DHQ: Digital Humanities Quarterly

Preview
2023
Volume 17 Number 3

DH as Data: Establishing Greater Access through Sustainability

Alex Kinnaman <alexk93_at_vt_dot_edu>, Virginia Tech https://orcid.org/0000-0001-8943-8946
Corinne Guimont <gcorinne_at_vt_dot_edu>, Virginia Tech https://orcid.org/0000-0001-9364-7708

Abstract

This paper presents methodology and findings from a multi-case study exploring the use of preservation and sustainability measures to increase access to digital humanities (DH) content. Specifically, we seek to develop a workflow to both prepare DH content for preservation while enhancing the accessibility of the project. This work is based on the idea of treating DH as traditional data by applying data curation and digital preservation methods to DH content. Our outcomes are an evaluation of the process and output using qualitative methods, publicly accessible and described project components on two Virginia Tech projects, and a potential workflow that can be applied to future work. By breaking down individual projects into their respective components of content, code, metadata, and documentation and examining each component individually for access and preservation, we can begin migrating our digital scholarship to a sustainable, portable, and accessible existence.

Introduction

Numerous Digital Humanities (DH) projects are developed in institutions and organizations without standards for access and preservation. However, various standards and best practices more closely associated with other fields exist for project management, data curation, access, and preservation, but have not been exclusively adapted for DH. The goal of this paper is to explore the concept of DH as data and the application of data curation methods to two DH projects at Virginia Tech (VT) in order to enhance access and preservation. By breaking down individual projects into their respective components and treating them as data, we can begin preparing digital scholarship for a sustainable, portable, and accessible existence. 1

Challenges in Maintaining DH

At Virginia Tech University Libraries (VTUL) there are currently eighteen DH projects that we currently maintain on a technical level, may potentially maintain, or plan to provide preservation services for project content. These projects are currently hosted on local library servers, private servers we do not have access to, different departmental servers, and two projects that are currently hosted outside of the university. Seven of the projects are built on Omeka, four on Wordpress, and many were built using HTML or other homegrown tools. These projects were created over a period of time ranging from 1996 to 2021 with some having been static and unmaintained since the mid-2000’s. A survey of these projects revealed that seven had some supplementary content (such as news articles, corresponding publications, presentations, and images) in our institutional repository VTechWorks, but only one of the projects had deposited raw data or content into that repository. Further, only four projects mention or document details regarding preservation and sustainability efforts for ongoing access. These projects are fragile with an average lifespan of two to five years without human intervention, and as the University Library we are responsible for the maintenance of the open source resources created by our VT community. In a larger effort to support these projects and other future digital work, we chose to explore methods to preserve DH projects with a strategy that could be reasonably implemented on a larger scale. This strategy is based on preserving components of the DH project for reconstruction of the content in the future and 2 documenting methods to educate our digital scholars to create and modify their projects with preservation in mind.

The problem with developing a reasonable approach to sustainable DH projects is manifold but can be summed up into a single concept. DH is uniquely located in the overlap between the humanities, archives, and information technology, but unless a resource is labeled DH-related, it may be overlooked. There are best practices, and general technical maintenance that are applied in several other fields, but may not cross into DH education and infrastructure [Barats, Schafer, and Fickers 2020]. By identifying these practices and testing their application specifically to DH without compromising the culture and purpose, we can avoid reinventing new ways to curate digital content. 3

For the purposes of this paper, we refer to the Digital Preservation Coalition definition of digital preservation as “the series of managed activities necessary to ensure continued access to digital materials for as long as necessary” [Digital Preservation Coalition]. This is a widely adopted definition utilized by libraries and archives as it allows for varied approaches to preservation based on needs. Although this paper is written in the context of an academic library, it is also a resource for those creating DH projects and sustaining and preserving the digital components of their work. Further, we define a sustainable project as one that can be supported financially, including the resources for storage, ongoing maintenance and monitoring, unique technology needs, and time and effort from personnel. Finally, infrastructure refers to the technical scaffolding and organized governance necessary to host, maintain, and preserve digital projects. 4

Digital Humanities as Data

Given the infrastructural issues discussed above, VTUL is exploring the idea of DH as data. In this case, we describe “data” as all of the digital content pertaining to a DH project or produced in the preservation process. Our approach is to preserve each DH project component as an individual entity with corresponding metadata for the purpose of reconstruction. Project components are distinguished by their file format or software type and enhanced for our purposes with documentation and metadata. 5

The concept of treating content as data was explored by the IMLS project *Collections as Data* [Padilla et al 2020], which sought to address the problem that “cultural heritage institutions have rarely built digital collections or designed access with the aim to support computational use.” This signals a cultural shift that treats non-traditional data, such as text, images, and audio, as data to be computationally explored. While definitions of DH are broad and varied, this concept can be applied to enhance access on small and large scales. The “Ten Principles” [Padilla et al 2020] developed by the *Collections as Data* project are a starting point to guide digital collections into a more engaged environment that promotes multiple forms of access. 6

We applied *Collections as Data* to the idea of DH as data in that it widely varies in tools, techniques, file formats, softwares, and hardwares, to generate, manipulate, and render information. Where DH lacks a community-supported archival best practice for creation and storage, there are definitive database structures built to store and maintain data, metadata standards to improve search and context, and ways to measure the success and use of the data. Rather than reinventing a new workflow for DH as a field, treating components of DH projects as pieces of a dataset serves to provide a best practice for development, access, and storage that increases the long-term preservation and usability of project material. Treating DH as data also serves to support DH as an academic scholarship, which it is not often considered at many institutions, and provides a more familiar structure to further legitimize the labor of developing effective DH work. 7

Similar efforts have been explored at the University of Oxford Sustainable Digital Scholarship Program in their Humanities Division that supports sustainable digital scholarship through the data repository platform Figshare. The Oxford instance of Figshare hosts exclusively humanities content supporting digital scholarship and adheres to the *FAIR Data Principles*. The *FAIR Principles for Data* [Wilkinson et al 2016] are guidelines for increasing the *findability*, *accessibility*, *interoperability*, *reusability* of data and its metadata, and the system’s infrastructure. This method applies a license to the data and increases the discoverability of the data through proposed metadata, persistent and unique identifiers, and open access. These basic discoverability and preservation parameters apply to all digital content and are already used as best practices. 8

Instruction on this topic is emerging as well. Outside of courses in formalized university programs, organizations like the Digital Humanities Summer Institute Technologies East 2021 hosted a class in May 2021 on “Making Research Data Public: Workshopping Data Curation for Digital Humanities Projects”, citing in its abstract that “a lack of formal training opportunities for data curation in multi-site DH teams means that the data produced in these teams is in danger of being lost.” This is a single example among many DH organizations moving towards technology education to enhance projects. 9

Many grant funders also require data management plans, sustainability plans, technical plans, and other relevant documentation in grant proposals. The National Endowment for the Humanities Office of Digital Humanities, for example, provides tips on writing final white papers that includes an entire section on outcomes and obstacles in technology, documentation, and platforms, and requires a section on Managing and Sustaining the Project Assets in level three DH Advancement Grants [Serventi 2019]. 10

Methodology

We used a multi-method approach, with a comprehensive literature review on the community’s practice in developing sustainable DH projects, as well as two case studies conducted over a year-long period. 11

Literature Review

Acknowledging that the rapid increase of digital scholarship also requires new methods of preservation and sustainability for long-term access has been an ongoing topic of discussion since the early days of DH. Discussion on best practices for preserving and curating DH scholarship became more common in the mid-2000s with the impending issue of preserving complex digital objects. In their article, Hunter and Choudhury 2003“focus on the problem of preserving composite digital objects consisting of multiple media types” and introduce the concept of emulation, migration, and metadata as it pertains to complex digital objects, which includes DH. They state, “As scholars in the humanities create increasingly sophisticated multimedia research and teaching resources, they need to capture, collect, and create the documentation or metadata – descriptive, administrative, and structural – necessary to migrate, emulate, or otherwise translate existing resources to future hardware and software configurations” [Hunter and Choudhury 2003]. Cantara (2006) performed an early literature review of digital humanities metadata best practices and tools to automate that metadata, which is significantly different from traditional cataloging [Cantara 2006]. 12

The bulk of individual digital objects being preserved are simple; images, text files, audio/visual materials, all of which we have established best practices for ensuring continuity. We have the Library of Congress Recommended Formats Statement to transform objects into the optimal format for long-term access. We use checksums to confirm fixity and ensure objects are unchanged and uncorrupted. We use audit logs like the PREMIS Data Dictionary for Preservation Metadata to document stages in an object's lifestyle, such as ingest, deletion, restoration, fixity change, virus checking, etc., and then use tools like the Metadata Encoding and Transmission Standard to encode the metadata for broader understandability and interoperability with common systems. We maintain copies, mirror the information technology method for redundancy. All of these practices are guided by digital preservation auditing metrics, most notably the ISO 16363:2012 Space data and information transfer systems — Audit and certification of trustworthy digital repositories (TDR) which also provides guidance on governance, security, and financial sustainability. Good and best practices for common digital objects are straightforward and have well-established guidelines, but when combined into a complex digital object like a website, interactive resource, or software, the basic digital preservation principle of ongoing preservation becomes increasingly complicated. 13

Another problem at large in hosting digital projects lies within the general scalability and sustainability of the services offered. Conway (2010) explored the concepts of digitizing for preservation and digital preservation and the transition from traditional preservation to digital preservations, leading to dilemmas in what objects are treated differently and how they are treated in a digital environment. Further, Vinopal and McCormick (2013) state that “these services should promote the development of reusable tools, platforms, and methods, and facilitate the creation of preservable, reusable scholarly content to ensure the long-term value of and access to the institution's research.” Their article provides an 14 overview of various researcher needs and the relationship between the services and Libraries of a given institution and provides a Proposed Model for Digital Scholarship Services that includes DH scholarship and how DH programs fit into other institutional services. Leslie Johnston of the National Archives and Records Administration expresses similar thoughts on strategies in preserving DH specifically, stating “sustainability and preservation depend on active management of a project. Digital humanities projects correctly put the highest focus on the content and scholarship that they present. But these projects should not let their technologies go stale” (2013) by relying on technologies and plugins that no longer receive maintenance or community support.

Project failure can occur in many ways, from lack of funding to complete, inaccessible data, little to no preservation or access storage, and lack of technical training, among others. While a project can fail for a number of reasons, we searched for examples of DH projects that had failed due to lack of preservation or sustainability planning. Surprisingly, there is little record of specific projects that have been discontinued or failed as they are no longer accessible. A study conducted by Timothy Vines and his team surveyed 516 articles published over the course of ten years and found that after the first two years of publication, access to scientific data falls by seventeen percent every year [Vines et al 2013]. Their summary was that “broken emails and obsolete storage devices were the main obstacles to data sharing” [Vines et al 2013]. This study examined accessibility and sharing and can be applied to other forms of digital content in that broken links and obsolete storage are obstacles to digital maintenance. In terms of DH, there are some but few examples of failed projects. James Cummings, for example, provided a list of projects he personally supported that had ultimately failed, five of six of which included lack of sustainability planning, funding, or limited technology in their reasons for failure (2020). Jasmine Kirby also discusses the failure of the Sophie 2.0 project at the University of Southern California, an international collaboration with over a million dollars in funding, noting “the focus was on building the tool with the most features and not something that actually could be sustained or even worked” (2019). Even a 2013 JISC study surveyed DH Center Directors regarding long-term maintenance of hosted projects and found difficulty in keeping these sites live [Maron and Pickle 2013]. The study suggests that centers can turn to museums and libraries for long-term hosting and access, but also notes that “many have yet to develop a formal digital strategy” (2013). 15

In contrast to unsuccessful projects, many long-standing DH projects share specific aspects that led to success. The accessibility of the raw data as a component of the project regardless of the platform and the ability to explore it independently appears to be the key to the success of these long-standing projects. These projects all ensure the data is platform-agnostic and manipulable in other platforms with multiple tools and exists in formats that are interoperable. Many of these projects also contain corresponding documentation to describe the data as another component of the work. For example, the Walt Whitman Archive is built on TEI/XML and simple HTML. The interface has not been updated in some time, however, the success of this project is based on the range and depth of the content, and the accessibility of the raw data. All of the manuscripts, correspondence, and most other content is available for direct download from the Git repository in their original TEI/XML markup. Similarly, the William Blake Archive provides a short API code to query the entire archive independently. Finally, the Charles Algernon Swineburne Project provides basic TEI/XML and XSLT files to explore and recreate the archive. 16

Comparing these successful and unsuccessful projects prompted us to review other guides on sustainable websites in beginning our case studies. There are several guides on creating sustainable websites available through the Library of Congress’ “Creating Preservable Websites” page, the blog post “On Preserving Digital Culture and Digital Projects” and “Digital Humanities for Tomorrow: Opening the Conversation about DH Project Preservation” (2016) from Digital Humanities at Berkeley, and Stanford Libraries’ “Archivability” guide (Stanford Libraries). This guidance provides some foundational best practices and again acknowledges the need for sustainable and preservable websites, but does not provide a comprehensive solution. Additional guides that are more thorough and include practical information such as the Socio-Technical Sustainability Roadmap at the University of Pittsburgh and the Digital Project Preservation Plan: A Guide for Preserving Digital Humanities [Miller 2019] have been recently developed. However, Kilbride (2015) notes that while the standards have been created to fill the gaps in digital preservation as it pertains to DH, these best practices have not been widely adopted. 17

The literature review of resources, guides, and projects indicate that increasing access to DH scholarship is a mix of infrastructural and developing best practices, as well as providing education for scholars to learn to adopt best 18 practices. The review also supports the theory of supporting the components of projects for long-term preservation and access in regards to common preservation practices and what has allowed long-standing DH projects to thrive.

Case Studies at Virginia Tech

Virginia Tech is home to dozens of DH projects ranging in legacy to recent, hundreds to thousands of items, and on a variety of platforms. Recently the Libraries have received requests from faculty and administration to develop sustainability plans for these programs and to promote greater use through easier access to data, formal analytics and branding, and advertising of these projects. The Libraries approached this request by experimenting with the concept of DH as data with comprehensive exports of project content, the creation of necessary metadata, and the creation of project documentation into a single data set. 19

There are several possible repositories and storage locations for this content. Projects that contain humanities content with additional scholarship and fit the criteria for the institutional repository, VTechWorks. While projects that contain mapping datasets fit the criteria for the Virginia Tech Data Repository. Additionally, any project that is a general product of Virginia Tech research could be ingested into our newer platform the Virginia Tech Digital Libraries Platform. For the both of our case studies, we found them most suitable for the Virginia Tech Data Repository as the final preservation datasets contained a layered file system with a mix of documentation and smaller datasets. 20

We explored two case studies in the last year. Social Networks in Georgian Britain (SNiGB), a new project branching off of a long-standing project, Lord Byron and His Times (LBT), to enhance the exploration of the content containing over 40,000 items. And the second project, Redlining Virginia, is an Omeka-based project with approximately 110 items that is represented in multiple instances outside of the Omeka site. 21

Social Networks in Georgian Britain

SNiGB originated in 2008 when it began as the prosopography for LBT, an archive of printed letters and diaries by and about Byron’s associates encoded in TEI-XML and rendered as the website https://lordbyron.org/ using XSLT transformations and XPath queries. These records, used to generate footnotes, included links to name authority files such as the LOC, National Archive (UK), National Portrait Gallery (UK), Oxford Dictionary of National Biography (ODNB), and the Virtual International Authority File (VIAF). In 2014, the LBT prosopography became an XML relational database with the addition of a demographic file containing fields for place of origin and professions, with linked information about family members, educational institutions, and correspondents. Fields were hand-coded using TEI elements and included place of origin, professions, and correspondents. 22

From 2008 to 2012, the LBT website was hosted by Performant Software in Charlottesville, and sponsored by Nineteenth Century Scholarship Online (NINES); in 2012 web hosting was transferred to VTUL. In 2017, work in LBT shifted from document markup to development of the prosopography exclusively, and Professor David Radcliffe began discussions with VTUL about data design and publication. With the 20,000 names originally collected from the letters and diaries, Dr. Radcliffe has since added an additional 25,000+ names collected from membership lists of learned societies, social and benevolent clubs, and political organizations. 23

In the winter of 2018, as data collection and description continued, four VTUL personnel and additional student workers began structuring a work plan to transform the TEI-XML into a freely accessible database with a publicly available, interactive user interface. In recent years, VTUL has increased support for faculty and student digital projects, including website and database hosting, consultations, and support for new technologies. However, the transformation and migration of a project at this scale has yet to be tested. The final version of SNiGB content comprising over 40,000 names has been completed by Dr. Radcliffe, and in collaboration with him, this new content has been exported into individual XML files by person and organization, bagged, and ingested into the Virginia Tech Data Repository as a new access point for the raw data which can be accessed at https://doi.org/10.7294/14849748 and have been made available in an HTML index at https://snigb.vt.domains/persNames_index.html. The development of an interactive prosopography site is forthcoming, and this currently serves as the exclusive access point to the entire network of names in its final format. 24

As a longer-standing DH project, SNiGB is reflective of similar projects mentioned in the Literature Review, such as the Walt Whitman Archive, and of the other projects VTUL may need to manage in the future. It is built on standard TEI-XML with a simple HTML/CSS interface that has maintained existence partially because of its simplicity. The maintenance issues that originally prompted the restructure of this project stem from all of the records being encoded in a single TEI document that could not be validated. 25

Redlining Virginia

The second case study is a practical application of our initial strategy on the Virginia Tech project Redlining Virginia led by history professor LaDale Winling. Redlining Virginia “explores the actions of Home Owners' Loan Corporation (HOLC) in Virginia and its impact on Virginia cities.” It is a child project of the larger “Mapping Inequality: Redlining in New Deal America” that specifically explores the Virginia and West Virginia eastern regions. This digital project consists of images, contextual text, videos, and georectified maps. It was also featured as a physical exhibit in Newman Library at Virginia Tech from December 7, 2016 - February 17, 2017, which broadened the content to images of the exhibit and post-it notes left by visitors to the exhibit. 26

This DH project was chosen for several reasons. It is much newer and less long-lived than SNiGB and therefore much smaller, approximately 110 items in the original website, but also contains a wider variety of file types for experimentation. It is hosted on Omeka where items and their metadata could be exported. Packages containing the georectified maps were already uploaded to the Virginia Tech Data Repository by Professor Winling, but the corresponding text is not replicated elsewhere. The digital exhibit hosted at VT contained post-it notes with feedback from visitors and were also uploaded to the institutional repository, VTechWorks. These factors indicated that it would be reasonable to assemble all of the project components into a single package and create additional documentation to recreate or reconstruct the website. 27

Components for Redlining Virginia spanned several locations between both Virginia Tech repositories, Omeka, the parent project Mapping Inequality, and Professor Winling’s personal machine. We also created several new pieces of documentation, including a preservation profile, a collection-level Metadata Application Profile, additional metadata in DublinCore for the images of the physical exhibit, and additional metadata in the Content Standard for Digital Geospatial Metadata for the georectified maps. To replicate the website, we exported the HTML and CSS, as well as exporting each page of text to a PDF and included several screenshots to preserve the original visual style and organization. We also prepared a user guide with details on each digital object, information on how to read and write content in open access software, and general project information. This package can be viewed in the Virginia Tech Data Repository at https://doi.org/10.7294/14597751.v1. The original website is static and live, and this additional access point provides all individual project components to a user. 28

Given that VTUL has at least seven Omeka-based projects from students and faculty that may require support, this aligns with our current field of practice. We chose this project because we know there are other Omeka based projects we will need to preserve; however, we ultimately want to serve other platforms and projects as well. Omeka is also a popular tool for DH projects [Rath 2016], as well as a tool used in other digital scholarship projects [Cobourn 2016] and therefore relatable to a wide variety of practicing digital humanists. 29

Results

These results include both the impact of the case studies on our work, as well as the key themes that arose over the year of research and implementing the case studies. 30

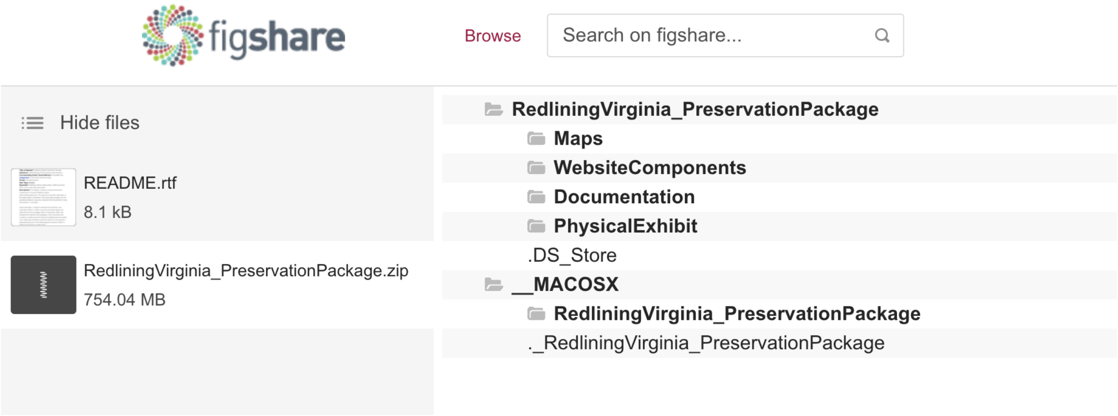
In terms of these case studies, we define success as first, preparing the data for long-term storage which included migrating formats, documenting workflows for open source software needed to open each format, and creating collection-level metadata; and second, depositing the resulting package of the original project objects and new documentation into a repository as an access point and into a third-party digital preservation service for preservation storage. 31

While our case studies were successful in practice, they did require a lot of bandwidth from us that is difficult to repeat at a larger scale for all DH projects at VT. We found two key themes, in our context at VTUL, that are the primary obstacles to increasing the access, usability, and preservation of DH scholarship: a lack of education on best practices from our DH creators, and a lack of sustainable infrastructure development to support DH projects in university libraries. 32

Case Studies

We employed data curation methods that resulted in described datasets and content packages consolidated to be portable, explorable, and reused. The two case studies we have conducted so far have produced another access point for the raw data, content, and code, as well as provided another method of tracking impact through the Virginia Tech Data Repository (VTDR) when the Libraries do not have the ownership to apply analytics to a live project. This accomplishes both goals of providing a permanent access location for data regardless of whether the project as a whole was successful (an example of Redlining Virginia is in Figure 1), and of providing a preservation environment for DH data. Impact can also be measured through the number of views and downloads shown in the VTDR interface. 33

VTDR is hosted on Figshare, which is backed up automatically to the digital preservation service Chronopolis and VTDR manually deposits all datasets into the Academic Preservation Trust digital preservation service of which VTUL are members. We successfully ingested each project's data into a stable storage location with enough context to reconstruct the entire project in its integrity and to explore individual components. This was accomplished by organizing each individual component into its own folder with associated metadata packaged into a single folder with collection-level documentation and guides on where and how to locate additional resources (an example of Redlining Virginia file structure is in Figure 2, Figure 3, Figure 4, Figure 5, and Figure 6). 34

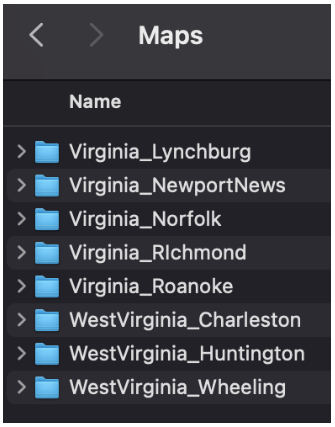


README.rtf
8.1 kB

RedliningVirginia_PreservationPackage.zip
754.04 MB

- RedliningVirginia_PreservationPackage
  - Maps
  - WebsiteComponents
  - Documentation
  - PhysicalExhibit
  - .DS_Store
- __MACOSX
  - RedliningVirginia_PreservationPackage
  - ._RedliningVirginia_PreservationPackage

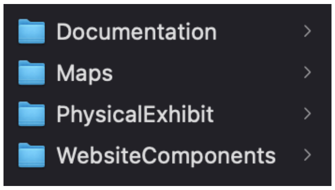
Figure 1. Screenshot of Redlining Virginia data in the Virginia Tech Data Repository hosted in Figshare


Maps

Name

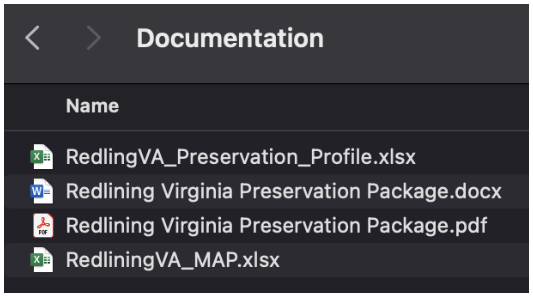
- Virginia_Lynchburg
- Virginia_NewportNews
- Virginia_Norfolk
- Virginia_RIchmond
- Virginia_Roanoke
- WestVirginia_Charleston
- WestVirginia_Huntington
- WestVirginia_Wheeling

Figure 2. Screenshot of the top level folder



Documentation
Maps
PhysicalExhibit
WebsiteComponents

Figure 3. Screenshot of the collection-level documentation

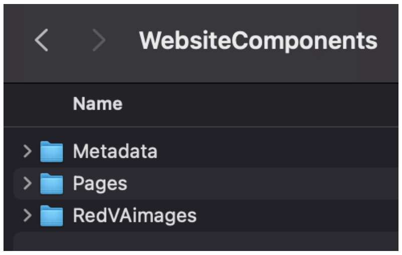


Documentation

Name

RedlingVA_Preservation_Profile.xlsx
Redlining Virginia Preservation Package.docx
Redlining Virginia Preservation Package.pdf
RedliningVA_MAP.xlsx

Figure 4. Screenshot of each map that contains georectified maps and associated metadata

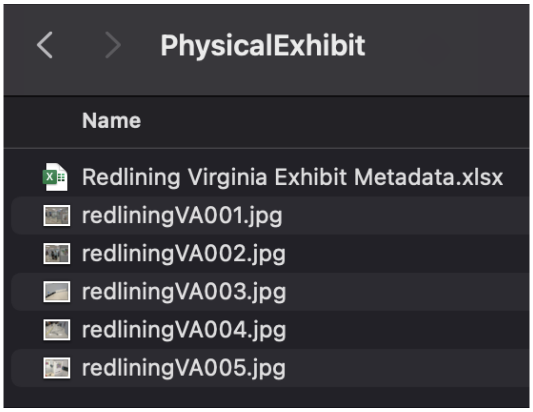


WebsiteComponents

Name

Metadata
Pages
RedVAimages

Figure 5. Screenshot of the images from the associated physical exhibit and its metadata



PhysicalExhibit

Name

Redlining Virginia Exhibit Metadata.xlsx
redliningVA001.jpg
redliningVA002.jpg
redliningVA003.jpg
redliningVA004.jpg
redliningVA005.jpg

Figure 6. Screenshot of the collected website components and associated metadata

Metadata development for SNiGB was straightforward as the raw data is a series of small XML files stored in zipped folders. Redlining Virginia contained georectified maps displaying the evolution of redlining over time. These maps, a combination of JPG, TIF, XML, TFW, OVR, and Shapefile bundles, required more research into appropriate metadata schemes. The final schema chosen was the Content Standard for Digital Geospatial Metadata modified for our georectified maps. From a technical standpoint, researching and developing this metadata was the single most time-consuming phase of this case study. We applied a newly developed Digital Exhibits Metadata Application Profile designed by the Metadata Coordinator and a Preservation Profile designed by the Digital Preservation Coordinator and Digital Preservation Technologist that will also be applied to additional DH projects. Applications of both of these Profiles can be viewed in the Virginia Tech Data Repository preservation packages for Redlining Virginia.

One limitation we did not explore is discoverability. The new documentation developed for these case studies is not discoverable in and of itself and is only found through the data repository via the associated title, keywords, and descriptions. While we established another location for access, we did not have any autonomy in the metadata development to be more efficient in search engine capabilities. 36

This practice is imperfect. We can only enhance and preserve what we are given or what can be found and exported. From these case studies, it is clear that defining a minimal standard of what is required from the project and researchers, and what is reasonable and possible for the Library to manage, is necessary. 37

Time and bandwidth were also challenges in these case studies. Two library faculty members dedicated over forty hours of time each to develop and implement workflows for each project. Given the dozens of projects that will require similar efforts in preservation and without considering additional enhancement and large-scale migration, this is not a reasonable task. The immediate outcome from these two case studies was increased evidence and advocacy in VTUL for a permanent support system for digital scholarship in the form of a standing committee . With sponsorship from the Associate Dean and Information Technology Executive Director and support from Library stakeholders and department heads, the newly established Digital Scholarship Technical Review Committee is designed to serve as an advisory and implementation body for both legacy and new Digital Scholarship and web-based projects hosted by VTUL. This committee is currently developing and will implement workflows, policies, and technical infrastructure for hosting digital scholarship, and will also include the appraisal and prioritization of migrating and developing projects. The committee will determine what qualifies as a digital scholarship project that the library is prepared to support, create migration procedures, identify responsible individuals, and implement workflows to transfer projects to the Libraries. This committee will also develop documentation and guidance for creating new sustainable digital projects. 38

Education and Best Practices

The first qualitative result we found is that accessible and sustainable DH suffers from a lack of education for researchers and a lack of standards and best practices that are usable by novices. Kilbride (2015) suitably remarks: “Consequently there are two abiding risks for those humanities scholars trying to ensure a long-term future for their precious digital cargo: that the emerging practices of digital preservation are a poor fit to the changing needs of humanities; and that jargon and miscommunication confound thoughtful efforts to engage.” The overlap of humanities critical analysis skills and the technical skills to implement a successful and usable project is difficult to attain, particularly for projects hoping to continue beyond a few years. Foundational DH courses in the undergraduate and graduate programs tend to focus on the multiple definitions of DH, tools and techniques, and impact of providing unique public access to otherwise unavailable information. Sustainability, usability, and project management may be introduced, yet the number of DH projects created early in a researcher’s profession that eventually break or disappear is high. 39

What DH standards and best practices do we introduce to students and new professionals? A traditional example is Nineteenth Century Scholarship Online which provides a formal, peer-reviewed system for evaluating projects, but is limited to projects concentrating on 19th century humanities. Other similar resources are equally contextual and not universally applicable, such as the University of Nebraska-Lincoln’s “Best Practices for Digital Humanities Projects” guide that offers technical best practices, but does not necessarily provide information on documentation or preservation infrastructure. 40

Within the DH field it is often necessary to learn at least one or multiple technologies from digitization, processing and quality control, metadata, website development, text mining, and so on to create successful digital scholarship. With our two case studies, as well as many other projects at Virginia Tech, researchers are largely self-taught and use popular platforms like Omeka to execute their work. Sometimes these projects are completed without direction or collaboration with the library and are later brought to VTUL for maintenance and preservation. In the case of older projects, many of these were built before VTUL had developed a preservation program, meaning that even those with the best intentions may have received little direction in the area of preservation. Humanists are not trained in information technology, 41 digitization standards, metadata standards, content management systems, storage management, and digital preservation, because that is not generally required in the field. There is a balance between what technology is community-supported, what is simple to adopt, and what is sustainable. What are considered foundational DH skills may not incorporate IT best practices, digitization guidelines, or project management. The problem is paradoxical - what services can a library provide to its faculty and students when there are no defined best practices to teach? The lack of consistency between institutions and initiatives in their approach to DH is partially what makes DH so dynamic, but also causes formalization to be even more difficult than in a more established field of study.

Infrastructure

The second result we found was an inconsistent approach to project building, indicating a lack of DH project infrastructure available to DH-ers. Like many other digital scholarship projects, website creation, and digital resources, DH projects are typically not designed to be sustainable and preservable. DH has grappled with insecure and unsupported technology because many creators are drawn to trendy technology that may not be sustainable [Owens 2014]. DH is often driven by new or easy-to-use technology that may not have a built-in infrastructure yet, when the goal of the project should drive the technology choices. Butler et al explored this integral infrastructure issue in 2019 in the two-part blog series “Archiving DH”. Taking a new technology and applying it to brand new research is a unique benefit for exploring options, but the lack of infrastructure leaves these projects unsustainable over time [Butler, Visconti, and Work 2019]. In “Archiving DH Part II: The problem in detail”, Butler et al also discuss what defines “authentic access and use over time for digital scholarship” as websites cannot last forever without human intervention and maintenance. The authors discuss what the ideal point in the life cycle of scholarship is to introduce “preservation, sustainability, and portability” [Butler, Visconti, and Work 2019]. The piece on portability is highly relevant to the existing DH projects at VTUL as the Libraries receive many requests to transfer ownership of projects, and in some cases, this can mean transferring the platform of the project. 42

As a culture, DH is open access in that many of the projects are openly available online. Additionally, funding from sources like NEH require grant awardees to produce open access projects [Brennan 2020], while publishers like the International Journal for Digital Humanities are exclusively open access, and more often than not, DH is designed to highlight new content and research that only benefit from open distribution. However, while a project and its corresponding data may be openly available, if the code, algorithms, documentation, and metadata are not, how impactful and usable is the project in the long-term? Preservation is tricky, you cannot just “preserve it” because the essence of what matters about “it” is something that is contextually dependent on the way of being and seeing in the world that you have decided to textual components alone are not enough to accurately enhance, reproduce, or preserve a DH project. 43

There are also the typical digital object collection problems to consider regarding storage, access platform, and responsibility for maintenance. For traditional scholarship we have journal databases and institutional repositories, and for data we have databases and data repositories, but a best practice for storing and maintaining DH projects, which is often a combination of scholarship and data, is still evolving. The choice we made to store our DH preservation packages in the Virginia Tech Data Repository was based on the best choice from the resources we had available, not because it’s a perfect DH storage solution. 44

At VTUL we want to provide access to a preservation package that contains the text, context, code, metadata, and infrastructure to reproduce and reuse. This preservation package is designed to provide all of the components of a project and guidance on how to reuse, rebuild, and validate DH projects and data. In terms of digital preservation specifically, in the Open Archival Information System (OAIS) Reference Model, the Submission Information Package would be created by VTUL, the Archival Information Package is stored in a preservation system (Virginia Tech Data Repository), and the Dissemination Information Package is hosted in a publicly accessible location (the live website). In many Virginia Tech DH projects, and many DH projects in general, the main deliverable is equivalent to the Dissemination Information Package with little regard to raw data or preservation and sustainability from the lack of technical infrastructure. 45

Solution and Suggestions

On a more universal level, these case studies have indicated the need for established, straightforward models for researchers to understand and implement. The Data Information Literacy Project from the Institute of Libraries and Museums provides a similar framework for a standard training package for DH. The project provides a framework for educating graduate students and researchers on following best practices in data management with their research. A similar approach could be taken for creating sustainable and preservable DH projects. Based on our experience attempting to develop a reasonable and scalable workflow for preserving our DH projects, we recommend the following priorities to others both creating and curating DH projects: documentation development, openness and raw data sharing, and implementing sustainable storage and digital preservation practices. 46

Our case studies required significant documentation to be created. This included specialized metadata, VTUL-established metadata, lists of file formats and open access tools and softwares to open them, and some additional content conversion for better portability and preservation. Establishing a method of documentation and examples of said documentation can increase usability from a wider audience. DH-ers creating new projects can begin documenting during project planning and design and develop content that is technology-agnostic and providing guidance on how to use and manipulate data and code increases the range of users that can access a DH project. Curators can work with DH-ers managing existing projects to create this documentation in a way that is most suitable for their access and/or preservation platform and for their infrastructure to migrate hosting to the Libraries. 47

The ultimate goal of preservation is access. DH creators can share raw data and documentation to increase transparency and further research validation. By providing additional contextual information, raw data, code, and more, other users can fully access the many pieces of a DH project. Additionally, while not every project has the ability to support data mining, visualization software, or other methods of data exploration, providing access to other features such as unique code, exportable metadata, and plaintext files can increase usability of a project even if it’s not in its final format. This is similar to our approach with SNiGB, for which the preservation package is a series of XML files that does not yet have a platform for visualization or mining. Curators can take this approach to assist current projects to create more usable files for access. 48

Furthermore, we recommend DH researchers focus on project planning and unique or specific project needs as early in the process as possible. Deciding what standards to use or best practices to follow after the project has begun can muddle progress and force researchers to backtrack in order to adopt a different strategy or undergo deduplication. Since there are few well-established standards specifically for DH as a field that have also been adopted by the community, choosing best practices that support individual project components can aid in developing an entire infrastructure of components that are supported, sustainable, and documented. As an example, in our case study with the Redlining Virginia preservation package, we needed to identify a metadata standard that aligned with the content, specifically for the georectified maps and their Shapefile bundles. We chose to find and use an existing schema rather than attempting to create a brand new metadata schema. While creators should be documenting as they develop, this is not a usual experience and the effort may fall largely to the curator for technical documentation development. 49

Finally, technical infrastructure, storage, and preservation must be prioritized by DH-ers creating projects. Identifying a pre-established location for storage, such as an institutional repository or an appropriate public repository and ensuring the data fits the standard for that repository early on ensures a form of access and direct line to preservation copies and activities without the previously mentioned challenge of revising content or metadata after development to conform to a repository. Preservation activities ensure long term use through migration into preservation file formats that are well-supported, non-proprietary, and likely able to be migratable in the future. Digital preservation supports records of provenance articulating if data has changed or needs to be migrated or updated. It also monitors the lifespan and viability of the technology being used. Preservation-centric projects ensure that the project will be accessible over time and can be migrated or updated more smoothly. Reaching out to a university library or comparable institutional body to see what options are available for hosting and storage early in the process will help DH-ers create more sustainable projects and provide curators with more information to successfully maintain projects. 50

Despite the success of creating a new iteration of the VT case study projects, the task of managing dozens of projects with the same methodology is not a sustainable process for two people to manage. Successful DH programs require dedicated and intentional infrastructure supporting the needs of researchers. Mutual responsibility for the Library to be transparent about what tools are supported and for researchers to choose a tool that can be supported in a location that can be easily accessed. However, many of our strategies can be duplicated at different levels depending on the resources available. 51

Conclusion

This paper explored the case studies performed on two Virginia Tech DH projects at varying stages of complexity and age for the purposes of access and preservation. These case studies aimed to compile project components, develop additional description and documentation, and prepare content for long-term access. The findings shaped specific interventions utilized by VTUL and will be applied to additional DH projecfts. The result was the appraisal of two DH projects for preservation and access needs and the compilation of preservation-prepared datasets and corresponding documentation stored and accessible through the Virginia Tech Data Repository. The other significant result was the improved advocacy and support for our current and legacy DH projects as we work to maintain and enhance VT digital scholarship through a newly established Library committee. 52

Our current goal is to ensure that our DH projects at VTUL can be reconstructed or recreated in the event that access to the original format is lost. In the preservation field, emulation is the optimal form of preservation for access. As organizations like SPN are already investigating the preservation and future accessibility of software and as we are currently employing web archiving through Archive-It, our ideal goal would be to ensure the usability of sites that cannot be migrated without losing integrity can be emulated and reused. Web archiving can preserve the look and feel of the project but may not be able to handle complex data manipulation software in a way that is meaningful to a user, thus the need to think of the components of DH projects and what we can do to preserve those as well as the look and feel. 53

Works Cited

Ammon 2019 Ammon, S. (2019) “Archiving DH part 1: The Problem. University of Virginia Scholars Lab”. Available at: https://scholarslab.lib.virginia.edu/blog/archiving-dh-part-one.

Barats, Schafer, and Fickers 2020 Barats, C., Schafer, V. and Fickers, A. (2020) “Fading Away... The Challenge of Sustainability in Digital Studies”, *Digital Humanities Quarterly*, 14(3). Available at: http://www.digitalhumanities.org/dhq/vol/14/3/000484/000484.html.

Brennan 2020 Brennan, S. (2020) *Planning Your Next DHAG 1: Idea, Audience, Innovation, Context*. Available at: https://www.neh.gov/blog/planning-your-next-dhag-1-idea-audience-innovation-context.

Butler, Visconti, and Work 2019 Butler, B., Visconti, A. and Work, L. (2019) “Archiving DH Part 2: The Problem in Detail”, University of Virginia Libraries Scholars Lab. Available at: https://scholarslab.lib.virginia.edu/blog/archiving-dh-part-2-the-problem-in-detail.

Cantara 2006 Cantara, L. (2006) “Long Term Preservation of Digital Humanities Scholarship”, *OCLC Systems & Services: International Digital Library Perspectives*, 22(1), pp. 38–42. Available at: https://doi.org/10.1108/10650750610640793.

Cobourn 2016 Cobourn, A. (2016) “Spreading Awareness of Digital Preservation and Copyright Via Omeka-Based Projects”, *Journal of Interactive Technology and Pedagogy*. Available at: https://jitp.commons.gc.cuny.edu/spreading-awareness-of-digital-preservation-and-copyright-via-omeka-based-projects/.

Conway 2010 Conway, P. (2010) “Preservation in the Age of Google: Digitization, Digital Preservation, and Dilemmas”, *The Library Quarterly*, 80(1), pp. 61–79. Available at: https://doi.org/10.1086/648463.

Cummings 2020 Cummings, J. (2020) “Learning how to fail better: Resilience in Digital Humanities projects”, in *Proceedings of the Centre for Data, Culture, and Society*, University of Edinburgh. Available at: https://slides.com/jamescummings/cdcs2020/.

Digital Preservation Coalition Digital Preservation Coalition. (2023) *Digital Preservation Handbook, Glossary*. Available at: https://www.dpconline.org/handbook/glossary.

Hunter and Choudhury 2003 Hunter, J. and Choudhury, S. (2003) “Implementing preservation strategies for complex multimedia objects”, in T. Koch and I.T. Sølvberg (eds) *Research and Advanced Technology for Digital Libraries*, ECDL 2003: Lecture Notes in Computer Science, 2769. Berlin, Heidelberg: Springer, pp. 473–486. Available at: https://doi.org/10.1007/978-3-540-45175-4_43.

Johnston 2013 Johnston, L. (2013) “Digital Humanities and Digital Preservation”, in *The Signal, Library of Congress*. Available at: https://blogs.loc.gov/thesignal/2013/04/digital-humanities-and-digital-preservation/.

Kilbride 2015 Kilbride, W. (2015) “Saving the Bits: Digital Humanities Forever?”, in S. Schreibman, R. Siemens, and J. Unsworth (eds) *A New Companion to Digital Humanities*. John Wiley & Sons, Ltd, pp. 408–419. Available at: https://doi.org/10.1002/9781118680605.ch28.

Kirby 2019 Kirby, J.S. (2019) “How NOT to Create a Digital Media Scholarship Platform: The History of the Sophie 2.0 Project”, *IASSIST Quarterly*, 42(4), pp. 1–16. Available at: https://doi.org/10.29173/iq926.

Library of Congress Library of Congress (no date) “Creating Preservable Websites”. Available at: https://www.loc.gov/programs/web-archiving/for-site-owners/creating-preservable-websites/.

Maron and Pickle 2013 Maron, N. and Pickle, S. (2013) “Sustaining our Digital Future: Institutional Strategies for Digital Content”, in *Ithaka S+R*. Available at: http://sr.ithaka.org?p=22547.

Miller 2019 Miller, A. (2019) “Digital Project Preservation Plan: A Guide for Preserving Digital Humanities / Scholarship Projects”, in *Middle Tennessee State University Digital Scholarship Initiatives Publications*. Available at: http://jewlscholar.mtsu.edu/xmlui/handle/mtsu/5761.

Owens 2014 Owens, T. (2014) “Digital preservation’s place in the future of the digital humanities”, in *Trevor Owens Personal Blog*. Available at: http://www.trevorowens.org/2014/03/digital-preservations-place-in-the-future-of-the-digital-humanities.

Padilla et al 2020 Padilla, T. et al. (2020) “Always Already Computational: Collections as Data”, in *Final report to Institute of Museum and Library Services* (LG-73-16-0096-16). Available at: https://doi.org/10.17605/OSF.IO/MX6UK.

Rath 2016 Rath, L. (2016) “Omeka.net as a Librarian-Led Digital Humanities Meeting Place”, *New Library World*, 117(3), pp. 158–172. Available at: https://doi.org/10.1108/NLW-09-2015-0070.

Ross 2012 Ross, S. (2012) “Digital Preservation, Archival Science and Methodological Foundations for Digital Humanities”, *New Review of Information Networking*, 17(1), pp. 43–68. Available at: https://doi.org/10.1080/13614576.2012.679446.

Samberg and Reardon 2016 Samberg, R.G. and Reardon, S. (2016) “Digital Humanities for Tomorrow: Open the Conversation about DH Project Preservation”. *Digital Humanities at Berkely Blog*.

Serventi 2019 Serventi, J. (2019) *Planning your next DHAG 3: Managing and Sustaining the Project Assets*. Available at: https://www.neh.gov/blog/planning-your-next-dhag-3-managing-and-sustaining-project-assets.

Stanford University Libraries Stanford University Libraries. (no date) “Archivability”. Available at: https://library.stanford.edu/projects/web-archiving/archivability.

Vines et al 2013 Vines, T.H., et al (2013) “Mandated Data Archiving Greatly Improves Access to Research Data”. “The FASEB Journal”, 27, pp.1304–1308.

Vinopal and McCormick 2013 Vinopal, J. and McCormick, M. (2013) “Supporting Digital Scholarship in Research Libraries: Scalability and Sustainability”, *Journal of Library Administration*, 53(1), pp. 27-42.

Wilkinson et al 2016 Wilkinson, M. et al (2016) “The FAIR Guiding Principles for Scientific Data Management and Stewardship”, *Scientific Data*, 3. Available at: https://doi.org/10.1038/sdata.2016.18.

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