Appraisal and Selection for Digital Curation

Jinfang Niu
School of Information
University of South Florida

Abstract

Based on traditional archival appraisal theories and existing appraisal/selection policies in libraries, archives, museums, social science and science data centers, this paper presents a generic appraisal/selection framework for digital curation and discusses how it can be implemented. The framework includes three selection methods: statistical sampling, risk analysis and appraisal. Details about the appraisal method are illustrated, including the objects of appraisal, appraisal criteria and appraisal decisions.

Introduction

Due to the large quantities of records, appraisal has been considered one of the most important functions in archives management for many years. However, in digital curation appraisal has not received as much attention as in traditional archives management. The Open Archival Information System (OAIS) reference model, which has been considered a key document in digital curation, does not explicitly include appraisal as one function or even one sub-function in its functional model. The 2012 version of the OAIS model only briefly mentioned appraisal implicitly:

> 'An organization operating an OAIS should have established some criteria that aid in determining the types of information that it is willing to, or it is required to, accept. These criteria may include, among others, subject matter, information source, degree of uniqueness or originality, and the nature of the techniques used to represent the information (e.g., physical media, digital media, format).' (ISO, 2012, p. 3·2).

One might argue that the OAIS reference model focuses on the internal functions and information packages of an archival repository, whereas appraisal primarily occurs before resources are ingested into the archival repository. However, even the Producer-Archive Interface Methodology Abstract Standard (CCSDS, 2004), which delineates detailed interactions between producers and the archival repositories, does not explicitly include appraisal/selection. While the Digital Curation Lifecycle model does include appraisal/selection and reappraisal, it does not provide details for the appraisal/selection action, and its definition of re-appraisal as '[r]eturn data which fails validation procedures for further appraisal and reselection' is not accurate from an archives management point of view (Higgins, 2008).

The archives and records management community has created and accumulated rich theories and methods for appraising both traditional and electronic records, such as Schellenberg's appraisal model for government records (1956), the macro-appraisal theory created by the National Archives of Canada (Cook, 2004), Bole's appraisal model that developed from appraising university records (Boles, 2005), the Minnesota method created by the Minnesota Historical Society to appraise business records (Greene and Daniels-Howell, 1997), the appraisal model created by the Appraisal Task Force of the InterPARES[1] project (InterPARES Appraisal Task Force, 2001), and the benchmark and baseline requirements created by the InterPARES project for assessing the authenticity of electronic records (InterPARES Authenticity Task Force, 2002). These theories and methods can be applied to other types of digital resources, such as digital publications, web resources, as well as social science and science data. Harvey (2006) has tried to create a generic framework for appraisal/selection in digital curation by adapting some archival appraisal theories and methods and reviewing existing appraisal/selection policies for preservation and digitization. While Harvey's review of existing appraisal/selection policies was extensive, the framework he created was surprisingly simple, containing only ten appraisal criteria including value, physical condition, resources available, use, social significance, legal rights, format issues, technical issues, policies, and documentation. It lacks some essential elements that should be available in a generic appraisal/selection framework. In addition, appraisal criteria in Harvey's framework are not mutually exclusive. For example, value and

---

[1] InterPARES: http://www.interpares.org

social significance are used as two separate appraisal criteria, although social significance is a kind of value. Similarly, format issues and technical issues are listed as two separate appraisal criteria although format issue is a kind of technical issues.

Using the similar methodology adopted by Harvey, this paper will create a more comprehensive appraisal/selection framework for digital curation. The framework synthesizes traditional archival appraisal methodologies and elements identified from existing appraisal policies. It intends to serve as high-level guidance for individual institutions to create their local appraising/selecting policies for digital curation.

Methodology

The framework was formulated based on traditional archival appraisal methods and elements identified from existing appraisal/selection policies from libraries, archives, museums, science and social science data centers. Appraisal policies were gathered from these many fields because the author believes that digital curation merges several previously existing information management fields, including electronic records management, digital library, digital archiving and data curation. The author first re-analyzed those appraisal/selection policies reviewed by Harvey (2006), such as the selection policy of the Cedars Project Team report (2001), the Decision Tree for Selection of Digital Materials in the Digital Preservation Coalition's handbook (2001) and United Nations Educational, Scientific and Cultural Organization (UNESCO) Guidelines for the Preservation of Digital Heritage (Webb, 2003). Please refer to the reference list in Harvey (2006) for a complete list of appraisal policies reviewed. Then the author identified additional policies through literature search, including:

- Inter-university Consortium for Political and Social Research (ICPSR) selection criteria for social science data (ICPSR, n.d.);

- Earth Resources Observation Systems (EROS) appraisal tool for earth observation data (2007);

- National Digital Stewardship Alliance (NDSA) selection criteria for geospatial data;

- Guidelines for selecting scientific data released by the Digital Curation Centre (DCC) (Whyte and Wilson, 2010);

- Guidelines for selecting digital television programs of the National Digital Information Infrastructure and Preservation Program (NDIIPP) project on preserving digital public television (Ide and Weisse, 2006);

- Penn State Institutes of Energy and the Environment Data Repositories data acquisition guidelines (Kelly, 2010);

- A selection of guidelines from the PANDORA web archive (National Library of Australia, 2005);

- Guide to assessing the significance of museum collections released by the Collections Council of Australia (Russell and Winkworth, 2009);

- Guidelines for the selection of geospatial datasets at the California State Archives (n.d.);

- Procedure for Scientific Records Appraisal and Archive Approval of the National Oceanic and Atmospheric Administration (NOAA) (2008);

- Appraisal polices of four national archives (United States, United Kingdom, Canada and Australia);

- National Archives and Records Administration (NARA) General Records Schedules (GRS)[2] and numerous records schedules from the NARA Records Control Schedule (RCS)[3] repository;

- NARA guidance on scheduling various kinds of electronic records, such as geospatial data and web records (NARA, 2007b, 2007c);

- NARA guidance for the transfer of various kinds of electronic records (NARA, 2005).

A preliminary appraisal framework was constructed based on the author's background knowledge of archival appraisal. This preliminary framework includes statistical sampling and appraisal as two selection methods, value as an appraisal criterion as well as context and resources as two objects of appraisal. This preliminary framework was iteratively enriched, adjusted and refined as more and more elements are identified through content analyses of appraisal policies and related literature. For example, the appraisal criterion "mission alignment" was found from the appraisal policy of EROS (2007). Since this criterion is different from value, it is added as a separate appraisal criterion in the framework. The author also found that some complex policies do not simply provide a list of appraisal criteria, they also define appraisal procedures that specify the steps and parties involved in each step of the appraisal process. For example, the appraisal procedure created by EROS include nine steps, starting with the proposal of a collection for review and ending with a final decision of the EROS director for accepting or rejecting the collection (EROS, n.d.). Based on this finding, a new section about the "implementation of appraisal/selection framework" was added.

A Framework for Appraisal/Selection

Three methods can be used for selecting digital resources for preservation: statistical sampling, risk analysis, and appraisal. Statistical sampling does not select based on careful assessment of the values of resources, instead, it relies on statistical methods to create a representative sample for a population. This method has been used for a long time by archives in selecting paper records for preservation (Kepley, 1984). Two kinds of statistical sampling have been used: systemic sampling that selects every nth item from a population and random sampling that selects items from a population with equal probability regardless of their values. In the digital world, statistical sampling is used by libraries to select resources for digitization (Ooghe and Moreels, 2009). It is also used by the National Library of France in selecting web pages for archiving (Lasfargues, Oury and Wendland, 2008).

A risk analysis method involves assessment of the kinds of risks that could possibly happen to digital resources, as well as the probability and consequence of each risk. Commonly mentioned risks in existing appraisal/selection policies include risk of

---

[2] General Records Schedules: http://www.archives.gov/records-mgmt/grs/
[3] NARA Records Control Schedule (RCS) repository: http://www.archives.gov/records-mgmt/rcs/

obsolescence of file format and storage media, risk of loss caused by not preserving the digital resources, and by storing resources in an unsafe area. Risk analysis usually is not listed as a separate selection method in selection policies. Instead, it is mixed together with other selection criteria. Here are some example selection criteria that involve risk analysis:

- "Are there significant costs or consequences to the program or the Government if the data are lost?" (GeoMAPP, 2011);

- "Will it be in an area where it is at risk of theft or damage?" (Harvey, 2006, p. 26);

- "Selection criteria should consider the 'obsolescence rating' for each format, such as lower risk for formats such as VHS, and critically endangered for 1-inch SMPTE." (Harvey, 2006, p. 28).

Appraisal is a term often used in archives management, museum studies, and art galleries to assess the values of archival materials, museum artefacts, and artworks. The following sections will talk about various aspects of appraisal, including objects of appraisal, appraisal criteria and decisions of appraisal.

Objects of Appraisal

Context of resources

Appraising the context of resource creation and use is an efficient way to select from large quantities of resources. In archives management, this is called macro-appraisal. The basic assumption of macro-appraisal is that resources which are produced either by important producers or during the conduct of important functions and activities are more likely to be valuable; therefore, given limited time and resources for appraisal, we can ignore resources that are from an insignificant context and focus on appraising resources that are from an important context. In addition to Library and Archives Canada, where the macro-appraisal theory originated, several other institutions also assess the significance of context in selecting resources. For example, the selection policy of the Edinburgh University Archives states that "data created by projects of major national or international significance" should be archived (Edinburgh University Library, n.d.). Similarly, the Penn State Institutes of Energy and the Environment acquire data associated with "Specific important events (Hurricane Ivan flood event)" (Kelly, 2010). In web archiving, many web pages are selected based on their association with significant events, such as the Iraq War web archive and the September 11, 2011 web archive of the Library of Congress[4].

Resources

In archives management, appraising records themselves is called micro-appraisal. For digital resources, not only content, the physical conditions and technical characteristics also need to be appraised, because they directly affect the costs and feasibility of preservation. Existing appraisal/selection policies contain an extensive assessment of the currency or obsolescence of file formats and storage media. For example, the DCC guide for selecting research data assesses whether data files are received in formats that meet designated technical criteria (Whyte and Wilson, 2010).

---

4 Library of Congress Web Archives: http://lcweb2.loc.gov/diglib/lcwa/html/lcwa-home.html

Harvey (2006)'s appraisal framework also involves the assessment of format and technical issues of digital resources.

Significant properties of resources

The UNESCO Guidelines for the Preservation of Digital Heritage (Webb, 2003) pointed out in appraising digital resources, we not only need to select resources, but also need to select the significant properties of resources. This means that the micro-appraisal of digital resources can be more granular than that of print resources. We may decide to preserve only important events recorded in a digital calendar or important layers in a digital map. Similarly, in appraising a website, we may want to remove the advertising banners and navigation bars and preserve only the textual content. The decision of what significant properties to preserve should depend on the designated user community, types of resources, and the mission of the preservers. Archives may mainly preserve the intellectual content of organizational records, whereas a museum may need to ensure users can play old computer games in ways that reproduce the original experience.

Metadata and documentation

The quality of metadata and documentation affects the values of resources, and the costs and feasibility of preservation. Thus many appraisal policies include the assessment of the sufficiency and adequacy of metadata and documentation. For example, ICPSR "strongly prefers data collections that have comprehensive technical documentation providing ample information on sampling procedures, weighting, recoding rules, skip patterns, constructed variables, and data collection procedures to allow users to assess the quality and analytical reliability of the data" (ICPSR, n.d.). Some appraisal/selection policies are also concerned with the interoperability (Fusco and van Bemmelen, 2004) and usability of metadata (Faundeen, 2010). Conformance with agreed standards is an indicator of interoperability. If the metadata as received is consistent with or easy to convert into the metadata format used by the preserver, the metadata is considered easy to use.

Metadata and documentation are also the targets of records scheduling. For example, the GRS of the NARA schedules various metadata and documentation related to electronic records. The NARA Records Control Schedules Repository also contains many schedules for metadata or documentation, such as the records schedule no. nc1-afu-76-74 which schedule the Air Reserve Pay and Allowance System (ARPAS) documentation, and the records schedule no. n1-412-07-044 which schedule EPA Pesticide Usages Survey Data and Documentation.

Depending on how essential they are for the preservation and use of associated records, metadata and documentation can be divided into two categories. The first category of metadata and documentation is useful for assessing the authenticity and reliability of data, but is not essential for the preservation, use, and understanding of records, such as system maintenance documentation relating to system performance testing, system usage, and some user identification, profiles, authorizations, and password files (NARA, 2012). In NARA GRS 20, this kind of documentation is scheduled as temporary and will be deleted when no longer needed by the records producers. This is the same for system security documentation (NARA, 2010c, Item 5), access control and handling records for classified records (NARA, 2010a).

The second category of metadata and documentation is essential for the preservation, use, and understanding of records. This includes technical documentation about software and hardware environment and other information that is essential to

decide how to open and run the digital records, as well as metadata and documentation that is essential to understand the intellectual content and context of records. In NARA GRS 20, this kind of documentation is defined as "[d]ata system specifications, file specifications, code books, record layouts, user guides, output specifications, and final reports (regardless of medium) relating to a master file, database or other electronic records" (NARA, 2010b). For social science data, this kind of documentation includes the codebooks, sampling procedures, weighting, recoding rules, and data collection procedures (Gutmann, Schürer, Donakowski, and Beedham, 2004).

This second kind of metadata and documentation has at least the same retention periods as associated records. In NARA GRS 20, system documentation for temporary systems is scheduled as temporary and system documentation for permanent systems is also scheduled as permanent and will be transferred to NARA along with associated records. This is the same for metadata in many other records schedules found from the NARA Records Controls Schedule Repository. For example, in the records schedule for Environmental Information Management System (records schedule no. n1-412-04-009), both the data files and supporting documentation are scheduled as disposable (delete when superseded or obsolete). In the records schedule for the Economic Directorate Document Management System (records schedule no. n1-029-12-004), both content and metadata are scheduled as permanent (while system is in operation, transfer a copy of records and associated metadata to NARA every five years). A records schedule from the University of Waterloo (2011) shows that some metadata and documentation are retained longer by the producer than associated records. In their OnBase system, after records are destroyed or transferred, some minimum metadata will be retained in the original information system, such as purge reports.

The appraisal of metadata and documentation not only happens upon or before the initial acquisition by an archival institution, it may also happen to metadata or documentation already under custody by the archival institution. During the preservation process of digital resources, metadata accrue and accumulate. Some of them may no longer be needed after a period of time and need to be pruned. For example, in the NARA records schedule with ID number DAA-0064-2009-0002, the review status metadata of records already under the custody of NARA is scheduled as temporary and will be retained for a minimum of six years and then deleted when no longer needed. Review status metadata means electronic information relating to reviews of records that are conducted in response to Freedom of Information Act requests or for various other reasons.

Appraisal Criteria

The first and foremost appraisal criterion is mission alignment: whether the resource supports the mission and falls within the scope of the collection policy of a preserver. Valuable resources may be rejected or de-accessioned because of this reason. For example, the EROS data center de-accessioned the Apollo film data set that consists of thousands of frames of film taken by the National Aeronautics and Space Administration (NASA) astronauts during the Apollo spaceflight missions. While this film set has significant historical value to the United States, most of the films were taken of spacecraft or the surface of the Moon and are not consistent with the mission of EROS who preserves only films and photos taken of Earth (Faundeen and Oleson, 2007). A national library that has the mandate to preserve all publications within the country, an institutional repository that aims to preserve all intellectual products of an institution, or a social science data archive that has the mandate to preserve all research

data generated through the support of a government funding agency, will select materials regardless of values and other appraisal concerns. The same piece of software may be turned down by NARA but accepted by a technology museum for the same reason.

The second most important appraisal criterion is the value of digital resources. Archives management theories categorize values of records into primary values and secondary values. Primary values are the values for the resource producers, including administrative value, fiscal value, and legal value (Tibbo, 2003). Secondary values are the values for users other than the producers. Schellenberg (1956) divided secondary values into evidential values and informational values. These values have been used as the guidance for the appraisal policies of many archival institutions. For example, the appraisal policy of the UK National Archives contains one appendix specifically about these values (UK National Archives, 2012). Similarly, California State Archives also assesses the potential evidential and informational value in selecting geospatial datasets for long-term preservation. The DCC guidelines for selecting research data includes scientific and historical value as appraisal criteria (Whyte & Wilson, 2010), and the appraisal policy of NARA mentioned the research value of records: "How significant are the records for research?" Scientific, historical and research values are all specific kinds of secondary values of digital resources. Appraisal policies of libraries and museums also consider aesthetic, artistic and cultural values, such as the guidelines for assessing the significance of Australian collections (Russell & Winkworth, 2009) and the UNESCO selection guidelines the preservation of digital heritage (Webb, 2003).

Values of resources are affected by various factors. Many appraisal/selection policies include these factors as appraisal criteria, although it is more accurate to categorize them as factors affecting values. The first group of factors includes the authenticity, reliability, integrity and accuracy of resources. These factors are found in the appraisal policies of NARA (NARA, 2007a) and the Digital Library for Earth System Education of Columbia University (Kastens and Butler, 2001). The second group of factors includes usefulness, usability and accessibility. The appraisal policy of ICPSR stated that it seeks data that "are useful in utilization of current and emerging research and statistical techniques" and data that permit the use of quantitative and/or qualitative social science research techniques. It also gives data that are in a format that facilitates ease of use a high priority for inclusion in the data archive (ICPSR, n.d.). Both the NARA appraisal policy and the DCC appraisal guideline for research data include usability as a concern. Accessibility is very much a legal and ethical concern. Access may be restricted due to copyright or privacy concerns. For data that meet the selection criteria, ICPSR gives data that are in the public domain and have cleared copyright a high priority for inclusion in the archive. The third group of factors that affect value includes uniqueness, diversity, and representativeness. In science data archives, the non-replicability of data, which is a kind of uniqueness, is used as a selection criterion. Data that are impossible or very expensive to re-collect are more likely to be preserved. This can be seen in the DCC appraisal guidelines (Whyte & Wilson, 2010) and Appendix 2 of the NARA appraisal policy (2007a). Some archival institutions acquire resources that fill gaps in existing holdings (Gutmann, et al., 2004). For example, one selection criterion of the State Library of Victoria is whether the acquisition will address an existing weakness (State Library of Victoria, 2011). These resources become valuable due to diversity concerns.

The third appraisal criterion is cost. Cost may become the decisive factor in selecting multiple versions of the same content. For example, the National Library of Australia prefers to preserve the online version of publications when both online and

CD ROM or floppy disk versions are available, because the technical obstacles to long-term preservation are fewer for online publications (National Library of Australia, 2005). The policy of EROS considers various kinds of cost in appraising earth science data: costs in acquiring, housing, preserving, processing the collection to make it accessible and de-accessing the resources. It also assess whether value of the data exceeds costs. For social science data that involve human subjects, the cost of redacting confidential information of human subjects should be considered. Probably due to this reason, ICPSR gives a high priority to datasets that adheres to standards for privacy and confidentiality (ICPSR, n.d.).

Feasibility is the fourth appraisal criterion. Feasibility of preservation is often determined by the technical capacity of the preserver. Some preservers refuse to accept digital resources in certain file formats or stored on certain storage media because they cannot preserve those files. In its guidance for scheduling and transferring electronic records, NARA states that it will determine if any technical considerations might impede its ability to process and provide future access to records. These technical considerations include metadata and documentation, file formats and protection mechanism, such as password, digital signatures or encryption (NARA, 2005) To some extent, feasibility of preservation is affected by the cost of preservation. It might be cost prohibitive to preserve certain digital resources even though it is technically feasible. Legal restrictions also affect feasibility of preservation. Digital preservation requires copying. This includes making duplicates and storing the copies in different locations for disaster recovery purposes, making copies in different file formats for preservation (e.g., master images in tiff format) and for access purposes (e.g., png format for online access), refreshing storage media, migrating to newer file formats, etc. If copyright or donor restrictions are so restrictive that it is not possible to make copies for preservation or access purposes, then it is not feasible to preserve the digital resource.

Appraisal Decisions

Archivists conduct appraisals primarily to decide whether to acquire certain resources for preservation. Many resource producers do not aim to preserve digital resources permanently, instead, they preserve the resources for a limited time period to support primary use and to fulfil legal mandates. Thus they need to decide the retention periods of digital resources. In fact, the NDSA report defines appraisal as "a process by which archivists and records managers assign administrative, legal, research, and historical value to records in order to determine retention period." (NDSA, 2013). In light of this, the "appraise and select" action in the digital curation life cycle model might be changed to "appraise, select, and schedule".

Another two kinds of appraisal decisions are made in the re-appraisal of resources under custody. One kind of re-appraisal is to decide resources that should be de-accessioned. Libraries may want to de-accession materials that are no longer used. Archives may need to re-appraise and de-accession materials that were acquired due to erroneous appraisal decisions, or materials that no longer fit in the scope of collection due to changed mission or collection policy, or materials that no longer considered valuable due to changed societal trends or designated community. For example, United States Geological Survey re-appraised some data sets already held by the data centre and decided to purge at least one large satellite collection (Day, 2004). Another kind of re-appraisal is called technical re-appraisal by Harvey and Thompson (2010). It is the technical assessment of the currency of file formats and storage media in order to decide

the next step in the preservation strategy. For example, if a certain file format is appraised as obsolescent, then a migration will need to be initiated.

Implementation of Appraisal/Selection

A preserver can create its own appraisal/selection policy by selecting a number of selection methods, appraisal objects and appraisal criteria presented above. For example, a national library with a legal mandate to preserve all websites in its country only needs one selection criterion: mission alignment. When multiple elements are considered in the appraisal/selection, the preserver needs to decide the priority and weight of each element. A preserver may choose a different set of elements, priorities, and weights for each type of resources. For example, NOAA produces both business records (e.g., financial records, human resource records) and science data gathered by earthquake sensors and satellites. It needs different appraisal/selection policies for these two different kinds of resources. Even for the same institution and for the same type of resources, the weight and priority of each element may change over time (Ayris, 1998).

In practical implementations, the appraisal criteria in this framework can be customized for specific type of digital resources, and converted into more operational guidelines or appraisal questions. Informational value can be judged based on whether the resources are related to specific subjects or topics. Scientific and research value can be judged based on citation counts. Evidential value can be judged based on whether the records document decisions that set precedents (GeoMAPP, 2011). Here are some appraisal questions from the NARA appraisal policy for GIS data (GeoMAPP, 2011):

- "Do the data involve or reflect any legal rights of the Government or individuals?"

- "Will the data be needed to defend the agency or the Government against charges of data fraud or misrepresentation?"

- "Can the data be usefully integrated with newer data resulting from resurveying or improved methods of data collection and interpretation?"

- "Will the data be useful for analyzing geographic distributions over time?"

The first two questions help determine evidential values and the latter two help determine information values of the GIS data.

A preserver needs to decide the parties that are involved in the appraisal process. Assessing the value of records based on real or anticipated needs of researchers has been a central methodology in traditional archival appraisal (Brichford, 1977). In the digital world, user needs continue to be an importance factor in selecting resources for preservation. Some digital repositories acquire resources based on user requests, such as the Penn State Institutes of Energy and the Environment Data Repositories (Kelly, 2010), the PANDORA web archive (National Library of Australia, 2008), and the National Taiwan University Web Archiving System (Chen et al., 2008). For digital resources that require specialized knowledge, such as science data, it is essential to involve subject experts to assess the current and potential future value of data. Subject experts can be involved directly through joining the appraisal team or as external peer-reviewers. Some institutions involve resource producers in the appraisal based on the rationale that producers have good knowledge of their data. For example, the NDIIPP

project on preserving digital public television incorporated the insights of television producers in the selection of television programs for long-term preservation (Ide and Weisse, 2006). In fact, relying on records creators for selecting archival records has been the argument of the famous English archivists Hillary Jenkinson (1966) and his followers (Duranti, 1994). However, as has been pointed out by the American archivist Gerald Ham (1993), records creators tend to make biased decisions that are favorable to themselves. Some institutions, such as EROS, also involve senior management in the appraisal (Faundeen, 2010).

Depending on who are involved in appraising/selecting resources and for whom, different appraisal decisions may be reached. For example, in the NDIIPP project on preserving digital public television, producers identified intermediary products, such as submasters, promo reels and graphic elements, as having significantly less value for long-term retention, whereas educators consider intermediate products as extremely instructive because they show what footage was left out as well as used in a program (Ide and Weisse, 2006). When multiple parties are involved in the appraisal, different parties may focus on appraising different aspects. For example, archivists appraise historical value, scientists appraise scientific value, and senior management appraises mission alignment and costs, etc. Finally, lawyers and technical experts may also be involved in assessing the legal and technical feasibility of preserving digital resources respectively.

When multiple elements and multiple parties are involved in appraisal/selection, in order to ensure that the appraisal is conducted on a consistent and orderly manner, it is necessary to define an appraisal procedure. An appraisal procedure is a workflow that specifies what elements are appraised by whom during each step of the process. Although the four national archives studied by Niu (2012) use basically the same appraisal methods, their appraisal procedures are different. The Library and Archives Canada actively approach government agencies to assess the value of records with the help of government agencies. However, at NARA and the National Archives of Australia, the government agencies make initial appraisals, and submit draft records schedules to the national archives. The national archives then appraise and approve the records schedules received. Some digital curation institutions have created appraisal procedures, such as the appraisal procedures of EROS (Faundeen, 2010), NOAA (2008), and the Decision Tree for Selection of Digital Materials for Long-term Retention created by the Digital Preservation Coalition (2006).

A preserver also needs to decide the timing of appraisal. Today it has been commonly recognized that the appraisal of digital resources needs to happen early in the lifecycle, upon or even before resource creation. However, it is not always possible to conduct appraisal upon or before creation. Early appraisal is easier to do when the resource producers and preservers have a close and direct relationship. It is more difficult for preservers who accept donations or solicit resources from those sources not pre-defined.

There can be multiple appraisal occurrences during the lifecycle of resources. For example, the UK National Archives reviews government records twice. The first review appraises primary values at five years from closure, and the second appraises secondary values at 25 years from creation (Niu, 2012). Appraisals that happen after acquisition are called re-appraisal. Traditionally, re-appraisals are conducted to de-accession materials that no longer fit in the collection of a preserver. Today, some scholars consider the technical assessment of the currency of file formats and storage media in order to decide the next step preservation strategies a kind of re-appraisal (Harvey and Thompson, 2010).

Technologies have offered new approaches to conduct appraisal. One approach is to crowd-source appraisals based on user views or rating. Today publishers automatically track the number of views and citations of each article; social media websites capture the number of views and votes of online videos or blog posts. Preservers can also actively solicit evaluations from the public. For example, to create a video archive for the Occupy Wall Street Movement, archivists asked members of the Occupy Wall Street Working Group to vote for the five most important YouTube videos about the Occupy Movement (Besser, 2013). Another approach is to use software tools to implement certain selection criteria automatically. A web crawler can automatically select web pages in a domain name, with certain file formats, or published after a certain date. The National Library of the Czech Republic uses software to analyze web pages and looks for pre-defined properties that characterize the Czech web. Every time a pre-defined property is found, a certain amount of points are added to the URL. When a certain threshold is reached, the web page is considered a part of the Czech web and will be archived (Vlcek, 2008).

Conclusion

This paper presents a framework for the selection of digital resources based on traditional archival appraisal theories and existing selection policies of various information management institutions. The following list summarizes the major components of the framework:

- **Selection methods**

    o Statistical sampling

    o Risk analysis

    o Appraisal

        ▪ Objects of appraisal

            • Context

            • Resources

            • Significant properties

            • Metadata and documentation

        ▪ Appraisal criteria

            • Mission alignment

            • Value

            • Cost

            • Feasibility

        ▪ Appraisal decisions

            • Retention period

            • Acquisition

- • De-accession
- • Preservation strategy
- • **Implementation of appraisal/selection**
  - o Decide selection methods, appraisal objects and criteria
  - o Convert appraisal criteria into operational guidelines
  - o Decide parties involved in the appraisal
  - o Decide appraisal procedure
  - o Decide timing of appraisal
    - ▪ Early appraisal before or upon resource creation
    - ▪ Appraisal before acquisition
    - ▪ Re-appraisal of resources under custody

This framework is in no way perfect and final. It intends to serve as a foundation upon which other people can criticize, adjust and build upon. Once a generic and commonly accepted appraisal framework is created, each individual institution can customize and extend the framework for local use. This will improve the consistency and completeness of appraisal/selection policies of individual institutions.

References

Ayris, P. (1998). *Guidance for selecting materials for digitisation.* Paper presented at Joint RLG and NPO preservation conference, Guidelines for digital imaging, University of Warwick, Coventry, UK, 28–30 July 1998. Retrieved from http://eprints.ucl.ac.uk/492/1/paul_ayris3.pdf

Besser, H. (2013). Archiving aggregates of individually created digital content: Lessons from archiving the Occupy Movement. *Preservation, Digital Technology and Culture 42*(1), 31–37. doi:10.1515/pdtc-2013-0005

Brichford, M. (1977). *Archives and manuscripts: Appraisal and accessioning.* Chicago: Society of American Archivists.

Boles, F. (2005). *Selecting and appraising archives and manuscripts* (1st ed.). Chicago: Society of American Archivists.

CCSDS. (2004). *Producer-archive interface methodology abstract standard* (Magenta Book No. CCSDS 651.0-M-1). Retrieved from http://public.ccsds.org/publications/archive/651x0m1.pdf

California State Archives. (n.d.). *Appraising geospatial records: Strategies and guidelines.* Retrieved from http://salt.unc.edu/eLegacy/docs/FinalReport/appraisal_of_geospatial_records.pdf

Cedars Project Team. (2001). *The Cedars project report: April 1998 – March 2001.* Retrieved from http://www.imaginar.org/taller/dppd/DPPD/146%20pp %20CedarsProjectReportToMar01.pdf

Chen K.H., Chen, Y.L., & Ting, P.F. (2008). *Developing National Taiwan University web archiving system.* Paper presented at the 8th International Web Archiving Workshop, Aarhus, Denmark. Retrieved from http://iwaw.europarchive.org/08/IWAW2008-Chen.pdf

Cook, T. (2004). Macro-appraisal and functional analysis: Documenting governance rather than government. *Journal of the Society of Archivists, 25*(1), 5-18. doi:10.1080/0037981042000199106

Day, M. (2004). The selection, appraisal and retention of digital scientific data: The ERPANET/CODATA workshop. *Ariadne, 39.* Retrieved from http://www.ariadne.ac.uk/issue39/erpanet-rpt

Digital Preservation Coalition. (2006). Decision tree for selection of digital materials for long-term retention. In *Digital Preservation Handbook.* York, UK: Author. Retrieved from http://www.dpconline.org/advice/preservationhandbook /decision-tree

Duranti, L. (1994). The concept of appraisal and archival theory. *American Archivist, 57*, 328–344. Retrieved from http://archivists.metapress.com/content/PU548273J5J1P816

Edinburgh University Library. (n.d.). Edinburgh University Archives archival selection criteria. Retrieved from http://www.ed.ac.uk/polopoly_fs/1.37798! fileManager/criteria.pdf

EROS. (2007). Records appraisal tool. Retrieved from http://eros.usgs.gov/government/ratool/

EROS. (n.d.). Acceptance of data collections by the USGS. Retrieved from http://www.usgs.gov/contracts/acq_opp/EROS_tech_library/TSSC%20Recompete /Reference%20Library/EROS-DMD-01%20Acceptance%20of%20Data %20Collections%20by%20the%20USGS%20v1.0.pdf

Faundeen, J.L. (2010, March). *Archiving strategy for USGS EROS Center and our future direction.* Paper presented at the US Digital Preservation Interoperability Framework Workshop, Gaithersburg, MD. Retrieved from http://ddp.nist.gov/workshop/papers/01_07_Faundeen_NIST_Paper_Submitted.doc

Faundeen, J.L., & Oleson, L.R. (2007, October). *Scientific data appraisals: The value driver for preservation efforts.* Paper presented at the PV 2007 Conference, Oberpfaffenhofen, Germany. Retrieved from http://www.pv2007.dlr.de/Papers/Faundeen_AppraisalsValue_for_Preservation.pdf

Fusco, L., & van Bemmelen, J. (2004). Earth observation archives in digital library and grid infrastructures. *Data Science Journal 3,* 222-226. doi:10.2481/dsj.3.222

GeoMAPP. (2011, July). *Appraisal of geospatial data.* Presentation given at the online GeoMAPP Informational Partners Meeting: Focus on Geospatial Appraisal. Slides retrieved from http://www.geomapp.net/docs/InfoPartner_Appraisal_presentation_final_20110728.pdf

Greene, M.A., & Daniels-Howell, T.J. (1997). Documentation with an attitude: A pragmatist's guide to the selection and acquisition of modern business records. In J.M. O'Toole (Ed.), *The Records of American Business* (pp. 161-230). Chicago, IL: Society of American Archivists.

Gutmann, M., Schürer, K., Donakowski, D., & Beedham, H. (2004). The selection, appraisal, and retention of social science data. *Data Science Journal 3,* 209-221. doi:10.2481/dsj.3.209

Harvey, R. (2006). Appraisal and selection. In S. Ross & M. Day (Eds.), *DCC Digital Curation Manual.* Edinburgh, UK: Digital Curation Centre. Retrieved from http://www.dcc.ac.uk/resource/curation-manual/chapters/appraisal-and-selection

Harvey, R., & Thompson, D. (2010). Automating the appraisal of digital materials. *Library Hi Tech, 28*(2), 313-322. doi:10.1108/07378831011047703

Ham, G.F. (1993). *Selecting and appraising archives and manuscripts* (1st ed.). Chicago: Society of American Archivists.

Higgins, S. (2008). The DCC curation lifecycle model. *International Journal of Digital Curation, 3*(1), 134-140. doi:10.2218/ijdc.v3i1.48

ICPSR. (n.d.). ICPSR collection development policy. Retrieved from http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/policies/colldev.html

Ide, M., & Weisse, L. (2006). Recommended appraisal guidelines for selecting born digital master programs for preservation and deposit with the Library of Congress. Retrieved from https://www.prestocentre.org/system/files/library/resource/appraisal-guidelines-final.pdf

InterPARES Appraisal Task Force. (2001). *Appraisal Task Force report.* Retrieved from InterPARES Project website: http://www.interpares.org/book/interpares_book_e_part2.pdf

InterPARES Authenticity Task Force. (2002). *Requirements for assessing and maintaining the authenticity of electronic records.* Retrieved from InterPARES Project website: http://www.interpares.org/book/interpares_book_k_app02.pdf

ISO. (2012). *ISO 14721: Space data and information transfer systems – Open archival information system (OAIS) – Reference model.* (Also published as Magenta Book No. CCSDS 650.0-M-2, available from CCSDS website: http://public.ccsds.org /publications/archive/650x0m2.pdf)

Jenkinson, H. (1966). *A manual of archive administration* (2nd ed.). London, UK: Lund Humphries.

Kastens, K.A., & Butler, J.C. (2001). How to identify the "best" resources for the reviewed collection of the Digital Library for Earth System Education. *Computers and the Geosciences, 27*(3), 375-378. Retrieved from http://www.ldeo.columbia.edu/DLESE/collections/CGms.html

Kelly, M.C. (2010, November). *Penn State Institutes of Energy and the Environment data repositories.* Presentation given at the Digital Geospatial Appraisal Meeting, Washington, DC. Slides retrieved from http://www.digitalpreservation.gov/meetings/documents/othermeetings/kelly.pdf

Kepley, D. (1984). Sampling in archives: A review. *American Archivist 47*, 237-242. Retrieved from http://archivists.metapress.com/content/KJ041746214P6U1V

Lasfargues F., Oury, C., & Wendland, B. (2008). *Legal deposit of the French Web: Harvesting strategies for a national domain.* Paper presented at the 8th International Web Archiving Workshop, Aarhus, Denmark. Retrieved from http://iwaw.europarchive.org/08/IWAW2008-Lasfargues.pdf

Niu, J. (2012). Appraisal and custody of electronic records: Findings from four national archives. *Archival Issues 34*(2), 117-130.

NARA. (2005). Transfer of permanent e-records to NARA. Retrieved from http://www.archives.gov/records-mgmt/initiatives/transfer-to-nara.html

NARA. (2007a). Strategic directions: Appraisal policy. Retrieved from http://www.archives.gov/records-mgmt/initiatives/appraisal.html

NARA. (2007b). Tips for scheduling potentially permanent digital geospatial data records. Retrieved from http://www.archives.gov/records-mgmt/publications /geospatial-tips.pdf

NARA. (2007c). Tips for scheduling potentially permanent web content records. Retrieved from http://www.fws.gov/irm/bpim/docs/na_webrectransfertips.pdf

NARA. (2010a). General records schedule 18: Security and protective services records. Retrieved from http://www.archives.gov/records-mgmt/grs/grs18.html

NARA. (2010b). General records schedule 20: Electronic records. Retrieved from http://www.archives.gov/records-mgmt/grs/grs20.html

NARA. (2010c). General records schedule 24: Information technology operations and management records. Retrieved from http://www.archives.gov /records-mgmt/grs/grs24.html

NARA. (2012). Frequently asked questions about using GRS 20 to schedule electronic systems. Retrieved from http://www.archives.gov/records-mgmt/memos/ac28-faq-on-grs20.pdf

National Library of Australia. (2005). Online Australian publications: Selection guidelines for archiving and preservation by the National Library of Australia. Retrieved from http://pandora.nla.gov.au/selectionguidelines.html

National Library of Australia. (2008). Services to researchers. Retrieved from http://pandora.nla.gov.au/researchers.html

NDSA, Geospatial Content Team. (2013). *Issues in the appraisal and selection of geospatial data.* Retrieved from http://hdl.loc.gov/loc.gdc/lcpub.2013655112.1

NOAA. (2008). *Procedure for scientific records appraisal and archive approval: Guide for data managers.* Retrieved from https://www.nosc.noaa.gov/EDMC/documents /NOAA_Procedure_document_final_12-16-1.pdf

Ooghe, B., & Moreels, D. (2009). Analysing selection for digitisation: Current practices and common incentives. *D-Lib Magazine 15*(9/10). doi:10.1045/september2009-ooghe

Russell, R., & Winkworth, K. (2009). *Significance 2.0: A guide to assessing the significance of collections.* Adelaide, Australia: Collections Council of Australia. Retrieved from http://arts.gov.au/resources-publications/industry-reports /significance-20

Schellenberg, T.R. (1956). *The appraisal of modern public records* (Bulletins of the National Archives No. 8). Washington, DC: US Government Printing Office.

State Library of Victoria. (2011). Collection and resources development policy: Digital resources. Retrieved from http://www.slv.vic.gov.au/sites/default/files /Digital%20resources%20CRDP_0.pdf

Tibbo, H.R. (2003). On the nature and importance of archiving in the digital age. In M. Zelkowitz (Ed.), *Advances in Computers: Vol. 57. Information Repositories* (pp. 1-67). doi:10.1016/S0065-2458(03)57001-2

UK National Archives. (2012). Appraisal policy. Retrieved from http://www.nationalarchives.gov.uk/documents/information-management/appraisal _policy.pdf

University of Waterloo. (2011). Electronic recordkeeping metadata standard implementation in OnBase. Retrieved from https://uwaterloo.ca/document-management/university-waterloo-electronic-recordkeeping-metadata-standard-implementation-onbase

Vlcek, I. (2008). Identification and archiving of the Czech Web outside the National Domain. Proceedings from IWAW '08: 8th International Workshop for Web Archiving, Denmark. Retrieved from http://iwaw.europarchive.org/08/IWAW2008-Vlcek.pdf

Webb, C. (Ed.). (2003). *Guidelines for the preservation of digital heritage.* Paris: UNESCO. Retrieved from http://unesdoc.unesco.org/images/0013/001300/130071e.pdf

Whyte, A., & Wilson, A. (2010). *How to appraise and select research data for curation.* Edinburgh: Digital Curation Centre. Retrieved from http://www.dcc.ac.uk/resources/how-guides/appraise-select-data