Margaret Hedstrom

Dr Margaret L. Hedstrom is an Associate Professor at the School of Information, University of Michigan. Before joining the faculty at Michigan in 1995, she was Chief of State Records Advisory Services and Director of the Centre for Electronic Records at the New York State Archives and Records Administration (1985–95). Dr Hedstrom earned Master's degrees in library science (1977) and history (1979) and a PhD in history (1988) from the University of Wisconsin-Madison. Research: Dr Hedstrom has been conducting research on the management and preservation of electronic records for nearly 20 years. She is Project Director and Co-Principal Investigator of the Camileon project, a joint effort by the University of Michigan and the University of Leeds to investigate the feasibility of emulation as a digital preservation strategy. She is also director of a study of electronic record keeping in collaborative work environments, funded by the National Historical Publications and Records Commission. Her current research interests include strategies for long-term preservation of digital information, the impact of electronic communications on organisational memory and documentation, and remote access to archival materials. Teaching: Dr Hedstrom teaches graduate courses at the School of Information on access systems for archival materials, social systems

Significant properties of digital objects: definitions, applications, implications [1]

Margaret Hedstrom, Christopher A. Lee

Introduction

Significant properties are those properties of digital objects that affect their quality, usability, rendering, and behaviour. In an ideal world, free from technical and economic constraints, libraries and archives would preserve their physical and digital collections in their original form with all significant properties intact. Resource constraints and technical limitations, however, often require libraries, archivists, and curators to preserve objects as facsimiles or surrogates in lieu of original materials and to weigh decisions about which properties to preserve against institutional priorities, available resources, limitations of available preservation methods and technologies, and anticipated use. Digital preservation usually entails converting or migrating digital information to current computing platforms, normalising objects and collections so that they are less dependent on specific hardware and software, developing tools to emulate obsolete platforms, or some combination of these approaches in order to cope with rapid obsolescence of computing technology. Any of these approaches may alter the structure, appearance, and behaviour of the original objects. The purpose of our research is to identify the significant properties of digital objects that affect their quality, functionality, and look-and-feel so that custodians can select appropriate methods which preserve those significant properties of digital objects that are deemed important by designated user communities. A formal articulation of significant properties also provides a framework for documenting the impact of transformations of digital objects that may be necessary to ensure continuing access to them and to explain the reasons for such actions.

The importance of retaining the significant properties of documents and artefacts has been raised in numerous contexts. Librarians and archivists created microfilm surrogates of millions of books, office documents, and newspapers to overcome physical limitations of highly acidic paper and to conserve storage space, even though the microfilm surrogates were not capable of preserving all of the significant properties of the original hard copy materials, such as accurate colour representation or the exact physical dimensions of the originals. Similar problems may occur when digital objects are converted from one file format to another or when they are normalised into a standard format upon ingest into a repository. Format conversions, for example, may change document structures, stylistic features, navigation, and appearance. Decisions about significant properties of digital objects are much more complex than decisions regarding physical items because properties are expressed at several different levels of abstraction and because there are many more options available for creating surrogates, facsimiles, and derivatives. Digital documents have many features and exhibit behaviors that may or may not be important to preserve.

Guidelines for selecting preservation formats for physical materials provide a starting point for identifying potentially significant attributes of digital objects. Previous discussions of significant properties have invoked needs such as authenticity, legal admissibility, artefactual value, ease or convenience of use, and aesthetic quality as reasons for preserving physical artefacts in their original physical form. According to a recent report on the role of the physical artefact in library collections, several cardinal principles guide decisions about preserving significant properties of physical artefacts, such as age, evidential value, aesthetic value, scarcity, associational value, market value, and exhibition value [2]. Courts, for example, might challenge the admissibility of copies of original documents as evidence in legal proceedings if the reasons for and processes used to produce surrogates were not well documented. Scholars studying the evolution of book publishing glean much information from physical evidence of paper, typographical methods, and binding techniques. Concerns about significant properties are not limited to written or printed works. Photograph and film curators have to decide when it is necessary to preserve original negatives and which medium and format best captures significant features such as resolution and tonal qualities [3]. Curators of recorded sound collections are concerned about fidelity, sampling rate, volume, and dynamic range when sound recordings are transferred to new analogue media or digitised.

(1) The research for this project is part of the Camileon project, an international digital library initiative, at the University of Michigan and the University of Leeds. This work is funded by the National Science Foundation, award number 9905935 and by the Joint Information Systems Committee (JISC) in the UK. Additional information about the Camileon project is available at: www.si.umich.edu/CAMILEON/

(2) *The evidence in hand: Report of the task force on the artefact in library collections*, Washington, D.C., Council on Library and Information Resources (November 2001), p. 9.

(3) Schwartz, J., 'We make our tools and our tools make us': lessons from 'Photographs for the practice, politics, and poetics of diplomatics', *Archivaria* 40 (1995), pp. 40–74.

Currently, we do not have a formal method for identifying which properties of digital objects are considered significant or for describing significant properties. We build on the concept of significant properties from the Cedars project that treats decisions about which significant properties to preserve as collection-management decisions ([4]). Our goal is to provide an empirical foundation for making collection-management decisions where choices of preservation strategies may eliminate or alter some of the properties of original objects. Our model is modular and extensible. We treat common data types, such as numeric data, text, and images, as separate modules in the model so that collection managers can direct their attention to those types of data that are present in particularly complex digital objects or collections. The model is extensible in that particular communities may develop much more fully articulated definitions of significant properties that are relevant to the types of materials they preserve or to the needs of particular designated communities. Our work is in a formative stage, not only because of the technical complexity of digital objects, but also because we lack the accumulated social experience to guide decisions about which properties matter and to whom they matter.

Methodology

To develop an inventory and formal expression of potentially significant properties, we first explored the literature from information and library science, archival science, human-computer interaction, and various forms of criticism to evaluate the various properties of preserved information that scholars and other users have considered significant in the past. We found the literature on legal admissibility, authenticity of documents and artefacts, and intrinsic value, especially useful developing an inventory of potentially significant properties ([5]). Several studies have investigated the role of specific characteristics of documents, such as colour of text, layout, annotations, and even the appearance of book jackets ([6]). From the literature on human computer-interaction we drew on the concept of 'affordances' which refers to the functions that various types of digital objects support ([7]). Features such as browsing, viewing, annotating, and visualising digital content are examples of affordances. When encountering a preserved digital object, users may find the original affordances of the object important for accessing and making sense of the data it contains as well as for understanding the functional context in which it was used. We also found guidelines for converting documents and analogue recordings to digital form useful because they specify attributes such as resolution, colour matching, and fidelity that may represent transcendent properties that are significant for print, analogue, and digital information ([8]).

We also explored several common technical guidelines and standards for defining the properties or attributes of digital objects. We considered the Internet protocol MIME (multipurpose Internet mail extensions) which provides for the use of file headers and file name extensions to specify common encoding formats. We also extracted some properties from format-specific specifications, such as the Moving Picture Experts Group (MPEG) standards for digital video and the draft NISO data dictionary of technical metadata for digital still images ([9]). The work within the World Wide Web Consortium on style sheets — cascading style sheets (CSS) and extensible stylesheet language (XSL) — has delineated hundreds of properties that can be used to format and render structured and semi-structured digital objects ([10]). We also reviewed several metadata standards and specifications developed for the management and preservation of digital collections, such the tag library from the Text Encoding Initiative, the Cedars digital preservation metadata model, and the recent work of the OCLC/RLG Working Group on Preservation Metadata ([11]). Our review of numerous technical specifications and descriptive guidelines yielded an inventory of more than 800 specific property references.

A model for expressing significant properties

We developed a conceptual model of complex digital objects and their components which identifies common data types that can be assembled to create digital objects such as multi-level, multimedia web sites. Drawing on the definitions from the OAIS reference model, we defined a digital object as an object composed of a set of bit sequences ([12]). A complex digital object is a digital object composed of more than one type of component. We then mapped significant properties to common types of components, including numeric data, text, images, graphics, audio, video, and executables. All complex digital objects have internal compositional properties that manage how the components are assembled to compose a complex object. Complex digital objects may also have explicit external relationships to other digital objects, such hyperlinks that

([4]) Cedars project, www.leeds.ac.uk/cedars/

([5]) Menne-Haritz, A. and Brübach, N., 'The intrinsic value of archive and library material', Marburg: Archivschule, 1997; *Authenticity in a digital environment*, Washington D.C.: Council on Library and Information Resources, May 2000, National Archives and Records Service, *Intrinsic value in archival material*, Staff Information Paper 21, Washington D.C., NARS, 1982; Bearman, D., and Trant, J., Authenticity of digital resources: Towards a statement of requirements in the research process, *D-Lib Magazine* (June 1998). Available http://www.dlib.org/dlib/june98/06bearman.html; and MacNeil, H., *Trusting records: Legal, historical, and diplomatic perspectives,* Dordrecht: Kluwer Academic Publishers, 2000.

([6]) Baker, N., *Doublefold: libraries and the assault on paper*, New York: Random House, 2001; Marshall, C.C., Annotation: from paper books to the digital library, *Second ACM international conference on digital libraries*, Philadelphia: ACM Press (1997), pp. 131–140, O'Toole, J. M., 'On the idea of uniqueness', *American Archivist* 57 (1994), pp. 632–658; Pettersson, E., 'Automatic information processes in document reading: A study of information handling in two intensive care units', in *European conference on computer-supported cooperative work*, London (1989), pp. 63–72; Cadiz, J. J., Gupta, A., and Grudin, J., 'Using web annotations for asynchronous collaboration around documents', *Computer-supported cooperative work*, Philadelphia: ACM Press (2000), pp. 309–318; Toms, E. G. and Campbell, D. G., 'Genre as interface metaphor: exploiting form and function in digital environments', *32nd Hawaii International Conference on System Sciences*, IEEE Computer Society,1999; O'Connor, B. C., and O'Connor, M. K., 'Book jacket as access mechanism: An attribute-rich source for functional access to academic books. *First Monday*, 3:9 (September 1998).

([7]) Norman, D. A., *The design of everyday things*, New York, Doubleday, 1990, Phelps, T. A., and Wilensky, R., 'Multivalent documents: Inducing structure and behaviours in online digital documents', *Proceedings of Hawaii International Conference on System Sciences '96,* IEEE, 1996. Levy, D. M., 'Fixed or fluid? Document stability and new media', *European Conference on Hypertext Technology (ECHT '94)*, Edinburgh: ACM (1994), pp. 24–31.

provide a means for navigating from one object to another. We have not attempted to define all possible component types or all possible methods of internal composition. Rather, we have developed a decomposable model with a tree structure that defines common properties for each component type. In developing the model, we have discovered that certain properties apply to all types of components, some properties are unique to particular component types, and some properties are applicable to multiple components but not necessarily to all types.

A number of properties that are common across all component types need to be considered when selecting preservation strategies so that the technical approach is documented sufficiently to support access, retrieval, and low-level data management.

These include the definition of the basic data unit, byte-level encoding, data typing, and the logical structure or data model. Although these attributes often are specified or implied for specific file formats we have attempted to separate these properties from both the ways that data are stored on physical media and from file formats. These basic elements are similar to the concept of an underlying abstract form (UAF) which Holdsworth and Sergeant use to 'encapsulate the recognition that the data has an existence and a content separate from the medium upon which it is written. This underlying abstract form contains all the significant properties of the data and is independent of the medium upon which the data is written.' ([13]) Although low-level properties, such basic data unit, byte-level encoding, data type, and logical schema are germane to all types of components, the way these elements are expressed varies considerably from one component type to the next.

The underlying abstract form contains all of the significant properties of data necessary to access a digital document where access means 'realising the UAF on the technology appropriate to the time of access in such a way that the desired form of access (which may not necessarily be viewing) can be achieved.' ([14]) While the UAF is a useful starting point for identifying preservation formats which, at the highest level of abstraction, do not discard significant information, our concept of significant attributes is broader in that it takes into account the various types of stylistic and aesthetic features discussed in the literature above. For example, features such as font, spacing, and layout may be important to preserve textual components, especially if variations in these features alter the intended meaning of the text. Our formal expression of significant properties is recursive because many of the properties can be decomposed further into sub-properties. The property 'font,' for example, can be further decomposed into sub-properties such as font type, style, family, size, or colour. If changes in any of these sub-properties alter the appearance or the meaning of a digital object, they might be considered significant and worth replicating as they were specified in the original object. Some sub-properties are relevant to several different component types. Representations of colour, for example, might apply to images, graphics, video, and text, but not to audio or numeric data. We express colour properties, such as colour space, bit depth and colour model, in such a way that a common formal expression is applicable to all types of components. Definitions of significant properties that affect the aesthetics, implied meaning, and affordances of digital objects tend to be much more subjective and tied to the context of creation and use. Whether or not colour, for example, is a significant property of a given digital object or collection will depend on the extent to which colour features affect the quality and usability of the preserved object for a designated community.

Our model also includes a means for expressing internal composition and external relations between complex objects. By internal composition we mean the methods used to assemble components into complex objects and the means for associating various components with each other. Specifying how a complex object is assembled is important because there are several viable options for preserving complex objects, provided that their internal composition is specified. For example, a repository could preserve all of the components of a complex object along with the methods or procedures that manage the relationships among components, or it could preserve the components as separate files along with specifications for how to reconstruct those relationships. Our model also specifies the types of external relationships between different components. These relationships are important in choosing preservation strategies, especially when components of digital objects are accessed by way of link or pointer to another digital object. The problem of linked content is pervasive for web documents, and collection managers are grappling with policies and methods for dealing with linked objects. A variety of strategies have been proposed for addressing the persistence of links, but most include some sort of string to serve as an identi-

(8) Kenney, A. R., and Reiger, O., *Moving theory into practice: Digital imaging for libraries and archives*, Mountain View, CA: RLG, Inc., 2000; and US Library of Congress, American Memory Programme, technical information and background papers, available at: http://memory.loc.gov/ammem/ftpfiles.html

(9) Martínez, J. M., ed., *Overview of the MPEG7 Standard (Version 5.0)*, Singapore: International Organisation for Standardisation, 2001, available at: http://mpeg.telecomitalialab.com/standards/mpeg-7/mpeg-7.htm; MPEG Requirements Group. MPEG-21 Overview. Sydney, Australia, 2001, available at: http://mpeg.telecomitalialab.com/standards/mpeg-21/mpeg-21.htm; and *Draft data dictionary: Technical metadata for digital still images,* Version 1.0. National Information Standards Organisation, 2000.

(10) World Wide Web Consortium, 'Cascading style sheets' http://www.w3.org/Style/ CSS/; and World Wide Web Consortium, 'eXtensible style sheet language', http://www.w3.org/TR/xsl/.

(11) Cedars project, 'Metadata for digital preservation: The Cedars outline specification, 2000', available at: http://www.leeds.ac.uk/cedars/OutlineSpec.htm; Phillips, M., Woodyard, D., Bradley, K., and Webb, C., 'Preservation metadata for digital collections', National Library of Australia, 1999, available at: http://www.nla.gov.au/preserve/pmeta.html; Lupovici, C. and Masanès, M., 'Metadata for long term-preservation', Bibliothèque nationale de France, 2000, available: http://www.kb.nl/coop/nedlib/results/D4.2/D4.2.htm; Sperberg-McQueen, C. M., and Burnard, L., eds. *Guidelines for electronic text encoding and interchange: XML-compatible edition (TEI P4)*: TEI Consortium, 2001, available at: http://www.tei-c.org/P4X/; and 'A recommendation for content information. A report by the OCLC/RLG Working Group on Preservation Metadata', (October 2001), available at: www.aclc.org/ research/pmwg/.

(12) Consultative Committee for Space Data Systems, 'Reference model for an open archival information system (OAIS)', *Red book*, Issue 1.2, Mountain View, CA, June 2001.

(13) Holdsworth, D. and. Sergeant, D. M., 'A blueprint for representation information in the OAIS model', IEEE Mass Storage Conference (March 2000), available at: www.leeds.ac.uk/cedars/.

(14) Holdsworth and Sergeant, *A blueprint for representation information in the OAIS model: 416.*

fier, a specification of a domain within which the identifier is meant to apply, and some mechanism for resolving the identifier as a reference to a specific location within a storage system.

Our approach simply provides a means to express the attributes of external relations. In our model, the properties of external relations include such factors as:

- cardinality (one to one, one to many, many to many)
- direction (outbound, inbound, third party, bi-directional)
- semantic relation
- locality (specified by a value, specified by a location)
- access type (ftp, afs, local file)
- target type (pointer, character string, binary object, calculated value, bounded area),

etc.

By specifying the properties of external relations, we provide a basis for breaking down the functions and behaviours of hyperlinks and pointers so that collection managers can decide which properties of the external relations are important to preserve. It is worth noting that the properties of external relations are also applicable to the internal composition of digital objects, if their components are associated with each other through pointers or links. Whether a link or pointer supports internal composition or external relations depends on where a repository draws the boundaries of a given digital object. For example, if a web page links to other web pages within the same site and that site is preserved as a single digital object within a repository, then the links would be considered properties of that digital object's internal composition. Links from those pages to other pages not preserved in the repository could be treated as external relations.

Significant properties and digital preservation strategies: applications

A formal expression of significant properties of complex digital objects has many general and practical applications. Such a model can be applied to appraisal and selection of digital materials, to assessing the risk of information loss associated with various preservation strategies, to the development of preservation metadata, to documenting the basis for preservation decisions, and to the automated management of complex digital objects. In future research, we will also test the model as a basis for developing cost comparisons and trade-offs associated with various digital preservation strategies.

Appraisal and collection management decisions about complex digital objects entail choices about which significant properties of objects and their components merit preservation. For example, a repository might decide to preserve the text, images, and graphics of a web page along with procedures to reproduce its original composition, but to forego the video segments. The repository may decide to preserve the images at a lower resolution than the original. No single set of decision-making rules can or should apply to all such choices. Rather, decisions about which significant properties to maintain will depend on institutional priorities, anticipated use, knowledge of the designated community, the types of materials involved, and the financial and technical resources available to the repository. Nevertheless, a formal means for expressing significant properties provides the basis for more clearly articulating the range of available options and for documenting the trade-offs among them.

Collection managers can also use a formal expression of significant properties to select preservation techniques and to guide choices about the timing of various preservation actions. Anne Kenney and her colleagues at Cornell University have completed two studies that applied risk management methods to digital preservation strategies ([15]). Their multi-stage process includes risk identification, risk classification, risk assessment, risk analysis, and risk management implementation. Our formal model of significant properties could be used for identification, classification, and risk analysis. A repository using a risk management approach could evaluate the risks of technology obsolescence for various component types, the risks of changes to objects that are externally linked, and the risk that a preservation action might degrade the object beyond the point of use-

and collections, electronic records management, and archives as memory institutions. She has also taught numerous continuing education workshops in the United States, Canada, Australia, New Zealand, Sweden, and South Africa.
Professional activities and service: Hedstrom has participated in numerous elected, appointed, and voluntary professional and service activities. She is currently a member of the National Digital Strategy Advisory Board which is advising the Library of Congress on development of a national infrastructure for digital preservation. She is a member of an expert reference group on IT questions for the Swedish Government. She served on the National Research Council Study Committee on the Digital Future of the Library of Congress, 1999–2000; the Commission on Preservation and Access/Research Libraries Group Task Force on Archiving Digital Information, 1994–96; on the Council of the Society of American Archivists, 1992–95; and as Secretary, Vice-Chair, and Chair of the New York State Forum for Information Resources Management, 1992–95. She has been a consultant to more than a dozen government archival programs, the World Bank, the International Council on Archives, and recently the University of Fort Hare in South Africa. Hedstrom is a fellow of the Society of American Archivists and she was the first recipient of the annual Award for Excellence in New York State Government Information Services.

Christopher Lee

Christopher Lee is currently pursuing a Ph.D. at the University of Michigan School of Information after previously obtaining a Master of Science

([15]) Kenney, A. R., et al, 'Preservation risk management for web resources: Virtual remote control in Cornell's project prism, *D-Lib Magazine* 8:1 (January 2002); available at: http://www.dlib.org/dlib/january02/kenney/01kenney.html; and Lawrence, G. W., et al, *Risk management of digital information. A file format investigation*, Washington D.C.: CLIR (June 2000), available: http://www.clir.org/pubs/reports/reports.html.

degree in information with a concentration in archives and records management there in 1999 and a Bachelor of Arts in Philosophy (summa cum laude with honours) with a concentration in public service at Albion College in 1995.
He has been a graduate research assistant in the Camileon (creative archiving at Michigan and Leeds: emulating the old on the new) project since August 2000. Prior to that he worked as an electronic records project archivist at the Kansas State Historical Society between May 1999 and August 2000. Other professional activities include: project member, LUIGUI (Linux/UNIX Independent Group for Usability Information), February–November 1999; co-investigator, Online Archival Description Project, University of Michigan School of Information, January–May 1999; graduate research assistant for Dr Michael Cohen, University of Michigan School of Information, September 1998–May 1999; project team member, Formshare, University of Michigan School of Information — September 1997–September 1998; webmaster, technical advisor and archival processor, University of Fort Hare, National Heritage Cultural Studies Centre, Alice, South Africa, May–June 1998.
Recent publications and presentations include: Guerilla ERM: Lessons learned from some time in the trenches, Ohio Archivist, Spring 2001; Open-source software: A promising piece of the digital preservation puzzle, electronic currents, Midwest Archives Conference (MAC) Newsletter, Volume 29, Number 2 (113), October 2001; Electronic records and government accountability:

fulness. As repositories accumulate experience with such decisions, a common vocabulary for expressing significant properties would provide the basis for sharing data and assessments about the impact of various preservation strategies.

The formal model can assist in selecting appropriate preservation strategies based on acceptable levels of risk defined by each custodial institution. In the Camileon project, we are investigating the feasibility of using emulation as a digital preservation strategy. The results of our research show that emulation is technically feasible, but it is not always most effective or cost-effective strategy. If the primary goal of preservation is to preserve the content and structure of a collection of textual documents, then converting the materials to structured text using a mark-up language such as SGML or XML, or to a page description format, such as PDF, may be more cost-effective and easier technically than emulating the original computer environment. If, on the other hand, properties such as navigational aids, optional viewing methods, macro-enabled procedures, or internal links were considered significant to the designated community, then emulation might be a better choice for preserving this range of functionality.

The model can be applied at several points in the digital preservation process. The open archival information system reference model (OAIS) provides a common set of terms for describing the processes and information flow in an archival repository. Using the OAIS terminology, the formal expression of significant properties is applicable to transformations that might be necessary when resources are extracted from their original creation environment and used to create a submission information package (SIP); when the SIP is transformed into an archival information package (AIP); when it is necessary to repackage or transform an AIP; or when a dissemination information package (DIP) is produced for a consumer. Our model for expressing significant properties is also closely related to the concept of a representation network in the OAIS reference model. The OAIS model defines a category of information called representation information that accompanies a digital object and is used to convert bit sequences into more meaningful information. Representation information is further divided into two types: structure information and semantic information. Structure information describes the format or underlying data structure, similar to the lowest-level properties in our model that are common across all data types, such as the basic data unit, byte-level encoding, data typing, and logical structure. Semantic information is information contained in or associated with an object that is necessary to interpret its meaning. In the case of textual information, for example, an indication of the language or languages represented in the text would be important for future interpretation of the content. Similarly, meaningful analysis of scientific data requires additional semantic information that describes how meaning is ascribed to data elements and how different pieces of data are related ([16]).

Although we found the distinction between structure information and semantic information useful, we agree with the model's developers that in some implementations, that distinction is subjective. Depending on how digital objects are created and managed, significant properties can be imbedded in the structure of an object or managed with additional semantic information, or both. For example, structured text uses mark-up languages and document type definitions to formally describe the meaning associated with various text formatting and layout practices. Well-designed mark-up makes a distinction between the structure of a document and its semantics. In proprietary word processing applications, such a formal distinction may or may not be present. Most contemporary word processing systems include a facility for using style sheets that manage the document structure and its associated semantics. If the stylistic features constitute significant properties of the digital objects, then exporting the object from its native software environment to another platform could eliminate or alter the way the significant properties are represented.

A formal model for significant properties also has many implications for preservation metadata. Our work is an extension to a number of existing metadata models rather than a new approach that requires new concepts, schema, or implementation methods. Both the OCLC/RLG Working Group on Preservation Metadata and the Cedars project recommend that metadata schema to support preservation should include elements that describe the significant properties of digital objects, but neither of these proposals includes a vocabulary for expressing significant properties ([17]). The metadata encoding and transmission standard (METS) provides a framework for exchanging metadata among repositories ([18]). The administrative metadata portion of METS, for example, provides a structure for maintaining technical metadata about the creation, format, and use characteristics of a file or object; intellectual property rights metadata; metadata about the

([16]) Consultative Committee on Space Data Systems, 4-20–4-21.
([17]) Cedars, 'Metadata for digital preservation: The Cedars project outline specification, draft for public consultation', available at: http://www.leeds.ac.uk/cedars/colman/metadata/metadataspec.html; and the OCLC/RLG Working Group on Preservation Metadata, 'A recommendation for content information'.
([18]) Metadata encoding and transmission standard (METS), 2001, available at: www.loc.gov/standards/mets/.

source of the object; and digital provenance for documenting migrations or transformations employed on files. A formal language for expressing significant properties could augment this model by indicating which significant properties were retained which were altered or eliminated in at ingest or through subsequent migrations.

The formal model has at least one other potential application for developing cost models and conducting cost benefit analyses of various digital preservation options. The field of digital preservation in general suffers from the lack of cost models that allow collection managers and decision makers to estimate immediate and long-term costs of digital preservation. We do know, however, that different technical strategies impact long-term preservation costs. A formal means for determining which significant properties are lost or retained using various preservation strategies could assist custodians in selecting strategies and in justifying them to users in the future.

Our work is in a formative stage and it offers only a general framework for expressing significant properties. Further articulation of significant properties is needed for specific types of digital objects and collections for particular designated communities. This work will require input from domain experts who are familiar with the semantics, use practices, and potential users within particular designated communities. We hope that specific communities will populate this model with context-specific cases of significant attributes.

Further work on significant properties also must contend with the role of interactions between digital objects, as they are structured and stored for access, and the user environment. Increasingly, digital objects are stored in some underlying format, but the way the objects appear when they are 'served' or delivered to users depends on a variety of factors in the user's environment. Choices of viewers or browsers as well as user-defined preferences can influence the appearance and behaviour of digital objects. Digital objects may have dynamic formatting effects, where user actions influence the behavior or representation of the content. These affordances raise new challenges for defining what constitutes a digital object and which of its properties are significant to whom.

Present practice and future prospects (with David Wallace) Computer professionals for social responsibility (CPSR) Annual Conference, October 20, 2001; Emulation, migration and long-term preservation of electronic records — ECURE 2001: Preservation and Access for Electronic College and University Records, 12–13 October, 2001; Beyond the byte stream: What to preserve, Society of American Archivists (SAA) annual meeting, 31 August 2001.

Propiedades significativas de los documentos digitales: definiciones, aplicaciones e implicaciones

Margaret Hedstrom, Christopher Lee

La conservación debería implicar la retención de las propiedades más significativas de los objetos digitales desde una generación tecnológica a la próxima. Desafortunadamente, en la actualidad no contamos con ningún medio formal para describir estas propiedades significativas.

El presente documento defiende que las decisiones sobre los métodos de conservación digital deben estar ilustradas por un análisis de las propiedades significativas de los objetos digitales, a fin de conservar las propiedades que justifiquen un uso futuro de los objetos digitales. Presentaremos la investigación inicial sobre los métodos destinados a identificar y expresar formalmente las propiedades significativas. Estos métodos formales contribuirán a una toma de decisiones eficaz sobre los métodos de conservación más adecuados para las distintas clases de objetos digitales que se conservan para comunidades concretas.

La importancia de retener las características originales de los objetos se ha planteado en diversos contextos. Cuando se cambian de formato los objetos físicos utilizando técnicas como la microfilmación, parte de las cualidades físicas de los documentos no son capturadas por la técnica de cambio de formato. Problemas similares pueden ocurrir cuando se con-