

Review of Perseus Digital Library

Perseus Digital Library, Gregory Crane (ed.), 1985-2017. <http://www.perseus.tufts.edu/hopper/> (Last Accessed: 10.12.2017). Reviewed by Sarah Lang (Zentrum für Informationsmodellierung (ZIM-ACDH)), sarah.lang (at) uni-graz.at.



Abstract

Perseus Digital Library is a well-established Digital Humanities project providing a library of out-of-copyright editions of canonical world literature with a focus on the Classical languages while expanding its scope beyond its original field. Apart from being a text collection, *Perseus* also provides a plethora of tools for the analysis of its text material and builds upon a sound base of methodological grounding. The digital library is made up of one main collection (Greek and Roman materials) and several smaller sub-collections spanning a range of other subjects. Data acquisition is achieved semi-automatically. The contents of the library consist of out-of-print text editions, dictionaries and commentaries being enriched by the supply of digital tools for analysis. *Perseus Digital Library* is reviewed as a whole but a special focus of the review lies on the Greek and Roman sub-collection and sub-projects.

Introduction

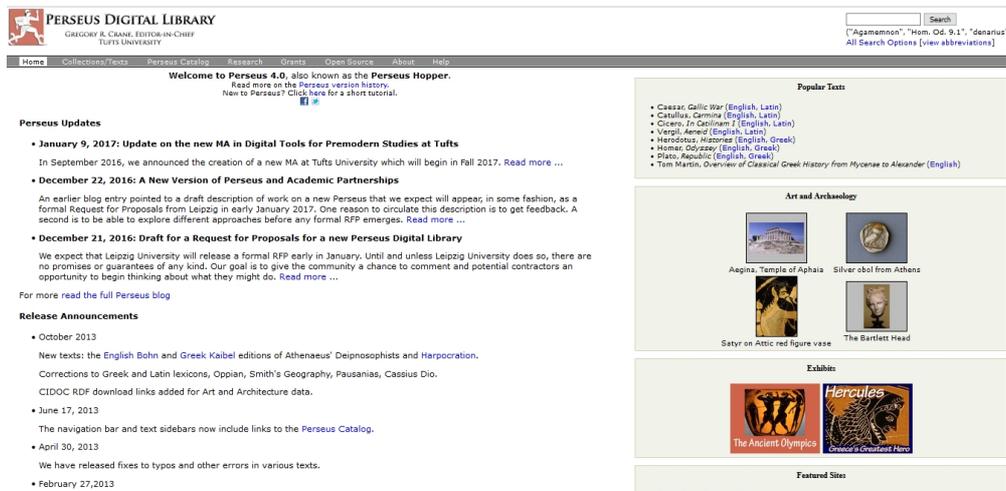


Fig. 1: *Perseus* home page.

1 The *Perseus Digital Library*¹ is a collection of out-of-copyright late 19th century to early 20th century text editions, started at the Department of Classics at Tufts University at Medford, Massachusetts (US). The current version is *Perseus 4.0*, also known as the *Perseus Hopper* (Fig. 1). Formerly known as the ‘Perseus Project’, *Perseus* defines and classifies itself as a digital library that collects primary and secondary text sources.

2 Following in the footsteps of 3rd century BCE scholars at the library of Alexandria, the self-defined mission of *Perseus* is ‘to make the full record of humanity – linguistic sources, physical artifacts, historical spaces – as intellectually accessible as possible to every human being, regardless of linguistic or cultural background’.² *Perseus* thus targets research-driven academic audiences just as much as the general public. Furthermore, its purpose is the promotion of Digital Humanities research as much as the democratization of cultural heritage through improved accessibility. Beside human readable information as provided throughout the various text collections, *Perseus* aims at creating a network of machine-actionable as well as machine-generated knowledge, as they state in their project description.

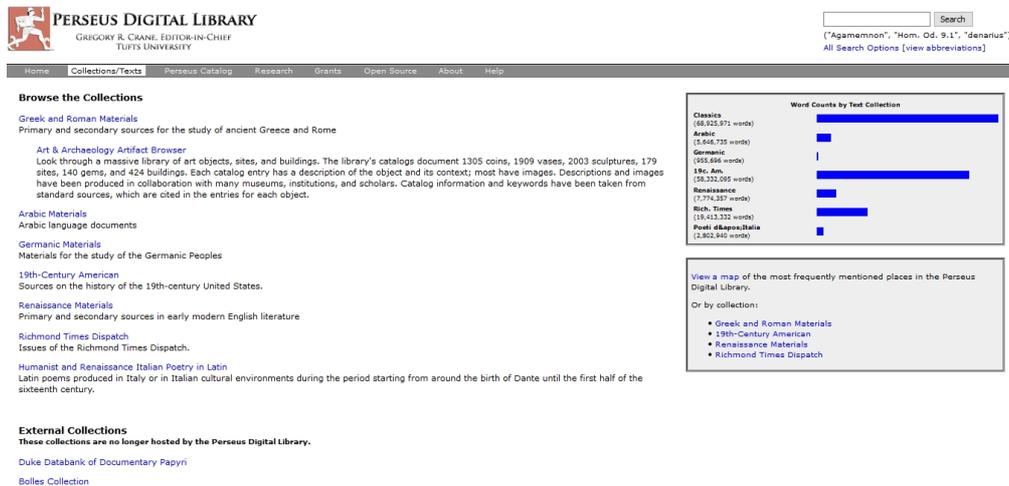


Fig. 2: Overview of the collections.

3 The digital library itself now contains, in addition to the original base of Greek and Roman materials Arabic, Germanic, Renaissance and 19th century American documents, Italian poetry in Latin as well as papyri (Fig. 2). As a part of the digital library, *Perseus* offers a digital collection of museum objects, the *Art & Archaeology Artifact Browser*. The scope of each component is visualized in a 'Word Count per Text Collection' Graph showing that after Classics (68 million words), 19th Century American (58 million words) and Rich Times Dispatch Issues (19 million words) dominate the overall collection. The opening of the project past the Western cultural sphere has begun with the inclusion of Arabic materials.

4 Responsibilities in the project are fairly transparent. Among others, Gregory Crane (Editor –in-Chief), Marie-Claire Beaulieu (Associate Editor), Bridget Almas (Software Developer), Alison Babeu (Digital Librarian) and collaborators from external institutions are mentioned. An extensive overview of past grants and projects documents the projects' resources. Originally started at the Classics Department of Tufts University in 1985, *Perseus* of today features a great magnitude of cooperation partners and extended its scope way past the original field of Classics. The project has collected a number of science and government funds alike as well as private donations. Thus it was able to adopt a broader research agenda in 1998 (Digital Library Initiative Phase 2). Its goal was to create a digital library for the humanities as a whole which would enlarge the digital infrastructure with a mass of non-text materials and side-projects supplying Digital Humanities technologies. A plethora of child projects has sprouted from *Perseus* (such as the *Perseus Catalog*,³ the *Perseids Project*,⁴) and the project now starts to span the world with greater international cooperation and a second base at Leipzig University

6 *Perseus* Editor-in-Chief Gregory Crane stresses that ‘access to the cultural heritage of humanity is a right, not a privilege’ (Crane 2002, 626-63). The project does not want to continue restriction of academic resources to institutions paying expensive fees and thus feed information to a limited audience. *Perseus Digital Library* thus represents a democratization of cultural heritage which is being made accessible to everyone. Already in the early 2000s it disseminated its contents ‘far beyond traditional academia’ (Crane 2002, 626-63). Crane also points out the need for user customization of digital library contents for diverse audiences in general. A logistic problem with this is that the practical use of libraries as places for reading for the public can sometimes conflict with the internal needs of cataloguing and organizing the material (Crane 2002, 626-63). *Perseus Digital Library* is showing good efforts in both user-customization and advanced forms of internal organization.

Corpus Design and Representativeness of the Corpus

7 The purpose of *Perseus* is to make texts in the field of the Humanities available online in a very complete way. The criteria for text selection seem to be the popularity and the (canonical) status of the texts. It is not specified, however, whether works largely ignored to this day are going to follow, such as, for example, early modern Latin literature that has not been re-edited ever since its beginnings. *Perseus* so far provides texts which were edited in the 19th century and beyond and which are in circulation. *Perseus* does not represent a library of rare works, but rather of commonly found ones in specialized academic libraries. Seeing that the canonical versions of ancient texts, sectioned according to today’s accepted order were established mostly in the 20th century, it can sometimes be problematic that *Perseus Digital Library* uses old text versions from the late 19th century. While certain text versions have undergone thorough corrections and intense research since the late 19th century, others have stayed virtually the same since then. Also, the exact wording isn’t perceived as essential for every research purpose. In these cases, the 19th century version remains the current accepted version until now. In the case of these not frequently re-edited, less frequently read ones especially – which might for that reason be more difficult to acquire and less available for a potential non-academic user (such as Dionysius of Halicarnassus, for example) –, *Perseus Digital Library* is a very valuable tool. *Perseus Digital Library*’s selection criteria are the epoch and the representativeness of the texts in the sense of being part of canonical ‘world literature’.

8 If the goal of *Perseus* is to provide a complete Digital Library for the Humanities, they have of course not yet reached a high degree of completeness. *Perseus Digital Library* remains a work in progress and further development can be expected. Compared to other text collection projects,¹¹ however, *Perseus* is an extraordinarily broad and ‘complete’ collection. In terms of representativeness, *Perseus* provides a good representative sample of Classical texts, which after all was the primary purpose of the now very much expanded collection. While being fairly representative of the Classical world, it has to be noted that the term ‘representative’ operates within the scope of the canonical Classical literature having been the focus of research until today. However, it is less balanced towards less frequently read works. Therefore the question of whether *Perseus* is a well-balanced text collection remains unsolved and depends largely on the definition of what is chosen as a point of reference for what we consider as *the corpus* of Classical language literature in the whole.

9 The texts *Perseus* offers outside its Classical collection cannot yet be considered representative or balanced within the ideological framework of a library for the Humanities in general. However, certain sub-collections are representative of their sub-fields,¹² such as the *19th century American materials*, the *Richmond Times Dispatch* and the *Humanist and Renaissance Italian Poetry in Latin* collection as their scopes are defined in a more narrow way. The Arabic collection contains only the Quran, the general *Renaissance materials* section contains 7 million words and well-known works but yet can be considered neither representative nor balanced nor complete. The *Germanic* collection though seems already fairly well developed considering the fact that this one is probably the globally least widely studied topic of the selection *Perseus* provides so far.

10 The *Perseus* collection is dynamic and growing, though the Classical literature part might be considered fairly complete.¹³ It has, however, to be stressed that the *Perseus* project already now offers three sub-collections that can claim relative completeness, balance and representativeness in a general sense. *Perseus* is very extensive and even very complete depending on the definition of completeness. It is rich in content, functionalities and sub-projects. It has reached sufficient maturity and consistency to serve as a base for further research and the integration of further sub-collections.

Usage scenarios

11 The Roman and Greek materials collection could be defined as a digital access to late 19th, early 20th century editions of Latin and Greek primary texts where each segment can be viewed in the context of its translations and dictionary entries. This provides a good base for studying the primary texts and probably the best one can possibly get without bigger copyright issues or more generally, for free. However, it has to be mentioned that the critical apparatus from the original editions is not contained in *Perseus*. Depending on the research interests in working with the material, the reused editions might not be sufficient in every case. From personal interviews, I received the feedback that scholars of ancient history see no problem whatsoever in citing *Perseus* texts in scientific publications, whereas Classics scholars with interest in literary texts insist that the provided editions are too old to be used in a philological publication, especially as the main portion of material deals with well-researched texts where there would be newer, more text critical editions available. For Classical philology, the *Perseus Digital Library* is a great study tool to be used in learning and teaching ancient languages but not of sufficient quality in order to do qualitative research on a certain text. An exception to this are the treebanks and linguistic features of course, although they provide a very specific type of research opportunity where they are immensely useful for quantitative research and the analysis of certain recurring grammatical structures. This is a very specialized branch of Classical philology where the *Perseus* project is a scientific pioneer but this cannot be said of Classics as a whole discipline.

Related projects

12 The Perseids side-project provides a platform for ,collaborative editing, annotation, and publication of born-digital editions of source documents in the classics' coupling different tools and services, inspired among others by the Harvard Homer Multitext Project which has laid the foundations for the Canonical Text Services (CTS) protocol, of which the CTS URNs are a part (Berti et. al. 2014, 3; Smith 2009). The CTS is built upon FRBR metadata. It allows for the citation of single verses and serves as the link between Perseus Catalog and the TEI XML based textual content of the Perseus Digital Library.

13 The *Leipzig Open Fragmentary Texts Series* (LOFTS) is developed in conjunction with *Perseus Digital Library*. Texts from fragmentary Greek historians in the LOFTS

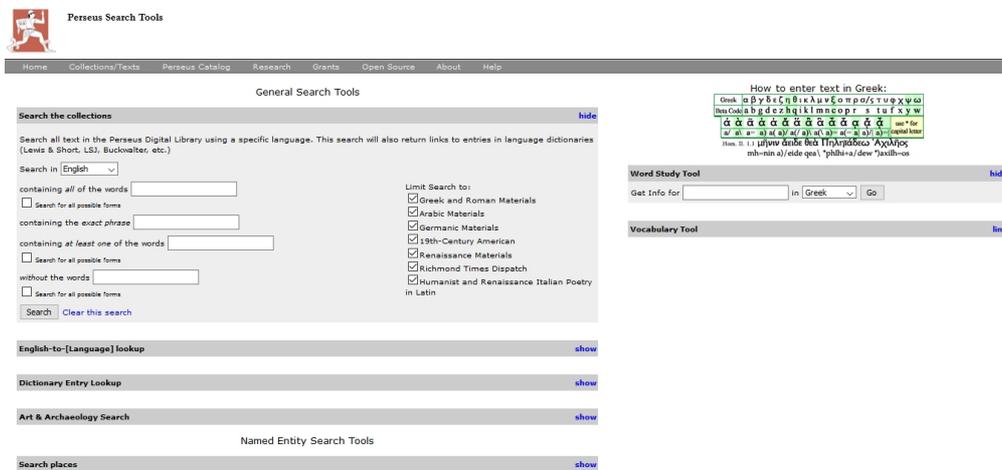


Fig. 5: View of the Perseus search, at the left-hand side, word study and vocabulary tools.

15 *Perseus* offers computationally processable text and specifies the creation of machine-readable and even machine-created knowledge as core goals. Engagement with the text of the Roman and Greek materials happens mostly in sub-projects or associated projects and not so much in the main library where we primarily find a broad text base that is provided book-wise or chapter-wise. It is integrated in a user interface with visualization and analysis tools and as XML documents with basic, mostly not semantically very rich, mark-up. Every word contains a link to its respective dictionary entry which probably is the greatest innovation, though these annotations are all stand-off and not part of the XML base of data (Fig. 4). The link will trigger a dictionary search, not the word itself is annotated or referenced. This feature makes of *Perseus* a text collection following the digital paradigm in the way that it links different materials, makes them searchable (Fig. 5) and machine-processable, even while it sticks closely to the form of the books which served as a text base.

16 All mark-up and annotations are created automatically which explains the not very rich mark-up in older generation texts, but shows impressive results in newer parts of the collection. A good example for this is the mark-up of historical names and places in the 19th century American materials section. Stable identifiers are available in the form of four different URIs (text, citation, work, catalog record) containing URNs. The choice to use a TEI-compliant SGML format was meant as a commitment to long-term usability. Ancient Greek is stored as beta code. Maybe Unicode would be a better option here. One regrettable fact regarding the XML representation is the lack of a TEI-header containing bibliographical information and metadata about the respective source. Although

admittedly this information can be found in the *Perseus Catalog*, metadata and transcription of a source should be stored together in a single file.

17 The actual text presentation is segmented in books, chapters and sections – full text versions are not available due to copyright reasons. The segments however include notes, a bibliographic reference of the resource, display preferences, a vocabulary tool, translations, commentaries and different versions of the text. The segments are enhanced further, if available, with stable identifiers, a Creative Commons licence and a TEI XML version of varied quality.

User Interface

18 The text collection provides a user interface where segments of primary texts can be viewed side by side with their translations and commentaries (if available). Apart from that, the word study tool and thereby dictionary entries are accessible. Notes from the used edition are also represented. The *Perseus* text collection engages with its texts in various ways but texts are not associated with their material carriers – *Perseus* is in that respect a library following the digital paradigm as opposed to a digitized library focusing on facsimiles or text presentation of the base documents.

19 Works (original and related) can be found via the list of authors as well as the simple (metadata, not full text) search. Once one version of a text is found, one can comfortably switch to related versions or access commentaries and the like. The simple search does not always yield results straightaway – if for example the user does not enter the English naming conventions for ancient names and titles –, and so it can sometimes be difficult to find the text one was looking for. The ‘view abbreviations’ link does help here though.

20 The help section features a *Perseus 4 Quick Start Guide*, information on copyright and reuse as well as an FAQ page last updated in May 2016. It also suggests starting points for the exploration of the platform as well as help with the texts, the vocabulary tool and *Perseus* search. With the lookup tool, *Perseus* provides a map to its own contents, enabling users to search topics they are interested in.

21 The user interface combines the view of the primary text with study tools such as translations, comments and the word study tools even if the project description stresses the fact that *Perseus Digital Library* focuses on their own Digital Humanities research

over services for an audience. *Perseus* does not provide introductory texts but it provides old commentaries to primary sources. Further introductions of the *Perseus* editors to the text collection could, however, still improve the overall impression and usability of resource.

Vocabulary Tool						hide
Currently showing the top 50% of words sorted by weighted frequency. Study all vocabulary in this passage.						
Word	Definition	Key Term Score	Min Freq	Max Freq	Weighted Freq	
et	and	0,0018	3	3	3	
ille	that	0,0029	1	1	1	
condo	to put together, make by joining, found, establish, build, settle	0,0118	1	1	1	
ob	towards, to	0,0193	1	1	1	
unde	Relat., from which place, whence	0,011	1	1	1	
vis	strength, force, vigor, power, energy, virtue	0,0063	1	1	1	
infero	to bring in, introduce, bring to, carry in	0,0193	1	1	1	
deus	a god, deity	0,0044	1	1	1	
urbs	a walled town, city	0,0068	1	1	1	
Juno	the planet Venus	0,0269	1	1	1	
Roma	Rome, the mother city	0,0165	1	1	1	
terra	the earth	0,0058	1	1	1	
ira	anger, wrath, rage, ire, passion, indignation	0,012	1	1	1	
atque	and	0,0036	1	1	1	
venio	to come	0,0039	1	1	1	
saevus	raging, mad, furious, fell, fierce, savage, ferocious	0,0145	1	1	1	
primus	the first, first	0,0053	1	1	1	
Italia	adapted to the climate of Italy	0,0225	1	1	1	
ab	all the way from	0,003	1	1	1	
profugus	that flees, fugitive, in flight	0,0336	1	1	1	
litus	to labor in vain, take useless pains	0,0143	1	1	1	
Albanus	the peace between the Romans and Albans	0,044	1	1	1	

Fig. 6: Word count tool result for the beginning of Vergil's Aeneid.

22 The word study tool provides statistical word counts ([Fig. 6](#)). Visualization is provided with word counts, but user empowerment in the form of personalization is provided only by the ability of adding additional materials to the primary texts by adding them to the view.

23 Navigating *Perseus* in its full complexity is not intuitive and does take some getting used to. This is partly due to the vastness and complexity of the collection but it still could be improved by a clearer layout. The visual patterns of *Perseus* also seem a bit old-fashioned and – while the project stresses the research focus over the presentation for the audience – it still is one of the biggest and best funded Digital Humanities projects out there and some more effort on the visual side would be desirable. While functional, the layout diminishes usability in parts, and especially the organization of the main

navigation containing the ‘About’ sections is very confusing. It seems redundant and is not intuitively logical. A user new to *Perseus* will have a very hard time finding orientation and will probably overlook good parts of the project trying to get to the desired content as fast as possible.

24 *Perseus* does not encourage accidental findings by browsing very much. The fact that one finds ‘news’ from 2009 on the page for current research is a bit irritating, especially as the project is in fact still very active today. It has to be admitted though that the publications section is perfectly up-to-date and provides the newest research papers on and by *Perseus*. But as even the sections supposed to explain the project are a bit confusing and although one usually finds what one was looking for in the end, there definitely is room for improvement.

Provision and sustainability

25 XML documents can be downloaded for single text segments, automated downloading is not permitted. *Perseus* offers download packages for their public materials and does not allow automated downloading so as not to endanger restricted items, such as images under reuse from external rights holders that *Perseus* offers in the *Artifact Browser*. Every single text segment has its own Creative Commons rights information indicating the conditions of reuse. But not all texts contain this information and, wherever it is not available, XMLs aren’t for download either. There is no other way of (automated) downloading such as an API.¹⁵ But each text as well as bundles of text (sub-collections) can be downloaded.¹⁶

26 In terms of reliability and sustainability, the *Perseus* project is convincing by its long history and the resulting vast amount of experience. The fact that the project has now spread from its base at Tufts to a whole network of institutions vouches for more sustainability. The *digital library* is also mirrored by the University of Chicago.¹⁷ In terms of long-term preservation, a Fedora backend is supposed to facilitate data management and thus – according to *Perseus*’ self-description – provide the base for good long-term preservation.¹⁸

27 The *Perseus Hopper* (the current version, released in 2007) is available as Open Source Java code that has been built with extensibility in mind, but the site also states major restructuring work since 2016 that is supposed to result in a completely new version.¹⁹

Conclusion

28 *Perseus Digital Library* is a well-established Digital Humanities project providing a library of out-of-copyright editions of canonical world literature with a focus on the Classical languages. Furthermore, it is expanding its scope way beyond its original field of Classics. *Perseus Digital Library* has lots of goals just as impressive as they are noble and shows great productivity in achieving them, thus creating – step by step – a digital library of and for the Humanities as well as humanity as a whole.

Suggestions for improvement

29 The Vocabulary tool already provides great help in language education but it would be nice to have a more print-friendly output available for the results of the word count tool as to make use of it for vocabulary building. It might also be interesting to provide a print-friendly option that includes a chosen segment of text and its vocabulary list which could serve as an automated work sheet-generator in teaching. Knowing from personal experience that *Perseus* is already now established as an irreplaceable tool in teaching Classical languages, I think that this is a great functionality and field of use of the library and am happy to see that this aspect is already being extended in the *Leipzig Open Greek and Latin Project*.

30 Text selection criteria of such an important library do actively judge the texts it provides and de-value those not chosen for publication in this digital library of the Humanities. It has to be noted that so far *Perseus* is a collection of canonical world literature mostly and so reinforces the paradigm of a canonical world literature. Works which do not fall into this narrowly defined scope remain endangered as a work not re-edited is prone to getting lost over time. *Perseus* could and should make a step in the direction of active preservation of less read, ‘rare’ texts by making room in its selection criteria for works that might be considered less representative but which are in danger of being lost or forgotten. Adopting the archival practise of including random samples might be an approach worth considering.²⁰ Still, I encourage the project to broaden its definition regarding the question what the cultural heritage of humankind is, what the responsibility of an influential project like *Perseus Digital Library* could be in that context, and to reflect on whether they really want to ‘write the history of the winners’ only.

Synopsis

31 The amount of material already available in *Perseus Digital Library* is vast; sub-collections can claim relative completeness and representativeness. Apart from the textual material made available, *Perseus* excels in its initiatives to provide tools for deep analysis of these texts. The project can already now look back on 30 years of experience and a wealth of ground-breaking achievements and publications in the promotion and theoretical grounding of Digital Libraries for the Humanities. *Perseus* remains one of the most interesting projects in the Digital Humanities and has been so for a long time. The project has had a great methodological impact on the discussion of best practises for the Digital Humanities and has been a pioneer of the field.

32 To sum it all up: *Perseus Digital Library* is a big and important Digital Humanities text collection project with lots of methodological background research, a good choice of texts available, and provides a very well-known and established online resource for learners and academics alike. The project involves a plethora of very interesting functionalities and data. *Perseus* already now is a most powerful tool and any further development is highly anticipated.

Notes

1. To be found at: <http://www.perseus.tufts.edu/hopper/>, archived link: <https://web.archive.org/web/20171210180948/http://www.perseus.tufts.edu/hopper/>.

2. See <https://web.archive.org/web/20171210181632/http://www.perseus.tufts.edu/hopper/research>.

3. <https://web.archive.org/web/20171210181722/http://catalog.perseus.org/>.

4. <https://web.archive.org/web/20171210181806/http://sites.tufts.edu/perseids/>.

5. https://perseusdl.github.io/treebank_data/ [Accessed 02.10.2017].

6. <https://web.archive.org/web/20171210182005/http://www.dh.uni-leipzig.de/wo/lofts/>.

7. <https://web.archive.org/web/20171210182145/http://www.dh.uni-leipzig.de/wo/projects/open-greek-and-latin-project/>.

8. <https://web.archive.org/web/20171210182305/http://www.dh.uni-leipzig.de/wo/open-philology-project/open-persian/>.

9. See <https://web.archive.org/web/20171210182441/http://www.perseus.tufts.edu/hopper/research/current> , <https://web.archive.org/web/20171210182638/http://www.perseus.tufts.edu/hopper/research/background>.

10. <https://web.archive.org/web/20171210182638/http://www.perseus.tufts.edu/hopper/research/background>.

11. See other projects reviewed in the RIDE digital text collection issue.

12. As far as the author of this review is qualified to judge. The Digital Library however does not explain how they understand completeness and what degree of representativeness they want to reach in the end.

13. Even though it lacks certain works of even Aristotle and Plato, probably no other similar text collection can claim Perseus' degree of completeness and representativeness.

14. <https://archive.org/> [Accessed 02.10.2017].

15. <https://web.archive.org/web/20171210184220/http://www.perseus.tufts.edu/hopper/help/faq>.

16. <https://web.archive.org/web/20171210184603/http://www.perseus.tufts.edu/hopper/opensource/download>.

17. To be found at: <https://web.archive.org/web/20171210184634/http://perseus.uchicago.edu/>.

18. <https://web.archive.org/web/20171210184755/http://www.perseus.tufts.edu/hopper/research/current>.

19. <https://web.archive.org/web/20171210184809/http://www.perseus.tufts.edu/hopper/opensource>.

20. This sample could, for example, be taken from the libraries and institutions providing the base (text) material.

References

- Berti, Monica, Bridget Almas, David Dubin, Greta Franzini, Simona Stoyanova and Gregory R. Crane. 2014. "The Linked Fragment: TEI and the Encoding of Text Reuses of Lost Authors". *Selected Papers from the 2013 TEI Conference, Journal of the Text Encoding Initiative* 8.
- Celano, Giuseppe G. A.. 2014. "A computational study on preverbal and postverbal accusative object nouns and pronouns in Ancient Greek". *The Prague Bulletin of Mathematical Linguistics* 101: 97–110.
- Crane, Gregory. 2002. "Cultural Heritage Digital Libraries: Needs and Components". *Proceedings of the 6th European Conference on Research and Advanced Technology for Digital Libraries*. London: 626-63.
<http://www.perseus.tufts.edu/~ababeu/ecdl2002.pdf>.
- Crane, Gregory (ed.). 1985—2017. *Perseus Digital Library*,
<https://web.archive.org/web/20171210180948/http://www.perseus.tufts.edu/hopper/>.
- Rydberg-Cox, Jeff. 2011. "Social Networks and the Language of Greek Tragedy". *Journal of the Chicago Colloquium on Digital Humanities and Computer Science* 1/3.
- Smith, David A. and Gregory Crane. 2011. "Disambiguating Geographic Names in a Historical Digital Library?" *Proceedings of the 5th European Conference on Research and Advanced Technology for Digital Libraries, September 04–09*. London: 127–136.
- Smith, Neel. 2009. "Citation in Classical Studies". *Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure* 3/1. Accessed 02.10.2017.
<https://web.archive.org/save/http://www.digitalhumanities.org/dhq/vol/3/1/000028/000028.html>.

Factsheet

Resource reviewed	
Title	Perseus Digital Library
Editors	Gregory Crane
URI	http://www.perseus.tufts.edu/hopper/
Publication Date	1985-2017
Date of last access	10.12.2017

Reviewer	
Surname	Lang
First Name	Sarah
Organization	Zentrum für Informationsmodellierung (ZIM-ACDH)
Place	Graz
Email	sarah.lang (at) uni-graz.at

General Information		
Bibliographic description	Can the text collection be identified in terms similar to traditional bibliographic descriptions (title, responsible editors, institution, date(s) of publication, identifier/address)? (cf. Catalogue 1.1)	yes
Contributors	Are the contributors (editors, institutions, associates) of the project documented? (cf. Catalogue 1.3)	yes
Contacts	Is contact information given? (cf. Catalogue 1.4)	yes
Aims		
Documentation	Is there a description of the aims and contents of the text collection? (cf. Catalogue 2.1)	yes
Purpose	What is the purpose of the text collection? (cf. Catalogue 2.2)	General purpose
Kind of research	What kind of research does the collection allow to conduct primarily? (cf. Catalogue 3.1.8)	Quantitative research

Self-classification	How does the text collection classify itself (e.g. in its title or documentation)? (cf. Catalogue 2.3)	Digital Library
Field of research	To which field(s) of research does the text collection contribute? (cf. Catalogue 2.2)	History, Literary studies, Art history, Archaeology
Content		
Era	What era(s) do the texts belong to? (cf. Catalogue 2.5)	Classics, Medieval, Early Modern, Modern
Language	What languages are the texts in? (cf. Catalogue 2.5)	Arabic, English, Greek, Italian, Latin, other: Germanic
Types of text	What kind of texts are in the collection? (cf. Catalogue 2.5)	Literary works, Newspaper/journal articles
Additional information	What kind of information is published in addition to the texts? (cf. Catalogue 2.5)	Commentary, Context material
Composition		
Documentation	Are the principles and decisions regarding the design of the text collection, its composition and the selection of texts documented? (cf. Catalogue 3.1.1-3.1.3)	no
Selection	What selection criteria have been chosen for the text collection? (cf. Catalogue 3.1)	Epoch
Size		
Texts/records	How large is the text collection in number of texts/ records? (cf. Catalogue 3.1.4)	> 1000
Tokens	How large is the text collection in number of tokens? (cf. Catalogue 3.1.4)	> 10 Mio.
Structure	Does the text collection have identifiable sub-collections or components? (cf. Catalogue 3.1.5)	yes
Data acquisition and integration		
Text recording	Does the text collection record or transcribe the textual data for the first time? (cf. Catalogue 3.1.6)	no

Text integration	What kind of material has been taken over from other sources? (cf. Catalogue 3.1.6)	Full texts
Quality assurance	Has the quality of the data (transcriptions, metadata, annotations, etc.) been checked? (cf. Catalogue 3.1.7)	yes
Typology	Considering aims and methods of the text collection, how would you classify it further? For definitions please consider the help-texts. (cf. Catalogue 3.1.8)	General purpose collection, Corpus
Data Modelling		
Text treatment	How are the textual sources represented in the digital collection? (cf. Catalogue 3.2.1)	Edited text, Translated text
Basic format	In which basic format are the texts encoded? (cf. Catalogue 3.2.4)	XML
Annotations		
Annotation type	With what information are the texts further enriched? (cf. Catalogue 3.2.2)	Semantic annotations
Annotation integration	How are the annotations linked to the texts themselves? (cf. Catalogue 3.2.2)	Embedded, Stand-off
Metadata		
Metadata type	What kind of metadata are included in the text collection? (cf. Catalogue 3.2.3)	Administrative
Metadata level	On which level are the metadata included? (cf. Catalogue 3.2.2)	Whole collection
Data schemas and standards		
Schemas	What kind of data/metadata/annotation schemas are used for the text collection? (cf. Catalogue 3.2.4)	General standardized schema
Standards	Which standards for text encoding, metadata and annotation are used in the text collection? (cf. Catalogue 3.2.4)	TEI
Provision		
Accessibility of the basic data	Is the textual data accessible in a source format (e.g. XML, TXT)? (cf. Catalogue 4.1)	yes

Download	Can the entire raw data of the project be downloaded (as a whole)? (cf. Catalogue 4.2)	yes
Technical interfaces	Are there technical interfaces which allow the reuse of the data of the text collection in other contexts? (cf. Catalogue 4.2)	General API
Analytical data	Besides the textual data, does the project provide analytical data (e.g. statistics) to download or harvest? (cf. Catalogue 4.3)	no
Reuse	Can you use the data with other tools useful for this kind of content? (cf. Catalogue 4.4)	yes
User Interface		
Interface provision	Does the text collection have a dedicated user interface designed for the collection at hand in which the texts of the collection are represented and/or in which the data is analyzable? (cf. Catalogue 5.1)	yes
User Interface questions		
Usability	From your point of view, is the interface of the text collection clearly arranged and easy to navigate so that the user can quickly identify the purpose, the content and the main access methods of the resource? (cf. Catalogue 5.3)	yes
Access modes		
Browsing	Does the project offer the possibility to browse the contents by simple browsing options or advanced structured access via indices (e.g. by author, year, genre)? (cf. Catalogue 5.4)	yes
Fulltext search	Does the project offer a fulltext search? (cf. Catalogue 5.4)	yes
Advanced search	Does the project offer an advanced search? (cf. Catalogue 5.4)	yes
Analysis		
Tools	Does the text collection integrate tools for analyses of the data? (cf. Catalogue 5.5)	yes

Customization	Can the user alter the interface in order to affect the outcomes of representation and analysis of the text collection (besides basic search functionalities), e.g. by applying his or her own queries or by choosing analysis parameters? (cf. Catalogue 5.5)	yes
Visualization	Does the text collection provide particular visualizations of the data? (cf. Catalogue 5.6)	no visualization
Personalization	Is there a personalisation mode that enables the users e.g. to create their own sub-collections of the existing text collection? (cf. Catalogue 5.7)	no
Preservation		
Documentation	Does the text collection provide sufficient documentation about the project in general as well as about the aims, contents and methods of the text collection? (cf. Catalogue 6.1)	yes
Open Access	Is the text collection Open Access? (cf. Catalogue 6.2)	yes
Rights		
Declared	Are the rights to (re)use the content declared? (cf. Catalogue 6.2)	yes
License	Under what license are the contents released? (cf. Catalogue 6.2)	CC-BY-NC, CC-BY-NC-SA
Persistent identification and addressing	Are there persistent identifiers and an addressing system for the text collection and/or parts/objects of it and which mechanism is used to that end? (cf. Catalogue 6.3)	other: CTS
Citation	Does the text collection supply citation guidelines? (cf. Catalogue 6.3)	yes
Archiving of the data	Does the documentation include information about the long term sustainability of the basic data (archiving of the data)? (cf. Catalogue 6.4)	yes
Institutional curation	Does the project provide information about institutional support for the curation and sustainability of the project? (cf. Catalogue 6.4)	yes
Completion	Is the text collection completed? (cf. Catalogue 6.4)	no
Personnel		

Editors	Gregory Crane Marie-Claire Beaulieu Lisa Cerrato
Programmers	Bridget Almas Shawn Doughty Anna Krohn
Contributors	Alison Babeu Frederik Baumgardt Tim Buckingham