

# Mapping Significance of Video Games in OAIS

Rhiannon S. Bettivia  
School of Information Science  
University of Illinois at Urbana-Champaign  
rbettivi@illinois.edu

## ABSTRACT

In this paper, I explore the concept of *significant properties* and how such properties do and do not fit within the Open Archival Information System (OAIS) Reference Model. Combining interview data from research about the deployment of OAIS in cultural heritage institutions with data about video game production and preservation from the Preserving Virtual Worlds II (PVWII) grant project, this paper maps stakeholder-identified significant properties onto the 2012 version of OAIS [4]. Significant properties have many definitions and even many names. Operationalizing this term broadly, many such properties do fit within existing OAIS entities. However, significant properties that are relational and external to digital objects' code and environments do not. This paper concludes that additional metrics are needed to begin shaping the process of documenting significant properties at scale.

## Keywords

Significant Properties; OAIS; Digital Preservation.

## 1. INTRODUCTION

I explore the concept of *significant properties* and how these do and do not fit within the Open Archival Information System (OAIS) reference model. *Significant properties* is not an OAIS-specific term. Operationalizing this term represents a point of tension between the various disciplines brought together to construct the sub-discipline and profession of digital preservation. Significant properties are important because they refer to some kind of information without which digital artifacts are unintelligible, even if the artifacts remain functional. Significance is determined by a variety of stakeholders [7], and digital repositories will not be able to engender the trust of users if they cannot communicate back those elements about a digital object that consumers find most important. Addressing significant properties poses a challenge because they are undefined, or perhaps over-defined: even the term itself is under dispute and there are a variety of alternatives that implicate significance, from Information Properties to significant characteristics to context information. This paper adds to the ongoing conversation by identifying significance of digital objects as defined by practitioners and content producers. By beginning with what is described by these stakeholders as essential in particular case studies, this paper attempts to operationalize the concept of significance rather than weighing in on the various definitions. In this way, this research is productive of new possibilities: it is not simply a criticism of existing information structures, but aims away from silo-ing this discussion according to institution or discipline type towards more macro understandings that can inform the creation of metrics to guide processes of documenting significance at scale.

I describe digital preservation as meta sub-discipline of the meta-discipline of information science [1]: in the same way that information science is imbricated across the traditional research disciplinary spaces of humanities, social sciences, and natural sciences/mathematics, thus incorporating and informing all of these areas, so too is digital preservation a meta-sub-discipline of information science: preservation work is part and parcel of all the work information professionals do, and so borrows terms and practices from all areas of the information professions.

Significant properties stem from library and archival traditions, yet need to be rendered functional in the broader space of digital preservation, and this poses a challenge that is expressed by the myriad definitions, readings, and projects that reject significant properties as unimportant or untenable. Webb, Pearson, and Koerbin [32] of Australia's National Library sum up this ethos within the general realm of digital preservation:

*"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."*

This simultaneous acknowledgement of the critical yet poorly understood nature of significant properties demonstrates both the importance of the term, but also the barriers to its productive impact given a lack of definitional clarity: significant properties have become an elephant in the room for digital preservation. I argue that one method of synthesizing these various definitions is to engage with how this term is used in practice.

This paper marries data from semi-structured interviews about the deployment of OAIS within memory institutions with interview data collected during the Preserving Virtual Worlds II (PVWII) grant project. The OAIS interviews cover a range of digital preservation scholars, practitioners, and OAIS authors, revealing insight into how 'insiders' perceive the significance of digital objects. The PVWII data explicitly examine significance as described by content producers, in this case programmers and others working on the creation of digital games and virtual worlds. I examine these data alongside the Transformational Information Properties proposed as an alternative to significant properties in the 2012 version of OAIS [4][13][30], to see how well user-described significance fits within the entities for Representation Information, Provenance, and within the OAIS conception of authenticity. This paper examines how complicated multi-part works like video games, virtual worlds, and other dynamic popular culture materials fit within OAIS. I work with OAIS given its ubiquity in the field. I argue through these data that some significant properties fit within the entities of the OAIS reference model, particularly those related the digital object itself and the software/hardware environments required to make an object functional. However, I also argue that OAIS, as currently scripted, cannot encapsulate all the types of significant properties derived from the interview data. The places where these mismatches occur are places wherein other preservation practitioners and scholars have identified weaknesses in the model related to the changing landscape of digital content towards more distributed models. By deriving importance and productive definitions of significant properties from practitioners, I locate significance in relation to the digital object and identify the types of significance not currently covered by prominent models and advocate for new guidelines that incorporate these.

## 2. LITERATURE REVIEW

Webb, Pearson, and Koerbin [32] sum up the consensus that significant properties are important, yet difficult to employ for preservation purposes: their description of significant properties

as a ‘stumbling block’ indicates that previous attempts to clarify this term and provide methods by which to make it operational have not been widely adopted. The lack of a simple and widely accepted definition is one 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 term *significant properties* has been used in digital preservation and curation literature for over a decade. The most commonly referenced definition, and also an early one compared to others I reference here, is the one by Hedstrom and Lee [14], who define the term as “those properties of digital objects that affect their quality, usability, rendering, and behaviour”. *Significant Properties* are described variously in many places, and Giaretta et al [13] and Dappert and Farquhar [7] discuss the difficulty in settling on a single definition by exploring some of the myriad definitions that currently exist in disciplinary literature. These definitions stem from various sources, across institutions, information types, and research disciplines. Of science data, for example, Sacchi et al [26] say:

*“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.”*

What precisely is meant by “demonstrated usefulness” is not entirely clear, as many practitioners in cultural heritage acknowledge the use-value of this notion without being able to advance either a concrete definition of what it means or how to account for it formulaically or machine-readably.

Within the interview data that I present in this paper, definitions of significant properties were similarly varied. One participant from my OAIS interviews, a manager of digital preservation at a European national library, suggested that libraries are well equipped to deal with significant properties, “because...as a library we have a lot of experience in describing things so we are very good at metadata”. This quote suggests that she perceives a relationship between descriptive metadata and significant properties. Demonstrating the salience of findings about the occasionally contradictory nature of various definitions of significant properties, another OAIS interview participant, a research and development officer at a European national archive, said “well, [the term *significant properties* refers to] just technical metadata, [doesn’t] it?”

The other difficulty with this term is that it represents a larger schism within the field of digital preservation between practitioners from computer science and those who come from archival or library science. Bradley [2] 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 encapsulates 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 the preservation of cultural heritage and popular culture materials, significant properties serve as a potential flash point within larger preservation discourses that arise around OAIS and the growth of the field of digital preservation.

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 endeavors [22]. Giaretta et al [13] examined the relationship of significant properties to existing entities in preparation for the 2012 revisions to OAIS. The authors proposed a number of existing, and thus more precisely or homogeneously defined, terms from within the OAIS reference model to act as an alternative to proposing a new definition for significant properties or reconciling existing ones. They also proposing the Information Property as an alternative. The Information Property and the resultant emphasis on authenticity relies heavily on the *Designated Community* term within OAIS, as authenticity does not exist in a vacuum but is instead a product of the relationship between a potential end-user and the data they might receive from an OAIS. This echoes work by scholars like Dappert and Farquhar [7] who posit that significance is not inherent to objects but determined by stakeholders. The term *Designated Community* is ‘weakly defined’ in OAIS, according to an interview subject, in the sense that the model does not concretely detail how to form and document such a community. While such specificity is not necessarily within the purview of a reference model, the missing piece with digital preservation practice is that standards subsequent to the reference model have not yet been developed, and that many institutions have not, at a site-specific level, formally defined their Designated Communities [17] [3].

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. Yeo [33] 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.”*

The situated nature of the Designated Community and the idea of ‘preservation for someone’ arise from the same discourses of place and time that inform conversations about significant properties. Struggles I identify here are due in part to changes in technological landscape the importance of which authors of OAIS were not able to predict. This is not new: for example, earlier versions of OAIS assumed migration to be the default preservation method, yet recent years have seen a shift away from migration and normalization towards a more mainstream acceptance of emulation and the importance of computing environments, particularly in reference to complex media like video games [15][23][9][8][6]. The 2012 OAIS revisions encapsulated this change. Recent developments in areas like linked data and other forms of distributed content pose a challenge to the current iteration of the OAIS reference model, and practitioners like David Rosenthal [25] have made calls for attention to this as OAIS heads into a new round of revisions in 2017.

The 2012 changes to OAIS resulted importantly in the definition of the Transformational Information Property, which does some work to capture significance in relation to stakeholders [4]. Sierman [30] 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 Transformational Information Property in the 2012 revisions of OAIS is meant to stand instead of significant

properties, rather than in place. During my interviews with OAIS authors, some noted that they decided to side-step the *significant properties* discussion entirely by creating a separate entity that would serve a distinct set of functions partly because of the sheer number of incommensurate existing definitions for significant properties. The key is that Transformational Information Properties are meant to work in conjunction with other existing features in OAIS, in lieu of actually defining significant properties, thus avoiding the need for authors and OAIS as a sociotechnical network to engage within this space. In practice, the outcomes are not so neat: by choosing not to wade into the significant properties debate, the OAIS authors are taking an effective stand indicating that the concept does not need to be incorporated within the major standard of the field: significant properties are not useful or important enough. This dictates in part how well significant properties can be taken up by others given the pervasiveness of OAIS and the ways in which practitioners in interviews struggled to envision alternative frameworks for their preservation work. Additionally, the solution conceived of by the authors to avoid the term has not stopped practitioners within the profession from continuing to call for OAIS to deal with significant properties more explicitly.

The treatment of significant properties within the literature is reflective of current discourses in digital preservation practice. As such, suggested models or practices fit squarely within existing models like OAIS and address property/value pairings in relation to aspects of digital objects that are better understood within the general field of digital preservation. This means that earlier literature focused on aspects of digital objects like semantic representation and functional bits; more current work incorporates the environment of the object as well. What this reveals is an additional difficulty in developing means of documenting significance at scale: theoretical approaches focus on significance of the features of digital objects that are prominent in the moment. If, as I will argue here, significant properties of digital objects are located elsewhere, then the current theoretical approaches will not be able to sufficiently account for significance.

### 3. METHODS

#### 3.1 Data Collection

This paper utilizes two sets of data that capture different views on significance from important stakeholders in digital preservation. The first is comprised of semi-structured interviews conducted with a variety of preservation practitioners, scholars, and OAIS authors. These interviews were conducted in Europe and North America as part of a research project that investigated the effects of OAIS on values and professional practice in cultural heritage institutions. Interviewees included 28 participants from 5 countries. These participants included digital preservation specialists who practice or research in public and private universities; public and national libraries; national and private archives; museums; and consulting firms. Also included were authors of OAIS and data curation scholars working predominantly in the sciences. Within the practitioner interviews, participants had a range of specialties and areas of expertise, including technology officers, research and development administrators, as well as some analog archivists and librarians who had little to no knowledge of OAIS despite working within institutions or departments that are heavily influenced by OAIS. In conjunction with the interview data, this dataset includes a variety of documents such as the various versions of OAIS itself as well as a number of procedural and policy documents submitted to me by interview participants. These interviews were qualitatively coded for characterizations of OAIS; discussions of particular OAIS terms; and descriptions of what is well-enabled by OAIS as well as what is missing or constrained.

The second dataset was gathered as part of the Preserving Virtual Worlds II grant. PVWII was funded by the Institute of Museum and Library Services (IMLS) and concluded in 2013. It included investigators from the University of Illinois, the University of Maryland, Rochester Institute of Technology, and Stanford University. Investigators 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 and a previous game preservation project, Preserving Virtual Worlds I (PVWI) [19][16][21]. Broken into two investigative phases, Phase 1 entailed a two-fold method for examining significance. Investigators performed technical and content analyses of a set of video game series. 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 qualitatively coded and analyzed by members of the research team across the various institutions involved in the grant 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. These included 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. PVWII suggested a layered model for looking at games, delineating different aspects of each system wherein different users might locate significance. I will discuss this model in greater detail later in this paper.

For this paper, I coded both datasets using NVivo software. There were three overarching nodes: explicit mentions of significance; implicit mentions of significance where participants mentioned terms identical or similar to those that appear in the myriad definitions of the term; and things that were explicitly defined by participants as not significant. Within these first two nodes, responses were further categorized according where significance was located within the layered model mentioned above as well as within existing OAIS entities according to Giaretta et al [13]. The last node recognizes that an equally important part of creating adequate preservation information packages is determining what information should not be saved, and this echoes on-going discussions in the realm of science data curation and media art preservation.

#### 3.2 Research Questions and Process

The research question for this paper is: given the ubiquity of OAIS, how do complicated multi-part works like video games, virtual worlds, and other dynamic popular culture materials fit within the model? I investigate this by allocating significant properties to existing OAIS entities and identifying those which do not fit within the model. I began the process with the hypothesis that all would fit despite the fact that video games and other complex digital objects pose a challenge to digital preservationists for two reasons. First, as mentioned above, the large and general category of significant properties is one that OAIS intentionally avoids. Second, while the term *Archive* in OAIS is very specific, it shares some foundational tenets with the study of traditional paper archiving practice and it is not the job of the traditional archive to collect or preserve external significant properties, those not contained within code or computing environment. Several interview participants expressed views about the traditional role of archives that indicated archives should not collect external significant properties. Within traditional archival practice, the term *selection* is used very narrowly: the scope and collection statement determine the type of content to be collected, and all such content

from the organization is archived rather than an archivist selecting certain materials for processing and preservation in a more colloquial sense [10][11]. Creating Information Packages for complex media requires some measure of this latter selection: the Archive must choose a set of things to include in the package that encompasses the most significant properties. The purposeful and transparent creation of artificial boundaries is at odds with foundations of archival practice which inform the authors of OAIS and how OAIS gets deployed. This is the second difficulty that arises when mapping video games into OAIS.

I 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, I parse the data to look specifically for information that could be modeled as Representation Information, especially for the documentation of Provenance; to act as benchmarks for authenticity; and what can be modeled as Transformational Information Properties.

Finally, I identify significant properties that do not easily fit within the Representation Information of particular digital objects and discuss why it is that these do not work within the current iteration of OAIS. Some of these properties are related to the tricky OAIS term Knowledge Base; others are distributed in a manner that challenges OAIS's requirement for adequate control of the content.

## 4. FINDINGS

The current interest in emulation as a preservation method 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 and this has been recognized in a number of research endeavors including some that specifically examine significance [7][9][6][12]. PVWII research painted a very complex picture of significance within the realm of games. 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. I argue for the consideration of even broader data about significance that may encompass social and cultural aspects and elements of the Designated Community's Knowledge Base. These terms within OAIS acknowledge 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. Singling an object out as divorced from its spatial and temporal context will not guarantee the understandability of the object over time, even if its rendering environment and bits are preserved: a digital library director at a private US university summed it up nicely saying, "I mean, files are not that useful without something." That is to say, we need something beyond even working files themselves.

The situation (that determines the situatedness) of an object needs to be preserved. I argue that this is what is encapsulated in OAIS by the terms *Knowledge Base* of the *Designated Community* within the sociotechnical complex of OAIS, even if the explicit definitions in the OAIS documentation do not indicate this. OAIS requires information packages to change with the Knowledge Base of the Designated Community [4]. The often described example of what this looks like in practice is a shift in the dominant language (Knowledge Base) of the

Designated Community that requires additional translational assistance in archival packages where none was needed before (altered IPs). Another example that stems from the PVWII data is the change in geography over time: Carmen Sandiego games involve chasing 'bad guys' across various geographic locations. The Knowledge Base of the 1985 game player contains the USSR rather than the de-federated former Soviet nations contemporary to the writing of this paper. Maps make for easy pieces of Representation Information to store along with the digital object, all under the OAIS entity *Content Information*.

But when a digital object like a game is seen as imbricated in a complex and ever-changing sociotechnical network, then there are subtler changes to its understandability that are more difficult to document than a dictionary or a map. In the content analysis data of games from PVWII, several Carmen games depict South East Asian countries by employing images of people in conical hats working in rice fields. Today, 31 years after the release of the first *Carmen*, this image still allows game players to identify a certain part of the world, but this knowledge will change rapidly. Water politics and rising sea levels associated with global warming trends mean that large swathes of the Mekong River delta, known as one of the top rice producing and exporting areas in the world, are at risk of being flooded with salt water from the sea. These climate changes threaten to end the farming of rice in these areas: if these trends continue unabated, within a few decades this region will no longer be the center of the rice growing industry. With it will go the cultural association of people in conical hats bent over rice fields as production shifts to Africa, where popular imaginaries suggest different visual markers to note time, place, and occupation. At this point, parts of these video games that rely on tacit knowledge that recognizes images of conical hats and non-descript green fields (this non-description being due largely to technological limitations at the time these games were produced) means that the games can no longer be played: the very behaviors of the digital object break down without enough understanding about the contemporary Knowledge Base of original intended users. And so this situatedness, I would argue, is a significant property in the sense that, without this kind of information, the game is not playable over time even if the bits and computing and rendering environment are preserved. A current presumption of game preservation is that a game, by its nature, is meant to be played, so if it cannot be played, we cannot be said to have preserved a working copy [20].

### 4.1 PVWII Interviews

Significant properties, as identified in interviews from PVWII, could be located at any point in the layered model developed as part of the grant project.

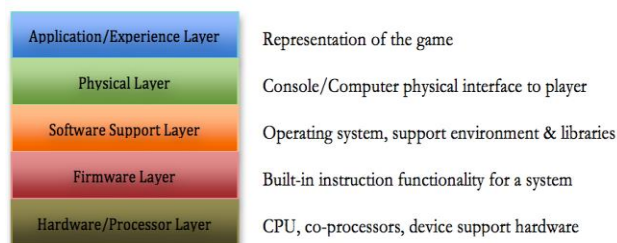


Figure 1: PVWII Layers of a Game

For example, some video games were designed around specific *software support layers*, layer three on the stack, such as the first *Civilization* game designed to work with early Windows operating systems. The functions of the then-novel eponymous windows were incorporated heavily into the game, and constituted a significant property to the developer we spoke with, who mentioned the role this operating system played in the

game's development. The Nintendo game *Duck Hunt* notoriously uses a special peripheral *hardware* piece, layer five on the stack. A light gun (as opposed to the normally used d-pad and four-button controller) allowed players to shoot at ducks, as the name of the game implies, and the game is not functional without this piece of equipment. The light gun only functions in conjunction with a cathode ray tube (CRT) television. A CRT monitor might be considered *hardware* or might be considered part of the *physical layer*, layer two in the stack, as part of the physical interface for the player. These twin external hardware dependencies, both of which are essential to a functioning version of *Duck Hunt*, might be considered significant by some Designated Communities.

Yet most significant properties identified by the interview participants in PVWII fell unambiguously under the top, *application* layer of the stack, which is the representation of the game. As a result, I divide the significant properties in this part of the data into three categories, according to where they can be located in relation to the layered model: two of these lie within the top layer of the stack and the third lies outside the stack altogether.

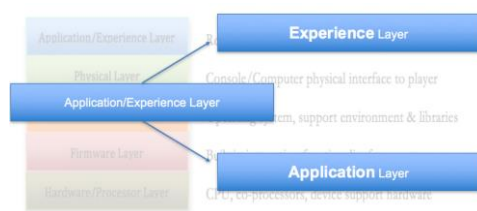


Figure 2. Application and Experience Layers

I firstly break the Application/Experience Layer into two parts: Application and Experience. These encompass many of the significant properties identified by PVWII participants. The application layer includes things like the game code itself, as well as items like jump tables for early Mario Brothers games or historical statistical mortality data that determined how likely a player was to die when playing Oregon Trail. The experiential layer encompasses the surface and affective experiences of playing the game: the fact that Carmen Sandiego is only kind of a one-player game despite its single avatar because of its situation in public schools, for example. I separate Application and Experience because I argue that they are not necessarily related. To be sure, the original code and a computing environment were necessary to manifest the original playing experience. But to recreate the mnemonic experience [31], to give an authentic representation of the experience of play, the original code is no longer necessary. PVWII investigators posed the question to game designers: how important is the original code if you can generate the same surface appearances and behaviors with a different backend? Most responded that they were not wedded to the original code, but more so to the experience of play. Some noted that the original code itself was ‘poor’, often due to time constraints. These two things can exist separately: because it is possible to save 1s and 0s and even consoles and media without saving the experience and it is possible to recreate the experience without the 1s and 0s, I separate this layer into two discrete layers.

Finally, I also argue that some kinds of significance, as described by PVWII participants, lie outside the stack altogether—that is to say, *external significant properties* cannot be found in the code or environment. These include significant properties like those I term *relationally significant*. PVWII investigators asked participants to name their favorite game franchise and to explain what made it so important. One point of significance that was mentioned was a game that was ‘leaps and bounds’ ahead of its predecessors and contemporaries. I term this

*relationally significant* because understanding this statement about what makes a game important requires placing it in context, almost like archival bond, with other games of its time. To understand this property of a game does not require a playable copy, although it might entail placing a playable copy up against playable copies of its contemporaries to demonstrate its advancements. But there are other ways to represent and benchmark this: for example, placing visuals from saved game files, videos of play, or machinima in relation to visuals of its contemporaries. The constant feature here is that is impossible to understand the “advancedness” of a game by looking at the game itself: it has to be seen in relation to other things.

PVWII interviews also raised other external significant properties of game play. For Carmen Sandiego and Oregon Trail, two franchises that are often termed *edutainment* games to the chagrin of their producers, interviewees expressed that understanding them in the educational context of the mid-1980s is important to understanding the experience of play. Like the tangible difference between playing the original arcade *Donkey Kong* and playing a game on a home console (one designed to be short to eat money, the other designed to be long to engender customer loyalty to a product), there is a tangible difference between playing Carmen Sandiego solo at home today with Google, versus playing it in its original environment: several kids around a tiny and expensive computer in a school, with one person at the keyboard and another working the accompanying encyclopedia. One interview participant who worked on programming for Carmen Sandiego said that seeing these games in context was how he envisioned ideal preservation for his games, while acknowledging the difficulty in manifesting something like the mnemonic impressions of a particular time and place.

## 4.2 OAIS Interviews

While coding the interviews with OAIS practitioners, authors, and scholars, there were only three instances in which the specific term *significant properties* was brought up by my interviewees, and this makes sense given that the dominant theme of these interviews was OAIS and the fact that significant properties is not an OAIS term. The explicit significant property instances echo the dominance of the OAIS authors in shaping how people within the realm of digital preservation continue to respond to and understand their work in relation to OAIS. In one instance, I asked an interview participant about significant properties specifically because I knew this participant had made public statements about them in relation to OAIS in the past. In this particular discussion, the interviewee mentioned significant properties in relation to enrolling analog professionals within libraries in digital work. The interview subject said:

“...we have more analog material... and lots of people were trained to deal with analog material and fewer people are trained to deal with digital material. And as you can't just give them the sack [laughs], you need to deal with them, you train them or whatever, so that takes a long time and I think that's one of the problems all main libraries are dealing now with that they have staff that's not quite prepared for digital material. So that the thinking about OAIS starts within a... small group of people... and we tell them 'I think you should interpret it like this or like that' and what you don't see is that they try to translate it to their analog environment and sometimes that does not work because it's digital. So it's difficult to translate I think, although the model itself is very clear, I think it's rather straight forward, but when you go the significant properties, well, endless discussions.”

This interview participant, someone who is both a library practitioner and actively involved in OAIS revisions and related

standards, describes OAIS as “relatively simple”. In this case, the designation of simplicity is meant, as much as anything, to indicate how not simple the concept of *significant properties* is. The situation in which she is working is already a fraught one to some extent: the library has a large analog collection and many analog employees, and moving into the digital space requires people to learn new skills. And it is under this umbrella discussion about employees who work with analog materials, who cannot make analogies between their previous work and their digital futures, and who struggle with a “simple” model often because they “only read the first 80 pages [of OAIS]” according to the same interviewee, that the subject of significant properties arises. As suggested elsewhere, this notion comes from library and archive traditions, and therefore clashes with data and systems design origins that dominate the construction of OAIS. This is the unresolved tension a reviewer noted in response to an article I submitted on the subject to a major preservation-oriented conference. And perhaps it is the perception by OAIS authors that significant properties come from libraries and archives that predicates its continued exclusion from OAIS.

A second mention of significant properties in relation to OAIS came from a US-based data scientist who said:

*“I mean, if OAIS didn't exist, you know, people would still need to preserve things and they would come up with some other framework, and obviously it would be not exactly the same as OAIS. It would probably have a lot of the same ideas in it. There were, you know, obviously... concepts that I used before I ever saw OAIS, but when I saw it, I thought, “Oh, yeah, this maps to this in OAIS.” And OAIS has concepts in it from earlier versions of OAIS that aren't the same anymore like format migration isn't called format migration anymore, it's called transformation. And significant properties are now like transformational information property, you know, and things like that.”*

This suggests a familiarity with the process of OAIS creation and revisions, such that this person is aware of the fact that Transformational Information Properties are the official term meant to deal with significant properties. This interview subject speaks from a place of privilege: as a science data scholar, this person was already familiar with the type of terminology that is contained within OAIS, and is happily fluent in its lingua franca. In fact, of all my participants, this one had the fewest complaints about OAIS, expressing most answers in form similar to the quote above.

The comment by US-based data scientist about the relationship between significant properties and Transformational Information Properties is a common misconception, if it can be called that. It may simply be a casual simplification. While Transformational Information Properties are meant to encompass some aspects of significant properties, they are not a replacement. Defined in the 2012 revisions [4] as an:

*“[i]nformation [p]roperty the preservation of the value of which is regarded as being necessary but not sufficient to verify that any Non-Reversible Transformation has adequately preserved information content. This could be important as contributing to evidence about Authenticity. Such an Information Property is dependent upon specific Representation Information, including Semantic Information, to denote how it is encoded and what it means. (The term ‘significant property’, which has various definitions in the literature, is sometimes used in a way that is consistent with its being a Transformational Information Property).”*

It is stated quite clearly that this definition is meant to cover only some definitions of significant properties. Depending on the definition of significant properties one employs from among the

myriad ones in existence, some of these properties are contained within entities that predate the 2012 revisions, including within the Digital Object itself as well as in places like the Preservation Description Information entity, without necessary reference to a Non-Reversible Transformation.

These are two distinctly interesting explicit mentions of significant properties from the OAIS interview data. The more populous node, however, was *implicit significant properties*. I applied this label to any discussions wherein an interviewee mentioned some aspect of a digital object without which that object would not be understandable, functional, authentic, or worth preserving; in other words, specific values labeled by the participants with any of the descriptors from the myriad definitions of significant properties at the outset of the paper. The findings from this node within the OAIS data include a number of references that echo the PVWII data. One practitioner mentioned a concern about the dependency on outside objects for understandability, in particular external technologies. This US-based museum practitioner also said:

*“Yeah, like Windows '95, we need a place to track that information and because there is a many-to-many relationship there, it makes sense to record that in a structured way where we have some kind of master record of all these technologies.”*

This comment was in reference to the difficulty of creating mutable AIPs within the software programs the institution uses for documenting art records. The substance of the comment mirrors discussions with video game creators who referenced the significance of the role of the operating system, coincidentally also Windows '95, in the creation of a title within an iconic video game franchise.

Likewise, the experiential aspects of digital objects also arose in the OAIS interviews. One participant, a digital preservation manager at a private US university, said:

*“...Maybe we need to be more clear about it's not just about providing [access] to the files, it's about providing an experience... I mean, I like to think about it as being able to present the same content to the user...we could have documented that content, regardless of the experience through which they receive that content, even if the content is an experience... I don't know. It's complicated...And I also feel like... just in general... there's so much interaction, and the experience of being able to work and build, something like that.”*

This is not to say that engaging with analog materials is not experiential: indeed, reading a paper book is an experience, and reading a Dickens novel as a set of serialized chapters over the course of months is not quite the same experience as reading the entire work at once when it has been collected into a single volume. But in this case, the interviewee is expressing something fundamental about the interactivity of many types of digital content. I take this ethos to be the same one that motivates the response on the part of video game programmers that the look and feel and even social experience of playing a game may be more important to preserve than the code. This is precisely the difficulty that preservationists face with dynamic and interactive content. Cases like video games offer heuristics that demonstrate one of the chief difficulties in the realm of preservation: it is very hard to predict the future. The difficulty is how to demonstrate, through the Dissemination Information Package (DIP), the temporal, spatial, and social aspects of content.

Conveying this information back to users is a function of multiple entities within the OAIS information model. First, an Archive must store sufficient information within its Archival Information Packages (AIP) to be able to convey external significant properties or to change Information Packages to match changing Knowledge Bases: this includes something like

the OAIS/FRBR mapping constructed as a result of PVWI wherein the model suggests linking to an outside source for Context and Provenance information [21]. Perhaps in a case like the ‘leaps and bounds’ advancement of a particular video game, the AIP would contain not only the game, but also references to popular articles, industry reviews, and fan content. The digital preservation manager at a private US university quoted above describing the interactive nature of technology also mentions the practice of documenting the experiences of users. For very complicated media that is one of the few (perhaps the only) options at this point in time. Another interview participant, a researcher at a European national archive, said:

*“I looked at technological hardware preservation. I looked at simulation—yeah, migration and emulation then documentation. Documentation is kind of like a separate thing but I felt because so many of these other things there are so many reasons why we can’t really do that yet. I feel like documentation is basically what we’re left with.”*

The second entity that is implicated in conveying mnemonic experience is the DIP. These types of experiential significant properties require creative work through DIPs to deliver authentic experiences to Consumers. The DIP is one of the more poorly defined entities within OAIS in large part, as one interview subject said, because it requires a prediction of the future. When Knowledge Bases change and people no longer understand how a d-pad works, the DIP for a Nintendo game has to go beyond simply providing a working console and cartridge to a Consumer. There is recent work that provides formal modeling of DIPs displayed as a set of services and exchanges with Consumers [12] and this work acknowledges the need for DIPs to change according to queries by Consumers; it suggests tracking different DIPs and the queries the spurred their generation and potentially adding them back into the AIP using the PREMIS standard for documentation; the most recent version of PREMIS even allows for the documentation of environments as their own objects, a move that recognizes that environments may be significant in the preservation of content beyond the bits themselves [6]. But even though PREMIS is a more specific and prescriptive standard that follows OAIS, it does not and perhaps cannot help to address what will need to be somewhat imaginative solutions for conveying the experiences of interactive and dynamic digital content. This entire concern is imbricated in the complexity of Designated Communities and Knowledge Bases. Archives are supposed to track Knowledge Bases and update content when Knowledge Bases change. This is a difficult task, not only because there are no current guidelines that deal specifically with this<sup>1</sup>, but also because change is both a hard thing to notice in the moment and a more difficult thing to document after the moment has passed. That there is no one solution is part of what makes this kind of thing hard to standardize; that there should be guidelines anyway is probably obvious given the complexity of the task.

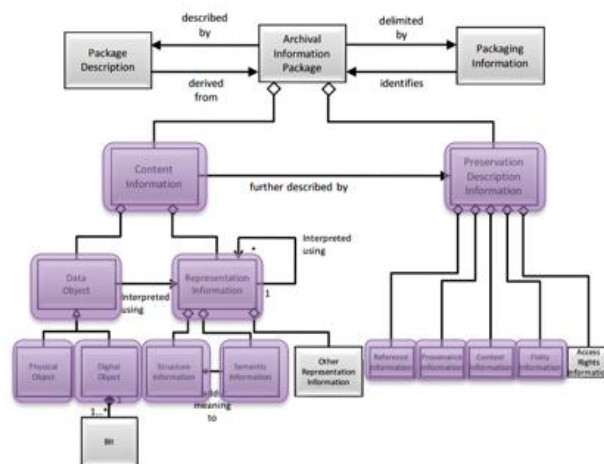
There are also ways in which it may be possible to overstate the difficulty of the digital preservation task: it may be that at this particular juncture, the preservation of surface and affective experience is not possible, particularly not at scale. One interview participant, a senior digital preservation consultant at a boutique US firm, noted that these preoccupations can serve to paralyze the field in such a way that getting to grips with what should be relatively simple tasks like bit-level preservation still have not been definitively addressed:

*“Yeah, I think it actually... and this isn’t OAIS’s fault, it’s just I think this field has suffered from -- in my opinion, it has suffered from too much fixation on those kinds of issues and not just doing the absolute minimum to get you to a point to have a future opportunity to visit those questions when the need really arises. We don’t even have good bit preservation nailed down, and that should be very easy. It’s really simple, it’s dumb, just do it, and stop talking about it, please. I’m so tired of it.”*

This participant also noted that concerns about significant properties are more challenging for some kinds of content than others. For audiovisual materials, she argued: “Watch it and listen to it, and look at it.” Another participant, a senior special collections archivist at a private US university, said, “So, for us to be able to push [a digital object] into something where we have, you know, huge, huge disk space, and to be able to say well, at least you know, it’s safe, the original is safe. I would think that would be like a big plus to people, just to be able to provide that as a service for their materials.” Keeping the 1s and 0s safe is a most basic requirement, and this might be seen as sufficiently significant in many cases, particularly if this is explicitly stated in users and donor agreements. Yet at the same time, multiple people have pointed out, including the authors of OAIS that I spoke to, that 1s and 0s alone are rarely sufficient, particularly when longer time scales are involved.

## 5. DISCUSSION

The previous section detailed some of the significant properties that arose in conversations with game programmers and OAIS practitioners and scholars. Here, I will demonstrate what maps well to the existing OAIS entities and what works less well. The figure below is an image from OAIS that details the contents of the AIP. I have highlighted in purple the entities wherein some significant properties could be located and I speak about some of these in the examples that follow.



**Figure 3. Significance in OAIS**

Some significant properties fit well within the Content Information entity in the AIP model. Content Information includes the Data Object itself, which can be comprised of both Digital Object(s) (bits) and Physical Objects. Source code of games fits here as do some physical ephemera essential for use, like *Carmen Sandiego*’s analog copy protection *World Encyclopedia*. Ephemera can also be documented as a separate

<sup>1</sup> Although one interview subject suggested the outcomes of the SCAPE project [28], suggesting “a lot of the idea in the SCAPE approach of preservation monitoring and planning is predicated

on evolution of and instruction of the designated community in technology, in semantics, in usage, in requirements.”

object and related to the digital data via the Context Information entity.

Access software and, by extension, access hardware may be documented as part of the Data Object itself or as Structural Representation Information. Changes in the Designated Community's Knowledge Base may be documented as Semantic Information, although there are limits. Including software as part of the digital object itself is something that OAIS does not do very well yet, according to some practitioners. One of the interview subjects has argued vociferously and publicly for its inclusion as part of the object itself in the 2017 revisions. Semantic Information can document significant properties like a language shift from English to Chinese, for example. Preservation professionals interviewed disputed whether or not it is the role of the repository to document changes in common knowledge, such as geographical names and borders or popular imaginaries in the case of *Carmen Sandiego*.

Sometimes what is significant about a game is its relationship to other games. One game programmer said, "*Doom*, for example, it made some of these huge graphics and texturing leaps and bounds, [these were] obviously... a product of its time." "Leaps and bounds" progress in one game necessarily relates it to a history wherein a game was markedly different than its contemporaries, as noted previously. Another significant property noted by interviewees is the relationship of a particular title to a larger franchise, for example a particular release of *Civilization* in relation to all versions. This was stated explicitly but is also tacitly implied when participants spoke about franchise games by collapsing an entire series into a single sociotechnical entity, saying things like, "Civilization is one of my really favorite games of all time," as opposed to naming a particular version or release of *Civilization*. In OAIS, this relationality can be mapped as Context Information within the Preservation Description Information entity. What is meant by Context Information is unclear to some interview participants; its description in the OAIS literature is similar to archival bond. Therefore, a repository can only express this Significant Property as Context Information if it holds enough games to demonstrate how a particular game relates to others.

Many interviewees acknowledged that preserving the affective and social aspects of games is a most challenging task. Playing games in arcades is a fundamentally different experience than playing at home; these locations impact game design, for example the simplicity of original *Donkey Kong* versus the deeper interaction of *Super Mario Brothers*. Creators and players describe the school-setting of the earliest *Oregon Trail* and *Carmen Sandiego* titles as a significant property. The need to understand the time and place in which a game was made and/or played might be easiest to understand with a game like *September 12<sup>th</sup>*, a news game predicated on the events of September 11, 2001. The twin difficulties are encapsulated by two quotes from different game developers. The first, a contemporary developer working in a US game studio, said, "...it's hard to differentiate between what is like your nostalgia and what is sort of useful, right?" A second quote, from a developer of a game series that is no longer in production, said, "So you really have to sort of capture the essence of the time. Now I don't obviously have a good answer for that, but somebody should think about it."

These Significant Properties do not fit well within OAIS. This may be because documenting this type of information in relation to a particular object has not always been seen as the province of the archive itself. In some cases, the preservation of some non-code significant properties of a game is more desirable than preserving working code itself: a video of game play, a textual narrative of a walk-through may better capture the

experience than working copies of obsolete technology. In fact, these expanded descriptions of what might be significant about a game challenge the very assumption that a baseline for a game's authentic preservation is its functionality.

## 6. CONCLUSIONS

Some significant properties, as suggested by interviewees in OAIS research and PVWII respectively, fit well within the existing OAIS entities. For others, one could argue for their inclusion within existing entities although it may mean stretching the capacity and meaning of these entities beyond what was envisioned by the designers of OAIS. This latter is not to indicate that such actions would be wrong: indeed, it is the role of a reference model to inform things in the future which likely entails moving into spaces the original authors could not envision.

Data from PVWII suggest that social and affective attributes of games are considered significant by designers and players. These significant properties are largely expressed as relational properties: they obtain in relation to objects, events, spaces, and times outside the object and often outside the archive or repository. These relationships are also nuanced in nature: certain properties are more important than others, or are only important in certain cases (for example, to particular Designated Communities). In fact, the situatedness of significant properties suggests that, for popular content like video games, the notion of Designated Communities is too vague and it is more important to think about archived objects in the context of Ranganathan's [24] third law: every [digital object] its [user]. Video games serve as an excellent case study for this type of research precisely because they are complex technological objects but also because their heterogeneous users offer up a complicated sociotechnical network within which to understand something like significance. But these findings are not specific to video games: rather, the case study serves to bring to the fore issues that are already present in long-standing preservation practices for analog materials and that are currently under debate for digital materials such as scholarly data, media art, web archiving, and the nebulous notion of digital archives more broadly.

Some significant properties identified from within the PVWII data fit within OAIS, such as semantic and environment information; others will require either new metrics or changes to the existing standard, like affective and relational values. These findings are echoed by similar comments from OAIS interview subjects, and this is all the more pertinent given both the variety of participants in this latter study and the fact that the conversations I had with them were very different in nature and subjects from the PVWII interviews. The similarities between the two datasets, PVWII and OAIS experiences, speak to the salience of these themes beyond the theme of video games and within wider digital preservation discourse.

What was surprising about this project was just how much data I struggled to map to OAIS: my original hypothesis when I began this mapping project was that all Significant Properties should fit within OAIS, given its commitment to changing Knowledge Bases over time. For example, the process of documenting context is nothing more than moving additionally pre-inscribed affordances of a digital object into the circumscribed setting of the Archive. In the language of OAIS, adding information from the Knowledge Base of the Designated Community to the AIP as additional documentation is taking what is normally afforded to the stakeholders forming the Designated Community and pulling it into the AIP. This finding is an extension, and not necessarily incommensurate, with earlier work done on significance in OAIS [13] and work on significant characteristics [7]. What this paper suggests is merely an



extension of an on-going balancing act, of finding the line for sufficiency in deciding how much to document: this is precisely why I call for the creation of metrics to help drawing these artificial boundaries so that this work can be made machine-actionable for digital preservation at scale. My conclusion is, therefore, that all significant properties do not fit within existing OAIS entities and I echo the calls of other preservation scholars that changes are needed in the ways in which we think about the responsibilities of repositories, especially given the potential for distributed digital preservation in linked data environments. Additionally, I posit that these difficulties will be exacerbated in areas where OAIS already does not work as well. A couple interview participants noted that the scripts within OAIS presume a level of infrastructure. While Seles [29] demonstrates how this plays out in situations where Archives are located in geographical regions where the legal, electrical, and network infrastructure are missing, some of my interview participants pointed out that, even in wealthy first world contexts, institutions wherein preservation is not a primary function will lack many of the structures presupposed by OAIS.

In this paper, I do not tackle the breadth of descriptions or definitions about what significance actually means, whether characteristics, properties, or anything else. In fact, this work encompasses many of the definitions from digital preservation literature. Instead, I locate claims that significant properties are situated and sometimes outside the digital object and its computing environment within a growing body of archival science literature that speaks to the situatedness of archival content and what is needed to contextualize it [18]. The juridical and legal undercurrents of archival conceptions of authenticity are balanced by work in practice, where archivists understand that evidence, for example, aids in interpretations of the world [5] and that archives may have the role of preserving mnemonic devices in addition to evidence [31]. What is necessary is for digital preservationists to decide whether what is wanted is particular bits of information or impressions of the past.

## 7. ACKNOWLEDGEMENTS

I wish to thank all the investigative partners from Preserving Virtual Worlds II, including colleagues at the University of Illinois, University of Maryland, Rochester Institute of Technology, and Stanford University. PVWII was supported by a grant from the Institute of Museum and Library Services; additional research was supported by the Graduate School of Library and Information Science at the University of Illinois, Urbana-Champaign. I also acknowledge Jerome McDonough, Gail Ferguson, and Nikki Weickum for their feedback on earlier drafts.

## 8. REFERENCES

- [1] Bates, M. (2015). The Information Professions: Knowledge, Memory, Heritage. *Information Research: An International Electronic Journal*, 20(1)
- [2] Bradley, K. (Summer 2007). Defining Digital Sustainability. *Library Trends* 56(1): 148-163.
- [3] Cargill, C. (1997). *Open Systems Standardization: A Business Approach*. Upper Saddle River, NJ: Prentice Hall.
- [4] Consultative Committee for Space Data Systems (June 2012). *Reference Model for an Open Archival Information System (OAIS)*. CCSDS 650.0.M-2, Magenta Book. Issue 2.
- [5] Caswell, M., & Gilliland, A. (2015). False promise and new hope: dead perpetrators, imagined documents and emergent archival evidence. *International Journal Of Human Rights*, 19(5), 615-627. doi:10.1080/13642987.2015.1032263
- [6] Dappert, A. (2015). Digital Preservation Metadata and Improvements to PREMIS in Version 3.0. Webinar.
- [7] Dappert, A., & Farquhar, A. (2009). Significance is in the eye of the stakeholder. In *Research and Advanced Technology for Digital Libraries* (pp. 297-308). Springer Berlin Heidelberg.
- [8] Dappert, A., Peyraud, S., Delve, J., Chou, C. (2012). Describing Digital Object Environments in PREMIS. In: iPRES 2012 - 9th International Conference on Preservation of Digital Objects, October 3-6, Toronto, Canada.
- [9] Delve, J. (2013). Introduction to POCOS e-Book 3: Preserving Gaming Environments and Virtual Worlds In The Preservation of Complex Objects: Volume 3. Gaming Environments and Virtual Worlds. JISC.
- [10] Duranti, L. (1994). "The concept of appraisal and archival theory" in *The American Archivist* 57, pp. 328-344.
- [11] Duranti, L. (2010). "The appraisal of digital records: the return of diplomatics as a forensic discipline." International Conference on Surviving the Digital Wild Frontier. Singapore, Singapore.
- [12] E-ARK. (2015). E-ARK DIP Draft Specification. D5.2, Revision 4
- [13] Giaretta, D., Matthews, B., Bicarregui, J., Lambert, S., Guercia, M., Michetti, G., Sawyer, D. (2009). Significant Properties, Authenticity, Provenance, Representation Information and OAIS Information. iPRES 2009: the Sixth International Conference on Preservation of Digital Objects. California Digital Library, UC Office of the President
- [14] Hedstrom, M. and Lee, C.A. (2002). "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](http://ec.europa.eu/transparency/archival_policy/dlm_forum/doc/dlm-proceed2002.pdf)
- [15] KEEP (Keeping Emulation Environments Portable). <http://www.keep-project.eu/ezpub2/index.php>
- [16] Knight, G. (2008). Framework for the definition of significant properties. Retrieved from <http://www.significantproperties.org.uk/documents/wp33-propertiesreport-v1.pdf>
- [17] Lee, C. (2005). "Defining Digital Preservation Work: A Case Study of the Development of the Reference Model for an Open Archival Information System." Doctoral Dissertation. The School of Information, University of Michigan, Ann Arbor.
- [18] MacNeil, H. and Mak, B. (2007). "Constructions Of Authenticity." *Library Trends* 56.1, 26-52.
- [19] Matthews, B., McIlwrath, B., Giaretta, D., & Conway, E. (2008). The significant properties of software: A study. *JISC report, March*.
- [20] McDonough, J. (2013). A Tangled Web: Metadata and Problems in Game Preservation In The Preservation of Complex Objects: Volume 3. Gaming Environments and Virtual Worlds. JISC.
- [21] McDonough, J., Olendorf, R., Kirschenbaum, M., Kraus, K., Reside, D., Donahue, R., Phelps, A., Egert, C., Lowood, H., & Rojo, S. (2010). Preserving Virtual Worlds Final Report. Retrieved from IDEALS online 2011-11-30.
- [22] Meghini, C. (2013) Data preservation. *Data Science Journal* Volume 12, 10 August 2013, pp. GRDI51-GRDI57

- [23] Pinchbeck, D., Anderson, D., Delve, J., Ciuffreda, A., Otemu, G., Lange, A. (2009). Emulation as a Strategy for the Preservation of Games: The KEEP Project
- [24] Ranganathan, S. R. (1931). The five laws of library science.
- [25] Rosenthal, D. (2015). The case for a revision of OAIS. Digital Preservation Coalition OAIS Wiki. Available at: [http://wiki.dpconline.org/index.php?title=The\\_case\\_for\\_a\\_revision\\_of\\_OAIS](http://wiki.dpconline.org/index.php?title=The_case_for_a_revision_of_OAIS)
- [26] 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.
- [27] Schwartz, J. & Cook, T. (2002). "Archives, Records, and Power: The Making of Modern Memory," *Archival Science*, 2:1/2: 1-19.
- [28] SCAPE. Scalable Preservation Environments: <http://scape-project.eu/>
- [29] Seles, A. (2016). The Transferability of Trusted Digital Repository Standards to an East African Context. Doctoral Thesis. Department of Information Studies, University College London. London, UK.
- [30] Sierman, B. (2012). "OAIS 2012 Update" in Digital Preservation Seeds Blog accessible at <http://digitalpreservation.nl/seeds/oais-2012-update/>
- [31] Smith, A. (Summer 2003). Authenticity and Affect: When is a Watch Not a Watch? *Library Trends* 52(1), pp. 172-82.
- [32] 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](https://doi.org/10.1045/january2013-webb)
- [33] Yeo, G. (2010). "'Nothing is the same as something else': significant properties and notions of identity and originality." *Archival Science* 10, no. 2: 85-116.