

# DHQ: Digital Humanities Quarterly

2014

Volume 8 Number 2

## Burying Dead Projects: Depositing the Globalization Compendium

Geoffrey Rockwell &lt;geoffrey\_dot\_rockwell\_at\_ualberta\_dot\_ca&gt;, University of Alberta

Shawn Day &lt;day\_dot\_shawn\_at\_gmail\_dot\_com&gt;, University College Cork

Joyce Yu &lt;jmyu\_at\_ualberta\_dot\_ca&gt;, University of Alberta

Maureen Engel &lt;maureen\_dot\_engel\_at\_ualberta\_dot\_ca&gt;, University of Alberta

### Abstract

In the digital humanities we specialize in imagining and launching digital projects, but we rarely consider how to end them. In this paper we propose to discuss the ends of a particular digital project as a case study for the planning of ending. The project we focus on is the Globalization and Autonomy Online Compendium that was developed as a digital outcome of the Globalization and Autonomy project. Specifically, this paper will:

1. Survey the general issues at stake when planning for the end of a digital project;
2. Provide some background on the project and the Compendium;
3. Discuss the underlying technologies that had to be dealt with;
4. Address the specific problem of ending and how we prepared this project for archival deposit;
5. And conclude by talking about some other ends that are really beginnings.

When can a digital scholarly project be considered finally "done"? Perhaps never. Something done is past, irrevocable, requiring nothing more and indeed immune from further action.

1

[Brown et al 2009]

### Introduction

In the digital humanities we specialize in imagining and launching digital projects, but we rarely consider how to end them. We imagine that all projects are dynamic and ongoing — taking on a life of their own. Rarely do we admit that some are moribund, or should be gracefully wound down. We don't want to think about their ends so we don't plan endings, and thus many projects are not properly documented and deposited. This is a shame, especially in a field that has thought so long and hard about preservation through the digital.

2

In this paper we propose to discuss the ends of a particular digital project as a case study for the planning of ending. The project we focus on is the Globalization and Autonomy Online Compendium that was developed as a digital outcome of the Globalization and Autonomy project.<sup>[1]</sup>

3

Specifically, this paper will:

4

1. Survey the general issues at stake when planning for the end of a digital project;
2. Provide some background on the project and the Compendium;
3. Discuss the underlying technologies that had to be dealt with;
4. Address the specific problem of ending and how we prepared this project for archival deposit;
5. And conclude by talking about some other ends that are really beginnings.

### The Problem of Endings

The vast expansion of digital resources and digitization technologies through the 1990's meant that much of the attention of librarians, archivists, and researchers was dedicated to establishing best practices for preserving newly digitized cultural heritage materials. While born-digital research projects were developing in parallel to these large-scale digitizations, the work of establishing best practices for concluding and archiving those research projects necessarily involved both an extension of, and a significant departure from, the practice of digitization. As Margaret Hedstrom noted in 1998, "the critical role of digital libraries and archives in ensuring the future accessibility of information with enduring value has taken a back seat to enhancing access to current and actively used materials" [Hedstrom 1998]. Still, Hedstrom is able to articulate two key departures for digital preservation: that preserving media alone cannot preserve "information with enduring value" and that technologies for mass storage do not translate into technologies for long-term preservation. Print-based projects found their conclusion in the physical medium ("book", "article" etc.), and libraries and archives preserved those objects by means of acid-free paper, climate control, etc. The technology for mass storage *did* constitute a technology for long-term preservation. Born-digital research, on the other hand, has yet to comprehensively establish the conventions that will be capable of dealing with the now separate requirements of doneness and archival object — form *and* content. Each new project reveals nuances and complexities that we must address as we develop new best practices.

5

In the last few years, there have been a growing number of organizations that have been making advancements in digital preservation methods. This includes the Reference Model for an Open Archival Information System (OAIS), CEDARS (CURL, Consortium of University Research Libraries, Exemplars in Digital Archives-UK), National Library of Australia (NLA), and RLG/OCLC (Research Library Group). Indeed, librarians, archivists and other information professionals have a vital role to play in developing and maintaining such methods and standards, and in providing the stable end-point for deposit. As Kretzschmar and Potter persuasively argue, "collaboration with the university library is the only realistic option for long-term sustainability of digital humanities projects in the current environment" [Kretzschmar and Potter 2010]. Given the exigencies of project funding, project leadership, and ever-shifting technological developments and demands, the stability of the institutional archive is vital.

6

Being aware of such efforts and debates, however, and successfully depositing a digital project, are not always the same thing. The Compendium did many things right, and in fact was forward-thinking in its approach to its own ending. Still, even with improved standards, thoughtful guidance, and more rigorous practices, important lessons were learned through the process. These will provide valuable guidance to future DH projects, and provide a significant supplement to the existing literature.

7

## Project Background

The Globalization and Autonomy Online Compendium was one of two major coordinated outcomes of the Globalization and Autonomy project. The other major deliverable was a ten-volume academic book series published by UBC Press.<sup>[2]</sup> The project was supported by a Major Collaborative Research Initiative (MCRI) grant of \$CAD 2.5 million from the Social Sciences and Humanities Research Council (SSHRC) of Canada, awarded in 2002. The project was "concluded" (insofar as the funding was concerned) in 2007. And yet, as papers in a special cluster of DHQ titled *Done* discuss: Is a project ever truly done?<sup>[3]</sup> Certainly the publication of all the print volumes took longer.

8

The Globalization and Autonomy project was led by William Coleman at McMaster University, and involved over forty co-investigators in twelve universities across Canada and another twenty academic contributors around the world, not to mention funded graduate students and staff. Geoffrey Rockwell was a co-investigator with the responsibility for the design and management of the Compendium; Nancy Johnson was the editor of the Compendium; and both Shawn Day and Joyce Yu were research assistants hired to help manage the deposit process. Unlike some projects, the digital deliverable was woven in from the beginning — it was written into the grant, budgeted, assessed at the mid-term review and peer reviewed by the UBC Press.<sup>[4]</sup> As part of weaving in a digital component, the team also planned for a wrapping up when the Compendium would be archived and deposited, even if it was also maintained online. While we planned for an end, what we failed to estimate properly was how difficult it is to wrap up a project. This forms one of the crucial learning outcomes of the digital side of this project and probably serves as a benchmark reminder of the crucial importance of budgeting and planning for the "end" of a project. In fact, as in so many cases, an end that allows for a sustainable continuity to the research is one of the most important outcomes of any project.

9

10

The goal of the Globalization project was "to investigate the relationship between globalization and the processes of securing and building autonomy". The project was designed from the beginning and administered to understand globalization in a collaborative and interdisciplinary way that avoided the often political and economic focus of globalization research. Hence the "autonomy" in the title — we were looking at globalization and resistance to globalization.

## The Compendium

The screenshot shows the entry screen of the Globalization and Autonomy Online Compendium. The header includes the project logo and title in English and French. A search bar is present in the top right. The main content area is divided into several sections: a welcome message under "Globalization and Autonomy", a "New to Globalization and Autonomy?" section with a link to a short introduction, and a "Bibliographic Database" section with a search box. A left sidebar contains "Topics" and "Quick Links" menus. A "Glossary" section is also visible with a search box and category links.

Figure 1. Entry screen to the Compendium

From a user perspective the initial interface of the Compendium was engineered to introduce the project, the compendium itself, and the UBC Press print series. Visitors to the online Compendium were presented with a sampling of the most recent contributions as well as an opportunity to browse within major topical areas. A provision was made for search, but browsing and discovery were presented as the principal ways for using the site through a prominent menu down the left. There was also, at the bottom of the entry page, information about one or another of the volumes in the coordinated publication series.

11

**Featured UBC Press Volume**

**Empires and Autonomy: Moments in the History of Globalization**

This volume synthesizes the recent outpouring of works on globalization and world history. Its essays depart from the existing literature by depicting forms of autonomy - individual, local, cultural, state, imperial - that were effaced, challenged, reconfigured, or even brought into being by emerging transcultural systems and empires.

[→ All UBC Press Volumes](#)

[G & A Home](#) | [Contact Us](#) | [Site Help](#) | [Site Map](#)  
 Copyright © MCRI Globalization and Autonomy. ISSN 1916-5579

Figure 2. Bottom of entry page with featured UBC Press volume

If users follow the left-hand menu to one of the topical pages, they see a list of all the research articles, position papers, summaries, and other materials associated with that topic. They can then click on an item to get an HTML view of the article. The views of the materials are dynamically generated from XML so that users can get different views of same materials. All these materials can be viewed in the normal HTML view and in a print view (without all the navigation and site design) as PDF, or accessed as XML. The print view and the PDF were to support global readers who might not have stable access to the web and therefore want to save and/or print materials. These views are generated dynamically so that only the XML has to be maintained. We also pull out a Table of Contents for the article (see Figure 3) and a list of relevant Glossary items from the article XML when we generate the standard HTML view.

12

entity	<p><b>American Foreign Policy and the Foundations of World Order</b>                  Louis W. Pauly, University of Toronto</p> <p style="text-align: right;"><a href="#">Print Article</a>   <a href="#">PDF Version</a>   <a href="#">XML</a></p> <p><b>American Foreign Policy and the Foundations of World Order</b></p> <p>Intense debate is underway within the United States and elsewhere on the reconstruction of world order under conditions of rapidly deepening economic and social integration. The debate is not new, but it has lately entered into a more pressing phase. This paper seeks to put that debate into contemporary historical context and examine the implications. One set of implications focuses on the adaptation of novel international economic institutions created in the aftermath of systemic catastrophe in the middle of the twentieth century. A related set concentrates on requisites for stability in the relationship between leaders and followers in a system reshaped by the myriad pressures now conventionally associated with the word "globalization."</p> <p><b>Leadership and Followership in the International System</b></p> <p>Global economic and social order is not spontaneously generated. Policy shapes it, although not necessarily in circumstances of policy-makers' own choosing.<sup>1</sup> The specific policies of some states are</p>	<p><b>Contents</b></p> <ul style="list-style-type: none"> <li>→ <a href="#">American Foreign Policy and the Foundations of World Order</a></li> <li>→ <a href="#">Leadership and Followership in the International System</a></li> <li>→ <a href="#">Foundations of World Order</a></li> <li>→ <a href="#">Contemporary Challenges</a></li> <li>→ <a href="#">Acknowledgement</a></li> <li>→ <a href="#">Works Cited</a></li> <li>→ <a href="#">Notes</a></li> </ul> <p><b>Glossary Terms</b></p> <ul style="list-style-type: none"> <li>→ <a href="#">International Monetary Fund</a></li> <li>→ <a href="#">World Bank</a></li> <li>→ <a href="#">global governance</a></li> <li>→ <a href="#">World Trade Organization</a></li> <li>→ <a href="#">League of Nations</a></li> <li>→ <a href="#">World War II</a></li> <li>→ <a href="#">democracy</a></li> </ul>
--------	---	--

Figure 3. View of the top of an article with the alternate views

In addition to the articles written by project participants there is a wide-ranging and searchable hypertext Glossary of terms and issues related to globalization as well as a searchable bibliographic database that includes all the references in the individual articles and the glossary items. There is also an interactive site map that can be used for navigating and understanding the Compendium as a whole (see Figure 4).

13

## Advanced Search

Terms:  Full Text ▾

Sources: 

- All Docs
- Glossary
- Position Papers
- Research Articles
- South\_North Dialogue

Hold down the Ctrl key (or Apple key) to select multiple entries.

Topics: 

- All Topics
- Community and Identity
- Culture
- Democracy

Hold down the Ctrl key (or Apple key) to select multiple entries.

**SEARCH**

Figure 4. Interactive site map to the Compendium

In addition to the search interfaces for the Glossary and the Bibliography, a global search and an Advanced Search page is also available (see Figure 5).

14

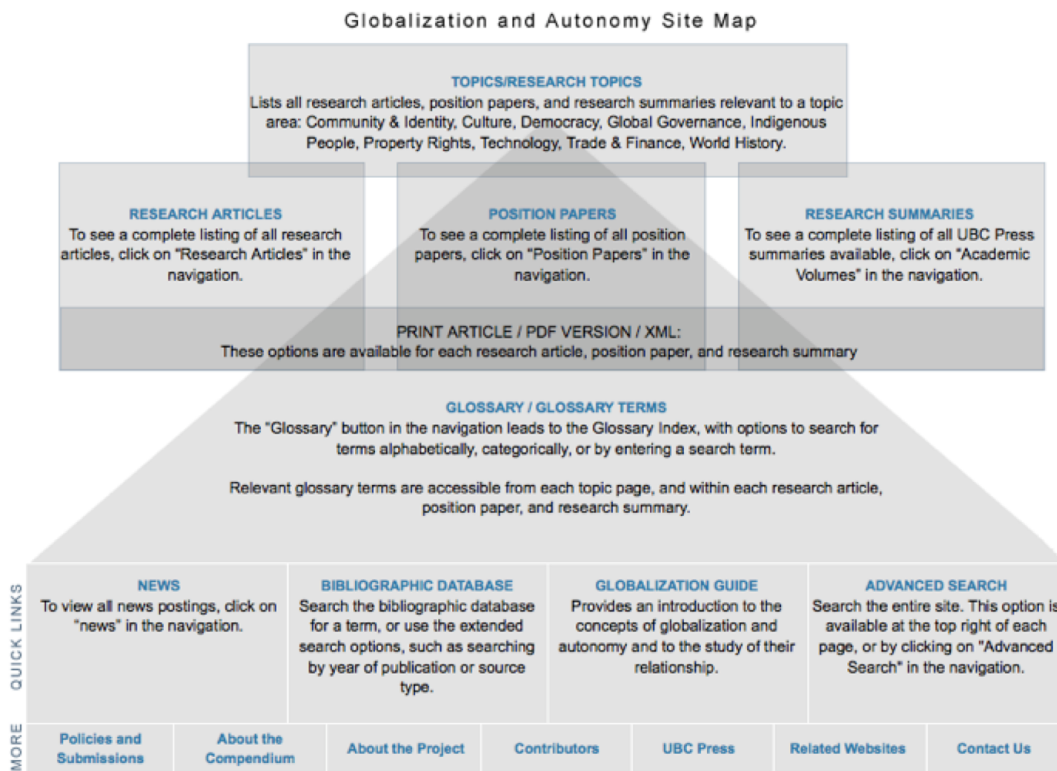
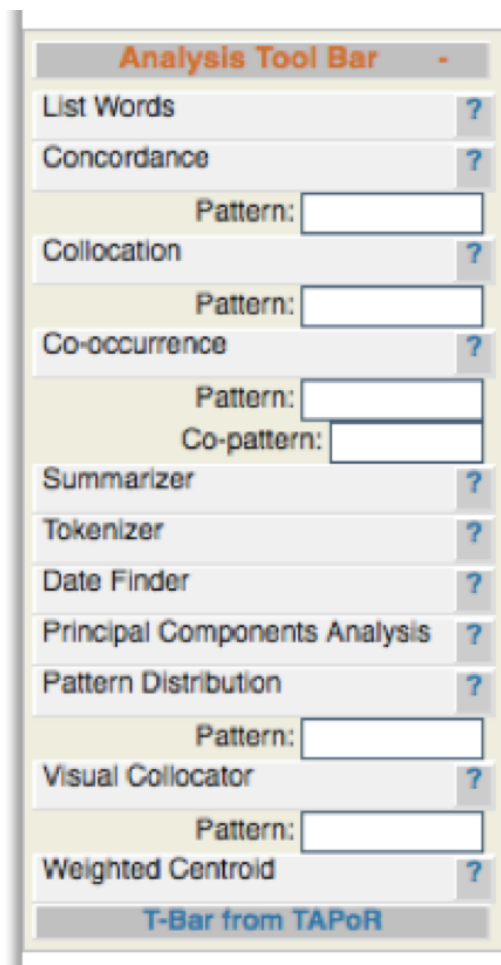


Figure 5. Advanced search facility

Finally, we built in a text analysis tool bar that shows up on every article, position paper, and research summary. This tool bar, when used, sends just the text of the article to a TAPoRware tool for analysis. [5]

15



...certainly no longer can  
invisible hand of the ma  
flows. It is an increasing  
process of neo-imperia  
intended to control, cor  
intense and multiple tra  
globalization. These pr  
connected with power,  
and Pauly 2005; Hedeto

The important implicati  
challenge to states, do  
world without states, wi  
interests. Rather, it cre  
economic environment  
relations, and a new di  
according to the resour  
adaptability that they c  
(Baylis and Smith 1997;  
particularly developing  
increasing inequality, p

Figure 6. Text analysis tool bar from TAPoR

## Underlying Technology

Digital humanities readers can probably imagine much of the technology involved in the construction of the Compendium, but a review of the technological structure of the project is essential to understanding what aspects needed to be wrapped up and the challenges faced in depositing it. 16

Most of the content in the Compendium was written by participants and submitted as Microsoft Word files for editing. A decision was taken early not to try to force contributors to learn our XML application. The editor and assistants translated and encoded the content using a Text Encoding Initiative (TEI) application for contemporary research developed with consultants including Julia Flanders and Syd Bauman. <sup>[6]</sup> During the encoding process, the editor inserted links from articles to glossary entries. This was done using a search and replace batch tool we built that allowed us to update the links regularly as we received more glossary entries. 17

For the TEI application — the Document Type Definition (DTD) we developed following the TEI guidelines — we followed best practices including involving outside consultants at regular intervals to advise and later review the system. The same applied to the technical structure and the interface design. For the interface design we followed a persona/scenario design process [Cooper 2004] that involved project stakeholders outside the development team. Interface personas and usage scenarios were presented to the full team at one of the annual conferences as were successive iterations of the Compendium. This was important to get feedback and buy-in from the larger team, many of whom were focused on their own research and the volumes planned for the UBC Press series. We also had to negotiate what would go in the volumes and what would go in the Compendium. In the end we included summaries of all relevant chapters from the print volumes in the digital Compendium — a compromise with the UBC Press that had the advantage of allowing the editor to rewrite summaries for the Compendium in language more accessible than the volumes and more suitable for the web. 18

Because our anticipated audience included students and policy workers, we strove to make the online Compendium accessible to non-academics interested in Globalization and a thorough Glossary was an important addition to that end.

To manage the Compendium as it was being edited, we developed an administrative interface with tools so that the editor could manage the TEI files. The administrative interface would parse, verify, and process files when uploaded. Upon uploading for the databases, metadata that dynamically generated tables of content would be extracted; the files were also indexed for searching. This administrative interface is hidden and less polished. One of the questions we had to ask ourselves was what should happen to this part of the system at the end. Should it too be documented and archived? More generally, we had to ask what was ending and what had to be documented. A project of this size has all sorts of ephemera from email discussion archives to conference materials. Even before the actual "content" was being produced, important questions of what ought to be documented and preserved arose.

19

The bibliographic records followed a different administrative process and were entered directly through a separate web interface to the database. At an early stage in production we encountered an interesting problem: how to synchronize a single bibliographic database for the project with all the bibliographies from the individual position papers, articles, and glossary entries as they trickled in. As documents came in, we found that different writers would enter slightly different information for the same reference. We wanted a way to normalize the bibliographies without having to wait until the end. What we settled on may seem somewhat complicated but it made sense in the context of the wider ongoing project and was what the larger team wanted when we presented the problem to them.

20

1. First, the editor did not encode the bibliography of an article in the XML file. Instead, she entered any new records into the online database and checked if any bibliographic entries (from a previous article) were already entered.
2. If an entry already existed, she checked the new author's entry against the existing record. In case of a discrepancy, she researched the reference and corrected the database entry as needed.
3. Once an entry was checked/added/updated in the database, she generated a stub tag with a key that corresponded to the database record. This, rather than the full record was put into the XML file for the article. Then the article could be uploaded to the Compendium.
4. When the XML version of the article was uploaded, the system replaced the stubs with a full TEI `<bibl>` entry from the database. Thus, should articles become detached from the bibliographic database (as happens when one deposits the materials), each article will include a full bibliography marked up in XML. The project team in one of our meetings on the Compendium felt this was important; they didn't want their writings dependent on dynamic generation from other data for completeness.
5. In the uploading we also kept online all the XML files with just the stubs. This allowed us to periodically rerun the process that added the full bibliography and replace the full XML files, thereby eliminating any inconsistencies that might occur as we correct entries over time. In short, we could regenerate the content on a regular basis to guarantee consistency.

As for the components of the Compendium, it is technically composed of:

21

- XML files with the content;
- A MySQL bibliographic database;
- A metadata database of the content for generating topical pages and for searching;
- A full text index for searching the text;
- The code that handles the dynamic generation of the site, the searching, linking, and the XSL transforms;
- Some HTML pages and CSS stylesheets;
- And various images that are embedded in pages.

The XML files are not on the site, and therein lies the problem of depositing the Compendium as a whole. The true, though simple, story we tell ourselves in the digital humanities is that ensuring that a project uses an appropriate form of markup (like the TEI) for content is sufficient to preserve the work. The experience of the Compendium is that the intellectual work is not only in the individual articles, or even in the bibliographic data – it is in the interaction between these, mediated by code and in the user experience. The Glossary is a prime example — the meaning is not just in the text of entries, but also in the searchable whole and web of articles linked to glossary entries. Likewise the interface design reflects decisions about the audience that is unquestionably important to understanding the work as a whole. These difficulties clearly instantiate the difference between what Paul Conway has named "digital preservation" and

22

"digitization for preservation". While they are "intimately related... the underlying standards, processes, technologies, costs, and organizational challenges are quite distinct" [Conway 2010, 63]. We initially thought it would be trivial to deposit the Compendium – in the grant proposal we promised that we would encode the content following the TEI guidelines and then deposit it at the Oxford Text Archives and other similar digital archives (digitization for preservation), but of course, the XML is not the Compendium. The Compendium is a work of its own that is more than the sum of the XML files. How do we deposit such a system (digital preservation)? What exactly are the boundaries in time and scope of the work that mark what should be deposited? Ultimately we realize that the choices we make in constructing the deposit reflect a major intent of undertaking the project itself and that the ability to capture user experience must increasingly inform digital project planning.

## What is Done?

As Matthew Kirchenbaum writes in the introductory essay to a Special Cluster of the Digital Humanities Quarterly on *Done*, "What does it mean to 'finish' a piece of digital work?" [Kirchenbaum 2009]. Large complex sites that reflect a group of researchers, like the Globalization Compendium, can always be added to as people write another working paper or glossary entry. The web allows one to publish a work online and keep on updating it. Users want and expect good resources to be maintained. They expect the interface to be refreshed, corrections to be made, and content to be updated. If not, a site looks stale and becomes a suspect resource. [7]

23

Susan Brown and colleagues nicely describe the tension between how we call these sites "projects" with ends and how they can yet take on a life of their own. A project is defined as something with a planned and anticipated end, and we call these things "projects" in that sense when we are applying for funding. But, when successful, we also want to keep on experimenting and adding to a "projection". Thus projects morph into projections with futures for knowledge. As Brown et al. put it, "This interplay between traditional humanities content and innovative methodologies means there is always more to be done" [Brown et al 2009].

24

So what did we consider "done" in the Globalization and Autonomy project? When and how did we plan to end the Compendium? First of all, as a grant-funded project, Globalization and Autonomy had a natural deadline as the funding was supposed to be spent by 2007. SSHRC does allow projects to apply for extensions if the funds are not spent and there were some funds secured from other sources that were independent of the grant, but effectively the plan was to finish the Compendium and deposit something by 2008 when the funding for researchers, research assistants, editors and programmers would wind down. We had anticipated that more work would be involved to deposit the project and had allocated funds to do so. What we underestimated was how long it would take to deposit the project. We finally deposited the materials in 2012 — four years following the official "conclusion" of the project. In retrospect, it was unreasonable to expect the project to be wrapped up and deposited during the last year of the project. It would be akin to archiving a book while the author is still writing the conclusion.

25

This raises a question of whether the model of grant-project funding needs to be rethought. Ideally, projects would properly account for the time needed to wind down and bury their dead data, but we believe that is unreasonable, at least in 3- to 5-year projects, which is the typical funded length in Canada. Given the rhythm of the academic calendar, the time it takes to get a project going and then to develop digital research sites makes it hard to do a good job that ties everything up. Any interesting project will change as it continues to do research that further challenges tidy timelines. Better would be small grant programs that provide funding for finished projects that secure trusted preservation partners to wind their project down, document the project, and deposit appropriate data. Previously funded projects, for which granting agencies have some responsibility, would then be encouraged to be open about what they had and how they could make sure the research data was available over the long term. Such a grant program would also draw attention to data preservation and foster research into preservation. But, for many this is simply the least glamorous aspect of the project process and therefore hardest to find internal motivation to accomplish.

26

As for what would be done in the ending of our project, it made sense to us to deposit the project as it was at the end of the grant. The state of the Compendium at that point represented a natural moment to deposit the materials. In fact, because it took us longer than expected to work out where and how to deposit the project, we were able to include materials from related conferences. These additional materials were added to the Compendium in 2009 as we were still working things out. Nonetheless, the goal was to deposit the Compendium as of the end of the grant in 2007 and into 2008.

27



As with many digital projects, we did not conceive of the version finished in 2008 as necessarily the end of any work on the Compendium. The SSHRC grant had funded the development of a useful resource and the project leader had ideas for future phases of the Compendium, but we felt we should nonetheless assume that the Compendium might never be funded again and deposit its state at the end of the grant. 28

It should be noted that a number of things have been done to the Compendium since 2007. William Coleman ran a conference in 2007 titled *Building South-North Dialogue on Globalization Research* that brought together researchers from the Global South and Global North, and then a follow up meeting in August of 2008. <sup>[8]</sup> Selected papers from the conferences were reviewed, edited, and added to the Compendium. More importantly, the conferences looked at the idea of developing the Compendium into a portal for research into globalization that would enhance South-North dialogue. The consensus was, however, that there wasn't sufficient need or support for such a portal. Another change to the Compendium happened as a result of Coleman leaving McMaster University for the University of Waterloo bringing the Compendium with him. The Compendium was redeployed to a server at the University of Waterloo and a news feed from Coleman's blog was added to create a new entry page that could stay current. <sup>[9]</sup> 29

A third consideration as to what should be done at the end of the project was that the funder (SSHRC) requires all projects to deposit their research data. To quote from the SSHRC Research Data Archiving Policy, "All research data collected with the use of SSHRC funds must be preserved and made available for use by others within a reasonable period of time." <sup>[10]</sup> All who accept SSHRC funds are obligated to do so, although a survey SSHRC conducted as part of the National Research Data Archive Consultation revealed that very few datasets could actually be found: 30

In any given year, as many as one-half of SSHRC-funded researchers produce research data. For those who responded to this consultation, the figure is 55 per cent. This extrapolates to approximately 1200 data sets created by SSHRC-funded researchers between 1998 and 2000, or an average of 400 each year. As of January 2001, only 7 per cent of those researchers surveyed had archived their data, and only a further 18 per cent reported that they intended to do so. Of the 18 per cent that intend to archive their data, less than one half were able to identify an actual data archiving service or agency. [NRDAC 2001, 8]

The even harsher reality when considering this mandated archiving of publicly funded research is that globally, far fewer funding agencies even demand that the research products be effectively archived, let alone provide the mechanisms to do so. <sup>[11]</sup> The uncomfortable truth of project ends is that we don't properly bury our projects, even though we know this is what we should do and often talk about using guidelines like the TEI. And that is what scholarly encoding following best practice guidelines like the TEI is about — encoding one's data so that others can understand the decisions and be able to reuse it long after the original researcher is gone. Projects should be designed from the beginning to die gracefully, leaving as a legacy the research data developed in a form usable in the future. An archived project must not be expected to be a live one, and yet, as Paul Conway argues, our expectations of the digital information environment require that our access to materials remain (apparently) unmediated (Conway 63). We are fooling ourselves if we think projects will survive over time as living, well-maintained projects. Ask yourself how many projects you have let lapse without a service. Likewise, we are fooling ourselves if we think we can always do the burial next year when we think we have more time. 31

But, why bother? SSHRC succinctly defines some of the reasons for archival deposit: 32

Sharing data strengthens our collective capacity to meet academic standards of openness by providing opportunities to further analyze, replicate, verify and refine research findings. Such opportunities enhance progress within fields of research as well as support the expansion of inter-disciplinary research. In addition, greater availability of research data will contribute to improved training for graduate and undergraduate students, and, through the secondary analysis of existing data, make possible significant economies of scale. Finally, researchers whose work is publicly funded have a special obligation to openness and accountability. <sup>[12]</sup> 33

They don't say it explicitly, but another reason to deposit is that our research, itself, is of its time and grist for the mill of future researchers who may want to study us. Our artifacts carry, despite our best intentions, hermeneutical baggage. That which we bury may be of interest to the archaeologists of knowledge of the future. SSHRC expects us to be open so that others can study the research process once we are dead, buried and history — a rather alarming prospect, but one of the features of an emerging philosophy of open research that advocates for exposing the research process rather than hiding the mess behind authoritative results. 34

## Depositing the Compendium

Having defined the complex digital and human nature of the Compendium, we were faced with determining what exactly we were going to deposit and where could we deposit it. What were the components in a seemingly straightforward technical construct that we considered essential to deposit and how could we package these in a means that would allow for their potential disinterment? Just as important to this process was the consideration of where to deposit that would allow for both preservation and access to the constituent parts. <sup>[13]</sup>

To build our deposit package we identified four key aspects of the digital project to attempt to capture and preserve:

**Content:** These are the original research articles and other documents (bibliographic database, HTML pages) created and published in the Compendium. This, of course, raises the question of exactly what is content. Is there not content to an interface independent of the text? Those questions are for another paper; in our case we considered content to be the texts, including bibliography, and glossary. We also considered the text on the HTML pages content.

**Code:** Although the underlying code that we have described may seem obvious, this is where things begin to become difficult to cleanly identify. While one can easily deposit code, it is difficult to imagine code being useful for people or usable for archaeologists who might want to reconstruct the Compendium. So we may ask, why deposit code? Our reason was that in the code lies the interactivity and interface — for us code includes the XSLT code that generated much of the interface. We decided that one of our objectives was to deposit materials that would allow for the reconstitution of the Compendium in its interactive form whether through the interpretation of code or, less likely, through the reconstitution of a working system.

**Process:** Related to code is process. We wanted to not just deposit the components that resided on the server (XML files, database files, and code), but also to deposit information that would trace the processes of the Compendium as a collaborative project. The Compendium is the result of various research, programming, and editorial processes and decisions — many of which are documented in instructions to authors and coders and other administrative documents, including documentation around the deposit process itself stored in the wiki. The process whereby we handle synchronizing bibliographic entries mentioned above is a case in point — it made a difference to the content that isn't apparent in the final XML files which hide the process whereby bibliographies were generated. Therefore we decided to deposit certain materials (but not all) that document the editorial processes, including the editorial backend that strictly speaking was not part of the Compendium as experienced.

**The User Experience:** Lastly we wanted to deposit information that allows people to get a sense of what the user experience of the compendium might be without having to reconstitute a working copy. We made a concerted attempt to deposit information about the experience of the Compendium as an interactive work by writing a narrative along with screen shots of typical use of the Compendium stored as PDFs. This has the added advantage that it could help someone wanting to re-implement some part of the interactive code.

It is also worth noting what we did not prepare for deposit. We did not prepare for deposit materials from the project that went into the UBC Press volumes because their printing preserves that research in its final state. To gather those materials would also have been a time consuming project as we would have had to contact all the authors, negotiate rights, and negotiate with UBC Press. Nor did we prepare for deposit materials from the running of the Globalization and Autonomy project itself; the only process documents chosen were those associated with the editorial process for the Compendium and materials concerning the decision-making around the depositing process itself. (We added, at the last moment, copies of the wiki where we documented our deposit project.) Again, depositing all process documents was beyond what we could do and it would have taken considerable work to gather, negotiate rights for, and document all the discussions. Finally, we did not gather draft documents for the Compendium itself, such as the first drafts of papers before editing and encoding. We focused instead on depositing the Compendium — the finished online publication and the experience of it. As useful as other materials might be, we believe the most important materials to deposit should be those that were carefully prepared for publication and public viewing. These represent what the team collectively wanted to pass on to the larger community as useful research.

As for how to organize the deposit, we decided to create a deposit collection with these four components (content, code, process, and experience) separated, each in the best preservation formats we could find. This was easy for the content; it was designed from the start in XML, which is, to a certain extent, self-documenting. But in the case of code it is less clear.

For the code, all the materials were output to a flat-file format, so things like the bibliographic database were output to XML. The code was then minimally commented so that it could be compiled, and documentation was generated in HTML or XML in an industry standard fashion where possible, though we note these standards are for documentation, not preservation. The point is that the documentation is embedded in the code and could be extracted to produce documentation assuming that future computer scientists recognize how to extract documentation.

We also created "Read Me" documents describing the environment and the technical arrangements needed to run the code. We realize this means we did not deposit a working system that someone could download, install, and run to recreate the Compendium. The databases were not stored in their native format; they will have to be regenerated and we didn't create a tarball (tape archive file) of the whole site that could be unarchived into a directory on a suitable machine to provide a working instance. Frankly, we doubt anyone would bother and we doubt server configurations will stay stable enough for unarchived tarballs to work years from now. Instead, we sought to deposit something that could be explored as is AND used to recreate aspects of the current Compendium in the more distant future if that matters to information archaeologists. Preservation is, after all, the protection of "information for access by present and future generation" (Conway, cited in [Hedstrom 1998], our emphasis). Access is not recreation, and over the long term the chances that someone can recreate the hardware and software platform on which an installation could work will approach zero. We are better off giving them something they can understand and re-implement, if needed, than something they can't install. Further, the purpose of depositing is not only so that people can recreate the original site, but also so that they can study the Compendium and reuse it in unanticipated ways.

43

Similarly, we were not trying to deposit something in a form where the interactivity could be maintained. There are models for preserving interactive objects so that they can be easily run on emulators. The most obvious would be to move all the interactivity into XSL or other XML standards for interactive processing like SMIL. The reasons we did not go that route are that it is too expensive, it probably won't capture all of the interactivity, and we do not hold much confidence in any of the candidate standards for interactivity. Think about what happened to HyTime as a standard (if you have even heard of it).<sup>[14]</sup> Instead, we created a deposit package with screen shots of the experience of using the Compendium so that someone who also had the code and content could at the very least understand the experience and, if they chose to re-implement things, could recreate it. There is nothing glitzy or flashy about this, and that is the point. We are trying to recreate the experience in as basic a form as possible to allow for appreciation of the experience using as simple tools as possible, for as long as possible.

44

Lastly, we should mention that we prepared the package for deposit in an Institutional Repository, but we also prepared other ways of archiving the Compendium. We printed selected parts of it on acid-free paper so there would be a paper archival copy and we burned the package to CD-ROM and distributed the CD-ROMs to participants so that the CDs might survive. Now we will turn to the deposit itself.

45

## The Question of Where to Deposit

Once the issue of *what* to deposit had been settled, the question became where to deposit the Compendium. We decided that multiple depositories were the most prudent solution, and one that presented the best possibilities for long-term preservation and future access. The question then was where to find Institutional Repositories (IR) willing to take the materials. Library and Archives Canada<sup>[15]</sup> is the official depository for published material in Canada, but they were not equipped to take digital materials. When it comes to funded research from institutions of higher education, a number of university libraries have been certified to receive research materials.<sup>[16]</sup> At the time, the project host institution, McMaster University, did not have an institutional repository, which meant we didn't have a local IR. Another option was to find an international IR. Unfortunately those we contacted were not able to help us at the time due to funding constraints. (Alas, it seems that some IRs have had their funding cut to the point of being barely able to operate, which does not bode well for preservation and is another reason for depositing to multiple IRs.)

46

The situation in 2008 thus seemed hopeless. We had done all this work to prepare a package and now we couldn't find anyone to take it. Fortunately time heals IRs, and since then both McMaster University and the University of Alberta have set up IRs able to take our data. With the help of the Digital Preservation Librarian Nick Ruest we deposited the package at DigitalCommons@McMaster.<sup>[17]</sup> We have also deposited it to the Education and Research Archive (ERA) of the University of Alberta.<sup>[18]</sup> As a side note, although Library and Archives Canada did have it in their brief to accept the package and were in fact anxious to learn by doing so, revisions to the Library and Archives Canada Act (2004) in 2007

47

removed *Databases or web-based applications* from their brief. <sup>[19]</sup> Seemingly they have backed away from the challenge despite the increasing use of such means to make research material available for reuse.

## Conclusion

This brings us back to ends. There are ends to the Compendium other than its deposit. One end that is not in the scope of this paper, but may interest readers, is managing the review of the Compendium so that those of us who worked on the digital design but are not published in the book series are recognized. When the project negotiated with publishers, one reason we selected UBC Press was that they agreed to conduct a peer-review of the Compendium along with the print series. We often talk about how digital work isn't reviewed, which causes trouble in the academy, but in this case the Compendium was reviewed and, in effect, accepted for publication as a companion to the print series.

48

A more important end, of course, is an anticipated beginning. We may be burying the Compendium, but we do so partly in order to bring it back to life in a new form. William Coleman has been negotiating with other organizations to develop the Compendium into an ongoing peer-review venue that would, like a journal, continue to accept and review new materials. But unlike a journal, these would be organized encyclopedically rather than chronologically in issues. With the addition of community tools like forums and comment features that are not peer-reviewed, we hope the Compendium could evolve into a venue for research and learning around Globalization, especially from the Global South — voices that are too infrequently heard in our circles. The quality, extent and even distribution of content we now have are, we believe, assets on which we can build relationships.

49

Finally, an end is what one learns. Some of the key lessons we learned from this project are:

50

- Researchers and other research stakeholders need to take research data deposit seriously and allocate it the necessary time and resources.
- Researchers should decide when something is done and what that means *early* in a project, not in the rush of a messy end. At the same time, researchers should be willing to redefine the when and what as the project evolves.
- The decision of *what* to do when a project (or phase) is done is not simple. At the very least the content published online should be deposited in such a way that it can be discovered and reused.
- Researchers should seek *advice* from local librarians about the deposit services available and current best practices.
- It is better to deposit something at some point than to perpetually put it off. Beware of developing the wrapping up process into such an ambitious project that it is never finished. Finishing a project should not become a new project.
- Researchers should consider depositing information about the deposit decisions themselves. That sort of metadata can help future users understand what was deposited and why.
- It takes significant time and effort to deposit a project. It is not a matter of spending a weekend uploading the XML at hand to a repository.
- That which you decide to finish and deposit will change as you deposit it. Be prepared for a long deposit process that will have to deal with a moving target.
- For grant-funded projects, you should in principle budget and plan a first deposit at the end of the funding. However, this is not practical as most projects are not really finished until the very last moment of the grant, if that. At the end of a grant there is usually a rush to finish what you said you would do and therefore no time to step back, consider what was done, and carefully document and deposit the project. For these reasons you should hold back some resources to be able to deposit a project after it has ended, not while it is ending.
- Grant funders should consider small post-award deposit funding grants that large projects can apply for once the project is over, the digital work done, and budget spent.
- Projects are often associated with individual academics. These academics change institutions. Digital humanists and institutions should therefore be willing to take ongoing responsibility even when people move away from institutions where they developed projects. At the very least, institutions should be prepared to help document and deposit projects even when faculty move on, and individuals should be prepared to finish off projects and deposit them even though they may have gone to different jobs.
- Institutional and government stakeholders need to expect and invest in the long-term preservation of data. The poor rate of depositing research data documented by the NRAC Working Group (2001) is due to a

number of factors including lack of trusted institutional repositories staffed to explain the importance of depositing data to researchers. There is also a lack of understanding of the long-term commitment needed for such repositories to fulfill their function. Repositories can be set up and then abandoned for other priorities. Unfortunately, we seem to be in a time where no one wants to make long-term commitments.

- Articles about projects, like this one, that are published and preserved by libraries are likely to be an important way that researchers in the future learn about projects. It may seem superfluous to document how a web site that is still online and usable works, but given how quickly a web site can disappear, published reports, white papers, and articles can be an important end.

The most important lesson from our experience is that it is difficult and time consuming to deposit research materials. Until more people actually start depositing datasets and institutions start encouraging deposit, it will continue to be so. There is precious little discussion and experience for guidance. A second lesson is that so much of the idea of long-term preservation is just that — an idea. In principle there are many places that can accept deposits, but in practice it isn't always easy. Repositories come and go; some have even had their funding pulled back. Support also comes and goes, leaving academics confused about the importance of this and how to do it.

51

We believe that by depositing in multiple forms, in multiple locations, and with rich documentation of the process and experience, we have buried the Compendium in a suitable open casket, ready for reanimation or reuse. Perhaps no one will be interested, but that isn't for us to say. Let the worms loose.

52

## Notes

[1] The Compendium is still running at <http://globalautonomy.ca/global1/index.jsp> For information about the project see [http://globalautonomy.ca/global1/about\\_project.jsp](http://globalautonomy.ca/global1/about_project.jsp); We would like to recognize for their assistance Dr. William Coleman, the project lead and Academic Editor of the Compendium, and Nancy Johnson, also Academic Editor of the Compendium. We would also like to thank the Social Science and Humanities Research Council of Canada and McMaster University for support of the project. This paper is based on a presentation given at *DH 2009* at the University of Maryland, College Park, Maryland.

[2] See [http://www.ubcpres.ca/books/series\\_globalization.html](http://www.ubcpres.ca/books/series_globalization.html)

[3] See DHQ Spring 2009, v3, n2, <http://digitalhumanities.org/dhq/vol/3/2/index.html>.

[4] The Compendium was reviewed as part of the peer review process run by UBC Press when they reviewed the proposal for a book series.

[5] TAPoRware can be used by itself at <http://taporware.ualberta.ca/> where there is also documentation on how build a similar tool bar for other projects.

[6] The Text Encoding Initiative (TEI) Consortium have developed Guidelines for the encoding of digital texts. See <http://www.tei-c.org/> for more.

[7] These expectations overturn the academic expectations we have for published materials where we want the stability of an edition in order to properly cite a work. Something that is perpetually a work in progress lacks the stability to be cited which makes it difficult to base other research on. See [D'Iorio 2010].

[8] For a summary of the conference see [Coleman 2008].

[9] The entry page and the home page of the Compendium used to be the same. Now the entry page <http://globalautonomy.ca> leads to <http://globalautonomy.ca/global1/news.jsp> while the home page is at <http://globalautonomy.ca/global1/index.jsp>. Coleman's blog is at <http://wdcoleman.uwaterloo.ca/>.

[10] See [http://www.sshrc-crsh.gc.ca/funding-financement/policies-politiques/edata-donnees\\_electroniques-eng.aspx](http://www.sshrc-crsh.gc.ca/funding-financement/policies-politiques/edata-donnees_electroniques-eng.aspx). SSRHC considers a reasonable period to be within two years of completion of the project.

[11] The UK may be a notable exception to this where, despite the demise of the ADHRC, the UK Data Archive (<http://www.data-archive.ac.uk>) has made available an open source policy framework for other fund granting agencies and combined this with a series of best practice publications to encourage the research deposit process.

[12] From the Research Data Archiving Policy, [http://www.sshrc-crsh.gc.ca/funding-financement/policies-politiques/edata-donnees\\_electroniques-eng.aspx](http://www.sshrc-crsh.gc.ca/funding-financement/policies-politiques/edata-donnees_electroniques-eng.aspx).

[13] The background research and discussion we carried out is openly available on a wiki at <http://tada.mcmaster.ca/Main/ProblemOverview> as a working document that was modified as we made decisions and

carried out steps in the process. The pages on this wiki were also included the package deposited.

[14] For more on HyTime (Hypermedia/Time-based Structuring Language), an SGML standard for multimedia see <http://www.hytime.org>. HyTime was, alas, ahead of its time and too difficult to use.

[15] For more on LAC see <http://www.collectionscanada.gc.ca/>.

[16] See <http://carl-abrc.ca/en/scholarly-communications/canadian-institutional-repositories.html> for the list of Canadian Institutional Repositories.

[17] See [http://digitalcommons.mcmaster.ca/global\\_coll/1](http://digitalcommons.mcmaster.ca/global_coll/1).

[18] See <http://hdl.handle.net/10402/era.26095>.

[19] See Legal Deposit of Publications Regulations (SOR/2006-337) <http://laws-lois.justice.gc.ca/eng/regulations/SOR-2006-337/index.html>.

## Works Cited

- Alemneh D. et. Al.** "A Metadata Approach to Preservation of Digital Resources: The University of North Texas Libraries' Experience", *First Monday* 7.8 (2002). <http://firstmonday.org/ojs/index.php/fm/article/view/981/902>
- Beagrie et al 2009** "Keeping Data Research Safe 2", *JISC Research Study*, 31 July 2009 [http://www.data-archive.ac.uk/media/1687/KRDS2\\_finalreport.pdf](http://www.data-archive.ac.uk/media/1687/KRDS2_finalreport.pdf) Accessed 28 October 2008.
- Brown et al 2009** Brown, S. et al. "Published Yet Never Done: The Tension Between Projection and Completion in Digital Humanities Research" *Digital Humanities Quarterly*, 3:2, <http://www.digitalhumanities.org/dhq/vol/3/2/000040/000040.html>, 2009.
- Coleman 2008** Coleman, W. D. and Johnson, N. "Building South-North Dialogue on Globalization Research: Report on Proceedings", *Globalization and Autonomy Online Compendium*, [http://globalautonomy.ca/global1/dialogueltem.jsp?index=SN\\_ColemanJohnson\\_MeetingReport.xml](http://globalautonomy.ca/global1/dialogueltem.jsp?index=SN_ColemanJohnson_MeetingReport.xml), 2008.
- Conway 2010** Conway, P. "Preservation in the Age of Google: Digitization, Digital Preservation, and Dilemmas" *The Library Quarterly* 80.1 (2010): 61-79.
- Cooper 2004** Cooper, A. *The Inmates Are Running the Asylum*, Indianapolis, Indiana: SAMS, 2004.
- D'lorio 2010** D'lorio, P. "Scholarly Information Management: A Proposal in Connexions", May 14, 2010, <http://cnx.org/content/m34327/1.2/> 2010.
- Hedstrom 1998** Hedstrom, Margaret. "Digital Preservation: A Time Bomb for Digital Libraries", *Computers and the Humanities*, 31 (1998): 189-202.
- Kirschenbaum 2009** Kirschenbaum, M. G. "Done: Finishing Projects in the Digital Humanities", *Digital Humanities Quarterly*, 3:2, 2009.
- Kretzschmar and Potter 2010** Kretzschmar, W. Jr. and Potter W. "Library collaboration with large digital humanities projects" *Literary and Linguistic Computing* 24.4 (2010): 439-445.
- NRDAC 2001** NRDAC Working Group. *National Research Data Archive Consultation: Phase One: Needs Assessment*, Ottawa: Social Sciences and Humanities Research Council of Canada, 2001.
- OPENDOAR** <http://www.opendoar.org/find.php> Accessed 28 October 2012.
- Rockwell 2011** Rockwell, G., Day, S., Yu, J. and Coleman, W. D. "Globalization Compendium Archive", *Globalization Publications*, Paper 1, [http://digitalcommons.mcmaster.ca/global\\_coll/1](http://digitalcommons.mcmaster.ca/global_coll/1), 2011.
- SHERPA-JULIET 2012** Research Funders Open Access Policies <http://www.sherpa.ac.uk/juliet/index.php>; Accessed 28 October 2012.
- UK Data Archive 2011** "UKDA Preservation Policy", 18 May 2011 <http://www.data-archive.ac.uk/media/54776/ukda062-dps-preservationpolicy.pdf> Accessed 28 October 2012.
- Van den Eyden et al 2010** "Data Management Practices in the Social Sciences", UKDA Report 10 August 2010 [http://www.data-archive.ac.uk/media/203597/datamanagement\\_socialsciences.pdf](http://www.data-archive.ac.uk/media/203597/datamanagement_socialsciences.pdf) Accessed 28 October 2012.