at the beginning of the volume, which has been split into sections corresponding to each author of the volume and inserted at the beginning of the relevant author section in the DFHG. In this case the DFHG follows the model of the other FHG volumes, where almost every author has a separate introduction in the relevant section.

a short or long commentary), (3) the Greek or the Latin text of the fragment, (4) the Latin (or French) translation/summary of Greek fragments, and (5) the Latin (or French) commentary to the text of the fragment;²³ c) two- or three-item rows for still surviving sources (e.g., Apollodorus' *Bibliotheca*, the *Marmor Parium*, and the *Marmor Rosettanum* in FHG I, or Diodorus Siculus in FHG II) with (1) the Greek text, (2) the Latin (or French) translation, and (3) the commentary sometimes with notes. The grey sidebar of the main page shows the original arrangement of pages in the FHG with links to the printed edition available through Google Books.

2. Search through the DFHG. The DFHG Digger filters the FHG according to authors, works, work sections and book numbers. By typing and selecting through a live search, users can display the desired part of the collection. It is possible to combine filters using logical AND/OR expressions to get a more precise selection.²⁴ DFHG contents (introductions, fragments, translations, commentaries and source texts) are searchable in two different ways: (1) by highlighting words in the DFHG main page of the entire collection or of a single volume, and (2) by searching words directly in the DFHG Search tool. Results show the number of occurrences in each DFHG author and are organized by authors and works, and searched words are highlighted in the texts of the DFHG. When available, results display also inflected forms and lemmata through Morpheus, the Suda On Line, and the Liddell-Scott Lexicon in the CITE Architecture (see below).²⁵
3. Integration with external resources. One of the main goals of the project is to integrate the DFHG with external resources such as textual collections, authority lists, dictionaries, lexica and gazetteers. The DFHG main page is currently connected to the printed edition of the FHG, to the 8,427 entries of the *Index Nominum et Rerum* (FHG I), to the 249 entries of the *Index Marmoris Rosettani* (FHG I) and to the OpenNLP POSTagger for Ancient Greek; the DFHG search tool is connected to the corresponding fragment in the main page, to Morpheus, the Suda On Line and the Liddell-Scott Lexicon in the CITE Architecture. These resources allow users to get information about the texts of

²³ On the OpenNLP POSTagger see Celano et al. (2016). On its integration in the DFHG, see below.

²⁴ Combining for example author name and work title, like CHARON and ΠΕΡΣΙΚΑ.

²⁵ Morpheus is the morphological parsing and lemmatizing tool of the Perseus Project, the Suda On Line is the digital version of the lexicon Suda, and the Liddell-Scott Lexicon in the CITE Architecture is the digital version of the LSJ lexicon published as a CITE collection (see the paper by Chistopher W. Blackwell and Neel Smith in this volume).

the fragments of the FHG by obtaining results concerning the morphology of words, their syntactic function, their meaning, and the disambiguation of named entities. As far as it concerns ancient Greek and Latin, all these resources already offer significant results, but are not complete and still require a work of disambiguation and correction. The goal is to make use of these resources to automatically disambiguate and annotate part of the DFHG data, which in turn offers a collection of parsed texts for enriching external libraries of Greek and Latin sources. In this regard, the DFHG project is working on named entities recognition and on the creation of a complete DFHG thesaurus by including other external authority lists. Figure 3 shows an example with some of the DFHG occurrences of the Greek word Εὐρώπη, which is both a personal and a place name. The lemmatization of the inflected forms automatically identifies the word both in the Lexicon of Greek Personal Names (LGPN) and in Pleiades.²⁶ A further work of analysis of the contexts of the DFHG fragments, where this word appears, provides an overview of the use of Εὐρώπη in Greek historiography both as a personal and a place name.

4. Data Citation. Each DFHG menu element has a unique identifier expressed as a URN (Uniform Resource Name). The syntax of each URN represents the editorial work of Karl Müller, who has arranged fragments in a sequence and has attributed them to fragmentary authors, works, work sections and book numbers. The following examples show different levels of granularity of these URNs, that are used to identify and cite fragmentary authors and works down to the fragment level.
 - urn:lofts:fhg.1.hecataeus identifies the author Hecataeus in FHG I;
 - urn:lofts:fhg.1.hecataeus.hecataei_fragmenta identifies the whole section of Hecataeus' fragments in FHG I;
 - urn:lofts:fhg.1.hecataeus.hecataei_fragmenta.genealogiae identifies Hecataeus' Γενεαλογία in FHG I;
 - urn:lofts:fhg.1.hecataeus.hecataei_fragmenta.genealogiae.liber_secundus identifies the second book of Hecataeus' Γενεαλογία in FHG I;
 - urn:lofts:fhg.1.hecataeus.hecataei_fragmenta.genealogiae.liber_secundus:350 identifies fragment 350 of the second book of Hecataeus' Γενεαλογία in FHG I.

²⁶ LGPN is originally a printed edition that collects all ancient Greek personal names attested on written sources from the eighth century BC down to the late Roman Empire (<http://www.lgpn.ox.ac.uk>; last access 2019.01.31). Pleiades is a community-built gazetteer and graph of ancient places (<https://pleiades.stoa.org>; last access 2019.01.31).

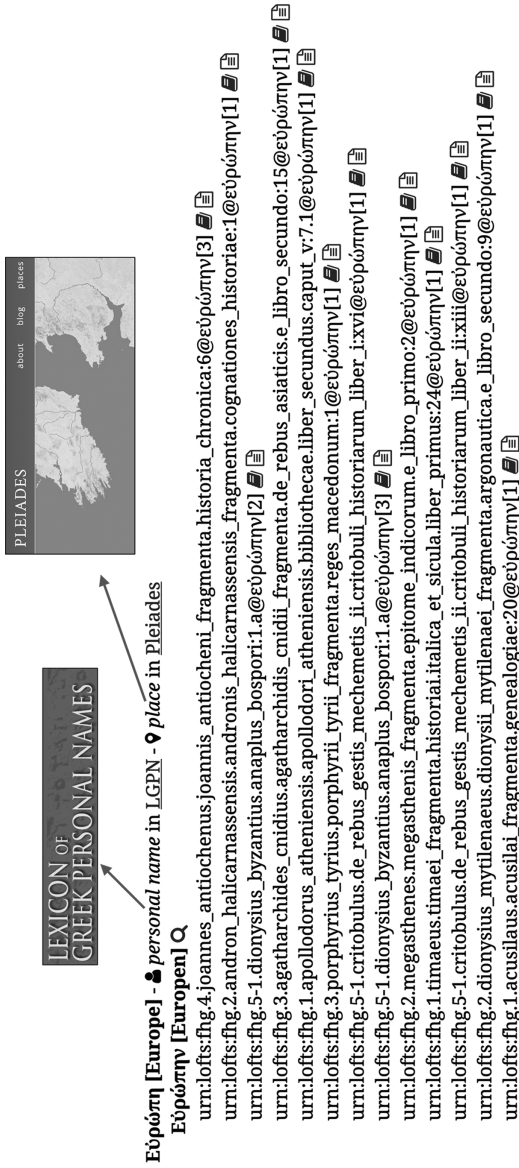


Figure 3: Named entity disambiguation in the DFHG.

A URN identifies itself as a urn in the LOFTS domain, whose acronym stands for the Leipzig Open Fragmentary Texts Series (LOFTS) and represents the domain of textual fragments.²⁷ Work titles in the URN are expressed in the Latin translation provided by Müller in the FHG. URNs are combined with a URL prefix (<http://www.dfhg-project.org/DFHG/#>) to generate stable links. The DFHG project provides also CITE URNs according to the guidelines of the CITE Architecture.²⁸ CITE URNs are accessible through the DFHG API, the DFHG Fragmentary Authors Catalog, and the Müller-Jacoby Table of Concordance (see below). By using URN identifiers, it is possible to export citations of DFHG fragments and source texts down to the word level. By selecting the desired portion of text, users get a URN that identifies the selection.²⁹ The DFHG provides also a URN Retriever, which is a tool for retrieving and citing passages and words in the fragments by typing the corresponding URN. For example:

- Hellanicus' fragment 1 corresponds to `urn:lofts:fhg.1.hellanicus.hellanici_fragmenta.phoronis:1;`
- the beginning of Hellanicus' fragment 1 (Ελλάνικος ὁ Λέσβιος τὸς Τυβόρηνοὺς φησι, Πελασγούς πρότερον καλουμένους, ἐπειδὴ κατώκησαν ἐν Ἰταλίᾳ, παραλαβεῖν ἦν ἔχουσι προσηγορίαν) corresponds to `urn:lofts:fhg.1.hellanicus.hellanici_fragmenta.phoronis:1@ελλάνικος[1]-προσηγορίαν[1]`.

5. Data export. The DFHG provides a web API that can be queried with author names and fragment numbers. The result is a JSON output containing every piece of information about the requested fragment.³⁰ The DFHG automatically exports data to CSV and XML files. XML files are generated both as EpiDoc XML and well formed XML. EpiDoc XML files represent the structure of the printed edition of the FHG and are based on guidelines specifically developed for the DFHG project as part of the EpiDoc community.³¹ Well formed XML files collect information about fragments and source texts of the FHG.
6. DFHG Fragmentary Authors Catalog. This tool searches and visualizes the 636 Greek fragmentary historians whose quotations and text reuses are collected in the FHG. The catalog enables users to search the database by

²⁷ (Berti et al. 2016a; Berti 2018).

²⁸ See the paper by Christopher W. Blackwell and Neel Smith in this volume.

²⁹ For example `urn:lofts:fhg.1.ephorus.ephori_fragmenta.historiae.liber_tertius:37@περιθοῖδαι[1]-ιξίουος[1]` identifies the sentence Περιθοῖδαι, δῆμος τῆς Οἰνηίδος φυλῆς, ἀπὸ Πειρίθου τοῦ Ἰξίουος in Ephorus' fragment 37.

³⁰ For example <http://www.dfhg-project.org/DFHG/api.php?author=ACUSILAU&fragment=10> (last access 2019.01.31).

³¹ (Berti et al. 2014–2015).

authors (e.g., Hippys Rheginus) and volumes (e.g., FHG II). Results display data about the exact location of authors in the FHG, their chronology according to the arrangement by Müller, pages with links to both the digital and the printed version of the FHG, CITE URNs of DFHG authors (e.g., urn:cite:lofts:fhg.1.hellanicus), and places corresponding to the geographical epithet of each FHG author used by Müller with links to Pleiades canonical URIs.³² This data can be also visualized in the Fragmentary Authors Map and in the Fragmentary Authors Chart, which represent the geographical distribution of FHG authors and their arrangements in the volumes of the printed edition.³³

7. DFHG Witnesses Catalog. This tool searches and visualizes authors and works (witnesses) that preserve quotations and text reuses of FHG fragmentary historians. The catalog allows users to search the database by FHG authors (e.g., Phanodemus) and works (e.g., ATTIKA), by witnesses (authors and works, as for instance Harpocration or the *Deipnosophistae*), and by editions, manuscripts and inscriptions cited in the FHG as sources of fragments (e.g., Bethe. *Pollucis Onomasticon* I. Lipsiae 1900, the *Codex Palatinus Graecus* 398, and *IG* XII.5.444). Results display witnesses (authors and works) with Perseus Catalog URNs (e.g., urn:cite:perseus:author.728 and urn:cts:greekLit:tlg0016.tlg001), literary and geographical epithets and dates of witnesses (authors) according to the TLG, the Perseus Catalog, Pleiades and the Brill's New Pauly, passages of works (witnesses) that preserve quotations and text reuses with detailed information about the corresponding *fragmenta* in the DFHG and links to the URN Retriever, and finally references to inscriptions, manuscripts and editions cited in the FHG as sources of fragments.³⁴ This data can be also visualized in the Witnesses Map, the Witnesses (Authors and Works) Charts, and the Witnesses

32 Authors in FHG I don't have geographical epithets, but places have been added in the DFHG because they are known. As for other volumes, missing geographical epithets in the FHG correspond to missing places in the DFHG.

33 Future work will also provide a catalog of fragmentary work titles, which is available in the *index titulorum* of the printed edition. In this case the goal is to extract data concerning these titles by annotating each of their occurrences in the text of the FHG.

34 On the difficulties of attributing literary and geographical epithets to ancient authors and on the issues concerning their chronology, see Berkowitz and Squitieri (1990, xvii–xxii). Links to resources concerning editions, manuscripts and inscriptions are progressively added to the DFHG Witnesses Catalog.

Timeline. These resources complement the printed edition of the FHG, which lacks an index of source texts of the fragments.³⁵

8. Müller-Jacoby Table of Concordance. This tool finds correspondences between fragmentary historians published in the FHG and in the FGrHist, including the *continuatio* and the BNJ. Given that Jacoby Online is a work in progress, as soon as new BNJ authors are published they are also included in the DFHG table of concordance.³⁶ Users can search the database by FHG, FGrHist, and BNJ (1 and 2) data. Results display, in addition to information from the DFHG and Jacoby Online, corresponding data in other editions of Karl Müller related to the FHG and links to the Perseus Catalog.³⁷ This table of concordance complements the FGrHist and Jacoby Online, which offer incomplete or abstent correspondences to FHG authors.³⁸ The goal is to go beyond these collections and generate expanded catalogs of Greek fragmentary historians with corresponding data from printed and digital editions.
9. Text Reuse Detection. The DFHG project offers experimental text reuse functionalities for automatic text reuse detection of FHG authors in their witnesses. Users can insert XML file URLs or select one of the PerseusDL / Open Greek and Latin editions available in the DFHG.³⁹ Results display quotations and text reuses of FHG authors within their source texts. The DFHG allows scholars to download complete XML files of the source texts of the fragments with *dfhg* attributes that mark up the presence of DFHG text reuses in the relevant passages of the source texts. DFHG text reuse detection is based on the Smith-Waterman algorithm that performs local sequence alignment to detect similarities between strings.⁴⁰

35 The need of complete indices of source texts of historical fragments has been shown by Bonnechère (1999).

36 I'm very grateful to the team working on Jacoby Online for sending me updates about new published authors.

37 Only for corresponding authors in the FHG, FGrHist, and BNJ. More information is available in the homepage of the table of concordance.

38 The FGrHist has incomplete *Konkordanzen*. Jacoby Online doesn't include correspondences with authors in the FHG.

39 PerseusDL is the Perseus Digital Library collection of Greek and Latin texts. OGL is the Open Greek and Latin collection, which includes also the Free First Thousand Years of Greek texts (see the paper by Leonard Muellner in this volume).

40 For an overview of the Smith-Waterman algorithm, see https://en.wikipedia.org/wiki/Smith-Waterman_algorithm (last access 2019.01.31).

The Digital Athenaeus: annotation of text reuse entities

The Digital Athenaeus is a project that provides scholars with experimental tools for accessing the text of the *Deipnosophists* of Athenaeus of Naucratis and getting information about citations of authors and works that are preserved in it.⁴¹ The reason for choosing this work is due to its importance as a rich collection of text reuses (*fragmenta*) of ancient Greek authors who belong to many different literary genres.⁴² The *Deipnosophists* offers the opportunity to experiment with a new way of representing fragmentary texts inside their context of transmission, which is the main concern when collecting evidence about reused authors and works. Textual fragments are a form of hypertext and a digital environment permits to annotate and visualize them as reuses within their context. This possibility allows to go beyond the limits of printed editions, where extended chunks of texts conserving *fragmenta* of other texts are extracted, decontextualized, and reprinted in other editions.⁴³

The Digital Athenaeus aims at providing an inventory of authors and works cited by Athenaeus and at implementing a data model for identifying, analyzing, and citing uniquely instances of text reuse in the *Deipnosophists*. This means extracting and annotating a wide variety of elements that pertain to text reuse, such as names of quoted authors, titles and descriptions of quoted works, and in general the language of the text reuse itself. The Greek text of the *Deipnosophists* in the Digital Athenaeus is based on the Teubner edition of Georg Kaibel (1887–1890) and the project is producing tools and services for reading the text and generating text reuse related data that are described in the following pages.⁴⁴

1. Casaubon-Kaibel Reference Converter. This is a tool for finding concordances between the two different reference systems used in the editions of the *Deipnosophists* by Isaac Casaubon (1597) and Georg Kaibel (1887–1890).⁴⁵ This resource is not only helpful for getting the correspondence between passages of the two editions, but most importantly for generating machine readable citations based on Kaibel references, because they are canonical, independent of any particular manifestation of the text, and valid across

⁴¹ <http://www.digitalathenaues.org> (last access 2019.01.31).

⁴² (Berti et al. 2016b).

⁴³ (Almas and Berti 2013).

⁴⁴ Tools and services are available through the homepage of the project with detailed descriptions and instructions.

⁴⁵ On the two systems, see Lenfant (2007).

editions and translations.⁴⁶ Casaubon citations are by definition tied to page-breaks in his particular edition and are therefore not logical. Kaibel citations are based on books and paragraphs corresponding to precise chunks of text and are well suited to a digital environment.⁴⁷ Given that in printed editions scholars traditionally make use of Casaubon citations, the Casaubon-Kaibel Reference Converter automatically converts Casaubon citations into Kaibel citations, in order to create URNs based on the CITE Architecture, as for example `urn:cts:greekLit:tlg0008.tlg001.perseus-grc2:1.4` (= Ath., *Deipn.* 1.4).⁴⁸ In addition to the converter, which also includes links to the printed editions of Casaubon and Kaibel, the tool provides a web API with a JSON output for integrating data into external services.

2. CTS URN Retriever. This tool allows to retrieve and cite paragraphs, passages, and words in the Greek text of the *Deipnosophists*. For example:
 - Ath. *Deipn.* 3.7 corresponds to `urn:cts:greekLit:tlg0008.tlg001.perseus-grc2:3.7`;
 - the second occurrence of the word βιβλου in Ath. *Deipn.* 1.1 corresponds to `urn:cts:greekLit:tlg0008.tlg001.perseus-grc2:1.1@βιβλου[2]`;
 - the quotation of the words of Antiphanes (ἀεὶ δὲ πρὸς Μούσαισι καὶ λόγοις πάρει, ὅπου τι σοφίας ἔργον ἐξετάζεται) in Ath. *Deipn.* 1.4 corresponds to `urn:cts:greekLit:tlg0008.tlg001.perseus-grc2:1.4@ἀεὶ[1]-ἐξετάζεται[1]`.

Each URN is combined with a URL prefix (<http://www.digitalathenaeus.org/tools/KaibelText/index.php#>) to produce stable links for visualizing every citation in the whole text of the *Deipnosophists*, which is browsable by books and paragraphs through a slide in/out navigation menu. The text is based on the edition by Kaibel and each paragraph is connected to the corresponding entries in the *indices scriptorum* of the *Deipnosophists* (see below) and to the OpenNLP POSTagger for Ancient Greek for getting automatic information about the morphology of each word. Using CTS URNs, it is possible to export citations of the *Deipnosophists* down to the word level. The Search tool allows to search the entire text and, when available, results display also inflected forms and lemmata from Morpheus, the Suda On Line, and the Liddell-Scott Lexicon in the CITE Architecture.

⁴⁶ (Berti et al. 2016b, 124–125).

⁴⁷ Every scholar of Athenaeus knows the ambiguity of Casaubon references, because it's difficult to identify with precision the beginning and the end of his paragraphs.

⁴⁸ `tlg0008.tlg001.perseus-grc2` identifies the edition by Kaibel in the Perseus Catalog. See Berti et al. (2016b).

3. Indices Scriptorum. One of the goals of the Digital Athenaeus is to experiment with semi-automatic annotations of data related to text reuse. This is the reason why the project has produced digital versions of indices of authors and works published in the printed editions of the *Deipnosophists* by August Meineke, Georg Kaibel, and S. Douglas Olson. SQL databases of these indices have been created starting from OCR outputs of the printed editions and have been enriched with automatically converted Kaibel references and with links to external resources for reading the whole context of each reference. Dynamic graphs generate graphic visualizations of the indices (Figure 4) and a web API with a JSON output allows to integrate data into external services. These indices offer lists of author names and work titles cited by Athenaeus and they can be considered as already disambiguated lists of named entities (author names and work titles) to be mapped on to the text of the *Deipnosophists* to obtain a first set of annotations pertaining to text reuse.⁴⁹
4. Book Stream. This tool shows an automatic alignment of index entries extracted from the indices by Meineke, Kaibel, and Olson. The resource is based on the alignment of Kaibel references that have been automatically generated by the conversion of Casaubon references included in the printed versions of the indices. Each entry in the book stream is linked to the database of each index. Each paragraph of the *Deipnosophists* is linked to Index to Text, which is an experimental tool based on the Levenshtein distance for producing an automatic alignment of the index entries with their corresponding forms in the Greek text of the *Deipnosophists*.⁵⁰ Given that index entries are in Latin or in English, the Levenshtein distance has to be adjusted to generate the closest possible results between the indices and the Greek text.⁵¹ A further work of manual correction and a comparison with data obtained from named entities extraction (see below) will enable to create a complete and correct alignment.
5. Named Entities Digger and Concordance. These tools allow to search inflected forms of detected named entities (with transliteration) as they appear in the

⁴⁹ This is not the case of the index by Olson, because it includes not only authors but also other personal names. The Digital Athenaeus offers also the index *dialogi personae* by Georg Kaibel, because this is a list of the names of the sophists who participate in the dialogues described by Athenaeus and who actually quote many authors and works.

⁵⁰ For an overview of the Levenshtein distance, see https://en.wikipedia.org/wiki/Levenshtein_distance (last access 2019.01.31).

⁵¹ The threshold can be changed by users in the online version.

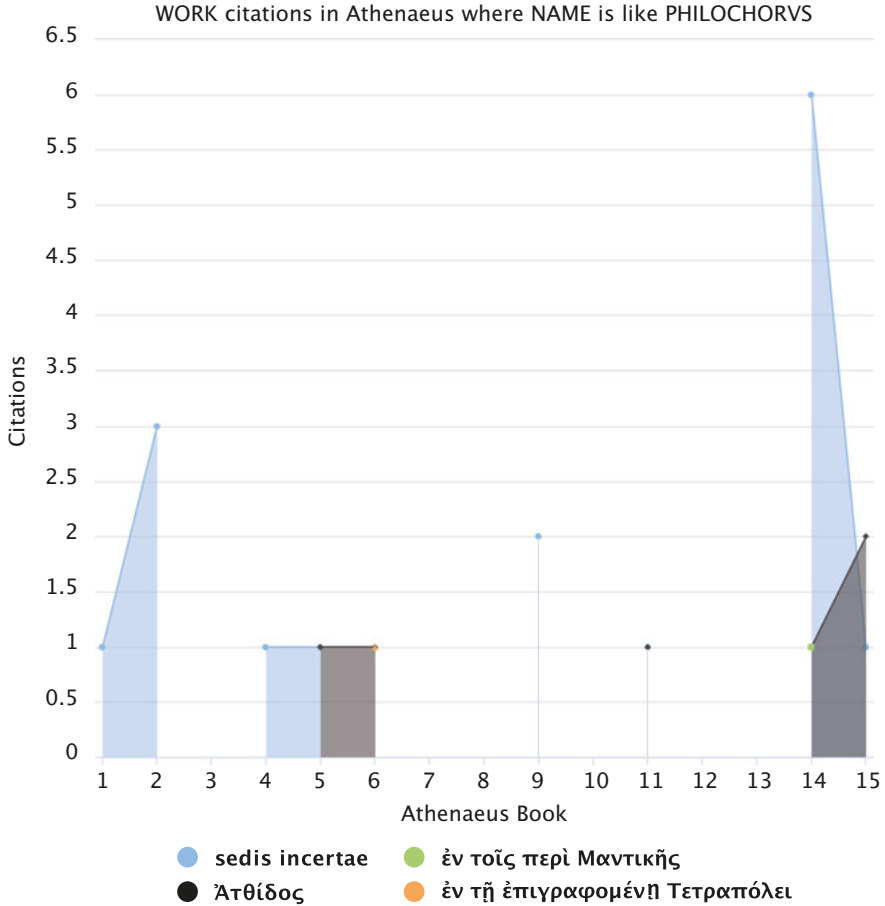


Figure 4: Digital Athenaeus dynamic graph.

Deipnosophists (e.g., Αἰσχύλου [Aischylou]) and visualize their immediate context. The tools are the result of semi-automatic extraction of named entities and are connected to external resources and authority lists: Logeion, the TLG, LGPN, Pleiades, VIAF, an annotated EpiDoc XML file of the *Deipnosophists* (ed. Gulick) in the PerseusDL, and the Index of Ancient Greek Lexica (DC3 – Duke Collaboratory for Classics Computing). Thanks to the lemmatization of detected named entities, it is possible to compare lemmata with the datasets of these external resources and obtain provisional lists of partially disambiguated named entities, such as personal names, place names, constellations, ethnic, festivals, groups, languages, months and titles. Every form of detected named entities

has a CTS URN and, if present, can be visualized in the *indices scriptorum* by Meineke, Kaibel and Olson for further disambiguation.

Conclusion

The two projects described in this paper show how many possibilities the digital environment offers for accessing and analyzing Classical sources that are preserved through quotations and text reuses in later texts. The digitization of Greek and Latin sources is increasing the number of textual data at our disposal, allowing us to work with big quantities of resources in a way that was not possible in a printed world. Language technologies offer techniques and models for accessing these resources, structuring their content, and extracting information from them. Classical fragmentary texts require a further effort to manage challenges and issues concerning their philological ambiguities and complexities. The projects presented in this paper aim at offering a first selection of these challenges, issues, and needs that future generations of scholars will be able to address, expand, and implement.

Bibliography

- Almas, B.; Berti, M. (2013): “Perseids Collaborative Platform for Annotating Text Re-Uses of Fragmentary Authors”. In: DH-Case 2013. Collaborative Annotations in Shared Environments: Metadata, Vocabularies and Techniques in the Digital Humanities. Florence, September 10, 2013. ACM Publication. DOI: 10.1145/2517978.2517986.
- Babeu, A. (2011): Rome Wasn’t Digitized in a Day. Building a Cyberinfrastructure for Digital Classics. Washington, D.C.: Council on Libraries and Information Resources.
- Berkowitz, L.; Squitier, K.A. (eds.) (1990): Thesaurus Linguae Graecae. Canon of Greek Authors and Works. 3rd ed. New York and Oxford: Oxford University Press.
- Berti, M. (2012): “Citazioni e dinamiche testuali. L’intertestualità e la storiografia greca frammentaria”. In: V. Costa (ed.): Tradizione e Trasmissione degli Storici Greci Frammentari II. Tivoli: Edizioni Tored, 439–458.
- Berti, M. (2013): “Collecting Quotations by Topic: Degrees of Preservation and Transtextual Relations among Genres”. *Ancient Society* 43, 269–288. DOI: 10.2143/AS.43.0.2992614.
- Berti, M. (2018): “Annotating Text Reuse within the Context: The Leipzig Open Fragmentary Texts Series (LOFTS)”. In: U. Tischer; U. Gärtner; A. Forst (eds.): Text, Kontext, Kontextualisierung. Moderne Kontextkonzepte und antike Literatur. Hildesheim, Zürich, and New York: Olms, 223–234.
- Berti, M. (2019): “Digital Fragmenta Historicorum Graecorum (DFHG)”. In: Digital Humanities 2019. Book of Abstracts. Utrecht, July 8–12, 2019.

- Berti, M.; Almas, B.; Dubin, D.; Franzini, G.; Stoyanova, S.; Crane, G.R. (2014–2015): “The Linked Fragment: TEI and the Encoding of Text Reuses of Lost Authors”. *Journal of the Text Encoding Initiative* 8. DOI: 10.4000/jtei.1218.
- Berti, M.; Almas, B.; Crane, G.R. (2016a): “The Leipzig Open Fragmentary Texts Series (LOFTS)”. In: N.W. Bernstein; N. Coffee (eds.): *Digital Methods and Classical Studies*. DHQ Themed Issue 10: 2. <http://www.digitalhumanities.org/dhq/vol/10/2/000245/000245.html> (last access 2019.01.31).
- Berti, M.; Blackwell, C.W.; Daniels, M.; Strickland, S.; Vincent-Dobbins, K. (2016b): “Documenting Homeric Text-Reuse in the *Deipnosophistae* of Athenaeus of Naucratis”. In: G. Bodard; Y. Broux; S. Tarte (eds.): *Digital Approaches and the Ancient World*. BICS Themed Issue 59:2, 121–139. <https://doi.org/10.1111/j.2041-5370.2016.12042.x>.
- Berti, M.; Jushaninowa, J.; Naether, F.; Celano, G.G.A.; Yordanova, P. (2016c). “The Digital Rosetta Stone. Textual Alignment and Linguistic Annotation”. In: M. Berti; F. Naether (eds.): *Altertumswissenschaften in a Digital Age: Egyptology, Papyrology and Beyond*. Proceedings of a conference and workshop in Leipzig, November 4–6, 2015. Universität Leipzig: Publikationsserver der Universität Leipzig. <http://nbn-resolving.de/urn:nbn:de:bsz:15-qucosa-201522> (last access 2019.01.31).
- Berti, M.; Stoyanova, S. (2014): “*Digital Marmor Parium*. For a Digital Edition of a Greek Chronicle”. In: S. Orlandi (ed.): *Information Technologies for Epigraphy and Cultural Heritage*. Proceedings of the First EAGLE International Conference. Roma: Sapienza Università Editrice, 319–324.
- Bonnochère, P. (1999): *Die Fragmente der griechischen Historiker*. Indexes of Parts I, II, and III. Indexes of Ancient Authors. Leiden and Boston: Brill.
- Borsuk, A. (2018): *The Book*. Cambridge, MA and London: The MIT Press.
- Celano, G.G.A.; Crane, G.; Majidi, S. (2016): “Part of Speech Tagging for Ancient Greek”. *Open Linguistics* 2:1, 393–399. DOI: 10.1515/opli-2016-0020.
- Grafton, A. (1997): “*Fragmenta Historicorum Graecorum*: Fragments of Some Lost Enterprises”. In: G.W. Most (ed.): *Collecting Fragments. Fragmente Sammeln*. Göttingen: Vandenhoeck & Ruprecht, 124–143.
- Jacoby, F. (1909): “Ueber die Entwicklung der griechischen Historiographie und den Plan einer neuen Sammlung der griechischen Historikerfragmente”. *Klio* 9:9, 8–123.
- Jacoby, F. (2015): “On the Development of Greek Historiography and the Plan for the New Collection of the Fragments of the Greek Historians”. The 1956 Text with the Editorial Additions of Herbert Bloch. Trans. by Mortimer Chambers and Stefan Schorn. *Histos Supplement* 3. Newcastle upon Tyne: Newcastle University.
- Lenfant, D. (2007): “Athénée: Texte et systèmes de référence”. In: D. Lenfant (ed.): *Athénée et les fragments d'historiens*. Actes du colloque de Strasbourg (16–18 juin 2005). Paris: De Boccard, 383–385.
- Martin, T.R.; Berti, M. (2017): “Open Greek and Latin Data for the Challenges of the Fragmentary State of the Primary Sources for the Pentekontaetia”. *Mouseion* (special issue on Open Greek and Latin) 14:3, 409–436.
- Most, G.W. (1997): *Collecting Fragments. Fragmente sammeln*. Göttingen: Vandenhoeck und Ruprecht.
- Most, G.W. (2009): “On Fragments”. In: W. Tronzo (ed.): *The Fragment. An Incomplete History*. Los Angeles: Getty Research Institute, 9–20.
- Müller, K. (1841–1873): *Fragmenta Historicorum Graecorum*. I-V. Paris: Ambroise Firmin-Didot.

- Petitmengin, P. (1983): “Deux têtes de pont de la philologie allemande en France: Le *Thesaurus Linguae Graecae* et la *Bibliothèque des auteurs grecs* (1830–1867)”.
In: M. Bollack; H. Wismann (eds.): *Philologie und Hermeneutik im 19. Jahrhundert*.
Volume 2. Göttingen: Vandenhoeck & Ruprecht, 76–107.
- Strasburger, H. (1977): “Umblick im Trümmerfeld der griechischen Geschichtsschreibung”.
In: *Historiographia Antiqua. Commentationes Lovanienses in honorem W. Peremans septuagenarii editae*. Volume 6. *Symbolae A*. Leuven: Leuven University Press, 3–52.