Gabriel Chrisman

LIS 539

May 7, 2010

Significant Properties of Digital Records

Within the field of research on digital preservation strategies, the significant properties of digital records have been a major focus of study over the past ten years. It is now widely acknowledged by researchers and practitioners that long-term digital preservation will require some degree of change in the digital record as it passes through migration, emulation systems, and any other likely preservation processes. Significant properties, which have also been discussed by the National Archives of Australia (NAA) as the 'essence' of the record and by other organizations under various other names, are defined by Andrew Wilson as: "the characteristics of digital objects that must be preserved over time in order to ensure the continued accessibility, usability, and meaning of the objects."[1] Is it possible to determine these properties in a general sense, or even in a specific case, and to encode them in order to assist in preservation? How important or valuable is it to conceptualize digital records through this lens? Are there alternative concepts or theories which have been suggested in recent publications?

The OAIS reference model, on which so much of the current thinking about digital archiving is based, does not explicitly define or discuss significant properties. The JISC argues that it strongly implies such a concept of detailed selection criteria by its distinction between the Submission Information Package (SIP), the Archival Information Package (AIP), and the

[1] Wilson, A., *Significant Properties Report*, JISC London, 2007, p. 8. http://www.significantproperties.org.uk/documents/wp22_significant_properties.pdf (accessed 6/1/09).

Dissemination Information Package (DIP).[2] By highlighting the implied differences between the data contents of these stages in the digital preservation process, the choices of which aspects of the record are preserved in each package are assumed. The criteria by which these choices are made are described in the OAIS model as subjective, and the model also indicates that these criteria may experience changes over time. In another publication, Gareth Knight wrote that "In an OAIS, significant properties are the characteristics of the abstract information object (e.g. an image), while representation information indicates characteristics of the data object (e.g. format, encoding scheme, algorithm)."[3] By this definition, the significant properties are essentially the informational content (or even the message) of the record. By any of these definitions, these significant properties appear at first glance to be highly important keys to successful digital preservation programs.

Accordingly, much has been made of the OAIS's fairly vague, implicit endorsement of the need to identify, specify, and codify significant properties. A 2004 article described the identification of significant properties as "a necessary first step" in the digital preservation process,[4] and this has seemingly emerged as a common consensus in the field. However, while the value of the OAIS model in describing the general framework of digital preservation, and the changes which records much undergo in order to be preserved over time, is undeniable, it is less convincing that administering this process requires such in-depth specific detail as is called for

---

[2] JISC InSPECT Project, *Framework for the Definition of Significant Properties*, 2008, p. 4. http://http://www.significantproperties.org.uk/documents/ wp33-propertiesreport-v1.pdf (accessed 6/1/09).

[3] Knight, Gareth, Pennock, Maureen, *Data Without Meaning: Establishing the Significant Properties of Digital Research*, 2008, p. 1. http://www.bl.uk/ipres2008/presentations_day1/16_Knight.pdf (accessed 6/1/09).

[4] Deken, Jean Marie. 2004. "Preserving Digital Libraries: Determining "What?" Before Deciding "How?"." *Science & Technology Libraries* 25(1/2): 227.

by the extraordinarily detailed JISC InSPECT project literature.[5] This paper will investigate how the research into significant properties developed in this direction, and will examine the value of the developing theories and body of literature.

In a 2002 NAA paper, the essence of a digital record is analyzed in terms of the performance model. Different types of documents are discussed, and some examples of significant properties are presented. Most importantly, each record is visualized as a performance which represents "a combination of characteristics, some of which are *incidental* and some of which are *essential* to the meaning of the performance."[6] While some elements of a digital record are to be preserved as important, others are seen as "inconsequential to the record's archival meaning."[7] The atomistic deconstruction of the digital record into components (some important for preservation, and some not) is a neat intellectual exercise in the abstract, but presents serious problems in practice. In order to select and segregate the important information, it seems clear that the place to start would be to first identify all of the characteristics of the record. This in itself would be a significant task, as it would include not only all of the technical aspects of the file format, creating software, operating system, etc., but also all of the intellectual components of the content. This deconstructive approach would inevitably be subjective, highly specific, and hence extremely difficult to generalize to other record types or formats. However, it represents only the beginning of the involved process which is required to identify significant properties.

---

[5] JISC InSPECT Project, *Framework for the Definition of Significant Properties*, 2008, p. 7-47. http://http://www.significantproperties.org.uk/documents/ wp33-propertiesreport-v1.pdf (accessed 6/1/09).

[6] Heslop, H., Davis S. , Wilson A., *An Approach to the Preservation of Digital Records*, Canberra, 2002, p. 13. http://www.naa.gov.au/Images/An-approach-Green-Paper_tcm2-888.pdf (accessed 5/29/09). Emphasis added.

[7] Heslop, H., Davis S. , Wilson A., *An Approach to the Preservation of Digital Records*, Canberra, 2002, p. 14. http://www.naa.gov.au/Images/An-approach-Green-Paper_tcm2-888.pdf (accessed 5/29/09).

Following this exercise, the next Herculean task would be to determine the creator's intent, and select the specific components which carried out that intent within the record. This act of judgment goes far beyond Schellenburg's doctrines supporting the necessity of archival appraisal and selection,[8] which essentially operate through selecting complete items (or, even more likely, groups of items) for preservation, into the realm of editorial license. Finally, as Gareth Knight and Maureen Pennock identify in a 2008 paper, "It is…important to identify the potential stakeholders and understand the functions that will be required of the information object and the environment in which it will be used, as criteria for evaluating…preservation strategies."[9] Trying to understand and address the needs of all potential future users adds yet another difficult dimension to the significant properties endeavor. As has been pointed out, "…the key question about significant properties is: for whom are they significant?"[10] Research has been conducted on various different perspectives on significant properties, from the archivist, to the creator, to the user.[11] Even if such research leads to locally useful solutions or programs, how generally applicable or reusable will all of this work be? Will it even be useful in the same institutional context over time, as formats and records change? The 2002 NAA paper states flat-out that "Determining the essence of records is not a science and is open to subjectivities and

---

[8] Schellenberg, Theodore R. 1956. *Modern Archives: Principles and Techniques*. Chicago: University of Chicago Press.

[9] Knight, Gareth, Pennock, Maureen, *Data Without Meaning: Establishing the Significant Properties of Digital Research*, 2008, p. 2. http://www.bl.uk/ipres2008/presentations_day1/16_Knight.pdf (accessed 6/1/09).

[10] Hockx-Yu, Helen, Knight, Gareth, "What to Preserve?: Significant Properties of Digital Objects." *The International Journal of Digital Curation* 1(3): 150.

[11] Hedstrom, Margaret L., Christopher A. Lee, Judith S. Olson, and Clifford A. Lampe. 2006. ""The Old Version Flickers More": Digital Preservation from the User's Perspective." *American Archivist* 69(1): 159-187.

archival interpretation…," but claims that "it is essential to an efficient, effective and accountable preservation program."[12]

This claim requires some closer scrutiny. First, the legalistic focus on accountability is very limited in the scope it envisions for the possibilities of digital preservation. Traditional, physical archives put forward many reasons for preservation of materials, and accountability was only one of them. Besides the limitations, I also question how the identification of significant properties contributes to accountability if they are acknowledged to be completely subjective constructions of the archivist. Theory aside, the other stated goals of efficiency and effectiveness would seem to me to be ill-served by investing time and energy in a potentially endless series of specifications for different sets of significant properties metadata encoding schemes for every new file format, as well as for every new record type within each institution generating records. This job would be a Sisyphean task, which would, even if it were completely successful, still result in only partial preservation of the digital artifacts in question because only those properties deemed significant would be retained – and it is easy to imagine situations in which the particular subjective assumptions made by the archivists could lead to preservation of the 'wrong' data from the perspective of future archival researchers.

Some objections do appear to have been raised at conferences, though not always the ones I would have expected: in 2006, one breakout group "proposed the funding of individuals (or working groups) to develop templates of the "significant properties" of different types of objects, to evaluate standards and develop consensus, and to facilitate staff-exchanges across

[12] Heslop, H., Davis S. , Wilson A., *An Approach to the Preservation of Digital Records*, Canberra, 2002, p. 14. http://www.naa.gov.au/Images/An-approach-Green-Paper_tcm2-888.pdf (accessed 5/29/09).

national and disciplinary boundaries."[13] However, it was noted that "Another subject that came up for discussion was the exact role of "significant properties" *vis-à-vis* costs, Caroline Arms (Library of Congress) and others being keen to see some identification of the different cost factors that would apply to preserving those characteristics of objects that are deemed essential."[14] The costs involved with setting up these working groups and selectively preserving content would certainly be large, and, again, the OAIS model reminds us of the inevitable subjectivity of the resulting schema. Even if these standards could be shared to some degree in order to spread out the costs between institutions, it is likely that the resultant standards would be either too vague to be useful in truly fostering interoperability and standardization, or else too specific to be widely applicable.

A recent publication identified seven major projects which have made "important contribution[s]" to the work on identifying significant properties in research datasets.[15] In surveying these widely varied groups, "the distinct methodologies adopted by each JISC project suggest that further work is necessary to encourage adoption of the Utility Analysis and Digital Diplomatics methodologies."[16] While this is diplomatically phrased, essentially this indicates that the current systems under development are not compatible or interoperable. Besides this basic incompatibility of methodologies,

---

[13] Day, Michael, Hockx-Yu, Helen, "Joint US-UK Digital Preservation Workshop, Washington, D.C., May 7-9, 2006." *International Journal of Digital Curation* 1(1): 68.

[14] Day, Michael, Hockx-Yu, Helen, "Joint US-UK Digital Preservation Workshop, Washington, D.C., May 7-9, 2006." *International Journal of Digital Curation* 1(1): 69.

[15] Knight, Gareth, Pennock, Maureen, *Data Without Meaning: Establishing the Significant Properties of Digital Research*, 2008, p. 1-2. http://www.bl.uk/ipres2008/presentations_day1/16_Knight.pdf (accessed 6/1/09).

[16] , Gareth, Pennock, Maureen, *Data Without Meaning: Establishing the Significant Properties of Digital Research*, 2008, p. 7. http://www.bl.uk/ipres2008/presentations_day1/16_Knight.pdf (accessed 6/1/09).

> "it is evident that there remains some difference in the understanding of properties that may be categorized as significant for the information object and those that may be classified as Representation Information and that further work is necessary to map the significant properties of an information object onto a conceptual and practical model in a consistent manner."[17]

Because the OAIS model is not specific about the position of significant properties, subjective interpretation has led to inconsistent assumptions and mismatched selections of even what can count as potential significant properties. Despite this evidence of the unavoidably subjective nature of significant properties suggested by the OAIS model and other publications, Gareth Knight and Maureen Pennock write of the seven major projects they reviewed that "Although each project has a distinct conceptual basis and methodology, the outputs of earlier work has [sic] contributed to the development of subsequent projects." Building on the field (or at least adding to it) does not seem to be a problem: the work of analyzing significant properties is as potentially endless as the number of record types and formats, but it is not clear that it is ultimately leading anywhere constructive.

The recent publications on the issue of significant properties of digital records are inconclusive, but do almost universally encourage further work in the area. For instance, Gareth Knight and Maureen Pennock write:

> "The review of projects and institutions that have made some contribution to the development of digital preservation strategies suggests that there is a great interest in the identification, analysis and extraction of significant properties…[However,] We have yet to reach the stage where a researcher or academic in an institution is

[17] , Gareth, Pennock, Maureen, *Data Without Meaning: Establishing the Significant Properties of Digital Research*, 2008, p. 7. http://www.bl.uk/ipres2008/presentations_day1/16_Knight.pdf (accessed 6/1/09).

> able to define the significant properties of their digital research without ambiguity."[18]

This ambiguity, after ten years of theoretical and practical work, is telling. While it seems as though there is still interest in pursuing the subject of significant properties in research, I found myself in complete agreement with Andrew Wilson's trepidations about the usefulness of the whole enterprise.[19] However, if we eliminate the possibility of determining the significant properties of digital records in a general sense, and question the efficiency and purpose of determining them in many specific cases, how does this affect our goals and models for digital preservation and archiving? Is there an alternative model that has been proposed, or is an alternative even necessary?

I have not found any directly mapped alternatives to the significant properties concept proposed in the recently published literature, but I do not think that this reflects universal acceptance of the theory. On the contrary, I feel that there is an easy alternative (addressed by Andrew Wilson's lectures during this course) which requires no added theoretical constructions or discussion.

Because the OAIS model does not specifically discuss significant properties, this most fundamental digital archive model is not affected when we simply remove this accumulated set of assumptions from our consideration. The distinctions between the SIP, AIP, and DIP are therefore understood to be produced with less conscious or pre-programmed deliberation, by

---

[18] Knight, Gareth, Pennock, Maureen, *Data Without Meaning: Establishing the Significant Properties of Digital Research*, 2008, p. 7. http://www.bl.uk/ipres2008/presentations_day1/16_Knight.pdf (accessed 6/1/09).

[19] Wilson, Andrew, Online lecture, University of Washington iSchool LIS 539, May 2008. http://uweoconnect.extn.washington.edu/p38966038/

simply preserving as much as it is technically possible to preserve of each record which is selected for 'permanent' archiving. Technical possibilities and technological limits of the archiving institution become the guiding forces behind what material is preserved. Whatever their format, the SIP, AIP, and DIP should all also include the original bitstream. This would ensure the preservation of all of the original information in that form at the very least, offering the possibility of later recovery or interpretation. When information is lost or unavoidably changed in the 'use copy' component of the AIP or DIP as a result of ongoing preservation activity such as reformatting, this loss is recorded and added to the preservation metadata for the object in its AIP. This OAIS-based interpretation seems very close to the actual practices of physical, traditional archives, and hopefully will result in similar outcomes as far as preservation success. We cannot expect more than that – digital archives should not be held to a higher standard than our past archival systems. Digital records may be more vulnerable to deterioration than many older record types or formats, but to me this should not influence or change the essential nature of what archivists do.

I do not think that ignoring or failing to concern ourselves with the analysis of significant properties will negatively impact the mission of digital archival institutions. On the contrary, an increasing amount of time and energy devoted to deconstructing file formats, record types, and user needs accompanied by increasingly selective and time-consuming editorial work on the materials stored in AIPs seems like it might detract from the more basic missions of the archival institution: simple preservation (to the best of our technical ability) and user service (offering unbiased research support). If there is value to in-depth analysis of either technical aspects of file formats, or the informational content of digital records, it does not seem like working archives would be the primary beneficiary of such research on a practical level. Especially at the

large scale that digital archives are expected to operate at, this practice seems like it invites an item-level, specific approach to processing digital records when time, cost, and the increasing volume of incoming material dictates a series-level, general approach.

The situation presented to us by the growing volume of digital records in widely varied formats and on different media types is a difficult one. Given the practical realities of the predicament of modern archival institutions, the realistic options available are limited. Essentially, the guiding philosophies that seem as though they would be most appropriate to this new digital records environment bring to mind a combination of Schellenberg's theories of selective preservation,[20] and Greene and Meissner's more recent theories of limited processing.[21] Both of these involve making decisions essentially at the series level, and rarely examining or processing items directly unless they are of extreme intrinsic value. While this approach does at least require an understanding of the most basic requirements of digital records for their preservation, and acknowledgement of the need for progressive changes in digital materials for preservation purposes, it does not require an in-depth analysis of each record type. As Andrew Wilson pointed out in lecture, it is both better and easier to simply preserve all that it is technically possible to preserve of the digital records selected for archiving. While the ongoing analysis of significant properties may be of interest to theoretical information science researchers seeking to understand the communication of information or message in digital formats, I don't see a future for it in the practical world of working digital repositories.

---

[20] Schellenberg, Theodore R. 1956. *Modern Archives: Principles and Techniques*. Chicago: University of Chicago Press.

[21] Greene, Mark A., and Dennis Meissner. 2005. "More Product, Less Process: Revamping Traditional Archival Processing." *American Archivist* 68(2): 208-263.

Bibliography:

Day, Michael, Hockx-Yu, Helen, “Joint US-UK Digital Preservation Workshop, Washington, D.C., May 7-9, 2006.” *International Journal of Digital Curation* 1(1): 60-76.

Deken, Jean Marie. 2004. "Preserving Digital Libraries: Determining "What?" Before Deciding "How?"." *Science & Technology Libraries* 25(1/2): 227-241.

Greene, Mark A., and Dennis Meissner. 2005. "More Product, Less Process: Revamping Traditional Archival Processing." *American Archivist* 68(2): 208-263.

Hedstrom, Margaret L., Christopher A. Lee, Judith S. Olson, and Clifford A. Lampe. 2006. ""The Old Version Flickers More": Digital Preservation from the User's Perspective." *American Archivist* 69(1): 159-187.

Heslop, H., Davis S. , Wilson A., An *Approach to the Preservation of Digital Records*, Canberra, 2002. Sections 1 and 4. http://www.naa.gov.au/Images/An-approach-Green-Paper_tcm2-888.pdf (accessed 5/29/09).

Hockx-Yu, Helen, Knight, Gareth, “What to Preserve?: Significant Properties of Digital Objects.” *The International Journal of Digital Curation* 1(3): 141-153.

JISC InSPECT Project, *Framework for the Definition of Significant Properties*, 2008. http://http://www.significantproperties.org.uk/documents/ wp33-propertiesreport-v1.pdf (accessed 6/1/09).

Knight, Gareth, Pennock, Maureen, *Data Without Meaning: Establishing the Significant Properties of Digital Research*, 2008. http://www.bl.uk/ipres2008/presentations_day1/16_Knight.pdf (accessed 6/1/09).

Schellenberg, Theodore R. 1956. *Modern Archives: Principles and Techniques*. Chicago: University of Chicago Press.

Wilson, Andrew, Significant Properties Report, JISC London, 2007. http://www.significantproperties.org.uk/documents/wp22_significant_properties.pdf (accessed 6/1/09).

Wilson, Andrew, Online lecture, University of Washington iSchool LIS 539, May 2009. http://uweoconnect.extn.washington.edu/p38966038/