



Document Name			
Document Title	Deciding factors: Issues that influence decision-making on significant properties		
Work Package			
Author(s) & project role	Gareth Knight, Digital Curation Specialist		
Date	12 February 2010	Filename	Deciding-factors.doc
Access	<input checked="" type="checkbox"/> Project and JISC internal		<input checked="" type="checkbox"/> General dissemination

Document History		
Version	Date	Comments
1.1	12 February 10	Copied text into revised template
1.0	10 June 08	First version



Table of Contents

Table of Contents	2
Purpose.....	3
Introduction	3
Understanding value	3
Atomic model of significant properties	4
Composition.....	4
Purpose	5
Investment.....	6
Capability.....	7
Summary.....	8
References.....	9

Purpose

The InSPECT Project is funded by JISC to investigate methods for maintaining the authenticity of digital resources across transformation processes and over time. It is developing a framework that will allow institutions to identify, measure, and declare the significant properties of a specified group of digital object types. This discussion paper identifies the factors that contribute to the evaluation of significant properties.

Introduction

Significant properties are those aspects of a digital record that must be preserved over time in order for it to remain accessible and meaningful. The identification of significant properties that contribute to these objectives may be considered an intellectual exercise, as opposed a technical challenge. In this paper we introduce four broad categories that have been identified as contributing to the understanding of value and are likely to have an influence on the decision-making process through which the properties to be maintained are identified. The InSPECT Project encourages the discussion of each category, and contributions from researchers, to clarify existing factors and introduce additional considerations.

Understanding value

The development of a broad understanding of value as a concept, and its subsequent application to digital resources, requires consideration of the terminology used for discussion.

The Oxford English Dictionary defines 'value' as a verb, indicating it is used to 'estimate or appraise as being worth a specified sum or amount' (OED, n.d.), or as a noun (e.g. value for money), to be 'equivalent for something else; a fair or adequate equivalent or return'. Subsequent reference to the definition of estimate, appraise and adequate equivalent imply that an evaluation is required. In philosophy it is common to find distinctions between intrinsic value, the concept that something has value "in its own right" exclusive of an external influence, and extrinsic value, the concept that value is derived from an external function (Zimmerman, 2007). The wider use of the term suggests it is not possible to create a definitive assessment of value that is applicable to all. Instead it is better to consider the application of the term as being determined by specified criteria. The value attributed to physical and digital resources may be considered extrinsic, influenced by factors assigned by an individual, institution, a community of users, or multiple communities in the wider world.

Atomic model of significant properties

The InSPECT Project has identified four broad categories that contribute to the understanding of value and the decision-making process of properties that must be maintained. Figure 1 does not present an exhaustive list of factors that may be relevant in considering significance and in practice there may be some level of crossover and internal decision-making for the factors with the greatest importance. However, it is likely to be useful for considering the significant properties that are required for specific uses.

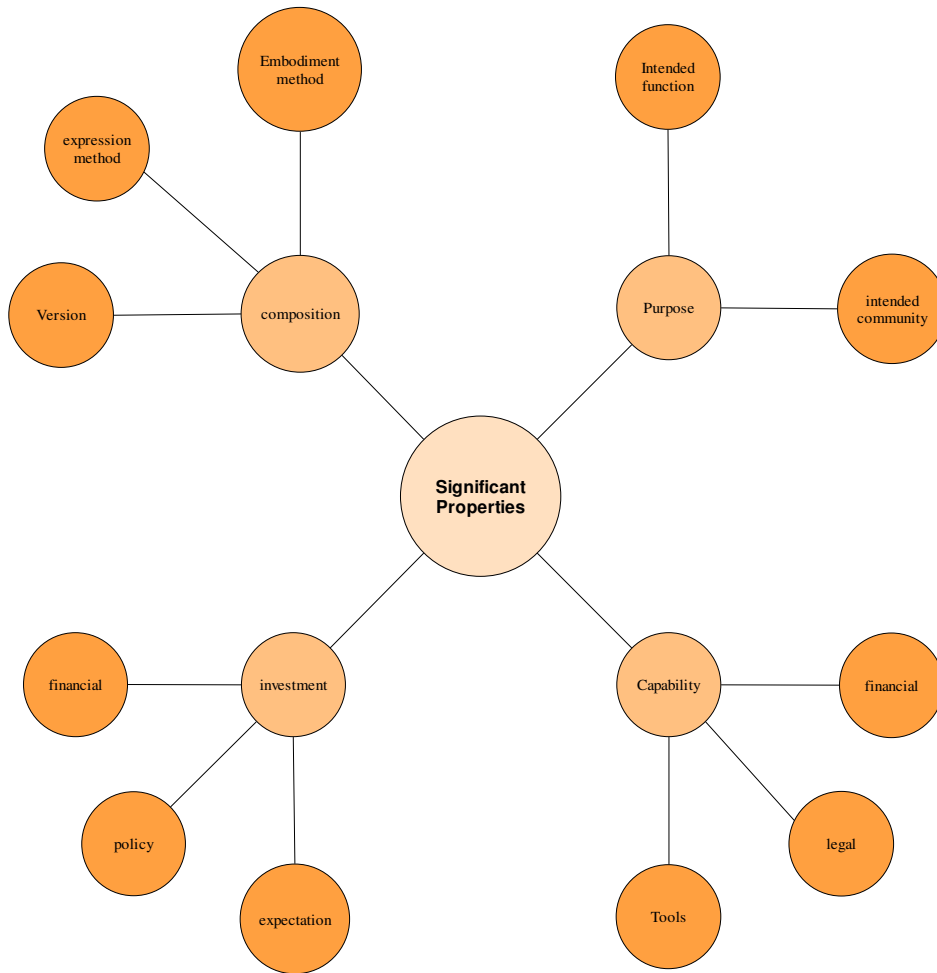


Figure 1: An atomistic model of factors that contribute to the definition of significant properties

Composition

Composition encapsulates the method in which a Creator expresses an intellectual or artistic idea and renders it to a particular media. The composition of a digital resource is likely to be influenced by several factors, including the chosen method of expressing an idea; the method in which it is created; and the process that the creator and other agents use to create and revise it. These factors are encapsulated in the three headings: Expression method, Embodiment method and version respectively.

Expression method

This is the method that is chosen to express an intellectual or artistic idea. Potential examples of expression include alphanumeric text, visual imagery, sound recording, animation, etc. The Expression Method is derived from the Functional Requirements for Bibliographic Records (FRBR) understanding of Expression. The expression method may be chosen by one or more agents (e.g. Creator, Curator, Publisher, etc.) for a number of reasons. Possible explanations may include the preference of the Creator when developing the intellectual or artistic idea (e.g. a painter may choose to deliver information in abstract, expressionist, cubist, or other painting style; a writer may choose to express a model as a spider diagram or system model), the medium that is used (e.g. a canvas, A4 lined paper, music sheet, computer system), the tools that were used to express an idea (e.g. a musical instrument, paint brush, software application) and/or the establishment of some preference in the method that an idea is expressed for interpretation by a Designated Community (Coyne et al, 2007). See Purpose for further information of the latter.

Embodiment Method

The Embodiment Method refers to any method of storing and presenting the expression that may possess properties that become interwoven with the significant properties of the Expression in its entirety. It is equivalent to the FRBR concept of a Manifestation. Potential examples that contribute properties may include the medium on which the Expression is physically embodied (e.g. paper type, audio cassette, video cassette, canvas, plaster, etc.) and the file format in which an Expression is manifested (e.g. JPEG image format, TIFF image format, Microsoft Word Document).

Version

Version is used to describe the instance of a digital resource that is the target for curation and preservation (Brace, 2008). A version of a digital resource may be related to another, as a result of a creation and development timeline, as well similarities in intellectual content and authorship. Version is considered important on the basis that a digital resource may possess a set of significant properties that differ from previous and subsequent versions resulting from events that have occurred during its lifecycle. For example,

- A research paper may include more or less words in the submitted version of a book chapter in comparison to the published version.
- A received version of an email will contain properties related to the received datetime and delivery path, in addition to existing information contained in the submitted version.
- A final version of an still or moving image may contain additional metadata that describes the information content.

A decision on significant properties may also be affected by a desire to establish the relationship between different versions of a digital resource, by maintaining a version number or other relevant information.

Purpose

Purpose indicates the anticipated outcome that is intended or that guides planned actions to be performed. The Intended Function and Community headings serve as a focused method to identify properties, based upon their ability to fulfil a specific purpose (e.g. for preservation or reuse). However, a number of potential issues may be identified: some resources do not have one or more clearly defined functions that they are intended to perform or intended communities that they must serve and it is often difficult to establish intention with any degree of accuracy; it is common for the same type of Expression to be created to perform different functions and serve different communities, which as a result may require different properties to be maintained; and the identification of a large number of functions/communities may make it difficult to establish a minimum set of properties that must be maintained (Moss, H & Hampton, J, 2003). Manual intervention may be necessary to qualitatively identify the purpose of certain types of resource.

Intended Function

The Intended Function indicates the reason for which the Expression exists and, potentially the Embodiment Method that has been used to store it. An Expression may fulfil one or more functions throughout its lifecycle, specified by a number of Agents. Its intended function is influenced by the aims and objectives of an Agent at a particular time, and on its history of use. For example, an Expression may have one or more of the following functions intended to achieve a desired result or objective;

- A Creator will produce an resource to fulfil an immediate aim or objective. For example, an author will write a research paper to submit to a journal or fulfil funding requirements:
- A hosting institution will maintain an resource and append additional information to fulfil an intended or desired result or objective. For example, a research paper will be stored and a cover page will be appended for the purpose of citation:
- A third-party (a student or academic researcher) may access a resource to perform research and analysis:
- A third-party may repurpose a resource to fulfil an intended aim or objective. For example, a teacher may repackage it in a learning object for use in class tuition or distance learning course (Ashley, Davis, Pinsett, 2008); a scientist may reformat data for cross-analysis; a software developer may reuse software code in a different context (Matthews et al, 2008).

When asked to assess the value of properties on the basis of an intended function, certain elements may be given more weight. For example, a Creator and third party that wish to reuse the information content will require the ability to edit the resource, but others may not.

Intended Community

An analysis of the Intended Community of an Expression may establish specific properties that must be maintained to support access, use and reuse by that user group. An Expression may possess one or more Intended Communities during its lifecycle that have specific requirements. For example:

- A Creator may produce a resource to fulfil aims and objectives or requirements established by an intended community in the short-term (e.g. an author will write a research paper for publication, which will be reviewed for suitability by peer reviewers and subsequently evaluated by journal readers);
- A Curatorial institution will maintain an Expression for the purpose of curation and preservation. The Intended Community may be itself, a successor organisation with similar requirements to curate and preserve or a funding body (see Assessor Valuation):
- A Digital Repository may repurpose the Creator's resource to fulfil the needs and expectations of its Designated Community. For example, medical research analysis may be repurposed as an artistic work;
- An image may be incorporated into an Learning Object for use in teaching (Ashley, Davis, Pinsett, 2008).

The properties that are valued as significant may differ and change, influenced by the Intended Community of the Expression during a specific time period. To effectively curate a resource, it is essential that properties necessary to support current and future Intended Communities are maintained.

Investment

Investment indicates factors that affect the commitment that an institution will make to an assessment of significant properties. Investment may be separated into two sub-categories:

Financial

The total money, time and resources that have previously been invested in performing activities associated with digital curation and preservation. An institution that has made a considerable investment in digital curation and preservation may consider it to be essential that they continue to invest and develop their practices, identifying a large number of properties that they consider to be essential.

Strategy

A strategy establishes the course of action that an institution may take to curate and preserve a digital resource, which may include specific recommendations on requirements that must be met. An institution may establish a strategy for curating a set of significant properties for several reasons, including a desire to establish trust in their operation or a mandate to maintain resources to a pre-defined quality level. The latter rationale, in particular, is likely to contribute considerably to the properties of a digital resource that an institution considers to be significant.

Expectation

These are the expectations of the Creator or Designated Community, in regards to the standard of digital curation and preservation that will be performed and the significant properties that will be maintained. In the context of this paper, it is recognised there may be some overlap between investment expectations and the purpose of the dissemination manifestation, as used by the intended community.

Capability

Capability indicates the ability or capacity that the assessor can demonstrate to identify, analyse and/or extract significant properties at a desired level of detail. It is influenced by three factors:

Tools

The software tools that are available and in use are likely to have some influence on the ability to identify and extract significant properties. The assessor may be required to fund or develop software code and scripts to identify significant properties, if existing tools are incapable of extracting significant properties at the preferred level of analysis.

Legal

Copyright is a form of intangible property that applies to certain types of works. It is automatically assigned to one or more creators at the point of creation. An creator may subsequently assign it to a third-party. The copyright of a digital resource may be owned by one or more agents that assign different criteria to the access and use of different properties. For example, an author may own literary copyright of words, while a publisher may own the typographical copyright of a research paper (British Academy, 2008; Knight, 2005). An institution with a commitment to curate and preserve the significant properties of a digital resource may be limited in its actions by the legal rights that have been assigned to it, which is likely to influence the properties that it considers sufficiently important to maintain.

Financial

This is the total money, time and resources that may be associated with the activity of identifying and evaluating significant properties. It may include the expenditure of money required to purchase a software tool to perform an data analysis; the staff time required to identify significant properties; and/or a combination of factors necessary to validate that the significant properties have been maintained in subsequent manifestations.

The cost of identification, analysis and extraction of significant properties may be reduced through use of existing research data. Several institutions have published guidelines on the significant

properties of specific types of digital resource. For example, the Arts & Humanities Data Service make available the Preservation Handbooks produced between 2004-2008 that is used as the basis for analysis of deposited data¹. The Global Digital Format Registry (GDFR) and The National Archives PRONOM service have also announced plans to provide preservation planning functions (GDFR, n.d; The National Archives, n.d.)

Summary

The atomic model presented in this discussion paper indicates a broad range of factors that an institution may consider when deciding the properties that must be maintained over time and subsequent conversion, the properties that cannot be maintained and those that are superfluous. A subsequent document will indicate how these factors will contribute to the practical decision making process of choosing significant properties.

¹ AHDS Preservation Handbooks may be viewed by visiting <http://www.ahds.ac.uk/preservation/ahds-preservation-documents.htm>.

References

- AHDS (n.d.). AHDS Repository Policies and Procedures. Retrieved on June 9, 2008 from: <http://www.ahds.ac.uk/preservation/ahds-preservation-documents.htm>
- Ashley, K. Davis, R & Pinsett, E (2008). Significant Properties of E-learning Objects (SPeLOs). Draft 0.91. Retrieved on June 9, 2008 from: http://www.jisc.ac.uk/media/documents/programmes/preservation/spelos_report.pdf
- Brace, J et al (2008). Version Identification Framework. . Retrieved on June 9, 2008 from: <http://www.lse.ac.uk/library/vif/Framework/index.html>
- British Academy, The & Publisher Associated, The (2008). Joint Guidelines on Copyright and Academic Research. April 2008. Retrieved on June 10, 2008 from: <http://www.britac.ac.uk/reports/copyright-guidelines/>
- Coyne, M et al (2007). Study on the Significant Properties of Vector Images. Retrieved on June 9, 2008 from: http://www.jisc.ac.uk/whatwedo/programmes/programme_preservation/2008sigprops.aspx
- Global Digital Format Registry (n.d.). Retrieved on June 9, 2008 from: <http://hul.harvard.edu/gdfr/>
- IFLA Study Group (1998). Functional Requirements for Bibliographic Records. Retrieved on June 9, 2008 from: <http://www.ifla.org/VII/s13/frbr/frbr.htm>
- Knight, G. (2005). SHERPA-DP OAIS Report: An OAIS compliant model for Disaggregated services. Retrieved on June 9, 2008 from: <http://www.sherpadp.org.uk/documents/sherpa-dp-oais-report.pdf>
- Library of Congress (2006). Sustainability of Digital Formats: Planning for Library of Congress Collections. Retrieved on June 9, 2008 from: <http://www.digitalpreservation.gov/formats/>
- Matthews, B et al (2008). The Significant Properties of Software: A Study. Retrieved on June 9, 2008 from: http://www.jisc.ac.uk/whatwedo/programmes/programme_preservation/2008sigprops.aspx
- Moss, H. & Hampton, J (2003). Conceptual Representation: Special Issue of Language and Cognitive Processes. Psychology Press. Retrieved on June 9, 2008 from: <http://books.google.co.uk/books?id=yV2JGnIZTnUC>
- Oxford University Press (n.d.). Oxford English Dictionary. Retrieved on June 10, 2008 from: <http://www.oed.com/>
- The National Archives (n.d.). PRONOM: The Technical Registry. Retrieved on June 9, 2008 from: <http://www.nationalarchives.gov.uk/aboutapps/pronom/>
- Zimmerman, M.J. (2007). Intrinsic vs. Extrinsic Value. Substantive revision (07/02/07) Retrieved on June 9, 2008 from: <http://plato.stanford.edu/entries/value-intrinsic-extrinsic/>