Approaching Appraisal

Guidelines and Criteria to Select for Digital Preservation

Nathan Tallman

Penn State University
nathan.tallman@psu.edu

Lauren Work

University of Virginia
lw2cd@virginia.edu

ABSTRACT

Digital preservation practitioners are often asked to make collection and preservation decisions about content they steward; more often than not this is simply because a collection contains or is solely composed of digital content. However, traditional preservation practitioners typically do not make these same decisions in a vacuum for analog collections; they are informed and prioritized by selectors, liaisons, or curators who make decisions based on content and collection priorities in concert with information provided by the preservation practitioners.

In this paper, we argue that this model should not change simply because the nature of the materials has changed. While digital preservation may be new to an institution, basic preservation decisions are best prioritized by the people building collections and selecting content, which is complemented by the knowledge and expertise from digital preservation practitioners.

Selectors and curators are collaborative preservation partners whose roles and expertise render them best able to judge the value of the content they collect. Digital preservation practitioners have an obligation to work with selectors and curators to help them understand the ways in which technical characteristics, descriptive records, and financial impact, along with value, affect selection choices for digital content and how these choices affect a digital preservation program. Digital preservation practitioners need to provide selectors and curators with guidelines and criteria to help them make informed selection decisions for digital content.

In this paper, we will address this issue in five parts. First, we review existing literature and workflows on the issue of selection for general collections. Second, we review existing literature on the topic of appraisal in special collections. Third, we provide an analysis of how these practices compare and contrast. Fourth, we discuss differences between preserving analog and digital content. Finally, we recommend specific guidelines and criteria geared toward selectors and curators to aid with content selection for digital preservation in both general and special collections.

CONFERENCE THEMES ADDRESSED

Collaboration and capacity-building; mapping out sustainable digital preservation approaches and communities.

KEYWORDS

Appraisal, selection, digital preservation, digital preservation practitioner, digital content, digital preservation librarian, collection development, and collection management.

1. INTRODUCTION

Creating strong collections that meet the needs of local users and maintaining them over time is the foundation of librarianship. Many models exist to guide the collection development process, but less is written on how the decisions of collection maintenance, such as preservation actions, should be evaluated in this process. These preservation decisions are as important as the decision to acquire because they have significant impacts on the collection over time. In traditional collections of tangible resources, the person in the role of developing collections generally makes these decisions, often in consultation with preservation professionals.

However, in the case of digital collections, there can be a tendency to rely on the digital preservation professionals to make these prioritization decisions on their own. This may be due to a sense of discomfort or unfamiliarity for selectors due to the technical nature and preservation needs of digital content, in contrast to well understood decision pathways about analog collections. It is the duty of the digital preservation professional to help our colleagues with these issues and to provide guidelines and criteria so they, the selector or curator, can make informed decisions in the best interest of the collection, no matter the format. The decisions should not be made from a purely technical perspective or by the digital preservation practitioner alone, but by the collection specialist, incorporating the same criteria used to choose the best materials for the collection in the first place.

We have reviewed the literature for selection frameworks for both general and special collections, incorporating digital content, and comparing and contrasting between the two collection types to understand shared concepts. Using this review as a baseline, selection guidelines and criteria are proposed to specifically address the role of content selectors and curators in digital preservation practice. While these guidelines cannot include every possible criteria for all institution types, they are a solid footing to establish a shared understanding between collection professionals and digital preservation professionals, who share the responsibility of stewarding collections over time.

The reader will note differences in terminology between general and special collections. While the terminology may differ, the terms often describe similar roles or actions related to the development of collections. General collections, often referred to as circulating col-

lections, are developed by selectors and the process is referred to as selection. The individuals in this role may also be referred to as bibliographers, subject specialists, or liaisons. In this paper, we refer to this group of professionals collectively as selectors. In special collections, (sometimes noted as non-circulating), the same process is often referred to as appraisal. Although not always in their job title, the individuals who appraise collections are most often referred to as curators, though archivists and other professionals who work in this setting also often conduct appraisal. In this paper, we refer to these professionals working in special collections collectively as curators. Professionals in the digital preservation specialist role may have various titles as well, but here we use this term to refer to those whose duties primarily revolve around digital preservation practices and programs for library and archival collections.

2. SELECTING FOR GENERAL COLLECTIONS

A review of nearly four decades of literature shows evolving approaches and philosophies to selecting resources for general collections in libraries. As Peggy Johnson put it in 2014, "Selection is both an art and a science" [19, p. 138]. Strong collection development policies rooted in needs assessments facilitate the science through approval plans, blanket orders and other automated mechanisms. The art of item-by-item selection, once the standard, is still used but to a lesser degree.

Philosophies for selection vary by library type, but usually include a combination of populist and traditional thinking. Populism, also referred to as demand theory [8] and the liberalist point of view [20], selects for what users want. This method gained wider adoption as a way of justifying the use of public funds. Traditionalist, also called value or quality theory [8] and educational theory [12], selects for what users need, based on the role of libraries as champion of knowledge and collector of resources for the benefit of users' education. Academic libraries use traditionalist thinking when basing collection development policies on the curricular needs of programs and their institutions, but also use populist thinking when accepting faculty request for particular items.

Some authors make distinctions between evaluating materials for their intrinsic value and selecting them within the context of the overall collection. Many authors provide general rules to guide the librarian during the selection process. Vicki Gregory suggests the collection, as an aggregate, should be the guiding principle in *Collection Development and Management for 21st Century Collections* [12]. Resources are to be suitable to the collection, unbiased, appropriate to the users' needs, and free from discrimination. Regardless of the approach or philosophy, librarians use a variety of criteria when selecting resources.

In his 1980 guidelines for print materials in *Collection Development: The Selection of Materials for Libraries*, William Katz, describes ten objective criteria, (1) purpose, scope, and audience; (2) difficulty; (3) authority, honest, and the credibility of author and publisher; (4) subject matter; (5) comparison; (6) timeliness; (7) format; (8) price; (9) curriculum support; and (10) demand [20, p. 91-96]. Six of these criteria directly relate to the content, while the other four are intrinsic to the resource and its creation. Although Katz does not state it as a criteria, he recommends reading reviews, "as much as possible", if the librarian is unable to decide on whether to select a book or not. He also uses these criteria for non-print materials, but contextualizes them for other formats such as audiovisual materials.

John Rutledge and Luke Swindler recommend a quantitative method in *The Selecting Decision: Defining Criteria and Establishing Priorities* [27, p. 126-128]. Rutledge and Swindler developed the model when, after conducting a literature review, they felt there

was no satisfactory model to follow for collection development at their institution. Their model uses only six ranked, mutually exclusive and weighted criteria, (1) subject, (2) intellectual content, (3) potential use, (4) relation to the collection, (5) bibliographic considerations, and (6) language. Interestingly, they concluded cost was a non-factor for selection. Using the quantitative scores, resources under evaluation can be assigned priority status and purchased in that order, as the budget allows.

Dan Hazen's 1982 article, *Collection Development, Collection Management, and Preservation*, focuses on selection for preservation, rather than selection for purchasing (or licensing). Hazen draws parallels between collection development and item preservation noting, that the same criteria should be used for both decisions as they both affect the collection. His five specific criteria include (1) academic activity or user demand, (2) historic precedent and tradition, (3) volume and cost of materials, (4) availability of alternatives to purchase, and (5) discipline-specific models of access to information [14, p. 7-10]. Hazen states, "Preservation specialist are best suited to identify the endangered materials within a particular collection, but subject specialist must then delineate priorities among those items" [14, p. 8], suggesting that while preservation specialists may provide information that informs a decision, it's the role of the selector to decide whether or not the conservation treatment needed to preserve the book is warranted for the collection.

Although not universal, preservation was a often a criterion or part of the general rules authors recommended, particularly where periodicals and serials were concerned. "When one selects a periodical, a long-term commitment is usually being made," Richard Gardner notes in his 1981 book, *Library Collections: Their Origin, Selection, and Development* [8, p. 187]. This was especially apparent when scholarly resources began to include electronic resources. In 2001, Timothy Jewel suggests that libraries may be required to maintain "both local print and electronic subscriptions while working toward long-term technical solutions," in *Selection and Presentation of Commercially Available Electronic Resources: Issues and Practices* [17, p. 13].

### SELECTING ELECTRONIC RESOURCES IN GENERAL COLLECTIONS

Electronic resources began to be featured in articles and books on selecting around the year 2000. Diane Kovacs and Kara Robinson provide guidelines that highlight the unique aspects of electronic resources such as the nature of the resource and how it was published, (since the internet made it much easier for people to self-publish), accuracy, bias, opinion stated as fact, and especially the authority of the source. As commercial publishers began to preference licensing models, the business model of the publisher became important, too [21].

In 2006, Thomas Leonhardt includes criteria on redundancy and duplicates of electronic resources versus print resources already in the collection in his *Handbook of Electronic and Digital Acquisitions*. More criteria are added that show the rising role of electronic resources, such as predictability of pricing, and stability of coverage [22]. Dibyendu Paul notes the criterion of ease of use both in navigating to the resource and the ability to have concurrent users of the same resource in *Collection Development Policy and Selection Criteria for Electronic Material: Indian Perspectives* [26]. Suzanne Mangrum and Mary Ellen Pozzebon, in *Use of Collection Development Policies in Electronic Resource Management*, began to distinguish between criteria of licensing from user perspective (reuse) as well as library management [24].

Using the criteria gathered from the review of selecting literature above, we have created a dataset of selection criteria for general collections found the thirteen reviewed articles or chapters [32], including physical and electronic resource specific criteria.

The resulting set of 203 separate criteria were mapped to 14 normalized criteria.

Figure 1 and Table 1 [32] show selection criteria overall. For general collections overall and physical resources specifically, criteria related to content, such as subject, coverage, organization, were most frequent.
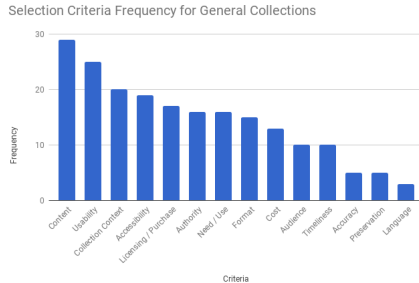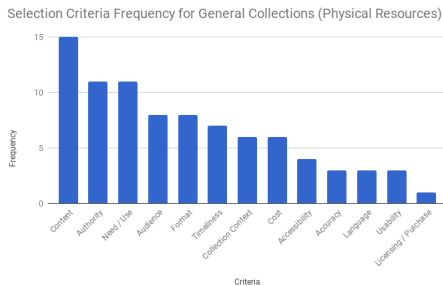

Figure 1. Selection Criteria for General Collections, Overall

Criteria	Frequency
Content	29
Usability	25
Collection Context	20
Accessibility	19
Licensing / Purchase	17
Authority	16
Need / Use	16
Format	15
Cost	13
Audience	10
Timeliness	10
Accuracy	5
Preservation	5
Language	3
n=13	

Table 1. Selection Criteria for General Collections, Overall Data

Figure 2 and Table 2 [32] show selection criteria for physical resources.

Figure 2. Selection Criteria for General Collections, Physical Resources

Figure 3 and Table 3 [32] show selection criteria for electronic resources. Electronic resources more frequently had usability, licensing/purchasing terms, and accessibility criteria than content. This suggests that the unique aspects of electronic resources, which can't be picked up off a shelf and examined and require computer hardware and software for use, the ability of users to access the content is more important than the content itself.

Criteria	Frequency
Content	15
Authority	11
Need / Use	11
Audience	8
Format	8
Timeliness	7
Collection Context	6
Cost	6
Accessibility	4
Accuracy	3
Language	3
Usability	3
Licensing / Purchase	1
n=6	

Table 2. Selection Criteria for General Collections, Physical Resources, Data
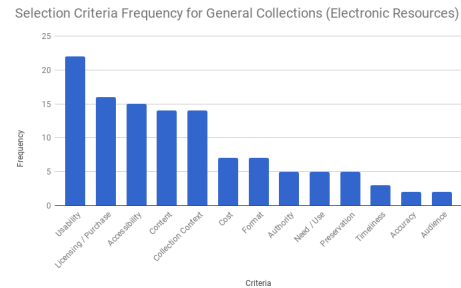
Figure 3. Selection Criteria for General Collections, Electronic Resources

Criteria	Frequency
Usability	22
Licensing / Purchase	16
Accessibility	15
Content	14
Collection Context	14
Cost	7
Format	7
Authority	5
Need / Use	5
Preservation	5
Timeliness	3
Accuracy	2
Audience	2
n=8	

Table 3. Selection Criteria for General Collections, Electronic Resources, Data

3. APPRAISAL AND SELECTION FOR SPECIAL COLLECTIONS

"Appraisal is, arguably, the most important thing archivists do. After all, determining what materials are preserved dictates the future of the archives." [33, p. iv].

Appraisal theory and practice in archives and special collections is cited as one of the core principles at the heart of modern archivy; the idea from which all other functions flow [7]. Combined with the important guiding principles of collection policy and scope, ap-

praisal theory over the last century reflects evolving archival practice that shifted from views of neutrality and passive acceptance of any evidentiary documentation to the modern necessity and acceptance of the active impact that archivists and curators have on the historical record by their appraisal and eventual preservation decisions, and the responsibilities of carrying out these decisions in accountable, documented, widely participatory, and transparent ways [7, 9, 16, 33].

For the purposes of this paper, we use the phrase "archives and special collections" to denote the variety of special collections and archives and the myriad materials they may collect, which can include non-commercial unpublished manuscripts and rare materials, university archives, and topical collections that distinguish these repositories from general library collections. Many of the frameworks and theories of appraisal first emerged around specific institutional or governmental constructs, particularly in archives, but a majority of the literature around the suitability for appraisal frameworks was emphasized by the article and book authors to be applicable beyond these specific organizational or repository contexts, and so we share them in a continuum here.

The intricacies of appraising materials for accession into archives and special collections has been well documented over the last century. Building on the *Dutch Manual for the Arrangement and Description of Archives* founding principles, archivists and curators have moved away from early appraisal theories of archivists such as Sir Hilary Jenkinson, who, in an effort to center the impartiality of archives in order to better serve as untarnished evidence, emphasized a very limited role of the archivists on appraisal. Jenkinson instead argued for the efficacy of Administrators to carry out appraisal tasks on their own, with the archivist simply as the protector of all documents committed to the archives.

This passive stewardship theory where archivists were expected to keep, but not select archives did eventually evolve over time, as archivists both iterated on their own theory and practice, and dealt with modernization and responsibilities of government administrations, institutions, and documentation. T.R. Schellenberg provided one of the next most significant theoretical frameworks around appraisal in the 1950s. As Schellenberg worked within the voluminous governmental and administrative structure of the U.S. National Archives, he both emphasized the exponential growth of the modern public records system, and used this experience to introduce frameworks around which archivists should take an active role in selection and appraisal within a deluge of modern records [5, 9].

Schellenberg emphasized the primary and secondary value of information and records, in which the primary value is associated with the record-producing institution and provides an important documentary function around the administrative, fiscal, legal and operating purposes for which an agency has been created. Secondary value is related to how these records retain both evidentiary and informational value even after they cease to be of current (primary) use, and therefore could provide both historical evidence on procedures, policies and decisions, as well as information on persons, places, and activities with which an institution or individual dealt [28].

Schellenberg's appraisal framework, particularly the secondary value concept, provided a robust backbone for curators and archivists to engage in more active appraisal efforts that focused on the value of records in terms of their future use for researchers, and emphasized the role of expertise by archivists and curators. However, the iterative nature of appraisal theory emerged again into the 1970s and 1980s, when increasingly diverse sets of institutional settings and collection mandates warranted approaches that did not fit frameworks originally produced in governmental or highly institutional settings. Frank Boles and Julia Marks Young addressed

this need by creating an appraisal model meant to allow for more diverse acquisition mandates across a broader range of institutional settings.

Their approach consisted of three general categories of decisions to be evaluated when appraising records. First was the value of information, second was the costs of retention, and the third related to the political and procedural implications of the appraisal recommendations. This framework, building on concepts of cost/benefit appraisal analysis work pioneered by G. Philip Bauer, introduced some important theory-to-practice concepts around the realities of appraisal and costs of collection to institutions over time by using a much more granular lens that previous appraisal theory [3, 9]. They also increased the emphasis on preservation considerations for materials during appraisal. For example, the cost of retention category alone is split into four distinct categories for consideration; storage costs, processing costs, conservation costs, and reference costs. However, Boles and Young made sure to clearly state their appraisal framework as relating to paper records only, maintaining a line between paper and "other formats" that would later be erased.

Appraisal theory was refined again into the early 1990s with the work of archivists such as Hans Boom, who brought a newer framework of societal paradigm, wherein archives and their appraisal should strive to reflect the values of the public at that time [5]. Canadian-led theory such as macro-appraisal emerged, which shifted archivists' view from appraising records to instead appraising record creating entities [6]. Another Canadian concept, total archives, argued that everything from organizational records and newspapers, books and family papers are all part of a larger documentary universe, and that special media should not be appraised and described out of context to more traditional material when likely both are linked by provenance [9].

F. Gerald Ham created an additional theory-to-practice framework in 1993 in *Selecting and Appraising Archives and Manuscripts* which provides five analyses (built on Schellenberg appraisal theory), for archivists to employ when defining records of enduring value. Ham also emphasized the calculation of preservation in his analyses, as his tenets included the functional characteristics of the record, the information in the record to determine its significance and quality, reviewing the record in context of parallel or related documentary source, potential uses that are likely to be made of the record and the physical, legal, and intellectual limitations on access, and the cost of preserving the record weighed against the benefit of retaining the information [13]. Ham's own definition of appraisal also notes the significance of preservation; "The process of evaluating actual or potential acquisitions to determine if they have sufficient long-term research value to warrant the expense of preservation by an archival repository"[13, p. 2].

APPRAISAL OF DIGITAL CONTENT IN SPECIAL COLLECTIONS

This stretch of archival theory and some archival practice alone was enough to have some archivists describe the state of appraisal as a "...stimulating but often confusing cacophony of ideas" [10, p. 167]. Add to this the dawn of the born digital or electronic record, and approaches to appraisal became even more complicated, as realities of networked communication, linked or distributed processes, increasing ability to create data, transient or opaque data, and the general movement away from the physical record and its arrangement prevailed. However, those familiar with the foundational theoretical structure of appraisal, such as David Bearman, saw only familiar concepts in these new digital realities, asserting that even virtual documents still closely correspond to important concepts such as provenance and context of creation for archivists,

regardless of format, though his focus was more narrowly applied to record systems for organizational accountability [1].

Though archivists and curators may have become familiar over the last twenty years or so with the reality of digital material, efforts to appraise and pragmatically acquire these records are still developing [2, 16]. The growth areas of related efforts such as organizational web archiving programs is encouraging, with technical viewpoints and approaches still underlining the familiar archival concepts of collection scope and appraisal, as well as demonstrating evidentiary value, provenance, and research value [25].

"In the end, the appraisal...must justify the value of the records against the costs of transferring, transforming, and maintaining them" [16, p. 65]. Geof Huth and others have added to the theory-to-practice methods in recent publications around the appraisal of digital materials in special collections and archives. Recent literature cites the use of technical appraisal checklists, which emphasize the increasing need of archivists and curators to understand how concepts like provenance and context of digital materials relate to realities of technical infrastructure, structured data, and networked environments. In this way, initial appraisal and digital preservation choices may be intertwined fairly early in the appraisal and acquisition process. Examples of technical appraisal criteria as enumerated by Huth include technical specifications, physical location of the system, size (and potentially expected growth), structure and relationships, database relationships, data quality, storage, and costs of custodianship (including recovery, transfer, transformation, and maintenance costs) [16].

4. SELECTION CRITERIA COMPARISON FOR GENERAL AND SPECIAL COLLECTIONS

As has been noted, there is a difference in terminology between general and special collections. For general collections, the term selection is generally used to describe the process of choosing materials. Sometimes, the term evaluation is used to distinguish between determining an item's value (evaluation) and making the choice to add it to the collection (selection). For special collections, although Ham defines archival selection as the process, by which archivists "... identify, appraise, and accession records of enduring value" [13, p. 2], the term appraisal is commonly used to describe the broader process of identifying materials offered to an archives or special collections that have sufficient value to be accessioned and is equivalent to the term evaluation used for general collections.

Interestingly, the Society of American Archivists' *Glossary of Archival and Records Terminology* notes the term selection is related to appraisal, but the specific definition in archives and special collections refers to "... the process of identifying materials to be preserved because of their enduring value, especially those materials to be physically transferred to an archives" [30]. Because the emphasis of this paper is on selecting from the aggregate, existing collections for digital preservation and not adding new materials to our collections, we therefore use the term selection here.

An additional difference in terminology between general and special collections is the way to which materials requiring computational devices are referred. In general collections, the term electronic resources is a generic term that may refer to a database of journal articles, a CD-ROM that accompanies a physical item such as a book, or a website. The Anglo-American Cataloguing Rules 2nd edition, 2002 revision, defines an electronic resource as consisting of "data (information representing numbers, text, graphics, images, maps, moving images, music, sounds, etc.), programs (instructions, etc., that process the data for use), or combinations of data and programs" [18, p. 9-2]. In special collections, the term

digital collections is generally a generic term that refers to both digitized and born digital materials. Born digital materials are created and used via a combination of computer software and hardware, often never existing in what would be considered a tangible, physical format. Sometimes, the word digital content is also used.

Both general and special collections selection criteria address rights and reuse. For general collections, the ability to fulfill any lending and interlibrary loan obligations an organization may have to consortia or other collaborative partners may have an impact on whether or not it makes sense to add something to the collection. Related, particularly for electronic resources, licensing terms may facilitate or constrain reuse and may impact the cost and research value.

Special collections are equally concerned with rights and reuse. As most special collections are unique, it is critical that a library has the appropriate rights to take preservation actions to protect the collection and make it available for research, including digitization and online access. Typically these details are worked out in the deed of gift, a legal instrument that specifies the terms and conditions of transferring a collection. Although the United States Code grants libraries rights to make preservation copies [35], it is a better defined process if the creator or donor can explicitly articulate assignments of ownership, copyright, and license for reuse in a deed of gift. While many older collections may not have deeds of gift that address these issues, they are now becoming standard [29].

One almost universal distinction between general and special collections is uniqueness. Special collections are overwhelmingly comprised of unique or rare materials whereas general collections are usually widely published and distributed. Though three of the thirteen sampled sets of selection criteria for general collections did include uniqueness as a criteria, the uniqueness referenced in the literature was in relation to the informational content and not the resource itself. For example, a selector may choose a single introductory calculus textbook from 2017 to add to the collection as opposed to all introductory calculus textbooks published in 2017 and a special collection curator may collect the manuscript notebooks of Gottfried Leibniz, who discovered calculus independently of Isaac Newton.

Resources in consideration for general collections are sometimes evaluated item-by-item, though more often they are selected, in part, through approval-based ordering and database packages. Special collections are generally collections in and of themselves, particularly with manuscripts, such as the Jane Goodall papers, and must be appraised as a whole. This is sometimes the case with rare books, but this distinction depends on the library. For example, the Thomas Fisher Rare Book Library at the University of Toronto collects entire personal libraries of notable individuals [34].

Subject matter is a primary selection criterion for both general and special collections; a resource must fit into the overall collection. Both can also include specific mandates or requirements to collect certain materials, such as Federal Depository Libraries or record retention schedules that specify certain resources should be sent to the archives based on primary or secondary values. Scope and coverage is criterion typical for general collections, usually referring to the depth and breadth to which a subject is covered in a resource. A related criterion in special collections is volume and scale, and is particularly applicable for digital collections.

Two closely related criterion are creator(s) in general collections and provenance in special collections. A subject specialist will be able to assess the authority of a particular creator, their reputation in the field, and their publication track record. Provenance can also include these criteria, but is equally concerned with the context in which the resources were created–who created them and why, what purpose did they serve, and how were they used. For example, if selecting a work on feminism, a book by Betty Friedan is probably

a good resource. If you were selecting an archival collection on feminism, Betty Friedan's papers might be a good collection, but if the particular collection you are offered was one box of material about Betty Friedan, collected by a graduate student in the course of creating a poster about Betty Friedan, the provenance isn't that significant.

Organization and/or structure is a criterion used for both general and special collections, but can have different methods for evaluation. How an author organizes the information in a resource, the chapters and sections, the indices and glossaries present can have a great effect on the usability and potential use of a resource. Similarly, the organization (or lack thereof) of an archival collection has a impact on its usability. However, a curator can take into account the amount of time required to process an archival collection so that it is organized in a way that is usable and decide if the collection is worth acquiring based on the present state and potential processed state.

Similar to organization, context is a criterion that is shared by general and special collections, but can also mean something different to both. General collections are most concerned with how the resource fits into the overall collection. Does it fill gaps? Is it part of a collection strength? Is it a duplicate? Is it available through interlibrary loan? Special collections are concerned with the context of its creation, not just as part of the provenance. Is the content meaningful on its own or are other resources necessary for interpretation? Does digital content require a specific software package to be installed? Is there related material that expands the meaning? All are questions the subject specialist and curator must take into consideration.

Cost is frequently a criterion for both general and special collections but is hard to assess independently, as it often an implied factor in other criteria. This difficulty causes some to leave it out all together and make it part of a secondary prioritization process [27]. However, models do take it into consideration, particularly for electronic resources in general collections. Cost is not limited to the expense of purchasing or licensing a resource, but reflects the long-term cost of adding something to the collection.

The long-term costs are often attributable to preservation actions, such as binding, storage, rehousing, and conservation treatments. For digital collections particularly, the volume and scale of content has a great effect on the cost as digital storage is often based on disk usage. There is also computing power and labor required to reformat or forensically process content. A secondary cost that is useful for making preservation decisions is the cost to replace. A new copy of a general resource may be more affordable than treating the existing copy. It is not uncommon for there not to be a replacement for special collections, but for digitized collections, where the original is still the preservation copy, it is necessary to determine the appropriate level of preservation for the digital files. If the digitization was particularly challenging or time consuming, or re-digitizing could damage collections beyond accepted limits, it may be appropriate to give the files a higher level of preservation, beyond backup.

Collection development policies guide the growth of both general and special collections and have long been the means to codify collecting goals and strategies [15]. Collection development policies are based on institutional user needs and can look very different across institutions and across collecting areas. Academic libraries usually base their collection development policies on the curricular and research needs of faculty and students at their institutions, though they may also be based on existing collection strengths. Collection strengths may be around a particular subject, creator, or even format or genre, such as the Special Collection Research Center's Archive of Recorded Sound and Music at Syracuse University Library [31]. Sometimes, collection development policies

may specify conditions under which materials should be transferred from general to special collections, such as monetary value, rarity, and condition.

Writing collection development policies requires understanding local users and their needs, as well as expertise in the relevant subject matter. These policies may also affect preservation decisions. As Hazen says in 1982, there is "significant overlap between preservation and such functions as building maintenance, collection management, and collection development" [14, p. 10].

5. DIFFERENCES BETWEEN ANALOG AND DIGITAL CONTENT

There are many differences between analog and digital content that affect digital preservation selection decisions. Although not the focus of this paper, the following brief discussion is meant to provide context and clarity for digital preservation selection criteria.

Analog preservation strategies can be easier and less expensive than digital preservation strategies. In cases where there are analog and digital versions of the same resource, consider which or if both are the preservation copy to be prioritized for preservation. The cost to replace the digital version may inform whether or not the digital copy is ultimately the preservation copy.

A primary difference between analog and digital content is that digital content is more dependent upon technology and other external factors that allow digital content to be used and interpreted. At its core, digital content is nothing more than a series of ones and zeros. So while an analog resource can often be used without any special equipment (though this is not universally true, as time-based media and microforms both require technology for use), digital content's technology requirements are complex. Typically, a combination of software and hardware is required to turn those ones and zeros into meaningful content that can be used for research. These layers of technology can be particularly challenging for older digital content. Additionally, technological innovations happen at a rapid pace, formats change, and the software that creates and renders digital content changes, as does the hardware it runs on.

However, digital content is intrinsically machine-actionable. As methods of digital scholarship continue to expand, evolve, and be adopted, digital content increasingly becomes valuable. Algorithms, artificial intelligence, natural language processing, and other techniques can be used to analyze enormous amounts of digital material in ways that are not possible with analog content. The digital nature of the content facilitates discovery of new ideas and knowledge.

Digital content is also more challenging to preserve. Because of those layers of technology, it's not as simple as rehousing materials and storing them in a climate-controlled environment over many years. The concept of benign neglect, sometimes taken when preserving analog content, is insufficient for digital preservation, which requires much more active management for digital material to endure. Setting aside the technology required for access and use, digital content itself is brittle and more at risk for damage and even preserving materials as-is is complicated. Physical degradation of hard drives and other digital storage media can lead to file corruption, as can simply opening a document or moving materials between servers.

Preserving digital content is not only more challenging, it can be more expensive. Best digital preservation practices include storing copies in multiple places in different geographic regions and regularly comparing them to make sure no degradation has taken place. Digital storage infrastructures needs constant management, as new systems and the content within them shifts. Security vulnerabilities are discovered regularly and must be mitigated. Policies and proce-

dures must be in place to control access to secure and confidential digital content against unauthorized access.

Finally, it is very easy for anyone to create digital content. Humans have a cognitive dissonance with digital content, because it lives on a hard drive or cloud servers and not a shelf where everyone can see and register the space it inhabits. It is easier to develop bad habits like not organizing files or inconsistently naming files. It's also very easy to replicate content. The mere volume and scale of born-digital collections can present many challenges to an archivist who has to appraise and process a collection. If too much is kept, financial resources are wasted. If too little is kept, important unique content may be lost.

6. GUIDELINES AND CRITERIA TO SELECT FOR DIGITAL PRESERVATION

Outlining these distinctions between general and special collections and analog and digital content is in no way meant to be an argument against contemplating the very difficult work of examining hard or unusual digital use cases or thorny formats. On the contrary, it is a way of emphasizing the importance of prioritization from subject specialists and curators on the content itself to ensure that complementary digital preservation action, in whatever form it may take, supports the priorities of the collection policy and the mission of the institution first.

Therefore, based on our review of the theory and practice of selection and appraisal for both general and special collections, we present a framework meant to guide these same subject specialists and curators to engage with their existing skill sets, and build on them, with the assistance of digital preservationists, so that they can make the selection of these collection materials for long term digital preservation. Formats, systems, and new technologies should not shift the paradigm of subject expertise, research needs, and collection prioritization that is the role of the selector or curator; even if the evolution of technology shifts how and when preservation actions for these digital collections may occur [16, 23].

Here, we aim to empower selectors by creating a shared framework that melds concepts from collection development/maintenance and digital preservation. These criteria allow for informed decision-making by selectors and curators that encourages them to adhere to their principles of appraisal and selection for collections, with additional digital preservation considerations in mind around these existing principles. In this section, we focus on guidelines specifically for the preservation of digital content; that is, content that may be born digital or has been digitized.

The proposed criteria take the shape of categories drawn from the literature review and various selection and appraisal models for both general and special collections. A condensed version of the criteria, meant for easy reference by both curators and digital preservationists, (including criteria examples), is also available at https://doi.org/10.18113/S1RM0J [36].

6.1 *Value*

The value of collections has various definitions across general and archival collection literature, but for the purposes of this framework we focus here on the research value of collections, for their institution and for society. This component of the criterion is at once the most important and most subjective of the criteria, and underlines the essential role of the subject specialist and curator in this framework due to their familiarity and collaboration with faculty and researchers at the institution. Research value is closely tied to collection policy and mission, as well as an emphasis and anticipation of the needs and interests of those who are expected or known to use collections held by the institution.

This type of selection allows for prioritization of content, systems and/or formats that may be part of this subject area as it relates to digital materials and could be clearly communicated to digital preservationists as an area for their desired expertise and consultation. For both special collections curators and general collections selectors, an emphasis on local research and curricular needs is paramount in the selection of materials for digital preservation, as expertise in the research needs of particular subject areas will help to inform preservationists strategy for preservation and access over time.

Consideration of this criterion could also allow for the strategic allocation of digital preservation resources related to best practices for distributed digital preservation (including concepts like geographic distribution and hardware heterogeneity) that may be reserved for top priority subject or collection areas. An example of this may be the annual allocation of particular amounts of data and content for distributed digital preservation systems.

Selecting for digital preservation using a research value criterion also aids with the completeness of record or subject focus to build on collection strengths and improve the connection and context of other digital materials. This criterion could be useful in special collections in areas of distinction to help grow the volume and depth of materials, of reappraisal of collection areas over time, or aid with the parameters for movement of general collections to special collections. For general collections, using this criterion could help justify or sustain involvement in particular consortia efforts for long term digital preservation of journals such as CLOCKSS, and grow the breadth of coverage of collections.

6.2 *Uniqueness*

Uniqueness as a selection criterion for digital preservation can draw on one, or both concepts of uniqueness from general and special collections. For general collections, this criterion tends to be more focused on the informational content of the digital material, rather than the material or item itself. Consideration for selection for digital preservation under this shade of uniqueness might include a focus on the impact of subject research developments in the field or new publication methods and digital formats that are unique to that material. Local considerations may also include individual research conducted at an institution in a subject area that an organization may be bound by institutional mission to prioritize and preserve.

Uniqueness selection criteria for special collections material may focus on both informational content and the uniqueness of the material itself. Given the ease of creating and copying digital material, uniqueness in the special collections context may be aided by stipulations in the deed or gift or other binding documentation that additional copies of material will not also be offered to other institutions. In other cases, even with the ease of creation and duplication, the digital content in our collections may indeed be the only known copy of that content. Unique digital materials for special collections may include the laptop and associated system files of a particular writer, for which both the digital content (the written material on particular subjects within the files, or correspondence), and the material itself (the physical or emulated environment of the laptop) could be considered unique. Uniqueness criteria could also relate to mediums or subject focus that by their nature produce unique or one-of-a-kind resources, systems or processes, such as born-digital works of art, or audiovisual materials. In the first example case, selection for digital preservation would include curatorial decisions regarding how this unique environment should be reflected in digital preservation action. Is preserving the unique written material in the files enough, such that the format of the files themselves may not matter over time, so long as the written word is accurately reflected? Or is it imperative and important to the uniqueness of the material that it be presented to researchers in an

environment that mimics the unique original as closely as possible? These are curatorial decisions that have major digital preservation implications.

Closely tied to both general and special collection concepts of uniqueness is the importance of technical, administrative, and descriptive metadata for digital materials in the selection for digital preservation. For general collections, the importance of having information regarding the creator of the unique content and their authority in that particular subject area, as well as information regarding the technical parameters of their unique content production to allow for preservation decisions is paramount. Similarly, special collections digital materials need metadata and documentation related to a creator or donor activities germane to the creation and modification of digital materials, as well as technical or descriptive metadata that will aid with questions related to provenance and authenticity of digital materials. For both types of collections, the importance and relation of metadata to the structure of systems and context of digital materials is imperative to long term digital preservation.

6.3 Cost

Cost must be considered as an important criterion for the selection of content for digital preservation. For general collections, this criteria may take the form of calculating the actual costs of replacing purchased or licensed materials (including general collections that may have been digitized, or will need to be re-digitized if a master digital file is not maintained), the cost of any preservation actions such as re-binding, as well as the researcher or user costs of having to wait an extra day for inter-library loan, or potentially not having immediate access to digital journal articles.

Costs related to the selection of both general and special collections material for digital preservation may include the actual costs of server storage, systems integration, development of systems to allow for the ingest, transfer, description, access and storage of born digital or digitized materials, or the time of the digital preservation practitioner in analyzing the content and coming up with preservation strategies, especially for less commonly used formats.

Costs specific to the selection of materials for special collections may include the costs of creating or maintaining unique systems to allow for the preservation of unique or rare digital content or the creation of emulation environments, the storage costs of large volumes of digital material such as audiovisual content or uncompressed master image formats, and the cost of subscription services or staff developer time to allow for the technical collection and preservation of born digital content such as web archives or born digital archives. Costs could also be associated with political or administrative costs associated with taking no action to preserve digital content, and allowing it to become unusable or disappear over time.

Maintenance of acquired digital materials over time incurs continued storage and migration costs, which should also be calculated as part of overall digital preservation selection priority, as well as consortia or hosted distributed digital preservation efforts. Costs of selection of content for digital preservation for both general and special collections digital material must also account for access to these materials over time in a way that encourages researcher discovery and use.

6.4 Legal / Fiduciary

Some institutions have legal requirements or institutional mandates to preserve specific types of materials over time, many of which have become available only in digital formats. For general collections, these types of collections may include government or state documents for which an institution is responsible. These types of considerations for digital preservation may also include materials

and data produced for federal grants and their preservation over time may also be important for consideration. For some academic special collections and archives, these types of collections of digital materials may relate university function and administration, as well as participation in institutional consortia. Often times, records retention schedules will classify some records for permanent retention. Selection for digital preservation for both cases may therefore consider and reference these specific institutional legal mandates, and how they might be fulfilled by digital preservation approaches.

6.5 Restrictions

Selection of materials for digital preservation must consider restrictions on access and use of the materials. If nothing else, we preserve for future access; and if this access is restricted, the benefit of preserving the content should be closely examined. For general collections, depending on institutional capacity for risk, restrictions could relate to the known or available rights information to make licensed or previously purchased digital materials available for researchers over time, which could impact the research value of these collections, as well as shift digital preservation selection decisions. The role of restrictions for general collections may also affect an organizational ability to fulfill consortia commitments of lending or interlibrary loan, as well as larger consortia efforts for maintaining complete journal runs.

For special collections materials, rights and restrictions are equally important. During the negotiation for legal custody of materials, rights and restrictions information has a large role to play for long term preservation and access approaches. Once special collections takes custody of materials, the responsibility for preserving and providing access to materials in a timely manner shifts to special collections, and should have an impact on curatorial decisions for digital preservation based on available rights and related restriction information that may affect preservation strategy, priority, and storage and access decisions.

Excessive donor restrictions on materials for long lengths of time may render them essentially unusable, and may potentially used as a mechanism to calculate whether these materials should be prioritized for selection. Restrictions of this kind may be written into policy that will enable repositories not bound by legal or other institutional mandate to actually reject collections that have excessive restrictions [2]. However, for digital materials bound by legal or institutional mandate to remain closed for some period of time, such as a ten year restriction on a past university president's email correspondence, digital preservation selection and strategy may be necessary when the collection is acquired to ensure that the records in question will be maintained, made available, and readable when the access restrictions are lifted.

6.6 Preservability of Content and Context

Both general and special collections selection criteria have represented the importance of preservation considerations in both theory and practice frameworks for collection building [3, 13, 14, 16]. Selection of materials for digital preservation adheres to the same traditional tenet of preservation; to extend the life of materials by preservation actions that will enable continued access over time.

However, for digital materials in both general and special collections, the preservability considerations of the content and context of digital materials for selection for digital preservation may take very different forms from traditional preservation approaches. Selection for digital preservation may occur very early in the material-creation or purchasing process, from legal agreements or contracts for the preservation or reuse of licensed digital materials in general collections. The preservability criterion may also reflect early discussions with creators or donors of born digital content around their creation or computing processes over time, user doc-

umentation or metadata relating to their digital materials, private or sensitive information, expectations for functionality of digital content, and rendering or importance of content of digital materials over the context of their creation. Digital preservation criteria for digitized rare materials, such as maps or books, may take into account the availability of these digitized resources via consortia or other repositories, the quality of digital preservation master files, and the quality of associated metadata.

Preservability may also involve earlier intervention or review of materials to enable an understanding and preservation approach for digital material content, and context of creation for use of that content. This preservation review may include networked or system environments necessary for the preservation of digital materials, as well as the potential creation or packaging of important metadata related to the technical, administrative or descriptive metadata of digital materials that will enable their continued identification, migration, emulation and access over time. Basic digital preservation actions such as bit level preservation are necessary earlier on in the selection process for digital materials, which will create checksums to ensure authenticity and monitor file integrity over time, and initial preservation action may be necessary for materials that digital preservationists identify as endangered or highly unstable, such as endangered audiovisual formats or rare digital formats. Preservability considerations must also take into account the necessary work to stabilize and render digital materials over time, which can take the form of digital appraisal workflows or digital forensics to identify file formats and systems as well as content that may be present on physical media carriers.

Digital format, code languages, and content encoding also plays a role in preservability. Open standard formats such as the Open Document Format are understood and documented, stable coding languages such as C, and codecs such as vorbis make it much easier to preserve digital content; even if they fall out of use, they are openly documented and software may be written to render the content. Widely-used and closed-formats, such as Microsoft Office Word's docx format, are easier to preserve, but may still need to be reformatted or normalized to guarantee preservation of the informational content.

Closed-source and proprietary formats, such as CorelDraw's cdr format, are more challenging to preserve and are closely related to rights, particularly for older content whose formats are no longer in common use. Without the creation software or documentation, it may only be possible to preserve part of the content. Older software and other encoded content, such as websites and applications, can present challenges as well. Some languages, such as PHP, don't always age gracefully, particular in regards to security vulnerabilities. It therefore may only be possible to preserve screen shots or create static web archives for these cases.

7. CONCLUSION

These guidelines and criteria are a first step towards a deeper understanding between selectors, curators and digital preservation professionals. Each role offers expertise that is necessary to make long-term decisions about the collections we steward. Selectors and curators have the domain expertise and familiarity with user needs that are necessary for establishing the value of a resource. Digital preservation professionals have the technical expertise that is necessary to make an informed decision on whether or not to select materials for digital preservation. These guidelines and criteria are an accessible basis for selectors and curators to make digital preservation selection decisions grounded in the theory and practice of selection. Their adoption may increase velocity towards achieving a cohesive digital preservation program. But they are just the beginning.

For instance, it has been documented that digital materials may not yet be well represented in the collection policies that enable subject specialists and selectors and curators to make sound decisions for selection; this is another area for potential reinforcement or examination [4]. How can we best document and make accessible and scaleable our decisions for the collection and preservation of content, particularly for items we may choose not to preserve or select in full [2]? Do the concepts of weeding and reappraisal/deaccessioning apply to past digital preservation selection decisions?

A framework can help to establish high-level policies and workflows that are capable of dealing with the majority of collections. With most content fitting into these suggested workflows, selectors, curators, and digital preservation professionals can focus their efforts on the exceptions and edge cases that require additional expertise and more in depth digital preservation actions. Collaboration towards preservation priorities contributes toward a cohesive and comprehensive collection strategy so that we are all able to meet the needs of our users.

References

[1] David Bearman. 1993. Record-Keeping Systems. *Archivaria* 36, (Fall 1993). 16-36. Retrieved April 6, 2018 from `https://archivaria.ca/index.php/archivaria/article/view/11932/12886`

[2] Frank Boles. 2005. *Selecting and Appraising Archives and Manuscripts*. Society of American Archivists, Chicago.

[3] Frank Boles and Julia Marks Young. 1985. Exploring the Black Box: The Appraisal of University Administrative Records. *American Archivist* 48, 2 (Spring 1985). 121-140.

[4] Sarah C. Choy, Nicholas Crofts, Robert Fisher, Ngian Lek Choh, Susanne Nickel, Clément Oury, and Katarzyna Slaska. 2016. *The UNESCO/PERSIST Guidelines for the Selection of Digital Heritage for Long-Term Preservation; 2016*. United Nations Educational, Scientific, and Cultural Organization. Retrieved April 10, 2018 from `http://unesdoc.unesco.org/images/0024/002442/244280E.pdf`

[5] Terry Cook. 1997. What is Past is Prologue. *Archivaria* 43 (Spring 1997). 17-63. Retrieved March 23, 2018 from `https://archivaria.ca/index.php/archivaria/article/view/12175/13184`

[6] Terry Cook. 2004. Macro-appraisal and Functional Analysis: Documenting Governance Rather than Government. *Journal of the Society of Archivists* 25, 1. 5-18. DOI: `https://doi.org/10.1080/0037981042000199106`

[7] Terry Cook. 2011. 'We are What We Keep; We Keep What We Are': Archival Appraisal Past, Present and Future. *Journal of the Society of Archivists* 32, 2 (Oct. 2011), 173-189.

[8] Richard K. Gardner. 1981. *Library Collections, their Origin, Selection, and Development*. McGraw-Hill, New York.

[9] Mark A. Greene. 2009. A Brief and Opinionated History of Archival Appraisal Theory, to 2005. Retrieved April 5, 2018 from `http://www.archivists.org/prof-education/pre-readings/FAA/A%20Brief%20and%20Opinionated%20History%20of%20Archival%20Appraisal%20Theory2008%20.rtf`

[10] Mark A. Greene and Todd Daniels-Howell. 1997. Documentation with an Attitude: A Pragmatists Guide to the Selection and Acquisition of Modern Business Records. *The Records of American Business*. James M. O'Toole (Ed). Society of American Archivists, Chicago.

[11] Vicki L. Gregory. 2006. *Selecting and Managing Electronic Resources: a How-To-Do-It Manual for Librarians*. Neal-Schuman Publishers, New York.

[12] Vicki L. Gregory. 2011. *Collection Development and Management for 21st Century Library Collections: an Introduction*. Neal-Schuman Publishers, New York.

[13] F. Gerald Ham. 1993. *Selecting and Appraising Archives and Manuscripts*. Society of American Archivists, Chicago.

[14] Dan C. Hazen. 1982. Collection Development, Collection Management, and Preservation. *Library Resources & Technical Services* 26, 1. 3-11.

[15] Frank W. Hoffmann. 2005. *Library Collection Development Policies: Academic, Public, and Special Libraries*. Scarecrow Press, Lanham, Md.

[16] Geof Huth. 2016. Appraising Digital Records. *Appraisal and Acquisition Strategies*. Michael Shallcross and Chrisopher J. Prom (Eds). Society of American Archivists, Chicago.

[17] Timothy D. Jewell, 2001. *Selection and Presentation of Commercially Available Electronic Resources: Issues and Practices*. Digital Library Federation, Council on Library and Information Resources.

[18] Joint Steering Committee for Revision of AACR. 2002. *Anglo-American Cataloguing Rules: Second edition : 2002 revision*. American Library Association, Chicago.

[19] Peggy Johnson. 2014. *Fundamentals of Collection Development and Management*. ALA Editions, Chicago.

[20] William A. Katz. 1980. *Collection Development: the Selection of Materials for Libraries*. Holt, Rinehart, and Winston, New York.

[21] Diane K. Kovacs. 2004. The *Kovacs Guide to Electronic Library Collection Development: Essential Core Subject Collections, Selection Criteria, and Guidelines*. Neal-Schuman Publishers, New York.

[22] Thomas W. Leonhardt (Ed.). 2006. *Handbook of Electronic and Digital Acquisitions*. Haworth Press, New York.

[23] Emily Maemura, Nicholas Worby, Ian Milligan, and Christoph Becker. 2018. If These Crawls Could Talk: Studying and Documenting Web Archives Provenance. Pre-print. Retrieved April 6, 2018 from `https://tspace.library.utoronto.ca/handle/1807/82840`

[24] Suzanne Mangrum and Mary Ellen Pozzebon. 2012. Use of Collection Development Policies in Electronic Resource Management. *Collection Building*; Bradford 31, 3. 108-114. DOI: `https://doi.org/10.1108/01604951211243506`

[25] Ian Milligan, Nick Ruest, and Jimmy Lin. 2016. Content Selection and Curation for Web Archiving: The Gatekeepers vs. the Masses. Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries. June 19-23 2016: Newark, New Jersey. Retrieved April 8, 2018 from `https://dl.acm.org/citation.cfm?id=2910913`

[26] Dibyendu Paul. 2011. Collection Development Policy and Selection Criteria for Electronic Materials. *International Journal of Information Dissemination and Technology*; Ambala 1, 3. 152.

[27] John Rutledge and Luke Swindler. 1987. The Selection Decision: Defining Criteria and Establishing Priorities. *College & Research Libraries* 48, 2. 123-131. DOI: `https://doi.org/10.5860/crl_48_02_123`

[28] T.R. Schellenberg. 1956. The Appraisal of Modern Records. *Bulletins of the National Archives*, No. 8 (October 1956). Retrieved March 20, 2018 from `https://www.archives.gov/research/alic/reference/archives-resources/appraisal-of-records.html`

[29] Society of American Archivists. *Guide to Deeds of Gift*. Retrieved April 10, 2018 from `https://www2.archivists.org/publications/brochures/deeds-of-gift`

[30] Society of American Archivists. Selection. *A Glossary of Archival and Records Terminology*. Retrieved April 10, 2018 from `https://www2.archivists.org/glossary/terms/s/selection`

[31] Syracuse University Libraries. Recorded Sound and Music. Retrieved April 9, 2018 from `https://library.syr.edu/scrc/collections/areas/sound.php`

[32] Nathan Tallman. 2018. *Selection for Digital Preservation Dataset*. DOI: `https://doi.org/10.18113/S1T334`

[33] Kate Theimer (Ed). 2015. *Appraisal and Acquisition: Innovative Practices for Archives and Special Collections*. Rowman and Littlefield, United Kingdom.

[34] Thomas Fisher Rare Book Library. Collections Overview. Thomas Fisher Rare Book Library. Retrieved April 13, 2018 from `https://fisher.library.utoronto.ca/resources`

[35] United States Code, Title 17, Chapter 1, Section 108: Limitations on Exclusive Rights: Reproduction by Libraries and Archives. Retrieved April 6, 2018 from `https://www.law.cornell.edu/uscode/text/17/108`

[36] Lauren Work and Nathan Tallman. 2018. Guidelines and Criteria to Select for Digital Preservation. DOI: `https://doi.org/10.18113/S1RM0J`