

SOLOMON R. GUGGENHEIM FOUNDATION

Electronic Records Management Start-Up Project Generously funded by the NHPRC

Three-tiered Plan for Managing Electronic Records

September 15, 2014 – Prepared by Anthony Cocciolo, Electronic Records Consultant

Introduction

The purpose of the NHPRC Electronic Records Management Start-Up Project is to plan an electronic records repository to store the digital records with enduring value (historical and/or legal value) produced by the Solomon R. Guggenheim Foundation (SRGF). This project includes five stages, with the final stage resulting in a three-tiered plan for an electronic records repository (this report).

During the earlier stages of this project, particularly Stage II: Survey of Electronic Records, the Electronic Records Consultant confirmed that electronic records represent a significant and growing portion of the Foundation's records. For example, over two-million files authored within the last seven years are available on SRGF's network storage. Because of the steep growth in electronic records in addition to continued use of paper records, preserving and making available select electronic records is essential to fulfilling the mission of SRGF Archives, which is to collect, preserve, and provide access to the historical documentation of SRGF for research and administrative use. Implicit in this mission is also to identify records that do not have permanent value and eliminate them in a timely manner, which is necessary both for legal compliance and for the sensible use of valuable IT resources. Records management responsibility was codified in 2007 when the Guggenheim's Board of Trustees approved its records retention schedule, which makes clear the record groups that should be permanently retained as well as those that should be kept only for a limited duration.

Just as collecting and preserving electronic records is essential to the mission of the Archives, the Archives are essential to the mission of the Guggenheim. The Archives collects and provides access to the documentation that illustrates the role played by the Guggenheim within art, architecture and visual culture communities. More broadly, the Archives makes evident the ways in which the Foundation contributes to the creative and intellectual life of the local and global communities it serves. By providing access to such documentation, the Guggenheim furthers its mission to create dynamic and rich exchanges around modern and contemporary art. The archives specifically advances this mission through engaging audiences such as scholarly researchers, art critics, and researchers from the general public, who visit the Archives nearly every week of the year.

Through the pilot projects made possible through this grant project, the project team was able to establish an electronic records repository that meets the minimum standards for being a trustworthy repository appropriate for storing content with permanent value.¹ Information on individual pilots is available in Appendix A. Tier 1 of this plan to manage electronic records reflects the repository established through the pilots, which only used the existing grant funds. However, as the grant will end in September 2014, it is essential that the Guggenheim continue to invest in the preservation of electronic records because of their centrality to the mission of the Archives and the Foundation. Tiers 2 and 3 indicate the ways in which the foundation can make investments with respect to its electronic records to ensure their persistence and availability to researchers. Tier 2 requires a minimal yearly investment to ensure that there is staff and infrastructure appropriate to supporting digital archives and repository, albeit a relatively thin foundation for building a robust program. Tier 3 provides a more sound foundation for infrastructure and staffing, and lays the groundwork for digital archives program that is appropriate to a world-class museum.

¹ This is contingent on continued use of the "Trustworthiness Checklist" developed in Pilot 1 and outlined in the "SRGF OAIS Repository" report.

Tier 1 – Current Status

Overview

Tier 1 necessitates that existing staff reorganize their work to take on more tasks related to managing their electronic records. This option recognizes that electronic records need to be processed by an archivist sooner than paper records because of their relatively fragile nature (e.g., obsolete file formats and media, ability to easily delete files, etc.). This option is not ideal as existing staff are already heavily taxed in processing the extensive backlog of paper-based records, as well as performing other duties such as responding to researcher requests, providing records management services and engaging the public through exhibitions and social media communications. Although staff have received professional development training with respect to digital archives through the funding made possible in this project, it is difficult to imagine how a robust digital archives program can be developed where there is no single person dedicated to making it a success.

Outcomes

- Make available a service for staff to deposit electronic records with the archives.²
- Provide limited access to deposited collections.

Risks

- Since this option requires that existing staff reorganize their work, processing of paper-based collections and responding to researcher requests will operate slower than in the past.
- Backlogs of inactive records, such as those of the Directors Office and the Special Projects department, will not be incorporated into the electronic records repository, putting them at great risk of loss or damage.
- Since uncompressed video requires extensive storage, it cannot be accepted into the electronic records repository due to lack of storage space. .
- Removable media—such as content on deteriorating DVDs—cannot be incorporated into the electronic records repository due to lack of staff time.

Technology Infrastructure

The technology infrastructure for Tier 1 was established in Pilot 1 using funds available in the NHPRC grant. More information on this infrastructure is available in the Appendix B.

Staffing

Tier 1 assumes no additional staffing. Existing staff would use the following approximate percentage of their work time dedicated to tasks related to electronic records.

- Director of Libraries and Archives (current staff) - 10%
- Associate Archivist (current staff) - 20%
- Limited project support from interns with educational component

² Uncompressed video that is not cannot be deposited because of disk space concerns.

Cost

Annual Operating Cost: 0.00 USD
Annualized Capital Cost: 1,000 USD *

* It should be noted that the storage hardware purchased with grant funds should be replaced in approximately 5 years to avoid hardware failure. Thus, approximately 5,000 USD would need to be reserved for this replacement every five years.

Tier 2 – Meets Minimum Requirements

Overview

Tier 2 provides minimal investment to establish a digital archives program by hiring an Assistant/Associate Archivist with specialization in digital archives and providing some minimal funds to make available hardware and software needed to preserve and make accessible the variety of digital files that SRGF employs.

Outcomes

- Make available a service for staff to deposit electronic records with the archives, including significant uncompressed video.
- Target specific departments for incorporating their backlog of records into the electronic records repository, including removable media.
- Free-up space on the IT department's network drives.
- Respond to staff and external researcher requests in a timely manner.
- Provide access to deposited collections.

Risks

- Since the backlog of records that need to be appraised for inclusion in the electronic records repository or scheduled for deletion is great, only select departments will be targeted for appraisal. Thus, some important documentation is at risk of being lost or damaged, and insignificant documentation is at risk of consuming valuable IT resources.

Technology Infrastructure

Tier 2 assumes that same technology infrastructure established in Pilot 1, with the following additions:

- Additional Storage space (with redundancy) to accommodate uncompressed digital video.
- Upgrades to hardware purchased in Pilot 1 needed every 5 years.
- Rhinoceros Software License: For researchers and archivists to access 3D models
- VectorWorks Software License: For researchers and archivists to access 2D drawings
- Miscellaneous supplies, services and equipment as required by Assistant/Associate Archivist to manage born-digital records (e.g., purchasing a drive to read obsolete media, buying a piece of software to render a file, etc.).

Staffing

- Assistant/Associate Archivist (TBD) - 80% digital archives; 20% related archives responsibilities
Brief Job Description – The Assistant/Associate Archivist with specialization in digital archives is responsible

for collecting, preserving and making accessible SRGF's born digital records with permanent value. Reporting to the Director of Library and Archives, he/she will identify records for permanent retention through regular collaborations with creating departments, incorporate born-digital records into the electronic records repository, remove content without permanent value from SRGF networks, ensure the persistence of digital information through infrastructure collaborations with the Information Technology department, and respond to researcher requests. He/she should have education and experience related to digital archives, including holding a MSLIS from an ALA-accredited institution of higher education with emphasis in digital archives (or related education) and at least 1 year of relevant work experience. He/she will supervise interns, and collaborate regularly with the digital asset manager and other staff in the Library & Archives department.

- Director of Libraries and Archives (current staff) - 10%
- Associate Archivist (current staff) - 20%
- Limited project support from interns with educational component

Cost

Annual Operating Cost: Salary/Benefits + 2,000 USD

- Assistant/Associate Archivist Salary (TBD)
- Assistant/Associate Archivist Fringe (36%) (TBD)
- Miscellaneous Supplies, Services and Equipment: 1,000 USD
- Continued professional development for staff around digital archives: 1,000 USD

Annualized Capital Cost: 5,800 USD

- VectorWorks Software License: 1,800 USD*
- Rhinoceros Software License: 1,000 USD*
- Upgrades to hardware purchased with grant funds every 5 years: 1,000 USD
- Purchase of additional storage space for AV materials: 1,000 USD
- Upgrades to additional storage space for AV materials every 5 years: 1,000 USD

* One time-purchase, with upgrades to licenses every five-years.

Tier 3 – Best in Practice

Overview

Tier 3 provides moderate investment in the digital archives program by hiring an Assistant/Associate Archivist with specialization in digital archives as well as a temporary Archives Assistant to assist with processing the backlog of digital records. Additionally, Tier 3 provides some funds for extending the infrastructure.

Outcomes

- Provide a service for staff to deposit electronic records with the archives, including significant uncompressed video.
- Comprehensively incorporate backlog of records into the electronic records repository, including removable media, from all departments.
- Systematically free-up space on the IT department's network drives.
- Important exhibition documentation contained with the DAM—such as installation photographs—will be systematically replicated in the electronic records repository. This is consistent with the notion that “lots of copies keep stuff safe.”

- Develop “post-custodial” digital archival services, which include assisting departments in maintaining digital records that are not physically located in the archives department (e.g., collection information held within the Curatorial, Conservation and Registration departments that are permanently active records which will not be transferred to the electronic records repository).
- Write grants and seek funding for projects around issues of post-custodial digital archives, DAM integration, and Audio Visual digital archives.
- Provide access to deposited collections.

Risks

- None.

Technology Infrastructure

Tier 3 employs the same technology infrastructure as Tier 2. However, it also includes employing the following services:

- Third Copy: The best practice in the preservation of digital records is to have three copies of the data that are in different geographic areas.³ Thus, there are two options to achieve this: either through the transferring of media to offsite storage (e.g., data tapes to a remote art storage facility) or use cloud storage. A cloud provider that could achieve this is DuraCloud, which is a service that manages the transfer of digital content to a variety of geographically-dispersed cloud providers (Amazon and Rackspace). Automated procedures for replicating data on the Archivemata server to DuraCloud would need to be completed to ensure information is made redundant without regular human intervention. A limitation of this approach is that it relies on Internet transfer, which could be problematic for large transfers of content to DuraCloud given the Internet bandwidth currently available at SRGM facilities.
- Archive-It⁴ – Rather than attempting to create web archives in-house as described in Pilot 10, this would outsource the creation of the web archives to the Internet Archive (IA) through the Archive-It service. The advantage is that this service is used by a wide-variety of cultural institutions and the IA staff have deep knowledge of web archiving, thus helping ensure high-quality web archives. Archive-It will be used to archive SRGF websites regularly (approximately one to three-times a year). Additionally, Internet Archive content from year 2000 through the year of Archive-It adoption will be incorporated into the Archive-It account (this is a service provided by Archive-It). This will allow enhanced preservation and access to some of SRGF earliest websites. Additionally, the WARC files (web archive files) will be downloaded from the Archive-It account and ingested into the electronic records repository for safe keeping. The use of Umbra—the new software from IA that can be used with Archive-It—should also be used for enhancing the quality of the web crawls with respect to interactive content such as Javascript.

Staffing

- Assistant/Associate Archivist (TBD) - 80% digital archives; 20% related archives responsibilities
Brief Job Description is same as Tier 2
- Archives Assistant (TBD) – 100% - Temporary Staff Member
Brief Job Description: Under the supervision of the Assistant/Associate Archivist, the Archives Assistant will appraise records for inclusion in the electronic records repository, schedule records for deletion in accordance with the Guggenheim’s Records Retention Policy, and transfer records in accordance with the

³ See NDSA Levels of Digital Preservation, Level 4: <http://digitalpreservation.gov/ndsa/activities/levels.html>

⁴ <https://www.archive-it.org/>

Guggenheim's Electronic Records Processing Manual. The Archives Assistant should have completed a BA/BS degree, and have an interest in further study of digital archives, information and library science, museum studies, or art history.

- Director of Libraries and Archives (current staff) - 10%
- Associate Archivist (current staff) - 20%
- Limited project support from interns with educational component

Cost

Annual Operating Cost: Salary/Benefits + Hourly/Benefits +13,100 USD

- Assistant/Associate Archivist Salary + Fringe (TBD)
- Archives Assistant: Hourly + Fringe (TBD)
- Continued professional development for staff around digital archives: 1,500 USD
- Third Copy: DuraCloud subscription (assuming 10 TB of storage): 8,800 USD or Tape/Media shipping to remote location (8,800 for technology, media and staff time).
- Archive-It Subscription (with Umbra and incorporation of year 2000 content to current, price range is 3K to 13K annually): 8,000 K
- Miscellaneous Supplies, Services and Equipment: 2,000 USD

Annualized Capital Cost: 6,800 USD

- VectorWorks Software License: 1,800 USD*
- Rhinoceros Software License: 1,000 USD*
- Upgrades to hardware purchased with grant funds every 5 years: 1,000 USD
- Purchase of additional storage space for AV materials: 1,500 USD
- Upgrades to additional storage space for AV materials every 5 years: 1,500 USD

Additional Infrastructure to Consider Adopting

A necessary component for the preservation of digital information is to have redundant copies of data since even high quality hardware occasionally fails (e.g., subtle bit-rot or bit-flipping, and more obvious failures). Thus, the field has discovered that “lots of copies keeps stuff safe.” According to the NDSA levels of digital preservation—sponsored by the Library of Congress—the minimum level of digital preservation should be two copies of the data that are not collocated.⁵ Ideally, copies should have some geographic separation. At the highest level, there should be three copies of data, each within geographic areas with different disaster threats (e.g., not all data stored near the coast). Since all SRGF locations are located relatively near the coast, it is recommended that SRGF invest in some geographic dispersal strategy. This could include more substantial investment in services such as DuraCloud, where data would be stored among a variety of geographically dispersed cloud providers. An alternative is to invest in tape technology (such as LTO 5 or 6), and sending tapes of critical data such as the digital archives on regular rotation to locations that are geographically dispersed (e.g., an art storage facility).

Another limitation discovered during this project is the difficulty in being able to move digital data among various SRGF locations because of relatively low-bandwidth connections between these various sites (100 Mbit/sec). A recommendation is that SRGF invest in network bandwidth between locations—and/or re-architect its wide area network (WAN)—so that it can more easily replicate content across locations. This would be useful not only for the digital archives, but for other initiatives that rely on digital preservation such as born-digital artworks, as well as conservation and collection documentation.

⁵ <http://www.digitalpreservation.gov/nds/activities/levels.html>

Appendix A: Pilot Project Summary

A major outcome of the NPHRC Electronic Records Management Start-up Project is executing on several pilot projects related to preserving SRGM electronic records. The proposed pilots are described below:

Pilot 1: Setup of an Open Archival Information System (OAIS) Compliant Repository

Objective: To setup, test, and evaluate an OAIS-compliant electronic records repository.

The Open Archival Information System (OAIS) reference model is a well-accepted standard for the long-term preservation of digital records. Several software packages have been developed to adhere to the OAIS model, such as Archivematica⁶, Tessel's Preservica,⁷ and OCLC's Content DM.⁸

Pilot 2: Digital Records Processing and Obsolete Media Ingestion Workstation

Objective: To plan for and install a digital records processing and obsolete media ingestion workstation that can double as an archive viewing workstation.

This workstation allows for the imaging of obsolete media (3.5 floppy, 5 ¼ floppy, Jazz and Zip drives), preparation of records for ingestion, and ingestion of digital assets into its electronic records repository as well as software for viewing a variety of file types used by SRGM over the last 2 decades. This workstation can be used by archives staff for imaging of media, as well as by SRGM staff for reading old media and viewing the contents of the electronic records repository. This machine could also be used outside researchers for viewing contents of the electronic records repository, although optimally a second workstation would be acquired for this purpose.

Pilot 3: Selection of Electronic Records from Departmental Network Storage for Transfer to the Archives

Objective: To develop guidelines and workflows for appraising electronic records and transferring them to the archives, especially those records on network storage over 10 years old.

In this pilot, methods will be explored from appraising and selecting department records from their departmental network storage for transfer to the archives, and non-permanent records scheduled for destruction.

Pilot 4: Workflows for Staff Transfer of Electronic Records to the Archives

Objective: To create processes and workflows for SRGM staff to transfer electronic records to the archives department.

This pilot will create processes and workflows for staff to transfer electronic records to the archives department. This will take into account the process for transferring paper records, and incorporate electronic records into this process. These workflows will also consider the relationship with the digital asset management system MediaBeacon.

Pilot 5: Copying Select Permanent Electronic Records to an Open Archival Information System

Objective: To copy select permanent records that originate in conventional file formats to an OAIS-compliant digital repository to test transfer workflows.

In this pilot, using the OAIS repository from pilot 1 and workflows from pilot 3, select permanent electronic records will be copied to an Open Archival Information System. This pilot will focus on selections of records that are commonly used standardized records, such as MS Office, JPEGs, and PDFs.

Since this is a pilot rather than the completed electronic records repository, these will only be copies of content (rather than transfers or moves of content). Aspects that will be explored are the ease of transfer, the ability to monitor the files, as well as access to them by archives staff. This process will also explore transfer/copying procedures to ensure the integrity and totality of the transfer/copy (e.g., Library of Congress' Bag-it protocol).

⁶ https://www.archivematica.org/wiki/Main_Page

⁷ <http://www.digital-preservation.com/>

⁸ <http://www.contentdm.org/>

Pilot 6: Problem Records: Obscure or Specialized Formats

Objective: To copy select files to the electronic records repository that originate in obscure or specialized file formats.

This pilot will investigate procedures for archiving obscure or specialized file formats, such as 2D drawings and 3D Models that were created in Vectorworks, AutoCad, and Rhino. This pilot will take into account the access needs by staff or researchers (impacting Pilot 2), as well as the conversion options (e.g., suitability of conversion of 2D drawings into PDF/A).

Pilot 7: Problem Records: Obsolete File Formats

Objective: To copy select files to the electronic records repository that originate in obsolete file formats.

This pilot will copy files to the electronic records repository that originate in obsolete file formats, and investigate procedures for conversion to new formats (e.g., Lotus 1-2-3 to MS Excel). It will also take into account the access needs by staff or researchers (impacting Pilot 2).

Pilot 8: Problem Records: Removable Media, including Obsolete Media and Removable Hard Drives

Objective: To copy files on removable media, such as floppy disks, CD/DVDs, and hard disks, to the electronic records repository.

This pilot will pilot the transfer of a variety of removable media, such as floppy disks and hard disks, to the electronic records repository. It will make use of the workstation from pilot 2.

Pilot 9: Problem Records: Preserving Significant E-Mail Correspondence

Objective: To develop guidance and workflows, and test software to preserve email.

SRGM records retention schedule identifies significant correspondence as being a permanent record group, and most of such correspondence has been conducted over email for the last fifteen years. This pilot will investigate some best practices in preserving email. It will investigate software for automatic transfer to the archives (e.g., integration with MS Outlook), and produce guidance for staff on what to transfer to the archives and how to do it. It will also investigate the former employee and other caches of email stored in PST or MSG files. It will also investigate pre-Exchange/Outlook email preservation needs. Research indicates that Novell Groupwise emails reside on the network, and research is needed if such email is significant correspondence, and the best ways to preserve/maintain such email.

Pilot 10: Problem Records: Web Archiving Guggenheim Websites

Objective: To develop procedures for preserving SRGF websites and microsites.

Currently, the SRGF websites are not archived (e.g., comprehensive web-crawls of relevant websites). Copies of the earliest iterations of the homepage are only stored on the Internet Archive. The website has become a major publication over the last 20 years, and this project will suggest that copies of the website get permanently retained, with copies made three times a year to correspond with the roll-out of major exhibitions. In this pilot project, web archiving software (such as Heritrix or wget) or services (Archive It) will be investigated and setup. This software can be installed onto the same server as used in Pilot 1. Research will be conducted on producing progressive downloads of the website to keep the size of the web archive as small yet as comprehensive as possible. Additionally, this project will investigate the downloading of web archives from the Internet Archive for earlier iterations of the website.

Pilot 11: Problem Records: Very large files (e.g., video)

Objective: To pilot the transfer of very large files to the electronic records repository, such as master video files.

This pilot will investigate the handling of large files within the archives, such as master video files. This pilot will make recommendations on how to handle this type of media, and where it should be stored, and how related derivatives should be stored (e.g., mezzanine copies, access copies, etc.).

Appendix B: Technology Infrastructure

See *Setup, Test, and Evaluate an OAIS-Compliant Trusted Digital Repository* PDF for more details.

Guggenheim Digital Archives Infrastructure

