



HAL
open science

TEI models for the publication of SSH journals: opportunities, challenges, and firsts steps towards a standardized workflow

Anne Baillot, Julie Giovacchini

► To cite this version:

Anne Baillot, Julie Giovacchini. TEI models for the publication of SSH journals: opportunities, challenges, and firsts steps towards a standardized workflow. 2019. hal-02420483v1

HAL Id: hal-02420483

<https://hal.science/hal-02420483v1>

Preprint submitted on 19 Dec 2019 (v1), last revised 3 Sep 2021 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



TEI models for the publication of SSH journals: opportunities, challenges, and firsts steps towards a standardized workflow

Anne Baillot (Le Mans Université) & Julie Giovacchini (Centre Jean Pépin UMR8230 CNRS-ENS-PSL)

Abstract :

The TEI Guidelines are developed and curated by a community whose main purpose is to standardize the encoding of primary sources relevant for Humanities research and teaching. But there are other communities working with TEI-based publication formats. The first goal of this paper is to raise awareness for the importance of TEI-based scholarly publishing as we know it today. The second goal is to contribute to a reflection on the development of a TEI customization that would cover the whole authoring-reviewing-publishing workflow and guarantee archiving options as solid for journal publications as we now have them for primary sources published in TEI.

Keywords : scholarly publishing, workflow, OJS, digital publications, schemas

The TEI Guidelines are developed and curated by a community whose main purpose is to standardize the encoding of primary sources relevant for Humanities research and teaching. This development, which is rooted in the history of the TEI, gives the community a great strength (the common focus on a research and dissemination goal), but it leads to a lack of recognition for other areas of digital textuality that also use the TEI. We mostly encode primary sources relevant to our research, and concentrate development efforts in their direction. But there are other communities working with TEI-based publication formats.

Looking at the quantity of encoded documents, the production of the TEI community itself is by far not the most numerous among those who currently use the TEI. The best known example for this uneven situation is that of the European Patent Office: all its patents are encoded in TEI, amounting to a total of 200 Mio documents and 2 bn TEI annotations.¹ This kind of magnitude is far from the one achieved by all TEI-based Humanities primary sources put together.

A lack of balance also exists within the Humanities ecosystem, where the TEI is not exclusively used to encode primary sources, but also secondary sources, especially journal publications. In that context, the quantity of TEI-based documents is quite impressive without there being much theoretical or research-oriented development

¹ Information from Laurent Romary, private communication.



made in that field, even though it is a growing one. With TEI publication formats for secondary sources evolving beneath rather than within the TEI community, it remains a relevant issue to consider their adequation with the scholarly evolution of the TEI for primary sources.

This paper draws on the questions that were at the core of Holmes & Romary (2010) and that initiated the creation of the jTEI format, now used among others by the *Journal of the Text Encoding Initiative*². It is not our purpose to discuss this schema, possible amendments to it, nor is it to discuss the respective technical benefits of TEI vs. JATS. We have two distinct goals by drawing on Holmes & Romary (2010). The first goal is to raise awareness for the importance of TEI-based scholarly publishing as we know it today. The second goal is to contribute to a reflection on the development of a TEI customization that would cover the whole authoring-reviewing-publishing workflow and guarantee archiving options as solid for journal publications as we now have them for primary sources published in TEI.

In a first step, we will present a panorama of the use of TEI in SSH scholarly publishing and consider advantages and challenges of using TEI-based formats in scholarly publishing in the Humanities. Which organs are currently doing so, what is the amount of document power that comes together in that area? Why are they using a TEI format, to what end? This paper will then draw on our experience as editors, with the main purpose to initiate a discussion within the community on the role of TEI-based formats for scholarly publishing at large, advocating that they have the potential to be a powerful leverage for the TEI to gain a greater impact on the scholarly community and for Humanities scholars to be better empowered with the dissemination of their own research. We will propose first ideas for encoding a complete publishing workflow of secondary sources that includes the reviewing and copyediting process.

1. Scholarly publishing in TEI: an overview

Scholarly publishing practices are in many ways connected to the existing research infrastructures that are available to researchers and librarians. The development of TEI-based infrastructures and services on the different continents and in the different countries is the result of the personal engagement by members of the TEI community and of the historical development of infrastructures in these different environments. It also has an economic dimension (who is paying how much for which service) that impacts greatly the digital publishing landscape as we know it today.

² As explained in the corresponding wiki page: https://wiki.tei-c.org/index.php?title=JTEI_Authoring_Tools.



The following is an overview over TEI-based journal publications that was gathered mainly by initiating an informal survey on the TEI-L.³

In Europe

The largest actor on a European scale is France, where a long history of political centralization led to the development of national infrastructures and services which allowed to deploy TEI-based publishing formats and compliant platforms on a national scale. France has three main providers for secondary scholarly content that rely on the TEI: OpenEdition, HAL, and Istex.

OpenEdition⁴ is a platform hosting four different services: an Academic calendar announcing events such as conferences and call for papers (calenda), a scholarly blog platform called hypotheses, a scholarly book service (OpenEdition books) and a scholarly journals platform (called revues.org in the past and now simply named journals.openedition). The *Journal of the TEI* is hosted by OpenEdition. Almost all journals published on the journals.openedition platform are TEI-based.⁵ The articles currently published amount to a total of 260.000 documents⁶. The book platform also provides a number of TEI-based documents (currently 114.000)⁷, which adds up to a total of over 350.000 documents, some of them being books, e.g. rather large documents.

Although the TEI files for these articles and books are theoretically available for reuse and can be used for research purposes, there is currently no direct access like a download link or button that would make harvesting easy. When asked about the feasibility of such an endeavor, OpenEdition mentions local legal issues but a clear willingness to support research projects that would require to, for instance, download these files, should such a research project issue a request (which never happened to this day).

The second French infrastructure to be TEI-based is not a platform for scholarly publishing stricto sensu but a publication repository that can be used either for pre-print or for post-print open access publication. Contrarily to OpenEdition, which was initiated by a research project and gained momentum due to its adequation to the needs of the scholarly community and to the national political agenda, HAL⁸ was from its onset a national infrastructure, conceived as a service to the French scholarly

³ See <https://listserv.brown.edu/cgi-bin/wa?A1=ind1908&L=TEI-L#9> (August 22nd, 2019).

⁴ <https://www.openedition.org/>

⁵ The schema documentation is to be found on one of the blogs hosted by the platform: <https://leo.hypotheses.org/10717>.

⁶ Documented on the search interface: <https://search.openedition.org/results?s=&pf=OJ>.

⁷ Current figures available through the search interface, here: <https://search.openedition.org/results?s=&pf=OB>.

⁸ <https://hal.archives-ouvertes.fr/>



community at large, with a specific effort directed towards Humanities disciplines that proved resistant to non-commercial archiving strategies even long after other communities relied on them. Scholars can either register metadata concerning their publication, or additionally archive one or several pre-print versions. It is the metadata that are backed up in a TEI format, potentially allowing to gain a wide overview of French scholarly publications.

The significance of the HAL data has grown over the past year when the French Ministry for Higher Education and Research issued an Open Science Plan that led the main funding agency ANR to require HAL publications of research output they fund from 2018 on⁹. It was followed in this constraining requirement by the national research evaluation organ HCERES, which issued a statement that they would only take into consideration full-text HAL publications when evaluating universities¹⁰. As for the CNRS (a national instance employing researchers and research assistants only) and Inria (national instance for research in informatics), they also require researchers to use HAL for evaluation purposes on lab level and on individual level. At the CNRS, HAL import functionalities are currently being transferred into the internal databases that harvest annual reports.¹¹ Moving to the TEI-based HAL is a political evolution that took some time to be implemented throughout the research ecosystem, but from 2020 there is virtually no single publication produced in France that will not have a TEI set of – at the very least – metadata associated to it. This does not go without protest related to the political control potential this represents, especially to those who are not familiar with the underlying technologies or the values of the community that develop them.

The third French publication organ hosting TEI-based information is Istex¹². Istex is a platform that integrates content from private scholarly publishers (OpenEdition being one of them) with whom national licenses are being negotiated. University Libraries usually contribute a yearly amount to benefit from the deal, that is managed at a national level by a negotiation consortium. Istex contains metadata from publications, but also full-text pdfs, that are being transformed into TEI and integrated to the server. The API is used by University Libraries to retrieve information as well as for bibliometric purposes. Istex currently hosts 23 Mio documents.¹³

⁹ <https://anr.fr/fr/lanr-et-la-recherche/engagements-et-valeurs/la-science-ouverte/>

¹⁰ <https://www.ccsd.cnrs.fr/2019/04/hceres-choisit-la-plateforme-hal-pour-archivage-et-la-diffusion-des-rapports-evaluation/>.

¹¹ Individual researchers are to use HAL imports in the CRAC system (<https://crac.dsi.cnrs.fr/index.php?session=74f55155598a9a8204370673bf3e3b08>) from 2019 on and in the RIBAC system for SSH scholars (<https://www.ribac-shs.cnrs.fr/>) from 2020 on.

¹² <https://www.istex.fr/>.

¹³ See <https://blog.istex.fr/american-chemical-society-ou-acs-desormais-disponible-sur-la-plateforme-istex/>.



Taking all of this together amounts to a considerable mass of documents, but not all of them are available in TEI and not all of them are available in full OA from the onset. There still are accessibility restrictions, while in the case of the examples that were gathered from other countries, all journals are completely TEI-based (metadata and text) and OA.

Germany has a different strategic approach, but it still presents an honorable series of initiatives in the field of TEI-based scholarly publishing. Until recently, Germany relied solely on the good (political and economic) will of its regions, and more specifically of their Universities, to build and maintain the infrastructures that are necessary to hosting reliable publication platforms. A funding scheme for a national infrastructure has recently being negotiated for hosting scholarly data at large in a stable environment.¹⁴ Building such an infrastructure will certainly change the overall approach of scholarly publishing in the middle and long run. At this stage though, it has not been implemented and we still are in a rather fragmented situation, despite the coordination efforts provided by the DARIAH¹⁵ and CLARIN¹⁶ ERICs¹⁷.

Several academic actors in SSH are positioned in the field of TEI-based scholarly publishing. The *Zeitschrift für Digitale Geisteswissenschaften*¹⁸ was created together with the German-speaking branch of ADHO, DHd (Digital Humanities im deutschsprachigen Raum). It focusses on Digital Humanities at large, the articles are mostly in German. 81 articles are currently published, all in TEI-P5.

R-I-D-E¹⁹ does not follow a similar institutional logic in the sense that it is not carried by one scholarly society, but by the virtual institute for documentology and edition sciences, the I-D-E.²⁰ After the institute concentrated for several years on teaching and research, its publication organ R-I-D-E gained momentum over the past years. All of their information is TEI-based and the data can be retrieved easily.²¹

Finally, the *Jahrbuch für historische Bildungsforschung*²² – a publication backed by a solid institution, in this case a Leibniz institute – has also been preparing its turn to a

¹⁴ <https://www.dfg.de/foerderung/programme/nfdi/>.

¹⁵ <https://www.dariah.eu/>.

¹⁶ <https://www.clarin.eu/>.

¹⁷ On ERICs in general, see:

https://ec.europa.eu/info/research-and-innovation/strategy/european-research-infrastructures/eric_en.

¹⁸ <http://www.zfdg.de/>.

¹⁹ <https://ride.i-d-e.de/>.

²⁰ <https://www.i-d-e.de/uber-uns-about-us/>.

²¹ See <https://github.com/i-d-e/ride>. There is also an evaluation of all questionnaires in XML/TEI available for further usage (https://github.com/i-d-e/ride/tree/questionnaire_data). The schema documentation was being revised at the time of writing this article, it is based on the jTEI schema and only deviates for specific aspects related to the nature of the publication

²² <http://opac.bbf.dipf.de/hk/jahrbuch.htm>.



full TEI-based, OA publication in an eXist database. While it will use the jTEI format for the scholarly journal, the goal is to generate a continuity between other scholarly resources that will also be published in TEI, and the journal itself.

In other European countries, Slovenia has one journal that is TEI-based: the *Contributions to Contemporary History*²³, following the jTEI schema. The Journal has published 181 articles so far.

On the American side

Among the scholarly journals Publishing based on a TEI format, *Digital Humanities Quarterly* is the one that comes to mind first, with strong personal ties to the TEI community, using a customization of TEI-P5. The *Scandinavian-Canadian Studies*, that currently has 189 articles, uses P4 converted to P5. Indiana University has two TEI-based journals: *Indiana Magazine of History* (IMH) and *The Medieval Review* (TMR).²⁴ Beth Mardutho: Syriac Institute of Texas A&M 's journal *Hugoye: Journal of Syriac Studies* has been migrated to TEI recently, currently converting all back issues in TEI.²⁵ *Romantic Circles* (University of Colorado Boulder) has been working with TEI since 2011.²⁶

2. Advantages and challenges of using a TEI-based format for scholarly publishing

Publishing scholarly journals in TEI offers several advantages on different levels. On an economic level, using the jTEI schema on top of an OJS workflow allows low-budget production of scholarly publications like journals. The development needed is limited, it does the job, and if it doesn't, you can mostly rely on the community to tweak minor issues. In other words, you don't need to invest a lot in order to publish your journal. This means in return that there is little economic gain to be expected from it.

The low cost and the easy transfer to open science can appeal to decision makers such as university presidents, especially for scholarly domains that are comparatively not too impact factor-dependent. More generally, there is, on a more political level, a growing interest for stable, sustainable publishing solutions that are now more and more being recognized as economically valuable.²⁷

²³ [Prispevki za novejšo zgodovino] <http://ojs.inz.si/pnz>.

²⁴ See <https://journals.openedition.org/jtei/1746>.

²⁵ See <https://hugoye.bethmardutho.org/index.html>.

²⁶ <http://romantic-circles.org/about/about.html>.

²⁷ See the COAR White Paper, "Pubfair – A Framework for Sustainable, Distributed, Open Science Publishing Services": <https://comments.coar-repositories.org>.



These arguments can explain why convincing decision makers, and more generally people from outside the TEI community, to “invest” in a TEI-based workflow for scholarly publishing has become an arguable case, especially in times where Open Sciences values are being pushed forward, a philosophy in which the TEI technology can easily be integrated.

For the TEI community, pushing forward secondary scholarly publications in TEI (as opposed to primary scholarly publications for which the TEI are designed and developed by the community) presents several advantages. The first one is a reality check. We have obviously reached more than a critical mass already, one that calls for a better coordination and sustainability in terms of how the community integrates its outputs.

Why does the TEI work so well for scholarly publishing of secondary sources? Several arguments can be taken into consideration there. First, the TEI is flexible in its vocabulary, which means that it allows to manage and bring together heterogeneous sources of information. Second, the TEI is not closed. On the contrary, it is conceived as to allow resources to communicate: it enables to avoid silos of internal formats developed for the use of one specific publication platform, that will need further specific development in order to remain sustainable, costing a lot to be maintained and keep being active. What makes the TEI valuable for secondary scholarly publications is what makes it valuable for any publication: its stability, its interoperability, its openness, its reusability. TEI-based scholarly publications are non-proprietary publications, that can also easily be transferred to OA publication models. There is one final argument, proposed by Laurent Romary in the course the TEI-L discussion on the topic, that is less obvious than the previous ones. To him, the main advantage of the TEI being the same basis format for primary and secondary sources is that it allows to use the same format for primary and secondary scholarly publications: e.g., for digital editions and for articles on the digital editions, etc. This continuity between the two major publication dimensions of SSH research activity allows both fluidity and solidity. Fluidity means here that integrating elements from primary resources to secondary resources (and reciprocally) is made particularly easy. Solidity means that the same people who have the expertise in one field can contribute to the other: the manpower available is considerable. But it also means that issues of nesting TEI structures are of central importance in this context.

In terms of research content, the continuum between primary and secondary scholarly publications could have another consequence: the TEI community could (or should?) become more attractive for scholars interested in less philological questions than the core community is, for instance sociological aspects of knowledge transfer,



community building or evolution of research strategies. Journal material could easily be turned into a primary source. And there are enough documents available at this point in terms of quantity to make this corpus interesting to sociologists, sociolinguists, historians of sciences, etc. What is still missing most of the time is a more obvious download button, e.g. a structural incentive to use TEI corpora.

The advantages of expanding TEI-based models to secondary publications would not only benefit a widening of the TEI community. It would also allow to go past the dead end of reputation mechanisms in the SSH at large. We are currently in a situation where scholars, assistants and research engineers are sacrificing a great deal of time and work for the profit of publishing houses that are negotiating reputation for money. Coming up with a TEI-based format has the potential to break these reputation rules because it is non-profit and low-tech in terms of technical maintenance, allowing to dedicate more time to actually reading papers than evaluating their impact factor or abiding by publishers' editorial guidelines (see Kosmopoulos & Pumain 2008).

What are TEI-based scholarly publications aiming at at large? Mainly at stable and wide dissemination. What makes reputation in the realm of such values is neither primarily quantity nor established publishing houses but mostly a dissemination strategy based on core virtues like FAIR²⁸, and a common set of values and improvements carried out by the community. Researchers need relevant papers to be accessible and easy to find, to evaluate and to reuse: TEI-based formats can help answer these needs. It is already the case for publishing in the situations described in Part 1, it is also sometimes the case for authoring, but never so far for reviewing and copyediting.

3. TEI-based workflow improvements

In order to be able to cover a complete publication workflow in TEI, we would not only need to improve and develop customizations for publishing and for authoring as they already exist.²⁹ It is necessary to conceive a complete TEI-based workflow, e.g. to take into account the reviewing and copyediting phase as well – phases that are currently dealt with in other formats.

Complex workflows are at the core of editorial work for scholarly journals. The coordination chain usually includes authors, editors, reviewers, and copyeditors. Depending on the journal and the reviewing format (single blind, double blind, open), this can easily add up to 10 people working on one text. The jTEI for instance foresees

²⁸ <https://www.go-fair.org/fair-principles/>

²⁹ See also Thoden, 2019, on the strategy of converting the workflow and document basis from a proprietary format to a fully standards-compliant system in the context of a publishing platform.



3 different reviewers for each paper and 3 copyedit rounds (some of which are done by the editors, but not all). The *Philosophie Antique* journal³⁰ has a printed edition in addition to the digital one; this implies, in addition to double blind review for each article, a double copy-edit for each of the two formats. With different issues running in parallel as it is the case in most online publications now, this means having an editorial interface that allows to deal with different workflow timelines at the same time and different editorial roles according to one single person.

jTEI is working with OJS like many TEI-based journals do. There has been some improvement of the OJS functionalities and interface with the recent update to version 3.0. but OJS still suffers from being developed for too many different uses, making it occasionally tricky to tailor for specific needs. For years, the overall management of jTEI was actually dealt with not in OJS itself, but in tables archived in separate googledocs (one googledoc for each issue of the journal), because OJS proved unable to offer such an overview in the way that was needed. This type of management has the inconvenient that the text and the information on the status of the text are separated, making an overview difficult to gain for the different actors involved in the process of text production. *Philosophie Antique* is even worse off. The workflow doesn't use any content management interface. The editorial team works by document exchange, archiving successive versions, with two parallel workflows for the preparation of the pdf to be printed and the uploading of texts in html, using the Lodel tool on the OpenEdition interface.

These two examples are symptomatic of the situation in many journals. Compared to this reality, it is clear that, ideally, the texts should contain metadata allowing to see at first glance as well as process automatically their editorial status in order to be able with the complexity of workflows as we know them today. This is one of the many things that the TEI can provide.

In order to conceive a TEI encoding that would fulfill the reviewing and copyediting function and allow for an overview of the editorial status of the text, let's first consider the different roles (author, reviewer, copyeditor) and the different types of interventions in the text. In an ideal world, each role would correspond to a type of intervention, but in the editorial practice, it might well be that the copyeditor who checks for content coherence also finds typos to correct. One can roughly divide into these two categories the types of interventions that will be done on the text: one intervention type encompasses content (editorial requirements and their application), while the other one deals with the form (ortho-typographical aspects). This corresponds to two workflows that run in parallel: one checking for the scholarship, adequacy, coherence of the article, and one checking for its adequation

³⁰ <https://journals.openedition.org/philosant/>

to the typographical guidelines of the journal. The TEI encoding for reviewing and copyediting should reflect these two aspects as well as the two work phases that are the reviewing and the copyediting processes.

The first step in a TEI-based workflow will be to define on the one hand roles (reviewer A, reviewer B, reviewer C, etc.) and on the other hand types of modifications that can be undertaken: an editorial schema (EdSchema), a tagset for modifications of formal aspects, and a redactorial schema (RedSchema) that allows to tag content-related modifications. In order to illustrate the prototype we have in mind, we used existing examples of articles already published or in the process of being published, and converted the copyediting/reviewing process from the word document to a TEI-based version.³¹

EdSchema allows to tag elements from the review process as well as from the copyediting process. The tags are attributed to the different resp defined in the header (see Example 3&5 Header).

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <?xml-model href="http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_all.rng" type="application/xml" schematypens="
http://relaxng.org/ns/structure/1.0"?>
3 <?xml-model href="http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_all.rng" type="application/xml"
schematypens="http://purl.oclc.org/dsdl/schematron"?>
4 <TEI xmlns="http://www.tei-c.org/ns/1.0">
5 <teiHeader>
6 <fileDesc>
7 <titleStmt>
8 <title [REDACTED]/title>
9 <author xml:id="aut">Anonyme</author>
10 <respStmt>
11 <resp xml:id="Ed1">reviewer A</resp>
12 <name>Anonyme</name>
13 </resp>
14 </respStmt>
15 <respStmt>
16 <resp xml:id="Ed2">reviewer B</resp>
17 <name>Anonyme</name>
18 </resp>
19 </respStmt>
20 <respStmt>
21 <resp xml:id="Ed3">reviewer C</resp>
22 <name>Anonyme</name>
23 </resp>
24 </respStmt>
25 <respStmt>
26 <resp xml:id="Ed4">reviewer D</resp>
27 <name>Anonyme</name>
28 </resp>
29 </respStmt>
30 </title>
31 </titleStmt>
32 <publicationStmt>
33 <p>[REDACTED]</p>
34 </publication>
35 </publicationStmt>
36 <sourceDesc>
37 <p>[REDACTED]</p>
38 </sourceDesc>
39 </fileDesc>
40 </teiHeader>
41 <text>
42 <body>

```

3&5 Header

³¹ We would like to thank the authors, reviewers and copyeditors who allowed us to use for this purpose data that is usually not made public but considered as part of the “black box” of academic publishing.



EdSchema contains primarily the `<lem>`, `<add>`, ``, `<choice>` elements.

Illustration: examples 1a and 1b.

In addition to the reference format, each collection is made available in a simple plain-text format automatically derived from the ~~TEI XML~~ TEI version, containing only the text included in the body of the narrative texts and plays (in particular, excluding prefaces and other paratext as well as notes) and with external metadata provided in tabular format. This format is especially suitable notably for direct use with the ~~stylo tool~~ package ~~style~~ for R (Eder, Kestemont, and Rybicki et al. 2016) and other tools operating on the surface level of texts.

1a

```
In addition to the reference format, each collection is made available in a simple plain <del resp="#Ed1">-</del>text
format automatically derived from the <del resp="#Ed1">XML</del>TEI version,
containing only the text included in the body of the narrative texts and plays
(in particular, excluding prefaces and other paratext as well as notes)
and with external metadata provided in tabular format. This format is
<choice> <sic resp="#aut">suitable notably</sic><corr resp="#Ed1">especially suitable</corr> </choice>for direct use with the stylo
<choice> <sic resp="#aut"> tool</sic> <corr resp="#Ed1">packages</corr></choice> for R
(Eder <choice><sic resp="#aut">et al.</sic><corr resp="#Ed1">, Kestemont, and Rybicki</corr></choice> 2016)
and other tools operating on the surface level of texts.
```

1b

RedSchema consists of a tagset that is in part similar to that of EdSchema, attributed to the corresponding role (author, reviewer, editor).

Both types of interventions are likely to involve short alterations (changes in punctuation marks for instance), bibliographical elements³², and longer text passages that need re-writing, this last category being more likely to be relevant to RedSchema (see examples 2a and 2b).

editions, but they may not be well adapted to ongoing work. Proposing a **visualization** for such encoding is not easy, because there might not be an identifiable critical text (yet), and the styles can be mixed in the apparatus. The Critical Apparatus Toolbox will display each style (lemma and reading({s}), or readings only) differently:

In any case, the content of **lem** and **rdg** are always highlighted, with a white background.

When an **<app>** element contains a **<lem>** and one or more **<rdg>**s, there is an easily identifiable critical text: only the content of the **<lem>** will be displayed in the text, and the readings will appear in a pop-up note. To see the note, the users simply need to click the upward-pointing arrow sign (↑) following the content of the lemma. Example:

Does this revision make sense?

First item in list?

Editors: Is boldface necessary (here and throughout)? We usually just use `monospace` to set off technical terms.

Seems like wrong image proportions...

2a

```
<app from="#Rev1beg" to="#Rev1end" resp="#Ed1"> <note> Does this revision make sense? </note></app>
<app from="#Rev2beg" to="#Rev2end" resp="#Ed1"> <note>First item in list?</note></app>
<app from="#Rev3beg" to="#Rev3end" resp="#Ed1"> <note>Editors: Is boldface necessary (here and throughout)?
  We usually just use monospace to set off technical terms.</note></app>
```

2b

mais il l'associe discrètement à son époux légitime⁴. Mais Lucrèce, dans un geste poétique que cette intertextualité révèle donc comme hautement **subversif**, choisit de transformer cette situation dangereuse et immorale en une relation paisible, joyeuse et bienfaitante – puisque l'union de Mars et de Vénus représente l'apaisement de la brutalité et de la sauvagerie par la recherche du plaisir selon la nature.

Ce choix poétique place dès l'incipit la sexualité sous le signe de l'ambiguïté⁵; elle est une nécessité non seulement physique mais éthique, elle est également un piège dont il faut se méfier – piège au sens propre du terme, la scène dépeinte par Lucrèce précédant immédiatement dans le récit homérique le châtement des amants, enfermés par Vulcain dans un filet.

Pour nuancer un peu le caractère subversif de ce geste, voir Françoise Gury, 'Mars et Vénus : l'amour dans la dignité : le dossier de la peinture campanienne', in De la théâtralité du corps aux corps des dieux dans l'Antiquité, ed. by Valérie Huet and Florence Gherchanoc (Brest: Centre de Recherche Bretonne et Celtique, 2014), pp. 143-161

4a

³² It should be possible to add bibliographical entries using the resp indicating it comes from a specific reviewer or a copyeditor (see Example 4 and 4b).

```
<p>Mais Lucrèce, dans un geste poétique que cette intertextualité révèle donc comme hautement <anchor xml:id="Rev1beg"/>
subversif<anchor xml:id="Rev1end"/>, choisit de transformer cette situation dangereuse et immorale en une relation
paisible, joyeuse et bienfaisante – puisque l'union de Mars et de Vénus représente l'apaisement de la brutalité et de la
sauvagerie par la recherche du plaisir selon la nature.</p>

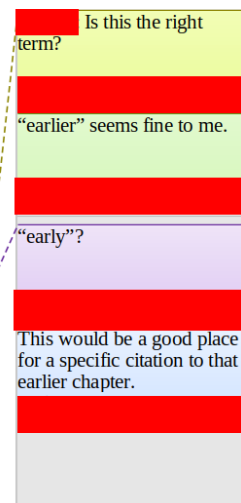
<app xml:id="Rev1" from="#Rev1beg" to="#Rev1end"> <note resp="#Ed1">Pour nuancer un peu le caractère subversif de ce
geste, voir <bibl >Françoise Gury, 'Mars et Vénus : l'amour dans la dignité : le dossier de la peinture campanienne', in
De la théâtralité du corps aux corps des dieux dans l'Antiquité, ed. by Valérie Huet and Florence Gherchanoc (Brest:
Centre de Recherche Bretonne et Celtique, 2014), pp. 143-161</bibl></note></app>
```

4b

In order to address remarks that require the rewriting of a longer text passage, RedSchema needs to include an anchor-based tagset allowing to point to a commentary, which in turns should allow the author(s) and/or editor(s) to answer this commentary (see examples 3a and 3b; 5a and 5c).

for the current status quo in the TEI Guidelines or directly influenced our own understanding of the gGuidelines and how they should evolve.

A major milestone in the digital dictionary era is the work by [Amsler and Tompa \(1988\)](#), which, together with the unifying contributions of Ide and Véronis (1994), led to the [earlier](#) TEI “Print Dictionaries” chapter. Focusing here on their contribution to etymology, example 1 shows how they have introduced a highly structured model based on etymons and links implemented as an SGML DTD. The underlying model is based upon a graph of etymons (<etymon>) connected with relations (<rel>), forming a more global etymological tree (reflected by the



3a

```
<div type="texte" resp="#aut">
<p>A major milestone in the digital dictionary era is the work by Amsler and Tompa (1988),
which, together with the unifying contributions of Ide and Véronis (1994),
led to the <anchor xml:id="Rev1beg"/>earlier<anchor xml:id="Rev1end"/> TEI “Print Dictionaries” chapter.
Focusing here on their contribution to etymology, example 1 shows
how they have introduced a highly structured model based on etymons
and links implemented as an SGML DTD. The underlying model is based upon a graph of etymons (@etymon)
connected with relations (@rel), forming a more global etymological tree
(reflected by the @es, etymological segment, element).</p>
</div>

<app xml:id="Rev1" from="#Rev1beg" to="#Rev1end"><note resp="#Ed1">[red box]: Is this the right term?</note>
<note resp="#Ed2">“earlier” seems fine to me.</note>
<note resp="#Ed3">This would be a good place for a specific citation to that earlier chapter.</note></app>

<div type="discussion">
<p resp="#aut"> <ptr target="#Rev1"/>“early”?</p>
</div>
```

3b



repertoire of a given language, the `<etym>` element will obviously appear as a child of `<entry>`. It can also be the case that one has to deal with the emergence of a new sense for a given word, in which case `<etym>` should be attached to the corresponding `<sense>` element. Clearly, this distinction is correlated with etymological types, borrowing being more likely to be related to lexical entries whereas metaphors would more often correspond to new senses.

Any change that occurs within a lexicon can be labeled in the `@type` attribute of the `<etym>` element in a TEI dictionary;¹⁶ and the fact that a change occurred within the contemporary lexicon (as opposed to its parent language) is indicated by means of `@xml:lang` on the source form.¹⁷

Oops! The ticket has been accepted by the council. See <https://github.com/TEIC/TEI/issues/1512>
We need to change the footnote to say that we made the proposal and that it is about to be implemented in a future version of the guidelines

Please see if my revisions to the footnote make sense.

Also, provide a URL for the ODD customization?

5a

```
<p> Any change that occurs within a lexicon can be labeled in the @type attribute of the @etym element in a TEI dictionary;
  <anchor xml:id="Rev2beg"/>
  <note resp="#aut" n="16"> Currently the use of @type in the @etym element is not permitted in the TEI schema as the
  aforementioned element is not a member of the att.typed class. <choice><sic resp="#aut"> The </sic> <corr resp="#Ed1">However, a
  </corr></choice> proposal
  <add resp="#Ed1">for this use</add> has been <choice><sic resp="#aut"> submitted in </sic><corr resp="#Ed1"> accepted by </corr>
  </choice> the <choice><sic resp="#aut"> TEI GitHub </sic> <corr resp="#Ed1"> TEI Council </corr></choice> <add resp="Ed1"> and
  will be implemented in a future version of the TEI Guidelines; meanwhile, the ticket </add> <del resp="#Ed1"> and </del> is
  available at https://github.com/TEIC/TEI/issues/1512. In our proposal, the @etym element has to be made recursive in order to
  allow the fine-grained representations we propose here. The corresponding ODD customization, together with reference examples, <
  choice> <sic resp="#aut"> are </sic> <corr resp="#Ed1"> is </corr></choice> available on GitHub.</note>
  <anchor xml:id="Rev2end"/>
  and the fact that a change occurred within the contemporary lexicon (as opposed to its parent language) is indicated by means of
  @xml:lang on the source form.2 </p>

</div>
  <app xml:id="Rev2" from="#Rev2beg" to="#Rev2end"> <note resp="#aut"> Oops! The ticket has been accepted by the council. See
  <ref target="https://github.com/TEIC/TEI/issues/1512"> TEIC/TEI/issues/1512 </ref>
  We need to change the footnote to say that we made the proposal and that it is about to be implemented in a future version of the
  guidelines </note> </app>

  <div type="discussion">
  <p resp="#Ed1"> <ptr target="#Rev2"/>Please see if my revisions to the footnote make sense. Also, provide a URL for the ODD
  customization?</p>
  </div>
```

5c

In order to generate a clean text in the end, the last modification would be considered as final, which means that the last copyeditor should validate with his/her `resp` attribute the earlier modification layers according to the final editorial decision. The main advantage of such an encoding is that it sums up in one file all the editorial evolution of a text, from submission to publication, displaying precisely its evolution and the contribution of each one of those who were involved in this process. It is therefore a question of including within the chain that structures and edits content the part corresponding to the evaluation and formatting, and thus making it visible and shareable. In this way, the data is made open to the point of the preparation of the data itself. It is worth noting that recent efforts made by scientific publishers to



automate workflows have focused on publishing content but did not include the preparation of such content.³³

All of this is summed up not only in one file, but also in one unique file format, guaranteeing conversion without information loss. It allows to split the editorial work on form and content, giving the editor the final hand on the last version of the text. And finally, it should be conceived as a fairly minimal combination of tagsets, meaning that these schemas should be easy to share with other scholars and journals.

It would have, on the downside, the inconvenient of its advantages: being multi-layered, such a document might quickly become complex. Transformation scenarios filtering specific tagsets in order to gain readability will be made necessary: both readability for the human eye and information extraction for digital tools would rely on the development of such transformation scenarios. But all in all, the development of such schemas and transformation scenarios seems in the realm of the doable considering what the TEI has been able to develop over the past decades.

```

39
40 <div type="texte" resp="#aut">
41
42 <p>A major milestone in the digital dictionary era is the work by Amsler and Tompa (1988), which, together with the unifying
contributions of Ide and Véronis (1994), led to the <anchor xml:id="Rev1beg"/>earlier<anchor xml:id="Rev1end"/> TEI "Print
Dictionaries" chapter. Focusing here on their contribution to etymology, example 1 shows how they have introduced a highly
structured model based on etymons and links implemented as an SGML DTD. The underlying model is based upon a graph of etymons
(@etymon) connected with relations (@rel), forming a more global etymological tree (reflected by the @es, etymological segment,
element).
43
44 [...]
45
46 Any change that occurs within a lexicon can be labeled in the @type attribute of the @etym element in a TEI dictionary;
47 <anchor xml:id="Rev2beg"/>
48 <note resp="#aut" n="16"> Currently the use of @type in the @etym element is not permitted in the TEI schema as the
aforementioned element is not a member of the att.typed class. <choice><sic resp="#aut"> The </sic> <corr resp="#Ed1">However, a
</corr></choice> proposal
49 <add resp="#Ed1">for this use</add> has been <choice><sic resp="#aut"> submitted in </sic><corr resp="#Ed1"> accepted by </corr>
</choice> the <choice><sic resp="aut"> TEI GitHub </sic> <corr resp="#Ed1"> TEI Council </corr></choice> <add resp="Ed1"> and
will be implemented in a future version of the TEI Guidelines; meanwhile, the ticket </add> <del resp="#Ed1"> and </del> is
available at <a href="https://github.com/TEIC/TEI/issues/1512">https://github.com/TEIC/TEI/issues/1512</a>. In our proposal, the @etym element has to be made recursive in order to
allow the fine-grained representations we propose here. The corresponding ODD customization, together with reference examples, <
choice> <sic resp="#aut"> are </sic> <corr resp="#Ed1"> is </corr></choice> available on GitHub.</note>
50 <anchor xml:id="Rev2end"/>
51 and the fact that a change occurred within the contemporary lexicon (as opposed to its parent language) is indicated by means of
@xml:lang on the source form.2 </p>
52
53 </div>
54
55 <app xml:id="Rev1" from="#Rev1beg" to="#Rev1end"><note resp="#Ed1"> [redacted] Is this the right term?</note>
56 <note resp="#Ed2">"earlier" seems fine to me.</note>
57 <note resp="#Ed3">This would be a good place for a specific citation to that earlier chapter.</note></app>
58
59 <app xml:id="Rev2" from="#Rev2beg" to="#Rev2end"> <note resp="#aut"> Oops! The ticket has been accepted by the council. See <ref
target="https://github.com/TEIC/TEI/issues/1512">https://github.com/TEIC/TEI/issues/1512</ref>
60 We need to change the footnote to say that we made the proposal and that it is about to be implement in a future version of the
guidelines </note> </app>
61
62
63 <div type="discussion">
64 <p resp="#aut"> <ptr target="#Rev1"/>"early"?</p>
65 <p resp="#Ed1"> <ptr target="#Rev2"/>Please see if my revisions to the footnote make sense. Also, provide a URL for the ODD
customization??</p>
66 </div>
67
68 </body>
69 </text>
70 </TEI>
71

```

3&5 complete

³³ See for instance the METOPES xml editorial chain which is based on the principle of single source publishing but leaves aside the preparation of content itself: http://www.unicaen.fr/recherche/mrsh/document_numerique/outils/metopes, which is not making its schemas openly available at this stage.



4. Structural workflow improvements beyond the TEI

The main reason why corrections are implemented, and the correction process hidden from the public eye, is a need for control. Editors and authors want to have control over each modification in the text, be it punctuation, bibliography formatting, or a sentence that seems a bit difficult to understand. This is all the more important in the case of papers written by non-native speakers. Those need a specific linguistic attention to reach the same level or readability as those written by native speakers.

But looking at it more closely, the whole workflow inherent to this control of modifications and corrections is based on reputation control. It is based on editorial black boxes that keep articles away from the public eye as long as they are not “perfect” or “finished”, e.g. “published”. It is based on the idea that work in progress can damage reputation. This is exactly what started to change when publishing digital scholarly editions online. The fact that it is possible to update a digital scholarly edition suggests that we could at least imagine that such a plasticity can be envisioned for journal articles too. Making it possible to work with sources that are not fixed in time is one of the greatest intellectual progresses made not only possible, but actually realized by the TEI. Taking it one step further for scholarly journals is a fascinating intellectual challenge in terms of data dissemination quality insurance. It is also the logical next step in terms of data empowerment.

This could easily be encouraged by two (infra)structural lines of action. The first one is to further fostering pre-print publications (jTEI encourages such publications for papers submitted to the Journal). The second one consists in improving the academic recognition of editorial expertise. There are already many experts in TEI working as research engineers or editorial assistant. They have a TEI expertise and are in general working for a specific journal. E.g., they work for an editor, or worse, a publisher. This expertise should be better recognized than it is now, and valued more explicitly in advertisements for editorial jobs. The TEI community should encourage this expertise to be better represented, for instance in the SIGs or through the awarding of prizes. The academic publishing market have been intensely professionalized for several years, particularly through the emergence of networks such as the Medici network³⁴, which help in the continued training of editors. The profession is now ready to integrate such a workflow and support its development.

If we were to turn the workflow in such a way that the versioning occurring alongside reviewing and copyediting was easily manageable, it would mean that the different updated versions could be published online in preprint archives until the journal publication appears. For instance, it would be possible to filter through customized

³⁴ <http://medici.in2p3.fr/?lang=fr>.



transformations what would be fit for pre-defined publication steps up to the final version. It would be possible to tailor the review process to make it completely open, open for EdSchema, open for RedSchema, or not open at all, according to the author's wish and/or the journal's policy. The same goes for copyediting.

In this situation where the dialogue between authors, reviewers and editors can be made transparent within the text, scholars would be in a different position than that of accepting or refusing a correction suggestion. The dynamics of the relationships between those involved in the process of generating the text could benefit from this change. The TEI community could be actively involved in providing schematrons, stylesheets and publishing environments for journals that would allow researchers to access first-hand TEI documents, be it their own for additional editorial work or others' for queries.

One low-key option to implement such a workflow consists in building overlay journals like those hosted on the episciences platform³⁵, that only provides an additional review layer on top of pre-print publications. The interface allows to set up a review process and link to the evaluated and selected pre-prints in order to constitute a journal issue. While this simplifies greatly the review process, especially due to the easy-to-tune interface, it doesn't really address the online presentation of publications in a reader-friendly way, for which other solutions have to be implemented.

Conclusion

There are options, some have been looked into, none of them is perfect at this point. Workflows are complicated because in theory they intend to address all the needs of the editorial process. In terms of scholarly publishing workflows, there is some development work to do in order to be able to disseminate an encompassing TEI schema for authoring, reviewing and publishing. But we have a good starter kit, potential users already working with a TEI-based publishing format, and a great deal of expertise to build upon. Looking at the many journals that are already working with a TEI-based publishing format, there is mainly one thing missing to take it to the next step: good documentation everywhere.

The TEI community is in a position to impact access to knowledge for the generations to come. This does not only concern our digital editions of manuscripts, but also the way we disseminate all the information we gather from working with manuscripts, and with digital editions. Fostering TEI-based scholarly publications is worth while: first, it has been done at different scales for publishing and is working; second, it is a field that is known to about all of us, allowing us to reflect on the specificities of the

³⁵

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



different disciplines we come from, and third, it is to us to initiate this change because nobody else can do it like the TEI community can. There is no reason why the coming generations should be plagued by requirements from publishing houses now that there is a real political awareness for solid archiving formats and that the TEI has established itself as a standard in so many fields already.

Bibliography

Martin Holmes, Laurent Romary. Encoding models for scholarly literature. *Publishing and digital libraries: Legal and organizational issues*, dir. Ioannis Iglezakis, Tatiana-Eleni Synodinou, Sarantos Kapidakis. IGI Global, 2010, [10.4018/978-1-60960-031-0](https://doi.org/10.4018/978-1-60960-031-0). [hal-00390966v2](https://hal.archives-ouvertes.fr/hal-00390966v2)

Christine Kosmopoulos, Denise Pumain. Révolution numérique et évaluation bibliométrique dans les sciences humaines et sociales. *Revue européenne des sciences sociales* [Online], XLVI-141 | 2008, Online since 01 May 2011, connection on 19 December 2019. URL : <http://journals.openedition.org/ress/151> ; DOI : 10.4000/ress.151

Klaus Thoden. Modeling scholarly publications for sustainable workflows. *ELPUB 2019 23rd edition of the International Conference on Electronic Publishing*, Jun 2019, Marseille, France. [10.4000/proceedings.elpub.2019.2](https://doi.org/10.4000/proceedings.elpub.2019.2). [hal-02143240](https://hal.archives-ouvertes.fr/hal-02143240)

Biographies

Anne Baillot is a Full Professor in German Studies at Le Mans Université. She is the editor of *Letters and Texts. Intellectual Berlin around 1800* and former managing editor of the *Journal of the Text Encoding Initiative*.

Julie Giovacchini is a Research Engineer in Classics at French CNRS. She is co-editor of *Philosophie antique*, a journal devoted to history of Ancient philosophy.