

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

Don Sawyer  
28 January 2018

**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.

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 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.

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 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. 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. 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. 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.

**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.