

**Drexel University**

---

**From the Selected Works of James Gross**

---

April 23, 2012

## Analytical essay: The OAIS model and preservation planning

James Gross, *Drexel University*



Available at: <https://works.bepress.com/jamesgross/32/>

Analytical essay:  
The OAIS model and preservation planning  
James Gross  
Drexel University, Info 756  
Professor Mark Matienzo  
4/23/2012

### Background

The week 4 readings addressed aspects of preservation planning and the role of Significant Properties.

Preservation planning is the core component or function of the OAIS model. It's purpose is to monitor a preservation environment and its surroundings. The goal is to be able to access and used the data without losing unacceptable loss of content or functionality. Preservation plans also address the Significant Properties of digital records, to ensure that the information will have continued access and viability over time. In this short paper I will take a look at the main topics raised by the authors and offer a few opinions about them.

The OAIS model is a reference model design for archival systems within a digital library. This is achieved by planning, developing, and monitoring strategies for the preservation environment of digital information. OAIS is designed with two key goals: (1) to be flexible so that a designated community can adopt it for their unique needs, (2) to offer open access and a non-proprietary interface.

### Main theme:

A main theme of this week's readings was that "one of the [OAIS] key functions is preservation planning which lies at the heart of any preservation endeavor" (Becker, et. al, 2009, p. 135).

### Major topics:

In this week's readings, Ross Harvey introduced the reader to the following concepts: planning data for curation, structuring data for better data management, structuring data for discoverability, and ensuring that the data is curation ready, usable, and accessible (Harvey, 2010). In addition, there were five assigned readings:

[1] The Giaretta article discussed concepts such as Significant Properties, authenticity, provenance, and representation, as they relate to the authenticity of digital objects (Giaretta, et.al., 2009, pp. 67-73). Giaretta argued that the OAIS model does not provide a definition, nor does it fully delineate how Significant Properties should be categorized and tested.

Giaretta explained that Significant Properties is a key concept within the library community but is rarely used in the context of the preservation of research data (Ibid, 2009, p. 67). An example given was the conversion of a word file to PDF without losing data (Ibid, p. 70). Giaretta suggested that the term Significant Properties could be changed to “Transformational Information Property” (Ibid, 2009, p.72).

[2] The Knight article also examined Significant Properties, as they related to the authenticity of data and research (Knight and Pennock, 2009, pp. 159-174). The article introduced the concept of Significant Properties and its role in maintaining the authenticity of research data across changing technological Environments (Ibid, 2009, p. 160). The author noted that, “...work should be performed that maps the significant properties of an information object ...in a consistent manner” (Ibid, 2009, p.172).

[3] The Yeo article, though optional, addressed important concerns relating to the potential loss of information from migrated records (Yeo, 2010, pp. 85-116). Yeo argued that the process of record migration is not a perfect exercise. Yeo quoted Giddens who noted that “structure is both enabling and constraining”(Ibid, p.110).

Yeo suggested that in the absence of loss-free conversion techniques, curators and institutions should try and adapt to the needs of the local user community (Ibid, p.110). And, flexible approaches to record migration or media conversion would help serve the needs of different communities.

Yeo also questioned whether the curator could fully implement data migration due to restricted budgets and finite budgets (Ibid, p. 111). Yeo quoted Lloyd in noting that the “concept of significance” and the process of “assessing significance,” should be recognized as “areas of tension and consternation“ (Ibid, 2010, p. 111).

[4] The Becker article examined the OAIS model, which it described as being theoretical, and supplied a suggested work-flow for its implementation (Becker, et. al., 2009, pp. 133-157). The Becker article concluded that “preservation planning needs to become a systematic and continuous management activity

as opposed to the prevailing ad-hoc decision making” (Ibid, 2009, p. 155).

[5] The Dappert article introduced a conceptual model and vocabulary for preservation guiding documents (Dappert and Farquhar, 2009, pp.119-134). The authors explained how their conceptual model represented, “...an expressive representation of the preservation planning domain” and could be shared and exchanged through software applications (Ibid, 2009, p. 132).

Harvey, in Chapter 9, focused on the conceptualization process of planning the intended creation and capture of data. “Conceptualise,” is the first of the sequential actions of the DCC Curation Lifecycle Model (Harvey, 2010, p. 105). According to Harvey, standards are “essential for reliable digital curation” (2010, p.108). Harvey discussed the process of conceiving and planning the creation of data, including decisions regarding preferred data retention and storage. Harvey noted, “...a focus on user requirements during the conceptualization and planning of a...project...” is important (2010, p.105).

In regard to user needs, identifying the user needs of the “designated community” was also addressed in the Knight article (Knight and Pennock, 2008, p.162). Knight noted curators should be aware of their own requirements, the needs of researchers, and the view of the resource’s creator (Ibid, 2010, p.162).

Curation planning is important to ensure the reusability of the data. Lynch, cited by Harvey, noted three important categories which should be planned at an early stage. These include: [1] ensuring the usage of appropriate metadata, [2] ensuring the provenance of the record is maintained, and [3] ensuring the storage formats are well documented (Harvey, 2010, p.109).

In Chapter 10, Harvey introduced the second stage of the DCC Curation Lifecycle and explained the activities of creating or receiving. These activities included: creating and receiving data, creating data for curation, structuring data for use and reuse, the concept and advantages of open formats and source, the importance of Significant Properties and data authenticity, proper documentation for digital objects, structuring data for management, structuring data for discovery, and receiving data for curation.

In chapter 10, Harvey also examined policies for creating, receiving, and structuring data for multiple uses.

These uses include discoverability, management access, and patron retrieval. As part of the planning process, Harvey noted that policies need to be developed and documented in regard to acceptable formats and size of files (Harvey, 2010, p. 116).

My reaction:

It was apparent from the readings that careful planning, developing, and monitoring strategies for the preservation environment of digital information is important. The main theme of the Chapter 9 and 10 of the Harvey readings were the first and second stages of the DCC Curation Lifecycle.

In reading the Harvey chapters, my initial impression was that the OAIS model and its usage by curators was devoid of any major confusion or problems. It wasn't until I read the assigned articles that I began to slowly see the gap between the seemingly ideal OAIS model and some issues related to its implementation.

For example, some authors, such as Knight and Giaretta, felt that the "Significant Properties" in the OAIS model were vague and in need of clarification. While I understand the importance of authenticity and integrity of digital information, how does the curator decide the priorities for data migration?

Dappert, in his paper, suggested using the term "Significant Characteristics," to refer to data properties. He also introduced a conceptual model and suggested the need to use software applications. Yeo presented questions regarding the ability to migrate data without the loss of some content or functionality. Becker offered details for OAIS implementation guidance.

In reviewing the author's arguments, I found myself in agreement regarding the lack of defined structure within the OAIS model. While a hypothetical data models were useful, there are undoubtedly some curators who will need more guidance to assist in revamping their digital libraries. The assigned readings appeared to primarily be hypothetical modeling as opposed to case studies. I failed to locate statistical data regarding the percentage of data loss from any particular conversion project. Concrete examples of past projects would have been helpful for me to better understand the issues in migration data loss.

5 Suggested questions for review:

[1] In the Dappert and Farquhar article, they stated: "The needs of the target community might be a deciding factor for the choice of preservation actions...the choice of preservation actions will shape and change the community" (2009, p.129). Also see "stakeholder requirements," (Knight, p. 162). In your opinion, which entity (records or community) should have the priority and why?

[2] This week's readings introduced the reader to a number of workflow models. For both the archivist and curator, data loss due to record migration can be an unavoidable reality. Is there a step, process, or procedure which you would initiate to try and address this problem?

[3] Yeo stated that "...migration is driven by expectations that originals (original records) will not remain accessible over time" (2010, p.109). As an archivist or digital curator would you want to expend resources to preserve or retain the original records, or concentrate on retention of the migrated digital records?

[4] The articles by Dappert, Knight, and Becker, introduced work-flow models. Knight stated that, "preservation planning needs to become a systematic...management activity as opposed to the prevailing ad-hoc decision making" (p.155). In my experience, many archives are under-staffed and have large backlogs. Would any of the work-flow models work in your local archive or library, or would you try and simplify (create short-cuts) aspects of it?

[5] Yeo observed that ethical dilemmas exist in regard to curation decisions in determining which records to retain (Yeo, 2010, p.111). Do you (as a future archivist or librarian) agree with Yeo's assertion or not?

In closing, while a useful model, it is apparent that a number of authors would like to see additional clarification and guidance on the use of the OAIS model for data migration. In addition, there appears to be some confusion as to how one defines, identifies, and makes use of "Significant Properties" of a record.

## References:

- Becker, C., et al., (2009). Systematic planning for digital preservation: evaluating potential strategies and building preservation plans.” *International Journal on Digital libraries*, 10, no. 4, pp. 133-157.
- Dappert, A. and Farquhar, A. (2009). “Modelling organizational preservation goals to guide digital preservation,” 4, no. 2, pp.119-134.
- Gareth, K. and Pennock, M. (2009). “Data without meaning: establishing the significant properties of digital research.” *International Journal of Digital Curation*, 4, no.1, pp.159-174.
- Giaretta, D. et al. (2009). “Significant properties, authenticity, provenance, representation information and OAIS information.” *iPRES 2009: Sixth International Conference on Preservation of Digital Objects*, pp.67-73.
- Harvey, R. (2010). “Chapter 9: Designing data,” In *Digital Curation: A how-to-do-it manual*, New York: Neil Schuman, pp 105-114.
- Harvey, R. (2010). “Chapter 10: Creating data,” In *Digital Curation: A how-to-do-it manual*, New York: Neil Schuman, pp 115-130.
- Yeo, G. (2010). “Nothing is the same as something else: Significant properties and notion of identity and originality,” *Archival Science*, 10, pp. 85-116.