CITATION

Creative Commons License: Attribution-NoDerivatives 4.0 International (CC BYND 4.0)
QUESTION

In your literature review #3 you state that "the conclusion from a variety of experienced repository managers is that the authors of the OAIS Reference Model created flexible concepts and common terminology that any repository administrator or manager may use and apply, regardless of content, size, or domain."

1) Does this one-size-fits-all model really work for repositories large and small? Please discuss.

2) You also note that Higgins and Boyle (2008) in their critic of OAIS for the DCC talk about the need for an OAIS lite. Please discuss what that might look like, who would be its primary audience, and how useful it could be.

3) Finally, how can repositories such as the US National Archives work with the concept of designated community as their mission is to serve all citizens. Is the notion of designated audience generally useful? Why or why not and under which conditions is it most valuable?

RESPONSE

Based on the feedback this author has received from participants attending the DigCCurr Professional Institute in May 2011, no, the one-size-fits all OAIS Reference Model recommendation does not work for repositories both large and small. The repository administrators in question were discussing digital curation concepts in general, but this may also be applied to the OAIS Reference Model, as this is one part of digital curation. The repository administrators wanted to know what part of what they had learned at the Institute they should apply, and which parts they could safely leave out. The attendees thought the information presented to them was useful, but that it would be “overkill” for their particular repositories.

Beedham, et al. (2004?) noted that the design of the OAIS Reference Model is for that of a repository within a large bureaucracy. The authors wrote that the Reference Model is not designed for a small archive collection with a limited audience with limited funding and personnel to build, maintain, and preserve the collections in the repository. It is designed for an institution with a team of personnel working on the repository, not one or two people responsible for all aspects of creating, maintaining, and preserving it. This author would add that the OAIS Reference Model is designed with an archive whose
collections consist of tens of thousands to $n$ number of objects. It is not designed for an archive of a few hundred or a few thousand objects with one person to administer it, who may or may not be trained in digital library/digital archive Information and Library Science (ILS) Best Practices.

The Reference Model has been designed such that it may federate with other OAIS archives, presumably to create access to one Really Large Preservation Repository. It has also been designed so that the object may have three different “versions”: the Submission Information Package (SIP); the Archival Information Package (AIP); and the Dissemination Information Package (DIP). As a concept, these are three different things, but in practice, a SIP may equal an AIP, which may equal a DIP. For a large repository with different audiences, the DIP may need to be different from the AIP. For a small archive with a homogeneous audience, the AIP and DIP may be exactly the same.

Therefore, with regards to my statement, “...any repository administrator or manager may use and apply, regardless of content, size, or domain”, the key is in the use of the word, “may”. They may use it. It is not that they must use it, or they will use it, it is simply that the repository administrator may use it. A repository administrator must take into account the rules and regulations that apply to their repository when applying the OAIS Reference Model. These rules and regulations may be domain Best Practices that differ from ILS practices, or federal, state, institutional or other local policies that differ from what the OAIS recommends. The OAIS Reference Model is a recommendation, not a requirement or a law. As Thibodeau wrote (2010?), any evaluation of a repository must be taken on a case-by-case basis. In other words, one size does not fit all.
The primary responsibility of a repository manager is to ensure the near-term availability of the objects in the repository, and the long-term as well, if that is part of the mission of the digital archive. This author has two views of what an “OAIS Lite” might look like. The first is to determine what is actually required to preserve content for the long-term, regardless of the model used. The second is how the documentation of the recommendation could be adapted to create an “OAIS Lite”. The primary audience for an OAIS-lite would be the managers of small- to medium repositories who do not operate within large bureaucracies, and, perhaps, have some kind of computer science knowledge, but who will generally not have an ILS background.

Jon Crabtree of the Odum Institute at the University of North Carolina at Chapel Hill supports the use of standards, but he has noted on several occasions that the Odum Institute “preserved” their digital data for decades without explicit preservation standards or policies. They did this because they hired competent people who did their job, and because it was understood that the data itself must be migrated, and the software and hardware must be migrated, replaced, upgraded, etc. This author’s own work experience seconds Crabtree’s comments.

At the time of this writing, the following must occur in order for data to be preserved without following any particular recommendation for preservation. Although this section is designed to be illustrative of “bare bones” preservation requirements, the “[]” designates the OAIS Reference Model section in which this would fit; i.e., either the “Information Model” or the “Functional Model”.

1. [Functional Model] Document the holdings of the archive and its system design. Update the documentation if and when there are any changes to numbers 2-4 below.
2. [Information Model] Ensure the appropriate metadata for the digital objects.
3. [Functional Model] Migrate & refresh the hardware and software periodically, as well as any software required to render the objects in the
repository (for example, CAD files). Upon ingest run integrity checks and virus scans. Periodically run these scans on the data. Set up at least 2 off-site back ups, and check that the back ups are actually backing up the data. Ensure all of the objects in the repository may actually be found and accessed, assuming access is permitted and desired.

4. [Functional Model] Find someone to take the data if the organization in charge of the data goes out of existence. Keep (1) above updated in order to facilitate a takeover of the archive’s contents.

5. [Functional Model] Hire competent people who ensure that numbers 1-4 above occur.

Additional steps a repository administrator may take are to take the documentation from (1) above, map it to the OAIS Reference model and identify gaps. Then, as time and resources permit, address any existing gaps within the current system design and content versus the OAIS Reference Model. At the least, identify that the gaps exist and document this in (1) above.

This author’s vision of an “OAIS Lite”, therefore, would be very general guidelines for the type of administration and management required to maintain a digital repository over time. This may not be what Higgens & Boyle (2006) had in mind.

However, if this author were to create an “OAIS Lite” based purely on the OAIS Reference Model recommendation itself, then it would be the current recommendation, but with each subsection designated as:

1. “Must have”/required.
2. “Nice to have”/recommended.
3. “Optional”.

The assumption is that if some part of the recommendation is not necessary, then it won’t be in the OAIS Reference Model recommendation at all. Thus, “not needed” is not provided as an option. This also assumes the same audience as outlined above for the “bare bones” preservation guidelines. This would have the advantage of breaking down
the Reference Model into manageable chunks. A repository manager of any size could begin by implementing the “must haves”; as time permits, add in the “nice to haves”; and, again, as time permits, add in any “optional” sections.

Another possibility is to divide the recommendations in the Reference Model by repository size, and then break those down by “required”, “recommended”, and “optional”. A committee of experienced repository administrators working with small repository owners could set up the Reference Model in this way. Either of these formats would be a useful version of the recommendation.

Thus, an “OAIS Lite” could consist of two types of recommendations. The first is a description of the bare bones functions required to maintain a repository and its contents over the long-term, mapped to the general OAIS models. The second version would be to take the recommendation itself, and break it down into required, recommended, and optional sections. Breaking down the recommendations would be useful to the managers of both large-and small repositories. The challenge would be get a committee of repository experts to agree on what constitutes “required”, “recommended”, and “optional” within the OAIS Reference Model.

The concept of a Designated Community is useful within the OAIS Reference Model, as it reminds repository managers that the goal of the repository is to serve a set of users. The goal is not necessarily to serve the needs of the repository managers! The concept is most useful when the users of a repository are homogeneous, and it is least useful when the users are heterogeneous. This is because the more heterogeneous the population using a repository, the less “one size fits all” fits all users. It is easier to serve a specific set of users (“scholars”) than all users (“hobbyists” and “scholars”).

Having said that, an organization like the National Archives may work around this limitation by aiming collections at specific users, once a baseline standard has been met.
So, for example, the Southern Historical Collection at UNC was initially put online for scholars and, to some extent, “to serve the people of North Carolina (NC)” (as that is also the stated mission of the University of North Carolina at Chapel Hill), but the administrators of the collection soon realized that K-12 educators were using the resource. Thus, the administrators of the digital library still serve their “generic” audience (“the people of NC”) and scholars of Southern history, but they have developed K-12 educational materials for teachers to use as part of the state curriculum.

This author believes it is possible for the National Archives to serve “the people of the United States” by breaking down the digital collections by themes, collections, etc., and determine who uses what collections, and how. They can thus better serve specific audiences, and tailor the site as needed. The administrators of an archive still must determine who their “general” Designated Community is, and set standards for that community, but can, as needed, serve targeted communities.

In conclusion, the “one size fits all” model of the OAIS Reference Model does not fit all. It is important to have standards for preservation repository design, but when the preservation repository design is more suited to a large bureaucratic institution than a small repository with fewer resources, then not all of those standards may be useful. If not all of the standards are applicable or seem like “overkill”, then the repository manager will need to decide which of the standards to use, and how. One way to ease this “cherry picking” of preservation repository standards is to determine the processes required to ensure preservation, regardless of repository design. A second way is for ILS and Computer Science experts to break down the OAIS Reference Model recommendations into “required”, “recommended”, and “optional”, also possibly based on a repository’s size. This would be useful to managers of repositories of all sizes, as it
would help the manager figure out what they have right so far, or where they need to start, and allow him or her to figure out what gaps remain.

A downside to this idea is that if a repository only implements the “required” recommendations of the OAIS Reference Model, then they may be only partially OAIS-compliant, and it might encourage laziness among repository administrators.

Regardless, “content is king”, so the important issue is that the content and its metadata, along with any required software to run it, are preserved. The model used to preserve it is secondary. Finally, while the concept of a Designated Community is important, it is a more valuable term when the users of an archive are more homogeneous, and less useful when the user base is heterogeneous. Large archives at the national level may work around this limitation by setting a baseline standard of quality for all users, and then targeting the archive’s collections to particular audiences who use those collections.