Building Capacity in Digital Preservation: Theoretical Issues, Practical Applications

1. Transfer: Current Approach (SFU Archives)

Richard Dancy, Staff Archivist
Simon Fraser University Archives and Records Management Department
Study Leave 2018 Project
January 28, 2019

SUMMARY

This paper provides an overview of SFU Archives' current management of digital transfers. It describes supporting infrastructure (hardware and software) and workflows. It identifies a number of problems with the current approach and articulates requirements for addressing them. Subsequent papers in the report series take up the different problems and propose solutions.

TABLE OF CONTENTS

1. INTRODUCTION

Between 2013-2016 SFU Archives developed a standard workflow for transfer of digital materials from producers (university departments or private donors) to the Archives. This paper gives a brief overview of the current approach. It provides a common point of reference and baseline for other papers in the project report series.

- Section 2 lists the hardware and software components that presently support the transfer process.

- Section 3 describes the current workflow.

- Section 4 identifies a number of specific problems and requirements for addressing them.

- Section 5 links the subsequent papers in the series (on transfer methods, utilities, and validation) to the problems they address and attempt to resolve.

For terminology, see the fuller comments in the *Project Overview* report, section 4. I generally follow the OAIS terms, using **producer** for the entity that transfers materials and **archive** (singular) for the entity that receives them. **Institution** is used generically for GLAM organizations (galleries, libraries, archives, and museums). **Archives** (capitalized, plural) is usually reserved for the name of a particular institution (SFU Archives, the Archives). **Digital materials** is used for the stuff transferred.

2. INFRASTRUCTURE

SFU Archives' current infrastructure employs several components to support digital transfer:

- Digital Records Transfer Request Form: webform on the Archives' website; maintained in the university's content management system, Adobe Experience Manager (AEM).

- Archives Information System (AIS) database: custom-built FileMaker database with tables for managing transfer and accession metadata; internal use only.

- SFU MoveIt: custom-built Python script with Windows and Mac graphical user interfaces for packaging transfers as standard BagIt-compliant bags.

- SFU Vault: the university's local implementation of the OwnCloud file transfer and storage service.

- Checksummertime: a custom-built Mac Automator app that validates a transfer package checksum.

- Archivematica: local installation of the open-source digital preservation software; manages ingest of material into the Archives' digital repository.

- SFU AtoM: local installation of the open-source Access-to-Memory software; used for archival description and online access to digital holdings.

- Networked servers, virtual machines (VMs) and disk space to run software and provide storage.

Responsibility for managing the system is shared:

- The Archives manages all content in the system.

- The university's IT Services department provisions the servers, VMs and disk space; manages and secures the network; and provides backup services.

- The Archives contracts with the lead developer of Archivematica and AtoM ([Artefactual Systems Inc.](#)) to install, maintain and upgrade those softwares, liaising with SFU IT as necessary. As resources allow, the Archives from time to time sponsors software development work by Artefactual on those applications.

3. WORKFLOW

The standard procedure for transfer of digital materials to the Archives uses transmittal over the SFU network.

1. Producers (university departments or private donors) complete the online Transfer Request form (created with AEM).

2. An archivist exports the form data from AEM and uploads it to the AIS database, which assigns a unique transfer number.

3. An archivist approves or rejects the request.

4. If approved, the producer uses SFU MoveIt to package the records for transfer as zipped bags, using the transfer number assigned in step 2.

5. The producer uploads the transfer package to a designated folder on SFU Vault.

6. An archivist validates the transfer, using Checksummertime to verify package checksum and manual inspection of transfer contents as needed.

7. If approved, an archivist registers the accession in the AIS database and uses Archivematica to ingest the transfer package into repository backlog.

8. An archivist communicates to the producer that the transfer is complete; the producer should delete the records from their active systems.

9. At a later date, an archivist retrieves the materials from backlog, arranges and describes them to create an Archival Information Package (AIP) for long-term preservation and storage; and a Dissemination Information Package (DIP) for public access via SFU AtoM.

In addition, the Archives has developed a procedure for export and transfer of email that is still in testing and has not yet been formally implemented. From time to time, the Archives also receives transfers via other methods (e.g. removable media, hard drives, disks), but has not instituted a regular procedures for these scenarios.

4. PROBLEMS

There are a number of gaps or difficulties in the current processes. This section identifies existing problems and articulates requirements for addressing them.

4.1 Capture of transfer metadata

Producers currently enter 17 pieces of information into the online Transfer Request webform maintained in Adobe Experience Manager (AEM). This represents useful contextual metadata for the records transferred. An archivist exports it from AEM and imports it into the Archives' AIS database. But the data does not find its way into the transfer package itself; nor does it get included in the eventual AIP.

Requirement 1
The Archives should include transfer metadata in the transfer package and / or AIP.

4.2 Retention of transfer metadata

An archivist deletes the webform data from AEM after it has been exported and uploaded to the AIS. This renders the data inaccessible to the ordinary AEM user (including Archives' staff), but in fact the system does not irretrievably delete it. The Archives has no control over AEM (it is maintained by SFU IT Services); so it has no control over the long-term retention or disposal of this data. While it is unlikely that producers will include confidential or sensitive personal information in the Transfer Request form, it remains a possibility.

Requirement 2
The Archives should have full control over disposal of transfer metadata.

4.3 Transfer via SFU Vault

An archivist creates a transfer folder on SFU Vault (a university-wide implementation of OwnCloud software) and provides a link to the producer for uploading the transfer package created by SFU MoveIt. This works well for internal, SFU transfers where the producer has installed OwnCloud on their local computer; installation is a simple and supported process with SFU Vault. However, non-SFU users who do not have access to a desktop version of OwnCloud must rely instead on the web version to access the transfer folder.

With the web version, users can only upload single files, not directories with sub-folder and files. But SFU MoveIt currently generates the transfer package as a folder containing two files. The package cannot be directly uploaded to SFU Vault via the web interface; each file must be uploaded separately. This is not a huge barrier, but it does add complications to the process for both private donors who wish to transfer digital materials and the archivist processing the transfer.

Requirement 3
The Archives should support the same functionality for both internal SFU and external non-SFU transfers.

4.4 Post-transfer validation – checksums

The current approach validates the checksum of the zipped bag as a whole. The pre-transfer checksum value is stored in a custom file (`meta.txt`) that SFU MoveIt creates outsides the bag. An archivist compares the value in the `meta` file against the post-transfer checksum. This can be done manually through a Terminal line command or with a Mac Automator app (Checksummertime) custom-built to automate this through a drag-and-drop interface.

This approach works fine as-is, but there are several limitations. The Archives is not performing a full bag validation, which would check the structure of the bag as well as the checksums of the individual files; the current methods employed cannot identify the problematic file(s) in the event of a validation fail. Checksummertime is

also a Mac-only solution. But the Archives also has a Linux machine (currently running BitCurator) that ideally could form part of a digital processing workstation.

Requirement 4
The Archives should do full bag validation with a tool that can run on different operating systems.

4.5 Post-transfer validation – contents

An archivist manually inspects transfer contents to ensure at a minimum that contents match the description on the Transfer Request form that was approved. The process is currently ad hoc and variable; with large transfers it has the potential to be time-consuming and / or unreliable.

Requirement 5
The Archives should articulate clear content validation criteria and look for utilities to automate some of the validation tasks.

4.6 Post-transfer accession metadata

Following successful validation, an archivist accessions the transfer in the Archives' AIS database. This creates some additional metadata about the transfer: accession number, descriptive information, validation information. The data is not included in the transfer package that is ingested to Archivematica and therefore is not included in the eventual AIP.

Requirement 6
The Archives should include accession metadata in the transfer package and / or AIP.

4.7 Archivematica backlog

An archivist completes a transfer by processing the materials in Archivematica to the backlog stage only.

Archival arrangement and description of digital transfers is a time-intensive process; most transfers are not (and for the foreseeable future will not be) arranged and described immediately following receipt. Judging by comparable paper backlog, materials may remain in digital backlog for considerable time periods.

The current approach has the effect of making Archivematica pipelines into storage depots. This is not their intended purpose, and over time it will strain the computing resources pipelines need for processing. Objects in backlog do not receive full preservation treatment (e.g. normalization); long-term storage in backlog leaves them at-risk. Access and delivery of materials in backlog is a cumbersome process.

Requirement 7
The Archives should not rely on Archivematica pipelines for long-term storage of unprocessed materials in backlog.

** Note that as January 2019, the Archives is working with Artefactual Systems Inc, lead developers of Archivematica, to tackle this issue by modifying the way Archivematica manages transfer backlogs.

4.8 Other transfer scenarios

The Archives has developed a procedure for "normal" transfers, but other transfer processes will be required from time to time. Currently these are largely handled ad hoc. While it is not possible to foresee all circumstances, the Archives should identify other transfer scenarios that may reasonably be expected and develop procedures for managing them.

Requirement 8
The Archives should identify a range of transfer scenarios and develop procedures for each.

5. SOLUTIONS

The other papers in this report series will take up these problems and propose various solutions.

- Paper 2 (*Transfer Methods*) addresses requirement 8.

- Paper 3 (*Transfer Utilities*) deals with requirements 1–4 and 6.

- Paper 4 (*Transfer Validation*) looks at issues relating to requirement 5.

- As noted above, requirement 7 (transfer backlog) is currently being addressed through an Archivematica development project with Artefactual.