Pathways Core: A Data Model for Cross-Repository Services

Jeroen Bekaert
Ghent University
Faculty of Engineering
Jozef Plateaustraat 22
9000 Gent, Belgium
+32 9 264 3911
jeroen.bekaert@ugent.be

Xiaoming Liu,
Herbert Van de Sompel
Digital Library Research & Prototyping
Research Library
Los Alamos National Laboratory
+ 1 505 667 1267
{liu_x,herbertv}@lanl.gov

Carl Lagoze, Sandy Payette,
Simeon Warner
Cornell Information Science
301 College Ave
Ithaca, NY 14850, USA
+1 607 255 9555
{lagoze,payette,simeon}
@cs.cornell.edu

ABSTRACT
As part of the NSF-funded Pathways project, we have created an interoperable data model to facilitate object re-use and a broad spectrum of cross-repository services. The resulting Pathways Core data model is designed to be lightweight to implement, and to be widely applicable as a shared profile or as an overlay on data models currently used in repository systems and applications. We consider the data models underlying the Fedora, Dspace and aDORe repository systems, and a number of XML-based formats used for the representation of compound objects, including MPEG-21 DIDL, METS, and IMS/CP.

At the heart of the Pathways Core data model (Fig. 1) are the *entity* and *datastream* elements. *entity* elements model the abstract aspects of digital objects and align with works and expressions in FRBR [1]. An *entity* can model anything from a digital object to a collection of digital objects (other *entities*), to a node created merely to express abstract properties. Core properties of *entities* are *hasIdentifier*, *hasProviderInfo*, *hasLineage*, and *hasProvider-Persistence*. If a repository attaches *providerInfo* to an *entity*, it provides a handle to access the *entity* from the repository, supporting its use and re-use. Persistence of this handle may be indicated with *providerPersistence*. The *hasLineage* property is used to indicate the *entity* (or *entities*) from which the *entity* to which the *hasLineage* is attached was derived. Other properties, such as *hasSemantic,* that convey the intellectual genre of the *entity* (i.e. journal article), can be added. *datastream* elements model the concrete aspects of a digital object; these align with items in FRBR, and can be thought of as aspects at the level of bitstreams. An *entity* may have any number of *datastreams*. Two properties of *datastream* have been defined as part of the Pathways Core: *hasLocation* conveys a URI that can be resolved to yield a bitstream; and *hasFormat* conveys the digital format of the bitstream. If a *datastream* has multiple *hasLocation* properties, resolution of the conveyed URIs yields bit-equivalent bitstreams.

The Pathways Core data model can be serialized in a variety of ways, and, an RDF serialization as well as a profile of MPEG-21 DIDL have been created as reference implementations. We have also conducted the following experiment to illustrate the power of the Pathways Core. A number of heterogeneous repositories implemented an OpenURL-based *obtain* interface from which, given the *providerInfo* of an *entity*, an RDF serialization of the *entity* compliant with the Pathways Core could be retrieved.

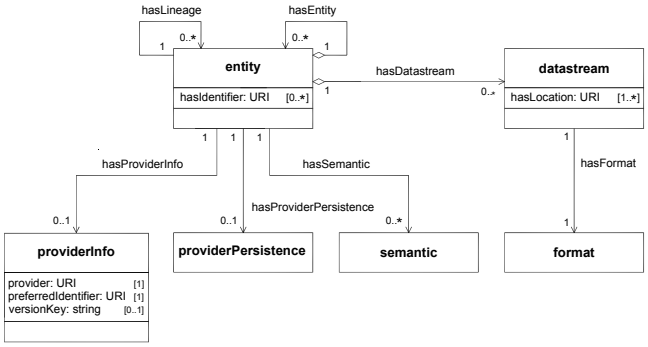

Figure 1 – UML structure diagram of the Pathways Core data model

Using this interface, an overlay journal can collect serializations of some *entities* (scholarly papers) from the different collaborating repositories, and assemble those into a new issue of the journal. The overlay journal then itself implemented the same *obtain* interface, and as a result, an RDF serialization of the entire journal, an issue, and an article could be extracted. This interface could then, for example, be used by a preservation repository to collect content from the overlay journal for ingest and mirroring. This experiment illustrates how cross-repository services and workflows can be facilitated through support of an interoperable data model (the Pathways Core) and an interoperable service interface (the OpenURL-based *obtain* interface).

Categories and Subject Descriptors
H.3.7 [Digital Libraries]: standards; system issues

Keywords
Data model, interoperability, scholarly communication

ACKNOWLEDGMENTS
This work was supported by NSF award number IIS-0430906.

REFERENCES
IFLA Study Group on the Functional Requirements for Bibliographic Records. (1998). *Functional Requirements for Bibliographic Records*. UBCIM Publications-New Series. Available at http://www.ifla.org/VII/s13/frbr/frbr.pdf

JCDL'06, June 11-15, 2006, Chapel Hill, North Carolina, USA.
ACM 1-59593-354-9/06/0006.