

 Open access • Journal Article • DOI:10.1353/LIB.2007.0044

Defining Digital Sustainability — [Source link](#)

Kevin Bradley

Published on: 01 Nov 2007 - Library Trends (Johns Hopkins University Press)

Topics: Digital preservation and Sustainability

Related papers:

- [Problems in the preservation of electronic records](#)
- [The Preservation Perspective](#)
- [Linking preservation metadata and collection management policies](#)
- [Chapter 9 – Digital preservation](#)
- [Preserving digital libraries: Determining what? before deciding how?](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/defining-digital-sustainability-16iq4gclb7>

Defining Digital Sustainability

KEVIN BRADLEY

ABSTRACT

This paper investigates what is meant by digital sustainability and establishes that it encompasses a range of issues and concerns that contribute to the longevity of digital information. A significant and integral part of digital sustainability is digital preservation, which has focused on one technical concern after another as issues and fashions have shifted over the last twenty years. Digital sustainability is demonstrated as providing an appropriate context for digital preservation because it requires consideration of the overall life cycle, technical, and socio-technical issues associated with the creation and management of digital items.

INTRODUCTION

If digital technologies had a sense of humor, a joke between them might run: There are ten types of technologies in this world: those that understand binary, and those that don't. Digital storage and delivery technologies allow the encoding of meaningful representations into two states, 0 and 1; a state of being and a state of not-being, of on and off, of plus and minus, or of falling below or climbing above a defined or given threshold. If the permanent maintenance of any given state, or set of states, was the definition of digital sustainability, then we could merely select a suitable technical strategy to permanently inscribe those states and entrust the objects to an appropriate storage and preservation strategy. However, the layers of dependencies and interdependencies, standards, agreements, understandings, technologies, strategies, workflows, and business models render that simple preservation model indefensible.

Thinking about some of the protocols associated with storing and accessing digital coding may help to illustrate these intricate dependencies. A bit, the lowest level of information, is meaningful only in relation to other bits with which it is associated; eight bits form a byte, and a word length might be 16-, 32-, or 64-bit depending on the operating system and the type of data. The word may exist, but it is just a seamless string of digits unless the system knows where the word or byte starts and finishes. The data is allocated a place on a disc that is formatted in a particular manner. The Microsoft disc operating system (MSDOS) uses a file allocation table (FAT), which may be either FAT 12, FAT 16, FAT 32, or FAT 64, depending on the memory space and partition size. In a UNIX environment the file system structure is managed by a protocol called inodes. Mac computers have used inodes as a sectoring protocol since the 2001 operating system OS X was released, and their own proprietary system for OS 9 and all earlier operating systems. As well as these there are many legacy disc structures associated with operating systems no longer supported; eventually all the current systems will also become legacy. Various tables and structures define the “address” at which data may be found.

Some systems, such as compact discs, use a small range of hard-coded words to describe the original word, and a lookup table is needed to associate the coded word with the stored word. If the data is backed up on tape, as is customary, then there are a different range of data storage protocols, tape standards, and potentially complex compression algorithms. Assuming the data can be found, and the appropriate word substituted where necessary, the operating chip will need to know if the word is big-endian or little-endian. The byte stream is described as little-endian when the low-order byte of the number is stored in memory at the lowest address, and the high-order byte at the highest address; big-endian is the reverse. This is an issue for the operating chip; the chip used in PCs have tended to be little-endian, while those used in Macs tend to be big-endian. As a consequence file formats developed on one platform or another may specify byte order. For example, a bitmap (.bmp) specifies a little-endian byte order, while a JPEG expects big-endian. TIFF image files can be big- or little-endian, and encodes in its metadata the form to which it conforms. The byte order can be reversed, but the system knowledge that this is necessary is essential.

The host computer must have access to enough coded information to allow it to recognize the binary file format and associate it with the appropriate piece of software. The version of the file is generally only known after the file is opened; then rendering software attempts to open the file if it is a version that it recognizes.

If the file is character-oriented, it will be necessary to decode the character set, which may be described in 7-bit or 8-bit ASCII (American Standard Code for Information Interchange), or UTF-7 or UTF-8 (Uni-

code Transformation Format), or a number of variants. Various lookup tables describe the relationship between the code and the text it represents. Characters associated with a particular language are an issue, and the character sets might contain Chinese, Japanese, or Korean characters (CJK) or Arabic characters, the transliteration to Roman code described in the standards ISO 233 or DIN 31635. Similar standards exist for other character sets.

The browser or rendering software, if it is needed, must not only be appropriate to the version of the file, but also to the operating system on which it will operate. If the file we are trying to preserve is an executable file, it too must have the appropriate operating system on which to run. The operating system must be the proper service pack, have the correct patch and install levels, and have the appropriate device drivers. A specific example of the level of compliance required, as well as an area of constant problems, might be the dynamic link library (DLL), a file that stores data used by Windows programs and links to those programs at "runtime." Often the DLL used by a particular program is missing, corrupted, or altered by the hardware or by another program that shares it in use. This generally produces an error message and requires a reinstallation of the DLL file. The number of DLL files available is very large and the process of identifying, tracking down, and installing the proper file is described by IT support staff as "DLL hell." Changes to the kernel, which is responsible for process and task management, and memory and disk management, can render a program inoperable, as can the inability to locate low level libraries in UNIX systems. The way in which operating systems and programs interact is complex, subject to change, and mediated by commercial interests; and faults or incompatibilities in any of these areas can make the whole system seem very fragile.

Besides the software interaction, file functionality also depends on standards, agreements, and understandings in interfaces, cabling, and hardware, and still this represents just a small set of examples of the complex interdependencies and detailed interaction that goes into making a digital object renderable. Though there are few who understand in detail each and every level, most IT professionals and support staff have a more than passing understanding of what roles each part plays. As long as the system operates transparently, that passing understanding is more than adequate to manage the system. In the event that the detailed infrastructure underpinning access to even the simplest digital object no longer functions, the level of knowledge would not allow support staff to rebuild it in new technologies.

Extensible Markup Language (XML) is perceived by most in the digital archiving community as an open, transparent and extensible way of encoding and accessing digital information. XML is the favorite format of those who are concerned with the longevity of their data. However, there

are layer upon layer of invisible technologies, standards, and agreements that enable XML documents to be transparent. At some level, as most people know, even XML is just a bunch of aligned magnetic domains on polyester tape.

So why am I using up paper and pen, as Sigmund Freud once said, "in order to expound things which are, in fact, self-evident"? (Freud, as cited in Derrida, 1996, p. 8). Simply this: to make the point that thinking of digital preservation as consisting of rendering a file or bit stream permanent is a pointless and futile exercise. The new field in which digital preservation plays a part recognizes that the infrastructure supporting the functionality of digital objects must itself be sustained in order to maintain access to their content and meaning. This may be the technology used to create and access the data contemporaneously, or the means to present old data with the new technologies as they emerge. However, meaning does not reside in the technology, and data streams cannot sustain themselves. In sustaining digital information it is necessary to consider the organizational, socio-technical and economic infrastructure, as well as the purely technical and structural issues associated with digital information.

This paper defines the concept of digital sustainability as encompassing the wide range of issues and concerns that contribute to the longevity of digital information. Digital preservation, a significant and integral part of digital sustainability, is shown to have changed its focus from one technical concern to another as issues and fashions have shifted. Digital sustainability, it is demonstrated, provides the context for digital preservation by considering the overall life cycle, technical, and socio-technical issues associated with the creation and management of the digital item.

A SHORT HISTORY OF DIGITAL PRESERVATION

Digital preservation has, at the least, a lexical link to preservation, and, at best, a philosophical and conceptual base embedded in the aspirations of traditional conservators. The profession of preservation and conservation matured both technically and philosophically in response to the 1966 disaster that saw the River Arno in Florence break its banks and wreak disaster upon a store of priceless cultural heritage objects. Practitioners and thinkers in the conservation field rallied in the salvage effort, and, in the aftermath of the flood, participated in a long reevaluation of traditional practices. The modern field of preservation has evolved from this process.

Water also played a role in one of the pivotal moments of digital preservation. In the early 1980s those concerned with keeping information on flexible magnetic media or tape recognized that the binder was subject to dramatic and catastrophic failure. The process of failure was identified as hydrolysis, the chemical decomposition of the binder by the addition of water in which the water reacts with a compound to produce other com-

pounds. The other compounds produced turned the tape binder into a sticky mass, which introduced a high level of errors into the digital system and eventually made the content of the tape completely irretrievable. Technical experts rallied, the process was explained, and a treatment was developed that made the content of the tapes temporarily accessible (Bertram & Cuddihy, 1982; Brown, Lowry, & Smith, 1983, 1984, 1986; Cuddihy, 1980). Publications began to include the word *preservation* in connection with the treatment of data and data carriers, and the audiovisual archiving community began to participate in the research and debate as audio and video tapes succumbed to the same syndrome. The question remained unanswered (and to some extent remains so today): did the failure of the tapes' binder point toward the eventual fate of all polyester urethane tape binders, or was it an aberration caused by inadequate manufacturing control?

The answer to the tape binder question, however, gradually began to be of less importance to the emerging field of digital preservation. The process of attempting to solve that problem led to a new set of questions for those concerned with preserving the growing archives of digital content. A treatment that provided temporary alleviation of the symptoms of hydrolytic binder degradation, and, therefore, made the data retrievable for as long as the solution retained its efficacy, was developed, but it produced a dilemma. Clearly the data had to be copied to a new carrier, but if the only viable storage technology, tape, had a limited life expectancy, how could the data be managed?

The established manufacturers responded to their customers' growing concerns about the life of their storage media and began to develop long-term carriers. Two companies, Creo and ICI, combined to produce the terabyte optical tape; Sony produced the much vaunted century media, an optical disc. Both of the new carriers measured their life expectancy in decades, and the Sony solution utilized WORM (Write Once Read Many) technologies, which, as it was inerasable, was marketed as an added preservation measure. The customers of the major companies, looking for solutions to their digital storage needs, did not support either of these technologies, or any of the others that appeared contemporaneously with the optical developments. Instead, the market adopted what Christensen (2000) would identify as a disruptive technology, pure iron particulate on polyester tape in cartridges.

The data cartridge tapes can be described as disruptive technology, from the view of digital preservation, for a variety of reasons. First, the manufacturers had not estimated the life expectancy of their tapes in decades, if at all; rather they described usable life in terms of number of passes. The useful life of the individual tape was, in other words, limited by the number of times the data could be accessed, a figure that could be easily measured. Additionally, manufacturers published development

roadmaps predicting when the current generation would be superseded and when they would become obsolete. The actual life of the tape was, to a large extent, irrelevant. The earliest generations of these cartridge data tapes, especially those manufactured using first generation metal evaporative techniques, were notoriously unreliable. Soon technologies developed around these tapes, such as tape robots, error measurement, and storage management systems, which compensated for the individual tape's inconsistencies. Instead of depending on the reliability of the carrier, data managers invested in the reliability of the system. Permanence was no longer in the carrier, but in the ability to migrate the byte stream from the superseded carrier to a new carrier within the system, and, ultimately, to the next and all subsequent storage systems.

The question of how to build a permanent carrier was never really answered; the answer changed the question, and the concerns and questions that occupied the minds of digital preservationists shifted to new issues. Though pockets of research still continue, and occasional news releases herald the latest everlasting media, the digital preservation community has, by and large, abandoned any interest in such enduring storage solutions. The goal of a permanent media has been wrecked on the rocks of relentless progress. Even if any media could be claimed and trusted, as permanent, the quandary is that within a short period of time the storage system would be technically superseded by storage media exhibiting superior performance specifications, and manufacturers would no longer support the old technology. Eventually there would be no functioning replay equipment to access the supposedly permanent media, and, even if there was, the technically slow performance of the old technology would make the transfer to a newer and faster storage media attractive. However, with the realization that carriers changed in response to the market came the recognition that the same was happening to file formats and access software—a threat that caught the attention of the second wave of digital preservationists.

Migration of the data from carrier to carrier was the solution to the problem of carrier failure, and a similar scenario was envisaged for the problem of file format obsolescence. The future of digital information would be linked to its past by a series of actions that would result in the current, transformed version of an item being accessible using current access technologies. The risks associated with cumulative migrations concerned many thinkers, and emulation was promulgated as an alternative, most notably by Jeff Rothenberg (1998). The risk posed by migration was seen not just as corruption of the data, but alteration of the “look and feel,” or a loss of “significant properties.” The value of emulation, Rothenberg argued, was that of always operating on the original byte stream; that is, the intentions of the document's creator would be better preserved by leaving the byte stream unaltered and introducing software instead to

make the old formats accessible on new technology. The relative value of the two approaches seemed to dominate digital preservation discussion from the mid-1990s, and the measure of potential success was quantified in terms of their ability to maintain the “look and feel” or preserve the “significant properties.” Permanence was shifting, in this debate, from concern with the bits to concern for the content.

The issues of migration and emulation no longer dominate the agendas of meetings and conferences. Most people involved in making decisions about digital collections are comfortable with the notion that it will be necessary to take one approach or the other, and they are content to make that decision when the time comes. They also recognize that that decision will be made more than once, and can be remade as required, at least for the first generation of changes. Major or quantum changes, of course, may well demand a definitive and final decision, but the current range of incremental technological changes means that the time of the disruptive and irreversible change is not yet here.

Discussion of “look and feel” and “significant properties” has similarly waned, not because these are not important or do not exist, but because there has yet to be found a way to automate and make this information machine-readable. “All God’s children got significant properties,” we can sing in unison, but this takes us no further if we cannot define its meaning in such a way that we understand what properties are under consideration, and describe them in a way that is machine-readable and automatically actionable. Defining significant properties runs up against the philosophical issue associated with any epistemology—knowing how we know these things in an objective way. It does not take long to reach a point where significant properties are those properties capable of being described as significant, and an object’s being is its significance. The pragmatism of technologists unable to resolve a philosophical dilemma leads to either broad and necessarily imprecise decisions about classes of materials, or to a position that aspires to preserve all of the properties that might exist in that digital object, and a recognition that a decision about what will be lost in a class of materials will have to be made at the time that an object-changing preservation action has to be taken. The pragmatic digital preservation community has moved on to the next wave of concerns.

The underlying and implicit conclusion of the discussion of the previous two digital preservation paradigms is that permanence in access is the critical measure. Using a term like access requires some explanation. It is not only about the ability to find and retrieve an item, but also the ability to use, view, listen to, interact with, display, or run the digital item in such a way that users can be assured that what they are viewing satisfies their needs. This may, for example, be a requirement to see exactly what the creator originally intended, the identical look and feel, or it may be the

ability to find and interrogate the same data, or simply to be able to read the same text.

At the same time as access to content was developing into the main debate, and migration and emulation were the topics under discussion, the concept of a Universal Virtual Computer (UVC), developed by IBM staff members Raymond Lorie (2002) and Henry Gladney (2003, 2004), began to be discussed. Rather than develop permanent carriers, the proponents of this concept argue for a simple and long-term approach that can be recompiled at a future date, and enable the extraction of the data so that behaviors that can be modelled on technologies that are in use at the time of access. The UVC concept seemed to cut through the issues associated with the ability to render the content on future systems and platforms. Apart from trial projects, such as the Koninklijke Bibliotheek's image archiving project (Van Wijngaarden & Oltmans, 2003), the UVC has not been widely embraced by the digital archiving community.

The UVC addresses the issue of future format obsolescence by isolating the digital object from mainstream systems in a form that expects to allow rendering at a future time in systems that are not compatible with present-day technologies. Similar in intent, though different in the level of implementation, are systems that seek to encode the archival digital object or encase it in a wrapper, generally a form of XML, that is so open and transparent that future actions to render the digital object in a new operating environment will present few problems. The National Archives of Australia's XENA (XML Electronic Normalising of Archives) is an example.

These systems solutions are underpinned by an episodic model of digital preservation, where the pressures of impending obsolescence force a quantum change in form at some given time. The unchanged data is retrieved from the permanent store and recompiled or re-rendered in a new environment. After making the leap to the new state, or format, the new form is stored in either a compiled or normalized XML-like form, quiescent until the need for another change occurs. Though I identify a philosophical similarity between the two approaches, they are not identical. The XML normalization approach is not dissociated from current technologies to quite the extent that the UVC is, and so the interval between necessary migrations is shorter. Also, XML normalization involves many upfront decisions about significant properties, or performance, that are left until later in the UVC.

In the repositories and digital archives, preservation is increasingly being defined as sustainable access. The Australian Partnership for Sustainable repositories (APSR) "has an overall focus on the critical issues of the access continuity and the sustainability of digital collections" (McGauran, 2003). The emphasis on access as a measure of preservation has led to a

natural alliance with those concerned with content delivery, and a growing awareness among repository managers and digital library personnel of the need to expose their data to a growing range of sophisticated users, with the ability to “feed back” to the host archive. The current digital preservation paradigm thinks of digital objects as parts of a complex relationship, continually changing their content as well as their form, constantly being required to interact in new ways in intricately constructed systems. The label-hungry marketing environment might say the new concern is Web 2.0, as compared to Web 1.0, but such labels themselves suppose a hard distinction that is not so easily drawn. Nonetheless, the concerns of the current environment are in the area of system architectures, standards, metadata, and tools. It is a sophisticated field requiring solutions to the challenges raised by long-term access to digital information that are more integrated than the earlier, more one-dimensional approaches. Rather than looking for quantum-level solutions, the current problems are addressed poco a poco, little by little, buttressing the existing approaches with solutions that address interoperability and access in ongoing systems. The critical tool in this process, one that is the center of today’s digital preservation debate, is the digital repository, which ideally holds the materials, provides access, tracks the changes, and maintains the authenticity of the item to the extent that is necessary in each individual case.

There is also recognition that the ability to preserve and provide access to digital information is linked to more than technical issues, and that economic, social, and other such factors will play a part in determining the useful life of any information encoded in digital form.

DIGITAL SUSTAINABILITY AND OTHER COMPETING LABELS

“Sustainable” and “sustainability” only recently have taken on the meaning that now seems so familiar to us, and have become a new part of the lexicon. In the earliest recorded usage of these words they meant something different. One might sustain a belief, or sustain an argument, but it was not until the 1960s that “sustainability” began to take on an economic as well as a temporal sense. By the early 1980s, “sustainable” had begun to be associated with concerns regarding the environment.¹ Since then the word has continue to expand in its usage; the Victorian Government has a Department of Sustainability and the Environment (<http://www.dse.vic.gov.au/dse/>), which links sustainability to the issues of environmental impact, but the Australian Capital Territory Government has an Office of Sustainability that is “committed to creating a sustainable Canberra” and “developing, facilitating and coordinating the implementation of guidelines, policies and procedures related to sustainability” (ACT Office of Sustainability, n.d.). It is difficult to read a paper, view a blog, or listen to a news broadcast without finding a new use of the word *sustainable* or a new context for its application. The popular meaning is derived from

the movement among the environmental groups to represent what were previously non-negotiable ideological positions as potential actions with economic costs. This change has opened negotiation between environmentalists and those who might exploit the environment by making the long-term costs of any particular case of environmentally degrading action a part of the present debate. Every describable aspect of the economic and socioeconomic consequences of a decision is included in the debate, including the economic value of the environment.

In the sustainable digital environment, the same inclusive debate is occurring, and here the word is used to mean building an economically viable infrastructure, both social and technical, for maintaining valuable data without significant loss or degradation. This includes the whole socio-technical composition of the repository, the short- and long-term value of the material, the costs of undertaking an action, and the recognition that technologies do not sustain digital objects: institutions do, using the available technology. Clearly it is not possible to preserve digital information without a sustainable organizational, economic, social, structural, and technical infrastructure, nor is it sensible to preserve material without sustained value.

Access to digital materials is maintained daily by data experts as they manage and modify content and react to the changing technical environment. The approach is neither sustainable nor in keeping with preservation requirements if it is not managed with the long-term accuracy and authenticity of the digital item in mind. Digital repository software, though in the early stages of development, must be able to manage and maintain records of change, original formats, and relationship and version information to describe the processes that led to the current form. A static copy will not satisfy the ever-moving present, and a changed copy without adequate documentation will not satisfy those concerned with authenticity. Clearly repositories, incorporating the sort of functionality and exchange standards necessary for long-term reliable use, are at the center of sustainable development.

However, the software that makes a digital repository is subject to the same changes and technical limitations as the data it manages. Practical repositories are products of the technology of the day and are consequently as much at risk as the content they manage. The DSpace Federation makes this explicit when it states that "it is an overt expectation that information assets managed by the DSpace system will outlive the current system, the current implementation of components within the architecture, as well as external implemented services that access and/or add value to the corpus" (Bass et al., 2002, p. 1).

A sustainable approach to repository design is one that considers, at its outset and through design and execution, future digital repository implementations that may not support or be supported by current standards

and technologies. It is clear that no repository will provide a complete solution to the problems of sustainability, but neither is it possible to envisage a workable solution that does not incorporate a viable, well-designed, digital repository.

Like the environmental movement, the sustainable digital community is defining its approaches in terms of economic factors. A maxim of digital preservation is that access to meaningful digital information will not be achieved through benign neglect, a strategy that has worked in physical collections for many decades. This is both self-obvious, as maxims should be, but also contestable. It may be that data, in the form of an ordered stream of bytes, will survive with minimal backup strategies for future users to decode. This, however, does not provide access to content, merely to bytes. The cost of providing meaningful access to the content through the use of digital archaeology skills and data experts who labor in the future to retrieve the meaning will almost certainly be cripplingly high, and not necessarily successful (Gladney, 2004). Little is lost forever, goes this argument, unless retrieving it is unaffordable. Archiving data as a basic byte stream and allowing the future to make the decision about whether to fund access is the logical extreme of the economic argument. It is not, however, sustainable by any of the definitions considered here.

The alternative to leaving access problems for the future to solve is to undertake a range of preservation activities in the present, which will facilitate access in the future. These activities might include developing preservation metadata schemas, normalizing encoding, creating multiple versions and copies, or migrating strategies or systems to enable future emulation. The pre-emptive strategies are probably much more cost-effective per digital item when compared to the projected cost of digital archaeology, but quite expensive when spread across the vast collections of potentially useful data.

A sustainable approach must navigate through the economic environment, determining whether it is more cost-effective to undertake a certain action in the future, or whether the present is the most economically propitious time to undertake some preventative task. Digital preservation, if it is to be sustainable, is an economic issue, one that advocates investment in the present to ensure access in the future. As some have noted, with the advent of digital authoring and distribution technologies, our developing capability to manage and sustain such information is being outstripped by our ability to produce it. Some have posited that, along with the necessary technological infrastructure for sustainability "must come the development of the associated *economic infrastructure*" (Lavoie, 2004, p. 46). Wise decisions will maximize economic resources and thus make access more sustainable.

Understood in these terms, digital preservation is as much an economic issue as a technical one. The requirements of ongoing sustainability de-

mand a source of reliable funding, necessary to ensure that the constant, albeit potentially low-level support for the sustainability of the digital content—and its supporting repositories, technologies, and systems—can be maintained for as long as necessary. It is not too strong to say the biggest single risk to sustained access to digital information is economic.

Because a sustainable approach is underpinned by continuing access, there is a need to ensure that economic decisions do not reduce the possibility of such access. A sustainable approach must also take account of other risks that any action or inaction might instigate through informed, though necessarily subjective, judgment. A sustainable approach must have accurate and informed risk identification and assessment, drawing on highly skilled or informed experts in the area. This is critically necessary in determining the risk to sustainability of digital objects, as the most likely failure mechanisms are not well understood, other than those caused by a cessation of funding.

Economic considerations are not limited to the cost of an action. The value of the content is another factor to weigh. A sustainable approach would be to ensure that the material acquired is of high significance to future researchers. It is very unlikely that a collection of low research significance will survive in the long term, as resources will always be allocated to high value materials first. However, associated with the choice of the most important collection items comes the risk of not selecting the content that will be most significant in the future. What we think is valuable now, may not be so in the future. Selection is a sustainability issue. Anthony Seeger made explicit this incongruity at a 2004 ethnomusicological research conference when he asked:

What is more valuable in the long run, researchers' theories or the by-products of research, like recordings and other collections? How many important theoretical articles published between 1900 and 1920 influence your current work? Wax cylinders recorded during that period are extremely valuable to both their original communities and contemporary researchers. Ironically, the by-products of our research may be more significant than our soon dated theoretical insights. (Seeger, 2004)

A sustainable approach must recognize and plan for future users as well as the exigencies of current demands.

Not all data will, or should, be sustained in perpetuity. Though costs are a function of many variables, not least the range of archival services, the archival period of retention is a significant factor (Lavoie, 2003). Planned retention of digital materials for the appropriate period is part of a sustainable approach. Certain datasets or learning objects may have intellectual, teaching, or research value for only a short period of time, possibly shorter than the life of the target sustainable repository. If sustainability is the primary aim of the repository, it may be valid to exclude

such materials, or to provide a limited type of service. Other materials may be considered valuable for a medium period of time, in which case the time between ingest and access may not be so great as to have incurred the problems caused by format obsolescence and impaired access. It may be possible to attach a reviewable lifetime rating to identified digital objects, and so reduce estimates of costs on objects so designated. The decision to delete after a given period can be reviewed, or the material can be assessed and deselected. It is worth considering, though, that the cost of expert review may well exceed the cost benefits of deselection and disposal, and would be, in these circumstances, an unsustainable strategy for managing the collection.

The sustainable repository must consider the barriers to participation and use. As economics is largely a matter of incentives and inhibitors, the use of such strategies can be applied to encourage users and depositors to participate in use of a digital repository. The economic incentives might be, for example, designing interfaces that facilitate deposit or access. Determining whether sufficient benefits were gained by participating in the use of the repository, and applying appropriate incentives where the benefits were insufficient, is a part of the same strategy. A collection of digital information is not sustainable if it has too few contributors or insufficient users to justify its existence.

The complex technical infrastructure that supports digital sustainability, the dependency on continued funding, and the likelihood that digital data will not survive extended periods of neglect means that digital repositories need stable technical support as well as resources. It follows that digital repositories are dependent on the ongoing existence of the sponsoring organization. This has been clearly recognized in an audit checklist for the certification of trusted digital repositories (RLG-NARA Task Force on Digital Repository Certification, 2005), and its prequel, where "organization" is a significant category. The sustainability of the organization, its funding and its business plan are critical to certifying a digital repository as trusted, at least equally with its technical infrastructure.

It is not only the business models and economic structures of organizations that are critical to the sustainability of digitally encoded content. An appropriate persistent identifier scheme and the ability to manage a resolver service that continues to locate digital objects intended for long-term use are also dependent on the sustainability of the institution or organization. A sustainable repository must be able to locate an item, but must also be able to resolve historic references to that item by those who use and cite it.

If a repository is at risk because of the vulnerability of the organizational structure that supports it, then the structure of the repository, the interoperability of the metadata and data formats, and the ability to seamlessly migrate to alternative repositories is an integral part of any plans

to manage sustainability. The metadata schemas, standards, and architectures must themselves be sustainable, and open and well described, so that their purpose and essence can be mapped and transformed to support the new systems that will emerge.

What distinguishes the contemporary sustainability approach from earlier aspirations to a “permanent” solution is the concentration on systems architectures and schemas that will aid in future management of digital information, rather than on the solution itself. The work on preservation metadata, the open archival information system (OAIS) producer model (Consultative Committee for Space Data Systems, 2002), the architecture and design of digital repositories all point to approaches that are designed to facilitate long-term access to digital information by enabling and informing future users so that they can maintain access to the digital content we are storing today.

There are two other terminologies that in part support the sustainability approach: curation and stewardship. Curation, as defined by the Digital Curation Centre (DCC), emphasizes the mutable and changeable nature of digital information by focusing on “maintaining and adding value to a trusted body of digital information for current and future use” (Giaretta, 2006, section 1.1). A sustainable approach must recognize the need to maintain access to content that may, for much of its life, be changing, and that change itself is a necessary part of that maintenance process. Clifford Lynch’s categories of stewardship, caring for information and cultural heritage, honoring our relationship to history, and preserving cultural heritage for the benefit of future generations, lays weight upon the scholarly traditions of selecting content for the future. The decisions lie in the present, though they may not be understood or realized until some future time.

The relationship between the approaches (sustainability, stewardship, and curation) may be best understood graphically, in a Venn diagram, showing the overlapping nature of the approaches (See Figure 1). All of these approaches recognize to some extent that the technical systems to preserve the information are necessary in order for there to be content to sustain. The common ground among the three can be described as preservation. Stewardship and curation share many aims, but the concepts embodied in sustainability overlap substantially with both.

A sustainable approach recognizes the need for society to support sustainable access to digital information, an economically expressed need, coupled with the resources necessary to undertake it, and an organized and structured community to need and support it. The technical systems and infrastructure must themselves be open and sustainable, but critically there must be a recognition that the processes put in place today are not the permanent solutions to digital access, but merely the tools that future users of the digital information will need to facilitate access to the content

Digital Sustainability, Preservation, Curation and Stewardship

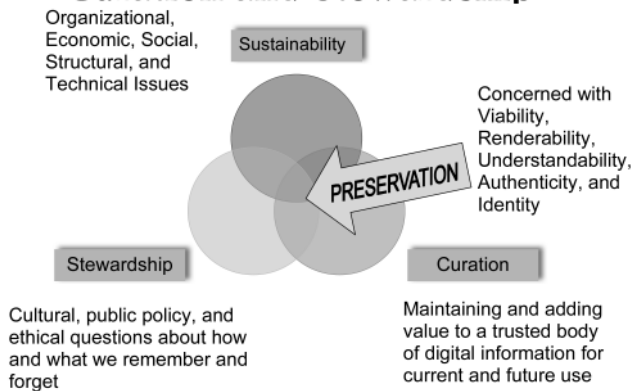


Figure 1.

encoded in those files, and to help them make a decision about its worth. Digital sustainability recognizes that the responsibility for access is shared by those in the present and the users of a future time, a time that may be as close as tomorrow, or in the dimly perceived future, and for as long as a society and a socio-technical system still exists and wishes to care for and sustain the information stored.

NOTE

1. The extensions of the meaning of the word "sustainable" are traced in *Oxford English Dictionary* (Simpson & Weiner, 2006). A source from the 1960s supports its usage in the economic sense:
 3. Capable of being maintained at a certain rate or level. 1965 *McGraw-Hill Dict. Mod. Econ.* 501 *Sustainable growth*, a rise in per-capita real income or per capita real gross national product that is capable of continuing for a long time. A condition of sustainable economic growth means that economic stagnation will not set in. Sources from the 1980s are used to support a 2002 draft addition to this definition "Ecol. Of, relating to, or designating forms of human economic activity and culture that do not lead to environmental degradation, esp. avoiding the long-term depletion of natural resources."

REFERENCES

- Australian Capital Territory Office of Sustainability. (n.d.) Home page. Retrieved February 28, 2007, from <http://www.sustainability.act.gov.au>.
- Bass, M., Branschovsky, M., Breton, P., Carmichael, P., Cattery, B., Chudnov, D., et al. (2002). *DSpace internal reference specification: Technology & architecture*. Retrieved December 15, 2006, from <http://www.dspace.org/technology/architecture.pdf>.
- Bertram, H. N., & Cuddihy, E. F. (1982). Kinetics of the humid aging of magnetic recording tape. *IEEE Transactions on Magnetics*, 18, 993-999.
- Brown, D. W., Lowry, R. E., & Smith, L. E. (1983). *Predictions of the long term stability of polyester-based recording media* (NBSIR 83-2750). Washington, DC: National Bureau of Standards.
- Brown, D. W., Lowry, R. E., & Smith, L. E. (1984). *Predictions of the long term stability of polyester-based recording media* (NBSIR 84-2788). Washington, DC: National Bureau of Standards.

- Brown, D. W., Lowry, R. E., & Smith, L. E. (1986). *Predictions of the long term stability of polyester-based recording media* (NBSIR 86-23474). Washington, DC: National Bureau of Standards.
- Christensen, C. M. (2000). *The innovator's dilemma*. New York: HarperCollins.
- Consultative Committee for Space Data Systems. (2002). *Reference model for an open archival information system (OAIS)*. Retrieved December 15, 2006, from <http://public.ccsds.org/publications/archive/650x0b1.pdf>.
- Cuddihy, E. F. (1980). Aging of magnetic recording tape. *IEEE Transactions on Magnetics*, 16, 558–568.
- Derrida, J. (1996). *Archive fever: A Freudian impression* (E. Prenowitz, Trans.). Chicago: University of Chicago Press.
- Giaretta, D. (2006). *DCC approach to digital curation—under development*. Retrieved February 28, 2007, at <http://dev.dcc.ac.uk/twiki/bin/view/Main/DCCApproachToCuration>.
- Gladney, H. M. (2003). *Trustworthy 100 year digital object: Evidence even after every witness is dead* (Preliminary version). Saratoga, CA: HMG Consulting. Retrieved February 28, 2007, from <http://eprints.erpanet.org/8/>.
- Gladney, H. M. (2004). *Principles for digital preservation*. Saratoga, CA: HMG Consulting. Retrieved February 28, 2007, from <http://eprints.erpanet.org/70/>.
- Lavoie, B. F. (2003). *The incentives to preserve digital materials: Roles, scenarios, and economic decision-making*. Dublin, OH: OCLC. Retrieved December 15, 2006, from <http://www.oclc.org/research/projects/digipres/incentives-dp.pdf>.
- Lavoie, B. F. (2004). Of mice and memory: Economically sustainable preservation for the twenty-first century. In *Access in the future tense* (CLIR Report 126) (pp. 45–54). Washington, DC: Council on Library and Information Resources.
- Lorie, R. (2002). *The UVC: A method for preserving digital documents: Proof of concept* (IBM/KB Long-term Preservation Study Research Series no. 4). Den Haag: IBM; Koninklijke Bibliotheek. Retrieved December 14, 2006 from http://www.kb.nl/hrd/dd/dd_onderzoek/reports/4-uvc.pdf.
- McGauran, P. (2003, October 22). *\$12 million for managing university information* (Media release). Retrieved December 14, 2006, from <http://www.dest.gov.au/Ministers/Media/McGauran/2003/10/mcg002221003.asp>.
- RLG-NARA Task Force on Digital Repository Certification. (2005). *Audit checklist for the certification of trusted digital repositories*. Mountain View, CA: RLG.
- Rothenberg, J. (1998). *Avoiding technological quicksand: Finding a viable technical foundation for digital preservation*. Washington, DC: Council on Library and Information Resources. Retrieved December 14, 2006 from <http://www.clir.org/pubs/reports/rothenberg/contents.html>
- Seeger, A. (2004, July). Handout notes given at the Symposium of the International Musicological Society, Melbourne, Australia.
- Simpson, J. A., & Weiner, S. C. (Ed.). (1989). *The Oxford English dictionary online* (2nd ed., with additions). Retrieved December 18, 2006, from <http://dictionary.oed.com>.
- Van Wijngaarden, H., & Oltmans, E. (2003). *Digital preservation and permanent access: The UVC for images*. Retrieved December 14, 2006, from http://www.kb.nl/hrd/dd/dd_links_en_publicaties/publicaties/uvc-ist.pdf.

Kevin Bradley is a world expert on digital and sound preservation. He contributed to and was editor of the *Guidelines on the Production and Preservation of Digital Audio Objects*, published in 2004 by the International Association of Sound and Audio Visual Archive, and is the author of *Risks Associated with the Use of Recordable CDs and DVDs as Reliable Storage Media in Archival Collections—Strategies and Alternatives* (Paris, UNESCO, 2006). He was a member of the OCLC/RLG Preservation Metadata Framework Working Group that developed the OCLC/RLG Metadata Framework to Support the Preservation of Digital Objects, and is currently a member of the Sub-Committee on Technology of UNESCO's Memory of the World Programme. Kevin was previously employed by the Australian Partnership for Sustainable Repositories. His current position is curator of oral history and folklore and director, sound preservation, at the National Library of Australia.