

APPRAISING DIGITAL RECORDS FOR LONG-TERM PRESERVATION

Terry Eastwood

School of Library, Archival and Information Studies, the University of British Columbia, 6190 Agronomy Road, Vancouver, B.C., Canada V6T 1Z3.

Email: eastwood@interchange.ubc.ca

ABSTRACT

This paper aims to extract lessons from archivists' experience of appraising electronic records that are likely to have wider application in the preservation of other digital materials, including scientific data. It relies mainly on the work of the Appraisal Task Force of the InterPARES project on long-term preservation of authentic electronic records to develop a picture of the process of appraisal. It concludes that the aspects of assessment of authenticity, determination of the feasibility of preservation, and monitoring electronic records as they are maintained in the live environment are likely to find counterparts in attempts to appraise digital objects for long-term preservation in the scientific community. It also argues that the activities performed during appraisal constitute the first vital step in the process of preservation of digital materials.

Keywords: Archival appraisal, Authenticity of records, Digital preservation, Electronic records

1 INTRODUCTION

As is by now well-known, long-term preservation of materials in digital form presents both organizations and individuals with a significant challenge to overcome the adverse effects of technological obsolescence and media fragility. Contemporary societies the world over need to keep all kinds of digital objects long enough to make this a problem in virtually every aspect of human endeavour, but it is most pronounced for archival institutions that are assigned responsibility for preserving the records of all facets of human endeavour for as long as we can preserve them. The difficulties of long-term preservation of digital objects grow in complexity with the complexity of digital technology. These difficulties, even for relatively simple digital objects, place a premium on our capability to identify those materials of enduring value, and demand that we develop a process of evaluation or appraisal adapted to the needs of the digital environment. At its essence, appraisal involves making a judgment or estimation of the worthiness of continued preservation. This paper reports on the experience and research in the archival community to come to grips with this problem of appraisal of digital objects. It will argue that appraisal in this realm is in some of its particulars significantly different from that which we have experienced for traditional documentary objects, that the process is vital to effective long-term preservation, and that there is more than meets the eye to the exercise. The aim is to extract from archival experience lessons likely to be worth pondering in spheres where selection and long-term preservation of digital assets are at issue.

2 THE APPRAISER AS ACTIVE AGENT OF PRESERVATION

The materials normally preserved by archival institutions were originally produced in the course of some human activity and derive their meaning from the context of their creation. Archivists appraise, acquire, preserve, and make accessible documents made and received in the course of the affairs of individuals and organizations, and regularly communicate knowledge about them and their context as part of the exercise of ensuring their continuing accessibility. By the 1980s, archivists recognized that digital documents, which they usually call electronic records, were being produced with little regard for the requirements for their long-term preservation. By the end of that decade, it had become evident that it was difficult if not impossible to preserve digital or electronic documents produced in so-called legacy systems operating just a

few years before, because little attention had been paid to identify important facts of their creation, relationships, and context, including their technological context. This was the era that McDonald (1995) has called the wild frontier, when people employed computer hardware and software to make documents, store them, and maintain them with almost no thought for their use outside the immediate environment in which they were created. Archivists concluded that record-keeping requirements, including provision for determining the disposition of records, had to be part of the design of systems software. In the last few years such software has appeared, but it has neither invaded all sectors of human endeavour nor is it adequate in all situations for all the kinds of digital objects, data, and information being created today. It is hard not to believe that archivists' experience will parallel that in any sphere where digital objects are being created and need long-term preservation. In particular, two things are evident. First, digital objects have to contain the means to identify them and their context of creation so that they can be managed beyond the active system in which they were generated, on the one hand, and have to be maintained so as to avoid corruption or loss of their essential characteristics, on the other. Second, the determination of their disposition, that is their appraisal, cannot without significant risk be delayed to some time in the future when the technology of creation is no longer extant.

This archival experience suggests that anyone responsible to select and preserve digital objects as records will have to seek materials actively in the here and now and be prepared to educate creators of them about the needs of long-term preservation. Seeking materials in the here and now is, given the proliferation of digital objects, an almost overwhelming task. Intervening with advice to promote the needs of long-term preservation is, practically speaking, a task it is all too easy to regard as being quixotic. Still, a literature survey conducted by the author (Eastwood, 2000) revealed that most authorities advocate actively appraising materials close to the time of their creation before it is too late to consider long-term preservation or effectively bring it about.

3 THE PROCESS OF ARCHIVAL SELECTION

In recent years, archivists have done a lot of work to understand and characterize the various activities performed during the archival appraisal process. Relying on this body of work and on the experience of archival institutions, the Appraisal Task Force of the InterPARES project on long-term preservation of authentic electronic records produced a model of the archival selection function.

The model takes the perspective of an entity responsible for long-term preservation of electronic records, for short the preserver. Selecting electronic records involves appraising them and carrying out their disposition, that is, carrying out the transfer of custody of records selected for preservation from the entity creating them, for short the creator, to the preserver. The preserver could be an office of the organization that created the records, or some institution that takes responsibility for long-term preservation of records created by other entities. One of the large problems the scientific community and the organizations of which it is composed face is to establish and sustain preserver entities, whether inside organizations or as "collecting" institutions, to look after scientific records. In any given case, the preserver needs to establish, implement and maintain a framework of policies and procedures guiding the selection function, such as an acquisition policy, appraisal criteria, or procedures governing actual physical transfer of records from their creator or custodian to the preserver. Perhaps the most difficult exercise in this sphere of management framework is to decide upon appraisal criteria, which are of course peculiar depending on the realm of records under consideration. Nonetheless, preserving records, and one would presume preserving scientific data responds to the creator's needs and broader societal needs to have continuing accessibility to the records, information, or data. In this sense, a large part of establishing appraisal criteria involves looking beyond the data itself to the ongoing needs of the entity that created it and to the continuing needs of society and in particular the scientific community for the data. One other factor constrains the exercise of selection, It is particularly necessary to preserve electronic records in such a way that their identity is known and their integrity is not impaired, in short that they are authentic, are what they purport to be. The framework of policies and procedures, the creator's and societal needs, and dictates of proper preservation all act as constraints on the process, as do the terms of laws or regulations that may apply to the records.

4 ARCHIVAL APPRAISAL OF ELECTRONIC RECORDS

Appraisal of electronic records comprises four distinct activities. I see no reason why you could not substitute scientific data when I say electronic records. The four activities are (1) compiling and analyzing information about the records and their contexts, (2) assessing their capacity to serve the needs of their creator and society, (3) determining the feasibility of preserving them, and, on the basis of the foregoing, (4) making the appraisal decision. Although evaluating digital objects other than those of archival nature for the purposes of long-term preservation may very well take on a somewhat different cast from that for records, it is worth looking at these four activities as they are likely to be very generally applicable to appraisal of all digital objects.

4.1 Compiling information

The first activity involves compiling information about the digital object or objects. Archivists normally appraise records in the natural aggregations to which they belong. These aggregations are natural in the sense that any given activity will produce records as part and parcel of conducting it. In the realm of science, it is probably a mistake to think that data on its own will be sufficient for future understanding, something that I trust will become clear as we examine how digital materials produced in the course of any activity are dependent on for their meaning on an understanding of their context. One can construe information relevant to the process of evaluation from the objects themselves, about their form, content, and so on, or one can compile information from sources external to the records about the various contexts of the objects relevant to their evaluation. The InterPARES project identified five distinct contexts of records. The first, the juridical-administrative context, refers to the legal and administrative system in which the creating body exists. It is the broad societal context of the country or territory where the objects were produced at the time they were produced. This context applies to the conduct of scientific investigations, which occur in particular legal and administrative context. The second, the provenancial context, refers to the creating body, its mandate, structure, and functions. For records produced by individuals, provenancial context would be the sphere or spheres of affairs and the various activities in which the person was involved that produced the records. Much of the data produced in “big science” would appear to fit nicely into an organizational framework, and understanding of which is necessarily a precondition of good appraisal. The third, the procedural context, refers to the business procedure in the course of which the records are created. Although business procedure suggests a bureaucratic environment, the concept can be extended, for instance, to encompass methods and procedures in scientific activity, whether in the context of the work of an organization or agency, a large collaborative project, or an individual scientist. The fourth, the documentary context, refers to the structure and interrelationships of the whole body of records produced by an entity. It is an indispensable condition of archival appraisal that decisions take into account the whole body of records, the meaning of individual records being dependent on their various relationships with other records in the whole body of records under consideration. The fifth, the technological context, refers to the technical components of the electronic systems in which the records are created, the hardware, software, data, and systems with which or within which records are created.

4.2 Assessing value

During this activity of compiling contextual information, archivists accumulate inferences about the likelihood the records under consideration will have continuing value, on the one hand, and about the degree to which they have enjoyed circumstances that suggest they have remained uncorrupted or unchanged over time or not, that is the degree to which we can presume them to be authentic. Continuing value simply refers to the capacity of records to serve the continuing interests or needs of their creator and society. As I said earlier, lurking behind any specific criteria used to identify long-term value will be suppositions, whether stated or not, about the needs preservation of the material will serve. Such suppositions are usually projected from past experience of the use of records. In the case electronic records, and one might propose in the case of scientific data as well, assessing value comprises assessing continuing value and assessing authenticity. Assessment of continuing value is of course the most critical

element, but that assessment is obviously affected by an assessment of the grounds for presuming records to be authentic. In large measure, archivists gather and evaluate evidence of what has happened to records during the course of their existence in order to assess authenticity. As the report of the Appraisal Task Force of the InterPARES Project (2002) put it:

In cases where the chain of custody and preservation has been broken or where migration has resulted in missing records, missing parts of records, or inadequate documentation of changes, there may be good reason to suspect the value of the records. If the appraiser has good reason to suspect that the records no longer reflect what they were at the time of their creation and primary use, he or she may decide not to preserve them.

Such evidence is not always easy to come by long after the fact of creation, which is one reason to intervene soon after digital objects are created. One might reasonably assume that assessment of the authenticity of other than archival digital objects will also affect judgments of continuing value. The proposition would go something like this: the greater the extent to which you can presume something to be the real thing, to use a common phrase, the more one is likely to accord it continuing value. Given the great facility with which to alter things either intentionally or unintentionally in the digital environment, this two-faceted assessment to determine value is very likely to become a regular feature of the process of evaluation of any kind of digital material for the purposes of long-term preservation

4.3 Determining the feasibility of preservation

Another aspect of appraisal in the digital environment is determining the feasibility of preservation. Archivists have of course always had to consider the feasibility of preserving records, but for electronic records (and also for scientific data, one suppose) this aspect of the process becomes a more critical element in the process than it ever was with traditional materials, where it had rarely been the case that it was simply not possible to preserve. This very outcome is more likely to occur with digital materials.

To explain what is involved in determining the feasibility of preservation, suppose that you have compiled all the information you need and analyzed it to make an informed judgment about continuing value, and have assured yourself that there is good reason to believe that what you have assessed is indeed authentic. You still have to ask yourself, can I preserve these digital objects in such a way that the essential elements conferring their identity and ensuring their integrity can be preserved? This is in part important because, in the digital environment, we do not actually preserve records, but rather digital objects and the capability to reproduce them. This is not a problem so long as the digital objects exist in their native environment, but short of perpetuating that environment or recreating it in the future, the only other strategies involve sustaining the capability of making copies. This involves determining that the digital components conferring identity and ensuring the integrity of the things you want to preserve *can indeed be preserved given current and future anticipated preservation capabilities*.

The InterPARES Appraisal Task Force broke this activity of determining feasibility of preservation into three phases.

1. The appraiser determines both the record elements containing informational content and those elements that need to be preserved according to the requirements for authenticity.
2. The appraiser identifies where these crucial record elements are manifested in digital components of the electronic record that must be preserved.
3. The appraiser reconciles these preservation requirements with the preservation capabilities of the entity that is responsible for the continuing preservation of the body of records being preserved.

The InterPARES project developed a set of requirements for authenticity for archival documents or records in the digital environment. No doubt requirements to ensure authenticity will vary for other kinds of digital documents and objects, but there seems little doubt that appraisers of such documents and objects will have to perform a parallel activity of determining feasibility. As the final report of the Task Force puts it, "this feasibility determination gathers and records technical information that is necessary to accomplish preservation of the individual elements conveying

both the intellectual content and the authenticity of electronic records being appraised.” As well, this part of the process gathers information about “the projected cost of preservation and an indication of whether or not the preserver has [or can foresee having] the capability to preserve the records in question.”

At this juncture, something needs to be said about the cost of preservation. Except to mention that it is a factor in determining feasibility, our Task Force did not dwell on cost. My own view is that the cost of preservation of digital objects of all kinds is both difficult to determine and will be rather more determinative of the outcome of appraisal than it was for traditional materials. Again, much as in the case of appraising continuing value, we have little experience to go on. However, we can reasonably project that cost will be incurred regularly to cope with technological obsolescence, as well as the normal storage and maintenance costs, including copying to new media. Many of the activities of long-term preservation will require intensive work by highly skilled persons, and so will be costly to perform in terms of the human resources needed. Because these costs of preservation are difficult to predict and will probably have an escalating quality to them the further one removes oneself from the native technological context, there is bound to be extra pressure on appraisers to select only those digital objects with continuing value that can be strongly justified. When you put together the lack of precedent for asserting that digital materials have continuing value with the need to justify the benefit to be derived in the cost-benefit analysis performed as part of determining the feasibility of preservation, it no wonder that we are as yet unsure of ourselves when appraising digital objects of all kinds. Archivists have been loath to reassess or reappraise records previously determined to have long-term value, that is, to question their own decisions. In the digital environment, we shall probably have to face up to the fact that we are in an era of experimentation, and will have to test the efficacy of decisions we make as against the cost of continuing preservation, at least until we have some solid experience on which to rely.

There is one other element of cost to consider. It seems pretty clear that developing and maintaining what almost by definition must be a cutting edge capability to preserve digital objects is very expensive. Few archival institutions in the world have yet developed that capability in part, no doubt, because any serious effort is costly both to establish and keep up. It is easy enough to say that every organization creating valuable digital materials should preserve them. Still, I think that the significant expense holds many organizations and institutions back as much as, if not more than, other factors. I would think that cooperative strategies would be the best way to address these cost concerns. In the archival field, the way we have been going is not resulting in widespread preservation of electronic records. Given that information technology is widely used to create and maintain records in both formal organizational circumstances and in personal affairs, to continue on the same path risks serious loss of our ability to sustain intelligent preservation of valuable digital materials, like scientific data, for instance.

5 MONITORING APPRAISED RECORDS

At the beginning of this paper, it was remarked that one of the earliest conclusions archivists reached about appraising electronic records was that, to be effective, it had to be done close to the time of creation or at least while the records were still active. It can be further supposed that if one appraises records long before one actually take them into the custody of the preserver, one is going to have to monitor what is happening to them from time to time. The purpose of monitoring appraised records is to determine whether the terms of an appraisal of records earmarked for continued preservation are still valid. Many changes affecting records will require little or no change to an initial appraisal. In some case, minor revisions, such as to the terms and conditions of transfer from the creator to the preserver, will be needed. However, in cases where the business processes and related computer systems are significantly revamped or rebuilt, it will obviously be necessary to consider initiating a disposition under the terms of the original appraisal *and*—for the two will likely go together—redoing the appraisal to take into account the radically altered

situation. The problem this kind of monitoring addresses is most familiar to archival preservation tied to the ongoing production of records of a parent organization, such as in the case of the designated archival institution of a government or the archival program responsible for long-term preservation of records of enduring value of a university or business. Of course archives are not unfamiliar with acquiring records from individuals and organizations in the community. In such circumstances, appraisal and its monitoring aspect will undoubtedly be more complicated, but, in

light of how difficult it is to arrive on the scene years after the fact of creation of digital material and conduct appraisal, some regimen of regular contact with prospective donors appears to be necessary in this realm as well.

6 CARRYING OUT DISPOSITION OF RECORDS

The activity of effecting disposition of electronic records according to the appraisal decision breaks down into three activities: preparing electronic records for disposition; preparing electronic records for transfer to the responsibility of the preserver; and transmitting electronic records to the preserver. As archivists see the process, appraisals set out the terms and conditions of transfer of records from the creator to the preserver, and indicate who is to do what and when they will do it.

The first step comprises copying and, if necessary, formatting those digital objects selected for preservation so as to prepare them physically for transfer. The next step is to associate the digital objects selected for preservation with the necessary information for their continuing preservation, such as the terms and conditions of transfer, identification of the digital components to be preserved, and associated archival and technical documentation needed for their continuing treatment. This information is usually compiled and recorded during the various stages of appraisal and monitoring. The task at this stage is to extract the information necessary for continuing preservation of the records from the mass of appraisal [and monitoring] documentation, and associate it with the records. The third step is to transmit the records selected for preservation along with accompanying information to the preserver.

Careful work in this phase of the process is absolutely vital. Archivists who are familiar with efforts to effect disposition of electronic records can recite a litany of horror stories. The wrong records were sent to them, or records have arrived on their doorstep in the wrong format or with no accompanying documentation. Imagine turning up a disk, tape, or CD in someone's desk drawer years hence, because some well meaning soul simply asked for a copy of something interesting.

7 CONCLUSION

Preservation of electronic records and, arguably, other digital objects depends to a greater extent than it did for traditional materials on effective appraisal in ways outlined here. This picture of archival selection undoubtedly differs in some of its particulars from what will be appropriate for scientific data, but one may suppose that anyone appraising and acquiring scientific data will have to go through similar steps or phases in the selection process in order to ensure effective their long-term retention. In this sense, we can see that appraisal is a necessary first step in the preservation process, or so it would seem to judge by the experience of archivists.

8 REFERENCES

Eastwood, T. (2000) Appraisal of Electronic Records: A Review of the Literature in English. Retrieved August 1, 2004, from the *InterPARES 2 Project* website:
http://www.interpares.org/book/interpares_book_1_app03.pdf

McDonald, J. (1995) Managing Records in the Modern Office: Taming the Wild Frontier. *Archivaria*, 39 (Spring), 70-79.

InterPARES 2 Project Appraisal Task Force (2002) The Long-term Preservation of Authentic Electronic Records: The Findings of the InterPARES Project, Part Two: Choosing to Preserve: the Selection of Electronic Records. Retrieved August 1, 2004, from the *InterPARES 2 Project* website: <http://www.interpares.org/book/index.cfm>