Customized OAI-ORE and OAI-PMH Exports of Compound Objects for the Fedora Repository

Alessia Bardi, Sandro La Bruzzo, and Paolo Manghi
Istituto di Scienza e Tecnologie dell'Informazione,
Consiglio Nazionale delle Ricerche, Italy

In order to address the needs of modern researchers to share and access research outcomes, Digital Library Systems (DLSs) evolved to manage new types of documents and collections which surpass the traditional publication-metadata document model to incorporate further entities involved in the research life-cycle. For example, *enhanced publications* [1] enrich the representation of traditional publications with other objects, such as metadata descriptions, research datasets, cited publications, etc. To cope with such representational requirements, modern Digital Library Management Systems (tools supporting DLS construction, ref. DELOS Digital Library Manifesto [2]) tend to offer *data models* capable of expressing DLS *document models* as "labeled graphs of information objects". As a consequence, DLSs evolved also in the way they export their objects to third-party applications. Typically, objects are organized into "packages of interlinked information objects", known as *compound objects*, and exposed through the standard APIs of the Open Archives Initiative protocols, which enable single or bulk access to compound objects respectively as OAI-ORE (RDF) aggregations [3] and OAI-PMH (XML) metadata records [4].

The Fedora Repository is a well-known DLMS platform [5], whose object data model is designed to represent arbitrary DLS document models by encoding them as graphs of *content model* objects, i.e., special Fedora objects declaring the "type of the objects" in a Fedora instance, i.e., the expected structure, relationships, and behaviors (Fedora methods) of the objects. This paper presents *OAIzer*, an extension of Fedora which is capable of exporting compound objects conforming to a given portion of the underlying DLS document model through the OAI-PMH or OAI-ORE protocols. The component implements a mechanism based on the concept of *"OAI view of a Fedora document model"*, that is a "sub-structure" of the document model which developers can provide to customize the shape of their compound objects. OAIzer interprets OAI views to automatically deploy web APIs capable of exporting compound objects compatible with the given structure and according to the preferred OAI protocol.

OAIzer is compared with other solutions for exporting compound objects in Fedora, namely *oreprovider*[1] and *Fedora2ORE*[2]. The former adopts an *object-oriented* approach, where ORE aggregations consist of sets of Fedora objects, the sets being identified by adding to the objects a pointer/reference to the relative ORE aggregation. The latter adopts instead an *object navigation-oriented* approach, where ORE aggregations consist of one Fedora object together with the objects reachable by navigating its relationships up to a given depth. The first solution is easy-to-use, but binds the aggregation logic to DLS applications and disregards document model relationships between objects. The second solution is independent from DLS applications and generates aggregations by following relationships between objects, but DLS developers can only define the boundaries of aggregations in terms of navigation depth rather than in terms of their preferred document model "sub-structure". Finally, both solutions do not support OAI-PMH exports of compound objects.

---

[1] *OREprovider* Fedora module, http://oreprovider.sourceforge.net/index.html
[2] *Fedora2ORE*, http://trac.eco4r.org/trac/eco4r/wiki/Fedora2ORE

References
1. Enhanced Publications, http://www.surf.nl/en/themas/openonderzoek/verrijktepublicaties/Pages/default.aspx, SURF Foundation
2. Candela, L., Castelli, D., Pagano, P., Thanos, C., Ioannidis, Y., Koutrika, G., Ross, S., Schek, H.J., Schuldt, H.: Setting the Foundations of Digital Libraries: The DELOS Manifesto. D-Lib Magazine 13 (2007)
3. Lagoze, C., Sompel, H. V. d., Johnston, P., Nelson, M., Sanderson, R., & Warner, S.. *Open Archives Initiative Object Reuse and Exchange*, http://www.openarchives.org/ore/1.0/toc.html (2008).
4. Lagoze, C., Van de Sompel, H.: *The open archives initiative: building a low-barrier interoperability framework*. In Proceedings of the first ACM/IEEE-CS Joint Conference on Digital Libraries, pp. 54–62. ACM Press (2001).
5. C. Lagoze, S. Payette, E. Shin, C. Wilper, *Fedora: an architecture for complex objects and their relationships*, International Journal on Digital Libraries, pp. 124-138, Vol. 6, Issue 2, (2006).

Bionotes

Alessia Bardi received her MSc in Information Technologies in the year 2009 at the University of Pisa, Italy.
She is a PhD student in Information Engineering at the Engineering Ph.D. School "Leonardo da Vinci" of the University of Pisa and works as graduate fellow at the Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" (ISTI), Consiglio Nazionale delle Ricerche (CNR) of Pisa, Italy. Alessia is currently involved in EC funded projects for the aggregation, curation and export of library, archival and museum digital objects and metadata records. Her research interests include Digital Library Management Systems, data interoperability, compound object management, and service-oriented infrastructures.
Email: alessia.bardi@isti.cnr.it

Sandro La Bruzzo received his MSc in Information Technologies in the year 2010 at the University of Pisa, Italy. He is working as graduate fellow at the Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" (ISTI), Consiglio Nazionale delle Ricerche (CNR) of Pisa, Italy. His research interests include Digital Library Management Systems, compound object management, and Service-oriented Data Infrastructures.

Paolo Manghi received his PhD in the year 2002 from the Dipartimento di Informatica of the University of Pisa, Italy. He is presently working as a researcher at the Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo", Consiglio Nazionale delle Ricerche (CNR) of Pisa, Italy. His research interest include Data Models for Digital Library Management Systems, Types for Compound Objects, data curation in Digital Libraries, service-oriented ICT infrastructures with special focus on data ICT infrastructures.
Email: Paolo.Manghi@ISTI.CNR.IT

CUSTOMIZED OAI-ORE AND OAI-PMH EXPORTS OF COMPOUND OBJECTS FOR FEDORA REPOSITORIES

Alessia Bardi, Sandro La Bruzzo, Paolo Manghi
name.surname@isti.cnr.it
http://nemis.isti.cnr.it/groups/infrascience

Fourteenth International Conference on Grey Literature - Rome, 2012
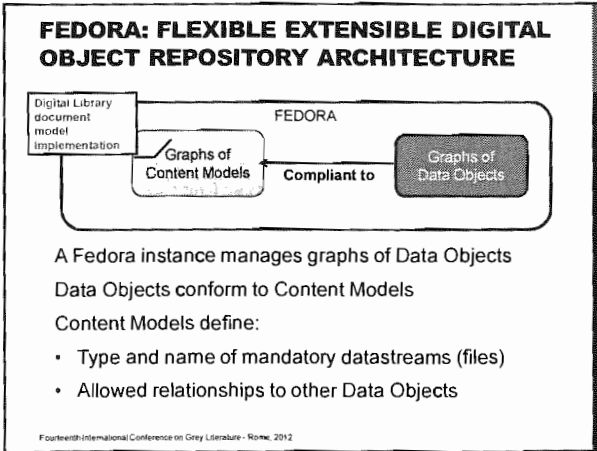

SCENARIO AND MOTIVATIONS

- Digital Library Systems (DLSs) need to export content
    - Export of compound objects: packages of information objects with an identity
    - OAI protocols (typical solution)
- DL management systems (DLMSs) need to support export protocols
- DLMS issues:
    - Absence of OAI-exports: some DLMS do not provide support for OAI protocols (Relational Databases)
    - Pre-defined OAI exports of compound objects: shape of objects as exported by OAI protocols is pre-defined by the DLMS (Fedora)
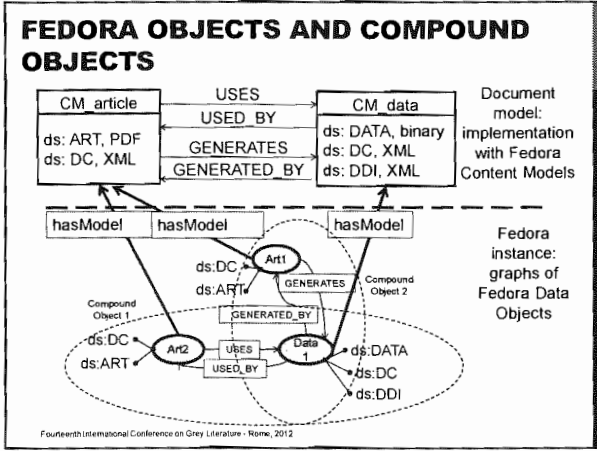
Fourteenth International Conference on Grey Literature - Rome, 2012

OBJECT EXCHANGE PROTOCOLS

	OAI-PMH	OAI-ORE
Exchange subject	Metadata	Resource
Graph-based data model	No	Yes
Data Model	Set Item MetadataPrefix	Aggregation Aggregated Resource Proxy
Protocol	6 OAI Verbs on HTTP	HTTP
Data Serialization	XML	RDF (XML, ATOM)
Deleted records management	Yes	No

Fourteenth International Conference on Grey Literature - Rome, 2012

FEDORA: FLEXIBLE EXTENSIBLE DIGITAL OBJECT REPOSITORY ARCHITECTURE

A Fedora instance manages graphs of Data Objects

Data Objects conform to Content Models

Content Models define:

- Type and name of mandatory datastreams (files)
- Allowed relationships to other Data Objects

Fourteenth International Conference on Grey Literature - Rome, 2012

FEDORA OBJECTS AND COMPOUND OBJECTS

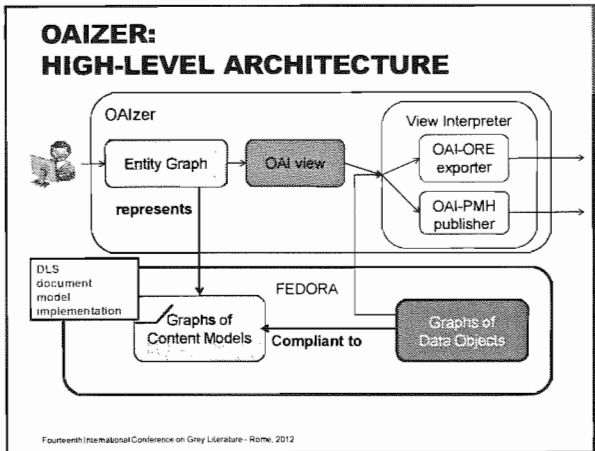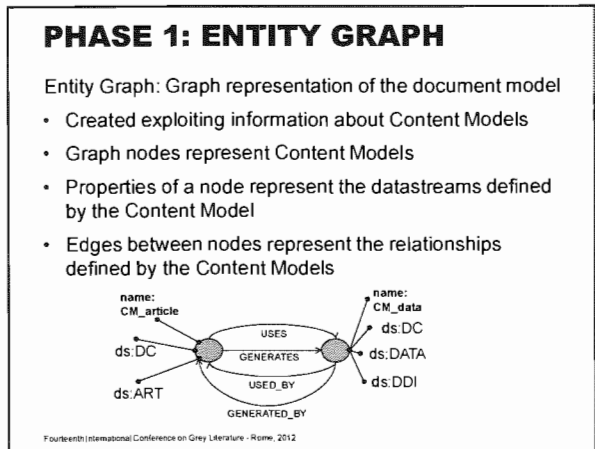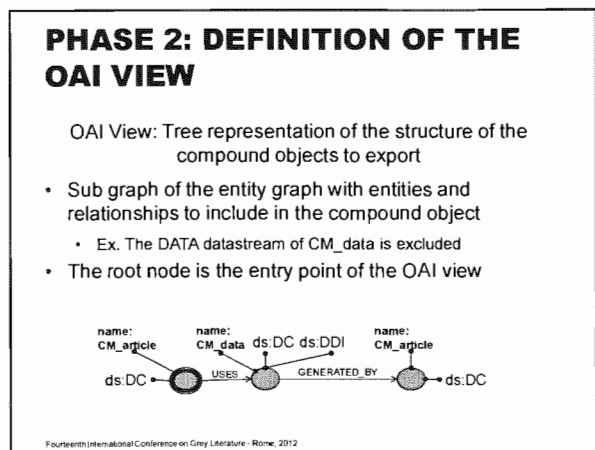Fourteenth International Conference on Grey Literature - Rome, 2012

EXPORTING COMPOUND OBJECTS IN FEDORA – SOLUTIONS AND LIMITS

- Current solutions for OAI-PMH exports on Fedora...
  - ...export OAI-items with the granularity of a Fedora Data Object
  - ...support the export of XML contained in datastreams
- Current solutions for OAI-ORE exports on Fedora...
  - ...exclude relationships and their semantics
  - ... allow limited customization of the structure of the compound objects (e.g., navigation depth)
  - ...pre-define the structure of the exported objects
- OAIzer works on any Fedora instance with Content Models:
  - ORE exports include relationships and their semantics
  - Structure of exported objects is fully customizable
  - OAI-PMH metadata formats can be generated on request (requires the configuration of an appropriate transformation rule)

Fourteenth International Conference on Grey Literature - Rome, 2012

24

OAIZER:
HIGH-LEVEL ARCHITECTURE

Fourteenth International Conference on Grey Literature - Rome, 2012

PHASE 1: ENTITY GRAPH

Entity Graph: Graph representation of the document model

- Created exploiting information about Content Models
- Graph nodes represent Content Models
- Properties of a node represent the datastreams defined by the Content Model
- Edges between nodes represent the relationships defined by the Content Models

Fourteenth International Conference on Grey Literature - Rome, 2012

PHASE 2: DEFINITION OF THE OAI VIEW

OAI View: Tree representation of the structure of the compound objects to export

- Sub graph of the entity graph with entities and relationships to include in the compound object
  - Ex. The DATA datastream of CM_data is excluded
- The root node is the entry point of the OAI view

Fourteenth International Conference on Grey Literature - Rome, 2012

PHASE 3.1: INTERPRETATION OF THE OAI VIEW FOR OAI-PMH

- Compound object represented as OAI-Item with several metadata formats

PHASE 3.2: INTERPRETATION OF THE OAI VIEW FOR OAI-ORE

- Compound object exported as OAI-ORE Aggregations.
- Aggregated resources in OZML format

CONCLUSIONS

- OAIzer's wizards guide DLSs developers through the realization of customized, domain-dependent exports of compound objects via OAI-PMH and OAI-ORE
- Future work:
  - Better graphical support for the definition of views in case of entity graphs with cycles
  - Implementations for different back-ends
    - RDBMS: ongoing work