Orbis Cascade Alliance
Content Creation & DIssemination Team
Digital Content Joint Working Group
Report on Surveys
2015 December 4

Isaac Gilman, Pacific University (chair)
Jodi Allison-Bunnell, Alliance staff

Digital Content JWG Members
Ping Fu, Central Washington University (Systems Team Representative)
Susan Hinken, University of Portland (Shared Content Team Representative)
Linda Crook, Lane Community College (Collaborative Workforce Team Representative)
Beth Blakesley, Washington State University (Discovery & Delivery Team Representative)
Rose Krause, Eastern Washington University (Content Creation & Dissemination Representative)

<u>**Overview of Surveys, Results, and Recommendations of the JWG**</u>
The Digital Content JWG (DC JWG) is charged with informing and supporting the Content Creation & Dissemination Team (CCD)'s work to identify shared practices that will facilitate the discovery, access, and use of unique Alliance digital content. More specifically, the DC JWG is exploring practices that will enable the aggregation, collocation, and shared discovery of content from across Alliance institutions.

There are two primary components to the DC JWG's current work: understanding the unique digital content landscape within the Alliance and gathering feedback on proposed shared metadata standards that would facilitate the aggregation of Alliance digital content for ingest into Primo, DPLA, or another discovery system. The environmental scan of the Alliance landscape includes (a) an inventory of the quantity and type of digital objects currently held by Alliance institutions and the platforms in use to manage those digital objects; and (b) the desire to manage digital objects in Alma and/or make those objects discoverable in Primo. Both this information and feedback on the proposed shared metadata standards were gathered in November 2015 through two separate surveys. CCD Representatives received notices of the survey schedule and structure starting September 25 (and repeated on October 23) so that they could be prepared to answer on behalf of their institutions. The first survey was open October 28-November 6 and the second was open November 9-24.

Survey Results Summary: Unique Digital Content Scan
The environmental scan survey of current digital content management revealed that there are over one million digital objects in Alliance DAMs/IRs. Of these, almost 500,000 are openly accessible, and could potentially be ingested into DPLA or other aggregation portals/systems. Many institutions expressed uncertainty as to whether they would want to be able to manage digital objects in Alma; however this uncertainty appears to be largely based on the lack of information available as to the management functionality that is or will be possible in Alma.

There was more explicit interest in making unique digital content discoverable in Primo, but it is clear from responses that Primo must be exposed to search engines in order for inclusion in Primo to be truly valuable to Alliance institutions. The value of unique digital content will not be fully realized if only shared within the Alliance; discovery must be possible by external users. This is a discovery need for unique collections that is distinct from the shared or purchased collections. There is also a clear need for support and shared practices as institutions struggle to create and maintain pipes to institutional and shared Primo.

A more immediate problem that should be addressed immediately is to resolve the question of MARC record analogs for collection in Archives West. The longstanding practice has been to create MARC records with 856 links to Archives West. A more efficient workflow would be to pipe those finding aids into shared and institutional Primo, but that requires both a de-duplication of existing MARC records in the Shared ILS and a resolution to the need to create MARC records in OCLC Worldcat. This is a known problem that was not able to be addressed during the implementation; it's now time to address it.

On the Center of Excellence project on managing digital objects in Alma, the comments indicate a strong interest, many questions, and an interest in seeing what emerges. There are strong themes that the Alliance should use as its criteria for success of the initiative as it works with Ex Libris:
- Primo must be optimized for search engine exposure, or managing digital objects in Alma is a non-starter
- Workflows for loading objects into Alma must be streamlined; they are currently too difficult to use

The Alliance should also share more information and provide updates on this initiative. (See https://www.orbiscascade.org/center-of-excellence-initiative-digital-formats/)

Survey Results Summary: Shared Metadata Standards for Digital Content
There is a general high level of support for the adoption of digital content metadata standards, which is consistent with Council's support for the initiative. However, it is clear that the standards require further revision before they would have full support for adoption (and before they would be fully usable). This will require the DC JWG to consult with Alliance metadata experts--primarily the working group that refined the draft standards in fall 2015--to prepare necessary revisions. The survey results also indicate that Alliance members are most comfortable with an output standard that allows for continuity in established local practices, but many institutions would appreciate and benefit from support and instruction on how to use their particular system to most easily produce the desired outputs.

The DC JWG and CCD Team should continue to facilitate conversations around the desirable Alliance role in supporting aggregation of digital content (which may or may not extend beyond the adoption of, and support for, shared metadata standards).

For more details on the outcomes of the surveys, see the detailed reports for each one that follow.

Recommendations
- Develop shared practices and support for piping digital collections from DAMs or IRs into institutional and shared Primo
- Resolve the question of MARC analogs for Archives West collections
- Advocate for search engine optimization of Primo with Ex Libris
- Use survey results to inform the COE on digital objects
- Modify the draft digital content metadata standards:
    - Re-focus as primarily an output standard, with optional support for inputs
    - Engage questions about specific fields raised in survey response comments and resolve appropriately with the assistance of metadata experts
    - Circulate to CCD Representatives and Teams again for review and approval
- Continue to engage questions around the Alliance role in preserving and exposing unique collections

Next Steps
The outcomes of these surveys will inform the following efforts:
- The inventory of DAMs and IRs will move to the Alliance's documentation center, when available, so that Alliance members can update the information there (thus eliminating the need for future surveys on this subject)
- The number of DAMs, IRs, amount of customization, and number of fully open digital objects will inform the CCD Team's work on proposing a DPLA hub for Alliance members
- Current and desired practices for managing unique materials in the Shared ILS will inform the DC JWG's work on recommending development of shared practices
- Desires for managing digital objects in Alma will inform the Center of Excellence work on digital objects
- Feedback on the draft digital content metadata standards will inform needed revisions to that draft and further higher-level conversations on the shared vision for aggregating digital content at the Alliance in fulfillment of the Strategic Agenda.

Survey on DAMs, IRs, and Unique Content in the SILS

This survey had two parts: one to gather base-level information on all digital asset management systems (DAMs) and institutional repositories (IRs) in use at the Alliance and the approximate number of digital objects; and one to ask about current management of unique items in Primo and Alma, and desires for managing digital content in Alma (which in turn informs the Center of Excellence project with Ex Libris; see https://www.orbiscascade.org/center-of-excellence-initiative-digital-formats/).

The survey was sent to the Alliance's 37 CCD Representatives on October 28, with reminders sent to those who had not responded on November 2, and responses due by November 6. We received responses from 34 of the 37 Alliance members.

DAM and IR Inventory

The DAMs and IRs in use at the Alliance are:
- ContentDM: 18 installations
- BePress: 11 installations
- Omeka: 8 installations
- Hydra: 2 installations
- Dspace: 6 installations
- Other (including Greenstone, homegrown, and other systems, some of which may not be DAMs or IRs): 12 installations

Most of the DAMs and IRs in use are customized to some extent; a few are highly customized, and a similar number use the default settings.

In total, there are 1,090,253 digital objects in Alliance DAMs and IRs. Of those 485,042 are open or unrestricted and potentially able to be contributed to the Digital Public Library of America. The remainder are either partially or fully restricted to campus use and could not be contributed to DPLA.

The inventory of systems in use, URLs, and numbers of objects will be transferred to a document that institutions can update as soon at the Alliance's documentation center is launched.

Unique Materials in the SILS

Part 2 was divided into three parts: questions about what unique materials Alliance members are piping into Primo; what unique materials that Alliance members are managing in Alma; and questions about desires for managing digital objects in Alma to inform the Center of Excellence project on digital formats.

On unique materials piped into institutional Primo, twenty-one institutions responded that they are doing nothing. Four institutions reported that they are piping in from a DAM, two from an IR, three are piping in finding aids from Archives West, and one is piping in finding aids from an

institutional site. Although it was in a different section of the survey, one institution commented that they have been unable to work with the MARC records for collections in Archives West and need assistance to resolve that problem. With little current activity and some thorny problems, this appears to be an area ripe for the development of shared practices.

On unique materials managed in Alma, rare books and yearbooks were the most prevalent formats, with 19 and 14 institutions respectively. Nine institutions manage yearbooks, seven manage no unique materials, six manage faculty publications, and five manage learning objects. Only one institution is managing data sets in Alma. Comments also indicated management of college and university publications, theses and dissertations, and collection-level MARC records for archival materials.

In answer to questions about managing digital formats in Alma, 13 institutions indicated that they want to be able to do this, with the understanding that this is a developing capability. The majority of respondents, 18 institutions, responded that they don't know. Two institutions indicated that they do not want to manage digital objects in Alma. Some commented that they need more information on this initiative and would appreciate regular updates. Some institutions mentioned a strong need for Primo to be optimized for search engines before management of digital objects wholly in Alma and Primo would be at all viable. Two institutions commented that the process of loading unique content into Alma and piping into Primo is time-consuming and complicated. Several comments indicated that their answer to this question depends on the functionality and ease that this development might bring them, but that switching from current DAMs or IRs would have to be for compelling reasons.

Survey on Draft Digital Content Metadata Standards

The draft digital content metadata standards
(https://www.orbiscascade.org/file_viewer.php?id=3171) support the first-priority goal assigned to the CCD Team by the Alliance's Board of Directors: Aggregate unique Alliance digital content. Adoption of minimal-level standards is the key to efficiently (e.g. without adding central staff to normalize metadata) aggregate digital content in any manner: in the Shared ILS, Digital Public Library of America, or anything else. Since there are no pre-existing standards, many institutions have developed very specific local practices as well as ones that tie into other partners with whom they work.

The standards were first created in fall 2014 by a working group associated with the IMLS National Leadership Grant with a specific focus on compatibility with DPLA's Metadata Application Profile and their emerging work on rights statements. The standards were extensively reviewed in fall 2015 by a group of Alliance metadata experts and revised to reflect new developments in the last year.

They were circulated to CCD Representatives and all other Alliance Teams through the Digital Content JWG starting November 9. A survey for feedback included questions on the required, required if applicable, and optional elements, the question of an input versus an output standard, metadata cleanup, and a general sense of support or not. Feedback was due from institutions and Teams by November 24.

Twenty-eight of the 37 Alliance member institutions responded. Of the Teams, only Systems provided formal feedback; the Collaborative Workforce Team provided a short email with overall support for standards.

If we relied only on the numbers, these draft standards could be adopted today. The number of respondents who were very satisfied or satisfied with the approach to the individual elements was high, and 25 of 29 respondents favored the adoption of these standards. However, a number of knowledgeable individuals put a great deal of thought into their narrative comments and indicated a need to rethink some major things and to make much more specific a number of other things. We will give that feedback the respect it deserves.

Resolving the question of input or output standard (or maintaining both) is the most important question. Many institutions were undecided as to input/output. There was good support for creating an output standard, but also providing elements of an input standard for optional use will assist institutions with dealing with specific systems and resolving the inevitable details. Whatever form the final standards take, there is an expectation that the standards will not simply be released by the Alliance, but that Alliance coordinated training and support will accompany the release of the standards. More specifically, interest in training or tutorials tied to the creation of appropriate output from specific DAM/IR platforms was requested (e.g. a tutorial specific to transforming Bepress OAI XML to Alliance-compliant XML). It was also noted that including

more examples--either incorporated into the standards or into training documentation--would assist institutions in applying an output standard appropriately.

While the decision between output or input standard (or the provision of both) is of primary importance (beyond the clarity of the standards themselves), critical questions about the application of the standards were also raised. As written, the standards are intended to apply to digital content writ large, but survey respondents identified important distinctions to be made that could affect their application of the standards. These distinctions fall broadly into two categories--content characteristics and policy considerations. On the content side, it was observed that the standards, and particularly required elements, need to address not just digitized physical content but also born-digital content--which it may not be possible (or desirable) to describe in the same way as digitized content. On the policy side, it is clear that local decisions will need to be made about (a) closed collections and (b) ongoing collections. To address the former, it was suggested that the standards only be required to apply to content that is intended to be shared or aggregated; individual institutions could choose to apply the standards to all content collections, but would only be expected to apply the standards to open content. The question of ongoing collections is less easily answered--for collections to which new content is being added on an ongoing basis, retrospective cleanup would be required in order to share/aggregate the entire collection. This reinforces the importance of addressing the existing question of what support, if any, should/will be available through the Alliance to support metadata cleanup (of any collections).

Ultimately, it appears that the most significant roadblock to adoption of the standards is in the details of the standard items. Detailed feedback was provided on clarifications/improvements that could be made to several fields--notably date, type, spatial, and source. Assuming that the consulted metadata experts are able to successfully address the concerns raised about these fields, adoption of the standards, especially as an output standard, should be broadly supported. Regardless of whether the Alliance's role as a DPLA service hub is established, or Alliance members elect to aggregate content in Alma/Primo, application of these standards will enable any future aggregation or shared discovery efforts for unique digital content in Alliance. They also provide an important starting point for further exploration of the ways in which unique digital content can be supported within the Alliance at the same level as purchased content. (For more details on the elements and the comments on needed modifications, see the attached survey report.)

Some expressed a desire to avoid creating additional metadata input fields to support aggregation (but that don't actually serve a local/user need).

Feedback indicated that rights statements should support fair use of materials. This desire is consistent with the direction of DPLA's set of rights statements, which both simplify statements and encourage use of materials where appropriate, which are included in the draft standards.

Respondents clearly desired more local or specific examples, essentially requesting that we support an optional input standard that assists institutions in use of local systems. Supporting

materials would need to include tutorials or documentation on using specific DAMs or IRs to produce the output needed for aggregation.