

Project Announcement

Annotating and Editing with Canonical Text Services (CTS) Project funded by the Andrew W. Mellon Foundation: 2016–2017

Christopher Blackwell, Charlotte Schubert

1 Goals

1. We plan to coordinate ongoing work on two specific projects: CITE/CTS and the DFG-funded project eComparatio (<http://ecomparatio.net/>).¹ CITE (Collections, Indices, Texts, and Extensions) is a protocol for identification and retrieval of digital data using concise citations in URN form.² The „Text” component of CITE is the well-developed CTS protocol (Canonical Text Services). CTS allows identification of passages of text using canonical citation:

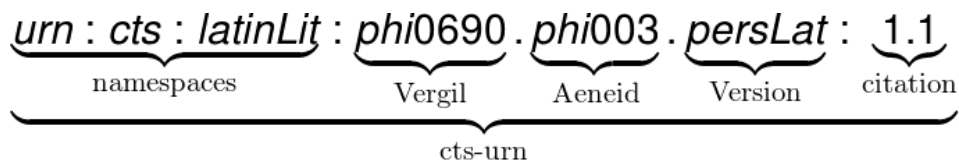


Fig. 1: The CTS-URN captures the semantics of „text” as „an ordered hierarchy of citation objects.”

CTS URN citations effectively capture the semantics of a text, and because CTS specifies a *service* for retrieving passages of text based on URNs, it facilitates comparison of different versions. For example, to compare Book 1, line 1 of two editions of the *Aeneid*, a scholar could simply request the text identified by these two URNs, which differ only in the *version-component*:

urn:cts:latinLit:phi690.phi003.deLaCerde1612:1.1
urn:cts:latinLit:phi690.phi003.ribbeck1860:1.1

eComparatio, developed at the University of Leipzig as part of the eAQUA suite of resources for philological commentary, is an online tool that implements sophisticated algorithmic approaches to analyzing *textual differences*. These may include „vertical differences”, in which one edition of a text has additional passages of text, or „horizontal differences” in which two texts have the same number of citation-units, but differ in the specific content of those citation units, or combinations of both.

¹ Login necessary. For a short description: <http://www.eaqua.net/>.

² W3C/IETF URI Planning Interest Group (21 September 2001). “URIs, URLs, and URNs: Clarifications and Recommendations 1.0”. W3C.

To realize fully the potential of this work, scholars need to be able to do the following:

- To capture canonical citations of passages of text more (or less) precisely than those defined by a text's citation scheme. For example, „The first letter *upsilon* in Book 1, line 2 of the *Iliad* as it appears on Manuscript A,” or „The first word (μῆνιν) of *Iliad* 1.1, and the first word (οὐλομένην) of *Iliad* 1.2, but *nothing in between*, because those two words are the first noun-phrase of the text.
- To capture and cite canonically an *analysis* of passages of text. In other words, if Manuscript M of Thucydes' *History* has at 1.10 οὐν („and so”), but all the other manuscripts have γοῦν („and so, at least”), we not only want to see this difference (which we will find and show with the tool eComparatio) and comment on it (which eAQUA allows), but to capture and *cite* the scholarly object that consists of three things: the differing editions of the text, the specific point of difference, and the scholarly interpretation.

CITE/CTS provides the flexibility for creating these citations; we are proposing to develop workflows that allow eComparatio to process texts based on URN citations, and to express its analytical finding by means of URNs. (Realization: C. Blackwell)

2. The German partners in the project will implement as a use case „Plutarch's *Life of Pericles* as a Network of Sources”, by which the specific benefit with regard to the technical, philological, and historical research will be demonstrated: In combination with the eAQUA tools, the *Life of Pericles* will be annotated with parallel passages, quotations and *testimonia* (as provided by the eAQUA citation tool), and textual variants (as provided by eComparatio), by storing variants and parallels via CITE-URNs, connected to the corresponding CTS-URNs. (Realization: Ch. Schubert)

3. Workshops with colleagues from the different research fields from the USA and from Germany in Leipzig and Greenville, an international exchange of American and German students, joint sections with expert conferences, publications in international journals are to ensure the visibility and dissemination.

2 Analysis, Tokenization, Graphs, and Canonical Citation

Any text is subject to a variety of analyses, such as paleography, morphology, prosody, syntax. Each of these demands a different view of the text, or more precisely, different divisions of the text into tokens. Palaeography views the text as characters and ligatures; syntax views the text as word-tokens; morphology views the text as prefixes, augments, stems, and ending; prosody views the text as syllables, which ignore word-boundaries. These analyses exist in parallel, but influence each other. A paleographic analysis may alter a single character, but in so doing completely change the syntactic graph of a sentence, or the meter of line of poetry.

Using CTS and the concept of *analytical exemplars* we hope to demonstrate precise, machine-actionable canonical citation that admits many different kinds of analysis, and allows us to address specifically conflicting, controversial analyses.

Use case: A text critic argues in favor of one manuscript variant over another; the distinction is a single character. Her argument is based not only on evidence from imagery of two manuscripts, but on a metrical and syntactic analysis of the text. She should be able to cite the regions-of-interest on the manuscript images, but also to *cite* the graphs of syntax, and the metrical analyses, that would result from accepting one variant over another. The citation of the graph of syntax must be different from a citation to the text, but must fully align, at every level, with the text.

Blackwell’s work at Furman will pursue workflows and technologies for canonically citable tokenizations, graphs, and alignment, in support of the work on Plutarch’s *Life of Pericles*.

3 Use-Case: Citation, Quotation, and Parallel Transmission in Plutarch’s *Life of Pericles* with CTS and eAQUA

Main goals

The main goal of this use-case is to show the possibilities of a combination of the citation tool developed during the project „eAQUA“, funded by the BMBF from 2008–2013, eComparatio (DFG funded 4/2014–3/2016) and the CTS/CITE–infrastructure provided by Blackwell and Neil Smith.

Blackwell will work on CTS-workflows for analytical records aligning text fragments at various levels of granularity, which is the heart of a critical apparatus. The eAQUA citation tool provides a full survey of the parallels between any given text and a given reference corpus, e.g. *The Perseus Digital Library*, based on a match of five words or more.³

CTS/CITE provides an efficient option to link the results of an analysis such as provided by the eAQUA citation tool to a text with CTS infrastructure.

eComparatio currently compares a text with specified exemplars – i.e. different editions of it – and generates an *apparatus criticus*. By implementing in eComparatio CTS URNs, which inherently align versions with their notional „work“ – any CTS-URN to a version of the *Iliad* is implicitly aligned with any other version, edition, or translation of the *Iliad* – we can extend the potential for new insights across a much wider corpus of texts.

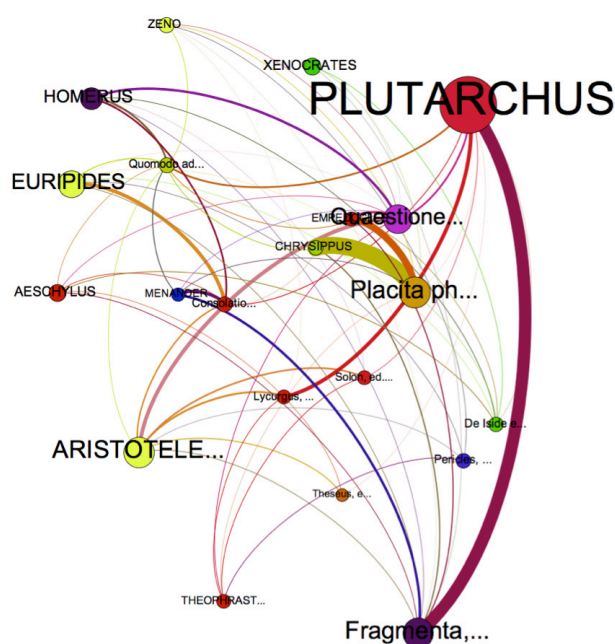


Fig. 2: The sources for Plutarch’s presentation of the *Life of Pericles*, drawn from data in eAQUA.

³ Ch. Schubert: Detailed description of eAQUA search portal, in: Ch. Schubert (Hg.): Neue Methoden der geisteswissenschaftlichen Forschung. Eine Einführung in das Portal eAQUA II (Working Papers Contested Order No.3). Leipzig 2011, ISSN 2191–3641, S. 33–53. eAQUA can draw on a variety of sources, e.g. the Thesaurus Linguae Graecae. In the United States text- and data-mining in non-commercial scientific context is possible without authorization based on the fair use doctrine text- and data-mining in non-commercial scientific context. This has been stated in the issue brief of the Association of Research Libraries as of June 2015. However our efforts will focus on digital libraries with content free from restrictive licensing.

Particular use for the scholarly community in classics and ancient history: diachronic and synchronic evaluation and automated creation of source networks.

Plutarch's *Life of Pericles* qualifies for being a use-case due to its relative brevity, making it a tractable problem. Its subject matter is the central figure of 5th century Athenian democracy in a very well-documented period of the history of Athens, with an explicit comparison to the Roman Fabius Maximus, victor over Hannibal in the Second Punic War. Furthermore, Plutarch is an author well known for his wealth of citations and parallel passages. He has already been the object of several studies based on analyses provided by the eAQUA citation tool.⁴ He also takes centre stage in ongoing scholarly debates about the habits of ancient authors with regard to their use and re-use of sources, citations, and quotations. Until now, metadata is required to create a network of sources. The example (above, created with Gephi) shows the authors which Plutarch quotes most frequently: In the works *Placita*, *Fragmenta*, *Questiones*, *Consolatio* and, most importantly, in the biographies of *Solon*, *Theseus* and *Pericles* especially *Homer*, *Euripides*, *Aeschylus*, *Theophrastus*, *Empedocles*. There are no possibilities yet, to point to such interdependencies on the level of single text passages and to visualize them, neither is there a standard for referencing.

A good example for the importance of studies in the use of sources in Plutarch in particular and the creation of new tools for equivalent surveys in general might be Plutarch's quotation in the *Life of Pericles* 26,4, the chapter, in which he describes the custom of the Athenians to mark their Samian prisoners' foreheads with the symbol of the *Samaina*, a characteristic Samian vessel with big prow, during the war of 441–439 B.C. Plutarch quotes the verse Σαμίων ὁ δῆμὸς ἔστιν ὡς πολυγράμματος („For oh! how lettered is the folk of the Samians!“).⁵ He also mentions that the tyrant Polycrates was the first to build this kind of ships, and that Aristophanes himself referred to the tattooing of the Samian prisoners with his verse.

The eAQUA citation-tool delivers five parallel passages, which show, on the one hand, a network of ancient text-transmission with its different steps of interdependence as well as various contexts, and, on the other hand, completely different appreciations by the modern fragmentary editions. (Those passages are neither treated in Ph. A Stadter's great commentary on the *Life of Pericles*, nor evaluated adequately yet.)

4 Text-Transmission in Antiquity

As one can see in the listing of the parallel passages, the saying has occurred first in the play *Babylonians* by Aristophanes (Hesych s.v. = fr. 64 Kock = 71 Kassel-Austin), of which only fragments have survived. Later, it has been picked up by Aristotle in his *Samian constitution* (fr.575 R). Plutarch, by putting the verse in the context of the Samian war and mentioning Polycrates, evokes a situation twice as shameful for the Samians: marked like slaves, with the sign of a tyrant.

⁴ Ch. Schubert: *Plutarchs Schrift De malignitate Herodoti (with Corina Willkommen)*, in: Schubert, Ch. / Klank, M. (Hrsg.): *Das Portal eAQUA – Neue Methoden in der geisteswissenschaftlichen Forschung III. Working Papers Contested Order No. 7*. Leipzig 2012, ISSN 2191–3641. Ch. Schubert / A Weiß, *Die Hypomnemata bei Plutarch und Clemens: Ein Textmining-gestützter Vergleich der Arbeitsweise zweier ‚Sophisten‘*, *Hermes* 143 (2015), S. 441–471; Ch. Schubert / F. Meins, *Galen und Plutarch*, erscheint in: K.-H. Leven, *Galen aus Pergamon*, (Philosophie der Antike, hrsg. von W. Kullmann) [in print].

⁵ Translation by Bernadotte Perrin.

The treatment of the passage in the ancient lexicography shows wholly different contexts. Photius knows the context of the Samian war and names the *Nostoi* of Lysimachus, the mythographer and grammarian, and the *Histories* of Duris of Samos as his sources. In another passage,

Author (TLG)
0007 - PLUTARCHUS
012 - Pericles

Found : 5

Original

Σαμίον ὁ δῆμος ἐστὶν ὡς πολυγράμματος.

Source : PLUTARCHUS Biogr. Phil. (0007) (Chaeronensis - A.D. 1-2): Pericles
Publication : Pericles, ed. K. Ziegler, *Plutarchi vitae parallelae*, vol. 1.2, 3rd edn. Leipzig: Teubner, 1964: 1-47.
Document citation : //26t/4/8 to //26t/4/10

Quotations

1. Photius <Σαμίον ὁ δῆμος> (ἐστὶν) <ὡς πολυγράμματος>:
Source : ARISTOTELES et CORPUS ARISTOTELICUM Phil. (0086) (Stagirites Pellaeus Atheniensis - 4 B.C.): *Fragmenta varia*
Publication : *Fragmenta varia*, ed. V. Rose, *Aristotelis qui ferebantur librorum fragmenta*. Leipzig: Teubner, 1886 (repr. Stuttgart: 1967)
Document citation : //8t/44t/575-578n/1n to //8t/44t/575-578n/2n
2. Photius Lex. v. [Σαμίον ὁ δῆμος ἐστὶν ὡς πολυγράμματος]:
Source : DURIS Hist. (1339) (Samius - 4-3 B.C.): *Fragmenta*
Publication : *Fragmenta*, ed. K. Müller, FHG 2. Paris: Didot, 1841-1870
Document citation : //59t50-68/111 to //59t50-68/211
3. Σαμίον ὁ δῆμος ἐστὶν ὡς πολυγράμματος:
Source : HESYCHIUS Lexicogr. (4085) (Alexandrinus - A.D. 5/6): *Lexicon (Π_Ω)*
Publication : *Lexicon (Π_Ω)*, ed. M. Schmidt, *Hesychii Alexandrini lexicon*, vols. 3-4. Halle: *n.p., 3:1861; 4:1862 (repr. Amsterdam: Hakkert, 1965)
Document citation : //sigma/150b/3t1 to //sigma/150b/3t1
4. πρὸς ταῦτα τὰ στίγματα λέγουσι καὶ τὸ Ἀριστοφάνειον ἠνίχθαι Σαμίον ὁ δῆμος ἐστὶν ὡς πολυγράμματος [Aristot.]
Source : MELISSUS Phil. (1494) (Samius - 5 B.C.): *Testimonia*
Publication : *Testimonia*, ed. H. Diels and W. Kranz, *Die Fragmente der Vorsokratiker*, vol. 1, 6th edn. Berlin: Weidmann, 1951 (repr. Dublin/Zurich: 1966): 258-267.
Document citation : //3/10 to //3/11 (Schema: Fragment/line) r
5. <Σαμίον ὁ δῆμος <ἐστὶν> ὡς πολυγράμματος>:
Source : PAUSANIAS Attic. (1569) (ca. A.D. 2): *Ἀττικῶν ὀνομάτων συναγωγή* Publication : *Ἀττικῶν ὀνομάτων συναγωγή*, ed. H. Erbse, *Untersuchungen zu den attizistischen Lexika* [Abhandlungen der deutschen Akademie der Wissenschaften zu Berlin, Philosoph.-hist. Kl. Berlin: Akademie-Verlag, 1950]
Document citation : //sigma/3*/1 to //sigma/3*/1

Fig. 3: Text-Transmission in Antiquity

the Athenians (403/2, cf. Ephorus FGrH 60 F106). Hesychius, too, (Σ 150) mentions the Samian war as well as the story about the alphabet, (cf. Fig. 3), whose tradition can be traced back from the Suda to Ephorus.

As we can see, at least three historical contexts of the quotation were known in antiquity: The one about the tyrant Polycrates, the Samian War, and the time after the Peloponnesian war.

Plutarch chooses the context of the Samian war (perhaps according to Duris), just mentions the tyrant Polycrates, and totally neglects the story about the adoption of the alphabet, which symbolizes the long-term good relations between Samos and Athens.⁶ The particularity of Plutarch's version and its own framing, which presents the Athenians from the perspective of the subjugated allies as tyrants, becomes apparent only through the citation-analysis of the parallel passages and their contexts. With the eComparatio tool, those can be visualized and examined for differences and cooccurrences.

⁶ Since he obviously knew the work of Ephorus, he must also have known the story about the alphabet.

5 The fragment in modern editions

Ascription of the fragment: Müller classifies the quote as a fragment of Duris (which seems justified by Photius' statement). Diels sees it as a *testimonium* about the fragments of Melissus, Erbse ascribes it to the atticist Pausanias.

Here, as well, the ascription to different authors leads to different contexts: Müller choses the frame of a historical work, Diels a specifically chronological context (the one of the Samian war, in which the philosopher Melissos was general of the Samians), Erbse sees it as a linguistic testimony.

As one can see from this example, an integration of the parallel passages as provided by the eAQUA citation tool would be highly useful. Performed for the whole *Life of Pericles*, a new quality of a digital *apparatus fontium* could be reached with this kind of diachronic and synchronic evaluation, hence the foundation of a more extensive annotation with CTS in terms both of a historical commentary and a commentary on the history of scholarship.

Autorenkontakt⁷

Prof. Christopher W. Blackwell

Furman University
Louis G. Forgione University Professor of Classics
Email: christopher.blackwell@furman.edu

URL: <http://www.furman.edu/academics/classics/about/Pages/FacultyandStaff.aspx>

Prof. Charlotte Schubert

Universität Leipzig
Historisches Seminar
Lehrstuhl für Alte Geschichte
Email: schubert@uni-leipzig.de

URL: <https://www.gko.uni-leipzig.de/historisches-seminar/seminar/alte-geschichte/professur.html>

⁷ Die Rechte für Inhalt, Texte, Graphiken und Abbildungen liegen, wenn nicht anders vermerkt, bei den Autoren.