

Cyberinfrastructure Steering Committee
Minutes
November 26, 2014

Present: G. Agnew (chair), K. Ananthan, I. Beard, F. Fan, Y. Lin, R. Jantz, R. Marker, C. Mills, A. Morgan, C. Radick, M.B. Weber (recorder), R. Womack, K. White (via audiobridge)

Absent: J. Brennan, P. Konin, L. Langschied, T. Purger

Guest: L. Palumbo

1. AUL Report- G. Agnew

RUL has been part of the Kuali OLE design team for the last five years. Lehigh University and the University of Chicago have gone live with their OLE implementations.

In August 2014 the Kuali Foundation announced the creation of a new for-profit entity. It will continue to offer an open source product. It was a struggle to sell their enterprise system to universities, and they are hoping that the commercial version is more financially viable. Consortia will have to use the commercial product. The OLE development team is committed that OLE will remain open source. They have been tasked by the Kuali Foundation with determining how they want to interoperate with KualiCo, with a recommendation for how OLE will operate in the new infrastructure.

G. Agnew reported on the Kuali Days conference. A number of libraries are going live in December 2014 with Kuali implementations.

The statewide VALID Open Library System hosted two OLE workshops at RUL. One was conducted by Carl Grant. Traditional services are on the decline and VALID was encouraged to explore and begin implementing the next generation of library services, to replace services that are not being used, such as circulation of physical materials and traditional reference. Staff need to be redeployed, and academic consortia can be leaders for their institutions and their states. Other state consortia are doing more than New Jersey. The second VALID workshop was conducted by Liz Bischoff to discuss a possible organizational structure for VALID that will support faster and better deployment of OLE.

G. Agnew, D. Hoover, C. Sterback and G. Smulewitz participated in a VALID trip to Lehigh to see their OLE implementation. There will possibly be an RU trip at a later date. RUL is going live with OLE on July 1, 2016. Other members of VALID are considering the same or a close date.

TAS is hosting a visiting Fulbright Scholar from Belarus, Darya Zvanarova, who will be here for four months. She is interested in how institutional repositories support research. Darya will have an office in the SCC, and R. Marker will coordinate her schedule. Everyone in CISC is encouraged to meet with her.

2. RUcore Content Model for Dark Archives-R. Jantz

Some older SCC projects have been decommissioned and some will be preserved (Alcohol History, Polish General Social Survey, for example). These projects have been taken offline because we are unable to continue to support them.

The Dark Archives content model has two basic characteristics: it supports materials that are only publicly available via an access copy external to RUCore and is intended only to preserve and restore the access copy as needed. Ingest uses the directory option. Each title is ingested as a separate archival master with technical metadata.

The dark archive is needed to provide full preservation for the long term for resources that are intended to persist but that are created and maintained for accessibility outside the RUCore preservation platform. A prime example is the Open Journal System journals that are created and managed by RUL outside RUCore. While these are backed up, preservation strategies that include metadata and preservation formats are not implemented. RUCore provides the structure and tools for complete archiving, including monitoring and replacement of master files. It was decided to create a dark archive for these resources in RUCore but to ensure that they are not publicly discoverable and accessible, as opposed to creating a separate dark archive. RUL is participating in collaborative dark archiving strategies such as DPN and in collaborative identifier assignment (DataCite DOI) and it was felt to be too complex to maintain two archives that are interoperable with other initiatives or to engage in tasks such as preservation format migration across two separate archives. The problem of ensuring that dark archive resources are available only to RUCore administrators had to be addressed, however. The dark archive content model addresses developing a standard for the information needed to restore access copies on their host platform as well as the strategies needed to ensure resources are not indexed and discoverable on the public facing user interfaces. The content model supports the datastreams needed to preserve and restore access copies on their host platforms such as multiple preservation formats (e.g., PDF and Base 64 exports for OJS), as well as a metadata application profile to support management and restoration of the files. In addition, ancillary information, such as readme files that are needed for restoration by future RUCore administrators. A readme file when provided will be documented in a descriptive documentation event. A content model is used instead of a collection because collections are associated with ownership and the resources in the dark archive are owned by the creators of the collection. For example, the editor who has entered into an MOU with Rutgers to publish the journal is the owner of that journal and will be acknowledged as such in metadata.

The issue of embargoes was raised. If a resource is embargoed past our lifetimes, it is a dark archive. However, while one of the characteristics of a dark archive—the inaccessibility for public use—is not, the other category, that of an authoritative access platform is not. Embargoed resources, even when indefinite, are still intended for ultimate access via RUCore. In most cases, full metadata is available, for example for a sensitive dissertation that might not be made available but that can still alert other researchers through the metadata that research already exists. Researchers can then select a different topic or contact the author to arrange for one on one access to all or part of an embargoed resource. This is substantively different than the intent of the dark archive, which is to provide preservation for resources where access is hosted outside the repository.

The Dark Archive will initially be populated with OJS journals published by Rutgers. It will be a generic application that will eventually be applied to other collections, but customized as needed to support other resource types, such as directory databases and websites. .

Each archival object will get a DOI. The DOI with associated metadata may be discovered through an open DOI registry. We want people to discover these resources, but to access them through the

appropriate access interface. A note in the metadata will indicate that this is a preservation resource only and not available for public access and will provide a link to the public access resource. A descriptive event will make clear to future administrators the relationship between the preserved resource and the access copy.

A question was raised about whether two DOIs are being provided for the same resource. Agnew clarified that the preservation and access copies are not the same identical resource. The preservation resource may have a different format, different navigation and will be bundled with readme files or tools necessary for file restoration.

The data stream needs to be faithful and a routine is needed to make sure that the archival master is accurate. We are not keeping a paper trail. The DOI will be open, and we will not offer accessibility in RUcore. When the process is implemented, it will be reviewed to ensure it is working effectively.

Content models behave in a certain way. Specs are needed as to how the WMS treats this content model when a user selects one of these resources. The Dark Archive content model and metadata will be tested in spring 2015 and will be vetted by MDWG and CISC. There are possibilities for using the Dark Archive content model for data as it has broad applicability.

R. Jantz will revise the document and post it to Sakai.

3. Restricted Access to RUcore Collections and Resources- R. Marker

A summary of access options is needed. Currently we have just embargo and Open Access, and need to add Dark Archive. There are a variety of restrictions that need to be codified into our documentation. The document R. Marker distributed includes an option to restrict by IP address or range, restrict to Rutgers netID users, restrict to defined groups or individuals, and restrict from discovery by search engine. Access options were raised following copyright questions. The current practice is to embargo and we are amenable to loosening the restrictions.

The document provides strategies that may be combined or implemented at various levels or on a case by case basis. How they are used will depend on the circumstances.

Other instances where access may be restricted include platforms (such as a mobile application to access resources), and restricting usage after a predetermined number of uses.

R. Marker will revise the document and post it to Sakai.

4. Draft Report Research Data Service Guidelines- L. Palumbo

The Task Force on Research Data Implementation has suggested that we begin with mediated acceptance of data. They reviewed thirty-seven other repositories as an environmental scan. Many of these repositories are not permitting self-deposit.

The task force recommends that Phase I of implementation begin with mediated data acceptance. It will be the researcher's responsibility to check the data to determine whether it is appropriate to be publicly shared. To avoid conflict, we won't accept data derived from human or animal subjects.

Projects in all disciplines and regardless of funding status will be considered. A suggestion was made to add the statement “Preference will be given to data mandated by a government agency to be publicly available”.

Projects will be restricted to Rutgers-led initiatives with a Primary Investigator (PI). We won't accept projects in which Rutgers personnel are subcollaborators.

The task force suggested that total data volume for Phase I will be 100 GB or less per project. After discussion, it was agreed that we can't turn away research and anything that exceeds 100 GB will be referred to the RUcore Collection Manager for review.

Phase I includes a recommendation about short-term embargoes. At CISC's suggestion, this will be deferred to a later phase.

The task force had suggested that Phase II address self-deposited and mediated data acceptance, but Cabinet rejected this suggestion.