**Perspectives on the Recent Evolution of the OAIS Reference Model**

Don Sawyer

28 January 2018

This paper suffers from a number of fundamental misunderstandings and fatal flaws which I attempt to clarify below.

 David Giaretta

**Purpose:** The audience for this paper is the set of active participants in the 5-year

review of the OAIS Reference Model (RM). The purpose of this short paper is NOT to

recommend changes to the near final version of the current draft although this author

would like to see some. Rather it is to suggest that there has been a partial breakdown

in the announced strategy for addressing updates to the OAIS RM and further this

breakdown appears to be largely unrecognized. It is hoped that this analysis will

resonate with the current participants and will result in an improvement in the next 5-

year review perspectives and approach.

**Context:** To address the purpose of this paper, it is first necessary to provide context

for this analysis. It is believed that most participants in the current review will largely

agree with this context.

The OAIS RM was initiated to provide a common communication framework to facilitate

discussion about the preservation of information, the role of archives, and their

implementations. A special focus was put on digital materials because of their fragility,

but since many archives include non-digital materials the model includes them in its

overall approach. The development strategy also asserted that the model should be

widely applicable in the belief that substantial commonality existed across many

disciplines. To this end, although led by the Space Agencies, participants with

experience in preservation from various disciplines, including science, national archives,

libraries, and others met over a period of years. Therefore the initial model reflected

both reality based on experience and a collective view on what could be a reasonable

view of a more idealized situation. The development strategy also attempted to cover

the functionality with sufficient detail to make clear what was meant by the key terms

and concepts, but it was explicitly noted that actual implementations may break out their

functionality differently. The same was not said explicitly about the information modeling

but many different choices for this modeling could have been adopted. As a

communication vehicle, it was assumed that when communicating using the OAIS

context, deviations from the OAIS RM would be addressed by explicit description. That

this was successful, despite being far from perfect, has been demonstrated by its wide

adoption across many disciplines. We all know this but it is useful to be reminded.

At its most basic level, the initial OAIS RM put the OAIS Archive (Archive) as an

intermediary between a Producer who has information to be understood by others over

the long term, and Consumers who are the ‘others’ that are expected to be able to

obtain this understanding at any point in the future. More recently there has been a

growing recognition that, in general, preserving an experience should also be included

as a possible objective.

I agree that “understandable” may be interpreted in a limited way – therefore we made the following clarifications:

1. The term Preservation Objectives has been introduced: *Preservation Objective: A specific achievable aim which can be carried out using the Information Object.*

In addition we extended “Independently Understandable” *Independently Understandable: A characteristic of information that is sufficiently complete to allow it to be understood by the Designated Community, as exemplified by the associated Preservation Objectives, without having to resort to special resources not widely available, including named individuals*

These changes allow us to describe things like a performance of contemporary music such as that of Pierre Boulet, to be preserved. This is something I was involved with in the CASPAR project.

1. The Designated Community (a subset of consumers) are the ones which should be able to understand

In order to provide a maximum degree of consistency over the long term, 5-year reviews

have limited the updates to correcting errors, clarifying concepts as needed, and adding

concepts only when it appears they are needed to accommodate widely based and

evolving community preservation concerns. Concepts are not to be revised just

because some think there are better ways to organize them. Unfortunately this author

finds that this approach to providing needed stability, in the context of the current 5-year

review and update process, has not been adhered to in the major update to the Archival

Information Package (AIP).

**Analysis:** The revised version of the OAIS RM recently made available

(650x0w2x1-20190114.doc) has changed the definitions of all of the Preservation

Description Information (PDI) components: Provenance, Context, Reference, Rights,

and Fixity. Whereas previously they were applicable to the Content Information (and

thus to all of its components), now they are only applicable to a Content Data Object.

Therefore when the Content Information is a digital object, the PDI is understood to be

no longer applicable to the Content Information’s Representation Information. When the

Content Information is non-digital, the PDI is no longer understood to be applicable at

all.

This is incorrect.

1. Representation Information is an Information Object. If this is to be preserved, which it must be, at least in the short-term, then it can be viewed as a Content Information Object (perhaps with the OAIS as the Producer) and so PDI will apply to its Data Object. A more detailed reply to the point about PDI is provided below.
2. Even if the Content Data Object is non-digital, it is nevertheless, by definition, still a Content Data Object and therefore would have to have PDI.

This is a major information modeling change, involving well over 100 edits to the

document, and it is not based on any error with the previous applicability of PDI to the

Content Information.

1. The previous applicability of PDI was clearly being misunderstood by archives, so this change is a clarification.
2. I would point out that in the previous revision we changed “archive” to “Archive” which resulted in several hundred changes.

This removal of PDI applicability to the Representation Information has significantly

degraded the importance of Representation information (generally regarded as a major

advance in dealing with digital information) as it has removed it, **in the context of**

**normal OAIS communication**, from explicit concerns about authenticity for better

understanding (think semantics as well as format information), reference for

independent searching, access rights, and fixity to help ensure it has not been

corrupted.

As pointed out above, this is incorrect. Every object that is being preserved by an archive must have some information that

1. Ensures that it has not been changed
2. Tells us where it can from and what has happened to it
3. Allows us to reference it
4. Says who/what can change it
5. The context of that object

In other words it has all the elements of PDI

As an attempt to limit this damage, a note has been added to the PDI

definition stating that provenance, context, etc. is still important to Representation

Information. This does not alter the fact that it is no longer part of the OAIS Information

Modeling and thus is no longer part of standard OAIS communication.

Incorrect!!

The rationale

given in the note is that this ‘is simply to ease discussion of these concepts at the

Content Data Object level’. Unfortunately this is not a valid rationale for two reasons: 1)

the previous model was not in error, and 2) one could simply refer, for example, to the

Content Data Object’s Provenance or the Content Data Object’s PDI for ease of

discussion.

It is further asserted, in the note, that provenance, context, etc. (all lower case and thus

not defined) should be applied to Representation Information and to any ’other’

Information that the Archives is preserving. What ‘other’ information is this referring to

and what is the justification that is relevant to actual Archives? On the one hand, it is

only in a note. But it is still interesting to consider what this ‘other’ information, apart

from Representation Information, might be. The OAIS does not define ‘preservation’

and clearly a real Archive can not put an equal effort into managing all the various types

of Information it is maintaining within the Archive. That is why Content Information from

the Producer has been singled out as the key information intended to convey an

understanding and/or an experience to future Consumers.

Remember the “Producer” has been defined from the original (2002) OAIS as Producer: The role played by those persons, or client systems, who provide the information to be preserved. This can include other OAISs or internal OAIS persons or systems.

In other words the Producer can be the OAIS itself, which is something which in our earlier emails you had clearly also ignored.

The PDI is (was) an

augmentation to long term preservation of the Content Information. ‘Other’ information

in the Archive could include password control files, descriptions enabling the finding of

information (Access Aids), Packaging Information, and the PDI itself. None of these,

with the possible exception of PDI, appears to have any real long term preservation

requirement involving the full suite of PDI components. This seems true in the general

case and thus hardly becomes a subject for inclusion in the OAIS (standard) Information

modeling. The possibility that PDI could or should have its own version of pdi can not be

dismissed out of hand. It has been discussed as a concept, particularly when one

wants to argue that the PDI components should get their own fixity. However my

analysis, submitted to David last September (private communication), suggests there is

very little utility in most cases with the exception of applying fixity. It would also be easy

to add provenance to the Provenance by including its source, etc.

As noted above one also needs to be sure that all the things which the OAIS needs to preserve

* has not been changed in an undocumented manner
* one knows where it came from
* how one can locate it,
* who can change it
* what the context is

In other words, it has all the components of PDI. Of course most of these things are provided by the computer system e.g. who can change something is a key function of an operating system.

It seems that a paper

on this subject, looking at real cases and examples of what pdi applied to PDI would

look like, is needed if this is to have any real validity. In short, I’m left to conclude that

the notion of applying pdi as a set of components to other information in the Archive is

premature at best.

As noted above, anything an archive is preserving requires these things. **Otherwise** you would be saying that you have some, let’s say Provenance

* + but you don’t know where it came from
		- presumably it just appeared from nowhere
	+ but you don’t know whether it has been changed
		- so what good is it
	+ but you don’t know who can change it
		- maybe the janitor can alter it
	+ but you don’t know how to locate it
		- so it is lost
	+ but you don’t know if it being used in the right context
		- so maybe you are applying the Provenance for a NASA image to a picture of a Picasso painting

**Conclusions:** The author was unable to attend all the relevant Telecons and

international meetings so a clear understanding of how this major change to the AIP

model came to be is not apparent. This is best left to those participating on a more

regular basis. Contributing factors seem to have included a very large number of

suggested changes from the wider community, some mis-understanding of the extent to

which the modeling of an association with an entity also applied to its components, and

pressure to conclude the review process. Perhaps most significant is that the role of the

OAIS RM as a standardized communication framework at a conceptual level appears to

have been unduly influenced by the use of the OAIS RM terms and concepts in

developing the auditing document 16363. This is just the author’s impression and it

may be incorrect. Finally, it may be argued by some that a more aggressive approach

to revising OAIS terms and concepts is warranted. This author believes that the long

stated approach to OAIS reviews will be most productive and that it should be the

starting point for the next 5-year review. Perhaps PDI will be put back in association

with Representation Information, at that time, based on the lack of this important

concept in the newly revised OAIS communication framework.

The guideline on the <http://review.oais.info> says:

*In order to ensure the continued usefulness of OAIS any revision must remain backward compatible with regard to major terminology and concepts. Further, for consistency the general level of detail should not be changed nor should the standard be changed from a reference model to an implementation design. Archive implementation standards or implementation profiles or detailed archival process standards or protocols should be addressed, but not in this document. They would become separate standards and would be developed through separate CCSDS projects. A particular interest for the current OAIS update is to reduce ambiguities and to fill in any missing or weak concepts and to add useful terminology.*

As noted above it was clear that there was a need to clarify the concepts of what, say, Provenance applies to – it was being misunderstood by all repositories.