

# COMPLEX DIGITAL OBJECTS

Dr Jane Hunter

## Background

Dr Jane Hunter is a Senior Research Fellow at the CRC for Enterprise Distributed Systems Technology with over ten years experience in research and development of digital library technologies for the cultural, scientific and educational sectors. Through collaborations with organizations including the Smithsonian, the Library of Congress, UK eScience Centres, and the UK Digital Curation Centre, she is developing data models, ontologies, metadata standards, indexing and annotation tools and search interfaces for the archival, management, knowledge mining and preservation of large mixed-media collections. Dr Hunter is also currently a CI on three ARC Special Research Initiatives and two ARIIC projects (DART and RUBRIC) funded in 2005.

Current research projects include:

- The [Sunago - Multimedia and the Semantic Web](#) project which is developing ontologies and annotation, inferencing, and information aggregation tools specifically for mixed-media resources (images, video, audio, text, SMIL, 3D, datasets) using emerging standards and technologies including OWL (Web Ontology Language), MPEG-7 and MPEG-21.
- The [FUSION](#) project – which is developing eScience tools and services to enable the optimization of nano-materials design through systematic capture, processing and correlation of manufacturing data, performance data and microstructural analyses.
- The [FilmEd/Vannotea](#) project - which is investigating the development of collaborative indexing, annotation and discussion tools for digital film/video, images, 3D objects and text over the GrangeNet broadband research network.
- The [Indigenous Knowledge Management](#) project - a collaboration with the Smithsonian National Museum of the American Indian, which is developing open source software tools to enable indigenous communities to define and enforce the ownership, access rights and traditional care and customary constraints associated with their cultural collections and Indigenous knowledge.
- The [PANIC \(Preservation and Archival of New media and Interactive Collections\)](#) project which is investigating techniques to semi-automate the archival and preservation of large collections of digital objects using preservation metadata and semantic web services.

## Topic of this Presentation: “Linking Publications to Data through Scientific Model Packages”

Scientific models provide a primary method for representing and encapsulating expert knowledge in many domains, including particle and high energy physics, astronomy, biological sciences, earth sciences, social sciences and economics. They provide simplified representations of real world objects, systems or phenomena – often in mathematical terms - enabling scientists to better predict how complex systems behave under different conditions or parameters, or change over time, in order to solve problems associated with a particular research focus. They also provide an ideal mechanism: for authenticating and tracking individual contributions to scientific collaborations; for publishing and disseminating scientific results; for integrating research into teaching; and for selective archival and preservation of scientific data.

Within this presentation, I will describe an innovative approach to scientific data management that we are developing that involves the representation of scientific models as complex digital objects called “Scientific Model Packages” (SMPs). The proposed SMPs encapsulate both:

- a) the complete set of contextual or lineage information associated with the model's development and
- b) semantic descriptions of each of the individual components that comprise and validate the scientific model and the relationships between them.

More specifically SMPs may contain any of the following components or references to them:

- Pre-existing data, models, hypotheses or publications;
- Large data sets generated from experiments, observations and instruments. This may include: numerical data, survey data, questionnaires, images, video, audio, maps, spectral data, real-time sensor data;
- Experimental and instrumental conditions, settings and parametric ranges or constraints;
- Formulas, rules, hypotheses, numerical models, mathematical functions;
- Conceptual models - paradigmatic, explanatory information or ideas in the form of axioms, models and metaphors;
- Software tools and services – that perform the analysis, interpretation, transformation, visualization, simulation and modelling of the data. This includes actual source code or executables, applets or links to web services as well as documentation describing the software;
- Hardware specifications – the instruments used to generate the data, the instrumental settings, and the computers that execute the analysis, processing, integration and visualization of the data;
- Workflows – steps involved in transforming the raw data into knowledge;
- Visualizations – 2D, 3D imagery, graphs, tables, charts, diagrams, animations;
- Textual data – notes, publications, reports, documentation, annotations, bibliographies, reviews etc..

This presentation will describe some of the tools, services and technologies that we are developing that will enable scientists to efficiently and easily capture, index, store, share, exchange, re-use, compare and combine different scientific models. The aim is to analyse and support the needs of a wide range of scientific communities, in order to support both discipline-specific and multi-disciplinary applications. The scale and dynamic nature of the problem will be tackled by determining commonalities and differences across communities and building the tools and services on top of an underlying, extensible, object-oriented infrastructure – the Semantic Grid. The aim is to also encourage the publication and sharing of “scientific model packages” by providing tools that will enable their easy construction and submission to either open access or access-controlled institutional repositories.