Towards Interoperable Preservation Repositories (TIPR)
IMLS Interim Report 2
April 1, 2008 – September 30, 2009

1. Progress towards activity targets

The TIPR project is proceeding according to the schedule projected in the grant application, shown in Figure 1 below.

Hold partner meetings: The third partner meeting was held at Cornell University on June 30, 2009, with 7 participants attending. In addition, partners have been holding bi-weekly conference calls.

Spec transfer format & use cases: The transfer format has been renamed the Repository Exchange Package or RXP. Version 0.89 of the specification is available on the TIPR website (wiki.fcla.edu:8000/TIPR). Use cases were finalized at the partner meeting in November, which is outside the scope of this reporting period.

Coding: In process. Two of the three partners have done the programming required to export an RXP for sample Archival Information Packages (AIPs) in their own repositories, and to ingest an RXP sent by another repository. Programming for Cornell’s repository has been delayed because the library is migrating to a new preservation system. Schematron validation routines were written for standalone PREMIS XML, METS XML, and RXP XML, and scripts were written to validate RXP files against BagIt, Schematron and the PREMIS and METS schemas.

Transfer testing and analysis: Two partners have tested the export of a native AIP in RXP format, and the ingest of a foreign RXP. A round-trip transfer test has also been accomplished: An AIP stored by FCLA was disseminated as a DIP, transformed into an RXP, transferred to NYU, ingested into an AIP by NYU, disseminated as a DIP and transformed into an RXP by NYU. (Future testing will include a re-ingest step by FCLA.) The test transfer raised a number of issues subsequently discussed at the November partner meeting.
2. Project Findings and Accomplishments

The initial specification for the RXP (previously called the TIPR Exchange Package, or TEP) was enhanced and refined based on trial transfers. The decision was made to require all PREMIS identifiers to be URIs to ensure uniqueness. Identifiers which are not currently URIs can be registered as info:URIs if need be. A history of the RXP package identifiers is maintained in package-level provenance, expressed using PREMIS Events.

A major technical stumbling block was resolved in the June partners meeting. At issue was how to associate PREMIS object identifiers with file objects as identified by flocat in METS. The specification was modified to require an OWNERID attribute in the METS file element containing the PREMIS object identifier, to serve as a link between the two schemes.

Discussions progressively clarified what information should be contained within each exchange package and which information would be better documented in an inter-repository service agreement. An outline of information recommended for a service agreement was drafted and posted to the website. It includes such topics as: RXP profile options used, RXP transfer logistics, target repository actions upon RXP receipt, inter-repository rights and permissions agreements, expected archiving and preservation treatment at the target repository, financial arrangements between source and target repositories, and legal aspects of the inter-repository relationship.

Although dissemination activities were planned for the second year of the grant, we took advantage of the DLF Spring Forum in May to give an overview of the TIPR project. The presentation by Joseph Pawletko is available on the project website.

3. Other findings
Rights in particular have presented a number of practical problems. First, although project partners agreed an expression of rights at the package-level was essential, they were uncertain if there was a legal basis for a sending repository to assert intellectual property rights over a package other than those adhering to the content as originally archived. In May we consulted with June Besek, the executive director of the Kernochan Center for Law, Media and the Arts at Columbia Law School and coauthor of the four-country report, \textit{International Study on the Impact of Copyright Law on Digital Preservation} (July 2008).

Ms Besek explained that it was unlikely a repository would have any intellectual property rights in an RXP by virtue of copyright law. One can have rights in a compilation of pre-existing works only if there is creativity in the selection and/or assembly of those works, which would not be true in this case. Additional metadata created for the package would also not be protected if it consisted only of facts. However, if the repository already had the right to reproduce, create derivatives of, or distribute the contents of the package, the repository could use contract law to restrict those rights when distributing the package to others. She used as a parallel example an archive that holds a manuscript that is in the public domain. Although everyone has equal rights to the manuscript under copyright, the archive can put conditions on its use, such as allowing access only during certain hours and prohibiting copying.

Given this information, it was clear that the source repository could want to a) communicate its own rights in the package, and/or b) communicate any restrictions imposed on the receiving repository. As this could vary from one package to another depending on the content, it is reasonable to express these as package-level rights, as opposed, say, to a general declaration of rights expressed in a service agreement. This in turn led to a second problem, that the most likely vehicle for expressing package level rights, the PREMIS rights entity, is not applicable at the package level.

According to the PREMIS data model, the highest level at which PREMIS metadata can be expressed is the representation. An RXP, however, can contain one or more representations. The closest entity to the RXP in the data model is the Intellectual Entity, but PREMIS Rights are not associated with it. (The problem also arises in regard to digital provenance – PREMIS Events.) In May 2009 we took a request to the PREMIS Editorial Committee (EC) to consider extending the PREMIS entity model to include intellectual entities in order to enable application of PREMIS Rights and Events information to intellectual entities. The British Library noted they have a similar situation on ingest when they receive bundles of representations and objects from the same source, e.g. from the same e-publisher. They treat an intellectual entity as a preservation object and would like to extend PREMIS to apply at that level.

At the time, the EC suggested that a descriptive metadata standard such as MARC 21 or MODS could be used to record both rights and preservation events; for example, events could be entered in MARC field 583 "Action Note." In subsequent meetings, however, the EC agreed to raise the question at the upcoming PREMIS Implementation Fair in November.
In the meantime, the TIPR partners decided to proceed as though it would be possible in the future to use PREMIS rights at the package-level, leading to yet a third problem. None of the participants used PREMIS rights within their own repository systems or felt they had enough rights experience to use PREMIS rights in a granular way. At the same time it was noted that the Rights Entity had undergone a major revision in version 2 of the PREMIS Data Dictionary and might still be in flux. Consequently partners agreed to use Rights metadata in a valid but limited way, to point to license terms held externally.