

National Digital Stewardship Residency | New York

Final Report

Genevieve Havemeyer-King

Host: Wildlife Conservation Society

***Piloting Workflows and Systems for Long-term Preservation of Born-digital
Content from the Wildlife Conservation Society's Education, Exhibit, and
Geospatial Analysis Departments***

Table of Contents

Description of the Project

Overview

Project Partners

Project Execution

Phase 1

Phase 2

Phase 3

Outreach and Dissemination Activities

Phase 4: Documentation & Reporting

NDSR Outreach

Analysis and Evaluation

Project Results

Standards Compliance

What Can & Can't Be Processed

Accepted Media Formats

Challenging Formats

Other Data Challenges

Final Ingest of Sample Collections

Metadata

Outcomes & Impact

Next Steps

Professional Development

Overview

Activities

Conferences, Workshops, and Memberships

Description of the Project

Overview

The project carried out with The National Digital Stewardship Residency (NDSR) Program at the Wildlife Conservation Society (WCS) aimed to address the needs of the WCS Library and Archives' growing born-digital collections by establishing a pilot digital archiving system based on national standards and best practices, and scaled to the organization's current resources. In addition to testing and implementing this system, the work also included updating archival workflows, policies, and documentation to be inclusive of electronic records. The project also has provided an opportunity to spread awareness within the organization about the digital lifecycle, the relationship between production and preservation of digital objects, and how the development of a digital archive directly supports the goals of WCS and the WCS:2020 Plan.

The work was organized into four phases in which I collaborated with the Education and Exhibits and Graphic Arts Departments (EGAD), and the Conservation Innovations department of the Global Conservation Division as case studies for configuring the pilot system. The NDSR-NY residency began on August 15th, 2015, with two weeks of in-depth reading and prerequisite workshops at METRO's office in Manhattan. Between September 11th, 2015 and May 27th, 2016, I worked with the WCS Library and Archives and participating departments at the Center for Global Conservation, which is located within the Bronx Zoo. The following is a list of my completed activities:

- Conducted a digital assets survey
- Identified archival needs and requirements
- Configured and tested a preservation platform with over 100GB of sample electronic records
- Updated and documented new workflows for managing a wide range and high volume of archival digital media
- Developed recommendations and guidelines for the creation, acquisition, and processing of electronic records at WCS.
- Contributed to wider recognition of and engagement with the WCS Library and Archives through regular blog and social media posting about the project
- Connected with and reported to the wider WCS community through department-wide presentations

Project Partners

An NDSR Steering Committee was developed before the project began, and was comprised of the following key staff from each department who served as ambassadors for the project and provided invaluable guidance and feedback through regular meetings and correspondence:

Leilani Dawson, *Processing Archivist (Primary NDSR Mentor)*

Kim Fisher, *Conservation Innovations Spatial Analyst and Lead Developer (Secondary NDSR mentor)*

Madeleine Thompson, *Institutional Archivist*

Amanda Lindell, *Director of Professional Development for Educators*

Eileen Cruz-Minnis, *Assistant Director of EGAD*

Other important contributors in the project include:

Kerry Prendergast, Director of WCS Library and Archives

Eric Sanderson, Conservation Ecologist

Mario Giampieri, **Chris Spagnoli**, **Jesse Moy**, Conservation Innovations

Our interviewees - especially **Kimio Honda**, **Nelson Then**, and **Erin Prada**

Al Moini, Information Technology Supervisor

Arul Chellaraj, Senior Systems Admin. and Security Officer

Project Execution

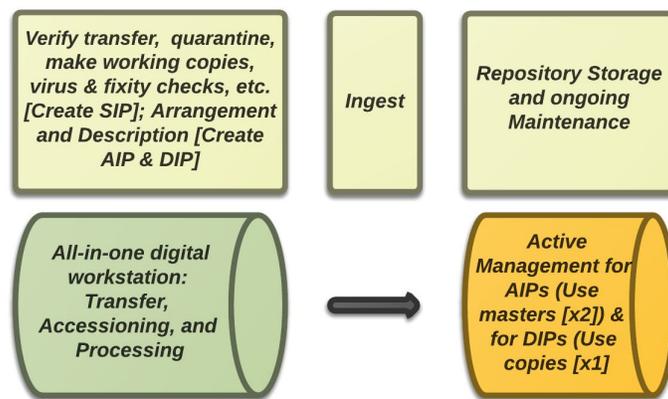
Phase 1

The interview and survey process took place between September and October of 2015, and involved the participation of a total of 15 key staff members from the participating departments. Interviews were generally 30 to 45 minutes in length, during which I inquired about the staff's day-to-day workflows and procedures for media management. These conversations were also an opportunity to better understand the relationships between the various departments and how media is worked on collaboratively, which became an important area of consideration for the organization's "Next Steps". With the assistance of Al Moini (Director of Information Support and Services), and Arul Chellaraj (Sr. System Administrator and Information Security Officer), I was also given access to the WCS networked server shares for each department, which enabled me to identify the common formats, media types, and organizational structures of each department. My report¹ for Phase 1 includes summaries of the departments' digital asset holdings and production workflow summaries, and also describes the challenges encountered while selecting and processing their media.

¹ *Phase1Summary_GHK_Draft1.2_201605_Revisions.pdf* (Havemeyer-King, 2015)

Phase 2

During Phase 2, which took place between November 2015 and January 2016, this data was used to define practical needs and functional requirements² of the digital archive. To aid in the evaluation and selection of a standards-based system architecture and digital preservation platform, I incorporated the Open Archival Information System (OAIS) Functional Entities³ and the activities outlined in the National Digital Stewardship Alliance (NDSA) Levels of Digital Preservation⁴ into a document outlining the functional requirements. The platform chosen during this project is a digital preservation software suite called *Archivematica*, which will be discussed in the following sections.



System Architecture: Minimal option, all-in-one digital workstation⁵

Following this, I met with Steering Committee members to select sample collections for testing the system with. This process helped lay the groundwork for developing a standard Selection and Appraisal procedure and updating the WCS institutional records retention policy to be electronic records inclusive. With EGAD and Education, the process involved meeting and identifying non-active media on their shared network server which would be considered historically significant according to the Library and Archives' current collection policy.

With Eric Sanderson's Welikia / Visionmaker development group, Conservation Innovations, I worked with Kim Fisher to create a data product inventory based on their production

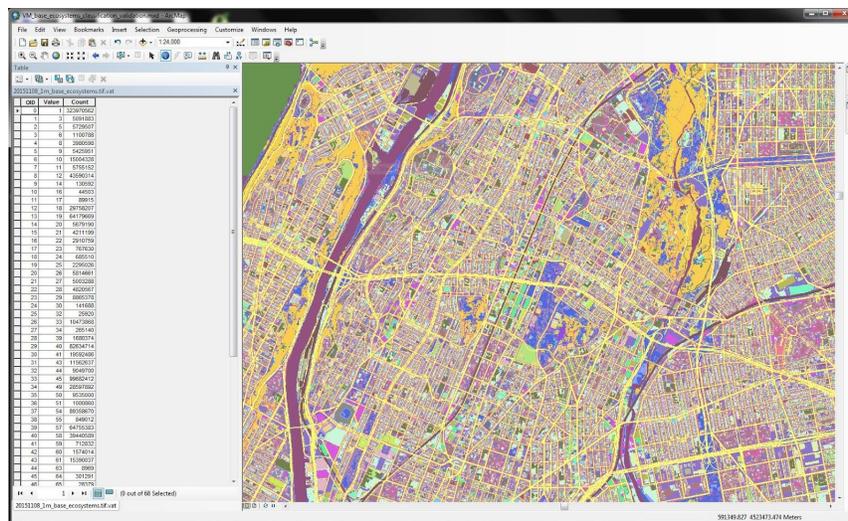
² *NDSR_Phase2-Report_Needs-Requirements_20151209-GHK-Final.pdf* & *Repository_Requirements_Checklist_GH-LD_201602.xlsx* (Havemeyer-King, 2016)

³ International Organization for Standardization. *Space Data and Information Transfer Systems, Open Archival Information System - Reference Model*. 2012. Web. http://www.iso.org/iso/catalogue_detail.htm?csnumber=57284

⁴ National Digital Stewardship Alliance. *Levels of Digital Preservation, v.1*. Web. http://ndsa.org/documents/Levels_v1.pdf

⁵ Dawson, Leilani. (Figure) *Minimal System Architecture*, 2016.

workflow⁶, which was used to locate and consolidate digital assets from the Mannahatta project. This inventory in particular was extremely helpful in determining what components of the project are the most significant and representative of the Mannahatta project as a whole. The final collection from this team consisted of core geospatial vector and raster data products, final imagery and publication materials, and lesson plans. Other materials that would be significant are correspondence, rough drafts of publications, notes, grant proposals, reports, and of course the website and Visionmaker application. For this project, we knew that, in general, standard documents probably would not pose any major complications in archival processing, and so we focused more on geospatial data. We had initially wanted to include Visionmaker source code and associated databases in the collection, but the preservation of software posed a number of issues regarding the scope of the digital archives, and we decided to prioritize spatial data in consideration of the project timeline.



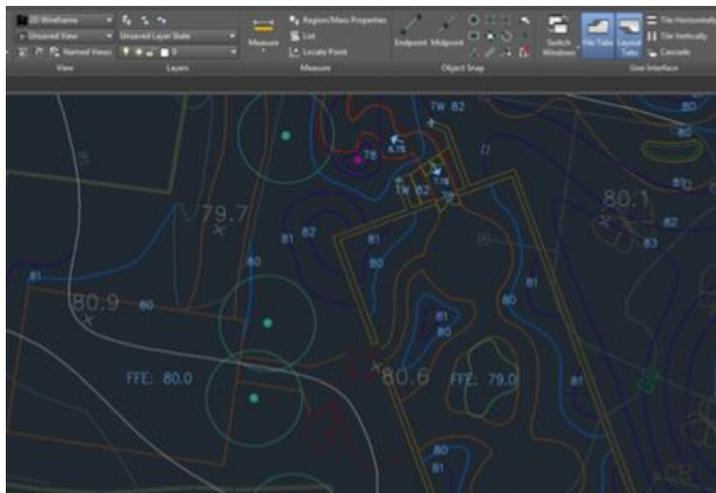
Contemporary geospatial raster and vector data for Mannahatta / Welikia / Visionmaker as viewed with ESRI ArcMap⁷

From EGAD, we collected project folders for four exhibits: World Of Birds; Butterfly Garden, African Plains Wild Dogs, and Queens Zoo Aviary. The contents of these folders included Photoshop, Adobe Illustrator, and CAD files, scans of hand-drawn illustrations, reference videos, photos and other image files, spreadsheets, presentations, text documents, and reports. These folders were pre-arranged according to EGAD’s file and directory structure policy⁸.

⁶ See *Mannahatta_data_product_inventory_GH-KF_201501.xlsx* (Havemeyer-King, 2016)

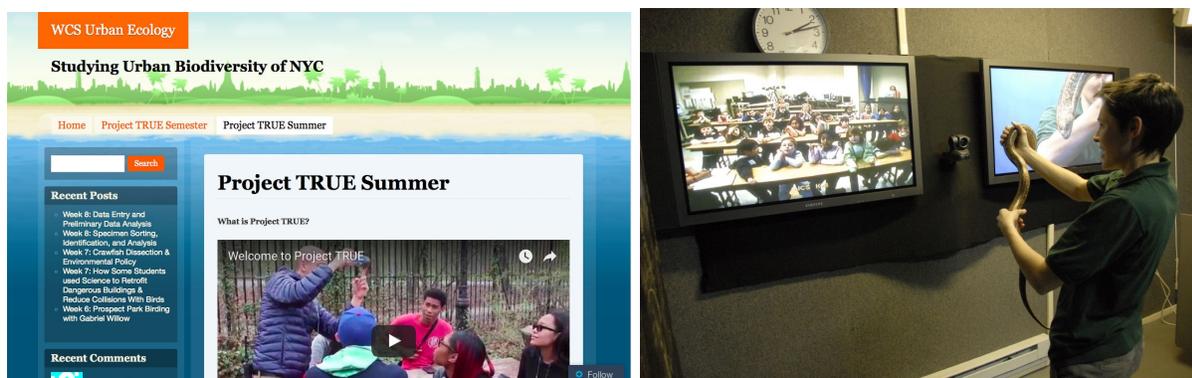
⁷ Image courtesy of WCS and Conservation Innovations.

⁸ See *EGAD_FileSetupManual-050714.pdf*, provided by Kimio Honda.



Example of CAD design file for Butterfly Garden exhibit⁹

From Education, we collected materials from two projects, the Distance Learning Program and Project TRUE, which included common document, image, and presentations files (relating to the administrative work on those projects), as well as geospatial vector data produced using Google Maps, and Prezis. The Project TRUE Blog was also partially captured using a new tool called Webrecorder¹⁰, which is currently in development and was a central focus of fellow NDS Resident, Morgan McKeehan’s project with Rhizome. Using the beta version, I captured 101.4MB of data in the form of a WARC file¹¹, which can be “played” using the accompanying Web Archive Player. This includes several months of posts by participants, project deliverables which are embedded in the site, and a curated selection of video and photo documentation.



Left: Screen capture of the Project TRUE Blog; Right: Photo documentation of Education’s Distance Learning Program in action.¹²

⁹ Image courtesy of WCS and the Exhibits and Graphic Arts Department

¹⁰ Kreymer, Ilya. Product of Rhizome, 2015. <https://webrecorder.io/>

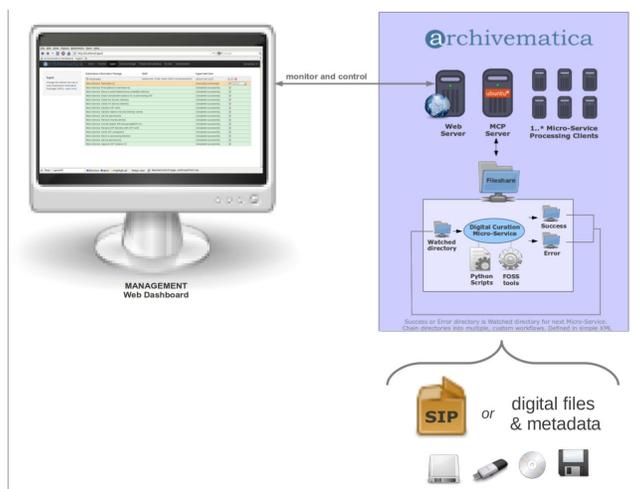
¹¹ Project TRUE Blog, WARC file:
 genfhk-wcsurbanecology-20151216184825978-00000-7-511c59e6ae57.warc.gz

¹² Images courtesy of WCS and the Education Department.

Obtaining this diverse cross-section of data definitely allowed us to explore a lot of preservation issues surrounding scientific research data, graphic and architectural design files, and documentation of the types that WCS produces more generally. During this selection process, I also began revising the Library and Archives' electronic records submission policies and transfer forms, and updating their Documentation Wiki¹³ to include general guidelines for submission to the Archives¹⁴.

Phase 3

Phase three involved the installation of Archivemata, first onto a WCS Archives laptop, and eventually onto a new desktop computer purchased during the project, which was carried out by Leilani Dawson. The system supports the preservation of digital objects through the use of open source microservices within a workflow that includes Transfer, Arrangement, Ingest, and Storage, and was structured in compliance with the OAIS Reference Model¹⁵. Archivemata is installed on the Linux based operating system, Ubuntu, and accessed primarily through a web-based dashboard. Details on the system requirements and installation process can be found at the Archivemata website¹⁶.



Archivemata system architecture¹⁷

¹³ Dawson, Thompson, Prendergast, Havemeyer-King. *WCS Archives Documentation Wiki*. 2015.

http://www.wcsarchives.org/documentationwiki/index.php?title=Main_Page

¹⁴ Havemeyer-King. *Electronic Records Submission Guidelines*. 2016. Appendix item [2]: *2_Wiki_Submission_Guidelines_WCS_Archives_Documentation.pdf*

¹⁵ Artefactual Systems. *What is Archivemata?*. Web.

<https://www.archivemata.org/en/docs/archivemata-1.4/user-manual/overview/intro/#intro>

¹⁶ Artefactual Systems. *Archivemata 1.4 Admin Manual, Installation*. Web.

<https://www.archivemata.org/en/docs/archivemata-1.4/admin-manual/installation/installation/>

¹⁷ Artefactual Systems. *Archivemata Technical Architecture*. Web.

https://www.archivemata.org/en/docs/archivemata-1.4/_images/AMarch.png

After a period of testing and research, it was determined that the WCS laptop could not support our medium-scale testing due to low processing power and inadequate memory and storage. The more robust desktop computer was purchased and has functioned well during additional test transfers.

A number of settings were adjusted within Archivematica to 1) ensure that files within each SIP are being processed properly and according to best practices and preservation standards (e.g. identified, indexed, converted - or not, and packaged), 2) prevent errors and failures, and 3) support an efficient rate of transfer of high volumes of data for the resources at hand. Below is a basic summary the configurations that were made:

- Storage pipelines and locations:
For the purpose of the pilot project, Kim Fisher provided the Library and Archives with access to a designated share on the Welikia / Visionmaker networked storage server, “Nostramo”. This share was used as a staging area for the sample collections. After multiple tests, it was determined that migration of files using an SMB connection to the staging area frequently fail due to character encoding issues with files made on multiple systems, and transferred over Linux. These two potential issues (Private Use Areas and UTF8 normalization rules, specifically) do not appear to cause errors with local or NFS storage connections, and therefore using one of these connection methods will be required for proper transfer.
- Format Policy Registry (FPR): Format entries, normalization rules & commands¹⁸:
Policies for the management, identification and normalization of certain file formats within Archivematica are defined by the Format Policy Registry¹⁹. After installing Archivematica, we ran a series of initial ingests to identify any immediate errors that occurred with the processing of a variety of file formats. The error logs from these tests were compiled into a list of formats to address in the FPR. Over a period of three weeks, I conducted research on recommended preservation formats²⁰ for various media types to inform the adjustment of format entries and rules for normalization in the FPR²¹.

This process also revealed Archivematica’s limitations in handling a number of complex formats, including CAD vector data and geospatial vector and raster data. For these files, the creation of a rule to bypass Archivematica’s default behavior was

¹⁸ See the FPR change logs on the WCS Archives Documentation Wiki for more details.
http://www.wcsarchives.org/documentation/wiki/index.php?title=Archivematica_Setup_and_Troubleshooting_Guide#Format_Policy_Registry_Configurations_.26_Logs

¹⁹ Artefactual Systems. *Documentation: FPR - Format Policy Registry*. Web.
<https://www.archivematica.org/en/docs/archivematica-1.4/user-manual/preservation/preservation-planning/#fpr>

²⁰ Library of Congress. *Recommended Formats Statement*. 2016. Web.
<https://www.loc.gov/preservation/resources/rfs/>

²¹ See *Archivematica_FPR_change_log.xlsx* and *Archivematica_Normalization_Plan.xlsx* for details on format policy changes.

necessary to preserve them according to best practices. Some formats are not yet supported by Archivemata's current normalization tools, namely .dwg and .dxf file formats. For these files, manual normalization with DWG TrueView or a similar application will be required to create access copies²².

- MySQL (heap / packet size) and Elasticsearch settings (memory):
Some of Archivemata's default settings do not adequately support the transfer of high volumes of data. Our initial tests were plagued by failures due to the system timing out because of the impact of these settings on the low processing power of the laptop. Upon setup of the new desktop workstation, we manually configured Archivemata's database settings using recommendations found on the Archivemata Users Google Group²³.

All of these above changes are now documented on the Archives' Wiki page, *Archivemata Setup and Troubleshooting Guide*²⁴, as well as within individual document files which reside on the digital workstation (copies are also included in the final NDSR final deliverables package).

Outreach and Dissemination Activities

Phase 4: Documentation & Reporting

During April and May, I compiled documentation on the transfer and ingest processes, which has been added to the WCS Archives Documentation Wiki. Below is a list of the documentation that has been created or updated²⁵.

- Digital Archives Policy
- Electronic Records Submission Agreement and Transfer Form
- Electronic Records Submission Guidelines*
- High-Level Workflow Outline & Diagram
- Archivemata Setup and Troubleshooting Guide*: Notes for general use, settings adjustment, and troubleshooting
 - Archivemata Normalization Plan
 - Archivemata FPR Change Log
 - Archivemata Ingest Log
 - Archivemata Issues Log

²² See the *Manual Normalization Instructions* on the Archives Wiki for more detail: (Appendix item [1]: [1_High-Level_Workflow_Wiki_WCS-Archives-Documentation.pdf](#))

²³ Archivemata Users Google Group: <https://groups.google.com/forum/#!forum/archivemata>

²⁴ Dawson, Havemeyer-King. *Archivemata Setup and Troubleshooting Guide*. 2016. See Appendix item [4]: [4_Archivemata Setup and Troubleshooting Guide - WCS Archives Documentation.pdf](#)

²⁵ Items marked with an * exist both as text documents and Wiki pages.

Final reporting on the project was done through a series of presentations given at the following meetings. These presentations provided an opportunity to educate WCS staff on the significance of the project outcomes as well as the relationship between the digital archiving initiatives and the 2020 Strategic Plan. They also sparked engaging conversations with leadership staff about how data management can benefit the organization and the specific departments we worked with during the project. The slide decks for these presentations are included with the final project deliverables:

- May 2nd: WCS Leadership
- May 10th: Global Programs
- May 13th: Public Affairs
- May 25th: Education Leadership
- May 27^h: EGAD

NDSR Outreach

Throughout the residency, I contributed regularly to the NDSR Resident Blog to share about my progress and explore preservation issues that came up during the project, promoted NDSR events and activities on social media, and published a two-part post on the Library of Congress' esteemed digital preservation blog, The Signal. Below is a list of links to my blog posts:

NDSR Resident Blog:

[*The Digital Ecosystem at the Wildlife Conservation Society*](#)

[*AMIA 2015 - Highlights*](#)

[*In With the Old*](#)

[*Trojan Dots and DIY Solutions*](#)

[*Let's Get Digital Recap*](#)

The Signal:

[*Blurred Lines, Shapes, and Polygons, Part 1: An NDSR Project Update*](#)

[*Blurred Lines, Shapes, and Polygons, Part 2: An Interview with Frank Donnelly, Geospatial Data Librarian*](#)

Analysis and Evaluation

Project Results

Standards Compliance

One main goal of the project was to build a system that adhered to national preservation standards and benchmarks provided by other institutions. To track progress towards this goal, we used an assessment tool called the NDSA Levels of Digital Preservation, which includes

four levels of activity or actions that can be taken to mitigate the risk of data loss. At the start of the project, Leilani Dawson had noted that the Library and Archives had not yet achieved level one, and with this system in place, we've met four out of five categories in both levels one and two. This essentially means that the Archives is maintaining good digital storage and management practices and is positioned well to provide guidance on how media is created. Additionally, by using Archivematica, the system is in close compliance with the OAIS reference model. Improvements to be made to achieve full compliance in these levels would include:

- Ensuring redundancy and geographical separation of archival copies with robust digital storage.
- Developing a schedule and method of monitoring file fixity, such as Fixity²⁶, which is a command-line tool developed by Artefactual for validating the checksums of AIPs created with Archivematica.
- Implement the use of secure file transfer protocols, such as the use of write-blockers and/or preservation-oriented transfer utilities, such as Exactly²⁷.

NDSA Levels of Digital Preservation

Category	Level 1: Protect your data	Level 2: Know your data
Storage & Geographic Location	2 Copies, Migration	3 Copies; one far away Storage workflow documented
File integrity	Check &/create fixity data	Check fixity; Write blockers; Virus Scan
Security	Permissions / Restrictions	Document access restrictions
Metadata	Inventory	Admin. & Pres. metadata
File Formats	Input on formats used in production...	Inventory of file formats in use.

What Can & Can't Be Processed

As a result of the changes made to the FPR and configurations made to the Archivematica system settings, the current system is able to transfer, identify, scan, and preserve most common file formats, as well as Geospatial data, most architectural and design files, common word documents, spreadsheets, raster image formats, and audiovisual files. Formats that remain troublesome include relational database files and flash-based file formats (specifically

²⁶Artefactual Systems. *Fixity*. <https://github.com/artefactual/fixity/blob/master/README.md>

²⁷ AVPreserve. *Exactly*. <https://www.avpreserve.com/avpsresources/tools/>

Prezi). The following tables illustrate which formats the WCS Library and Archives can now accept and process²⁸:

Accepted Media Formats

General Media Formats

Media Type	Formats	Collection Examples
Audio	MP3, WAV, WMA	Music, interview recordings
Email	PST, MBOX	Outlook email archives, Gmail email archives
Text Documents	TXT, RTF, DOC, DOCX	scripts, notes, plain text documents
Spreadsheets	XLSX, XLS, XLSB, NUMBERS	Inventories, schedules
Portable Document Format	PDF, GeoPDF*	Documents / slideshows / forms
Presentations	PPT, PPTX, Keynote	Slideshows
Camera files	3FR, ARW, CR2, CRW, DCR, DNG, ERF, KDC, MRW, NEF, ORF, PEF, RAF, RAW, X3F	Digital negatives, raw video and digital images
Raster images	BMP, GIF, JPG, JP2, PCT, PNG, PSD, TIFF, TGA	Pictures, graphics, Photoshop projects
Vector images	AI, EPS, SVG, DXF, DWG	CAD files and other architectural design projects (geospatial vector data, see below)
Video	AVI, FLV, MOV, MPEG-1, MPEG-2, MPEG-4, SWF, WMV, MKV	WCS-Produced promotional videos, event documentation
Web content	HTML, CSS	(Web pages)

Geospatial Data and formats of geospatial-related media:

²⁸ These tables are included in *WCS-ElectronicRecordsSubmission_Guidelines_20150523.docx*, created during this project.

Geospatial Data Type	Formats	Collection Examples
Databases, GeoDatabases*	DB, SQL, GDB	PostgreSQL
Raster data	GeoTIFF, ESRI GRID, GeoPDF	Georeferenced images with/without and associated auxiliary files
Vector data	SHP, KML, (+ associated files (ESRI World Files, .aux, .mxd etc.))	Shapefiles, Google Map files

* The Library and Archives can currently only preserve bit-level copies of database files; Section 3.5.3 of my WCS final report²⁹ details recommendations for how to more adequately preserve and provide access to this kind of data.

Challenging Formats

Proprietary formats posed challenges for processing during this project (most notably the Prezi file format). This was due to complex directory structures and dependence on other proprietary applications. In consideration of this, I chose to develop an ongoing list of formats which the Archives would prefer not to collect, and some notes on how to transform these formats into a more desirable media product. For example, with Prezi's specifically, it was recommended that they be downloaded as PDFs or recorded as a video using a built-in screen capture utility (to capture the motion-graphic nature that is unique to Prezi).

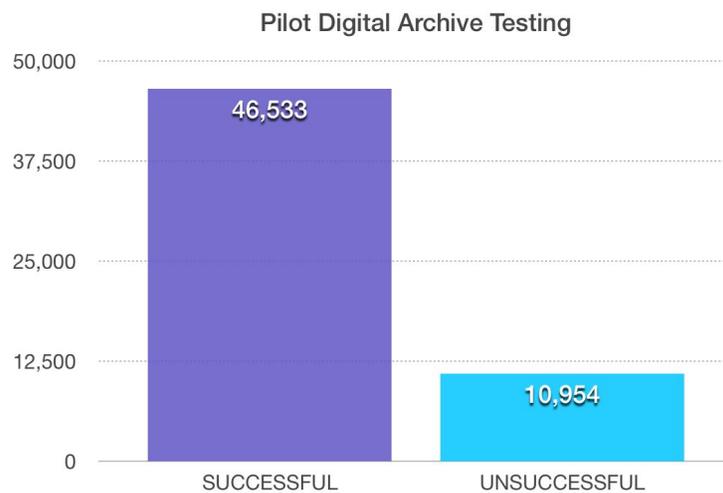
Other Data Challenges

- Relational databases & Software applications
The preservation of relational databases and software was a point of discussion throughout this project, and especially in regards to the Visionmaker web application. In the end, we did not collect these materials. After careful consideration, it seemed impractical to accept these types of data, as the resources required for the Digital Archives to adequately reproduce or maintain the databases are beyond what may be available in the foreseeable future.
- File and folder naming
Some sample files contained problematic characters in their file and directory names, which initially caused failure during transfer when Archivemata was connected to the staging area over SMB. Specifically, the main problem was with Private Use Areas

²⁹ Havemeyer-King, Genevieve. 2016. "WCS-NDSR Final Report". [WCS-NDSR_FinalReport_GH_20160527.pdf]. Note: The "WCS-NDSR Final Report" is the in-depth final report created as a Phase 4 deliverable for WCS staff. It can be found in the folder titled WCS-NDSR_Final_Deliverables_Packet_GH_20160527.

(PUAs)³⁰. Our first attempt in addressing this issue involved using a Python script created by Kim Fisher intended to “clean” the entire collection by replacing PUAs and other invalid characters with “_” (underscore). After further testing, it appears that the script is not catching certain files or PUAs. However, with Archivematica connecting to the staging area over NFS, these characters currently do not cause any issues during transfer. They are however still a cause for concern, as it’s not certain how they will behave or affect transfer or preservation in the future. This script should be further developed if it is still to be used to “clean up” folder and file names.

Final Ingest of Sample Collections



After nearly 2 months of testing and configuring the system, the sample collection was narrowed down to approximately 101 GB of data, or 57,487 digital objects. This final collection consists of one large project from each department (three total).

At the time of this report, approximately 81% of this collection has been successfully processed in Archivematica. The remaining 19% was contained within a single folder which failed upon ingest most likely because a majority of the files therein did not contain file extensions³¹ and could not be identified or processed correctly.

Ingest Volume & Time:

The current system can handle approximately 5,000 digital objects per transfer, at a rate of approximately 1,500 objects per hour (depending on the processing rules and services used for various formats). However, the size of our final archival information packages (AIPs) was consistently two to ten times larger than the original transfer. This is due to the variation in

³⁰ The Unicode Consortium. *Private Use Characters, Non-characters, and Sentinels FAQ*. Web. May, 2016. http://www.unicode.org/faq/private_use.html

³¹ *Research into solutions for handling files without extensions is currently underway, and it is assumed that this is the only barrier to processing these remaining files.*

processing and normalization activities, which for some files involves the creation of preservation versions that are often larger in size, as well as the creation of the METS metadata record which logs the processing history of every single file in a transfer (so it can become quite large for transfers of thousands of files). Although SIP size is an important consideration, the number of digital objects has proved to be more important than the size of individual files or individual collections of files (transfers) in factoring the amount of time it takes to process and the estimated increase in AIP size.

Overall, it took approximately 31 hours, or about one full work week, to process these three collections and log each test. This final phase of ingest, however, did not include Appraisal & Selection meetings with staff, collection and transfer of files to the staging area, quarantine of files, or cataloging. It is likely that including these workflow processes would add roughly one to two weeks to the overall timeline for a similarly-sized collection, but the volume and rate of transfers will likely vary.

To increase processing efficiency, the Archives could continue experimentation with different installation configurations, or upgrade their digital workstation in the future to add more processing power.

Metadata

During this project, metadata became a focal issue mainly in reference to complications with the METS metadata errors we encountered when configuring Archivematica. We did not however, focus extensively on metadata standards for the purpose of cataloging or discoverability - an important aspect of preservation that is often neglected.

As the project began to wrap up, the importance of developing policies for the extraction and inclusion of geospatial metadata resurfaced and we were able to meet with Kim Fisher and GIS Analyst, Mario Giampieri, to discuss the most efficient workflow for generating standards-compliant metadata records from ArcCatalog. Balancing the basic requirements of the Archives with recommendations put forth by the OpenGeoportal (OGP)³², we developed a schema³³ using select Federal Geographic Data Committee standard elements³⁴ and OGP elements, which outlines a streamlined set of "critical elements for discoverability and identification"³⁵. The general guidelines we developed for creation of geospatial metadata at WCS satisfy the basic informational needs of the Archives while requiring only a minimal amount of time investment from the staff who are creating such data.

³² OpenGeoportal. *Metadata Creation Guide*. Web. May, 2016.

<http://opengeoportal.org/working-groups/metadata/metadata-creation-guide/>

³³ Dawson, Fisher, Giampieri, Havemeyer. 2016. See Appendix item [3]:

[3_WCSArchives_GIS_Metadata_Schema.pdf](#)

³⁴ FGDC. *Metadata Quick Guide*. 2006. Web. May, 2016.

<http://www.fgdc.gov/metadata/documents/MetadataQuickGuide.pdf>

³⁵ OpenGeoportal Metadata. 2013. Web. May, 2016.

http://opengeoportal.org/wp-content/uploads/2013/12/OpenGeoportal_Metadata.pdf

Outcomes & Impact

My involvement in this process of planning and implementing a system for selecting, appraising, and preserving digital records has been formative in my development as a media archivist. In addition to becoming more comfortable with decision-making and application of best practices for digital preservation, I have found that the success of a project hinges on buy-in, and particularly on a shared understanding of the distinct, but overlapping, priorities of IT departments and digital archivists.

By keeping a dialog with two different IT groups at WCS and sharing our concerns with them, I believe we, as a team, were able to establish a kind of partnership that will enable the WCS Digital Archive to be more visible to WCS as a whole, and position the Archives to be an integral part of WCS's larger activities in conservation.

Next Steps

The WCS Library and Archives plans to continue the efforts of this project by migrating their collection of electronic records that reside on legacy digital media, such as CDs and floppy disks. My deliverables for the project also included a list of recommended steps to carry out in short and long-term time frames, which will be used to pick up where my work has left off. These include practical measures for filling the gaps left in their goal of reaching NDSA Level 2, as well as conducting quality assurance tests and refining their Access policies and workflow for digital media. Long term recommendations which were most important emphasized the need to ensure the sustainability of the project through internal support and stronger engagement with media creators at WCS as a whole.

This project was extremely enriching for me, and I fully aim to bring the sense of advocacy and inclusiveness that I gained at WCS with me to future roles. In consideration of the Archives' current plans, I'm confident that the results of the project will have a lasting benefit at WCS as the digital archive becomes more established over time.

Professional Development

Overview

Generally speaking, the line between our "20%" time and the time that was required for participating in NDSR workshops hosted by METRO, and NDSR outreach activities, such as blogging and social media, was somewhat blurred. This made it difficult to separate my work in these areas from my originally stated personal goals for the residency, which included becoming more familiar with scripting and programming languages such as Bash and Python, as well as

learning more about GIS and geospatial data preservation and librarianship. For a large portion of my 20% time, I decided to focus primarily on GIS, which was both integral to my project and also of personal interest. The other portion was spent on completing day-to-day NDSR activities and taking time away from WCS to attend conferences, events, and workshops related to digital preservation more generally.

In regards to my exploration of geospatial data, the learning curve was a challenge. Coming to the subject with zero experience, I gave myself a crash course in basic terminology and benefitted from Kim Fisher's patience and generosity in explaining the basic concepts during my first few meetings with the GIS team. With this basic understanding, I was able to better navigate the topic of preservation for this kind of data, and became more comfortable conversing with professionals in the field outside of WCS. The result of my research into this area has been that I'm more aware of the changes taking place in the field, particularly in regards to the creation and management of geospatial metadata and the creation of more dynamic access methods for GIS collections.

Activities

Below is a summary of the activities I participated in which pertained to the NDSR-NY cohort's individual, group, and external professional development requirements:

- Group field trip to Princeton:
 - ReCAP: Tour of ReCAP facilities and group discussion
 - Princeton University Center for Digital Humanities: Group discussion with Associate Director, Jean Baur, about the Center's activities and approaches to engaging with students on digital humanities projects.
 - Princeton University Archives: Tour of the archives and digital preservation facilities, and in-depth demonstration of BitCurator software suite with Digital Archivist, Jarret Drake.

- {Let's Get Digital} NDSR Symposium
 - Planned and co-hosted a day long symposium of talks and workshops on digital preservation topics.
 - Co-presented in two sessions: *Planning to Preservation: Method and Madness*

Conferences, Workshops, and Memberships

The following table describes my usage of the professional development funds provided for the residency. Overall, the ability to engage with working professionals and across disciplines in the field of preservation was extremely valuable. The sessions I've attended at various conferences throughout the year have provided important examples of others' techniques and

methodologies for specific preservation activities (such as migration, digitization, and development of large-scale digital asset management systems).

They have also made space for discussion with a larger preservation community about how we as archivists can continue to make professional development more accessible by bringing back information to our communities through skill-sharing and smaller event hosting. Having the opportunity to attend so many conferences also directly impacted our ability to successfully host our own event, after learning from the models of others.

Event / Membership	Date
TechFocus III - NYC	9/ 25-26/2016
AMIA - Portland	11/ 18-21/2016
AMIA Membership	2015-2016
Code4lib - Philadelphia	3/7/2016
Digital Asset Symposium - NYC	5/4/2016
METRO Annual Conference - NYC	1/21/2016
Orphan Film Symposium - Culpeper	4/6-9/2016
GIS Workshop at Baruch College - NYC	4/30/2016
Society of American Archivists Membership	2015-2016