Winter April 25, 2014

Digitization At Western

Nicholas Rapp

Digitization At Western

By: Nicholas Rapp

MLIS Candidate
Faculty Of Information & Media Studies
Western University

April 25, 2014

Digitization At Western Libraries

* A new digitization subcommittee has been formed at Western and began meeting in April of 2014. At the time of writing, it is unknown how the current digitization policies and procedures will be affected by the establishment of this new subcommittee.

Introduction

I am a MLIS candidate with the Faculty of Information & Media Studies at Western University. I am currently completing a co-op with the Library Information Resource Management division of Western Libraries. This report will document various practices, challenges and considerations which are involved in the digitization process at Western Libraries and at various other institutions around the world.

Digitization is a major part of the work which has not only been completed but is also ongoing at Western Libraries. Some of the previous projects have included the digitization of the Western Mustangs' football footage and the Occidentalia yearbooks. The current and ongoing digitization projects include the aerial photographs of London and the surrounding region, the UWO Medical Journal, London Free Press materials, and Canadian Tire commercials. Each of the past and present digitization projects is unique and comes with various challenges and expectations.

Summary Of Challenges and Considerations

Even though digitization has become quite common throughout Western, and the rest of the academic community, there are many challenges that individuals and libraries face when they undertake these types of projects. A summary of the challenges are listed below, with 1-8 being examined in greater detail throughout the report. Even though Staffing and Funding are very important aspects to digitization, they are beyond the scope of this report and will need to be determined by the project managers on a case by case basis.

1. Types Of Materials

   Different types of materials will present unique challenges to digitization. The techniques used in scanning, displaying, and creation of metadata for theses and dissertations will not be the same as those for photographs and videos. As a result before one can proceed with the digitization process they must first examine what type of materials they have and the best options for handling those materials. Examples of some of the most common materials that have been and could potentially be digitized at Western are listed in Table 1.

Table 1 – Common types of materials for digitization

Type Of Materials	
Theses And Dissertations	Musical Scores
Videos	Presentations
Photographs	3D Data Sets

Geographical Information Systems (GIS) Data and Maps	

2. File Formats

   Each of the materials that are listed in Table 1 could potentially be formatted in several different ways. For example, a video file can be converted into AVI, MPEG, MOV or a variety of other formats. The type of format that a department chooses will largely be determined by how the file will ultimately be used. Some departments at Western prefer to use MPEG because of its accessibility but for preservation and long term access Library and Archives Canada recommends JPEG 2000. (File Format Guidelines, pg. 8). Some of the most widely used file formats are listed in Table 2.

Table 2 – Common File Formats For Digitization

File Formats	
PDF/Word/Other Text based formats	TIFF
AVI	MPEG, MPEG2
JPEG, JPEG 2000	.ppt
MPEG	NETCDF, HDF5, FITS (Scientific Data)

3. Hosting and Displaying Materials

   When undertaking a digitization project it is important to have a plan about where and how the final product will be hosted. There are a variety of options for hosting materials including institutional repositories, internally created web pages or partnerships with outside vendors.

4. Storage

   One of the biggest challenges when undertaking a digitization project at Western is the issue of storage. With the ever increasing size of digital files and large scale projects, storage will continue to be a factor. Some of the storage options that will be discussed throughout this report are the cloud, file servers, external hard drives, CD and DVDs and microfilm.

5. Infrastructure

   If a digitization project is to be completed in-house, there are several pieces of infrastructure that should be in place. This report will highlight a few aspects of infrastructure including scanners, docking stations and solid state drives.

6. Copyright

Copyright is a complex and important issue when dealing with digitization. This is especially true when handling materials that were not produced at Western. It is essential to understand what can be done to an object, how it can be displayed, and how the rights holder information is going to be presented.

7. Metadata

   Designing a metadata schema for a digitization project is essential. This includes choosing a metadata standard such as Dublin Core, MODS or EAD. Since each project is going to be unique it is also necessary to design a schema for both the collection and item levels of the digitized materials.

8. Preservation

   Even though this report is primarily concerned with digitization and the accessibility of digitized materials, it is still important to be aware of the preservation implications of these types of projects. Certain file formats and storage options are only appropriate for the short term and lack the capabilities to adequately preserve digital materials for the long term.

9. Staffing

   When undertaking a digitization project it is important to ensure that there is sufficient staffing to complete the project. This could include full time, part time, student, or volunteer staff. Even though staffing is essential to any digitization project, the issues surrounding this topic reach beyond the scope of this report and have to be examined internally by each department during the project planning process.

10. Funding

    As with staffing, it is necessary to consider funding before undertaking any digitization project but funding options go beyond the scope of this report. Each project is going to be unique and departments have to determine if each is financially feasible.

Hosting and Displaying Information – Institutional Repositories

One of the most common ways for academic institutions to host materials is through an institutional repository (IR). Most IRs are established to handle materials produced or sponsored at a particular institutional. At Western the IR, Scholarship@Western, is primarily intended to be a home for materials developed or sponsored by various departments throughout the university including theses, dissertations and journals. The software that is used to set up Scholarship@Western, Digital Commons, has the diversity to also host research data, image galleries and various types of multimedia. As a result the current IR could potentially be expanded to include other types of digitization projects.

Digital Special Collections (DISC)

- http://ir.lib.uwo.ca/disc/

One of the steps taken to expand the scope of Scholarship@Western was the creation of the Digital Special Collections (DISC). Unlike the other pages on the IR, DISC was established to host materials which were not originally produced at Western. This could include books, musical scores, journals, videos and photo galleries. As a result, DISC could also be used to host larger scale digitization projects that do not necessarily fall under any of the categories in Scholarship@Western.

Other Institutional Repositories

Along with Digital Commons there are many institutional repository software packages and hosted platforms available. In 2014, UNESCO published their Institutional Repository Software Comparison which was "divided into eleven categories to help librarians identify the features that are most important to building a successful institutional repository program at their institution." (3) The report compared five of the most widely used platforms: Digital Commons, DSpace, EPrints, Fedora and Islandora. Even though the report did not make any recommendations about which IR software would be the most effective, an examination of the findings did seem to indicate that Digital Commons had as many or more attributes to offer than the other platforms in most of the categories.

T-Space
- https://tspace.library.utoronto.ca/

Another example of a possible IR structure is the University of Toronto's T-Space. There was an attempt to reach the team at T-Space in order to discuss the repository but at the time of writing a response had not been received. T-Space is a build your own IR, which is designed exclusively for work that is "produced, submitted or sponsored by University of Toronto faculty." (Content Guidelines) T-Space could potentially be used as a guide for a future institutional repository at Western if the university chose to proceed in this fashion. Rather than having individual contributors uploading work independently, T-Space divides their content into communities. A T-Space community "is a unit at U of T that produces research, has a defined leader, has long term stability, and can assume responsibility for setting community policies." (Collection Policies) Each community is provided a specific amount of storage space based on the size of the community. For a list of community sizes and storage space, refer to Table 3.

Table 3 – T-Space Communities (Collection Policies)

T-Space Communities			
Community Size	Small	Medium	Large
Number of Submitters	<50	51-149	>150
Annual Storage	3 GB	6GB	10GB

Hosting And Displaying Materials – In-House

CulturePlex

- http://www.cultureplex.ca/

For projects that are to be kept in-house, one of the possible options at Western Libraries could be to partner with CulturePlex. Produced at Western, CulturePlex is a series of projects dealing with Cultural Complexity and Digital Humanities. Some of the information that is contained on the site is the background on the projects and those involved. CulturePlex does not host the entire project but instead provides links to the dedicated webpages for each. The site could be a potential place to display digitization projects from other departments. The advantages of using CulturePlex are that it is a unique way to display data and each project could have its own dedicated webpage. The potential disadvantages are that there currently is no link to Western Libraries and even though it may be an effective platform for some projects, it may not be appropriate for others.

Create A Dedicated UWO Digitization Project Website

One of the best options for displaying materials would be the creation of a web page dedicated to digitization projects at Western. This page could ideally be a central place for digitization at the university. Even if the projects were not hosted on this particular page, the site could be used to link all the projects together. As with CulturePlex, a dedicated digitization page could contain abstracts of the various projects and short profiles on who is working on each one. The page could also contain links to where the projects are hosted. This would give digitization a home at Western and a shared project among the library staff and other departments.

Hosting and Displaying Materials – Partnerships and Outsourcing

Scholars Portal

- http://www.scholarsportal.info/

One of the most logical avenues for a digitization project would be an expanded partnership with OCUL's Trusted Digital Repository, also known as Scholars Portal. The repository "seeks to provide a robust long-term preservation environment for all the materials in its collection." (Trusted Digital Repository, 2013) This site contains over 35 million scholarly articles, and more than 500,000 commercial and open access books, and a large collection of geospatial data.

Scholars Portal offers a wide variety of services for not only theses and dissertations but other types of research data. If it is determined that Western would not have the resources or capacity to host all their own data, Scholars Portal could continue to be a valuable partner.

Internet Archive

- https://archive.org/

The Internet Archive is already a partner with Western Libraries on several projects. They currently host past editions the Western yearbook, Occidentalia, and other materials. As well,

they are also involved in digitizing various items for Western Libraries. The Internet Archive could continue to be a partner for digitization projects.

Disadvantages of partnerships and outsourcing

Organizations such as The Internet Archive are not necessarily equipped to handle every project that Western may have. As well, sending the Internet Archive materials costs time and money. It is true that if this work was done in-house costs would still be involved. Even so, rare and fragile materials would not have to be transferred off-site as often. As well, if the work was conducted in-house, students and faculty would still be able to use the materials for research purposes while the digitization process was underway. For example the materials could be set aside a few hours a day for digitization and be available for research the rest of the time.

Hosting And Displaying Videos

Some projects may require the hosting and the displaying of various types of video files. The main options for hosting and displaying videos are streaming the files from websites such as YouTube or Vimeo, or allowing users to download or link to the file. The major advantages and disadvantages to both are listed in Table 4. It is recommended that the streaming option be used because the files cannot easily be copied, redistributed or manipulated.

Table 4 – Options for hosting and displaying videos

Streaming	Downloading Or Linking
Could be hosted on websites such as YouTube or Vimeo	Videos are available in a variety of formats
Much faster and more straightforward for viewing	Files can be stored on a server
Reduces the number of file formats	Do not have to rely on an outside source or foreign server
Do not need compatible software	Do not lose file or link if video host crashes or is shut down
Video can be added onto a server prior to upload	Copyright concerns - If users are able to download videos there is a risk the file could be copied, redistributed and manipulated.

In-House Infrastructure

Scanners

There are a variety of scanners available if the university decides to handle projects in-house. The Map and Data Centre have recommended a 56” scanner that could be shared among the various departments. Along with the maps that are currently unable to be scanned in-house, the machine could also be used for score sheets in the music department. The score sheets are just one example of what is possible with equipment such as this. This scanner would not only raise the status and profile of Western but could also be used in partnership with other institutions to

expand what is possible in various areas of research. McMaster University was able to increase their output and web presence greatly by investing in this technology. Another major advantage of having such equipment on-site is that rare and fragile material would not need to be transferred off-site as often. This would not only reduce the potential for damage but would also minimize the time spent transferring materials. In turn, students and faculty could then take advantage of materials even when the digitization process was underway. For example the materials could be scanned at certain hours of the day or week and made available for research at other points.

Docking Stations

As will be discussed later, the use of external hard drives is a very popular option for storage. Docking stations could provide an alternate form of combined storage and file transfer across various departments. Docking stations are being considered among some at Western as a result of them being less bulky than multiple external hard drives and require less technical support than an internal server. The docking stations traditionally allow multiple SATA hard drives to be connected to a computer through single USB connection. If compatible docking stations were set up throughout the various departments it could potentially streamline the transferring of files across Western Libraries and other departments.

Solid State Drives

Another potential option as an alternative to Hard Disk Drives (HDD) could be the use of Solid State Drives (SSD). These devices are currently cost prohibitive for many, with 1TB of storage potentially being upwards of $600 or eight-cents-per-GB. This is opposed to $75 for the same amount of storage on a HDD. (Domingo, 2014) Even so, the benefit of other features could outweigh the costs. Speed, durability, and the ability to be compact and take up less physical space are all superior aspects of SSD. (Domingo, 2014) Therefore if the budget is not a concern or the products drop in price, SSD may be a viable option.

Storage

Once a digitization project is put in place it is essential to establish a plan for short and long term storage. Many aspects have to be considered when establishing a storage strategy. Different formats require different types of storage and need to be upgraded at various times.

Microfilm

Still one of the most preferred methods of storage is microfilm. A list of the major advantages and disadvantages of microfilm are located in Table 5.

- Table 5 – Microfilm (Brown et. al, 2012 and Skinner, 2010*)*

Advantages	Disadvantages
Preferred method of storage – commonly used	Lack of ability to capture the digital

	functionality of dynamic web materials
Longevity and sustainability	Levels of complexity and interactivity of born digital objects are likely to increase
Reduces the frequency of interventions that may required	If microfilm is damaged or destroyed the information will most likely be lost forever
Extremely difficult to alter without detection	Some materials have to be taken apart in order to put them onto microfilm. This is a problem when trying to preserve items

File Servers – Hard Disk Storage

File servers or hard disk storage is a common method of storage as a result of its ability to store large amounts of data for many people and be a platform where all devices can be stored on one network. Even so, many servers cannot expand unlimitedly to improve storage ability. According to staff members, the Western Libraries Equinox server could be limited in its ability to store every project because its capacity needs to be divided among the various departments. Servers and hard disk storage system also require a certain level of technical ability to not only create but also to maintain. (Information partially collected from Ran, 2011)

External Hard Drives

As mention previously in this report, external hard drives are one of the most common forms for storing materials, especially throughout Western Libraries. The major advantages of external hard drives are that they can store various file formats and are available in a variety of capacities. Even so, the devices might not always be adequate as a result of the size of some projects, with a single project potentially being the same size as some external hard drives. As well, external hard drives can only be considered a short term options as a result of the data which is on them needing to be transferred every 5 years.

CD/DVD

Using for CDs and DVDs for storage is quite common, which is at least partially due to the medium being relatively inexpensive. Even so, there are some major concerns when using this method of storage. One of the major questions that must be considered is how long is CD and DVD technology going to be existence? Some new laptops are not even equipped with CD and DVD drives. As a result, CDs and DVDs should only be considered a short term storage option. As well, there is an problem with physical space when dealing with CD and DVD storage.

Cloud Storage

One of the increasingly more popular methods of storage is the cloud or online storage. Some of the major reasons for this is the fact that cloud storage can provide higher capacity options than many physical storage options and can be mirrored between multiple locations. On the other hand cloud storage can be expensive and there is always a potential of network bandwidth issues and a single point of failure. (Banach et. al, 10, 2011). Even if the department has materials backed up, after a major network crash they would potentially have to find a different provider or

medium to host the materials. A list of the major advantages and disadvantages of cloud storage is located in Table 6.

- Table 6 – Cloud Storage (Flexible Storage Option and Banach et. al, 2011*)*

Advantages	Disadvantages
High Capacity	Expensive
Mirrored between multiple locations	Network bandwidth issues
Can archive storage for large files such as HD videos and high resolution images	Potential single point of failure
Multiple file formats	Complex
Variety of options available	Must trust service provider and have a clear service agreement in place
	Susceptible to hackers and viruses

The Ontario Digital Library Research Cloud (ODLRC)

Even though it is not yet fully operational, The Ontario Digital Library Research Cloud (ODLRC) may become of the best options for cloud storage. The "project is a collaboration of Ontario's university libraries to build a high capacity, geographically distributed storage and computing network using proven and scalable open source technologies." (Marks, 2014) One of the main potential advantages of the ODLRC is that it comes out of a partnership of major Ontario institutions. The project is a product of 11 of the provinces universities, and with the goal that "all 21 university libraries will ultimately be able to benefit from this collaboration by having a new option to acquire high-capacity archival storage." (Marks, 2014) Even though the ODLRC appears to be an adequate solution, at the time of writing it was not available for use. As a result, the site cannot currently be considered the main option for hosting materials.

LOCKSS

Lots Of Copies Keep Stuff Safe (LOCKSS) program takes content from one location and distributes it through a network of LOCKSS participants, ensuring that dozens if not hundreds of copies of the content are mirrored and stored across a wide geographical network. (Nadal, 2011) A partnership with LOCKSS or a participating institution could be a valuable storage solution for Western Libraries.

Migration Schedule

Even though there is no set standard for migration of digital materials, there are several recommendations that are possible. The University of Massachusetts Amherst (UMass Amherst) suggest checking files and refreshing storage on a regular basis, including creating a new media copy every five years or when it is necessary to avoid the loss of data. (Digital Images, 2012) For example, if it is foreseeable that a particular storage medium will soon be obsolete or unfashionable, it is essential to transfer the material to a more sustainable platform. Also, several sources say that information on an external hard drive should be transferred at least every 5 years.

UMass Amherst also recommends migrating files to newer file formats if needed. This includes attempting to stay current with the major shifts in standards, software, hardware, platforms, and formats to ensure that files are in a common, readable form. (Digital Images, 2012)

Storage – Multiple Copies

When storing materials it is also essential to keep multiple copies both on and off site. Below are three examples of strategies undertaken by various institutions.

UMass Amherst (from Digital Images, 2012)

- Keep multiple copies of your images in different locations and on different media forms to ensure the best chance of long term preservation.
- One copy may remain on your computer, but putting several other copies on separate media such as DVDs, CDs, portable hard drives, thumb drives, or Internet storage will be the best protection against loss
- Store copies in different locations that are physically as far apart as practical
- To be safe as possible create 2 copies of an image. Put one away as your master copy without modifying it at all and use the other one for your color-correcting, cropping, emailing, etc. When you need another copy to manipulate, make a copy of the master file and work with the copy. Many people use TIFFs as their master files, but created JPEGs for emailing, presenting, etc.

Yale (from the Digital Coffee Group, 2010)

- Maintain a copy of your photo archive off-site
- They explain that one potential problem of an online copy is that if there is a corruption in your primary copy and you have your system set to automatically clone it to the back up, you may corrupt the backup
- Server
    - Rescue Repository or Digital Preservation Repository
    - Redundant server with tape back up
    - Digital Asset Management System – with backup
- Physical
    - CD/DVD duplicate copies stored in separate locations
    - External Hard Drives
    - Should only be used for short term or as a non-primary method
    - Media and Files should be verified and migrated on a regular basis

Smithsonian Institutional Archives (Digitization Standards For Images, 2009)

*Since this information was collected the Smithsonian Institutional Archives has re-organized their website and this page is no longer available. It is still included in the report because it contains some very practical and useful information regarding storage techniques.

- Digitized images are to be stored redundantly on DVD-R/DVD+R with a minimum of two offline copies
    - A preservation master and a preservation back up
    - The master is to be held offsite
        - Contains technical information regarding the creation date of the disks and software used for the creation of the disks
- A additional set could be created for reference/public access as long as copyright and internal policy permits it
- In total they suggest having a minimum of 3-5 copies in various locations

Metadata

Determining which metadata fields to use will vary depending on the project. Even so, there are several basic fields that have become standard for most projects. Miller (2011) wrote that one "may want to create a collection level metadata record that describes or represents an entire collection," along with an "item level metadata representing each individual resource within the collection." This distinction would be important to distinguish the larger collection from the individual items. As Table 7 indicates, many of the same metadata fields could be used for both levels of a collection.

The Metadata Working Group at the University of Massachusetts Amherst (UMass Amherst) created a list of required, and required if applicable metadata fields. Their guidelines, presented in Table 8, have some of the same characteristics as Miller's list including Title, Identifier and Date. The lists by both Miller and the UMass Amherst Working Group could be enough to cover some digitization projects at Western. This is especially true when one is handling more text based materials such as theses, dissertations or journals. Even so, these fields may not be sufficient when describing more complex items, collections or other research data. Some departments such as GIS and science produce research data that requires a much broader set of metadata fields. For example it is also important for GIS data to include scale, producer, project and co-ordinate system information, limitation of use and on occasion accompanying codes. It is also important to consider other factors, such as metadata for a series of photographs of the same object. It is important for the metadata to represent each angle and not just repeat the same who, what and where information.

Another important piece of metadata that was absent from the examples in Tables 7 and 8 is copyright or rights holder metadata. For many, displaying the rights holder information is essential when hosting materials. This is especially true when the materials which are being digitized did not originate within a particular institution. Miller (2011) does point out that "different metadata schemes include such information as copyright statements and other more complex statements of rights, restrictions can access, terms and conditions of use and provenance." (85) As a result, there are various ways in which copyright or rights holder information can be displayed. Miller (2011) goes on to write that "many digital collection have separate webpages that state general intellectual property rights for resources in the collection as a whole and policies for use for those resources." (85) In that case, information regarding copyright does not necessarily have to be listed on every item but there can be an overall copyright statement for the collection. This report would recommend that rights holder or

copyright be a required metadata field for all digitization projects, whether it is displayed at the collection or item level. A list of encoded examples of how rights holder fields could be displayed using the major metadata structures is located in Table 9.

Table 7 – Collection Level and Item Level Metadata from Miller (2011)

Metadata	
Collection Level	**Item Level**
Title	Title
Identifier	Identifier
Resource Type	Resource Type
Date	Date
Creator	Creator
Creator (additional authors)	Topical Subject
Publisher	Geographical Subject
Topical Subject	Part of the collection
Geographical Subject	

Table 8 – Required and Required If Applicable Metadata from the Metadata Working Group (8, 2013)

Metadata	
Required	**Required If Applicable**
Date Created or Date Published	Creator
Identifier	Extent
Institution Name	Language of Resource
Title	Related Item
Type of Resource	

Table 9 – Encoded Examples of Rights Holder Metadata from the Metadata Working Group (8, 2013)

Encoded Examples	
Dublin Core	<dc:rights>
EAD	<accessrestrict>
MODS	<accessCondition type=restriction of access>
VRA	<rightsSet>

Dublin Core

Dublin Core is one of the most widely used metadata structures. One of the reasons for this is that “its elements are broad and generic, usable for describing a wide range of resources.” (Dublin Core, 2012) As a result Dublin Core could be appropriate for some digitization projects at Western. Even so, despite Dublin Core’s universality it is not necessarily suited for more

complex projects. As Vogel (2014) pointed out, while Dublin Core's "simplicity makes it easy to apply, the lack of specificity makes it less than ideal for documents that may require more detail." (4-5) As a result more complex projects may require looking beyond Dublin Core. Vogel (2014) went on to write that "the lack of detail present in the original core elements required other metadata standards to be created to better serve the needs of the specialized sectors." (4-5) Therefore if Dublin Core needs to be enhanced by another standard, than it may necessary to adopt a schema that is more in line with the needs of the project.

MODS

Another widely used and diverse structure is Metadata Object Description Schema (MODS). MODS is a "general metadata schema intended for description of a wide range of resources, rather than a specific type of object or in a specific discipline or subject domain." (Miller, 163, 2011) As a result MODS would be ideal for a wide variety of digitization projects and collections. Other major advantages include its compatibility with MARC and how it provides "implementers a great deal of freedom to be as simple or as detailed as they wish in the use of many MODS subelements, attributes and vocabularies." (Miller, 164 and 168, 2011) A number of institutions have begun to see the value in MODS and have implemented the structure into their systems. As Miller (2011) points out, "many digital libraries that formally used Dublin Core have moved and continue to move to MODS because of its capacity for richer resource description and because peer institutions had moved to it." (163) Even though Dublin Core could still be useful for some projects and collections, MODS could be a quality alternative if more complexity was required.

EAD

Encoded Archival Description (EAD) is the standard for encoding archival finding aids using XML. (About EAD, 2012) These finding aids allow for much more in-depth description within the metadata and can be used for a wide variety of items. Even though EAD might not be appropriate or necessary for all library materials, it could potentially be very useful in large scale or complex digitization projects.

VRA Core

Visual Resource Association's Core (VRA Core) is a metadata structure that may go beyond the needs of Western Libraries but could be appropriate for some projects. Miller (2011) says that VRA core categories are "arguably most useful when describing works of visual art and architecture" and "demonstrates some of the primary characteristics that art museums consider important for representation and retrieval of works of art and architecture, and of digital collections of images of those works." (213)

Preservation Metadata

When undertaking a digitization project it is also important to keep in mind the preservation implications of the work. If the intention is to preserve the material, then the metadata should be representative of that plan. Preservation metadata has been described as "the information

necessary to carry out, document, and evaluate the processes that support the long-term retention and accessibility of digital materials." (Banach, et. al, 12, 2011)

The Preservation Metadata: Implementation Strategies (PREMIS) was established to document the specific attributes related to preservation. PREMIS has stated that there are five entities that were particularly important in regards to digital preservation activities: Intellectual Entities, Rights, Objects, Agents and Events. (PREMIS Editorial Committee, 5, 2012) Contained within these five entities would be the information necessary for preservation metadata.

Preservation

If preservation is part of the goal of a digitization project than it is important to understand which file formats are appropriate for long term storage. Table 10 is a list of the file formats that Library and Archives Canada "recommends for the preservation of and long term access to digital content." (File Format Guidelines, 5)

Table 10 – Recommended File Formats For Preservation And Long Term Access (File Format Guidelines, pg. 7)

Content Type	Recommended
Audio	BWF WAV
Digital Video	Motion JPEG 2000
Still Images	JPEG, JP2, TIFF, GeoTiff
Web Archiving	ARC, WARC

Recommendations

Hosting and Displaying Materials

The various libraries and other departments at Western, for the most part, seem to be going their own way when it comes to digitization. It is recommended that Western create a centralized webpage for digitization projects. This page would not have to host all of the projects but could link all of the projects and departments together. As with CulturePlex this centralized webpage could contain an abstract about the project and a profile on who is involved. This would create a focal point for digitization and a common project that could be shared among the various departments.

Storage

When it comes to storage, a combination of online and offline storage would be ideal. For online storage, a cloud service, such as the future OCLRD would most likely be the best option. It will also be important to have multiple physical copies stored in various locations. There would be one master copy that would not be accessed unless absolutely necessary. This copy would also contain all of the technical information regarding the disk and its contents. The information

included would detail not only what is on the disk, but also how it was formatted and the potential formatting issues that could occur with this type of file in the future. At least two copies should be stored on CDs, DVDs, or external hard drives in various locations. If economically feasible, an expanded centralized server could also be beneficial. In total, it would be recommended to store a minimum of five copies across the various platforms.

Works cited

1. "About EAD," *The Library Of Congress*, July 11, 2012, http://www.loc.gov/ead/eadabout.html.

2. Banach, Meghan, Brian Shelburne, Kelcy Shepherd and Aaron Rubenstein, "Guidelines For Digitization," *UMass Amherst Libraries*, May 2011.

3. Brown, Heather, John Baker, Walter Cybulski, Andy Fenton, John Glover, Paul Negus and Jonas Palm, "The Role of Microfilm in Digital Preservation," *Microform and Digitization Review*, Volume 41 (2), July 2012.

4. "Collection Policies," *University of Toronto Research Repository*, https://tspace.library.utoronto.ca/policies/comm-and-coll.jsp#community.

5. "Content Guidelines for T-Space at U of T," *University of Toronto Research Repository*, https://tspace.library.utoronto.ca/policies/content.jsp.

6. *CulturePlex*, http://www.cultureplex.ca/

7. Digital Coffee Group, "Digitization Shared Practices – Still Images Version 1.0," *Yale University*, August 2010.

8. "Digital Images," *UMass Amherst Libraries*, May 8, 2012, http://www.library.umass.edu/services/services-for-faculty/preserving-your-digital-materials/digital-images/.

9. "Digitization Standards For Images," Smithsonian Institution Archives, Updated March 2009, http://siarchives.si.edu/records/electronic_records/records_erecords_digitization_images.html.

10. Domingo, Joel Santo, "SSD vs. HDD: What's The Difference?" *PC Mag*, February 20, 2014, http://www.pcmag.com/article2/0,2817,2404258,00.asp.

11. "Dublin Core Metadata Element Set, Version 1.1," *Dublin Core Metadata Initiative*, June 14, 2012, http://dublincore.org/documents/dces/#DCTERMS.

12. "File Format Guidelines for Preservation and Long-term Access," *Local Digital Registry (LDFR), Library and Archives Canada (LAC)*, http://www.collectionscanada.gc.ca/obj/012018/f2/012018-2200-e.pdf.

13. "Flexible Storage Options," *USC Digital Repository*, http://repository.usc.edu/storage/.

14. Marks, Stephen, "About the ODLRC," https://spotdocs.scholarsportal.info/display/ODLRC/About+the+ODLRC.

15. Metadata Working Group, "Metadata Guidelines," *UMass Amherst Libraries*, May 2013, http://www.library.umass.edu/assets/aboutus/attachments/Metadata-Guidelines-v4.pdf.

16. Miller, Steven J., *Metadata for digital collections: a how-to-do-it manual*, (London : Facet Pub., 2011).

17. PREMIS Editorial Committee, *PREMIS Data Dictionary for Preservation Metadata, version 2.2*, July 2012.

18. "Preservation," *Smithsonian Institutional Archives*, http://siarchives.si.edu/services/preservation.

19. . Ren, Yazhong, "Data Storage in Digital Library using Multi-Storage Technology," IEEE, 2011, http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6013968.

20. Skinner, Julia, "Learning From The Past: Digitization and Information Loss," *B Sides: UofI SLIS Journal*, May 14, 2010, http://ir.uiowa.edu/cgi/viewcontent.cgi?article=1008&context=bsides.

21. "Trusted Digital Repository," *Ontario Council Of University Libraries*, 2013, http://www.ocul.on.ca/node/97.

22. Vogel, Dustin M., "Qualified Dublin Core and the Scholarly Works Application Profile: A Practical Comparison," Library and Philosophy and Practice, 2014, http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=2685&context=libphilprac.

23. Witt, Michael, Defining and Deploying an Institutional Data Repository Service at Purdue (PURR), webinar, February 5, 2014, http://rds-sdr.cisti-icist.nrc-cnrc.gc.ca/eng/events/webinars/index.html.

Interviews

Information was also obtained from interviews with the following individuals:

- Tom Belton
- Bev Brereton
- Alyssa Evetts
- Christine Homuth
- Elizabeth Mantz
- Cheryl Woods

Acknowledgments

I would like to thank the following individuals for their assistance and guidance on this project:

- Tom Belton
- Bev Brereton
- Alyssa Evetts
- Christine Homuth
- Joan Kammerer
- Elizabeth Mantz
- Cheryl Martin
- Joanne Paterson
- Cheryl Woods