Remember, Restructure, Reuse - Adding value to compound scholarly publications in a digital networked environment *

Boulal A. , Iordanidis M. , Quast A.

boulal@hbz-nrw.de, iordanidis@hbz-nrw.de, quast@hbz-nrw.de

Abstract In this paper we present a formal technical approach which structures, disseminates and reuses complex digital objects of potentially any format. The applicability to existing scholarly publications systems is paramount for the *eco4r* project in which these results were achieved. The technical approach introduced utilizes high level descriptions of complex digital objects which are used for the exchange of contextual aggregations and may be reused by digital preservation technologies. Practical outcomes of the approach introduced in this paper are plug-ins for the repository systems Fedora and OPUS and an Overlay Journal demonstrator, respectively.

Introduction

Within the last decade, the nature of scientific publications has changed increasingly from monolithic to more complex digital objects. New techniques, guidelines and frameworks have emerged, enabling researchers to enhance their publications with any kind of supplementary digital material. Thus, researchers can provide online access to their scholarly works along with any associated material and results illustrating the entire discovery process. In a best-case scenario, this leads to more comprehensible and documented publications. Scientists can provide intermediate outcomes and materials during research in order to illustrate the genesis of their final results.

Scientific publications then appear as bundles of digital objects including text, visualizations, research data, supplementary materials and any kind of other digital

* in: Anderson D., Delve J., Dobreva M., Konstantelos L. The Preservation of Complex Objects (Volume 1): Visualisations and Simulations (forthcoming) . (Series Editors)

objects. In turn, they bring along an internal complexity of object-to-object relationships and object level metadata, which needs to be handled by digital research environments. Electronic publication systems are challenged by this increased complexity within their data sets, as they have to ensure that all parts of the publication are provided in correct structure, context and meaning. For instance, a file system randomly storing files in a directory cannot reproduce the relations between files named “data1.xsl”, “data2.xsl” and their combined results in a file called “result_data.xsl”. Thus, much of the semantic information stored on the level of data structure is being lost.

Instead, it is recommended to store data in a structured way in order to capture the components, their metadata information and the relationships between them. Ideally, any existing infrastructure in the field of scholarly publishing will be adapted and expanded for this purpose. In this article, we will use the term Compound Scholarly Publication (CSP) to describe complex forms of scientific publications, although we are aware of the existence of many other terms for similar publications (for more details see the Defintion and Scope section).

Several organizations and projects (DRIVER[1], Europeana[2], SURFfoundation[3], OAI-ORE[4]) have addressed the issue of Compound Scholarly Publications. However, the domain of complex scientific publications is still in its infancy. The current infrastructure for academic research still focuses on storage and dissemination of individual resources. This is because many repository systems do not provide functionality to create, store and manage Complex Objects. On the other hand, the creation of complex publications starting from existing data can be difficult, time-consuming and costly. Since practical implementations within existing repository infrastructures are not well established yet, we put special emphasis on the availability of software plug-ins in the eco4r[5] project.

Technical approach

The aim of the eco4r project is to address and meet practical requirements in the field of Compound Scholarly Publications. The project examines the impact of complex shapes of publications on existing information systems, on semantic technologies such as Linked Data[6] as well as on Long Term Preservation.

---

[1] http://driver-repository.eu

[2] http://www.europeana.eu/portal

[3] http://www.surffoundation.nl

[4] http://www.openarchives.org/ore

[5] http://www.eco4r.org

[6] http://www.w3.org/DesignIssues/LinkedData.html

The eco4r approach was started with a context analysis of existing complex objects in two exemplary scholarly repositories, Fedora[7] and OPUS[8], the latter being a wide-spread repository system at German universities. Based on these results, we have formulated guidelines for generating OAI-ORE resource maps from existing repositories (Boulal et al, 2010). At this early stage, resource maps were created manually in order to discover the requirements of data completeness as well as the technical possibilities to represent them. From the perspective of digital preservation, this first stage of analysis addresses the definition of significant properties at a structural level. Secondly, we developed a data model for Compound Scholarly Publications on the basis of OAI-ORE and the FaBiO ontology. Based on this data model, we created linked open data sets which can aggregate data that originates from several repositories and their resource maps.

In order to display aggregated content from multiple resources we recognized the demand for an integrated visualization layer. Being within the domain of scholarly publishing, we developed components for an 'Overlay Journal'[9] that dynamically creates topical data sets according to DDC classification. These components are as follows:

- OAI-ORE plug-in for generating resource maps from Fedora repositories
- OAI-ORE plug-in for generating resource maps from OPUS repositories
- OAI-PMH based harvester for OAI-ORE Resource Maps
- RDF triple store for storing the Resource Maps.

A proof of concept for the Overlay Journal will be available by the end of the eco4r project.

Definition and Scope

Several terms have been used to describe identical or very similar concepts of Compound Scholarly Publications in technical literature. The term Enhanced Publications (Place et al., 2008) was introduced to describe publications which combine heterogeneous, but related web resources. Other researchers propose the concept of Enhanced E-Theses (Ruijgrok, Slabbertje, & Van Luijt, 2009) and continue the concept of traditional academic publications. Both terms refer to the idea of a primary publication - i.e. in terms of a full text - which is enhanced by supporting materials. Our approach is based on the concept of aggregated web resources which each may act as a primary publication themselves.

[7] http://www.fedora-commons.org

[8] http://www.opus-repository.org/index.html

[9] http://www.earlham.edu/~peters/fos/guide.htm#overlay

Since we think that CSPs represent a special subset of Complex Digital Objects, we use the term to describe an aggregation of distributed web resources relevant to scholarship.

Like any other Complex Digital Object, a CSP is considered as a logical unit with boundaries defined by the referenced resources. CSPs might be created explicitly by operators of publication systems, dynamically by automated web services or, if supplied with aggregation editing systems such as SCOPE [10], manually by researchers themselves. The component parts of an aggregation are interlinked with each other through semantic relationships. Each part is either a conceptual construct (e.g. a conference) or a more concrete entity (e.g. a text, image, video file or visualization) stored on a web server. They may have different content and semantic types, vary in their manifestations or mime types and can be distributed over different locations on the web.

So, what are the implications of these sophisticated digital objects on digital preservation, academic research, digital publishing and semantic web technologies? Before addressing these questions, it is crucial to understand how to construct, organize and exchange CSPs.

Standards for representing Complex Digital Objects

To the human eye, a website containing publications with associated materials and metadata using hyperlinks seems to be an appropriate presentation format. However, software services cannot interpret information given in a web page the same way a human reader does. For example, a search engine cannot distinguish between raw data and an instructive video that is to be found at the end of a conventional hyperlink – certainly not without additional semantic information.

In order to give human users as well as software services profound information about contextual content, Complex Digital Objects must be represented as comprehensively as possible. This requires functionality that allows for interpretation of inherent relations, file types, semantic content models and unique identification schemes (Boulal et al, 2010).

In this chapter we will briefly present two standards for data representation (OAI-ORE[11] and METS[12]) and evaluate their applicability to software services such as data exchange and Long Term Preservation.

Open Archives Initiative Object Reuse and Exchange: OAI-ORE

[10] http://www.ijdc.net/index.php/ijdc/article/download/84/55

[11] http://www.openarchives.org/ore/1.0/

[12] http://www.loc.gov/standards/mets

OAI-ORE defines a standard for the description and exchange of aggregated web resources. It introduces an abstract data model on top of the RDF model, in which each of four main entities is represented by a Uniform Resource Identifier (URI) (Van de Sompel, 2009).

An OAI-ORE-Resource Map (ReM) describes an Aggregation (A) of Aggregated Resources (AR) and retains all information about it, including provenance information, technical and rights metadata as well as corresponding structural information. As OAI-ORE is compliant with RDF, a ReM can be serialized in different formats such as RDF/XML, Atom and N3 which makes it applicable to a broad range of data exchange on many digital platforms. Furthermore, the OAI-ORE data model supports nested Aggregations. They are able to represent Aggregations of *other* Aggregated Resources and thereby cover a typical use case in publication infrastructures for e-journals. OAI-ORE also allows the extension of its data model with third-party-vocabularies and ontologies.

Metadata Encoding and transmission Standard: METS

METS was primarily created in the digital libraries and archiving environment. It provides a framework to wrap several metadata types and, at times, an arbitrary number of file formats stored as byte stream. A METS file can be considered a container format that typically holds descriptive, administrative, structural and technical metadata related to a digital object. A unique internal identifier references each content item in a METS section. Interlinking the sections within a METS document provides the structure of a Complex Object described in METS. The content of a section may either be stored inside the METS document or held externally and referenced from a main METS file. METS' strong abilities as a wrapper format and its capability to store binary data make it well compatible with the OAIS Reference Model[13]. It can be used for creating packages for submission (SIP), archival storage (AIP) and dissemination (DIP) within a digital archive.

The OAI-ORE and METS frameworks are both well-suited for representing Complex Digital Objects. However, they put emphasis on different purposes. While the OAI-ORE framework offers ideal tools for the interoperability of web resources within networked semantic services, METS is a convenient format for packaging binary data along with its metadata.

OAI-ORE can turn Complex Digital Objects, such as those stored in repositories, into reusable and exchangeable web resources. As OAI-ORE relies on a standardized and flexible RDF-based data model, it became the framework of choice for tasks using the Linked Data approach. By comparison, METS appears to be an ideal preservation format. However, the preservation of dynamically changing

[13]http://public.ccsds.org/publications/archive/650x0b1.pdf

data sets raises new questions about versioning and the idea of an 'ideal time' for a preservation snapshot.

Remember

From the definition given above, we have concluded that Compound Scholarly Publications are a subset of Complex Objects. The statement has certainly an impact on the way objects are handled in Long Term Preservation. In addition to the conventional features of web resources or other digital objects relevant for preservation, even more criteria have to be considered. These include the object structure as well as the technical and semantic accessibility of each component within the Complex Object.

Compared to self-contained scholarly publications, CSPs require additional efforts in preservation. At the same time, they entail crucial improvements for the preservation of scholarly works as well as other Complex Digital Objects. (Cheung et al., 2008) note that scientific publications „*inadequately represent the earlier stages* [of the scientific process] *that involve the capture, analysis, modelling and interpretation of primary scientific data*".

This statement implies that a publication does not merely consist of final results, but rather of an aggregation of outcomes and results from the discovery process. In this sense, a publication goes through multiple stages until it becomes a "final" version. Each stage of research produces one or more outputs. Capturing these outputs in a well-structured form is likely to be interesting to a number of stakeholders (i.e. researchers, historians, archivists, authorities) concerned with the genesis of scholary works.

Compound Scholarly Publications may adapt some of the changeable nature of web resources. Used within digital research environments, they are suited for representing the discovery process close to real-time. Within the eco4r project, the actual appearance of Compound Scholarly Publications is represented within Resource Maps. They may be overwritten periodically or maintained for reuse at any desired time (**Fig.** 1).

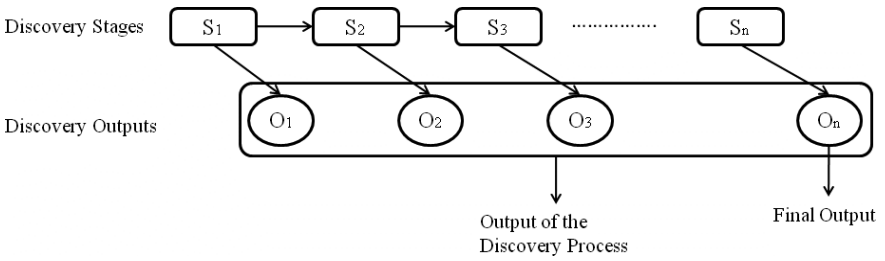

Fig. 1. Discovery stages reflecting the discovery process

In order to benefit from the information provided within CSPs on the level of Long Term Preservation, some key information must be contained within them. The next section will elaborate on the minimum requirements according to good practice in digital preservation.

Relevant Features for Preserving CSPs

Descriptive Metadata

A common shortcoming of data stored in repositories a lack of granular metadata (Boulal et al. 2010, Place et al. 2008). The "main object", for instance a text file, is usually well described, whereas supplements are hardly ever equipped with descriptive metadata. Since a CSP can include materials created by different authors stored in different repositories, the lack of metadata can bring about a proverbial 'invisibility' of data. Descriptive Metadata which are exclusively associated with a full text have no use whatsoever for supplementary materials once they reside beyond the boundaries of the repository. As an example from the eco4r project, much supplementary material is stored in repositories without explicitly describing its content. Without basic metadata such as contentual, rights and authorship information the objects are not very likely to be reused in other environments. We recommend to include the following minimum descriptive metadata for any resource used in a CSP, even if some effort must be taken to meet these recommendations in existing repositories:

- classification
- creation date
- last modification date
- authorship
- rights information.

Persistent Identifier

From the perspective of Long Term Preservation, both the integrity of digital objects and the coherence between its constituents and their persistent identification are relevant (Doorenbosch & Sierman, 2010). A global persistent identification mechanism guarantees durable validity of the relationships between the constituents as well as independence from local changes (i.e. storage location or removal) within repositories. As stated in the OAIS model, the use of stable worldwide unique identifiers is most beneficial for Long Term Preservation. Persistent identi-

fiers allow for authentic referencing and assure a reliable assignment of metadata to the corresponding digital object. Common Internet addresses like an URL are not recommended because they usually change over time.
Moreover, the German DINI initiative[14] requires generating one persistent identifier for each object version whenever the content has changed (DINI, 2010). We recommend the use of standardized and well-established persistent identification schemes for both the aggregation (CSP) and every aggregated resource. A list of the most used persistent identification systems, specifications and standards can be found at (Neuroth et al, 2009).

Structural Information

The networked structures of CSPs can be described through semantic relationships. Place et al. (2008) classify different kinds of relationships which are able to express containments, sequences, lineages, versioning, manifestations and bibliographic citations. Many repository systems do not explicitly designate the relationships between objects (i.e. OPUS). This can lead to a major drawback. Without the ability to set relationships between objects, a CSP is not much more than loose parts of data. A piece of information such as

- "result of experiment X *is_derived_from* results of experiments Y and Z" or
- "document A *is_Annotation_of* doctoral thesis B"

is neither recorded nor processed. This way, crucial information for present and future semantic software services remains unused.

Semantic Information

In general, the semantic of a Compound Scholarly Publication is provided by a set of properties (metadata) and relationships that describe its internal and interlinking structure. The CSP data model is a suitable concept from which this information can be instantiated. A data model defines classes, properties and relationships, which are applied to objects from a specific domain. Examples from the domain of scientific publishing include

- classes, i.e. "Journal_Article", Proceeding", "Thesis", "Annotation" etc.
- relationships like "has_annotation", "is_manifestation_of" etc.
- properties such as "title", "subject", "creator" etc.

A data model can be expressed in different ways, for example as an Entity Relationship Model or as an Ontology. For our purpose, we use the ontological approach since ontologies can be expressed through formal languages (e.g. OWL

[14] http://www.dini.de/english

and RDFS) and serialized in a machine-readable form. Another useful characteristic of ontologies is the ability to deduce additional (implicit) knowledge automatically. This mechanism is known as *inference* (Allemang & Hendler, 2008).
For example, if we state that a "Thesis" is a "Publication" and that a "Master Thesis" is a "Thesis" we can deduce that a "Master Thesis" is also a "Publication" (**Fig.** 2). For Long Term Preservation, both the explicit and implicit information in a data model can be of crucial importance. Thus, we recommend translating the underlying data representation inside the repository into ontology serializations and to archive them along with the data described.

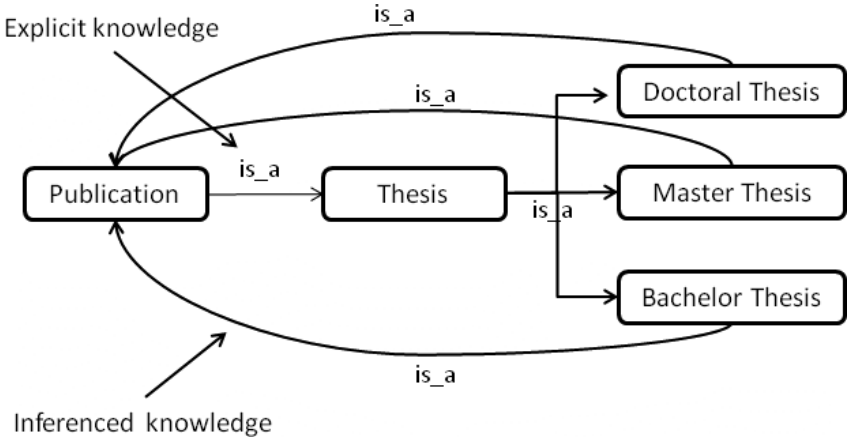

Fig. 2. Example for inference

Restructure

From the scientist's perspective, reusing all kinds of information provided by a CSP is in his main interest. The ability to find, access *and* understand *all* the content of a CSP gives him the means to use the new publication form most effectively. To enable this practical reuse, we will need to find a serialization form that encapsulates key properties of CSPs to capture technical, descriptive and structural information.

Both standards discussed above are frameworks that provide tools for the construction of standardized representations of Complex Objects. However, further adaptations and refinements are needed to fairly describe objects from a specific domain.

In this section we will discuss exemplary how to use OAI-ORE in order to extract an interoperable and machine-readable representation of CSPs stored in a Fe-

dora[15] repository. Then we will show how to include bibliographic information in an abstract OAI-ORE representation (Resource-Map).

Fedora to OAI-ORE Resource-Map

Fedora is a repository system that is well adapted for the deposit and management of Complex Objects. The underlying data model is flexible enough to depict relationships between digital objects and between different parts of an individual object as well.

Based on the Fedora data model[16], a digital object is represented as a container with different content objects known as "data streams"[17]. In addition to some reserved data streams (e.g. for storing relationships, Dublin Core[18] metadata and versioning information), they may be used to store any kind of digital material (PDFs, videos, software, visualizations, etc.). Object-to-Object relationships or relationships between content items are stored in special data streams called RELS-EXT and RELS-INT. The relational information is persisted in an RDF store as well.

Therefore, a mapping between Fedora and OAI-ORE data models is straightforward as shown in **Fig.** 3.

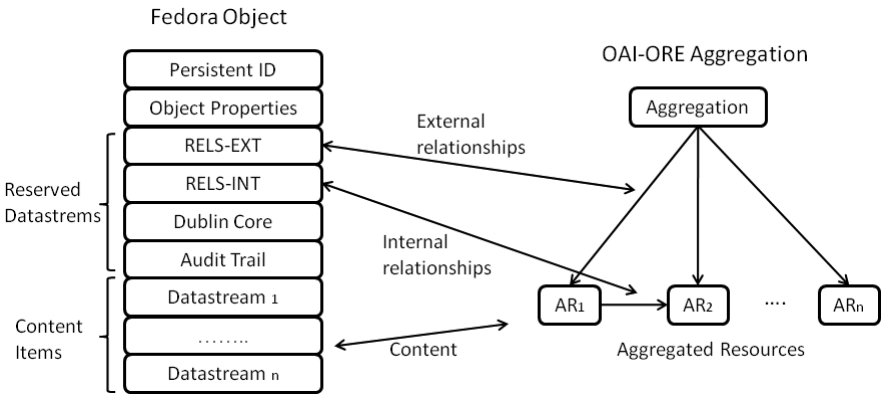

Fig. 3. Mapping between Fedora object and OAI-ORE Aggregation

[15] http://fedora-commons.org/

[16] http://fedora-commons.org/documentation/3.0b1/userdocs/digitalobjects/objectModel.html

[17]http://fedora-commons.org/documentation/3.0b1/userdocs/digitalobjects/objectModel.html#data

[18] http://dublincore.org/

Including Bibliographic Metadata in an OAI-ORE Resource-Map

As mentioned earlier, OAI-ORE defines a data model to describe aggregations of web resources. It defines an entity called Resource-Map that holds all information about an aggregation. The data model is highly generic but also expandable to enable an accurate description of objects of specific domains. OAI-ORE recommends the use of a variety of additional common vocabularies as Dublin Core or FOAF[19].

However, to create OAI-ORE representations of Compound Scholarly Publications we will need to integrate bibliographic information in the OAI-ORE representations (Resource-Map). This can be accomplished by incorporating one of the bibliographic vocabularies as The Bibliographic Ontology[20] or the FRBR-aligned Bibliographic Ontology[21] (FaBiO). The result will be an enhanced data model that is strongly based on the OAI-ORE data model and that simultaneously includes bibliographic concepts and properties for describing CSPs. Furthermore, using OAI-ORE as a basis of a data model for CSPs involves a good deal of interoperability. Every service on the web that understands OAI-ORE is able to handle CSPs as well. For example, in order to get a visualization of a CSP we can use every Resource-Map visualization service on the web (e.g. surf-incontext[22] visualization service).

The same approach has been adopted in the eco4r[23] project, where the FaBiO ontology has been used to refine the OAI-ORE data model towards a resulting data model. **Fig.** 4 shows a simplified example of a journal article represented according to the eco4r data model. A more accurate example can be found in the project wiki[24].

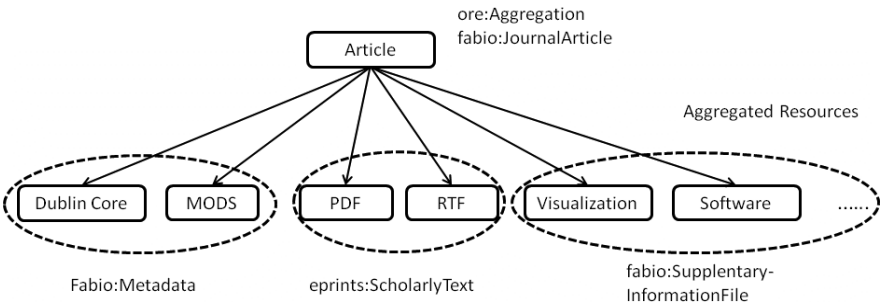

Fig. 4. A simplified representation of journal article according to the eco4r data model

[19] http://xmlns.com/foaf/0.1

[20] http://bibliontology.com

[21] http://speroni.web.cs.unibo.it/cgi-bin/lode/req.py?req=http:/purl.org/spar/fabio

[22] http://code.google.com/p/surf-incontext

[23] http://www.eco4r.de

[24] http://trac.eco4r.org/trac/eco4r/wiki/DataModel

Reuse

We define the reusability of CSPs as the ability to reuse entire publications or their component parts in other contexts. In prior sections we have seen that the process of reusing existing knowledge from a research process is crucial in order to generate new insights. Furthermore, capturing this knowledge in a well-structured form suitable for Long Term Preservation enables for durable access and reusability. We have also defined some criteria that enable CSPs to be reused by Long Term Preservation-technologies.
(Achtung, hier fehlen die Einrückungen wie im restlichen Text. Formatproblem?)
CSP that are represented by OAI-ORE Resource-Maps can easily be integrated in the Linked Open Data (LOD) infrastructure since they are RDF-based representations. LOD services such as lobid.org[25] (Linking Open Bibliographic Data) can reuse information about a CSP while data providers can in turn take advantage of this service to enrich their CSPs. For example, authorship information of a CSP can be enriched by connecting the corresponding Resource-Map to one ore more authority services as VIAF[26] (Virtual International Authority File). Subject information like a DDC-classification that is described as a simple character (e.g. ddc:600) in a CSP can be enriched by connecting the Resource-Map to a LOD service, see fig. 6.

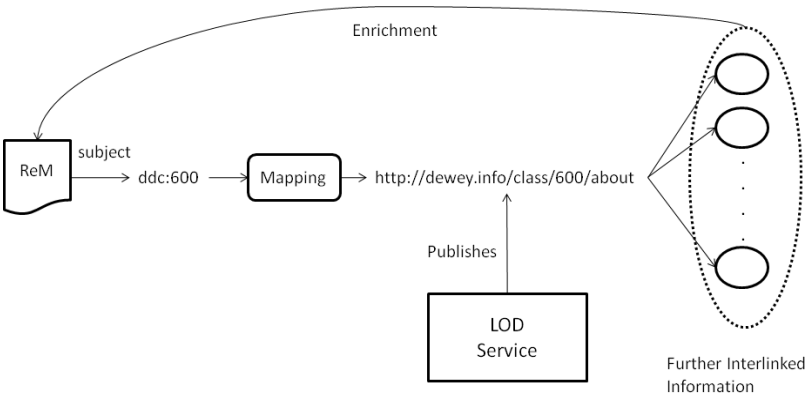

Fig. 5. Metadata enhancement with Linked Open Data Services

Although OAI-ORE requires common vocabularies such as Dublin Core together with domain specific information, it does provide a data model that may be used as a generic interoperability layer by different repositories. For example, in the

[25] http://lobid.org/de
[26] http://viaf.org

eco4r project, we show that Resource-Maps can be collected dynamically from different repositories in order to build new aggregations according to the DDC[27]-classification. By using a common data model (OAI-ORE), the integration and processing of CSP representations by a presentation layer[28] are both straightforward.

Case Study

With respect to its practical approach, the project's first effort was to analyse real life publication material stored in two different repository systems. Both Fedora and OPUS are hosted at the projects partners, hbz NRW[29] and Bielefeld University Library[30], respectively. We took the results from this evaluation to define practical requirements for an implemented prototype system that exposes and reuses real life CSPs. The system follows the principle of an "Overlay Journal". Technically, it will demonstrate the ability to gather information resources (including CSPs) from existing repository systems as well as the dynamic aggregation and creation of new CSPs on top of the gathered resources. In consequence, the Overlay Journal gives access to aggregated and newly arranged pieces of information for human researchers as well as automated systems. If metadata and semantic information are provided sufficiently, the information segments remain useful regardless to their actual composition.

In more detail, the prototype system collects OAI-ORE Resource Maps representing CSPs stored in the source repositories. After storing the collected representations in a RDF store, the Overlay Journal performs a sequence of processing steps. They include metadata enhancement through Linked Open Data services and the creation and exposure of new thematic aggregations. **Fig.** gives an overview about the architecture of the Overlay Journal.

[27] http://www.oclc.org/dewey

[28] In the eco4r project the presentation layer is an Overlay Journal

[29] http://www.hbz-nrw.de

[30] http://www.ub.uni-bielefeld.de

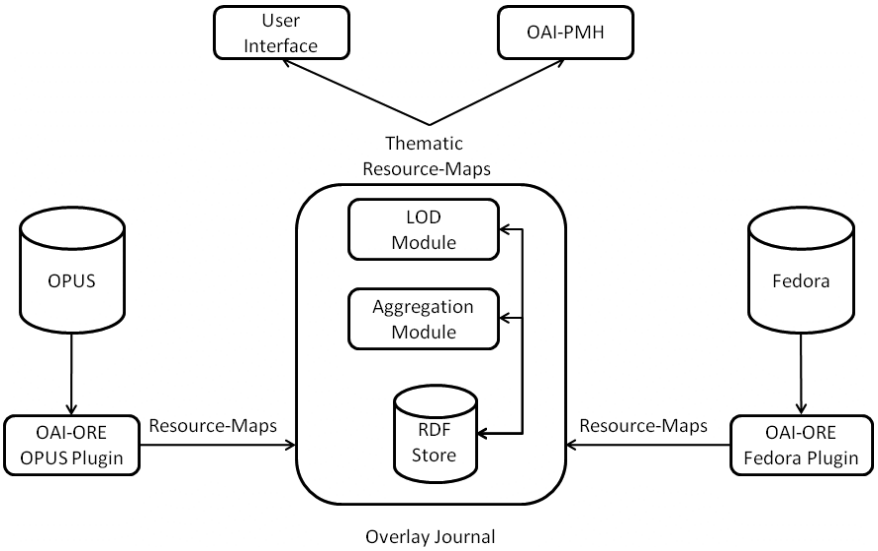

Fig. 6. Overlay Journals architecture and repository interaction

The repositories

The repositories used in eco4r are quite different in terms of the software systems used and the materials stored within them. The Bielefeld University Library uses the OPUS[31] repository software. Doctoral theses are the predominant content followed by some post-prints of published articles. The Library Service Centre in Cologne (hbz) operates and manages a publishing infrastructure for Open Access journals[32], wherein a Fedora[33] repository system stores the digital objects, representing journal articles and various supplementary materials (see Boulal et al. 2010).

The OAI-ORE Repository Plugins

The main practical outcomes from the project are OAI-ORE software plug-ins for each of the repository systems. Their function is to generate OAI-ORE Resource-Maps from the repositories' content according to the recommendations given be-

[31] http://www.opus-repository.org/index.html

[32] http://www.dipp.nrw.de

[33] http://fedora-commons.org

fore. Nevertheless, both plug-ins are designed to be generic and independent from the specific applications and data models used on top of Fedora and OPUS. By using a configuration mechanism, adjustments can be made to determine exactly how the generated Resource-Maps should look like. For example, the Fedora plug-in can be configured to select metadata, relationships and resources that should appear in the Resource-Map. Furthermore, a mapping mechanism is implemented to substitute repository-specific data structures with globally standardized relationships such as DC terms relationships, the FRBR-based FABIO ontology or the FOAF namespace. This way, we have harmonized the need for a deployment of standardized metadata with the practical requirements of proprietary repository systems.

The OAI-PMH Harvester

In practice, the OAI-ORE Resource-Maps are exposed through an OAI-PMH Interface. This way, an OAI Harvester can easily communicate with the repository's OAI-ORE plug-in. In our case, the harvester exposes the generated Resource-Maps as RDF/XML serialized items to the business logic of the Overlay Journal. By using the OAI-Protocol we introduced a rather simple but very robust technique to manage - i.e. find, harvest and update - CSPs in repository systems.

The RDF Store

Any Resource-Map provided by the Overlay Journal is stored in a RDF Store. In eco4r we use for the software Sesame 2.4.

The Linked Open Data (LOD) Module

With the LOD module, we aim to show that the information stored in CSP can be enhanced when connecting the associated Resource-Map with Linked Open Data services. In eco4r we are connecting to the lobid.org[34] service. The LOD module fetches the RDF store periodically and attempts to find intersections with information published in lobid.org. For example, in a Resource-Map stored in the RDF store we can find subject information according to the DDC[35] classification scheme, such as ddc:600 which represents the category "Technology". The LOD

[34] http://lobid.org/de

[35] http://www.oclc.org/dewey/about/default.htm

module fetches lobid.org and finds a resource for this category (e.g http://dewey.info/class/600), which is in turn connected to other information resources. The module will update the information stored in the RDF store by replacing the character "ddc:600" with the URI "http://dewey.info/class/600/". Thus, a metadata enhancement has been performed (**Fig.** 5).

Aggregation Module

The Aggregation module is the second use case to demonstrate the reusability of Resource-Maps. Here, new Aggregations are generated according to thematic properties such as the DDC classification. These new Aggregations will then be visualized in a user interface and, in turn, exposed via OAI-PMH.

Conclusion and Outlook

Scholarly communication is undergoing a major change. The emergence of modern technologies is opening new possibilities to the way scientists perform and disseminate their research. However, current publication infrastructures still focus on processing single monolithic resources within isolated data silos. They are not able to satisfy the growing demand for interlinked information resources that connect existing research data infrastructures using Semantic Web technologies.

For example, the High Level Expert Group on Scientific Data[36], which was assigned to prepare a "Vision 2030" [37] for the evolution of e-infrastructure for scientific data, recommends a larger degree of integration within scientific infrastructures. It states that the emerging Linked Data technologies have the potential to fulfil this requirement (Bizer, 2011).

Compound Scholarly Publications are a good example to illustrate a practical use case for linking different data resources using Linked Data and Semantic Web technologies. In the eco4r project as well as in associated projects, many theoretical and practical results have been achieved. However, we have discovered a lack of practical and real-life applications that allow for the processing of Complex Objects in scientific environments.

Having presented an executable technical approach for restructuring and reusing these valuable resources, we encourage both repository managers and data curators to affiliate their data pools as much as possible. Using tools and frameworks

[36] http://cordis.europa.eu/fp7/ict/e-infrastructure/high-level-group_en.html

[37] Charged by the European Commission's Directorate-General for Information Society and Media.

for data integration and Open Access publishing will not only revalue today's data but also ensure its continued existence.

Acknowledgments The eco4r project is a German Research Foundation (DFG) funded collaboration between Bielefeld University Library and North Rhine-Westphalian Library Service Centre (hbz). We would like to thank our Bielefeld colleagues Wolfram Horstmann, Jochen Schirrwagen and Friedrich Summann for their specialist support before and during the project. We also owe our thanks to the international participants of the 2010 eco4r workshop, which brought about very valuable insights into the way complex digital objects are - and should be - handled in future technical environments.

References

Allemang, D., Hendler, J. (2010). Semantic Web for the Working Ontologist - Modeling in RDF, RDFS and OWL

Bizer, C. (2011). *Expert Report on Linking Data & Publications*. Freie Universität Berlin.

Boulal, A., Quast, A., Iordanidis, M., & Schirrwagen, J. (2010). *Report on Enhancing Interoperability between existing Open Access Publication Infrastructures*. Library Service Center NRW, Cologne & Bielefeld University Library, Bielefeld.

Cheung, K., Hunter, J., Lashtabeg, A., & Drennan, J. (2008). *SCOPE - A Scientific Compound Object Publishing and Editing System*. The University of Queensland, St Lucia, Queensland, Austria.

DINI, Deutsche Initiative für Netzwerkinformation e.V., Arbeitsgruppe Elektronisches Publizieren (2010): DINI Certificate – Document and Publication Services 2010. http://nbn-resolving.de/urn:nbn:de:kobv:11-100182800

Doorenbosch, P., & Sierman, B. (2010). *Institutional Repositories, Long Term Preservation and* the changing nature of Scholarly Publications. Koninklijke Bibliotheek, the Netherlands.

Maslov, A., Creel,J. Mikeal,A. Phillips, S., Leggett, J. (2009).Adding OAIORE Support to Repository Platforms URI http://hdl.handle.net/1853/28448

Neuroth, H; Oßwald, A.; Scheffel, R.; Strathmann, M.; Huth, K. (2009). nestor Handbuch - Eine Enzyklopädie der Digitalen Langzeitarchivierung

Place, T., Van der Feesten, M., Hoogerwerf, M., Bijsterbosch, M., Slabbertje, M., Hogenaar, A., et al. (2008). *Report on Object Models and Functionalities - DRIVER II*. Leiden University.

Ruijgrok, P., Slabbertje, M., & Van Luijt, M. (2009). *A candidate semantic representation for enhanced e-theses: Guidelines for modeing enhanced e-theses*. Knowledge Exchange - University of Utrecht.

Van de Sompel, H. (2009). *The OAI-ORE Interoperability Framework in the Current Scholarly Communication Context*. Los Alamos National Laboratory, USA.