Abstract

In recent decades, university research has become bound up in the expectation of innovation, often at the expense of sustainability. As a guiding principle, innovation has shaped the goals of universities and funders; the recent turn to sustainability, signaled by the rise of data management plans and data deposit requirements, may indicate a welcome discursive shift. This article explores the interplay of knowledge creation and mentorship in the development of digital humanities research projects, with the aim of articulating how to resolve the artificial tension between innovation and sustainability. While often framed in opposition to each other, or used in a neoliberal framework to force researchers to do more with fewer resources, innovation and sustainability can, this article argues, be dual goals in the development of lasting digital humanities scholarship, with a focus on XML serializations' role in this scholarly ecosystem.

Scope, Scale, and Sustainability at the turn of the Twenty-First Century

Throughout the 1990s, most North American digital humanities work fell into one of two camps: edition and online resource creation or software development. Many of us started as junior faculty or graduate students in the former camp even if we now work in the latter. There were few graduate or undergraduate courses in the digital humanities in Canada at the time. Most of the graduate training that shaped the scholarship in which we now engage was offered through workshops or research assistantships. In addition to the challenge of accessing training, the generation of graduate students from the first decade of the 2000s in Canada were working when digital humanities methods were occasionally hampered by disciplinary mistrust. In North America, the digital humanities, when they garnered attention, were dogged by concerns that they did not constitute “real” humanities scholarship [Kirschenbaum 2012].[1]

The digital humanities’ outsider status notwithstanding, in the period from the mid 1990s to 2010, large-scale digital humanities publishing projects flourished online. Digital humanities scholars reveled in the affordances and pleasures of online publishing. Hypertext's ability to let readers move through texts in non-linear ways attracted a good deal of...
TEI, the XML-based language of the Text Encoding Initiative Consortium for modeling documents and formalizing text in a computationally tractable way, has been a major part of the digital humanities’ efforts to engage in sustainable long-running publication and textual analysis. The TEI was conceived by a multidisciplinary group of scholars and students at a Vassar-hosted meeting in 1987; over 50 scholars worked on its initial release in 1990 [History – TEI: Text Encoding Initiative]. The development of TEI-SGML (now TEI-XML) for representing textual material was an area of major intellectual effort and innovation that provided a useful tool for online edition production, and responded directly to the needs of editors and readers in a way that print publication could not. The proliferation of TEI impacted the growth of the digital humanities at the turn of the twenty-first century: many junior scholars were introduced to TEI through workshops and research assistantships. Not only is the TEI designed to meet the needs of textual editors, it also offers a number of advantages to users who have no prior technical background. TEI is human-readable in a way that other formats, for example JSON or Turtle, are not. TEI is platform-agnostic, which means that anyone can start writing valid TEI without the expensive overhead and resource-intensive setup of other tools, such as databases, Hadoop clusters, Adobe Creative Suite software, or 3D printers. Finally, TEI is not a language or tool borrowed from another discipline to be put to a humanities’ purpose. Instead, it is a community-led humanities-specific language. The language continues to grow in response to TEI-sponsored Special Interest Groups and in response to requests from regular users who want the language to expand to include new use cases. Volunteers govern the TEI Consortium through a Board of Directors and a Technical Council who undertake the demanding governance and stewardship of the TEI Guidelines, services, and documentation. The TEI Consortium provides tools to analyze and customize TEI such as the eXtensible Stylesheet Transformation Language (XSLT) scripts, the XML-ODD format for TEI customization, and tools for users to display TEI—including the TEI Archiving, Publishing, and Access Service (TAPAS) and CETElcian. For many humanities

scholarly attention [McGann 2001] [Hayles 2007]. More important, perhaps, for questions of sustainability and innovation, was the sheer scope of the projects that scholars undertook to recuperate and share material that traditional publishers could not provide in print. Many flagship projects from this era had a huge remit. For example, Orlando, started in 1995, provides interpretive and biographical information about “women’s writing in the British Isles from the beginnings to the present”; the Women Writers’ Project (WWP), started in 1988, is dedicated to early women’s writing in English; and Voices from the Gaps, started in 1996, aims to “share the works of marginalized artists, predominantly women writers of color living and working in North America” (see [Brown et al. 1997], [Hockey 2004], [Flanders 2006], [Earhart 2012], and [McNaron & Miller]). These projects are distinct from those with the more narrow scope that characterize current digital humanities projects. The directors of smaller projects with short development cycles can more easily frame a project as novel or innovative. Larger, longer-running projects, which are in greater need of sustaining maintenance, must develop new micro-projects to frame their work as innovative. These projects are often central to humanities research and teaching, and their long-term stability is key to maintaining access to the cultural record (for example, the disappearance of information about the cultural conditions of generations of women’s writing in the British Isles would be a significant loss, and yet there are few funding mechanisms to support the maintenance of such a corpus). As Jessica Otis has pointed out in her discussion of how the Roy Rosenzweig Center for History and New Media decides which projects to continue to support, principal investigators from the turn of the century often promised funders not only a broad project scope, but also that digitized material would stay up online indefinitely. The move from early hypertext-only sites to database-backed sites and content management systems that now need to be migrated forward or need virtual environments to run has made keeping the sites functional sometimes difficult and often impossible (see Holmes & Takeda and Otis in this issue).

In the age of large-scale online publishing projects at the turn of the twenty-first century, scholars took the mechanisms of publishing into their own hands. Many “scholars invested in early work on race [and class and gender] in digital humanities insisted on building editions and digital texts as an activist intervention in the closed canon” [Earhart 2012, 317]. This move was not always an indictment of traditional publishers: many scholars recognized that, for example, publishing 400 texts written by women between 1526 and 1850, as the WWP has, represented a greater financial burden than most print-based publishers could undertake or hope to recover through sales. Furthermore, the affordances of hypertext and of databases to reorganize and connect texts nearly endlessly offered advantages that printed texts could not supply. The development of easily archivable mark-up languages, such as TEI, has been key to the development of corpora that are easy to sustain.
scholars, TEI has been the introduction to digital humanities scholarship and learning to use these scripts and tools has been key to analysis and online publication.

While there have been advances in artificial intelligence and application design that support some branches of digital humanities scholarship, the TEI continues to be the gold standard for edition encoding. Confronted with a plain text transcription of a primary source, a computer will know nothing about either text structure (for example, where paragraphs start and end, or where the boundaries of paragraphs are) or about more nuanced content (for example, who added the interlinear glosses to a manuscript, how many years separate the original scribe and the glossator, how abbreviations ought to be expanded, or which calendar system is used in the text). Broadly, encoders use TEI to add what they know or wish to argue about a transcribed text in a computationally parsable form. The TEI was originally designed to help scholars describe what they know about a source text; the language has since expanded to let scholars describe images and any number of other cultural artifacts and ancillary information in a robust and systematic way. Best of all, TEI creation does not rely on databases, content management systems, or any other technology that requires perennial migration.

A Personal Account of Mentorship and Project Development in the Canadian Context

A part of the TEI-user community consists of scholars who are engaged in largely DIY encoding and publishing practices. In the Victorian Studies community in which I trained, this ethos grew out of the inaugural Networked Infrastructure for Nineteenth-Century Electronic Scholarship (NINES) workshop in July 2005, which catalyzed the use of TEI to facilitate the aggregated search of a number of now-iconic digital humanities projects, including the Nineteenth-Century Serials Edition[2] [Brake, Armstrong et al. n.d.], The Poetess Archive [Mandell n.d.], and The Vault at Pfaff's [Whitley n.d.]. The initial NINES projects each ran on its own project-specific stack: at the time there were no content management systems or prefabricated hosting services for TEI. That said, NINES was a community-led standards success: projects federated within NINES used TEI as a common encoding format, shared NINES keywords, and featured stable project URIs. These shared standards allowed for listing and linking within NINES, leaving the question of interface and backend development to individual projects. My formative training was on one of the projects initiated at the inaugural NINES summer workshop, the Yellow Nineties Online (1890s.ca). Relaunched as Yellow Nineties 2.0 in 2015, the Yellow Nineties and its federated sister projects feature eight fin-de-siècle little magazines, with supporting biographies and essays that focus on the context and networks of the little magazines’ production and reception; teaching resources and student scholarship; and tools for visual culture analysis [J most recent digital humanities

Smaller digital humanities teams rely on project members’ understandings of their shared project as a whole. This holistic knowledge is a boon for sustainability. As Frank Tough has argued, humanities research is a craft, and, at the graduate level is often taught through intensive, one-on-one research assistantships [Tough 2021]. In Canada in the 2000s, this model continued in the digital humanities, with many graduate students learning from principal investigators on small teams. The Canadian funding system, particularly the support provided by the mid-sized SSHRC Standard Research Grant (1998-2011, $21,000-$250,000 over three years), and the uptake of digital humanities predominantly at smaller universities, added structural elements that encouraged the creation of small teams, as opposed, for example, to the broader, multidisciplinary, multi-university projects with dedicated staff of more recent years. Research assistants on these small projects got to learn how projects were built, to understand what each member of a small and intimate team was doing, to contribute to the infrastructural and content development of each project, and to see how the technical infrastructure and subject matter of each project were knit together in the intellectual labor of each project [Engel & Thain 2015] [Anderson et al. 2016]. Both Matthew Kirschenbaum and Julia Flanders suggest that this refusal to separate out infrastructure and content development from the intellectual effort was a distinctive feature of digital humanities in the United States at the time (this approach is certainly central to the development of TEI as a language), as indeed it was in Canada. While the parallel development of infrastructure and content may be framed as the result of happenstance, this feature is consonant with humanities approaches to research as well as to the craft-aligned mentorship of the next generation of scholars. As Lorraine Janzen Kooistra—one of the two original principal investigators of the Yellow Nineties—has pointed out when comparing the project to more recent digital humanities
projects: “we did not know that we were supposed to have a technical manager or even collaborators with a background in computer science” [Janzen Kooistra 2021]. The project was instead developed by the principal investigators and students with extensive consultation with librarians and with hosting support (initially on a desktop computer running as a server) from a single Faculty of Arts systems administrator.[3]

In this craft system of humanities research training, graduate students tend to go on to reproduce the methods of their supervisors and to work in ways that build on the conditions of their training. This has certainly been the case in my career: having learned the technical aspects of digital humanities work from humanities scholars, often on small teams, I came away with the sense that each project should not have any elements that the principal investigator did not understand herself, and whose creation or deployment she could not undertake at a high level. In short, no one working on the project would do anything the principal investigator could not have done herself, given sufficient time. This approach is a useful safeguard against any changes in university affiliation or changes in local IT policy that might require moving a project, although it is not necessarily a safeguard against the demands of upgrades. The approach also makes it easier to communicate with funders, since the principal investigator is conversant with all parts of the project. Finally, and perhaps most importantly, it is enormously intellectually fulfilling [Roued-Cunliffe 2016]. Mentorship on digital projects that fit this model is indeed like traditional humanities mentorship, in which research assistants join principal investigators in the archives, on field research trips, in literature review preparation, in discussion of argumentation, and in manuscript revision. If the assistantship lasts long enough, assistants can get to know every part of the project, and can graduate with the knowledge they need to develop their own projects.

Balancing Sustainability and Innovation in Digital Humanities Projects

A single principal investigator who understands all parts of a project reduces project dependencies and so can safeguard a project. However, this is not the only, or even necessarily best, model for digital humanities project development. The model has two central drawbacks that may work against sustainability: fragility and non-scalability. If the documentation is poor and there is little buy-in from others, it is hard to maintain the project if the principal investigator has to leave the team for any reason. The project will also be fragile if the principal investigator does not update the project’s technology stack, especially once the technology reaches an end-of-life stage and options to ensure backwards compatibility become cumbersome (for more on this topic see Jessica Otis’ article in this issue of DHQ; for a more general discussion see [Barats et al. 2020]).

The second challenge of the model of a sole principal investigator who could perform every project task is that of scalability. I do not mean to use the term scalable as a neoliberal dog whistle meaning to do more with less, or meaning to get students to pay for what were once government- or endowment-supported low student-to-researcher ratios enjoyed by previous generations. Instead, I would like to point to the benefits of larger, more collaborative research projects. They are often less fragile, with more opportunities for leadership and therefore more people who could steward the project if the principal investigator has to step back. They are also more scalable in terms of knowledge creation, in that they will have multiple research questions, spanning several domains, with distinct areas of investigation of interest to, for example, collaborators in computer science, information science, the social sciences, and the humanities.

No matter what the leadership or collaboration structure of a research project may be, the central goal of primary research is the creation of new knowledge. This goal is the hallmark of university-led research, and yet it may be at odds with the discourses of innovation and sustainability in the contemporary research funding landscape. In exploring these tensions in project development, I will use the Lesbian and Gay Liberation in Canada (LGLC) project, which I co-direct with Michelle Schwartz (Toronto Metropolitan University), as my case study.[4]

The goal of the LGLC project is to analyze the emergence and expansion of gay liberation as an intellectual movement backed by informal and formal political action, in order to better understand the conditions that foster political change and to recover gay liberation history for a popular audience. The project has been built around two books by our collaborator Donald McLeod, Lesbian and Gay Liberation in Canada volumes 1 and 2 [McLeod 1996] [McLeod 2017]. McLeod’s chronologies consist of 3,100 chronological events spanning 1964 to 1981. Using TEI, the LGLC team has
encoded each event, marking up names, places, dates, organizations, and periodicals. The core events are contextualized by, to date, a further 32,000 entity records about people, location, publications, and organizations. The team is now involved in further archival research to uncover more events spanning 1960 to 1985. The TEI serves as a discovery mechanism for our analysis, letting us explore the relationships between the entities, and as the basis for our public-facing web app at lglc.ca. The lglc.ca web app consists of the standard set of tools, or stack, for web publication: a server, a database, a programming language to retrieve material from the database, and front-end scripts to create HTML for display in a browser. lglc.ca is hosted and supported by Toronto Metropolitan University Library's Collaboratory. It consists of a neo4j database, a node.js app for retrieval, and jade.js/pug.js-templates to create the HTML front end. It offers visitors to the site a graphic user interface that enables many of the exploratory and analytical features that otherwise would be available only to users with the skills to process the TEI that feeds the neo4j database.

I must say that, to this day, working with TEI and its user community is one of my greatest professional pleasures.\[5\] Moreover, for the purposes of this special issue, I must stress the safety and archivability the TEI lends to my projects, including the LGLC. From the first, TEI has been central to the LGLC preservation strategy. Unlike our database, TEI is human readable without specialized software or an integrated development environment (IDE). We are certain that our validated data can be preserved as a series of flat files, microfiche, or even as a printout that will be legible to future generations [Holmes 2019]. The code itself does not need more than text editor software to view, and the ubiquity of XML leaves us assured that if our TEI is not decipherable in the future it will be because of a more significant failure of digital systems well beyond the scope and control of academia, commercial companies, or even governments.

But what of our web app? Databases and their interfaces are, as this special issue underscores, notoriously difficult to maintain after the end of a project's active development. We have, therefore, never expected that the lglc.ca web application would exist forever. However, having been trained on the Yellow Nineties, which is part of the generation of digital humanities projects that are the enduring focus of a principal investigator or a pair of investigators for decades, we too expect to work on the LGLC project for many years to come. Unlike many more recent digital humanities projects, like the original collaborative Torn Apart map project, the Serendip-o-matic web app, or other shorter-term projects, our plans for decades of active development may, on the one hand, be attractive to funders, but on the other hand, have kept us from devoting resources to infrastructural changes that we would need for the graceful end to the project beyond periodic or final deposits of our TEI in national and institutional repositories.

Our attention to innovation in the development of lglc.ca at the expense of sustainability has been shaped by one of the best parts of the university research culture: the commitment to creating new knowledge. We are not alone in this approach. While replication and verification are central parts of the research ecosystem, most research is meant to break new ground and benefit the public through sharing the knowledge they supported us in creating. That said, in the last twenty years in the Canadian funding context, that drive to create new knowledge was coupled with a focus on innovation. Cultural and governmental privileging of innovation is, for example, hardwired into the main source of institutional research infrastructure funding in Canada, the Canadian Foundation for Innovation (CFI). And, in as much as it dovetails with the pleasure of new knowledge creation, innovation is no bad thing. A key drawback, however, to the assumption that innovation constitutes a good end in itself, perhaps at the expense of sustainability, is that innovation and sustainability have been discursively opposed to one another [Kirschenbaum 2009, ¶3 \[3\] [Russell & Vinsel 2016]. Principal investigators' perceptions that they may be turned down for research funding if they focus on using tried and true methods or on maintaining projects, rather than innovating, materially entrenches this discursive opposition.

There is, however, a sea change under way. In Canada, our public funders are increasingly requiring the open-access publication of results, the deposit of data where appropriate, and the creation of sustainability plans. Happily, on the LGLC project, some of our innovations have led us back to sustainable research development plans: one of our initial innovations in 2013 was to use a then-new graph database, neo4j, as the backend for our public web app. This choice involved thinking through how best to convert our TEI, which relies on a hierarchical tree structure, for representation in neo4j's non-hierarchical node-and-edge-filled graph structure. The exercise of thinking through how to model our data as both a tree and a graph led my research lab into work on how to represent TEI as linked data, another format that relies on graph structures. Linked data has, unlike a neo4j instantiation of the data, the option of remaining in flat files
for knowledge creation and long-term preservation.[6]

Sustainability and innovation: it may be possible to have it both ways. Indeed, Project Endings methods represent a marriage of sustainability and innovation. One of the project’s principles brings together these two concepts: “no boutique or fashionable technologies: use only standards with support across all platforms, whose long-term viability is assured,” [The Endings Project Team (n.d.)] [Carlin 2018]. The recommendation resolves the tension between sustainability and innovation: as languages, HTML, CSS, and JavaScript are sustainable and secure, but the Project Endings recommendation that researchers use them from the first, rather than producing database-based backed sites that are not sustainable in the long term, is a methodological innovation. While flat files that researchers deposit in repositories will be preserved, and in the case of TEI will remain human readable, these deposited files are less readily accessible to the members of the public who fund our research than a searchable web-version of the same material. The creation of project sites that are accessible (by virtue of being on the web) and that will remain accessible (by virtue of being lightweight, secure, and easy to host) is truly innovative. As someone who trained in the creation of digital humanities projects that are intended to run for decades under a model of development using tools that any principal investigator could understand, I welcome the Project Endings’ recommendation. As someone who is keen to see digital humanities projects of any duration persist in a format that is of maximum utility to the public, I am doubly pleased. The challenge will be to continue to use Project Endings principles to keep digital scholarship accessible, even after the Project Endings recommendations move from being innovative to being the gold standard.

Notes

[1] This dismissal of the non-traditional scholarly output of digital humanities projects is distinct from the later dismissal of digital humanities scholars as cool kids who have undue access to funds. For more on these debates see Matthew Kirschenbaum’s “What Is ‘Digital Humanities,’ and Why Are They Saying Such Terrible Things about It?” and “Am I a Digital Humanist? Confessions of a Neoliberal Tool” [Kirschenbaum 2014] [Kirschenbaum 2016].

[2] Since I trained on the Yellow Nineties, which initially had two principal investigators, I cannot be surprised, on reflection, that the Lesbian and Gay Liberation in Canada project, which I co-direct with my research partner, Michelle Schwartz, also has two principal investigators, owes much of its structures and ways of working to the Yellow Nineties Online. I will refer to sole principal investigatorship below even though joint investigatorship is relatively common in Canadian digital humanities (see, for example, the investigatorship of Patricia Clements and Isobel Grundy on Orlando, Geoffrey Rockwell and Stéfan Sinclair on Voyant, and Ian Milligan, Nick Ruest, Jeremy Lin, and Jefferson Bailey on Archives Unleashed).

[3] European digital humanities scholars and others have made great strides in the creation of linked data that focuses on cultural material central to the humanities. Some notable projects, such as Social Networks and Archival Context Project (SNAC), originate in the United States. In Canada, where there is real new growth in cultural linked data creation in the digital humanities, libraries, and government. For more information see work by the Canadian Heritage Information Network, Linked Infrastructure for Networked Cultural Scholarship, and the Single Interface for Music Score Searching and Analysis.

[4] LGLC has been supported by two Social Science and Humanities Research Council of Canada Insight Grants (2014–20, 2021-25) and by the Ministry of Canadian Heritage’s Canada History Fund (2021-2023). Grant partners include Susan Brown director of the Canadian Writing Research Collaboratory (CWRC, University of Guelph), Don McLeod (University of Toronto Libraries), El Chenier director of the Archives of Lesbian Oral Testimony (ALOT, Simon Fraser University), Fabien Galipeau (Archives gaies du Québec), and Fangmin Wang and M.J. Suhonos (Toronto Metropolitan University Library & Archives). The project has been built with the support of 16 paid research assistants since 2014. Their scholarship has included archival research, encoding, UX, data conversion, frontend design, and implementation. We are also grateful for the support of Toronto Metropolitan’s Centre for Digital Humanities, the Toronto Metropolitan Library Collaboratory, the University of Ottawa’s Labo de données en sciences humaines/Humanities Data Lab, the Canadian Foundation for Innovation, and Compute Canada (for more please see https://lglc.ca/about).

[5] Like many digital humanities scholars of my generation, I learned to write and process TEI in graduate school, but was enrolled in my PhD before most Canadian universities had a digital humanities course curriculum. My formal education in TEI creation and project development came from workshops offered through the Women Writers Project (now at Northeastern University), the Digital Humanities Summer Institute (DHSI, University of Victoria), and the Initiative for Digital Humanities, Media, and Culture (Texas A&M). The Yellow Nineties Online and the Toronto Metropolitan University Centre for Digital Humanities funded my travel and registration fees for many of these workshops. I now teach
TEI at the DHSI and in undergraduate classes. Not only is it a useful language for representing textual and visual sources, it helps me teach undergraduates how to think like editors and how to knit together primary and secondary sources in their analysis of text. While teaching others how to create and process XML in various formats likely appeals to the misguided neoliberal “hard skills” approach to higher education, it is in practice a key tool to empower students how to think about editing, research, analysis, and about how they can literally shape the systems that underpin our experience of technology.


Works Cited


McNaron & Miller McNaron, T. and Miller, C. (n.d.) Voices from the Gaps. Available at: https://hdl.handle.net/11299/164018.


The Endings Project Team (n.d.) The Endings Project Team (n.d.) “Principles”. Available at: https://endings.uvic.ca/principles/.


This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License.