A Persistent Digital Collections Strategy for UBC Library

Bronwen Sprout and Sarah Romkey
University of British Columbia

Abstract

In early 2011, UBC Library began work on creating a digital preservation strategy in collaboration with Vancouver-based Artefactual Systems. UBC Library and Artefactual ran a number of pilot projects to ensure that the strategy is pragmatic, tested, and involves proven technical solutions and business processes that work in our environment and towards our goals. The persistent digital collections strategy developed for UBC Library consists of using the open-source Archivematica digital preservation system to provide preservation functionality for the Library's digitized and born-digital holdings. The strategy identifies the software requirements, existing and new system components, staffing and business processes that can be implemented to establish operational digital preservation systems and processes by the end of 2012. The paper will discuss the strategy generally and cover three areas of implementation in greater detail: UBC Library's Rare Books and Special Collections, cIRcle, our DSpace-based institutional repository, and CONTENTdm, UBC Library's access system for digitized objects.

Authors

Bronwen Sprout is the Digital Initiatives Coordinator at UBC Library. She coordinates the development and management of the Library's locally created digital collections, including digital preservation activities. She has over ten years' experience working on digitization projects and managing digital collections. Prior to joining UBC Library in 2005 as Digital Initiatives Librarian, Bronwen worked as a cybrarian at a dot-com, and as a consultant for small businesses and non-profits, assessing information needs and implementing a variety of web-based solutions.

Sarah Romkey graduated from University of British Columbia's joint Masters in Archival Studies and Masters in Library and Information Studies program in 2008. She has since that time worked in UBC Library's Rare Books and Special Collections division initially as the Librarian and Archivist for the Chung Collection and since 2009 as the Rare Books and Special Collections Archivist. Sarah is responsible in this role for acquiring, providing access to and preserving the archives of individuals and organizations external to UBC.

The University of British Columbia (UBC) Library's digital preservation strategy supports the Library's strategic focus on managing collections in a digital context and reinforces the University's commitment to make UBC research accessible in open access repositories. In 2010, the Library's newly launched strategic plan guided the development of a Digital Initiatives unit and related decisions to hire a Director of Digital Initiatives and to outfit and staff a digitization centre. The Digital Initiatives unit is a key part of the Library's effort to adapt to the evolving needs of faculty and students and to support teaching, research and learning at UBC. The unit's goal is to create sustainable, world-class programs and

processes to make the collections and research at UBC available to the world and to ensure the authentic, long-term preservation of these digital holdings for the future. As the new unit became established, one of the obvious functional gaps was a preservation system for digital master files. As we began work on a digital preservation strategy however, it quickly became apparent that there were crucial digital preservation gaps with other Library-created and/or managed material as well.

This paper will address the pilot and early stages of the implementation of UBC Library's digital preservation strategy. The Library has been working on this strategy with Artefactual Systems Inc, the Vancouver-area company behind the open source Archivematica and ICA-AtoM software, since early 2011. The first year of the project was devoted to developing and testing the preservation strategy and included installation and testing of the Library's Archivematica instance and the development of copying workflows and procedures. The second and current year of the project is the implementation phase; we hope to be able to offer preservation services by the end of 2012.

Archivematica is one of the key components of our persistent digital collections strategy. The software "uses a micro services approach to provide an integrated suite of free and open-source tools that allows users to process digital objects from ingest to access in compliance with the OAIS model"[1]. Archivematica will provide preservation functionality for the Library's digitized and born-digital material, managing functions and activities such as checksums, virus checking, preservation metadata, and file format normalization. It will interface with our existing systems (DSpace, CONTENTdm, ICA-AtoM) and we will use existing Library staff resources to integrate digital preservation activities into our workflows. Although Archivematica is a key component, our preservation strategy is not solely focused on this software. The strategy also includes software requirements, existing and new system components, staffing and business processes.

Artefactual's initial task in establishing our strategy was to carry out was a gap analysis. The gaps in the Library's digital preservation activities were largely identified using the ISO-OAIS reference model[2]. The consultants looked at the digital material we were creating or managing locally and, referencing OAIS, asked: "what is not being adequately preserved?" Following that, several areas were identified for pilot projects, including:

1. Rare Books and Special Collections and University Archives – both pilot projects involve born digital records and are working with legacy and newly acquired material
2. cIRcle, our DSpace-based institutional repository – contains more than 40,000 items, mainly theses and dissertations but also a variety of research and teaching materials
3. Digitization projects in Digital Initiatives – CONTENTdm is UBC Library's access system for digitized objects and currently contains about 80 collections and over 200,000 images

Once the gaps were identified, a project team was struck. The team includes Artefactual staff as well as representatives from the areas of the Library identified and/or affected by the pilot projects: Digital Initiatives, Library Systems and Technology, University Archives, and Rare Books and Special Collections. A student assistant from the School of Library, Archival and Information Studies was also a part of the project team and assisted hugely with the testing. The team meets monthly and UBC Library

---

[1] "What is Archivematica", accessed August 16, 2012, https://www.archivematica.org/wiki/Main_Page
[2] "Open Archival Information System", accessed August 16, 2012, http://en.wikipedia.org/wiki/Open_Archival_Information_System

members work closely with Artefactual to test the software and iterate the development of workflows that will integrate with our local practices.

One of the major outcomes of our digital preservation program will be the ability for Rare Books and Special Collections (RBSC) to archive born digital material. The mandate of this division of the UBC Library includes the collection of archival material from private creators about the various aspects of life and industry in British Columbia. As of 2012 the division has in its holdings almost 700 archival fonds created by a wide variety of people and organizations, dating from the gold rush era to the present day. Since the time when digital technology became ubiquitous in homes and offices, RBSC has acquired born-digital records on external media (primarily floppy disks, CD ROMs and DVDs) only in the context of the media arriving unexpectedly in primarily analogue acquisitions. While this media was presumed to be of archival value and kept with the fonds, one could say that the digital media was not processed in the sense that no attempt to preserve the digital records separately from the physical media was made.

The University Archives division of UBC Library, responsible primarily for university and university-related records, had similarly begun to receive digital media, or requests to acquire digital media, from university departments. In 2011, the archivists from both divisions had the opportunity to participate in UBC Library's persistent digital collections strategy pilot project. The aim for these two divisions in the pilot project has been to develop a method for both processing legacy media in existing collections, and acquiring and processing born-digital material from new or ongoing collections.

After identifying collections in RBSC for which legacy digital media existed, the project team prepared to test the preservation of the files. Appropriate hardware was procured, including a 3.5 inch floppy disk drive and, with more difficulty, a 5.25 inch floppy disk drive. A basic workflow was developed which involved creating a Submission Information Package (SIP) from the files on the external media and feeding it through Archivematica, which would create an Archival Information Package (AIP) and a Dissemination Information Package (DIP) for access in ICA-AtoM archival description software.

Unsurprisingly, technological challenges have abounded during testing of external media, particularly in regards to floppy disks. A number of the disks tested were found to be unreadable, and possibly have been unreadable since their initial accession to the archives. The 5.25 inch floppy drive has proven to be particularly challenging to operate. A number of files which could be copied from the disks have been in outdated formats for which the proprietary software no longer exists. Some optical disks were found to be created from proprietary sources (such as photo finishing stores) which also required proprietary software to open or copy the files, even if the files themselves were in commonly used formats. At this point, we have sufficient experience processing external media ourselves to feel comfortable approaching an external company to take on some of the processing. Our pilot testing has allowed us opportunity to create documentation for several different use cases, including transferring media from optical disks, 3.5 and 5.25 inch floppy drives, and external drives, which we can use to guide future work as well as our negotiations with vendors.

Processing legacy digital media has also raised a number of intellectual problems. One is the arrangement of the digital media into series or files within the collection. Although archival theory and practice tells us that the physical container of records may or may not reflect the intellectual arrangement of the records, the processing archivists who had accessioned these records were limited in their options for arranging the media into series, being unable to separate the records from the media itself. Thus it is not uncommon in our finding aids to find series titled, "Electronic media" or similar. Actual reading of the records through a file viewer often reveals evidence of original order (the principle by which an archivist would ordinarily arrange records into series and files) such as folders and naming conventions

applied by the creator. Furthermore, it may appear that records on the digital media share "archival bond[3]" with analogue records in the same collection by virtue of their association with the same activity of transaction of the creator. Does one re-arrange the collection into series based on this newly discovered evidence, or maintain the arrangement provided by the first processing archivist? At this juncture, decisions are being made on a case-by-case basis.

It is one of an archivist's core duties to *select* records for permanent retention within a collection. Similar to the challenge of arrangement described above, the processing archivists of the past did not have an effective way of selecting the records contained on digital media, lacking a method for opening the files and inspecting their content. They were therefore only able to use what metadata existed on the label or container for the media and make assumptions about the content. While some disks have yielded archival documents such as book manuscripts and digital photographs of which no other copies are known to be extant, other disks have yielded routine administrative records that would not have normally met an archivist's criteria for retention, lacking sufficient evidential or informational value to be of use to researchers.

Finally, an examination of the rights transferred with the digital material must occur in order to determine the institution's ability to preserve and provide access to the material. While UBC Library's standard donation form for archival material transfers copyright to the institution, understandably not all archival creators are comfortable with such an arrangement and so they are accommodated by transferring the legal ownership of the physical material while maintaining their intellectual property rights. Because this legacy digital media was acquired in a primarily analogue context, no licensing was written into the agreements to allow online access or the creation/migration of digital copies for preservation reasons.

While testing was still occurring with the legacy media from RBSC and University Archives collections, it seemed that sufficient lessons had been learned to pursue what could be considered our "first" born-digital archival acquisition; first in the sense that we proactively sought a primarily born-digital acquisition and would therefore have greater control over many aspects of the acquisition process. RBSC had been approached by a photo-journalist who desired to donate his own born-digital photographs from the past seven years. The acquisition was from our perspective a desirable test case because first, the subject matter of the photographs (fishing boats and fishing practices) complements other collections in our holdings, and second, the donor has been able and willing to give answers to our questions regarding the creation of these images.

The information gathering stage was important in order to address the same issues we had faced with our legacy media project. A questionnaire for a donor interview was developed, adapted from *Born Digital Collections: An Inter-Institutional Model for Stewardship (AIMS) survey for Personal Digital Archives*[4] and the *Paradigm records survey*[5] published by the Bodleian Library. While the AIMS and Paradigm surveys provided ample opportunity to ask the donor about his technological and organizational processes, additional questions were added to address questions of rights and access: did the donor wish to retain his copyright? If so, what measures was he willing to allow UBC Library to take in order to preserve and provide access to his born digital records? Were there any rights attached to the records that

---

[3] Archival bond is defined as "The network of relationships that each record has with the records belonging in the same archival aggregation." From "The InterPARES Project Terminology Database", accessed August 23, 2012, http://www.interpares.org/ip3/ip3_terminology_db.cfm?letter=a&term=6
[4] "AIMS Born Digital Collections: an Inter-Institutional Model for Stewardship", accessed August 31, 2012, http://www2.lib.virginia.edu/aims/
[5] "Welcome to paradigm", accessed August 31, 2012, http://www.paradigm.ac.uk/

he did not have the ability to transfer or license to the institution? (See Appendix A for our survey instrument).

The donor's answers to the questionnaire allowed us to prepare both technologically and intellectually for the transfer of the material. From a technological point of view, the feasibility and appropriateness of using digital forensics tools and processes was explored. The donor wished to transfer the records on an external hard drive, which he had purchased explicitly for the purpose of donating his digital records. Initially we felt that using a write blocker, a piece of hardware used to prevent the inadvertent alteration or damage of files and documents when external devices are used to transfer files, would be an appropriate safe-guard. However, it became necessary to delete files from the external drive before the transfer could take place due to an accidental inclusion of files not intended for donation, and so the use of a write blocker became an inappropriate step[6]. This step was also deemed not *as* necessary as perhaps would be appropriate in other use cases: in the future, we anticipate that entire working hard drives (as opposed to copies of digital records saved to external media) will be imaged in order to preserve the records. In that use case a write blocker would be necessary to ensure the records, being the original records created by the donor, will be imaged authentically.

Based on our experience processing legacy media and the external drive we began to perceive there to be different *levels* of digital preservation that might be appropriate to guide future acquisitions:

Level of Preservation	Use case	Minimum tools
Level 1	Imaging files from working hard drives, directly from creator	Write blocker, file synchronization software (e.g. Grsync)
Level 2	Imaging/copying files from external media	File synchronization software (e.g. Grsync), write blocker if possible
Level 3	Preserving external media only	Hardware necessary to open media (floppy drive, optical drive, etc)

"Level 1" preservation would entail imaging of working hard drives, when a creator is prepared for the archivists to examine all of their digital records and image files directly from their working computer. Digital forensics tools such as write blockers would be necessary for this level of preservation to ensure the authenticity of the original files is maintained. "Level 2" preservation would entail making authentic copies from external media such as floppy or optical disks and external hard drives. Software tools such as Grsync[7] and functions such as checksums will be used to ensure the copied files are authentic copies of the files provided on the external media. Ideally, a write blocker would be used, but would not be considered as necessary as in Level 1. "Level 3" preservation may be a last resort to maintain bit-level

---

[6] Using a write blocker and making a forensic image of the hard drive contents would have inevitably meant including the files which the donor requested us to delete. We decided that this was not acceptable from a donor-relations point of view.
[7] "Grsync: Home", accessed August 31, 2012, http://www.opbyte.it/grsync/

records on external media that cannot be copied authentically due to technological problems. We have developed use case documentation for acquisitions made through external media which include the use of a write blocker, even though it has not been used for this particular pilot test.

To ensure that we have the necessary rights or licenses to preserve and provide access to the donor's material, we developed a donor agreement form specific to the issues presented by born-digital material. It was our initial hope that our existing donation agreement form would suffice with some alteration. Our realization though was that born-digital acquisitions go beyond the realm of physical, and even intellectual, transfer of material in the analogue world; what is really needed is the ability to gain a license for the rights needed to make copies of records for preservation purposes, migrate the records to new formats now and in the future, and provide access to users either on the internet or through a computer terminal in our reading room. Once again, AIMS and Paradigm documentation provided models, which were combined and modified with additions from a non-exclusive digitization and distribution agreement developed by UBC Library for managing rights associated with digitization projects.[8]

At the time of writing, we are at the stage of using Grsync to copy files from the donor's external hard drive, with the intention of running the acquisition through Archivematica in the fall of 2012. One question that remains to be explored in this pilot project is the extraction of creator-embedded metadata for use in the access system. The creator of the records meticulously adds geographical and subject indexing terms as well as brief descriptions to many of his photographs, which would be helpful for access purposes if they can be re-purposed.

Whereas digital acquisitions are still a fairly minor aspect of RBSC and Archives' holdings, cIRcle, UBC's institutional repository, is solely concerned with digital material. And whereas processing legacy digital media raised a number of intellectual questions, developing the Archivematica integration with DSpace and CONTENTdm has so far raised mainly technical issues.

cIRcle is a DSpace-based open access repository which currently contains over 40,000 items, about half of which are theses and dissertations. In addition to the comprehensive retrospective digitization of theses dating back to 1919, UBC has offered optional online submission of electronic theses and dissertations (ETDs) since 2008 and mandatory online submission for the last year. Thus, the ETDs in cIRcle are considered the University's official copies and were a major driver for the development of a digital preservation strategy for the repository. Because we use a distributed submission model to add new content to the repository, in terms of our digital preservation plan, DSpace will continue to function as the submission and access tool and full preservation management will take place separately in Archivematica. Archivematica will accept submissions from DSpace and preserve them without creating access versions. Submission to Archivematica will take place when submission is made to DSpace, without affecting the user interface. The identifier field (handle) will link the DSpace access record and the Archivematica preservation version. Archivematica 0.8 was designed to recognize DSpace exports and parse them into the Archivematica preservation version.[9]

Though it is not uncommon to refer to DSpace as a preservation system, the software does not have full preservation functionality. Archivematica performs many preservation actions on the DSpace files: it unpackages them, verifies their checksums, assigns unique universal identifiers, checks for viruses,

---

[8] Readers may also be interested in consulting *Research Library Issues (RLI) no. 279*, recently published by the Association of Research Libraries, which discusses digitization of analogue collections and provides a model gift agreement for mixed intellectual property rights (http://www.arl.org/news/pr/rli279-7aug12.shtml).
[9] "DSpace exports", accessed August 16, 2012, https://www.archivematica.org/wiki/DSpace_exports

identifies and validates formats, extracts technical metadata and normalizes the files to preservation formats. All metadata are captured as fully standards compliant PREMIS –METS. At this point, a couple of items regarding the DSpace integration are incomplete; our next steps include a retrospective migration of DSpace content into Archivematica and the development of an automated export/ingest trigger for newly-submitted items (so that items approved for submission in cIRcle are automatically exported to Archivematica).

A companion access system to cIRcle, CONTENTdm is UBC Library's access system for locally digitized objects. The digital collections available in our instance of CONTENTdm date back over 10 years to projects that were created long before the development of the new Digital Initiatives unit or the consideration of digital preservation. Collectively they document a diverse range of people and places, activities and events, and serve as a resource for students, historians, genealogists, and other researchers. The content of these collections is largely drawn from RBSC and Archives and ranges from rare maps and books to historical newspapers to images from British Columbia's forestry and fishing industries. A large part of the collection consists of images which present a visual record of UBC's growth and development over the past century. Like most sites using CONTENTdm, we digitize analogue materials, store the master versions of the digitized objects on network drives and upload the access copies, along with descriptive metadata, to CONTENTdm for public access. In our case, the master files were not adequately linked to our access copies (a filename based on the location of the physical item loosely links the two versions, but is prone to human error). And although we do participate in a Private LOCKSS Network (PLN) that has the capacity to store our CONTENTdm collections, it will not preserve master copies because we do not make them available online for LOCKSS to crawl. Therefore, one of our pilot projects includes the development of an API to link Archivematica and CONTENTdm.

We plan to ingest master digitized objects into Archivematica which will generate the access version for automatic upload into CONTENTdm. The SIP is ingested into Archivematica and when it has moved through to become an AIP the user is given the option of using CONTENTdm to make the DIP available to end users. Two methods for submitting the item to CONTENTdm are available: a one-click import into CONTENTdm as well as the option to save the item as a set of files suitable for importing into CONTENTdm using the Project Client (so Optical Character Recognition can be applied using the Project Client, etc.). Work is currently underway on a bulk DIP upload with metadata attached to each object via a user-supplied CSV file. With either method the regular CONTENTdm approval/indexing needs to be applied before the item appears to end users. Although the option exists to add metadata in Archivematica, most of the descriptive metadata for uploaded objects will be created and edited in CONTENTdm because it better facilitates batch processing. (Descriptive metatada will be added to the AIP by syncing the DIP in CONTENTdm and the AIP in Archivematica.)[10] This will change our workflow considerably as all material will have to go through Archivematica to be added to CONTENTdm. However, we plan to add this step to existing workflows and build the work into existing job descriptions rather than hiring or designating specific staff to take on the role of digital preservation.

Mid-way through 2012, many implementation tasks remain. Archivematica testing is ongoing as are the DSpace and CONTENTdm integrations. A large retrospective project to add DSpace and CONTENTdm content to Archivematica is necessary. We are interested in undertaking a pilot with the Council of Prairie and Pacific University Libraries (COPPUL) PLN we participate in, in which objects ingested into Archivematica would be uploaded to the PLN for distributed, geo-remote storage. We also

---

[10] "CONTENTdm integration", accessed August 16, 2012,
http://archivematica.org/wiki/index.php?title=CONTENTdm_integration

plan to undertake a research data pilot project using Archivematica to ingest and preserve one or more datasets. We are excited about the potential development of a unified discovery interface for UBC Library digital collections based on the contents of our Archivematica index. Additional effort is needed on our website archiving plan.[11] And finally towards the end of 2012 we plan to work with Artefactual to undertake a self-audit based on the ISO Standard for Trustworthy Digital Repositories (an international standard for auditing compliance with the OAIS reference model).[12]

Despite all of the work underway and still to come, we feel we have made tremendous progress over the past year and a half towards implementing a comprehensive digital preservation program. We have a solid plan for our locally digitized material and for the content of our institutional repository in terms of the Archivematica integration with our two primary access systems. Equally as important, we have a greater comfort with and understanding of the challenges around archiving born digital material as the archivists at RBSC and University Archives are anticipating that most, if not all acquisitions in the future will include a born-digital component. Currently, most newly acquired archival collections to these divisions are still primarily analogue, but frequently include external digital media that the donor either created in their normal business or creative practices, or made purposefully with transfer to the archives in mind. This pilot project has helped to solidify what the normal practices and procedures will be, so that digital acquisitions can be more or less systematic, and fewer decisions made on a case-by-case basis. During the time of our contract with Artefactual, the software was still in alpha. This meant that UBC Library had the opportunity to contribute to the development of the full release (expected for early 2013). We are greatly anticipating this release and are hopeful that our contributions to the development of Archivematica will benefit other institutions as well.

---

[11] Part of our digital preservation strategy is the ability to capture and preserve websites (mainly UBC sites but possibly external sites as well). We tested the open source tool Heritrix, which was developed by the Internet Archive. The software crawls selected sites and stores them as Arc or WARC files. Heritrix comes with a web-based admin module and sites are rendered by Wayback Machine. We found it fairly easy to use but it does require staff time to configure. A similar service called Archive-It is also available from the Internet Archive which includes hosting and remote storage, and this is likely the option we will implement. A major task still outstanding is to assess our library and university websites and decide on the parameters for the crawl.

[12] "TRAC/TDR Checklist", accessed August 16, 2012, http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying-0

Appendix A: Donor Survey Instrument

Adapted by the City of Vancouver Archives from *Born Digital Collections: An Inter-Institutional Model for Stewardship (AIMS) survey for Personal Digital Archives* and the Paradigm records survey published by the Bodleian Library and John Rylands University Library, and further adapted by UBC Library.

The survey is designed to be used by the archivist(s) during a face-to-face or phone interview with a donor/creator of born-digital records being considered for acquisition.

Donor name:
Donor affiliation (if any):
Archivist name:

1.   Digital Material Creation
1.1.   Are you the only person responsible for creating your digital files?
         1.1.1. If not, who else is involved and what is their role? (for each one list last, first, role [author, editor, secretary, proofreader, admin, ...])
1.2.   Do you maintain digital files created by others?
         1.2.1. If yes, how do you separate your files from files created by others?
1.3.   Do you separate your personal files from your work files?
1.4.   What is the earliest creation date (roughly) of your digital files?
1.5.   What is the latest creation date (roughly) of your digital files?
1.6   What software and computer system was used to create these files? Did this change over time?
1.7   What kind of camera did you use to shoot these images? Did this change over time?

2.   Varieties of Digital Material
2.1.   What kinds of materials are you donating today? Are they in specific formats (TIFF, JPEG)?
2.2.   Do you create files in both digital and paper formats?
         2.2.1. If yes, which files or file types?
         2.2.2. Do you also have prints of your photographs that you will donate?
2.3.   What do you consider the first, best copy of the file(s)? (this could differ by content type)
2.4.   Roughly how much of each type exist? (in MB, GB, etc)
2.5.   Do you distinguish in the files or the filing system between published, non published and re published photographs?
2.6.   Do you think you will want to reuse or republish the photographs that you donate?
2.7.   Do you have signed rights or clearances from subjects in your photographs?

3.   Digital Material Organization
3.1.   How are digital files named?
3.2.   Is some kind of version control used?
         3.2.1. If yes, list examples
3.3.   How are digital files organized? Can you give a brief summary about the organization of your digital files? (obtain classification scheme if available)
         3.3.1. Are digital files destroyed in regular intervals?
                3.3.1.1. What is the interval? (obtain retention schedule if available)
3.4.   Do you use more than one computer? (e.g. office desktop, office laptop, home desktop, etc)
         3.4.1. If yes, how do you synchronize files between different computers? (server, etc)
3.5.   Please explain fully your process of adding or creating metadata for your photographs.
         3.5.1. Which programs do you use to create metadata in them?
         3.5.2. Do you use templates?

3.5.2.1. If so, do you use them for one group, all, some?

3.5.2.2. Is there one template or many?

3.5.2.3. Can we get a printout of your templates?

3.5.3. Do you use the same metadata for all the photos, or do you customize it for all, some, etc.?

3.6.   Do you wish to retain the copyright for the material you are donating?

3.6.1. What rights are you able/willing to assign to UBC Library (access, migration and copying for preservation purposes, etc)

3.6.2. Are there any photographs for which the copyright has been assigned to someone else (e.g. a magazine, etc).

4.   Digital Photographs
4.1.   Which do you consider the original (jpg, raw, tif...?)
4.2.   Do you use geocoding?

5.   Mobile devices and tablets
5.1.   Do you use computing devices besides your desktop computer to create this material? (e.g. Blackberries, iPhone, iPad or other tablet, Android phone, etc.).
5.2.   Do you store photographs on these other devices?

6.   Email
6.1.   Do you have multiple email accounts?
6.2.   Which email program(s) / service(s) are you using? (e.g. Email program provided by your work place, Outlook, Mac Mail, Hotmail, Gmail, Yahoo! Mail, etc.)
6.3.   How is email organized? (e.g. in self-created email folders, etc.)
6.4.   How is email saved? (e.g. untouched in the email program, a copy in your PC, printed out in paper, etc.)
6.5.   Are email and paper correspondence managed together or separately?
6.6.   Do you use address books?

6.6.1. Please list address books

7.   Digital Files Storage / Backup
7.1   Do you have a backup routine for your files / emails?
7.2   What media are used for backup files? (e.g. optical disk, hard disk, file server, web based backup service such as SugarSync., cloud service etc.)
7.3   How recently have you changed computers? Do you transfer files in your old computer to your new computer?

7.3.1 If yes, what types of files are transferred?

7.3.2 Did you encounter any problems in transferring the files?

7.3.3 Please list problems encountered.
7.4   Do you keep your old computers?

7.4.1 Please list details. [computer name(s), operating system(s), version(s)]
7.5   Have you ever experienced a serious hardware failure (e.g. hard-drive crash, loss of files, accidental deletion)?

7.5.1 If yes, were the files in the affected computer recovered?

7.5.2 If no, would like us to attempt to recover the files as part of your donation?

7.5.2.1 If yes, which ones?
7.6   Are any digital files stored in unusual storage media? (e.g. punch cards, 8 inch. floppy diskettes, etc.)

7.6.1 Please list media types.

8. Work Habits

8.1   Can you tell us about your work habits of using computers / mobile device? (e.g. work online, work offline, use mobile, etc.)

8.2   Do you share computer with other people?

      8.2.1 If yes, how are files created by different people separated?

8.3   Since a visual representation of working space may provide researchers additional information about your works, do you mind we take photos of your computer with surrounding space?

11. Privacy and security

11.1   Are some digital file types of a sensitive nature? (e.g. tax records, medical records, peer-review comments, letters of recommendation, student records, etc.)

      11.1.1 Please list categories of files and their restrictions.

11.2   Are there files that you would want destroyed?

      11.2.1 If yes, please provide details so that we can act upon when we encounter such files when processing your files.

11.3   Do any digital files require passwords?

11.4   Where are user names and passwords kept?

      11.4.1 What service / software are used to save them?

11.5   Do you use digital watermarks?

      11.5.1 Please list files and watermark rationale.

12. File Transfer Arrangement

12.1   Do you want to delete any files / re-organize the files before the transfer?

12.2   Are there files you would like to transfer to us later?

      12.2.1 Which files?

      12.2.2 When?

12.3   Can we take the original computer(s) or storage media?

      12.3.1 If yes, do you want the original media back once we've processed the donation? If no, we will destroy the original media once the donation is processed.

      12.3.3 What storage media will you deliver? [original hard drive, copy on external drive, dvd, cd, etc]

      12.3.4 If external drive, how is it formatted? (operating system, version)

12.5   Can we make a copy of the entire hard drives or just specific file folders?

12.6   Were these hard drives used for anything else, for example personal documents?

Digital Material Survey (Part II)

Note: This part of the survey is designed to be filled out by digital archivists regarding technical details of the tools used to create digital material.

1. Hardware

1.1 List the hardware configurations of each computers / mobile device. (e.g. manufacturer, model no, cpu, ram, hard drive capacity, video card, etc.)

1.2 Find out if the computers have USB ports or CD writers which could be used to copy the digital files.

2. Software

2.1 List the operating system and other system software with version number, installed in all the hardware.

2.2 Check if system date and time are set correctly. List the time zone used, if any.

2.3 With the help of the donor, list the main application software, with version no., used to create digital files.

2.4 If Microsoft Office is used, find out if the "User Name" field is set to the name of the donor. Find out similar setting for other main application software used.

3. Internet Access

3.1 Find out if the digital archivist can use the Internet access in the donor's office using the digital archivist's portable computer?

4. Networking

4.1 With the help of the donor, confirm if the computer is connected to file servers. Confirm if the donor save files in the file server. How much file server space is used by the donor?

5. Security

5.1 With the help of the donor, confirm if login is required to access desktop computers / mobile devices?

5.2 With the help of the donor, confirm if a digital certificate is used by the donor to login / sign digital files / encrypt digital files?

5.3 With the help of the donor, confirm if digital files are encrypted?

6. Comments (per section and per survey)