An Empirical Study of Versioning in Digital Scholarly Editions

Martina Bürgermeister
University of Graz
`martina.buergermeister@uni-graz.at`

Abstract

English. This paper contributes to increasing the reliability in Digital Scholarly Editions (DSE) by emphasizing the use of versioning. These mechanisms have the potential to render DSE more trustworthy. Within an analysis of existing DSE projects it will be shown, what kind of versioning strategies have been implemented and why. Versioning strategies are assessed according to a definition of versioning, which has three criteria: creation of versions for each change, documentation of changes, and availability of previous versions.

Italiano. Questo articolo contribuisce a sviluppare una migliore valutazione dell'attendibilità delle edizioni scientifiche digitali (DSE), enfatizzando l'uso delle versioni. Questi meccanismi hanno il potenziale per rendere la DSE più affidabile. Nell'ambito di un'indagine sui progetti DSE esistenti verrà mostrato che tipo di strategie di versioning sono state implementate e perché. Le strategie di versioning sono analizzate secondo una definizione di versioning, che ha tre criteri: creazione di versioni ad ogni modifica, documentazione delle modifiche, e disponibilità delle versioni precedenti.

Digital publishing and a textuality that is dynamic, collaborative, distributed and interdependent lead to the digital scholarly edition (DSE) facing additional technological challenges in contrast to print editions. The *MLA Committee von Scholarly Editions* takes care of this and recommends the use of technologies appropriate to the goals of the edition, "in recognition of the fact that technologies and methods are interrelated in that no technical decisions are innocent of methodological implications and vice versa (MLA 2016, 1)." The committee also suggests a design of the DSE that will be as durable and sustainable as possible. This applies not only to DSE, but also to the entire scientific practice. In the "Guidelines of good scientific practice", like the one from the University of Parma, it says: "I ricercatori dell'Ateneo devono aver cura che tutti i dati, primari e secondari, generati dalle loro attività di ricerca siano archiviati e conservati in modo corretto ed appropriato, garantendone la sicurezza e l'accessibilità […]" Besides being responsible for data safety and access, researchers should support a prompt data exchange and reuse: "Con l'intento di rendere la ricerca più aperta, globale e collaborativa e garantirne un controllo di qualità, i dati dovrebbero essere messi a disposizione dei colleghi che vogliano replicare lo studio o elaborare nuove ricerche a partire da essi […] (Università di Parma, 2018)".

The reuse of data is a matter of trust in data itself and the IT infrastructure as maintainer and provider. In this case, transparency is a key factor. In the following it will be argued, that versioning is more than a measure to guarantee data authenticity, integrity and accessibility. Versioning as an integrated part of the DSE creates transparency in an ongoing scholarly discourse. This objective is achieved, firstly, when new versions of the DSE or components of it are created if changes are made to it. Second, if what has been changed is also communicated; and third, if previous versions are made available. These three criteria form the basis of my analysis of versioning DSEs. Using an empirical approach, I would like to find out why versioning plays a role in these projects and how it is implemented.

The implementation of versioning in a DSE is not yet state-of-the-art. In order to find out which DSEs have a versioning strategy and what it looks like, the "Catalogue of digital editions" (ÖAW) was used as a finding aid. Since the list is very extensive, the analysis includes edition projects mainly from the past 15 years, which have developed over a relatively longer period of time. Since there are far more DSEs without versioning than with it, the empirical basis was mainly provided by Bleier's preliminary work (2019). He has evaluated 100 DSEs on criteria such as citation recommendation, permalinks and versioning. Finally, those cases that assign version numbers, but neither convey what has changed, nor make past versions available, were excluded from further consideration. Their exclusion from this analysis is justified by the fact that they fulfil only one of three conditions for versioning DSEs discussed here. The review has shown that so far three different versioning strategies have been applied, which meet at least two of the three conditions formulated above. An overview of the discussed strategies is provided in section 5 (Table 1).

1 Versioning as documentation

DSE projects of this type implement the assignment of sequential version numbers for changes and document what has been changed. A list of past version numbers with descriptions of the revisions and where they were made is provided. In these projects, the DSE is seen as an ongoing process, which is documented.

The strategy is pursued in the DSE "Auchinleck Manuscript", edited by David Burnley and Allison Wiggins. The project started in 2000 and was launched in July 2003 with version 1. The edition has a citation recommendation with version number indication. The editors recommend citing version 1.1 because it corresponds to the current version (March 15, 2004). The version number applies to the entire project. There is a version documentation with the title "Archive of site updates". It consists of three lists: One with version number and date, another with "Corrections to texts and to textual notes" and an extra column "Record of other changes", which lists changes to the bibliography and the side menu. Allison Wiggins explains the version documentation theme as follows:

> All changes made to the content of this site are recorded here in the archive of site updates. Each time a batch of updates is added, the site is designated a new version number. This system ensure that users can keep track of any changes made and can reference the site materials accurately (Burnley and Wiggins, 2019).

At about the same time (March 2003), the DSE "The Old Bailey Proceedings Online" was published. It is an editorial long-term project that presents London processes from the period 1675-1913 and Tim Hitchcock, Robert Shoemaker, Clive Emsley, Sharon Howard and Jamie McLaughlin are responsible for the edition. A version documentation can be found in the "What's New Archive". As in the previous example, an introductory text refers to the non-final, process-like state of the DSE. It is followed by a listing and description of what has changed in previous versions: the current status is March 2018 with version 8.

A version documentation is also available in the edition project "The Online Froissart". The editors Peter Ainsworth and Godfried Croenen cite another reason for documenting versions:

> The Online Froissart is a collaborative, interdisciplinary and incremental project. Given the sheer size of Froissart's Chroniques, the number of surviving manuscripts and their dispersal across many libraries in several different countries, it has been decided not to delay publication of available transcriptions until all of these materials have been transcribed and annotated. The Online Froissart, therefore, publishes all currently available transcriptions and other materials produced by the project team, plus updates, and augments the website and its underlying datasets on a regular basis (annually from 2012 onwards). (Ainsworth and Croenen 2018)

Due to the large amount of material and its scattered provenance, an intentional decision was made that the publication of the DSE should be as early as possible, but the edition will be regularly updated.

2 Versioning as version management

Versioning is implemented in the same way in the DSE as it can be found in software development. Version management systems save all changes to text documents as versions. All versions can be restored, and these systems are also designed to handle collaborative writing processes. Examples of DSEs that implement this versioning strategy are "Papyri.info" and "The Devonshire Manuscript".

"The Devonshire Manuscript" is a Wikibook edition, whose main editor is the Devonshire Manuscript Group.

The aim is to discuss the edited sources as widely as possible and to change the role of the scientific editor from the sole authority for the text to that of a moderator: "The social edition is a work that brings communities together to engage in conversation around a text formed and reformed through an ongoing, iterative, public editorial process." (Wikibooks, 2014) One of the main contributors to this project is Ray Siemens (2012, 453), whose motivation to support the DSE with the help of social media is as follows:

> Such tools facilitate a model of textual interaction and intervention that encourage us to see the scholarly text as a process rather than a product, and the initial, primary editor as a facilitator, rather than a progenitor, of textual knowledge creation. (ibid.)

It is therefore about the editorial process as an iterative and collaborative activity. In this respect, it is essential for the progress of the project to keep all iterations available in the form of commented versions.

Like all pages of Wikipedia, all pages of a Wikibook are based on the same software MediaWiki (since version 1.5) and have a revision history (under the tab View history). The revision history in the form of a table contains all edits of a page in the wiki. Each change to a page creates a change line that contains information about the person who made the edit, the time when the edit was made, and a reference to the new wiki text in the text table. Elements of the revision table are preserved permanently, unless the page is deleted.

The versions of Papyri.info are accessible in a similar way, but via a different interface. Papyri.info aggregates papyrological resources from different databases and makes them available for editing. This DSE has been in existence since 2006 and interested editors can still add or change data today. A peer review of the revisions ensures the quality of the content. The implementation of the strongly social and collaborative project approach is made possible by a version management software that manages different users and their contributions. A method that was developed to facilitate the software development process is used here to manage the editorial processes. The DSE is stored in a Git repository. All editorial processes are recorded, versioned and recoverable. A look at the repository in Github shows 100 contributors[1] and more than 100,000 commits[2].

Git is a version control system that is used for collaborative software development. As already mentioned, the changes to the files are tracked. These programs provide access to any version of the file so that any changes can be undone. Each version has a timestamp and an author. It is always possible to see who changed what and when.

In the case of Papyri.info, the edited texts are saved as XML files in Git. Via the platform Github the repository can be viewed and the different versions of the texts can be displayed. The version comparison is done line by line and any changes to the file will be recognized by the software and automatically saved as a new version.

It makes no difference which version management software is used, the understanding of what a version is remains the same in the software development domain. In the examples mentioned so far, which have version documentation, the same version number stood for a whole series of revisions. If versioning takes place via version control systems, every saved change becomes a new version, which has no further semantic meaning. It makes no difference for the system if you make many changes or just fix a typo, it is always a new version. This strategy certainly has its advantages especially in a collaborative editing process.

3. Versioning as retrievable milestone versions

In contrast to the open DSEs mentioned above, which are geared towards a high frequency of changes, DSEs of this type are updated at intervals under new version names. These versions can also be called milestone versions, because the question of when to publish the next update is a project specific decision. Former versions are findable via a permalink and can be retrieved. A version name or number applies to the entire edition. This versioning strategy is evident in teams of editors who deliberately publish changes, revisions, and enhancements collectively. Every single editing step is not shown. What is desired as a research process in the case of Wikibooks is not part of the intention to publish an edition in this case.

This group includes the DSE "Der Sturm. Digital Source Edition on the History of the International Avant-Garde", developed and edited by Marjam Trautmann and Torsten Schrade since 2018. This edition project is still a "work in progress": the team of editors will gradually open up new sources and publish them promptly. The motive for this approach is explained as follows:

> Dies kommt den interessierten Forschungscommunities zugute, da somit ein schneller Zugriff auf eine beständig wachsende Gesamtedition gewährleistet ist. Ein weiterer Vorteil dieses iterativen Vorgehens ist, dass sich somit auch das Forschungsdatenmodell und die benötigten Softwarekomponenten kontinuierlich und in Einklang mit den hinzukommenden Quellen und ihren jeweiligen Spezifika weiterentwickeln können. (https://sturm-edition.de/projekt/methodik.html)

All developed sources have several permalinks that represent the versions. The permalinks are constructed in such a way that the identifier ends with the version information:

Version 1: https://sturm-edition.de/id/Q.01.19140115.FMA.01/**1**
Version 2: https://sturm-edition.de/id/Q.01.19140115.FMA.01/**2**

"Humboldt Digital" also integrates this type of versioning. The edition "Humboldt Digital" is a publication of the Academy Project "Alexander von Humboldt auf Reisen – Wissenschaft aus der Bewegung" at the Berlin-Brandenburg Academy of Sciences and Humanities. In this project, each entry offers the possibility to view past milestone versions. The specific version number is inserted after the domain name. E.g. "https://edition-humboldt.de/**v4**/H0002656".

---

[1] Github Glossary: "A collaborator is a person with read and write access to a repository, who has been invited to contribute by the repository owner."

[2] Github Glossary: "A commit, or "revision", is an individual change to a file (or set of files). It's like when you save a file, except with Git, every time you save it creates a unique ID (a.k.a. the "SHA" or "hash") that allows you to keep record of what changes were made when and by who. Commits usually contain a commit message which is a brief description of what changes were made."

4. Mix type: Versioning as retrievable, documented milestone versions

If milestone versions are also documented, then all three of the above criteria for versioning DSEs are met: changes will be published at intervals as new versions. Old versions remain viewable and what the version stands for and what has been changed are documented.

An example of this is the "August Wilhelm Schlegel Edition". This DSE is about bringing together August Wilhelm Schlegel's correspondence. The project runs until 2020 under the direction of Jochen Strobel and Claudia Bamberg. The first beta version was published on 2 June 2014; version-07-19 was published on July 1, 2019. Specific versions can be addressed by entering the name after the domain. For example, a letter in the up-to-date version has a Permalink like this:

`https://august-wilhelm-schlegel.de/`**`version-07-19`**`/briefid/1599`".

Under the menu item "Versions" is stated that every three months a new version with numerous resources will be published, while all previous versions remain fully accessible in the version archive.

The DSE "Johann Wolfgang Goethe: Faust", edited by Anne Bohnenkamp, Silke Henke and Fotis Jannidis, also has a version archive. The entire project is versioned at regular intervals and the current version is 1.2. In this case, the version is specified in the subdomain:

"`http://v1-2.faustedition.net/document?sigil=B.a&page=59&view=print&section=5#l813`"

5. Overview of applied versioning strategies

	version name	work in progress	collaborative	documented changes	former versions available
Versioning as documentation	X	X		X	
Versioning as version management	X	X	X	X	X
Versioning as retrievable milestone versions	X	X			X
Mix type: Versioning as retrievable, documented milestone versions	X	X	X	X	X

Table 1. Overview of applied versioning strategies. X means: is provided.

6. Conclusion

The empirical analysis of DSE has shown that integrating versioning is still relatively rare in DSE projects. This is the case despite the fact that there are manifold practical motives for implementing a versioning strategy. Such as the idea that the editorial process should be shared with the public; the project has a long duration and editors would like to publish interim results; the material to be edited is too extensive and will be made accessible in publishable stages; or the editorial team and the collections are so scattered that it would make sense and be beneficial for the overall project to publish partial results. For whatever reason, all projects intend to communicate changes and to be transparent in such a way that makes the DSE more reliable and trustworthy.

All projects presented here clearly pursue the strategy of making the DSE available to the public, although the editions are work in progress. The most technologically undemanding practice to communicate the changes is to maintain a version documentation. Many more ways to make changes transparent are given when the edition is linked to a version control system. In these cases, it is also easy to make a version comparison and

to understand the contributions of individual editors. Each resource has its own version history and can be retrieved. In the presented projects of the type 'versioning as retrievable milestone version' this is instead not the aim of the editors. The version number refers to the complete edition. Individual contributions can be searched under the milestone version number. A version control system is not absolutely necessary for the technical implementation, but it is an advantage. It is part of the workflow to have a published version and at the same time to have a new version in processing. The encapsulated data storage in the system is important for the addressability and availability of past versions. In addition, the information resources gain in quality if the traceability of the changes is also collectively available as documentation. By this means, versioning works like an apparatus in a broader sense, and one which verifies editorial decisions by making the work in progress transparent to the users.

References

Ainsworth, Peter, and Croenen, Godfried, eds. 2018. *The Online Froissart.* Version 8. Accessed 1 September 2019. `https://www.dhi.ac.uk/onlinefroissart/apparatus.jsp`

Burnley, David and Wiggins, Allison, eds. 2004. *Auchinleck Manuscript.* Version 1.1. Accessed 1 September 2019. `https://auchinleck.nls.uk/editorial/refer_site.html`

Bleier, Roman 2019: "How to cite this digital edition?" Forthcoming.

Bohnenkamp, Anne, Henke, Silke, and Jannidis, Fotis, eds. 2019. *Johann Wolfgang Goethe: Faust.* Version 1.2. Accessed 1 September, 2019. `http://v1-2.faustedition.net/document`

Driscoll, Matthew James, and Pierazzo, Elena eds. 2016. Digital Scholarly Editing: Theories and Practices, edited by, Cambridge, UK: Open Book Publishers. DOI: 10.11647/OBP.0095.02.

Ette, Ottmar, ed. 2019. *edition humboldt digital.* Berlin-Brandenburgische Akademie der Wissenschaften, Berlin. Version 5. Accessed 1 September, 2019. `https://edition-humboldt.de/index.xql?l=de`

Github, Glossary: *Contributors* `https://help.github.com/en/articles/github-glossary` Accessed 14 August 2019.

--- *Commit* `https://help.github.com/en/articles/github-glossary` Accessed 14 August 2019.

Hitchcock, Tim, Shoemaker, Robert, Emsley, Clive, Howard, Sharon, and McLaughlin, Jamie, eds. 2018. *The Old Bailey Proceedings Online* Version 8. Accessed 1 September, 2019. `https://www.oldbaileyonline.org/`

MLA Committee on Scholarly Editions. MLA Statement on the Scholarly Edition in the Digital Age. Web publication, May 2016, `https://www.mla.org/Resources/Research/Surveys-Reports-and-Other-Documents/Publishing-and-Scholarship/Reports-from-the-MLA-Committee-on-Scholarly-Editions/MLA-Statement-on-the-Scholarly-Edition-in-the-Digital-Age`

Österreichische Akademie der Wissenschaften (ÖAW), eds.2019. *Catalogue of digital editions.* Accessed 1 September 2019. `https://dig-ed-cat.acdh.oeaw.ac.at/`

Siemens, Ray et al. 2012. *Toward Modeling the Social Edition: An Approach to Understanding the Electronic Scholarly Edition in the Context of New and Emerging Social Media.* In: Humanities Commons: 445 - 461 `https://hcommons.org/deposits/item/mla:83/`, `http://dx.doi.org/10.17613/M6KS3G`

Strobel, Jochen, and Bamberg, Claudia, eds. 2019. *August Wilhelm Schlegel Edition.* Version-07-19. Accessed 14 August, 2019. `https://august-wilhelm-schlegel.de/version-07-19`

Trautmann, Marjam, and Schrade, Torsten, eds. 2019. *Der Sturm. Digital Source Edition on the History of the International Avant-Garde.* Accessed 1 September 2019. `https://sturm-edition.de/`

The Duke Collaboratory for Classics Computing and the Institute for the Study of the Ancient World, eds. 2019. Papyri.info Accessed 14 August 2019. `http://papyri.info/`

Università di Parma: Linee Guida per la Buona Pratica Scientifica e Disseminazione della Ricerca. 2018, `https://www.unipr.it/sites/default/files/albo_pretorio/allegati/19-06-2018/linee_guida_dr_emanazione.pdf`

Wikibooks,The Free Textbook Project. 20 Sep 2018. *The Devonshire Manuscrip*t. Accessed 15 September 2019. https://en.wikibooks.org/w/index.php? title=The_Devonshire_Manuscript&oldid=3469218

Wikibooks, The Free Textbook Project. 23 May 2014. *The Devonshire Manuscript/A Note on this Edition.* Accessed 15 September 2019. https://en.wikibooks.org/w/index.php? title=The_Devonshire_Manuscript/A_Note_on_this_Edition&oldid=265930_6