InterPARES 2 Project

International Research on Permanent Authentic Records in Electronic Systems

International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records

APPENDIX 20

CREATOR GUIDELINES
Making and Maintaining Digital Materials: Guidelines for Individuals

by

Philip Eppard
University of Albany, State University of New York

Status: Final (public)

Version: Electronic

Publication Date: 2008

Project Unit: Domain 2 Task Force

URL: http://www.interpares.org/display_file.cfm?doc=ip2_book_appendix_20.pdf

How to Cite: Phil Eppard, Domain 2 Task Force, "Appendix 20: Creator Guidelines – Making and Maintaining Digital Materials: Guidelines for Individuals," [electronic version] in *International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records*, Luciana Duranti and Randy Preston, eds. (Padova, Italy: Associazione Nazionale Archivistica Italiana, 2008). <http://www.interpares.org/display_file.cfm?doc=ip2_book_appendix_20.pdf>

CREATOR GUIDELINES
Making and Maintaining Digital Materials: Guidelines for Individuals[1]

Introduction

Most information today is created and stored in digital form. The advantages of the digital medium are by now familiar to everyone. Documents can be created quickly and edited and revised with ease. Thanks to the Internet, they can be distributed globally with lightning-like speed. They can be manipulated in ways that allow them to be used for multiple purposes. The digital medium also solves the longstanding storage problems associated with large files of paper records.

The blessings of the digital era, however, are not without their costs. Only in recent years have people begun to fully grasp the many problems inherent in the digital medium. For example, there is the fact that digital information can only be accessed using a computer. Furthermore, the computer must be equipped with the necessary software to be able to read the bit strings contained on the disc or tape. Ease of reproduction and the proliferation of copies make it more difficult to identify a complete or final version of a digital document. Easy distribution of information on the Internet makes the preservation of intellectual property rights difficult. Finally, all digital materials are vulnerable to viruses and simple technology failure, as well as to the rapid developments in software and hardware that risk making them inaccessible very quickly.

With all of these problems, it is little wonder that some people yearn for the comforting tangibility of paper. Yet although our systems for creating and maintaining information will likely continue for some time to be hybrid systems—that is, containing both paper and digital materials—there is clearly no turning back from the digital revolution. Consequently, everyone should be aware of the risks faced by digital materials and know how best to minimize these risks.

These guidelines have been developed for individuals who create digital materials in the course of their professional and personal activities to help them make informed decisions about making and maintaining these materials in ways that will help ensure their preservation for as long as they are needed. They may also be useful for small organizations or groups of individuals, such as medical offices, consulting groups or teams of research scientists.

Although these guidelines can be applied to various kinds of digital publications, documents and data, they are especially important for digital records. Records are the documents that you make, receive and use in your activities, and that you keep because you may need them later or because you want to have reliable evidence of what you have done. Therefore, you need to be especially careful in maintaining and preserving them. These guidelines are applicable to records that need to be maintained for only a short period of time as well as to those that require long-term maintenance. Adherence to these guidelines will help ensure that records that merit long-term preservation in an archival repository will be accessible when they are turned over to the care of a trusted custodian.

[1] These Guidelines have also been issued in an illustrated booklet form that is freely available at http://www.interpares.org/display_file.cfm?doc=ip2(pub)creator_guidelines_booklet.pdf.

Definitions

Before presenting recommendations to guide you in making and maintaining digital materials, it will be both necessary and helpful to clarify the meaning of some of the terms used in this document.

For the purposes of these guidelines, a *record* is defined as any document created (i.e., made or received and saved for further action or reference) by a physical or corporate person in the course of a practical activity as an instrument and by-product of that activity. A *publication* is defined as a document intended for dissemination or distribution to the public at large. All records and publications are documents and contain data. A *document* is information affixed to a medium in a fixed form; *information* is an assemblage of data intended for communication over time or space; and *data* are the smallest meaningful and indivisible pieces of information.

These guidelines aim at providing recommendations for the creation and maintenance of reliable digital materials in general, and records in particular, that can be accurately and authentically maintained and preserved over time. To facilitate their application, however, the terms "reliability," "accuracy," "authenticity" and "authentication" need to be defined.

For the purposes of these guidelines, *reliability* is the trustworthiness of digital materials as statements of fact or as content. It is the responsibility of the author of the materials, be that author an individual or the corporate person in whose name an individual is writing, and is assessed on the basis of the material's completeness and accuracy and of the degree of control exercised on the process of its creation.

Accuracy is the degree to which the data in the materials are precise, correct, truthful and free of error or distortion. To ensure accuracy, one must exercise control on the processes of creation, transmission, maintenance and preservation of the materials. Over time, the responsibility for accuracy shifts from the author to the keeper of the materials and later to the long-term preserver of the materials (if applicable).

Authenticity refers to the fact that the materials are what they purport to be and have not been tampered with or otherwise corrupted. Thus, with respect to records in particular, authenticity refers to the trustworthiness of records as records. To ensure that authenticity can be presumed and maintained over time, one must define and maintain the identity of the materials and protect their integrity. Authenticity is at risk whenever materials are transmitted across space and time. Over time, the responsibility for authenticity moves from the keeper to the long-term preserver of the materials.

Authentication is a declaration of authenticity, resulting either from the insertion or the addition of elements or statements to the materials in question, and the rules governing it are established by legislation. Thus, it is a means of proving that materials are what they purport to be at a given moment in time. Digital authentication measures, like the use of digital signatures, only ensure that the materials are authentic when received and cannot be repudiated, but not that they will stay authentic afterwards.

Recommendations

1. Select hardware, software and file formats that offer the best hope for ensuring that digital materials will remain easily accessible over time.

Accessing digital materials depends on having the appropriate software. Software that is not compatible with previous versions (backward compatibility) or with future versions (forward compatibility) makes it difficult to access records over time. Software for one application also needs to work well with that of other applications and systems (interoperability). Paying attention to the following factors can help ensure that your software and hardware maintain accessibility.

A. *Choose software that presents materials as they originally appeared*. Ideally, materials should keep the same look over time to be fully intelligible and accessible. Be sure that new software will be able to read your older materials in the software format in which you kept it and display it on the screen in the same documentary form in which it was originally displayed. In other words, new software should be backward compatible with older software.

B. *Choose software and hardware that allow you to share digital materials easily*. Software should be able to accept and output files in a number of different formats. The ability to interact easily with other technology is called *interoperability*. It will make it easier to access your materials and also to move them to other systems.

C. *Use software that adheres to standards*. This is one of the best things you can do to ensure your material will last. Standards endorsed by national and international organizations are best. These are called *de jure* standards.[2] If these do not exist for your material, you can help ensure longevity by adopting software that is very widely used. In the absence of an official standard, such software is often referred to as a *de facto* standard.[3] Open source software; that is, freely available non-proprietary software, is preferable (see subsection G on the next page).

D. *Keep the specifications of software*. This kind of documentation (e.g. the owner's manuals or any other more detailed description of the software you might have) will be essential in the future to access the materials or to migrate them to a new computer environment as technology advances. It is particularly important to fully document any software that you build yourself.

E. *If you customize software, make sure you document the changes you make*. Give detailed information about the changes and describe clearly the characteristics and features of the material these changes produce, as well as the outcomes you are trying to achieve by

[2] Defined as: A standard adopted by an official standards-setting body, whether national (e.g., ANSI), multi-national (e.g., CEN) or international (e.g., ISO). For computer file formats, two recent de jure standards are PDF/A (PDF standard for archiving) and ODF (OASIS OpenDocument Format).

[3] Defined as: A standard not adopted by any official standards-setting body, but nevertheless widely used and recognized by its users as a standard. Well known and widely used computer file formats that are considered de facto standards include PDF, TIFF, DOC and ZIP.

customizing the software. A good way to do this is to include the information as comments in the software code. The information will not get lost, as it is part of the file, and it will be very helpful to those who need to make adjustments later, as technology advances.

F. *Document the construction of your system as a whole to help ensure its accessibility*. You should document your system's structure and functions. This means identifying its hardware and software components, including peripherals, its operating system and software packages. Such documentation will identify how the software packages represent information, and how they process it and communicate it to each other and to users. These basic specifications will ensure that those who come after you understand the context in which you are working now. They will provide the information necessary to update the system as hardware and software evolve.

G. *Choose widely-used, non-proprietary, platform-independent, uncompressed formats with freely available specifications where possible*. These are often called "open formats," which means that their specification is published and freely available. However, it may also mean that the format is free of patent or royalty fees or the possibility of such fees being applied in the future, and/or that it is widely adopted. It should be noted that "open" formats are not necessarily the same as formats produced by *open source software*, as the latter term describes software for which the code is made freely available and can be modified. Open source software does not always produce non-proprietary formats. Distinguish between file formats, wrapper (or container) formats and tagged formats such as XML-tagged files, and ensure that version, encoding and other characteristics are clear and fully specified. For XML files, make sure that the files are well-formed and valid and accompanied by the relevant DTDs or schemas. If it is not convenient for you to follow this recommendation, consult with an archives that accepts digital materials and choose among the formats that it recommends for long-term preservation. You should not compress your digital materials, if at all possible, since this can lead to problems for their long-term preservation. If you need to compress them, choose lossless compression techniques that conform to accepted international standards.

2. Ensure that digital materials maintained as records are stable and fixed both in their content and in their form.

One of the great advantages of digital materials is the ease with which information can be edited, revised or updated. But this also means that important information can be changed or even lost, accidentally or on purpose. This is a particularly important problem for records, because one of the characteristics of a record is that its content is unchanged and unchangeable. This implies that the information and the data in the record cannot be overwritten, altered, deleted or expanded. A system that contains fluid, ever-changing information or data does not really contain records until someone decides to make them and save them with *fixed form*[4] and *stable content*.[5]

[4] Defined as: The quality of a record that ensures the documentary appearance or presentation is the same each time the record is retrieved.

[5] Defined as: The quality of a record that makes the information and data contained in it immutable, and requires changes to be made by appending an update or creating a new version.

Although the idea of stable content is fairly simple, the concept of fixed form is more complex. Essentially, it means that the message conveyed by a digital record (or other digital object) can be rendered with the same documentary presentation it had on the screen when it was made or received and first saved. The bit streams that compose the digital record and determine its digital presentation (i.e., its file format) may change, but its documentary presentation must not change. A simple example is when a document created in Microsoft Word is later saved as an Adobe PDF file. Although the document's digital presentation has changed—from a Microsoft Word .doc file format to an Adobe .pdf file format—the documentary presentation of the document—also called its *documentary form*[6]—has not changed, and therefore we can say that the document has a fixed form.

In some cases, digital materials can be presented in several different ways—in other words, the information they convey can take different documentary forms. For example, statistical data can be presented as a pie chart, a bar chart or a table. However, the possible variations of these displays are usually limited by the system. In such cases, we can regard each documentary presentation as having stable content and fixed form, since the information is selected from a fixed store of data within the system and the system's rules govern the form of its documentary presentation(s).

A similar situation occurs when the selection of both content and form is from a large store of fixed information that is only partially accessed every time a user queries the system. If the same query always produces the same output as to content and documentary form, the output can be regarded as having stable content and fixed form. Thus, if you, as the author of the record, establish fixed rules for the selection of its content and of its documentary form that only allow for a known and stable range of variability— that is, endow it with *bounded variability*[7]—then you can claim that your material has stable content and fixed form.

The concern for the documentary presentation of digital materials is particularly important for maintaining and assessing the reliability and accuracy of records. Future upgrades, conversions or migrations of data may result in changes to the documentary form. Therefore, you would be wise to first establish the documentary form of records associated with each activity or procedure and then identify the essential characteristics (i.e., the essential *intrinsic* and *extrinsic* elements[8]) of each documentary presentation or form. This will help alert you to any changes in the future that would imply a loss of identity and integrity of the record, especially if you are active in the sphere of digital art, where a certified description of those essential characteristics by the artist would help support the recognition of the intellectual property rights linked to work so described.

[6] Defined as: The rules of representation according to which the content of a record, its administrative and documentary context and its authority are communicated. Documentary form possesses both extrinsic and intrinsic elements.

[7] Defined as: The quality of a record that ensures that its documentary presentations are limited and controlled by fixed rules and a stable store of content data, form data and composition data, so that the same user activity, query, request or interaction always generates the same result.

[8] *Intrinsic Elements* are defined as: The elements of a record that convey the action in which the record participates and its immediate context, including the names of the persons involved in its creation, the name and description of the action or matter to which it pertains, the date(s) of creation and transmission, etc. *Extrinsic Elements* are defined as: The elements of a record that constitute its external appearance, including presentation features such as font, graphics, images, sounds, layouts, hyperlinks, image resolutions, etc., as well as digital signatures, seals, and time stamps and special signs (digital watermarks, logos, crests, etc.).

3. Ensure that digital materials are properly identified.

Giving a meaningful name to a computer file helps identify its content and makes it easier to find. The full identification of records is more complex than just naming files, however. Full identification is essential in distinguishing records from each other, in distinguishing different versions of a single record and in providing evidence of the identity of a record from the moment of its creation through its long-term preservation.[9]

The information about digital materials that supports their identification and retrieval is commonly referred to as *metadata*.[10] Most software applications automatically tag all digital materials with some data about their identity because this information is necessary to locate documents effectively. Without metadata, it would be nearly impossible to find a document without opening and reading through a folder or several directories. Metadata describe the properties or attributes of digital materials. In the case of records, however, these properties or attributes are also necessary to maintain and assess their authenticity, and that is why it is important to ensure that all the essential ones are recorded and that they are correct.

The properties or attributes conveying the identity of digital materials are referred to as identity *metadata*.[11] These include:

a. *Names of the persons involved in the creation of the digital materials*. These include:
    - the *author*—the physical or corporate person(s) responsible for issuing the materials;
    - the *writer*—the physical person(s) or position(s) responsible for articulating the content of the materials;
    - the *originator*—the physical person, position or office responsible for the electronic account or technical environment where the materials are generated and/or from which it is transmitted;[12]
    - the *addressee*—the physical or corporate person(s) for whom the materials are intended; and
    - the *recipient*—the physical or corporate person(s) to whom the materials may be copied or blind copied.
b. *Name of the action or matter*—in other words, the title or subject.
c. *Documentary form*—in other words, whether it is a report, a letter, a contract, a table, a list, etc.
d. *Digital presentation*—in other words, format, wrapper, encoding, etc.
e. *Date(s) of creation and transmission*. These include:
    - the *chronological date* written on the materials or on which the materials were compiled;

[9] In this context, *identity* is defined as: The whole of the characteristics of a document or a record that uniquely identify it and distinguish it from any other document or record. With integrity, a component of authenticity. (See also Recommendation 4.)
[10] Defined as: Information that characterizes another information resource, especially for purposes of documenting, describing, preserving or managing that resource.
[11] Defined as: The properties or attributes conveying the identity of a digital object that is to be kept as a record. (See also Recommendation 5.)
[12] Identification of the originator is only important in cases where the person, position or office responsible for physically creating and/or transmitting the materials is neither the author nor the writer, and when the presence of the originator's name appearing on, or in association with, the materials calls into question the actual author and/or writer of the materials. This is most commonly associated with e-mails in instances where the name of the originator appears in the header of an e-mail and/or its attachments that were in fact authored and/or written by another person, but physically manifested and/or transmitted on behalf of that person by the originator.

- the *dates of transmission and/or receipt*; and
- the *archival or filing date*—in other words, the date when the materials were associated with a computer folder or directory, or other classification scheme or filing plan (see Recommendation 5).

f. *Expression of documentary context*—for example, a classification code, or the name of the computer folder or directory, or comparable filing unit within the classification scheme or filing plan to which the materials are associated, and the name of the broader group of records in which the materials belong (see also Recommendation 5).
g. *Indication of attachments*, if applicable.
h. *Indication of copyright or other intellectual rights*, if applicable.
i. *Indication of the presence or removal of a digital signature*, if applicable (see Recommendation 6, Technology-dependent Authentication section).
j. *Indication of other forms of authentication*, if applicable. This could include, for example, the presence of a corroboration (i.e., an explicit mention of the means used to validate the record); an attestation (i.e., the validation of a record by those who took part in the issuing of it, and by witnesses to the action or to the 'signing' of the record); a subscription (i.e., the name of the author or writer appearing at the bottom of the document), or a qualification of signature (i.e., the mention of the title, capacity and/or address of the person or persons signing the record).
k. *Indication of the draft or version number*, if applicable.
l. *Existence and location of duplicate materials outside the digital system*, if applicable. If multiple copies of a document exist, you should indicate which one is the official or *authoritative copy*.[13] If the document is certified by the author as an "approved reproduction" of a work (for example, a digital work of art), indication of the existence of such certification is required. If the document comprises material copyrighted by different author(s), indication of copyright clearance (or lack thereof) with related dates is necessary.

4. Ensure that digital materials carry information that will help verify their integrity.

Although the identity metadata help distinguish digital materials from one another, another set of metadata allows users to infer that the materials are the same as when they were created (although not to verify or demonstrate it, because this would require comparison with a copy of the materials kept elsewhere). These metadata can be referred to as *integrity metadata* (see below). Digital materials have *integrity*[14] if they are intact and uncorrupted, that is, if the messages that they are meant to communicate to achieve their purposes are unaltered. This means that the physical integrity of digital materials, such as the proper number of bit strings, may be compromised, provided that the articulation of the content and its required elements of *documentary form* (see Recommendation 2) remain the same. The content and the data in it are considered to be unaltered if they are identical as to the value and presentation (i.e., position on the screen) of the content and data in the first saved manifestation of the material. The attributes that relate to the integrity of digital materials have to do with the maintenance of the materials, including the responsibility for their proper handling, such as overseeing and documenting any

[13] Defined as: The instance of a record that is considered by the creator to be its official record and is usually subject to procedural controls that are not required for other instances.
[14] Defined as: The quality of being complete and unaltered in all essential respects. With identity, a component of authenticity.

technological transformations or transfers of the materials to other systems. The integrity metadata include:

a. *Names of handling person/office*—the person or office using the materials to carry out business.
b. *Name of person or office with primary responsibility for keeping the materials*—this may be the same as the handling person/office.
c. *Indication of annotations added to the materials*, if applicable.
d. *Indication of any technical changes to the materials or to the application(s) responsible for managing and providing access to the materials*—for example, change of encoding, wrapper or format, upgrading from one version to another of an application, conversion from several linked digital components to one component only (e.g., by embedding directly in the materials digital components that were previously only linked to the materials, such as audio, video, graphic or text elements like fonts).
e. *Access restriction code*—indication of the person, position or office authorized to read the materials, if applicable.
f. *Access privileges code*—indication of the person, position or office authorized to annotate the materials, delete them, or remove them from the system, if applicable.
g. *Vital record code*—indication of the degree of importance of the record to continue the activity for which it was created or the business of the person/office that created it, if applicable.[15]
h. *Planned disposition*—for example, removal from the live system to storage outside the system; transfer to the care of a *trusted custodian* (see Recommendation 10); scheduled deletion.

5. Organize digital materials into logical groupings.

The management and retrieval of your digital materials can be enhanced if you can handle them in large sets, rather than one by one. Therefore, it is important that you group your digital materials in some logical manner. The categories chosen may reflect the way you work, your activities, procedures, thematic areas, or some sort of structural organization. Separating your records from other digital materials is an important first step. The organization of your records may be based on the different types of records or the length of time for which certain kinds of records need to be kept. These groupings can be related to each other in a hierarchical or flat way, as best suits your needs. Generally, this structure should be consistent with the organization of any paper records you have (or records in other media), so that all records related to the same activity or subject, or of the same type, can be easily identified and retrieved as part of one conceptual grouping, as needed. Your organization scheme should be recorded in a document that shows all the groupings of materials, describes them in a brief sentence and indicates how they are related. In this document, which is called a *classification scheme*[16] or filing plan, each grouping of records can be assigned a code or a name that should be linked to each record belonging in the same grouping no matter what the medium or location: thus, the records assigned to each grouping will share such code or name, followed by a number that indicates

[15] The *vital record code* only pertains to specific communities of practices, such as legal and medical offices, who must identify the records that are vital to the continuance of their business in case of disaster and who would therefore exercise special protection measures on those records.

[16] Defined as: A plan for the systematic identification and arrangement of business activities and records into categories according to logically structured conventions, methods and procedural rules. (See also Recommendation 3.)

their sequence. This identifier should be recorded among the *identity metadata*[17] of your digital records and on the face of your paper records belonging to the same grouping and should be unique for each record.

Identifying how long groupings of records need to be retained will facilitate their management while they are regularly needed and help ensure that records that need or merit long-term preservation are tagged early and given proper protection to ensure their survival.

You will find it easier and more efficient to assign a retention period—the length of time you want or need to keep materials—to a grouping of materials, rather than to individual items. Trying to ensure that some things are kept as long as needed while weeding out things that are no longer needed is simply too cumbersome at the individual item level. Although you may think that within a grouping some records should be kept longer than others, not only will you save time if you keep the whole grouping, but you will also have more complete information when you need to refer to the records. However, for some types of records, you can create subgroups within each given grouping on the basis of the retention period.

6. Use authentication techniques that foster the maintenance and preservation of digital materials.

The authenticity of digital materials is threatened whenever they are transmitted across space (i.e., when sent to an addressee or between systems or applications) or time (i.e., either when they are in storage, or when the hardware or software used to store, process or communicate them is updated or replaced). Because the acts of setting aside digital materials for future action or reference and of retrieving them inevitably entail moving them across significant technological boundaries (from display to storage subsystems and vice versa), the inference of the authenticity of digital materials must be further supported by evidence that they have been maintained using technologies and administrative procedures that either guarantee their continuing identity and integrity or at least minimize risks of change from the time the records were first set aside to the point at which they are subsequently accessed.

Technology-independent Authentication

Presumption of Authenticity. A presumption of authenticity is an inference that is drawn from known facts about the manner in which a document has been created and maintained. Adoption and consistent application of the recommendations presented in this document provide the best evidence to support such a presumption. The recommendations are cumulative: the higher the number of satisfied recommendations and the greater the degree to which an individual recommendation has been satisfied, the stronger the presumption of authenticity.

Successful implementation of the recommendations presented in this document is predicated on establishing and continuously applying effective administrative policies and procedures.[18] Ideally, you should strive to implement authentication techniques supported by administrative policies and procedures that are as technology-independent and/or neutral as possible.

[17] Defined as: The properties or attributes conveying the identity of a digital object that is to be kept as a record. (See also Recommendation 3.)

[18] See Appendix 19, "A Framework of Principles for the Development of Policies, Strategies and Standards for the Long-term Preservation of Digital Records."

Technology-dependent Authentication

Technology-dependent authentication techniques, such as cryptography, are used to provide a technological mechanism to guarantee the authenticity of digital materials. One such cryptographic technique is the digital signature, which can be used when transmitting documents between persons, systems or applications to declare their authenticity at a certain point in time. Such technologies have been given legal or regulatory value by some bodies, like the European Commission and the Securities and Exchange Commission.

Caution! Digital signatures are subject to obsolescence themselves and, by virtue of their purpose and inherent functionality, cannot be migrated to new or updated software applications together with the documents to which they are attached. In fact, the life of digital signatures and other authentication technologies may be much shorter than the length of time that even a temporary document not requiring migration may need to be maintained, because authentication technology is changing rapidly. Unless or until further development of digital signature technology enables such encrypted authentication information to be preserved over time with the document, you should, when you receive a document with an attached digital signature, detach the signature whenever possible and add information to the integrity metadata to indicate that the document had an attached digital signature when received and that the signature was verified, detached and deleted.

7. Protect digital materials from unauthorized action.

The accuracy and authenticity of digital materials cannot be presumed if there is any opportunity for modifying them without leaving a trace. You need to be able to demonstrate that it was impossible for anyone to tamper with or manipulate your digital materials without that person being identified. Security includes restricting physical access to places where computers are kept, as well as restricting access to the digital materials on the computers themselves. The latter can be accomplished through various means, including the use of passwords and/or biometric authentication to log on to the system.

It is also important to set up a structure of access permissions (also called access privileges—see discussion of *integrity metadata* in Recommendation 4) for all users of the system. For example, some users may only be able to read materials, while others may have permission to modify them. In any case, it should be impossible to modify any record once it has been filed according to the *classification scheme* or filing plan (see Recommendations 3 and 5), and only the person who has been given responsibility for recordkeeping and maintenance should be able to transfer or delete materials from the system. In addition, the system should maintain an audit trail to track access to the materials to control the administration and use of access privileges.

This recommendation may appear to be a tall order for individuals who may be working out of their homes, or even for those working in very small offices or communities of practice. But it is important to remember that if you cannot demonstrate that it was impossible for anyone to tamper with and manipulate your digital materials without being identified, your assertion that your records are de facto accurate and authentic becomes irrelevant. In this regard, it might be useful to keep copies of at least the most important digital materials offline and to establish some routine by which materials stored offline are randomly compared with their counterparts online on a periodic basis.

8. Protect digital materials from accidental loss and corruption.

Computers are not foolproof, and any of a number of factors can cause corruption or other accidental loss of records or data. The best way to ensure against accidental loss or corruption is to make backup copies regularly and often. If you store such copies off-site, additional protection is obtained against fire or theft of equipment. Many backup techniques, software packages, and services are available, including ones that automatically create the backup materials and then transmit them to a secure off-site location.

a. *Develop a rigorous policy or routine that ensures your system is backed up daily*. Your system is only as good as its last backup, so you need to make sure it is backed up often, at least once daily, using proven methods that will ensure that if something goes wrong, you and/or your business will be able to recover quickly. Such regular backups should be destroyed on a rotational basis according to a strategy or schedule that is most appropriate for your requirements, since they do not contain records but only exist for recovery of the system if it fails. Note that we are talking here about a comprehensive *system backup*, which includes the operating system, the software applications and all the digital materials in your system. If, in addition to a system backup, you need to have a security copy of your digital materials in case your computer is stolen or some of your records become corrupted, then you should backup those materials only on another computer, an external hard drive or other portable digital media and store these security copies in an off-site location away from the computer with the "original" copies.
b. *Choose and install the best backup technology for your situation*. Study the technology and services available, and choose what works best for your particular situation. Many different systems are available, ranging from those covering one-person operations to those able to back up very large systems. The backup system needs to include an audit trail, in case the system fails between backups and you need to recover the records or other digital materials created during the time for which there is no backup.

9. Take steps against hardware and software obsolescence.

The speed with which hardware and software become obsolete poses severe challenges to the maintenance and long-term preservation of digital material. One strategy to address this problem is to eliminate dependence on hardware by transferring hardware functionalities to software (i.e., use a software application to simulate the actions of a piece of hardware). This provides a more stable way to retain the function when the hardware becomes obsolete.

The rapidly changing technology environment means that both individuals and offices should regularly upgrade their digital systems as well as all the records within these systems and those that have been moved to another storage medium, such as CD, DVD or tape. In other words, when parts of the technological environment in which you are working begin to become obsolete, they should be upgraded to the most advanced technology available according to your particular requirements and constraints, and all digital materials inside and outside the system should be migrated to the new technology. When replacing hardware, it is important for the replacement hardware to have capabilities at least equal to the hardware it is replacing. For example, a new monitor needs to display a graphic record in a way that retains the documentary form of the original record. Planning for regular technology upgrades on a rotational basis will

help ensure that your technology does not become out of date and also help prevent large and unexpected technology expenses.

Sometimes digital records produced by or maintained in systems that are becoming obsolete need to be retained for a long time, but they are not expected to be accessed often. If such records are textual records and need to be read sequentially rather than randomly, you could convert them from their digital form to computer output microfilm. This will protect them from accidental loss or corruption better than any other measure. Another good protective measure is duplication—creating a second copy of groups of vital records and keeping it on another computer, on a second hard drive, on DVD, with another office or individual or in remote storage. When digital records or other entities are removed from a live system, for storage on magnetic or optical media outside the system, for example, it is essential that documentation about the system and about the digital materials (for example, the records' metadata) is also removed and kept with them. For more detailed information about the types of documentation in question here, see Recommendation 1, subsections D, E and F.

10. Consider issues surrounding long-term preservation.

Although the focus of this document has been on the creation and maintenance of all kinds of digital materials while they are needed on a regular basis by their creators, it is important to consider how best to preserve important digital materials for the long term. Typically, only a small percentage of materials need to be preserved for the long term, but the ability to provide ongoing, long-term care for materials, especially digital materials, is often beyond the capability or interest of individuals and small organizations. There are real costs—both financial and human—in retaining materials for the long term, but such preservation efforts are essential for establishing and maintaining our cultural heritage, for accountability purposes and for informing managerial decision-making.

To begin this process, you should identify someone who will take charge of your digital materials once they are no longer needed for regular personal or professional purposes. This person would take the role of *trusted custodian*.[19] A trusted custodian is a professional—or a collection of professionals, as in an archives or a community historical society—who is educated in recordkeeping and preservation, and who ideally has no stake in the content of the records and no interest in allowing others to manipulate or destroy the records.

In the case of small organizations or offices, this person could be the one responsible for keeping the records and organizing and storing them during their active use. In the case of individuals who manage their own recordkeeping, the person fulfilling the preservation function may be an archivist or a librarian in a documentation centre, or simply themselves. In either case, a preservation strategy should be established as soon as possible, because digital materials that have not been targeted for preservation early and taken care of in a proactive way will not be preserved. Close adherence to these guidelines will therefore facilitate long-term preservation.

[19] Defined as: A preserver who can demonstrate that it has no reason to alter the preserved records or allow others to alter them and is capable of implementing all of the requirements for the authentic preservation of records.

Conclusion

This document has outlined a series of activities for individuals and small organizations to carry out to create and maintain digital materials that can be presumed to be authentic, accurate and reliable. For individuals the burden may seem great, but the alternative—loss of records or the emergence of corrupt and unverifiable data—would be an even greater problem in the long run. Small organizations will benefit by making a clear designation of the individual or individuals responsible for overseeing the maintenance of the organization's digital records. Bear in mind, however, that not all recommendations presented in this document need to be implemented in each circumstance; you should be able to select and adopt the measures that address your particular problems in the specific context in which you operate. There may also be cases in which additional measures are necessary because of legislative or regulatory requirements specific to your field, or because of the characteristics of the activity and hence of the records that it produces. In such cases, consultation with experts may be required. Among such experts are the archivists of city, provincial, state or national archives, as well as local archival associations. Individuals, offices and small organizations should not hesitate to contact such experts for advice on any issues relating to the creation and maintenance of their digital materials.

Finally, this set of guidelines is but one of the documents issued by the InterPARES Project, an international research project studying the long-term preservation of authentic digital records. Additional material that will support the understanding of the nature of digital records and the development of methods for their reliable creation and accurate and authentic maintenance and preservation can be found on the InterPARES Web site at www.interpares.org.