
Implementing New Knowledge Environments (INKE)

Publications

The Changing Culture of Humanities Scholarship: Iteration, Recursion,
and Versions in Scholarly Collaboration Environments

Susan Brown, John Simpson, the INKE Research Team, & CWRC Project Team

December 16, 2014

© 2014 Brown, Susan, Simpson, John, INKE Research Team, & CWRC Project Team. This Open Access article is distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc-nd/2.5/ca>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article was originally published at:

<http://src-online.ca/index.php/src/article/view/191/354>

Brown, Susan, Simpson, John, the INKE Research Team, & CWRC Project Team. (2014). The Changing Culture of Humanities Scholarship: Iteration, Recursion, and Versions in Scholarly Collaboration Environments. *Scholarly and Research Communication*, 5(4): 0301191, 16 pp.

The Changing Culture of Humanities Scholarship: Iteration, Recursion, and Versions in Scholarly Collaboration Environments

Scholarly and Research
Communication

VOLUME 5 / ISSUE 4 / 2014

Susan Brown

University of Guelph & University of Alberta

John Simpson,

University of Alberta

the INKE Research Team, & CWRC Project Team

Abstract

The non-linear and iterative nature of scholarly research processes presents complexities with respect to how online collaborative systems manage versions both within interfaces and at the back end. This article maps out a two-part framework for thinking about versions and versioning in the context of contemporary scholarship and data preservation. The first presents four notable qualities of digital textuality that are intensified by the digital turn, and the second considers technical considerations flowing from these characteristics. The authors argue that the management of large humanities data sets and the design of associated interfaces, tools, and infrastructure need to recognize and preserve the dynamic, living nature of digital cultural artifacts and of scholarship on culture.

Keywords

Versioning; Editing; Textuality

Susan Brown is Professor of English at the University of Guelph and Visiting Professor in English and Film Studies, and Humanities Computing, at the University of Alberta. Email: sbrown@uoguelph.ca

John Simpson holds a PhD in Philosophy and is a postdoctoral fellow at the University of Alberta pursuing research on the Semantic Web and Linked Open Data. Email: jes6@ualberta.ca

INKE Research Team: Implementing New Knowledge Environments is a Major Collaborative Research Initiatives research grant funded by the Social Sciences and Humanities Research Council of Canada.

CWRC Project Team: Jeffery Antoniuk, programmer and systems analyst, Michael Brundin, data integrity and metadata coordinator, and Mihaela Ilovan, project manager, with the Canadian Writing Research Collaboratory infrastructure project.

CCSP Press

Scholarly and Research Communication

Volume 5, Issue 4, Article ID 0301191, 16 pages

Journal URL: www.src-online.ca

Received August 1, 2014, Accepted August 25, 2014, Published December 16, 2014.

Brown, Susan, Simpson, John, the INKE Research Team, & CWRC Project Team. (2014). The Changing Culture of Humanities Scholarship: Iteration, Recursion, and Versions in Scholarly Collaboration Environments. *Scholarly and Research Communication*, 5(4): 0301191, 16 pp.

© 2014 Brown, Susan, Simpson, John, INKE Research Team, & CWRC Project Team. This Open Access article is distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc-nd/2.5/ca>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Versioning is endemic to cultural scholarship. Mash-ups are associated with the types of social and participatory media that have increasingly characterized Internet culture since the turn of the century, and it has been argued that there are quite important specificities to mash-ups that are traceable to the affordances of digital technologies (Sonvilla-Weiss, 2010). But before digital tools we had the typewriter, which Hannah Sullivan (2013) blames for the modernist bent toward revision. Yet Sullivan overstates the case in arguing that revision was not valued among the pre-moderns. There is a strong argument that current mash-up practices are merely a further, technologically enabled stage in the continual process of remediation, recycling, and renewal that has always constituted culture and engagement with cultural artifacts.

Jerome McGann and Lisa Samuels propose that the “deformance” flowing from critical engagement with the cultural object is “as ancient as our currently more normative practices” (McGann, 2001, p. 106). According to the futurist and technology visionary Ted Nelson (1999), “The real work of writing is rewriting; and especially in big projects, is principally the overview and control of large-scale rearrangement—a rearrangement process that used to be called ‘cut and paste’ until those terms were redefined by the Macintosh in 1984” (p. 6). Literary theorists agree. Read Bakhtin, or Barthes, or Kristeva, who argue that there is no such thing as an original utterance. Read Woolf, Joyce, Shakespeare for confirmation. Ariel Katz puts it beautifully: “Civilization is an open-source project” (n.d.).

A scholar’s ode to a work of culture, *pace* Elizabeth Barrett Browning, might commence: “How do I change thee? Let me count the ways.” Counting revisions, actual or possible, takes on new connotations within a digital context, and the number of ways in which we can change, transform, remix, annotate, revise, edit, remediate, or otherwise version a text has undoubtedly increased in the half century or so since Father Busa began systematically, with the help of a computer, to chop up the works of St. Thomas Aquinas in the service of scholarship. Jerome McGann (2001), Steve Ramsay (2011), and others see in digital scholarship another version of the “deformance” of texts that has always been involved in the hermeneutical process.

Technological developments have not only allowed for a proliferation of methods for versioning, editing, and changing a text, but also caused a sea change in the practice of versioning—or saving particular instances of digital objects at various points in the process of being altered—related to the materiality of digital culture. When a new version does not involve repeated inscription or imprinting and excessive consumption of further storage media, but rather the repeated rearrangement of electromagnetic data on the same medium, versioning becomes legion. The ease with which Web materials could be first copied¹ and then changed by people without technical expertise was of course exponentially enhanced by the shift to what we now call Web 2.0 technologies. This shift was effected by interfaces that made updating Web content simple, fast, and accessible in a wide range of platforms and contexts. Matthew Allen (2013) has argued that the “jumble of competing, but irreconcilable, differences of perspective and purpose” (p. 262) in our understanding of the Web makes versioning inextricable from the language of Web 2.0. The versioning that Web 2.0 enables turned reflexively inward and became a defining component of the term:

Web 2.0 more generally brought to the web the *discourse of versions*, a discourse that created a 'history' of the internet, constructing what it claimed to describe, and influencing our collective, public understanding of the Internet through this historicization as much as in any other way. (p. 261, Italics in original)

In this light, the implications of the practice of versioning for how we understand the transformative technologies of our times, and their impact on the sustainability of culture and cultural memory, are brought into relief. If the word processor is, as Matthew Kirschenbaum argues, changing the history and culture of authorship, how is the mutability of digital data altering the culture of scholarship? How we produce and account for versions becomes a matter demanding attention from a wide range of stakeholders.

This article maps out a preliminary two-part framework for thinking about versions and versioning in the context of contemporary scholarship and the related fields of archiving and data preservation. The first part presents four notable facets of textuality that are intensified by the digital turn: dynamic textuality; collaborative textuality; granulated and distributed textuality, and interdependent textuality. The second considers technical considerations flowing from these characteristics regarding control, cost, collaboration, conflicts and management, and representation. Our insights are drawn most immediately from work at the intersection of literary scholarship, tool and interface development, and infrastructure development taking place in several large digital humanities projects.² Given this immediate framework and its focus on the literary, this article will refer most often to digital texts while discussing the objects and the products of cultural scholarship, but many of the implications are general enough to apply to digital cultural artifacts more broadly, to other forms of scholarship in and beyond the humanities, and to digital data sets. Despite the fact that versioning beyond the world of DJs, samples, and remixes is far from sexy, we hope the version of this "paper" offered here, whether on an LED screen or processed wood pulp, proves a useful stimulus to reflection and debate.

Digital textuality: Ch-ch-ch-changes

Scholarly processes, perhaps humanist ones in particular, are far from linear, involving a dialectic between source materials and scholarly engagement that results in revisiting sources, revising sources, and producing iterative versions of materials as they are refined and revised over time. The meaning of some cultural artifacts, particularly modernist ones such as Eliot's *The Waste Land*, seems inextricable from the history of their revision. However, the nature of versioning today is significantly different. This is not to say that contemporary practices of versioning are completely dissimilar to earlier ones, but that current technologies facilitate versioning in much digital writing to such an unprecedented degree that they seem to bear a symbiotic relation to the mode of writing itself. The relation of writing to documents has changed, as has the relation of documents to instantiation. This section will direct attention to four aspects of digital textuality that are particularly pertinent to a consideration of versioning: it is dynamic, increasingly collaborative, granulated and distributed, and interdependent with other text or data.

DYNAMIC TEXTUALITY

The dream of modern word processing was first realized on December 9, 1968, when Doug Engelbart and his lab provided a live demonstration of the oN-Line System (NLS) (see Engelbart, 2008). In addition to introducing the computer mouse as an input device, the team provided the first public demonstration of collaborative editing in an interactive word processing environment that bore an uncanny resemblance to what we are familiar with today. It took decades for this technology to find its way into the lives of scholars, replacing many a typewriter in doing so. As a *writing* tool the typewriter was a marvel when first introduced in the later nineteenth century, conveying the power to produce standard, legible text more swiftly than handwriting to anyone fortunate enough to afford one and patient enough to keep writing despite jams and tangled ribbons. As an *editing* tool, it left much to be desired, especially since the first correction fluid was not invented until 1951, when then-secretary Bette Nesmith Graham decided that the choice faced by typists of living with their mistakes or retyping the page was itself in need of correction.

Contemporary writing and editing environments – the extent to which they are co-equal is instructive, as is the replacement of the noun “writer” with “editor” – by contrast, allow one to edit seamlessly as part of the process of inputting a text, making changes as simple as using a backspace key or selecting and modifying. Indeed, editing environments increasingly perform low-level corrections automatically as we type, situating the author as one who must vigilantly block any automatic changes that are not desired, a particularly sore point for anyone who has hurriedly composed a document on a smartphone. So texts of many types are ceaselessly changing in the very act of composition, but there is often little record of that process. Many new forms of cultural and scholarly output, such as blogs, databases, and websites, change repeatedly post-publication, preserving no record on the site of the changes that have occurred. Scholarly citation protocols rely on being able to point to a stable text as a source, but now must point not only to the date of initial publication but to the date of access of a Web resource, signalling awareness that the source as cited—or even the source in its entirety—may not be precisely the same as the one that was cited, which may be irretrievable by a future reader. David Greetham (1992), in an introduction to textual editing, anticipates that the rich record of changes to manuscripts prior to publication in the past few centuries may turn out to be a historical blip, “a brief anomaly in the history of textual evidence” (p. 75), but Hannah Sullivan (2013) muses whether the very notion of a “‘finished product’ to which a manuscript or pre-publication version stands in opposition has meaning only within print culture?” (p. 313, n. 92).

Matthew Kirschenbaum notes:

In the particular realm of literature and literary scholarship, this means that writers working today will not and cannot be studied in the future in the same way as writers of the past, since the basic material evidence of their authorial activity—manuscripts and drafts, working notes, correspondence, journals—is, like all textual production, increasingly migrating to the electronic realm. (Kirschenbaum, Farr, Kraus, Nelson, Stollar Peters, Redwine, & Reside, 2009, p. 3)

Certainly there are technical means of preserving the changes to digital artifacts, and, particularly in the case of text, the storage involved is cheap. Australian author and technophile Max Barry (2011) has released the entire edit history of the novel *Machine Man*, which was released first as an online serial and later revised, right back to the initial notes. Scholarly editing, that area of textual studies devoted among other things to tracking the minutest changes in the evolution of a text, has itself been transformed by the versioning capacities of digital media, so that, as Elena Pierazzo notes, digital editions can provide diplomatic, semi-diplomatic, or genetic editions providing richly detailed information about minute changes to a text more cheaply and effectively than print ones (2009). The digital medium allows for, as she puts it with reference to the Text Encoding Initiative's model for genetic editions, "the encoding of time" (Pierazzo, 2009, p. 170). Quite a contrast to David Bowie's claim, "Time may change me / But you can't trace time" (1971).

However, notwithstanding the capacity of digital technologies to support the tracking and the representation of minute changes, the practicalities of doing so constitute a major challenge, particularly in conjunction with the other three aspects of digital textuality we identify here.

COLLABORATIVE TEXTUALITY

The ability to collaborate, including in real time, is one of the most exciting and transformative aspects of digital writing environments. Both online and offline writing platforms increasingly support collaboration. Wikis, blogs, Google apps, and the Git and Subversion revision control systems are all content creation systems that support and help to manage the collaborative process, and traditionally closed systems by Adobe and Microsoft have been following suit. The process of producing scholarly editions is increasingly a social process, involving collaboration not only among scholars but also between scholarly communities and the public through such means as editing and crowd-sourcing (Causer, Tonra, & Wallace, 2012; Siemens, Timney, Leitch, Koolen, & Garnett, et al., 2012; Terras, 2011). The participation of multiple individuals adds another layer of complication to versioning. Again, this is by no means new in the history of cultural or scholarly production, but in the context of real-time collaboration – in particular the imbrication of changes by different contributors to an artifact – can be continuous. Within numerous contexts, and in particular the scholarly one, attribution of authorship remains very important to how credit and rewards are allocated, and collaborative textuality complicates the versioned digital text. Online collaboration intensifies the nested and recursive aspects of both digital textuality and collaborative writing, considerably complicating both what might be considered to constitute a version and the relationship of that version to the concept of authorship (Lowry, Curtis, & Lowry, 2004; Speck, Johnson, Dice, & Heaton, 1999).

GRANULATED AND DISTRIBUTED TEXTUALITY

The opposition is breaking down between data-centric and document-centric conceptualizations and treatments of data, particularly humanities data. In the age of mash-ups, digital resources are becoming more granular, with a sense that documents are comprised of multiple components that may be stored separately and even in different locations. This aspect of digital textuality makes editing akin to producing magnetic poems on a refrigerator door but where the content of each magnet is

constantly expanding and retracting. As Christian Wittern argues, “Large texts cannot be conveniently handled by today’s computers, so they have to be split into smaller parts” (2002).

The Text Encoding Initiative, for instance, supports the incorporation of references to external files, containing components such as images that might be considered integral to a document, within the markup of XML documents (<graphic>, 2014). Digital scholarly resources such as the Orlando Project “textbase” compose apparently cohesive digital objects from multiple files, and their systems blend the representation of the contents of entire documents with the search, retrieval, and display of components of those documents based on databasing the XML encoding. Digital scholarly editions can be and increasingly are comprised of components from a range of sources and locations. Indeed, the Shared Canvas data model offers “a linked data based approach for describing digital facsimiles of physical objects in a collaborative fashion” (Sanderson & Albritton, 2013, Abstract), as a basis for creating and rendering composite digital objects such as a medieval manuscript recompiled, virtually, from digitized images of sheets of vellum held in scattered archives.

INTERDEPENDENT TEXTUALITY

Linked data provides a nice segue into the extent to which dynamic, collaborative, and granular textuality is also increasingly interdependent. Whereas once upon a time the assumption was that documents or digital artifacts stood alone, this is decreasingly the case. Versioning becomes of paramount importance in a linked and interdependent universe of online textuality, given that any digital object can change independent of others to which it is interrelated. In an ideal world, we would have the kind of “docuverse” conceived of by Ted Nelson, who coined the term “hypertext” and who is through his Project Xanadu one of the most persistent critics of the World Wide Web, with its “one-way ever-breaking links and no management of version or contents” (Project Xanadu, 2012).

What is happening, however, is that we are moving toward tantalizing prospects of interoperability as a result of the increasing uptake of Linked Open Data approaches beyond and within academic research projects (Brown & Simpson, 2013), without having thought through the implications of versioning in a linked data environment. To take a fairly straightforward example, what happens if a scholar annotates a portion of a webpage using a linked data tool such as Pundit, and the page subsequently changes? Depending on how the annotation has been anchored in the webpage, the annotation may appear irrelevant or inappropriate, and nothing will indicate that the page has changed in the interim since it was created. The more that the Web is comprised of interdependent, granular, and dynamic bits of data that together produce composite digital objects, the more pernicious dead links and inadequately versioned data will become.

Technical and cultural considerations

Separately and together the dynamic, collaborative, granular, and interdependent qualities of digital data expose versioning as a major, multifaceted challenge as each prompts different answers to the question, “What will count as a version?” With a range of answers bookended by *everything* and *nothing*, this is by no means a question to set

aside as trivial. This question similarly forms the foundation for technical considerations related to versioning, with answers giving rise to subtle yet important implications for the preservation of cultural artifacts as well as the formation of the behaviours that contribute to the formation of culture in the first place. The cycle of influence between culture and technology is easily seen in hindsight by looking at how technologies such as gunpowder, the steam engine, the cotton gin, or more recently the Internet and the cellular telephone have left deep and lasting changes on the people of the world and the ways that we interact and think today.³ While it may seem that the technologies and technological considerations surrounding versioning could not possibly have the explosive and world-changing consequences of gunpowder, it should be recalled that versioning amounts to a record of the manipulation of representations of ideas, and the transmission of ideas effectively constitutes a series of versions. The development of ideas, legislation, and policy through time can be known only to the extent that traces of that vast chain of versions are preserved.

In what follows we consider the question “What will count as a version?” through the lens of the technical considerations that must be addressed in the production of a versioning system for text-oriented cultural scholarship. For each of these considerations we also suggest what cultural consequences and influences are at play. The technical considerations are control, costs, collaboration, conflicts and management, and representation.

CONTROL

In terms of controlling the versioning process, the options available amount to 1) allowing authors or other contributors to version as and when they see fit, 2) allowing a machine to automatically create versions at predetermined intervals, or 3) some hybridization of these. The first of these has historically been the method employed by writers and artists, simply because it was the only option reasonably available when keeping a version of a creative work amounted to ceasing to work on it any more in its current state and materially starting over. This changed somewhat with the advent of the printing press and typewriter, but near-effortless versioning was not widely achievable until the advent of modern computing. Computing technology has made it easy to copy electronic files, allowing for a copy of a work to be retained while another copy is further edited and revised. This is, in essence, what versioning amounts to in digital environments.

As easy as it is to version digital artifacts, it remains something that we are rarely good at without the application of conscious effort. It would be a rare individual indeed who did not know the frustration experienced at the loss of all the changes made to their work since the last time they chose to save it—possibly hours ago—due to a program or system failure. The same frustration is felt by the author who uses only one file to save the day’s writing, repeatedly writing over the previous day’s copy until accidentally erasing previous edits deemed to be of value. In the face of such events, the benefits of automating the versioning process are clear: it saves us from ourselves. Automation comes with its own potential challenges, however, among them the processing and storage costs, which we will set aside briefly. Another is the challenge of producing *meaningful* versions, which is particularly relevant in most scholarly contexts, where it is hoped that preserved versions of objects would represent significant changes worthy

of analysis. As Matthew Kirschenbaum notes, autosave programs and Microsoft Word's Track Changes make no judgments between changes trivial or momentous, saving according to their own algorithms or manual setting (Kirschenbaum et al., 2008). In a Web editing environment, where texts are being modified in a user's browser and changes sent to a server, every modified character could conceivably be considered to constitute a new version.

For instance, it is possible within some repository software, such as the Fedora Commons framework, to employ "built-in versioning." Turned on, as it is by default, this feature saves every version of every datastream for an object any time it is modified. Fedora is quite sophisticated insofar as it stores not only the earlier version of the content, but also its look and feel (Fedora Content Versioning, 2005). Nevertheless, it is an all-or-nothing system: versioning is either on or off. But if Fedora is not being used to store static digital objects but instead dynamic ones that are being edited in a Web-based tool, the number of versions can easily get out of hand. For instance, depending on the relationship among browser, server, and repository, automatic versioning might result in thousands of versions of a single object within a day, each one possibly representing only a tiny and insignificant change, such as the addition of the space that now follows the period at the end of this sentence.

On the whole, automated versioning of the kind conducted by Google apps, or even by the Way-Back machine as it snaps copies of pages around the Internet, may seem excessive, and indeed may be prohibitive in terms of transaction and storage costs. Some kind of aggregation of minute versioning activities seems preferable in terms of the readability and significance of versions. Such aggregations might be arrived at algorithmically, but at the very least allowing creators to define what might be considered *milestone* versions seems crucial.

The GitHub versioning environment, as *Wired* magazine observed, offers an interesting model for dealing with versioning well beyond the context of socially networked code development (McMillan, 2012). Jentery Sayers of the University of Victoria is using it to track materials related to his Makers Lab, his courses, and some scholarly work. Such idiosyncratic use of GitHub is spreading to everything from wedding invitations to contracts to Gregorian chants (McMillan, 2013; Sayers, 2014). Although GitHub displays changes, provides clues to what may be meaningful revision, and has the virtue of doing so in public unless one pays for a private repository, it is not really a user-oriented interface, however: notwithstanding the claim that it is now mainstream, it is unlikely to be adopted widely by scholars.

In summary, at this point quite a number of infrastructures for tracking changes and managing automatic versioning exist in a wide range of contexts, and yet we know very little about what constitutes effectiveness in such environments: systematic reviews and comparison of them would hugely benefit our understanding of what works and what does not both from a user perspective and with a view to optimizing the considerable costs involved in implementing and maintaining versioning systems.

COSTS

Versioning, whether done manually or through an automated system, comes with inherent costs that can be mitigated but never fully avoided. The majority of these costs are not fundamentally monetary, although they may be expressed in this form, and are instead tied directly to the consumption of storage space and human time. Time is perhaps the first cost that any development team considering implementing a versioning system must consider, because adding versioning stands to significantly extend the time needed to design, implement, and deliver the final production environment. Time will also be a factor in terms of the potential wait time associated with producing or retrieving a version, particularly if doing so interrupts the experience of the end user, who is also likely to avoid versioning systems if they impede the scholarly workflow. Space becomes an important factor because every version must be stored somewhere. While there are clever ways to prevent each version from consuming as much space as a full copy of the document (Soules, Goodson, Strunk, & Ganger, 2003), it remains the case that the more versions there are, the more space is consumed. The consideration of costs is importantly a cultural consideration, since cultural considerations will finally determine how much it matters to preserve versions and who will be responsible for bearing the associated costs.

COLLABORATION

Among the most significant changes to the authoring process introduced by digital media are the new capacities that they offer for collaborative authorship. Collaboration with others was previously much more difficult, and material limitations tended to constrain and contain interactions between authors. It was necessary to share tangible copies of documents or other works with potential collaborators to solicit their input, necessitating a process that was almost always sequential. While sequentiality might be (though many would argue that it often is not) representative of reading practices, it is not representative of conversational practices, which typically involve various moves – many of them non-verbal – that participants may simultaneously exercise. Digital environments offer the opportunity for many creative activities to be undertaken in parallel in ways that mirror the construction of conversations.

This parallelization brings with it the synergistic benefits of a conversation and fuses them with the semi-permanence associated with many traditional acts of creativity. One of the fullest experiences of this affordance is currently available through Google Docs, which allows users instantaneous access to all changes made by anyone accessing the document with the appropriate privileges. The result of using a tool like this to author a paper is a unification of authorship that can make it difficult for even the contributors themselves to track or account for who exactly wrote what and why. Although the relationship of specific changes to particular authors can be tracked, most tools including Google Docs lack the functionality to expose such information effectively in the end product. As a consequence, materials written in such tools more fully represent the fusion of authorial contributions rather than their conjunction. Authors who have made collaboration a significant component of their work, such as Edith Emma Cooper and Katherine Harris Bradley, surely would have chosen to author in a tool like Google Docs had it been available.⁴ Those of us who write in such collaborative fusion are Field's cultural heirs, though we are unnamed and legion. The new crop of collaborative creation tools challenges our sense of authorship and

perhaps even our individuality, allowing us to experience more acutely the ways in which our writerly identities are already fluid.

CONFLICTS AND MANAGEMENT

Versioning makes it possible for conflicting versions to occur, and may necessitate a decision about which version(s) to privilege or how to merge versions appropriately. This problem becomes particularly likely when separate copies of a work can be created and edited simultaneously. The consequence of such a conflict is almost invariably a human intervention to determine what should be kept and what should be discarded. The often contentious question of who should be making these decisions follows. Recognizing this, the focus of versioning tools is on preventing such conflicts in the first place and/or assisting in the adjudication process as much as possible. Git stands out as a particularly strong example of this among a number of revision control systems. When a new version of a document is presented, the system algorithmically presents anyone with the authority to merge the changes with the files under its domain with a list of the changes to be made and what they are meant to replace or revise, if anything. The authorized user can then sift through the changes and accept or reject them as appropriate. Wittern (2013) advocates Distributed Version Control Systems as a means of returning the text to the reader.

Version management also comprises saving snapshots or milestones. Typically these are intentional, user-determined, states, but they can also be produced automatically nightly or weekly by some software. Distinguishing between the versions can be accomplished by determining the file names according to some predetermined standard such as dates and times of the capture. Human-produced versions that should be distinguished by the actual content of the files, however, often elude such systematic solutions. When working with large volumes of dynamic and interdependent files, tools for attaching metadata to each file become an attractive option for retaining information related to workflow stages, responsibility information, or notes on content revisions, and where collaboration is extensive and workflow processes are complex, tools for tracking significant versions within collections become essential. While there exist a number of systems to support business processing and scientific workflows, these tend to be very complex to set up, and they are often geared toward automated workflows rather than flexible, recursive flows that incorporate human judgment and intervention, which may be more typical of born-digital resources (Dergacheva, Brown, Roeder, Peña, & Knechtel, 2013).

The greater the granularity, distribution, and interdependence of versions, the more significant the challenges presented by versioning. We have been trying to think through a practical model for a versioning ecosystem involving various kinds of digital objects, but chiefly text (HTML pages, XML pages), metadata for those texts, RDF annotations of those texts, and XML entities to which the RDF points. Any of these interdependent objects can and will change as a result of scholarly processes, requiring links between related objects of the same temporality, but also links forward (and perhaps also backward) between versions. Orienting users in relation to all the versions of a text will prove a major representational challenge given that our reading methods are still so heavily influenced by print paradigms that we find it hard to adapt to interfaces that depart even in apparently quite trivial ways from the standard

organization and presentation of printed texts (Brown, Adelaar, Ruecker, Sinclair, Knechtel, & Windsor, 2013).

REPRESENTATION

A further aspect of costs, implicit above in the assertion that projects incorporating versioning will take more time, is how the versions will be made accessible and represented to users. Versioning poses a number of interface challenges. Versioned data may require management, such as the approval of changes, during the production process, which means that management and the associated labour need to be built into workflows. Many textual workflows, for instance the production of a creative work or a scholarly journal issue, may be fairly characterized as the process of creating and managing versions.

Where versions matter to the end user, revealing them in a reading interface becomes important. A long history of managing the visual presentation of textual differences in scholarly print editions is informing work on digital editions and collation tools. We can learn much from tools like the Versioning Machine (see Schreibman, 2003), CollateX (2010), and Juxta (n.d.) about how to present multiple versions of texts, as well as from experimental interfaces such as Ben Fry's beautiful animation of the variations in the six editions of Darwin's *On the Origin of Species*, "The Preservation of Favoured Traces" (2009). However, as members of the Modernist Versions Project observe, few generalized and well-documented systems have developed (Huculak & Richardson, 2013). Moreover, whether we are talking about line-by-line comparisons, side-by-side collations, or text variant graph models, these are interfaces developed to foreground and allow one to focus on differences between versions as the aim of the reading experience (Andrews & van Zundert, 2013; Schmidt, 2013), whereas this is not likely to be desirable in most generic reading or viewing interfaces. What an interface should do ideally is to flag in some subtle way the existence of other versions but bring them to the fore only if needed or if the reader elects to see them. So we need something less insulated from our consciousness than the history pages of wikis, but something that indicates, for instance, whether a more recent version of a page is available.

Not only reading interfaces but also production interfaces and paratextual interfaces will need to incorporate version management. The INKE Interface Design team has been developing a visual interface to handle workflows for such processes as journal editing or the collaborative production of digital scholarship, and an earlier project pioneered the representation of credit for contributions to wiki texts (Arazy, Stroulia, Ruecker, Arias, Fiorentino, Ganev, & Yau, 2010; Frizzera, Radzikowska, Roeder, Peña, Dobson, Ruecker, Rockwell, Brown, et al., 2013). Yet aside from the context of textual editing, focused attention on problems of versioning is still rare within the context of humanities infrastructure and tool development, even as we turn for interoperability in the direction of linked data approaches that intensify the challenges of versioning. The new scholarly infrastructures, which Susan Schreibman characterizes as "emerging distributed, interactive production and processing environments that go well beyond traditional working paradigms in the scholarly culture of the humanities" (Schreibman, Gradmann, Hennieke, Blanke, Chambers, Dunning, Gray, Lauer, Pichler, & Renn, 2013, p. 5), are devising versioning strategies with little guidance from data models such as the Open Annotation framework (Eckert, 2012; Open Annotation Data Model, 2013).

Consequences

In short, digital objects do not repose. Yet the terminologies, the conceptual frameworks, and the functionalities of repositories assume that they do. If we are to move away from having to freeze digital texts into simulations of dead trees through the PDF format, we have to come to terms with the challenges of versioning dynamic digital resources. For much of the actual work we do as scholars, it makes more sense to talk about collaboratories than repositories, and to move from a focus on fixed documents to an understanding that many digital objects reflect ongoing collaboration and labour, and as a consequence are subject to modification, remediation, revision, and the like within a digital ecology where textuality is increasingly granular, distributed, and interdependent as well. There is a gap between what digital texts are and how they should be handled within the context of ongoing scholarly production as opposed to in an archival context. The lack of adequate infrastructure for versioning online material has serious implications that extend well beyond the domain of scholarship into law, policy, history, and above all culture, which is constantly in flux thanks to creativity. The risk is the loss of the inestimable knowledge we glean from tracking changes.

Moreover, knowledge work is not always in the form of writing. Digital work can be tied into things like ontologies or tools. There is currently no standard procedure for versioning ontologies on the Semantic Web, although the need for this has been identified (Klein & Fensel, 2001; Kotowski & Stacey, 2012). Repositories are essential, but we need these and other infrastructures to be built with the capacity for managing and tracking change built into them, and we need representational systems capable of conveying the complex versioning of a singular or composite digital object.

The main point here, beyond the basic observation that the “lasting change” of digital culture and scholarship poses some far-reaching challenges (Bretz, Brown, & McGregor, 2010), is two-pronged. First of all, both infrastructures for scholarly work and longer-term repositories have to be able to accommodate the extent to which primary and secondary materials are increasingly in flux, which has ramifications for systems architecture and for the kinds of tools we will need to use as scholars. There is a general awareness of this in the digital humanities community, but we need to begin to grapple with it more purposefully. We should learn what we can from archive-oriented initiatives such as ResourceSync (Klein, Sanderson, Van de Sompel, Warner, Haslhofer, Lagoze, & Nelson, 2013), while recognizing that ongoing online cultural and scholarly production poses some different challenges. Secondly, we should not let the significant challenges associated with the back ends of our systems obscure the extent to which we lack mechanisms for dealing with versioning within interfaces either: these are equally important, since they will crucially impact our understanding of artifacts’ relations to their earlier instantiations.

The need to address, at the level of information architecture, tools, and interface, the flux in both primary and secondary digital sources in the humanities is analogous to the kinds of shifts that are already underway in our thinking collectively about cyber-infrastructure more generally. Whereas once heavy iron batch processing was the only model for high-end computing, the research community now recognizes that certain kinds of research, much of it in the humanities, demand infrastructure that is more

interactive and dynamic. We need to understand that the repository models that have, for good reasons, dominated the ways we manage large humanities data sets need to be modified to recognize the dynamic, living nature of digital cultural artifacts and scholarship on culture. Doug Reside's (2011) warning, with respect to born-digital materials, of "the very real possibility that a large portion of our cultural history will be lost unless we solve it quickly," applies equally to cultural scholarship. If the academic community allows the gap between increasingly dynamic and interoperable textuality, on the one hand, and the tools and environments with which we manage scholarly work, on the other, to persist, it will widen and become a gulf into which much early digital scholarship will fall, and much that remains will be deficient because it is out of sync with other components of the dynamic digital system. The culture of scholarship is also proving dynamic and susceptible to change, so there is hope that the community can meet the challenge of closing this gap, and thereby make a major contribution to the sustainability both of cultural scholarship and a portion of the cultural record.

Acknowledgments

Thanks to our research colleagues, collaborators, and partners in the INKE (inke.ca), Orlando (www.ualberta.ca/orlando), and Canadian Writing Research Collaboratory (cwrc.ca) projects. Many thanks to Abigel Lemak for research assistance and for help revising and preparing the manuscript for publication.

Notes

1. As Lawrence Lessig notes, consumption in a Web or e-reading context is fundamentally copying (*ReMix* 98-99).
2. These are the INKE Project (<http://inke.ca/>), Orlando Project (<http://orlando.cambridge.org/>), and the Canadian Writing Research Collaboratory (<http://www.cwrc.ca/en/>).
3. Bertrand Russell (1951) provides a succinct and balanced summary of the earlier technologies listed here in *The Impact of Science on Society*.
4. Cooper and Bradley worked closely together, co-authoring works as "Michael Field" to an extent that anticipates the intensity of digital collaboration described here. More than just a pseudonym, Michael Field was intended to represent an artistic collaboration around the fusion of two people into a single whole. So successful was their attempt that it led to friends referring to the pair as "both of him," and even with modern digital tools like the Semantic Web, it can look like it was Field who was real and Cooper and Bradley only pseudonyms (Brown & Simpson, 2013).

References

- Allen, M. (2013). What was Web 2.0? Versions as the dominant mode of Internet history. *New Media & Society*, 15(2): 260-275. URL: <http://nms.sagepub.com/content/15/2/260.full.pdf+html> [January 12, 2014].
- Andrews, T., & van Zundert, J. (2013). An interactive interface for text variant graph models. *DH2013 Abstracts*. URL: <http://dh2013.unl.edu/abstracts/ab-379.html> [January 24, 2014].

- Arazy, O., Stroulia, E., Ruecker, S., Arias, C., Fiorentino, C., Ganev, V., & Yau, T. (2010). Recognizing contributions in wikis: Authorship categories, algorithms, and visualizations. *Journal of the American Society for Information Science and Technology*, 61(6), 1166-1179. URL: <http://onlinelibrary.wiley.com/doi/10.1002/asi.21326/abstract> [January 24, 2014].
- Barry, M. (2011). *Nuts and bolts*. Max Barry. URL: <http://maxbarry.com/2011/10/05/news.html> [January 20, 2014].
- Bowie, David. (1971). Changes. *Hunky Dory*. RCA Records.
- Bretz, A., Brown, S., & McGregor, H. (2010). *Lasting change: Sustaining digital scholarship and culture in Canada: Report of the Sustaining Digital Scholarship for Sustainable Culture Group*, (n.p.). URL: <http://www.cwrc.ca/wp-content/uploads/2011/05/Lasting-Change-Knowledge-Synthesis.pdf> [January 20, 2014].
- Brown, S., Adelaar, N., Ruecker, S., Sinclair, S., Knechtel, R., & Windsor, J. (2013). Text encoding, the index, and the dynamic table of contexts. *Digital Humanities Conference Abstracts 2013*. University of Nebraska, Lincoln. URL: <http://dh2013.unl.edu/abstracts/ab-231.html> [January 24, 2014].
- Brown, S., & Simpson, J. (2013). *The curious identity of Michael Field and its implications for humanities research with the semantic web*. *IEEE Conference on Big Data 2013*: 77-85. Web. DOI: 10.1109/BigData.2013.6691674.
- Causser, T., Tonra, J., & Wallace, V. (2012). Transcription maximized; expense minimized? Crowdsourcing and editing *The Collected Works of Jeremy Bentham*. *Literary and Linguistic Computing*, 27(2), 119-137. URL: <http://llc.oxfordjournals.org/content/early/2012/03/28/llc.fqs004.short?rss=1> [January 24, 2014].
- CollateX. (2010). *The Interedition Development Group*. URL: <http://collatex.net/> [January 24, 2014].
- Dergacheva, E., Brown, S., Roeder, G.G., Dobson, T., Peña, E., & Knechtel, R. (2013, September). *Prospects and pitfalls of workflow management in born-digital projects*. Panel presentation at conference of the Japanese Association for Digital Humanities, Ritsumeikan University, Kyoto.
- Eckert, K. (2012). *A Linked Data based infrastructure for DM2E (WP2)*. URL: <http://www.slideshare.net/DM2E/kai-eckert> [January 20, 2014].
- Engelbart, D. (2008). *Doug's 1968 demo*. Doug Engelbart Institute. URL: <http://www.doungengelbart.org/firsts/dougs-1968-demo.html> [January 20, 2014].
- Fedora Content Versioning. (2005). *Fedora Commons*. Fedora Project. URL: <http://www.fedora-commons.org/documentation/3.ob1/userdocs/server/features/versioning.html> [January 20, 2014].
- Frizzera, L., Radzikowska, M., Roeder, G., Peña, E., Dobson, T., Ruecker, S., Rockwell, G., Brown, S., & the INKE Research Group. (2013). A visual workflow interface for the editorial process. *Literary and Linguistic Computing*, 28(4), 615-628. URL: <http://llc.oxfordjournals.org/content/28/4/615.abstract?sid=14ad3661-3eef-49d7-8dbf-d3e730304bbc> [January 24, 2014].
- Fry, Ben. (2009). *On the origin of species: The preservation of favoured traces*. URL: <http://benfry.com/traces/> [January 24, 2014].
- <graphic>. (2014). *Text encoding initiative guidelines*. Text Encoding Initiatives. URL: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-graphic.html> [January 20, 2014].
- Greetham, D.C. (1992). *Textual scholarship: An introduction*. New York, NY: Garland.
- Huculak, J.M., & Richardson, A. (2013). *White paper: A survey of current collation tools for the Modernist Versions Project*. URL: <http://web.uvic.ca/~mvp1922/wp-content/uploads/2013/10/WhitepaperFINAL.pdf> [January 20, 2014].
- Juxta: Collation Software for Scholars. (n.d.). *Juxta*. URL: <http://www.juxtasoftware.org/> [January 24, 2014].
- Katz, A. (n.d.). *Ariel Katz on intellectual property, competition, innovation, and other issues*. URL: <http://arielkatz.org/> [January 20, 2014].

- Kirschenbaum, M.G., Farr, E., Kraus, K.M., Nelson, N.L., Stollar Peters, C., Redwine, G., & Reside, D. (2008). *Mechanisms: New media and the forensic imagination*. Cambridge, MA: MIT Press.
- Kirschenbaum, M.G., Farr, E., Kraus, K.M., Nelson, N.L., Stollar Peters, C., Redwine, G., & Reside, D. (2009). *Approaches to managing and collecting born-digital literary materials for scholarly use*. White paper to the National Endowment of the Humanities Office of Digital Humanities. URL: http://mith.umd.edu/wp-content/uploads/2012/03/whitepaper_HD-50346.Kirschenbaum.WP.pdf [January 20, 2014].
- Klein, M., Sanderson, R., Van de Sompel, H., Warner, S., Haslhofer, B., Lagoze, C., & Nelson, M.L. (2013). A technical framework for resource synchronization. *D-Lib Magazine*, 19(1/2), 3. URL: <http://www.dlib.org/dlib/january13/klein/01klein.html> [January 24, 2014].
- Klein, M.C.A., & Fensel, D. (2001). Ontology versioning on the semantic web. *Semantic Web and Web Services, SWWS*, 75-91. URL: <http://secs.ceas.uc.edu/~mazlack/ECE.716.Sp2011/Semantic.Web.Ontology.Papers/kleino1ontology.pdf> [January 20, 2014].
- Kotowski, D., & Stacey, D.A. (2012). Ontology library—A new approach for storing, searching and discovering ontologies. *KEOD 2012*, 271-277. DOI: 10.5220/0004145702710277 [July 25, 2014].
- Lessig, L. (2008). *Remix: Marking art and commerce thrive in the hybrid economy*. London: Bloomsbury. URL: <https://archive.org/details/LawrenceLessigRemix> [December 4, 2014].
- Lowry, P.B., Curtis, A., & Lowry, M.R. (2004). Building a taxonomy and nomenclature of collaborative writing to improve interdisciplinary research and practice. *Journal of Business Communication*, 41(1), 66-99. URL: <http://ows.edb.utexas.edu/sites/default/files/users/jl3525/Taxonomy%20of%20Collaborative%20Writing.pdf> [January 20, 2014].
- McGann, J. with L. Samuels] (2001). Deformance and Interpretation. In *Radiant textuality: Literature after the World Wide Web*. New York, NY: Palgrave Macmillan.
- McMillan, R. (2012, February 21). Lord of the files: How GitHub tamed free software (and more). *Wired*. URL: <http://www.wired.com/2012/02/github-2/all/> [November 25, 2014].
- McMillan, R. (2013, September 2). From collaborative coding to wedding invitations: GitHub is going gainstream. *Wired*. URL: <http://www.wired.com/wiredenterprise/2013/09/github-for-anything/> [January 20, 2014].
- Nelson, T.H. (1999). Xanalogical structure, needed now more than ever: Parallel documents, deep links to content, deep versioning, and deep re-use. *ACM Computing Surveys*, 31(4es). DOI: 10.1145/345966.346033 [January 20, 2014].
- Open Annotation Data Model. (2013, 8 February). W3C. URL: <http://www.openannotation.org/spec/core/20130208/index.html> [January 20, 2014].
- Pierazzo, E. (2009). Digital genetic editions: The encoding of time in manuscript transcription. In M. Deegan & K. Sutherland (Eds.), *Text editing, print and the digital world* (pp. 169-186). Farnham, UK: Ashgate.
- Project Xanadu*. (2012, October 4). Internet Archive Wayback Machine. URL: <http://web.archive.org/web/20121004184443/http://xanadu.com/> [January 20, 2014].
- Ramsay, S. (2011). *Reading machines: Toward an algorithmic criticism*. Champaign, IL: University of Illinois Press.
- Reside, D. (2011, 22 April). *No day but today: A look at Jonathan Larson's word files*. URL: <http://www.nypl.org/blog/2011/04/22/no-day-today-look-jonathan-larsons-word-files> [January 20, 2014].
- Russell, B. (1951). *The impact of science on society*. New York, NY: Columbia University Press.
- Sanderson, R., & Albritton, B. (2013). *Shared Canvas data model*. SharedCanvas. Open Annotation Collaboration. URL: <http://www.shared-canvas.org/datamodel/spec/> [January 20, 2014].
- Sayers, J. (2014). *GitHub*. GitHub Inc. URL: <https://github.com/jentery> [January 20, 2014].
- Schmidt, D. (2013). Collation on the Web. *DH2013 Abstracts*. URL: <http://dh2013.unl.edu/abstracts/ab-108.html> [January 24, 2014].

- Schreibman, S. (2011). *Versioning machine 4.0: A tool for displaying and comparing different versions of literary text*. (Originally published in 2002 & 2010) Digital Collections & Research Project. URL: <http://v-machine.org/> [January 20, 2014].
- Schreibman, S., Gradmann, S., Hennicke, S., Blanke, T., Chambers, S., Dunning, A., Gray, J., Lauer, G., Pichler, A., & Renn, J. (2013). Beyond Infrastructure—Modelling scholarly research and collaboration. *Digital Humanities 2012*. URL: <http://hal.inria.fr/docs/00/80/14/39/PDF/DH2013DM2EProposal.pdf> [January 24, 2014].
- Siemens, R., Timney, M., Leitch, C., Koolen, C., & Garnett, A., with the ETCL, INKE, & PKP Research Groups. (2012). Toward modeling the *social* edition: An approach to understanding the electronic scholarly edition in the context of new and emerging social media. *Literary and Linguistic Computing*, 27(4), 445–461. URL: <http://llc.oxfordjournals.org/content/27/4/445.abstract?sid=22348d03-47cc-4a78-961a-1467c301685b> [January 24, 2014].
- Simpson, J., Brown, S., Quamen, H., Bath, J., Saklofske, J., Sayers, J., Goddard, L., Barclay, A., Christie, A., Elliott, M., & the INKE Team. (2012, December). *E/Merging models for the production of online research through linked data*. Paper presented at Implementing New Knowledge Environments (INKE) conference, Havana, Cuba.
- Sonvilla-Weiss, S. (2010). Introduction: Mashups, remix practices and the recombination of existing digital content. In *Mashup cultures* (pp. 8-23). New York, NY: Springer Wlen.
- Soules, C.A.N., Goodson, G.R., Strunk, J.D., & Ganger, G.R. (2003). *Metadata efficiency in versioning file systems*. *FAST 03 Proceedings of the USENIX Conference on File and Storage Technologies*: 43-58. URL http://pdf.aminer.org/000/202/444/metadata_efficiency_in_versioning_file_systems.pdf [December 4, 2014].
- Speck, B.W., Johnson, T.R., Dice, C.P., & Heaton, L.B. (1999). In B.W. Speck, T.R. Johnson, C.P. Dice, & L.B. Heaton (Eds.), *Collaborative writing: An annotated bibliography*. Westport, CT: Greenwood Press. URL: http://books.google.ca/books/about/Collaborative_Writing.html?id=IWRH-cKoHwYC&redir_esc=y [January 24, 2014].
- Sullivan, H. (2013). *The work of revision*. Cambridge, MA: Harvard University Press.
- Terras, Melissa. (2011). Present, not voting: Digital humanities in the Panopticon: Closing plenary speech, Digital Humanities 2010. *Literary and Linguistic Computing*, 26(3), 257–269. URL: <http://llc.oxfordjournals.org/content/26/3/257.abstract?sid=b30bc8fa-b15a-4889-8903-4ea92e3310f2> [January 24, 2014].
- Wittern, C. (2002). *WWW database of Chinese Buddhist texts*. WWW Database of Chinese Buddhist Texts. URL: <http://www.kanji.zinbun.kyoto-u.ac.jp/~wittern/can/can4/ind/canwww.htm> [January 20, 2014].
- Wittern, C. (2013). Beyond TEI: Returning the text to the reader. *Journal of the Text Encoding Initiative*, 4, 1–14. URL: <http://jtei.revues.org/691> [January 24, 2014].