The International Journal of Digital Curation
Issue 2, Volume 3 | 2008

SCOPE: A Scientific Compound Object Publishing and Editing System

Kwok Cheung, Anna Lashtabeg, John Drennan

AIBN, The University of Queensland,

St Lucia, Queensland, Australia

Jane Hunter,

ITEE, The University of Queensland,

St Lucia, Queensland, Australia

November 2008

Abstract

This paper presents the SCOPE (Scientific Compound Object Publishing and Editing) system which is designed to enable scientists to easily author, publish and edit scientific compound objects. Scientific compound objects encapsulate the various datasets and resources generated or utilized during a scientific experiment or discovery process, within a single compound object, for publishing and exchange. The adoption of "named graphs" to represent these compound objects enables provenance information to be captured via the typed relationships between the components. This approach is also endorsed by the OAI-ORE initiative and hence ensures that we generate OAI-ORE-compliant Scientific Compound Objects. The SCOPE system is an extension of the Provenance Explorer tool – which supports access-controlled viewing of scientific provenance trails. Provenance Explorer provided dynamic rendering of RDF graphs of scientific discovery processes, showing the lineage from raw data to publication. Views of different granularity can be inferred automatically using SWRL (Semantic Web Rules Language) rules and an inferencing engine. SCOPE extends the Provenance Explorer tool and GUI by: 1) Adding an embedded web browser that can be used for incorporating objects discoverable via the Web; 2) Representing compound objects as Named Graphs, that can be saved in RDF, TriX, TriG or as an Atom syndication feed; 3) Enabling scientists to attach Creative Commons Licenses to the compound objects to specify how they may be re-used; 4) Enabling compound objects to be published as Fedora Object XML (FOXML) files within a Fedora digital library.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. ISSN: 1746-8256 The IJDC is published by UKOLN at the University of Bath and is a publication of the Digital Curation Centre.

Introduction and Objectives

In recent years it has been increasingly acknowledged that traditional scientific publications only represent the outcome of the final phase of the scientific discovery process – on the whole, they inadequately represent the earlier phases that involve the capture, analysis, modeling and interpretation of primary scientific data. A record of the complete scientific discovery process (captured through detailed provenance information) is valuable because it enables verification and repeatability of results and peer review of the methodology. Lack of access to high-quality scientific data with the associated provenance information is an obstacle to interdisciplinary and inter-organizational research. As a result, many organizations and funding bodies are actively encouraging or even mandating the publication of scientific data together with traditional publications across many domains (Arzberger et al, 2004; Berlin Declaration, 2003; Klump et al, 2006; National Institutes of Health [NIH], 2003). But there are a number of barriers that need to be overcome, before many scientists willingly publish their raw and derived datasets. These include: a lack of simple tools for publishing data with provenance information; lack of motivation for scientists to spend time and effort preparing their data for publication; concern with intellectual property rights; a lack of standards for publishing datasets and provenance; and discipline-specific tools that prohibit cross-disciplinary sharing and exchange.

To date, a number of different approaches have been implemented that link raw scientific data to scientific publications. Some online publishers, including Acta Crystallographica Section E - Structure Reports Online (Helliwell, Strickland & McMahon, 2006), Nature (Nature, n.d.), and the ePIC Earth System Science Data and Methods (2007) journal support the association of supplementary datasets with scholarly papers. Murray-Rust and Rzepa proposed the concept of datuments – XML documents, that combine the data and the document using formal markup to allow processing and rendering in different ways via XSLT (Murray-Rust & Rzepa, 2004). A number of initiatives and projects (e.g., Protein Data Bank (PDB) (Berman, Bourne & Westbrook, 2004), CombeChem (Coles et al, 2005), eBank (Lyon, 2003), NCBI[1], Virtual Observatory[2] and GBIF[3]) have spearheaded the development of infrastructures that facilitate the online publication of electronic scientific datasets.

Although these existing approaches have advanced the publication of scientific data, they also have a number of limitations including:

- The relationships between the datasets and publications are one-to-one, relatively fixed and involve hyperlinks with little or no support for semantics or provenance information.
- A lack of flexibility or extensibility – scientists require simple, interactive GUIs that enable them to interactively select a set of resources generated from an experiment or investigation, relate them to each other and publish them as a package.
- Lack of support for multi-level access to data or information. Existing systems seem to support open access only.
- Lack of rule-based or template-based systems for rendering different

[1] National Center for Biotechnology Information http://www.ncbi.nlm.nih.gov/
[2] International Virtual Observatory Alliance: Documents and Standards http://www.ivoa.net/Documents/
[3] Global Biodiversity Information Facility (GBIF) Data Use Agreement http://data.gbif.org/datasets/

presentations dynamically, based on the context, user's needs or access rights.

- Difficulty discovering and retrieving components that are deeply embedded or hidden within HTML pages or the deep web (Bergman, 2001).

The full potential of compound objects cannot be realized unless the structural and component information is both human-understandable and machine-interpretable. The Open Archives Initiative (OAI) (Lagoze & Van de Sompel, 2001) is proposing a new standard named Object Reuse and Exchange (OAI-ORE)[4] that aims to make the information within compound/complex digital objects (Kahn & Wilensky, 2006) discoverable, machine-readable, interoperable and reusable. The objective of the OAI-ORE Initiative is to develop an interoperability layer across cooperating digital repositories, registries and services for the reuse and exchange of compound digital objects, based on the Web architecture (W3C Technical Architecture Group, 2004). The OAI-ORE white paper (Lagoze & Van de Sompel, 2007) recommends Named Graphs (Jeremy et al, 2005) as a means to publish compound digital objects in order to clearly state their logical boundaries and typed relationships between their components. Named Graphs consist of nodes and arcs. When applied to compound objects, the nodes correspond to component resources; and the arcs correspond to typed relationships. Named Graphs, their nodes and arcs are all web resources – so they can be identified and referenced unambiguously by HTTP URIs. This provides the basis for the reuse and exchange of the OAI-ORE compound objects and their components because URIs provide the handles.

Our hypothesis is that OAI-ORE Named Graphs provide the ideal mechanism for representing scientific compound objects that encapsulate the raw data, their derivative products and outputs/publications as well as the intervening processing steps. They do this in a way that is discipline-independent but provides hooks to include rich semantics, metadata and discipline-specific vocabularies, ontologies and rules.

Hence the primary objective of this work is to develop an intuitive, simple, easy-to-use system that enables scientists to quickly author scientific compound objects with built-in provenance and to publish them to a repository with associated metadata and a Creative Commons license – we call this the SCOPE (Scientific Compound Object Publishing and Editing) system. If SCOPE can deliver on these objectives, then the system overcomes some of the current barriers to scientific data publication including: a lack of incentive; lack of tools; difficulty preparing data for publication; difficulty providing an appropriate level of provenance data; concern with intellectual property rights.

The Provenance Explorer tool (Hunter & Cheung, 2007) provides dynamic rendering of provenance trails (captured from workflow engines) via the visualization of RDF graphs. In addition, users can expand or collapse arcs to generate fine or coarse-grained views. Because Provenance Explorer provides a subset of the functionality required for SCOPE, we decided to exploit this existing work and extend it through the addition of new functionality. More specifically, SCOPE extends Provenance Explorer through the addition of the following functionality:

[4] Open Archives Initiative Object Reuse and Exchange http://www.openarchives.org/ore/

- A Web browser which supports the discovery and importation of digital objects over the Internet – as new components of compound objects;
- Enabling the compound objects (named graphs) to be saved in a variety of serializations (RDF/XML, TriX (Carroll & Stickler, 2004) and TriG (Bizer, 2007), as well as the Atom syndication feed (Sayre, 2005)) in order to strengthen the dissemination of the compound objects over the Web;
- Converting the compound objects with their components into Fedora (Lagoze et al, 2006) Object XML (FOXML) files, and ingesting them to Fedora digital libraries.
- Enabling scientists to attach Creative Commons Licenses to the compound objects to specify how the compound object may be reused;
- Enabling existing compound objects to be reloaded and edited by authenticated users with appropriate access permissions via the SCOPE GUI.

The remainder of this paper is structured as follows: a description of related work; the SCOPE system's architecture and components; a description of the case study we used for evaluation and testing; implementation and user interface; a concluding section that comprises an evaluation, discussion and future work plans.

Related Work

There are two primary methods by which research data are linked to scholarly publications:

1. The first method involves including a reference from the paper to a unique accession number in a database or embedding a hyperlink in the paper that links to data held within a database or a dataset via a unique identifier (e.g., many publishers use Digital Object Identifiers (DOIs)).
2. The second approach involves embedding the data within the scholarly publication via a formal markup language.

Examples of publishers who support the first approach include: Nature and American Chemical Society – who require that papers about proteins, DNA sequences or molecular structures must uniquely identify the subject via an accession number assigned by designated publicly accessible databases such as Genbank (Benson et al, 2007), the Protein Data Bank (PDB) (Berman, Bourne & Westbrook, 2004), and SWISS-PROT (Boeckmann et al, 2003). This approach depends on the long-term availability and accessibility of large-scale online databases of scientific data. The Protein Data Bank (PDB) (Berman, Bourne & Westbrook, 2004) is just one example of a public database that is built from user submissions. Similar large-scale online databases have been established and maintained by organizations such as, NASA, NIST, NCBI, STD-DOI, GBIF and NOAA – for research domains including global atmospheric and climatic research, computational chemistry, genomics, earth sciences and analytical physics. Typically these organizations collect data in their own database schema and if others want to upload their data, they must first convert it to the organizational database schema and specified formats. There are a number of problems associated with this approach, including:

- The link from the paper to the data is usually uni-directional and does not include any semantics or provenance information. Discovering the data via web crawlers is not possible because it is part of the *deep web* (Bergman, 2001).

- The procedure of submitting papers and/or data to online publishers and publicly accessible databases are database-specific and rigid. Understanding those procedures can frustrate, or demotivate scientists from publishing their data.

The second approach to publishing raw data linked to publications, involves using some form of XML to mark up a scientific publication structurally and semantically – to distinguish between and interpret the publication text and different types of embedded or related data. Examples of this approach include Murray-Rust's datuments (Murray-Rust & Rzepa, 2004) – XML documents that are machine-readable and can be rendered in different ways using XSLT. The eCrystals Crystal Structure Report Archive (Coles et al, 2006), a sub-project under CombeChem and eBank, publishes first-hand but non-peer-reviewed crystallographic data online. All information about a single crystal is dynamically generated as a highly structured web page with detailed provenance information and links to related citations. Acta Crystallographica Section E - Structure Reports Online (Helliwell, Strickland & McMahon, 2006) also binds hyperlinks to the paper and supplementary material under the one title. The German Scientific Drilling Database (SDDB) project (Klump & Conze, 2007) is also investigating the use of XHTML to integrate geological sample information with derived data and published studies in which the data are interpreted.

The major drawback associated with the second approach is that many web spiders/crawlers cannot determine the semantic relationships between raw data and HTML text bound within a single web page. Explicitly typed relationships, as defined within Named Graphs, are required to raise the relationships between components to first class objects which have their own provenance information.

However, some of the limitations of the XHTML approach to scientific publishing may be overcome through the adoption of emerging technologies such as Microformats (Khare & Celik, 2006), RDFa (World Wide Web Consortium [W3C], 2007) and GRDDL (Gleaning Resource Descriptions from Dialects of Languages (W3C, 2004)). Microformats and RDFa enable semantic tags to be embedded within XHTML documents to tag the content or link to related documents or data (via the *rel* tag) – without affecting the display of the HTML text. The inclusion of these light-weight semantic tags supports machine understanding, interpretation and processing of the publication. GRDDL is a mechanism for deriving formal metadata in RDF by using XSLT to process XHTML documents, to extract the embedded semantics (e.g., RDFa). The future may well see scientific publication authoring systems that use RDFa to embed tags or annotations in (X)HTML files and use RDFa-aware browsers or GRDDL to extract this, convert it to RDF, store it in an RDF triple store and search it using SPARQL.

System Architecture and Overview

Figure 1 below illustrates the overall system architecture. It comprises:

- The SCOPE Java Application and GUI – which has four components:
  - The provenance viewer;
  - The web browser;
  - The publishing interface;
  - The metadata input and editing interface.

- The knowledge base which consists of SWRL.OWL files that contain the provenance instance data, the metadata and the inference rules.
- The provenance visualization tool – JGraph and Jena are used to convert an RDF graph into an image consisting of nodes (objects/classes) and arcs (relationships/properties).
- The Algernon rule-inference engine – for inferring new indirect relationships based on SWRL rules.

The Provenance Explorer paper describes in detail the interactions between the knowledge base and the provenance view (via JGraph and Jena) and between the knowledge base and Algernon (via Protégé-OWL). The development of SCOPE has required the development of a number of additional functionalities: importing digital objects via the web browser; the transformation of compound objects into various formats; and the ingestion to a Fedora digital library.

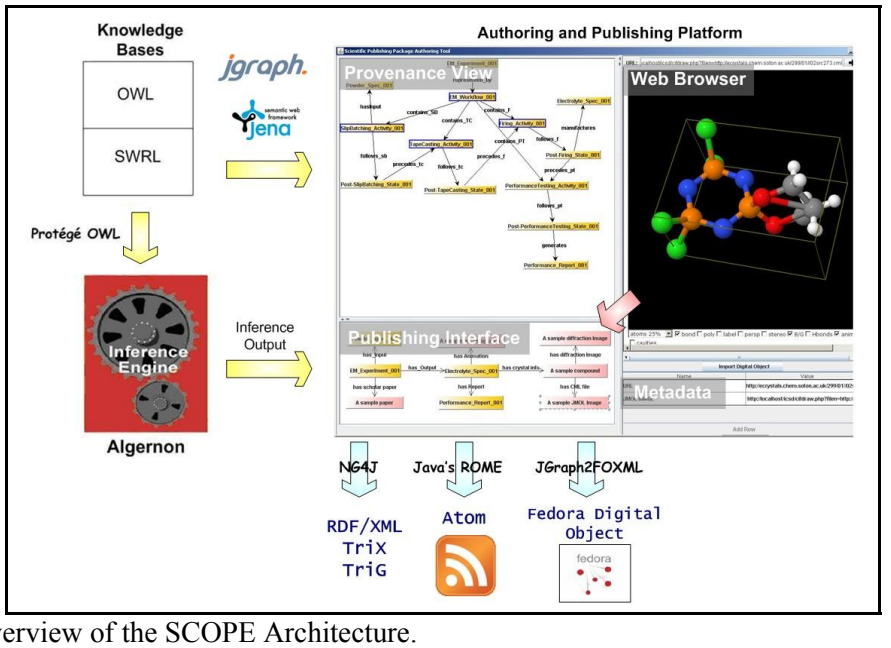

Figure 1. Overview of the SCOPE Architecture.

In the top right-hand side of the Authoring and Publishing Platform on Figure 1, there is a web browser implementing Java's JDesktop Integration Component (JDIC)[5]. In many situations, scientists want to incorporate links to external web resources from within their compound objects. An "IMPORT" button has been provided at the bottom of the web browser window. This adds a new node, representing the object, to the Publishing interface. The red arrow between the Web Browser and Publishing Interface indicates the import route and the red nodes on the interface are the representations of imported web information.

The nodes and arcs displayed in the Publishing Interface represent the current components of the Named Graph/compound object that is being constructed. Users can save this object in a range of different formats including: RDF/XML, TriX, TriG (using the Named Graphs API for Jena (NG4J) (Bizer, Cyganiak & Watkins, 2005)); and Atom syndication feed (using Java's ROME, the RSS/Atom syndication and publishing tools[6]).

[5] JDesktop Integration Component https://jdic.dev.java.net/
[6] ROME, the RSS/Atom syndication and publishing tools https://rome.dev.java.net/

Finally, the compound object can also be converted into a FOXML file, and ingested into a Fedora repository using JGraph2FOXML, a Java API developed by University of Queensland's eResearch Group, using Fedora Access and Management Web Services

At the time of publishing to Fedora, users are also required to enter metadata (creator, data, title, description) associated with the compound object. This is entered via the metadata input interface (bottom right-hand side). At this point users may also choose a Creative Commons license and attach it to the FOXML file.

Case Study

Although this system has been designed to be used within any scientific discipline, we have been evaluating it within the materials science domain at the University of Queensland. In particular, we have been collaborating with a group of fuel cell scientists within the Australian Institute for Bioengineering and Nanotechnology (AIBN). They are investigating novel metal oxides for Solid Oxide Fuel Cell components which exhibit high conductivity within relatively low temperature ranges.

The composite powder that is produced from a complex processing and sintering procedure is characterized using X-ray diffraction techniques. In addition thermal expansion coefficients and electronic conductivity are measured over a range of operating temperatures. During the characterization process, significant amounts of data are generated in a range of formats including images, numerical data and graphs. Figure 2 below provides a simplified view of the powder manufacture and characterization process. It also shows a subset of the datasets generated from X-Ray diffraction and characterization – which are processed to generate graphs which are included in the final publication. For illustrative purposes, we also assume that the publication contradicts earlier results published in another previous publication.

The challenge is to provide a system that enables the fuel cell scientist to quickly and easily package up the relevant datasets, images, graphs and papers into a publishable compound object that also contains an explanation of the relationships between the components, the method of derivation, and allows easy fine-grained discovery of components.

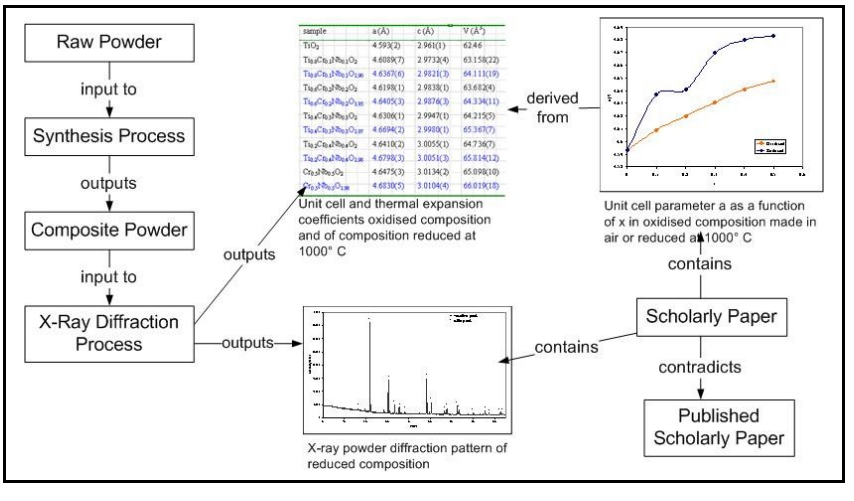

Figure 2. Simplification of the Scientific Discovery Process for Novel Oxide Conductors.

Implementation and User Interface

In this section, we discuss the implementation of the SCOPE system, in the context of the above case study. We assume that the experimental steps and digital objects generated during the Synthesis Process and X-Ray Diffraction Process are captured and stored as RDF using one of the existing scientific Workflow Systems that generate RDF – such as Kepler (Altintas, Barney & Jaeger-Frank, 2006), Taverna (Oinn et al, 2004), Triana (Majithia et al, 2004), or one of the e-Lab notebook systems such as the Collaborative Electronic Research Framework (CERF)[7], SmartTea (schraefel et al, 2004) or MyTea (Gibson et al, 2006) systems. The RDF graph corresponding to the scientific experimental workflow provides the starting point for the scientific compound object that is to be published. This ensures that any available provenance information will be leveraged in the published object.

Authoring

After users log on to the system and are authenticated, they are presented with a simple SPARQL search interface which supports search and retrieval of existing RDF experiments. For example, a user can search for and retrieve a particular experiment via a unique ID e.g. EXP280818. Initially users are presented with the basic default view of the experiment provenance in the top left-hand side of the application window. The blue nodes indicate the characterization processes that can be expanded to reveal further fine-grained information (represented by light gray nodes shown on the right-hand side of Figure 3 below).

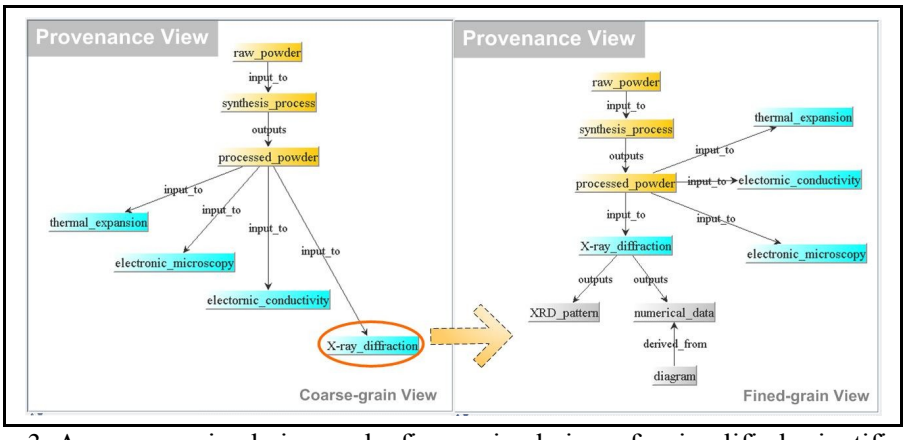

Figure 3. A coarse-grained view and a fine-grained view of a simplified scientific process.

One of the advanced functionalities within the Publishing Interface is automatic inferencing of direct relationships between indirectly related non-expandable nodes (light gray or yellow) via an inferencing engine. Users can drag and drop any two non-expandable nodes from the Provenance View panel down to the bottom Publishing panel, draw a link between them, and infer their relationship. This supports the streamlined publishing of coarse-grained views of the scientific method. In Figure 4 the inferencing result circled in blue is shown along the linkage on the bottom pane while the inferred path is highlighted in blue on the top pane. The inferencing rule is as follows:

IF *(processed_powder input_to X-ray_diffraction)* ***AND***
(X-ray_diffraction outputs XRD_pattern)
THEN *(processed_powder characterized_by XRD_Pattern)*

[7] CERF - Collaborative Electronic Research Framework http://www.rescentris.com/

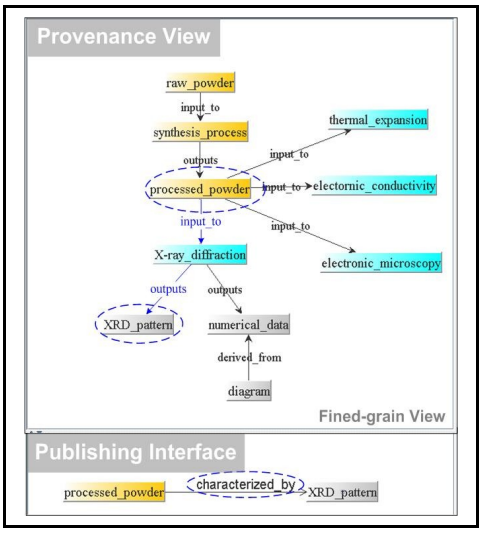

Figure 4. An inferencing result.

Users can also discover external objects on the web via the embedded browser, and import these external objects as components of publishable compound objects. Figure 5 below demonstrates the three-step process. The web browser (top right-hand side) displays the resource to be imported.

1. the *Import Digital Object* button is clicked to create a new red node in the Publishing Interface;
2. The user draws a line from the *Scholarly Paper* node to the imported node, and then labels the relationship e.g. *contradicts*; and
3. the Metadata view displays metadata extracted from the file's header (if available).

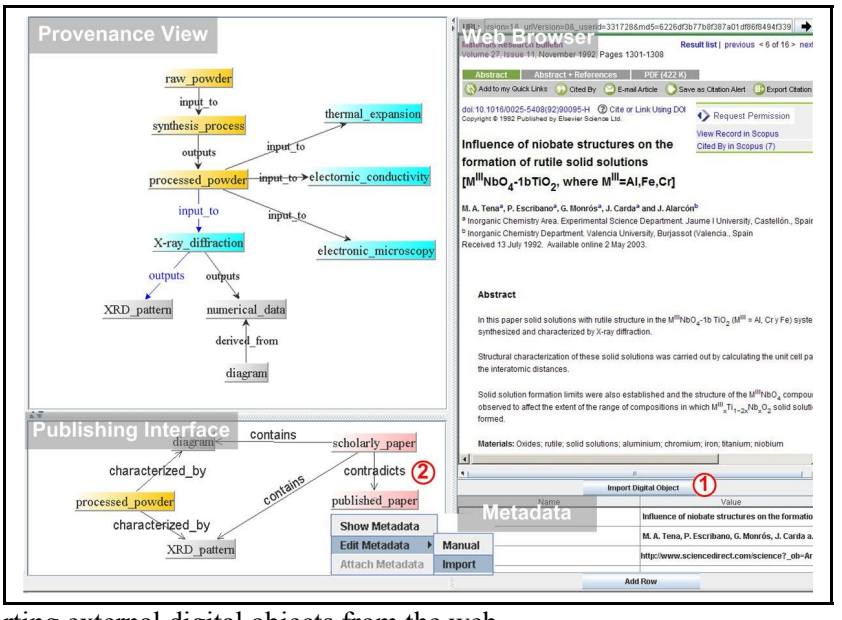

Figure 5. Importing external digital objects from the web.

The nodes and arcs displayed in the Publishing Interface represent a compound object. Figure 6 demonstrates the three-step process of creating/editing and attaching metadata to a compound object:

1. When the user clicks in the Metadata editing/input window, the background of the Publishing Interface turns gray;
2. The Metadata view displays fields for the entry of Dublin Core Metadata;
3. After the users have attached the new metadata, the background returns to the original color. Users may also edit the metadata associated with component objects if they are stored locally.

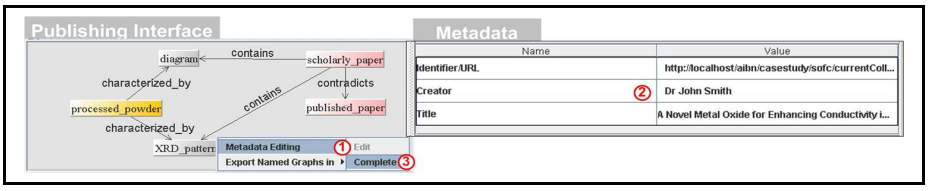

Figure 6. Creating/editing and attaching metadata to a compound object.

Publishing

The system enables a compound object to be saved to a variety of web formats including: RDF/XML, TriX and TriG, and the Atom syndication feed, respectively. Figures 7 and 8 below illustrate the conversion to TriX and the Atom syndication feed. The colored highlights on the right-hand side correspond to the coloured nodes within the RDF graph in the Publishing Interface on the LHS.

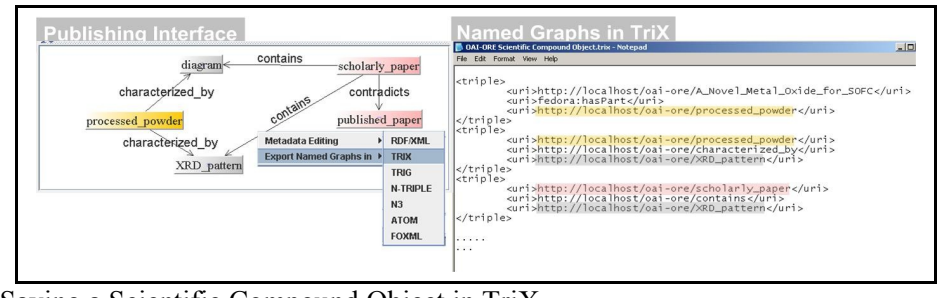

Figure 7. Saving a Scientific Compound Object in TriX.

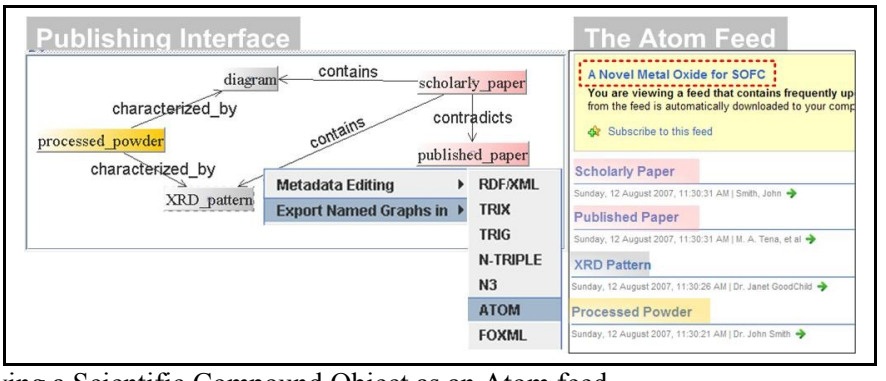

Figure 8. Saving a Scientific Compound Object as an Atom feed.

Finally, the system also enables users to publish a compound object by ingesting it into a Fedora digital library. Figure 9 demonstrates the ingestion process on the left-hand side and the resultant Fedora (compound) digital object on the right-hand side. The user chooses FOXML output from the Publishing Interface. The Fedora digital object (right-hand side) indicates that it is a compound object with both *Composite Relationship in RDF* and the *Component to Component Relationship in RDF*.

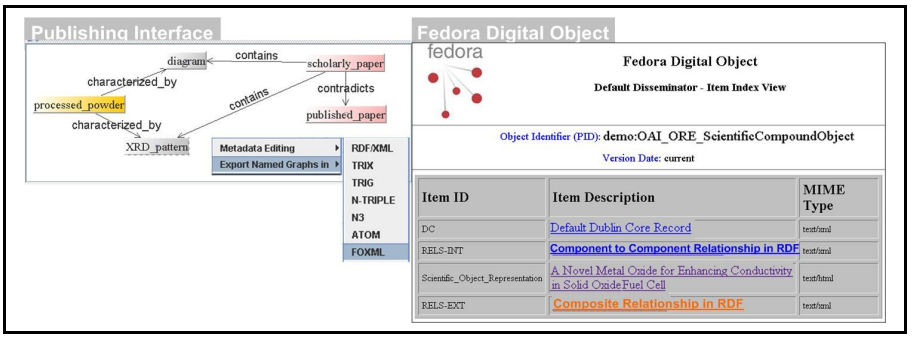

Figure 9. Generation of a Fedora Compound Digital Object.

Discussion

User Feedback

Feedback from the fuel cell scientists with whom we have been collaborating has been very positive. They particularly liked the ability to graphically link internally generated provenance trails to external resources, that can be discovered and retrieved via a web browser. This allows authors to include other relevant research outcomes to strengthen the claims of their findings, thereby making their publication package more comprehensive but still self-contained – facilitating the peer-review process. The ability to interactively generate coarse-grained views of scientific experimental processes, via automatic inferencing was also very popular – as was being able to attach attribution and Creative Commons license. Further collaboration with the fuel cell scientists is required to develop better, discipline-specific inferencing rules and a specific relationship ontology, grounded in solid scientific research and knowledge.

Limitations and Future Work

The system developed to date is a working prototype that demonstrates the benefits of named graphs and the OAI/ORE approach to scientific data publishing. However further effort is required to improve the system's usability and robustness and to overcome existing limitations that include:

- Currently, only Dublin Core metadata is supported for describing the FOXML files at publication time. Support for other metadata schemas should be an option;
- New typed relationships defined through the Publishing interface can currently only be labeled with free text labels. We are working on an ontology that defines a hierarchy of relationships between information objects within the scientific domain. We are also focusing more effort on the inferencing rules that apply to these relationships;
- The system currently only supports uni-directional relationships. We would like the ability to define bi-directional relationships – and symmetric, transitive and reflexive relationships within the relationship ontology.
- Some of the script behaviours on Web pages could prove frustrating to users. For example, clicking on a hyperlink within the Publishing page may trigger the launch of a new browser window outside the system.
- At this stage, the system does not support searching, reloading and editing of published OAI-ORE scientific compound objects. This capability is currently under development.

A number of weaknesses were also identified in the different serializations of the named graphs:

1. TriX and TriG are new and still in an early stage of their development and have not yet been widely adopted;
2. RDF/XML represents inferencing rules as triples, thereby confusing the rules with other triples;
3. The Atom syndication feed is less expressive and cannot indicate the relationships between entries/components;
4. The FOXML document relies on RDF/XML to represent the composite and component-to-components relationship, so inferencing rules cannot be represented within Fedora repositories effectively.

Future plans also include discussing possible deployment of SCOPE within the Materials Science domain, in conjunction with the NSDL Materials Digital Library – to support the publishing of compound scientific publications or e-learning resources associated with materials science research. We also plan to evaluate the system more thoroughly using case studies and user groups from other disciplines such as bioinformatics, earth sciences, crystallography and the humanities and social sciences.

Conclusions

In this paper, we have described the SCOPE tool for authoring and publishing *OAI-ORE*- compliant scientific compound objects, that we have developed by extending the functionality of our existing Provenance Explorer tool.

The SCOPE system provides solutions to some of the current barriers to scientific data publishing. It provides a simple tool by which scientists can author and publish scientific compound publications that encapsulate raw data, derived data, provenance and publications in a single package. Authors can also attach metadata to the individual components and the compound object and save the package in a variety of formats, in order to maximize the discovery, dissemination and reuse of the publication or its components. With the worldwide efforts for open access to publicly funded research[8], scientists are under increasing pressure from funding agencies to publish the experimental and evidential data with the related traditional scholarly publication(s). SCOPE can help facilitate this.

Although we have chosen to adopt the OAI-ORE recommendation of Named Graphs for representing these compound objects, the emergence of Microformats, RDFa and GRDDL are of significant interest and an area we would like to explore further. We believe that the future may well see a struggle for dominance between these two approaches to scientific data publishing:

1. the OAI-ORE Named Graph approach – which defines scientific publications as compound objects that are comprised of a set of component objects with typed relationships between them;
2. the XHTML with RDFa/Microformats approach – publications that are HTML documents with embedded data, images and links to external files – but that are interpretable due to the light-weight semantic markup.

[8] The Alliance for Taxpayer Access http://www.taxpayeraccess.org/international.html

References

Altintas, I., Barney, O., & Jaeger-Frank, E. (2006). Provenance Collection Support in the Kepler Scientific Workflow System. *International Provenance and Annotation Workship (IPAW'06)*. Chicago, Illinois, USA.

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., et al. (2004). Promoting Access to Public Research Data for Scientific, Economic, and Social Development. *Data Science Journal, 3*, 135-152.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Wheeler, D. L.. (2007). GenBank. *Nucleic Acids Research, 35*(suppl_1), D21-25.

Bergman, M. K. (2001). The Deep Web: Surfacing Hidden Value. *Journal of Electronic Publishing, 7*(1).

Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, (2003). *Conference on Open Access to Knowledge in the Sciences and Humanities*. Berlin, Germany.

Berman, H. M., Bourne, P. E., & Westbrook, J. (2004). The Protein Data Bank: A Case Study in Management of Community Data. *Current Proteomics,* 2004, *1,* 49-57.

Bizer, C. (2007). The TriG Syntax. Retrieved August 25, 2007, from http://sites.wiwiss.fu-berlin.de/suhl/bizer/TriG/

Bizer, C., Cyganiak, R., & Watkins, E. R. (2005). NG4J - Named Graphs API for Jena. In *The Second European Semantic Web Conference*. Heraklion, Greece.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, MC., Estreicher, A., Gasteiger, E., et al. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research, 31*(1), 365-370.

Carroll, J. J., Bizer, C., Hayes, P., & Stickler, P. (2005). Named graphs, provenance and trust. *Proceedings of the 14th international conference on World Wide Web*. Chiba, Japan: ACM Press.

Carroll, J. J., & Stickler, P. (2004). TriX: RDF Triples in XML. *Extreme Markup Languages 2004*. Montreal, Canada.

Coles, S., Frey, J., Hursthouse, M., Light, M., Carr, L., DeRoure, D., et al. (2005). The 'end to end' crystallographic experiment in an e-Science environment: From conception to publication. *UK e-Science All Hands Meeting 2005*. Nottingham, UK.

Coles, S. J., Frey, J. G., Hursthouse, M. B., Light, M. E., Milsted, A. J., Carr, L. A., et al. (2006). An e-Science environment for service crystallography - from submission to dissemination. *Journal of Chemical Information and Modeling, 46*(3), 1006-1016.

Gibson, A., Stevens, R., Cooke, R., Brostoff, S. and Schraefel, M. C. (2006). myTea: Connecting the Web to Digital Science on the Desktop. *World Wide Web Conference 2006*. Edinburgh, Scotland.

Helliwell, J. R., Strickland, P. R., & McMahon, B. (2006). The role of quality in providing seamless access to information and data in e-science; the experience gained in crystallography. *Information Services & Use, 26*(1), 45-55.

Hunter, J., & Cheung, K. (2007). Provenance Explorer-a graphical interface for constructing scientific publication packages from provenance trails. *International Journal on Digital Libraries*, 2007, *7*(1-2), 1432-5012.

Kahn, R., & Wilensky, R. (2006). A framework for distributed digital object services. *International Journal on Digital Libraries*, 2006, *6*(2), 115-123.

Khare, R., & Celik, T. (2006). Microformats: a pragmatic path to the semantic web. *Proceedings of the 15th international conference on World Wide Web*. Edinburgh, Scotland: ACM Press.

Klump, J., Bertelmann, R., Brase, J., Diepenbroek, M., Grobe, H., Höck, H., et al. (2006). Data publication in the open access initiative. *Data Science Journal, 5*, 79-83.

Klump, J., & Conze, R. (2007). The Scientific Drilling Database (SDDB) - Data from Deep Earth Monitoring and Sounding. *Scientific Drilling, 2007*(4), 30-31.

Lagoze, C., Payette, S., Shin, E., Wilper, C. (2006). Fedora: an architecture for complex objects and their relationships. *International Journal on Digital Libraries, 6*(2), 124-138.

Lagoze, C., & Van de Sompel, H. (2001). The open archives initiative: building a low-barrier interoperability framework. *Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries*. Roanoke, Virginia, USA: ACM Press

Lagoze, C., & Van de Sompel, H. (2007). Compound Information Objects: The OAI-ORE Perspective. Retrieved November 18, 2008, from http://www.openarchives.org/ore/documents/CompoundObjects-200705.html

Lyon, L. (2003, July). eBank UK: Building the links between research data, scholarly communication and learning. *Ariadne*, 2003, *36.* Retrieved November 18, 2008, from http://www.ariadne.ac.uk/issue36/lyon/

Majithia, S., Shields, M., Taylor, I., & Wang I. (2004). Triana: a graphical Web service composition and execution toolkit. *Proceedings of the IEEE International Conference on Web Services*. Washington, DC, USA: IEEE Computer Society.

Murray-Rust, P., & Rzepa, H. S. (2004). The Next Big Thing: From Hypermedia to Datuments. *Journal of Digital Information,* 2004, *5*(1).

National Institutes of Health. (2003). *Final NIH Statement on Sharing Research Data*. Retrieved August 25, 2007, from http://grants2.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html

Nature. (n.d.). Editorial policies about availability of data and materials. Retrieved August 28, 2007, from http://www.nature.com/authors/editorial_policies/availability.html

Oinn, T., Addis, M. J., Ferris, J., Marvin, D. J., Senger, M., Carver, T., et al. (2004). Taverna: a tool for the composition and enactment of bioinformatics workflows. *Bioinformatics, 20*(17), 3045-3054.

Pfeiffenberger, H., Richter, A. (2007, 18-20 April). An Earth System Science Data Publishing Journal, *5th Workshop on Innovations in Scholarly Communication (OAI5)*. CERN, Geneva, Switzerland. Retrieved August 31, 2007, from http://epic.awi.de/Publications/Pfe2007c.pdf

Sayre, R. (2005). Atom: the standard in syndication. *Internet Computing*, IEEE, *9*(4), 71-78.

schraefel, M. C., Hughes, G. V., Mills, H. R., Smith, G., Payne, T. R., & Frey, J. (2004). Breaking the book: translating the chemistry lab book into a pervasive computing lab environment. *Proceedings of the SIGCHI conference on Human factors in computing systems*. Vienna, Austria: ACM Press.

W3C (2004). *Gleaning Resource Descriptions from Dialects of Languages (GRDDL)*. Retrieved August 31, 2007, from http://www.w3.org/2004/01/rdxh/spec/

W3C (2007). *RDFa Primer 1.0 - Embedding RDF in XHTML*. Retrieved August 31, 2007, from http://www.w3.org/TR/xhtml-rdfa-primer/

W3C Technical Architecture Group (2004). *Architecture of the World Wide Web, Volume One*. Retrieved August 25, 2007, from http://www.w3.org/TR/webarch/