Mapping Significance Properties in OAIS: A case study with video games Extended Poster Abstract

Rhiannon S Bettivia

Graduate School of Library and Information Science
University of Illinois, Urbana-Champaign
501 E. Daniel Street, MC-493, Champaign, IL 61820-6211
rbettivi@illinois.edu

EXTENDED ABSTRACT

The term "Significant Properties" is commonly used in the fields of digital preservation and curation. It is variously defined, which puts it in tension with more precisely delineated standards for preservation, such as the broadly used Open Archival Information System reference model (CCSDS, 2012).

This poster proposes mapping properties identified by both creators of video games as significant into the OAIS reference model. This research uses the framework provided by Giaretta et al. (2009), outlined as a clarification about the relationship between OAIS and significant properties, as a starting point. Utilizing data from the Preserving Virtual Worlds II grant and data from interviews with digital preservationists working in libraries and the cultural heritage sectors as case studies, this project aims to provide a real world example of how well user-defined significance maps onto commonly employed preservation models. Data from PVWII suggest that social, surface, and affective attributes of games are considered significant by designers and players. These properties encompass the totality of what really makes a game beyond the code and computing environment, even while they do not appear to fit within the precisely defined categories existent in OAIS. This poster argues that OAIS can be made to accommodate such properties because of its requirement to update archival packages as the base knowledge of the designated community changes over time.

ASIST 2015,November 6-10, 2015, St. Louis, MO, USA.

1. INTRODUCTION

The term 'significant properties' has been used in preservation and curation literature for over a decade. Hedstrom and Lee (2002) define the term as "those properties of digital objects that affect their quality, usability, rendering, and behaviour". It is described variously in many places, and previous papers have summed up the sheer quantity and variety of uses of this term (Giaretta et al, 2009). Such discussions occur across institutions and information types.

Of science data, Sacchi et al (2011) proclaim,

> *"Although this notion has clearly demonstrated its usefulness in cultural heritage domains its application to the preservation of scientific datasets is not as well developed."*

Webb, Pearson, and Koerbin (2013) of Australia's National Library say,

> *"We have come to a tentative conclusion that recognising and taking action to maintain significant properties will be critical, but that the concept can be more of a stumbling block than a starting block, at least in the context of our own institution."*

Remarks like these illustrate the dual-edged difficulty entailed in attempting to use significant properties as metrics for preservation: there is a general consensus that they are important and there are no widely adopted methods by which one can determine what is significant in this sense for preservation purposes. This lack of definition is one part of the difficulty in actually evolving the term 'significant properties' into concrete preservation and curation strategies. General discourse on the topic refers to properties that are most essential to the understandability of digital objects over time. That is to say, significant properties recognize both the situatedness of digital artefacts and the fact that it may not be possible or practical to save every aspect of every object over time.

The other difficulty with this term is that it represents a larger schism within the field of digital preservation between practitioners and designers from computer science

and those who come from archival or library science. Kevin Bradley (2007) presciently said,

"'All God's children got significant properties,' we can sing in unison, but this takes us no further if we cannot define its meaning in such a way that we understand what properties are under consideration, and describe them in a way that is machine-readable and automatically actionable."

This echoes the tension between the social, the human and the technical. Because all of these elements are at play in preservation, particularly when it comes to cultural heritage, significant properties serve as a potential flash point within larger preservation discourses.

1.1 Significant Properties and OAIS

The OAIS reference model has long and wide adoption within the digital preservation community. Further, the terms contained therein have come to function as boundary objects across different types of preservation and curation. As such, mapping significant properties to established entities from OAIS is a promising project for moving concepts of significance into the realm of practice. Giaretta et al (2009) take on this task by firstly examining various uses of the term significant properties and by proposing what more precisely defined terms from within the OAIS reference model might be used in their stead. This is an important project: significance in this way does not appear in earlier version of the model, and even in the more recent changes in 2012, it is mentioned without being sufficiently addressed for some audiences (CCSDS, 2012). Barbara Sierman (2012) compares the most recent version of OAIS with its predecessors and notes:

"The Information Property is related to the commonly known but not always clearly defined term "significant property", but I think more discussion is needed to define better where the differences and similarities between the two concepts lie and how to translate this into the daily practice."

The Information Property in the 2012 revisions is meant to stand instead of significant properties, rather than in place. During interviews the author conducted with OAIS designers, some interviewees noted that they decided to side-step this discussion entirely by creating a separate entity that would serve a distinct set of functions. The key is that information properties are meant to work in conjunction with other existing features in OAIS, in lieu of actually defining the significant property concept.

Giaretta et al (2009) point to the tensions between the design and origins of OAIS and the daily experiences of practitioners when they say:

"Clearly the uses of Significant Properties of necessity focus on those aspects of digital objects which can be evaluated in some way and checked as to whether they have been preserved. However, the meaning associated with a value of the Significant Property is nowhere defined. Therefore it must be the case that the Significant Properties, while useful, do not strictly contribute to understandability of the Information Object."

Those authors propose a framework pointing to existing OAIS entities as containers for the types of data that might be considered significant. A particularly important part of this move is the emphasis on the designated community, as authenticity does not exist in a vacuum but is instead a product of the relationship between an end-user and the data they might receive from an archive.

Here, '*an* end user' is deployed purposefully to refer to the broadest set of possible users, because the term 'designated community' refers to a specific audience rather than general users. Within OAIS, the designated community is,

"An identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities. A Designated Community is defined by the Archive and this definition may change over time." (CCSDS, 2012, page 1-11)

This distinction is particularly pertinent for institutions such as libraries, whose users are a vast group. In such a case, the designated community is necessarily artificially constructed in order to scale preservation practices so they are manageable within the resources of the institution. OAIS does not say how broad or narrow a designated community must be. In interviews with OAIS designers conducted by the author, the interviewees describe the necessity of building a system for a constructed user rather than all the users of a public library, for example, because information packages for the latter would need to [improbably or impossibly] include an entire educational system along with the content to be preserved.

Work that does not address the designated community cannot address the significant properties elephant. In dealing with a concept like significance, it becomes necessary to ask *significance for whom*, something that is often implied but not always specifically addressed in discussions of significant properties. Geoffrey Yeo (2012) sums this up eloquently:

"However, the determination of 'significant properties' is no less problematical than the debate about notions of value '…not least because different user communities will bring different perceptions of what constitutes significance."

This poster will marry existing data about significant properties with the framework set out by Giaretta et al, to see how well user-designated significance fits within the entities for representation information, provenance, and authenticity with a focus on the notion of the designated community. This work also allows for an exploration of the concern laid out by Sierman above in examining the

extent to which transformational information properties can serve as a substitute for significant properties.

PRESERVING VIRTUAL WORLDS II

This work will utilize data gathered as part of the Preserving Virtual Worlds II grant as a sample data set. PVWII, which was funded by IMLS and concluded in 2013, examined the concept of significant properties as it applies to video games with the aim of informing preservation practices for complex media, building on previous projects that examined the significant properties of software (Matthews et al, 2008; Knight, 2008). Phase 1 of PVWII entailed a two-fold method for examining significance. Investigators performed content analysis of a number of video games series that spanned a time frame from 1981 to the present and ranged across different game genres. Simultaneously, investigators conducted interviews with people involved in the design and dissemination of games from the case set; with designers working in other game design studios; and with fans and programmers who have worked on more well-known modifications (mods) of some of the games from the case set. These interviews were coded and analyzed by members of the research team across the various institutions involved in the project.

Phase 2 of PVWII focused on the development of tools and metrics to assist in the preservation of the significant properties identified from the research in phase 1, including an examination of how such properties could inform decisions about the emulation, migration, and re-implementation of games as well as defining benchmarks for authenticity in playback. The phase 1 research painted a very complex picture of significance within the realm of games, and a key finding was, unsurprisingly, that significance is highly situated. The research data indicated that what is significant about games may not be something inherent to the game's code (bits) or even computing environment (platform, operating systems, controls), but could include elements as varied as underlying data models or general surface affective experiences. As such, PVWII suggested a layered model for looking at games, delineating different aspects of each system wherein different users might locate significance.

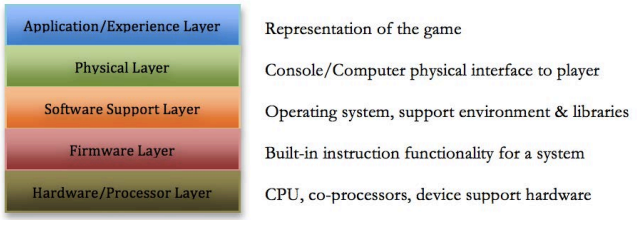

Figure 1: PVWII layers of a game system

A second deliverable was the creation of a survey tool which borrowed from earlier projects such as the Variable Media Questionnaire (Ippolito, 2003) that aims to capture significant properties as defined by various stakeholders, including designers, players, archivists/preservationists, and curators. This tool is designed as a wiki for the purpose of collecting and automating the analysis of large quantities of data that will serve as record of the knowledge base of different user communities. In combination with game-specific contributions to format registries, another PVWII phase 2 goal, the hope is that preservation of games can be enhanced by crowd participation in the process gathering and centralizing previously dispersed but necessary information about games.

CURRENT PROJECT OUTLINE

This poster seeks to operationalize the work done by Giaretta et al (2009) using PVW II interview data as a case study for mapping what participants deemed significant into the OAIS entities identified as more precisely defined alternatives to the rather nebulous term 'significant properties'. This poster will focus on interview data related to two games franchises from the PVWII case set: *Carmen Sandiego* and *Civilization*. For both games, multiple creators were interviewed, painting a broad and varied picture of significance as determined by creators. In order to determine how well this data can be captured by the high-level entities detailed in the OAIS documentation and further explicated by Giaretta et al (2009), it will be parsed to look specifically for information that could be modeled as representation information, especially for the documentation of provenance and to act as benchmarks for authenticity, and what can be modeled as transformational information properties.

In keeping with the requirements of OAIS, a designated community will be defined based on self-descriptive information derived from the interviews with the broader notion that, for this case study, the designated community is comprised solely of the interview participants. While this designated community is therefore quite limited, the process of defining this community in relation to notions of significance and thus bounding and describing a general knowledge base for this community should prove a useful example for those working in large generalist organizations such as libraries, where defining a particular designated community can at times be difficult when viewed from the perspective of the larger population of users such institutions wish to or are mandated to serve.

Finally, it is the long-term aim of this investigation to map any significant properties that do not easily fit within the representation information of particular digital objects into separate, discrete information packages designed to document the knowledge base of the designated community. While OAIS requires information packages to change with time as knowledge bases change, few of the many implementation studies available on the application of OAIS in particular organizations touch on the methods or strategies to document this base or understand when and how it changes. As such, this aspect of the poster may prove most useful to the preservation community.

CONCLUSIONS

This project is in its very early stages, so any conclusions at this juncture are tentative. The final aim of the project as previously described already presupposes, however, that it may be difficult to sufficiently map all the significant properties noted in the interview data onto existing entities used widely in many preservation repositories today. This study examines a difficulty in the day-to-day deployment of earlier frameworks, whose basic constructions often insufficiently account for significant properties that are not inherent to the digital object itself, such as those that are not intrinsic to the code yet essential still to some kind of long term understanding. These elements are drawn from interview data with game creators and enthusiasts, as well as from interviews with those engaged in the preservation of cultural heritage audiovisual materials using OAIS. In recent years, the dominant discourse in preservation has shifted from being about migration to considering alternatives such as universal virtual computers (UVCs) or, more commonly, emulation. This is reflected in the changes made to OAIS in 2012 as well as in interview data with OAIS designers. This does in some measure move the preservation community towards an acceptance that things beyond the object themselves are significant and require preservation – in the case of emulation, significance is found in the behaviors of the original computing environment. The work here looks to extend this by encompassing even broader data about significance that may encompass aspects of the social and cultural, aspects of the designated community's knowledge base, as significant properties of the objects that necessarily must be preserved. These data about what really makes a game are not, and should not be, incompatible to the precisely defined categories existent in OAIS because the model requires updating AIPs as the base knowledge of the designated community changes over time. This requirement acknowledges that there is more involved in understanding objects than simply recreating the objects themselves: artefacts are a product of a particular place and time, and are understandable as such. If what is called 'significant' by participants in the PVWII survey and interview data does not map well into the representation information particular to individual artefacts, this does not absolve the archive from the need to treat and preserve this kind of information. This poster aims to provide examples of how archives might begin to address the significant property conundrum using tools and frameworks with which they are already quite familiar.

ACKNOWLEDGMENTS

My thanks to all colleagues involved in the Preserving Virtual Worlds II project, as well as interview participants for that grant and those involved in my work on OAIS.

REFERENCES

[1] Bradley, Kevin (Summer 2007). "Defining Digital Sustainability." Library Trends 56(1): 148-163. Available at: http://muse.jhu.edu/journals/lib/summary/v056/56.1bradley.html

[2] Consultative Committee on Space Data Systems (June 2012) *Reference Model for an Open Archival Information System (OAIS).* CCSDS 650.0.M-2, Magenta Book. Issue 2.

[3] Giaretta, D., Matthews, B., Bicarregui, J., Lambert, S., Guercio, M., Michetti, G., and Sawyer, D. 2009. Significant Properties, Authenticity, Provenance, Representation Information and OAIS Information. In *Proceedings of iPres 2009: the Sixth Internationl Conference on Preservation of Digital Objects.* 67-73. Permalink= http://escholarship.org/uc/item/0wf3j9cw

[4] Hedstrom, M., and Lee, C. "Significant properties of digital objects: definitions, applications, implications", Proceedings of the DLM-Forum 2002. Retrieved from http://ec.europa.eu/transparency/archival_policy/dlm_forum/doc/dlm-proceed2002.pdf

[5] Ippolito, J. 2003. Accommodating the Unpredictable: The Variable Media Questionnaire. In Alain Depocas, Jon Ippolito and Caitlin Jones (Eds.) The Variable Media Approach: Permanence Through Change, pp. 47-55. Available at: http://www.variablemedia.net/pdf/Permanence.pdf

[6] Knight, G., 2008, Framework for the definition of significant properties. Retrieved from http://www.significantproperties.org.uk/documents/wp33-propertiesreport-v1.pdf

[7] Matthews, B., McIlwrath, B., Giaretta, D., Conway, E.. The Significant Properties of Software: A Study. JISC report, March 2008

[8] Sacchi, S., Wickett, K., Renear, A., and Dubin, D. 2011. A Framework for Applying the Concept of Significant Properties to Datasets. In *Proceedings of ASIS&T 2011.*

[9] Sierman, B. (2012) "OAIS 2012 Update" in Digital Preservation Seeds Blog accessible at http://digitalpreservation.nl/seeds/oais-2012-update/

[10] Webb, C., Pearson, D., and Koerbin, P. 2013. 'Oh, you wanted us to preserve that?!' Statements of Preservation Intent for the National Library of Australia's Digital Collections. *D-Lib Magazine 19:1/2* DOI= doi:10.1045/january2013-webb

[11] Yeo, Geoffrey. 2010. "'Nothing is the same as something else': significant properties and notions of identity and originality." *Archival Science* 10, no. 2: 85-116.