Journal of the Text Encoding Initiative

Issue 4 | March 2013

Selected Papers from the 2011 TEI Conference

Beyond TEI: Returning the Text to the Reader

Christian Wittern

Electronic version

URL: <http://journals.openedition.org/jtei/691>

DOI: 10.4000/jtei.691

ISSN: 2162-5603

Publisher

TEI Consortium

Electronic reference

Christian Wittern, « Beyond TEI: Returning the Text to the Reader », *Journal of the Text Encoding Initiative* [Online], Issue 4 | March 2013, Online since 25 February 2013, connection on 20 April 2019.

URL : <http://journals.openedition.org/jtei/691> ; DOI : 10.4000/jtei.691

This text was automatically generated on 20 April 2019.

TEI Consortium 2013 (Creative Commons Attribution-NoDerivs 3.0 Unported License)

Beyond TEI: Returning the Text to the Reader

Christian Wittern

AUTHOR'S NOTE

I would like to thank the anonymous readers of this journal for their suggestions and comments, which helped improve this contribution.

1. Introduction

- 1 Much research and practical effort has gone into the development and maintenance of a digital format that could provide a stable foundation for texts in the digital age. The results of this work in the form of the *Guidelines for Electronic Text Encoding and Interchange* have been widely adopted in the community.
- 2 Considerably less effort has gone into the question of how the texts thus encoded will be published. There is no common model—indeed, there is not yet a credible theory—of reading in the digital age. It should also be noted that the interests of creators of the texts, which at the moment also determine how texts are disseminated, do not necessarily match—in fact most surely mismatch—the interests of those desiring to read the texts, be they scholars interested in studying and close reading of a primary text or casual readers in search of reading material. It should also be noted that ownership and distribution gets quite a bit more complicated with electronic texts compared with printed books, which it is possible to buy, lend, borrow and resell as one pleases.
- 3 In the following article, I will first look at common distribution methods for digital texts as they are currently seen, then make a brief digression to acts of reading in the world of print. This will lead me to a proposal for implementing what I will call "active reading" in the digital medium, a proposal that includes both an architectural part and ideas for a

concrete implementation. The texts in question are digital critical editions of primary sources, intended to work as a primary reference for studying the work in question.

2 Distribution Methods for Digital Texts

- 4 There are three main distribution methods for digital text.

2.1 Publication on Distributable Media

- 5 Before the advent of the Internet, distribution on machine readable media was the norm, in most cases on CD-ROM or more recently on DVD-ROM. A few projects still distribute their results on such media, although due to the increasing size of materials, DVDs now seem to be used more often. The Chinese Buddhist Electronic Text Association (CBETA) publishes a yearly distribution of its collection of Buddhist canonical scriptures, which contains material that is also available for online reading at the website and for download in a number of formats. The main reason for maintaining a DVD for distribution is the censored and otherwise limited access to the Internet in China, especially for material of religious character.
- 6 Another recent example for the distribution of a scholarly DVD is the digital Klagenfurt edition of the works of Robert Musil (Musil 2009). Here the specialized viewer application, the number of digital facsimiles, as well as limits on distribution rights for these facsimiles, made a DVD the preferred option for the editors.

2.2 Web Publication

- 7 By far the most common distribution method for digital editions and in fact the default that comes to mind first is a web publication. This gives the editor a lot of freedom to explore ways in which to present the material, as this form of distribution is clearly still in an early, formative stage comparable to that of the incunabula of early print. Examples of such publications are:
- Vincent van Gogh—The Letters
 Edited through a collaboration between the Huygens Institute and the Van Gogh Museum, *Vincent van Gogh—The Letters* contains critically edited texts with facsimile, translation, and copious notes displayed in a variety of ways according to the needs of the users. There is also a printed version in 6 volumes.

The screenshot displays the Van Gogh Museum Amsterdam website interface. At the top, there are navigation links for 'Help | Updates | Credits | Contact | Home'. Below this is a search bar with options to search by 'period', 'correspondent', 'place', or 'sketches'. A 'Search' button is present. To the right, there are links for 'Van Gogh as a letter-writer', 'Correspondents', 'Biographical & historical context', and 'Publication history'. The main content area shows a sample facsimile of a letter, its metadata (Br. 1990: 024 | CL: -), and a transcription of the text. The letter is dated 'Helvoirt, Tuesday, 7 July 1874' and is addressed to 'Betsy Tersteeg'. The transcription is in Dutch and includes a list of line endings.

Figure 1. Sample facsimile, metadata, and transcription from the Letters of van Gogh.

- Mark Twain Project Online (MTPO)

The MTPO at the Bancroft Library, University of California is trying to "produce a digital critical edition, fully annotated, of everything Mark Twain wrote" (<http://www.marktwainproject.org/>). There is also an accompanying print edition, but in this case the critical apparatus is only available online. The online edition also provides some functionality for the user to add notes and bookmarks to the site, which can be read on subsequent visits.

The screenshot shows the Mark Twain Project Online (MTPO) website. The main content area displays a sample letter to Elisha Bliss, Jr., dated '8 August 1876 • Elmira, N.Y.'. The letter text is shown with various annotations, including 'cite' markers and 'Explanatory Notes | Textual Commentary'. The website also features a search bar, navigation links, and a footer with contact information and copyright details.

Figure 2. Sample letter with annotations from the Mark Twain Project.

8 Besides the idiosyncrasies of the interfaces, there are some serious problems for scholarly users in this form of publication:

- A user never "owns" the publication in the same way a book (or even a CD-ROM) can be owned. This means not only no marginal notes, but also the risk that access to the site might become impossible due to an interruption on the network, discontinuation of the service, or some other failure.
- A user cannot insert paper slips or scribble in the margins of the publication.
- If there is a feature for leaving notes, as in the MTPO publication, these notes, which constitute an essential part of the work of a scholar, are stored on servers out of her control, and indeed not owned by the scholar who created them.

2.3 Download of Text Files

- 9 One of the oldest ways to distribute electronic texts over the Internet is to place them on a site where they can be downloaded. The oldest project to do so is *Project Gutenberg*, founded by the late Michael Hart in 1974. Since then, it has produced tens of thousands of texts and is the first destination for many readers looking for material to read. The texts offered here are not scholarly editions; in fact in most cases no information about the edition used to transcribe the text is available. These texts are thus not usable for many scholarly purposes.
- 10 Another project offering texts for download is the above-mentioned CBETA, which offers texts derived from the critically edited texts in various formats for download.
- 11 This model of distribution has been re-gaining popularity with the recent advent of dedicated ebook readers like Amazon.com's Kindle or smartphone or tablet applications like Stanza or iBooks, which allow the user to assemble a library of ebooks to be carried along and accessed at will.

3. Traces of Reading

- 12 The physical copy of a book printed on paper not only offers on its pages the text for the consumption of the reader, but also serves as a canvas to keep traces of the interactions between the book and the reader. These traces can be simply indications of the rhythm of the text, as in figure 3, or they can be notes the reader puts down when trying to understand a text, which are especially necessary if the text is not in the reader's native language, as shown in figure 4, or anything in between. One of my teachers used to paste additional paper slips onto the page when he ran out of space in the margin, in addition to maintaining a list of noteworthy textual locations at the front or back cover of the book. If a book is used in such a way for an extended period of time, such marks will form a record of previous readings and interactions with the book.

Figure 3. A 16th-century woodblock print of the Chinese historiographical work *Zizhi tongjian*, with punctuation marks added in red by an unknown reader. *Zizhi tongjian* (literally “Comprehensive Mirror to Aid in Government”), by Sima Guang (1019–1086), was first published in 1084 and covers the history of the Chinese empire from 403 BCE to 960 CE. The copy shown is held in the Library of the Institute for Research in Humanities, Kyoto.

Figure 4. Page from a book on Erich Fromm, annotated in Japanese by its previous owner. I found this book in a used bookstore in Kyoto; it contains articles and commentary in German on the German-born philosopher Erich Fromm.

- 13 With the advent of digital distributed editions of texts, the ability to make use of such marks has decreased. For one thing, in most of the distribution forms it is simply technically impossible for the reader to leave traces of readings in a text. Even where this is made possible by the publication itself, as in the case of the Mark Twain project, or can be achieved through third party tools like the "Awesome Highlighter" (<http://www.awesomehighlighter.com/>) or the Open Annotation Collaboration (<http://www.openannotation.org/>), it is highly doubtful whether all these tools, together with the supporting infrastructure, will stay in place for 20 or 30 years. A different approach to distribution and annotation thus seems necessary. It should be emphasized from the outset, however, that this new approach is not intended to *replace* existing models of scholarly publication, but rather is meant to provide *additional* avenues for publication.

3.1 Traces of Reading in Digital Texts

- 14 It goes without saying that all kinds of "traces of reading" that are possible in printed texts should also in some way be transferrable to digital texts. Part of making this possible will be an interface that allows the reader to interact with digital texts. My concern here, however, is not with implementations of such an interface, but rather with an architectural framework, that will make the implementation of interfaces possible.
- 15 Margins in digital texts are theoretically without limits, so annotation of unlimited length could be added to a text. Thus, among other things, it would be possible to scribble an entire translation of a text into such digital margins. This might seem to overstretch the notion of annotation, but has the advantage that the text is automatically linked to the translation.
- 16 Whereas printed texts might occasionally see revised editions or have a new preface added in a later edition, they are for all practical purposes static texts, which do not get updated. Digital texts, on the other hand, are much more likely to be updated frequently, even if only to correct misprints noticed since the last publication. In fact, as a reader of digital text, my fingers become itchy if I spot an error, and I would love to read it in the context of a system that easily allows me to update such a text for the benefit of later readers. Any system that tries to solve the problem of digital annotations has thus to take into account the fact that texts may change; it has to ensure that changes can be published and picked up by readers, and that these changes interoperate well with annotations that readers might have made to their own versions of the text.
- 17 Another key point that distinguishes traditional, personal annotations from their digital equivalents is that the digital variant has the potential to be shared. Again, this depends largely on the protocol and underlying architecture of the digital system, but it could be set up in a way that allows several levels of access to annotations: they could be completely private, shared in one or more groups, or public. Groups of scholars could thus be set up to collaboratively annotate or even translate a text.¹

4. Distributed Version Control Systems (DVCS)

- 18 The proposal made here is of an architectural nature, which tries to point out how the goals outlined above can be achieved, namely, how scholarly editions could be distributed in a way that ensures long-term usability by and free interaction with the reader. It is by

no means a finished system ready for distribution and adaption,² but rather a proof of concept which I hope will improve ongoing attempts to explore the fascinating landscape of electronic reading.

- 19 There are different models for achieving such a distribution. The model proposed here does not rely on the continued existence of a central authority or the continued existence of a sophisticated infrastructure and is thus designed in a way similar to the Internet itself. It does however assume the existence of the Internet and the ability to communicate over the protocols it provides. It goes without saying, however, that if the infrastructure goes missing, not all of the functionality will be preserved, but at the very least, scholars already set up with the texts they are using will be able to continue to work with them.
- 20 The enabling piece of architecture for this model is a so-called "Distributed Version Control System" (DVCS). Such systems are currently mainly used in software development, but other uses are already beginning³ and will likely become more widespread.
- 21 Open source software projects typically are distributed across the globe with little or no hierarchy or control among the developers. Central repositories of source code with tight access controls, which had been used widely until a few years ago, proved to be a mismatch to this mode of operation. Consequently, DVCS, which support a different style of organization, have been developed. There are a number of different solutions, with procedures that vary slightly.⁴ Here I will use the program Git, developed by Linus Torvalds for the maintenance of the Linux kernel, as an example.
- 22 Git distinguishes a remote repository, which is accessible by all developers, and a local repository, which is only accessed by the user of the local machine. In addition to that, there is a place for doing the actual editing work, which I will call the "workspace." Both the remote and the local repository can and typically will contain several branches, that is, different versions of the codebase, just as a text might have different versions.
- 23 A typical workflow in development will be that a developer will "clone" the repository or a part of it. Cloning copies the files, together with associated information about the development history and the various branches in existence, from the remote to the local repository and usually also creates a workspace copy. The developer can then edit and change the content, keeping track of this process by committing (i.e. saving) the changes back to the local repository. At any time, the developer may look at what changes have been made in the remote repository and "pull" others' changes, merging them into the local repository. When ready, the developer might "push" edits to the remote repository, given that the rights to do so are available. However, frequently developers will announce their changes to their peers and invite them to pull them into their own repositories. The maintainer of the remote repository might also pull them. They might then be merged into one of the branches of the remote repository or continue to exist as a separate branch on the remote repository, making them available to all other users who might be interested.

4.1 DVCS for Scholarly Publishing

- 24 This workflow would translate quite nicely to work on electronic editions if they were made available as Git repositories. Users could clone texts or repositories they are

interested in, and from that time onwards pull any changes the publishers of the edition might have made. If they have made local corrections or annotations or added more witnesses, those will not be overwritten, but can be merged, which will allow the users both to work with their local versions as their research requires, and to keep track of the remote changes. In addition they could share their own work, whether annotations or corrections, with the editors of the edition, or with groups of other researchers that share a similar interest. For example, the text of a work could be in one branch of the Git repository and the translation in another branch. Additionally, branches could be set up to maintain other versions of the texts, such as those produced by other projects, or to reflect specific historic editions of the texts.

- 25 In addition to recording the changes, Git keeps track of who made the changes, thus allowing for a very fine-grained attribution of the changes to the original editor. It would thus be immediately clear and verifiable who introduced what changes.

4.2 DVCS Can Record Multiple Editions

- 26 Another important advantage of a DVCS is that it can record different versions of a text in so-called *branches*. In the case of an existing XML edition it might be desirable to publish such a "branched" version of editions. The first step here would be to set up a repository that publishes the XML "master" version as it is. In the example used here, the master edition contains information about several other editions via a text-critical apparatus, encoded using the parallel segmentation method.⁵
- 27 In addition, the publisher of these texts, a community interested in these texts or some other party can transform the texts to a "flat" version, where most of the markup is removed and all editions are represented in branches of the Git repository,⁶ as shown in figure 5 (in this example the versions are represented by their sigla in Chinese characters).

Figure 5. XSLT based transformation from TEI P5 to established text and multiple textual witnesses.

- 28 If now a new, additional edition is added to this repository, instead of laboriously encoding a new edition in XML, this system makes it possible to simply create a new edition with exactly the same page breaks, line breaks, prefaces, colophons, etc., as the original.

- 29 In addition to recording the additional edition, the system can also be asked to produce a list of differences between editions. If that proves insufficient, a full collation can be performed and a collated edition produced.

Figure 6. Cloning of branches from CBETA to private repository.

- 30 Figures 6 and 7 show schematically how a repository can be cloned and then a separate edition be added to the private repository. Additional branches, which are meant to hold private "annotations" (in this case a translation into English and German), are also created.

Figure 7. Adding new private branches in Chris's repository.

- 31 If the user **Chris** decides to publish from this repository the "trans-en" branch—the branch that holds the text with its English translation—and the new edition 〔東禪寺〕, then **Alice** will be able to get not only the public branches from the CBETA repository, but also the new 〔東禪寺〕 edition and the English translation, thus combining what is of interest to her in a new private repository on her own computer, as is shown in figure 8.

Figure 8. Alice clones from the public repositories CBETA and Chris.

- 32 Finally, CBETA might be interested in the new 〔東禪寺〕 edition added by Chris and add this to its own repository, which will make it more easily discoverable and might also indicate an endorsement of the new version.

Figure 9. CBETA adds branch 【東禪寺】 to repository.

5. Wrapping Up

- 33 The details of the format used for the editions and the application implemented as a prototype will be left to another publication,⁷ but just to show how radically the term "annotation" is used here, figure 10 shows a screenshot of the application with the text on the left-hand side with punctuation and paragraphs added, a translation on the right hand column of the text, and additional notes between lines starting with :zhu: and :END:.

云：	Er antwortete:
「如某所明，	"Soweit ich das verstehe,
二信之中是名緣信。」	ist es das, was die Verbindung der beiden Möglichkeiten genannt wird."
師曰：	Changsha fragte:
「依何教門得生緣信？」	"Wo hast du denn etwas über diese Verbindung gelernt?"
大德云：	Der Mönch antwortete:
「據《華嚴》云：	"Im <i>/Avatamsaka Sutra/</i> heißt es an einer Stelle:
『菩薩摩訶薩以無障無礙智慧，	'Wenn Bodhisattvas und Mahasattvas mit unbegrenzter Weisheit erkannt
werden，	
信一切世間境界。	dann glaubt man an die ganze Welt,
是如來境界。』	und diese Welt ist das Reich der Bodhisattvas.'
:zhu:	
Taisho Tripitaka, Electronic version, No. 0279 大方廣佛華嚴經 T10n0279:0273c07	
statt 信 heißt es dort 知。	
:END:	
又《華嚴》云：	Weiterhin wird im <i>/Avatamsaka Sutra/</i> gesagt:
『諸佛世尊悉知世法及諸佛法性無差別，	'Alle Buddhas wissen, daß die Dharmas der
Welt und die Buddhanatur	
決定無二。』	nicht unterschieden sind.'
:zhu:...	
又《華嚴》云：	Und schließlich, wiederum im <i>/Avatamsaka Sutra/</i> findet man:
『佛法世間法，	'Die Buddhas und die Dharmas der Welt,
若見其真實，	wenn man ihre wahre Essens ansieht,
一切無差別。』	so erkennt man, daß kein Unterschied zwischen ihnen besteht.'"

Figure 10. Text, punctuation, translation, and notes in the **chris-de** branch of the repository.

- 34 With some simple conventions for using a public distributed version control repository, an infrastructure can be put into place that will serve as a backbone for the distribution of scholarly editions. While they might not have the same bells and whistles that users have become accustomed to in modern web-based publications, they will require much less maintenance and will be serviceable long after the project that initially created the texts has been discontinued.
- 35 There is, however, another precondition for this to work, which is of a legal rather than a technical nature. In order for this to work, the cloning and copying mentioned above has to be not only technically possible, but also legal: the texts need to be made available under licensing terms that allow, or maybe even encourage, re-use in the way indicated here, which seems to me to be the precondition of any scholarly discourse involving digital text.

5.1 Distributed Ownership

- 36 As a researcher interested in a text, such as the 景德傳燈錄 *Jingde chuandeng lu* (CDL)⁸ used in the examples above, I will want to access all versions of the text, collate them, create annotations for certain parts of the text, link from the text to similar sections in other texts, and translate the text. In addition, I might want to mark names of places or persons in the texts, or enhance them by looking up georeferences and biographical information. All of this activity will create new digital objects that ideally I should be able to make available to other readers interested in the text. But since I do my work as part of my job as researcher, I cannot simply give up ownership of these additions, but need to be able to tag them with my name, and I need to be able to control how these objects are used. This will also serve as an indicator of their trustworthiness (or lack thereof), based on my reputation among the groups of users of these texts. Use of the DVCS as indicated above does exactly support this kind of publication model.
- 37 "Web 2.0" has enabled many people to contribute to websites and enhance their content, but only a few of these sites have acknowledged the contributors and made their contributions trackable.⁹ In many cases, registering at a site as a user will make all information created by this user owned and controlled by the site owner, not the user. A new way to enable better sharing and adding of content is needed that does not force the user to give up any rights.
- 38 It might also be worth pointing out here that the model made popular by Wikipedia, which relies on one central website, where only one version of a certain fact can be stated, does not seem to be very well suited for scholarly discourse, where a multitude of opinions and theories will coexist without an arbiter.

5.2 Other Models

- 39 It should also be obvious that the model sketched here is not the only possible development we might see. While the web matures into a hive of interlocking and thriving communities of users, some parts of the web are cordoned off as "walled gardens" that operate under a completely different model. One example of such a walled garden is Apple's AppStore,¹⁰ which maintains a strict division between sellers and consumers and tightly controls what developers are allowed to sell—a completely different mode of operation and a completely different set of infrastructure elements

from the one I am proposing (although they are ultimately built on the same foundations as the web itself).

- 40 This underlines the fact that there is no inherent infrastructure in the digital world; our imagination and the momentum it creates will determine future developments. But even here, it is the sequence of the decisions of the users and all stake-holders that influence the direction these developments take at every given moment.

6 Conclusions

- 41 These are not merely arcane technical issues of little relevance to end users; they do in fact influence the directions that all digital media and the internet are taking. Free flow of information is essential for the development of scholarship; any attempt to limit this freedom should not be allowed to succeed.
- 42 Although scholarly communication as such does not seem to be under immediate threat from these developments, it still seems to be advisable to build infrastructure for scholarly communication independent of short-sighted and self-serving interests of private, for-profit companies. The architectural model proposed in this paper tries to contribute to a fundamental protocol that could form the base for applications serving the long-term needs of research and scholarship.

BIBLIOGRAPHY

- Bradley, John. 2006. "A Response to the B2C 'Cultural Hegemony' in Humanities Computing: Pliny." Poster presented at Digital Humanities Conference, Paris, July 5–9.
- Caton, Paul. 2007. "Distributed Multivalent Encoding." In *Digital Humanities 2007: Conference Abstracts*, 2nd edition with corrections, 30–32. Urbana-Champaign: University of Illinois Graduate School of Library and Information Science. <https://www.ideals.illinois.edu/handle/2142/2368>.
- Jackson, Larry S. 2010. *RDF-encoding Pliny Annotations in the Open Annotation Collaboration Project*. Technical report #ISRN UIUC LIS—2010/2+OAC. Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign. <http://hdl.handle.net/2142/15418>.
- Jansen, Leo, Hans Luijten, and Nienke Bakker, eds. 2009. *Vincent van Gogh—The Letters*. Amsterdam: Van Gogh Museum. <http://www.vangoghletters.org/vg/>.
- Musil, Robert. 2009. *Klagenfurter Ausgabe. Kommentierte digitale Edition sämtlicher Werke, Briefe und nachgelassener Schriften. Mit Transkriptionen und Faksimiles aller Handschriften*. Edited by Klaus Amann, Walter Fanta, and Karl Corino. Klagenfurt: Robert Musil-Institut der Universität Klagenfurt. CD.
- Robinson, P. 2009. "Towards a Scholarly Editing System for the Next Decades." In *Sanskrit Computational Linguistics: First and Second International Symposia. Rocquencourt, France, October 29–31, 2007; Providence, RI, USA, May 15–17, 2008: Revised Selected and Invited Papers*, edited by Gérard Huet,

Amba Kulkarni, and Peter Scharf, 346–57. *Lecture Notes in Artificial Intelligence* 5402. Berlin: Springer.

Schmidt, Desmond, and Robert Colomb. 2009. "A Data Structure for Representing Multi-version Texts Online." In *International Journal of Human-Computer Studies* 67(6): 497–514. doi:10.1016/j.ijhcs.2009.02.001.

Shillingsburg, Peter. 2006. *From Gutenberg to Google: Electronic Representations of Literary Texts*. Cambridge: Cambridge University Press, 2006.

———. 2009. "How Literary Works Exist: Convenient Scholarly Editions." In *DHQ: Digital Humanities Quarterly* 3.3. <http://www.digitalhumanities.org/dhq/vol/3/3/000054/000054.html>.

———. 2010. "How Literary Works Exist: Implied, Represented, and Interpreted." In *Text and Genre in Reconstruction: Effects of Digitalization on Ideas, Behaviours, Products and Institutions*, edited by Willard McCarty, 165–82. Cambridge: Open Book Publishers.

Siemens, Ray, Cara Leitch, Analisa Blake, Karin Armstrong, and John Willinsky. 2009. "It May Change My Understanding of the Field": Understanding Reading Tools for Scholars and Professional Readers." In *DHQ: Digital Humanities Quarterly* 3.4. <http://www.digitalhumanities.org/dhq/vol/3/4/000075/000075.html>.

TEI Consortium. 2012. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*, edited by Lou Burnard and Syd Bauman. Version 2.2.0. Last updated October 25. N.p.: TEI Consortium. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>.

NOTES

1. It should be noted that some electronic reading devices, for example the Kindle, already implement some of the functionality for sharing highlights and annotations. While taking notes and highlighting passages was possible on a Kindle early on, in May 2009 it also became possible to view these annotations via amazon.com. They are also shared between different devices of the same owner. Since these annotations are also stored on the device in a separate, user-accessible file, it is possible to share them with others. In addition, there is a "Popular Highlights" feature, where amazon.com aggregates all highlights to a text into one view.

2. This is a difference from, for example, the Pliny project developed by John Bradley (see Bradley 2006 and <http://pliny.cch.kcl.ac.uk/index.html>, but also Jackson 2010). The Pliny project takes the development of an application for scholarly annotation built on a specific platform as the point of departure, whereas my emphasis here is on a cross-platform protocol.

3. The Integrating Digital Papyrology project, for example, started to make its data (some 55,000+ texts) available on the website github.com in January of 2011: see <http://digitalpapyrology.blogspot.com/2011/01/idp-data-available-on-github.html>.

4. The list in Wikipedia (http://en.wikipedia.org/wiki/List_of_revision_control_software) as of December 10, 2012, had fourteen entries for the distributed model in the open source category alone.

5. See the *TEI Guidelines*, chapter 12: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TC.html>.

6. The texts will retain an implicit connection to their XML sibling, but this connection is currently only spelled out in the application logic that transforms to and from the XML version. It should also be noted that this setup makes it impossible to use the mechanism for annotation and commentary introduced below to point to the markup itself.
 7. Some information is also available on the project web page, <http://www.mandoku.org>.
 8. The *Jingde chuandeng lu* is one of the defining texts of the Chan school of Chinese Buddhism. It was submitted to the Song emperor in 1004 and immediately slated for inclusion in the officially sanctioned authoritative collection of Buddhist texts, the Chinese Buddhist Canon. The oldest existing print of this text is from the so-called Dongchan Tempel edition in Fujian province, of which one set is preserved at Tofuku Temple in Kyoto. This text differs in some respects from the versions that have been transmitted in other editions of the Canon, namely the Taisho edition, which has been used by the CBETA project to create an electronic version of the text.
 9. Wikipedia (<http://en.wikipedia.org>) is a good example of a website that does indeed acknowledge contributions; nevertheless, it forces its users to give up the rights to their contributions and publish them under the conditions set by the Wikipedia project.
 10. Apple opened a web store for its mobile devices in July 2008. Access to the store is tightly controlled—for developers, who have to pass a review process for every application (and update), as well as for end users, who need to have an account and a valid payment method to access. This policy has received heavy criticism and in 2010 even an investigation by the Federal Trade Commission of the US. For more information, see <http://en.wikipedia.org/wiki/Appstore>.
-

ABSTRACTS

Much research and effort has been invested in creating a versatile format for digital texts and the TEI is now widely used in many communities. Much less consolidated thought has been spent to publish and distribute digital texts in ways that are most useful to scholars.

To remedy this situation, this paper proposes new, additional publication forms for digital texts through distributed version control systems. This will allow publication and maintenance of several different versions of a text. In some respects, this will be similar to publishing a college or paperback edition of the text established in a critical edition. In addition to this, the user of a text published through such a system can subscribe to later changes or corrections of an edition. The architectural model proposed in this paper tries to contribute to a fundamental protocol that could form the base for applications serving the long-term needs of research and scholarship.

INDEX

Keywords: electronic publication, scholarly communication, distributed version control, annotation

AUTHOR

CHRISTIAN WITTERN

Christian Wittern is Professor at the Center for Informatics in East Asian Studies, Institute for Research in Humanities, Kyoto University, Japan