Building a Semantic Knowledge-Base for Painting Conservators

Jane Hunter
School of ITEE
The University of Queensland
Brisbane, Australia
j.hunter@uq.edu.au

Suleiman Odat
School of ITEE
The University of Queensland
Brisbane, Australia
s.odat@uq.edu.au

***Abstract*—The *Twentieth Century Paint* project is a collaboration between the Asia Pacific Twentieth Century Conservation Art Research Network (APTCCARN) and the eResearch Lab at the University of Queensland. It is a collaborative effort to explore the preservation of twentieth-century paintings in Asia and the Pacific. One of the key objectives is to establish an online knowledge-base that will provide conservators with access to integrated, structured information and a portfolio of experiments and case studies that document the different causes of paint degradation and the optimum conservation treatments. This paper describes the knowledge-base and the associated ontology and services developed by the eResearch Lab in collaboration with APTCCARN. This work provides a flexible but robust framework that will enable future expansion of the knowledge base through both harvesting of structured data and collaborative input by domain experts.**

Keywords- paint conservation, e-Research, semantic tagging; ontologies, knowledge base

I. INTRODUCTION

The aim of the Asia Pacific Twentieth Century Conservation Art Research Network (APTCCARN) [1] is to explore the preservation of twentieth-century paintings in Asia and the Pacific. Modern paintings, in particular, are highly susceptible to problems such as aging, cracking and fading due to the increased instability of modern synthetic organic pigments and paint formulations. Therefore, it is vitally important that these modern pigments, along with their synthetic binders and additives, are characterized before and after problems arise to determine optimum conservation treatments and environmental conditions for storage, display and transport. Non-invasive analytical methods such as Scanning Electron Microscopy (SEM), X-ray diffraction and Raman spectroscopy enable improved identification of modern synthetic organic pigments in acrylic and alkyd paint formulations and oil media.

However, due to increasing access to such sophisticated techniques, painting conservation has evolved into a highly multidisciplinary research topic that requires the integration of knowledge about art history (artworks, artists, artistic techniques), the physical and chemical properties of paint and pigments, paint conservation and cleaning methods and different characterization techniques that can be used to determine the precise cause of the degradation or discoloration that is occurring.

The challenge is that relevant data and metadata is highly heterogeneous and distributed across databases, scholarly publications and the Web. Expertise, also, is distributed across art galleries, conservation centres and universities around the globe. Although it is possible to find some concentrated authoritative collections of information on this topic on the Web (e.g., Journal of the American Institute of Conservation, Smithsonian Museum Conservation Institute (MCI), Getty Conservation and Research Institutes, CAMEO materials database [2] and Forbes Pigment database [3]), the information is often embedded within databases or within highly unstructured textual documents and the relevant information is difficult to extract, re-use, interpret, correlate or compare. Moreover, it is often the case that the raw images or the raw spectrographic data associated with the analysis of a particular painting or paint samples, are not accessible via the related publication. For example, the experimental data underpinning publications that describe the long term effects of different environmental conditions (humidity, temperature, UV light) on different paints, is not accessible, verifiable or re-usable. It is also the case that mistakes made by conservators are often hidden rather than published. Marincola (Professor of Conservation at NYU) recently urged conservators to admit to their mistakes and to share case studies that document their conservation errors – to help avoid future similar mistakes by others [4].

The distributed, unstructured, heterogeneous nature of the relevant data, makes it extremely difficult for conservators to search and aggregate information to find answers to the problems that they face. For example, consider the following hypothetical example. An art conservator at the Queensland Art Gallery (QAG) recently wanted to know: "what is the best solvent for removing the surface coating from the painting *Epiphany*". The QAG database reveals that *Epiphany* was painted by Ian Fairweather in 1962, and purchased by the QAG in 1984. The Dictionary of Australian Artists Online (DAAO) database [5] tells us that during this period, Ian Fairweather frequently used Dulux acrylic paints coated with shellac. The CAMEO (Conservation & Art Material Encyclopedia Online) database tells us that the best solvent for removing shellac is methyl ethyl ketone. But the process of discovering these different pieces of information and linking them to answer the original question, is both extremely time consuming and cumbersome and involves reading through long textual resources (e.g., a biography of Ian Fairweather).

Ontologies have been successfully applied in many fields (biomedical, environmental sciences, literature etc.) to enable data integration, knowledge acquisition, automated annotation and cross-linking for knowledge discovery purposes. A literature survey undertaken by the authors indicated that there is no existing ontology designed to support knowledge management and discovery for art/painting conservators. However, there are a number of existing ontologies or vocabularies that provide data models for describing particular aspects of art conservation. For example, CIDOC/CRM [6] provides a data model for describing the provenance of artworks. The ChEBI ontology [7] describes chemical entities of biological interest. ChemAxiom [8] provides an ontological framework for Chemistry in Science. The Materials ontology [9] provides a common data model for exchanging information about materials (their structure, composition and properties). The Art and Architecture Thesaurus (AAT) [10] provides a structured, controlled vocabulary for describing artworks, but does not include support for describing the materials they are composed of (e.g., paint types). For example, in AAT "oil painting" is a technique, not a material. Our aim is to build an ontology that describes all of the information classes, properties and relationships of relevance to painting conservators, by drawing on existing onotologies and vocabularies where available – but also extending and refining them as required.

Given this underlying ontology (which we call OPPRA (Ontologies for Painting and Preservation of Art)), our second goal is to evaluate it by applying it to:

- the capture of new information being generated by our collaborators within APTCCARN. Experiments carried out by team members will be described using the terms from the ontology; and
- the extraction of structured information from past publications on painting conservation through the semantic modeling of the information in each publication. The ontology will be used to identify and markup key entities and concepts described in the publication and to build a structured case study.

The remainder of this paper is structured as follows. We firstly describe the specific objectives of this e-Research project. In section III we compare this work with previous related work. Section IV describes a case study that illustrates a typical scenario, the key challenges, the workflow and outcomes. Section V describes the ontologies that we have developed. Section VI describes the semantic markup tools. Section VII describes the experimental data capture interface. Section VIII provides a discussion of the outcomes to date and Section IX provides a conclusion.

II. Objectives

A workshop held in April 2010 at the Art Gallery of NSW, between the members of APTCCARN and the UQ eResearch Laboratory determined that the first priority was to develop a Web Portal that provides the team of conservators with fast, easy and authenticated access to structured knowledge stored in an integrated semantic knowledge base. High priority requirements were:

- The ability for team members to document and describe their own experiments and upload their experimental data to the knowledge base - so it can be shared and re-used between team members and/or made accessible to the public after publication of the results;
- The ability to extract structured data about past research and experiments from relevant publications and web sites on art conservation, and to ingest it into the knowledge base to enable fast, easy access to and comparison of related cases.

Both of the above functional requirements depend on the availability of a well-structured data model/ontology that provides a common, machine-readable formal representation of the knowledge in this domain (OWL/RDF). Hence our first objective was to develop the OPPRA ontology that draws on existing ontologies and vocabularies to describe:

- Paintings – title, artist, period, technique, genre, condition, owner, custodian, provenance;
- Paint – source (manufacturer, year, paint name, identifier, supplier), paintType, structure, chemical composition, properties, pigment, additives (thickeners, stabilizers, preservatives, surfactants, coalescing solvents and defoamers);
- Paint decomposition – types of degradation (cracking, peeling, fading, discoloration, mold growth) and the causes (humidity, light, temperature, water, technique) and physical/chemical processes/reactions;
- Paint analysis methods –macroscopic, scanning and transmission electron microscopes (SEM, TEM), Fourier Transform Infrared spectroscopy (FTIR), Raman spectroscopy, X-Ray Diffraction, X-ray Spectroscopy (EDS), X-Ray Fluorescence (XRF) and Synchrotron radiation;
- Paint conservation/preservation treatments – cleaning, surface coating, relocation;
- Experiments – experimenter, objective, paint samples, sample source, parameters, results, images, analysis data;
- Publications – identifier, author, title, date, publisher, subject, experiment.

Given this ontology, our next objective was to develop the required services – that enable the capture of new experimental data from team members and the extraction of legacy knowledge from past publications, creating RDF instances that provide a common semantic view of the information. These RDF statements are uploaded to the underlying knowledge base (a Sesame RDF triple store) where they can be accessed/queried via a Web interface. Hence the final objective was to develop the search, browse and analysis tools over the RDF triple store – that enable users to find answers to their questions via ontology-based SPARQL queries. Sections V, VI and VII describe our results to date and the extent to which we have met these objectives.

III. PREVIOUS RELATED WORK

In the past five years, a number of research projects have focused on improving access to cultural heritage collections by moving towards a deeper semantic representation of the stored data, through ontologies and semantic annotation. Examples include: MultimediaN E-culture[11], CultureSampo [12], ResearchSpace [13] and the Europeana Athena project [14]. These systems aim to improve the discoverability of cultural heritage content by integrating heterogeneous digital collections via rich semantic metadata and common models. But they don't focus on the specific requirements associated with the conservation of cultural artefacts, as outlined in the report from a recent NSF workshop on conservation [15].

The Andrew W. Mellon foundation also funded six pilot projects in 2009 that focus on issues specifically associated with conservation documentation: The Master of the Fogg Pietá [16], Cranach Digital Archive [17], Rembrandt Database [18], Merlin Database [19], Raphael Research Resource [20] and Southworth & Hawes Daguerreotypes [21]. These projects have primarily developed databases that integrate information about one particular artist or genre. However Mellon also recently funded the ConservationSpace project [37], which "aims to develop an open-source software application that will address a core need of the conservation community for a shared solution to the problem of documentation management". It appears that this project is still be in the planning phase and has not yet begun implementation or published any results.

A number of online databases have been developed that document particular types of information of relevance to paint conservation (Winsor and Newton Archive [22], INCCA Database for Artists' Archives (IDAA) [23], Getty Research Institute's Vocabularies [10], the IRUG Spectral Database [25], the IR-Spectra database [26], the Forbes Pigment Database [3], the Bibliographic Database of the Conservation Information Network (BCIN) [27], CAMEO: Conservation and Art Material Encyclopedia Online [2], Conservation Online (COOL) [28]). However all of these databases and online web sites are designed to provide a specific type of information to art conservators – they don't provide services to support the ingest, search or retrieval of structured, standardized information describing the provenance and results of experiments focused on paint conservation.

Some relevant related research has been undertaken in both the chemistry and the cultural heritage informatics domains. Of particular interest is the oreChem ChemXSeer project [29], which combines the OAI-ORE data model and ontologies to describe chemicals, chemical processes and experiments. Chemistry publications can be analysed and tagged based on the chemistry and other ontologies. Experiments are extracted and represented as OAI-ORE compound objects with other document-related metadata.

Also related to our work, is the CIDOC-CRM [6] and recent applications of the CIDOC-CRM to documenting the provenance of cultural artefacts [30]. Our work extends this previous work by Theodoridou et al, by adding support for describing paints (physical and chemical properties, structure and composition), paint degradation processes, paint analysis/characterization methods and preservation/ conservation methods. Our goal is to use the CIDOC-CRM as the upper ontology but to also combine it with ontologies from the oreChem project [31] to describe experiments, chemical compounds and chemical processes.

IV. CASE STUDY

Consider the following case study. Gillian Osmond is an art conservator at the Queensland Art Gallery and one of our collaborators from APTCCARN. She has been investigating the appearance of metal soap formation (surface lump aggregation) in some of the nineteenth- and early twentieth-century British and Australian paintings in the Gallery's Collection. The soaps are caused by a reaction between the oil medium and metal ions present in the paint due to lead and, more recently, zinc-containing pigments or additives. As part of her investigation, Gillian has taken paint samples from the painting "Woolshed NSW", an oil painting by R.Godfrey Rivers (painted in 1903) and purchased by the QAG in 1903. She examined the paint samples using SEM-EDX and UV fluorescence. This characterization has shown that zinc is consistently found at the centre of fluorescent regions, and her hypothesis is that a white pigment (commonly known as zinc white), contains zinc oxide that is reacting to form organic soaps. These soap compounds have a larger surface area and volume which cracks the original paint. Gillian has published her preliminary findings [32].

In order to better understand why the soaps form, their long term stability and how to prevent/reduce the formation of the soaps, Gillian is now undertaking a series of experiments to simulate the aging and degradation process She has obtained samples of zinc oxide, is mixing them with paint samples – and is subjecting them to a range of different environmental conditions (humidity, temperature, UV light) to determine which compositional and environmental parameters have the greatest impact on the formation of zinc soaps. After subjecting them to different controlled environmental conditions for varying durations, Gillian is examining the treated samples using a range of characterization techniques including SEM, TEM, UV fluorescence and FTIR, to determine the presence and/or nature of zinc compounds that form and their detrimental effects on works of art.

In order to store, analyse and interpret the results of her experiments, share the results with her collaborators and eventually publish the results (both the data together with the traditional textual publication), Gillian requires an online repository where she can describe each investigation, the source of the paint samples, the experimental conditions and the characterization results – using standardized, machine-processable metadata schemas, vocabularies and formats. She needs to be able to compare her research outcomes with similar research described in related publications. She also needs to be able to protect her results (through authenticated access control mechanisms) until she is ready to publish them. Finally she needs to be able to publish persistent links (URLs) from her publication that enable readers to retrieve the raw images or spectrographic data.

The next three sections describe the work that we have completed to date, in an effort to satisfy the requirements of Gillian, the APTCCARN team members and painting conservators in general.

V. ONTOLOGIES

The first step in developing the knowledge-base involved developing the OPPRA ontology. Our approach was not to develop OPPRA from scratch but to draw on previous related work undertaken in the cultural heritage and chemistry informatics domains. Hence, the OPPRA ontology comprises the following sub-ontologies:

- CIDOC-CRM [6] – this provides the top-level classes as well as the classes and properties required to capture the provenance information about a painting and its conditionState as well as the conservation/preservation activities that it undergoes (See Figure 1);
- OreChem [31] - was used to model the chemical compounds, chemical reactions and experiments;
- OPPRA-specific ontologies were developed to describe:
  - Paints – source (manufacturer/supplier, year, paint_name, identifier, bottle label), paintType, structure, chemical composition, formula, properties, pigment
  - Additives - thickeners, stabilizers, preservatives, surfactants, coalescing solvents and defoamers
  - Paint degradation – types of degradation (cracking, peeling, fading, discoloration, mold growth); causes (humidity, light, temperature, water, technique); and associated physical/chemical processes/reactions
  - Paint analysis methods –macroscopic, microscopic, SEM, TEM, FTIR, Infrared, Raman, X-Ray Diffraction, X-ray Spectroscopy (EDS), X-Ray Fluorescence (XRF), Chromatography, Synchrotron
  - Paint conservation/preservation treatments – cleaning, protective coatings, environmental conditions

Figure 1 below, illustrates how the CIDOC-CRM ontology can be applied to document the condition assessment and cleaning of the painting “Epiphany”.

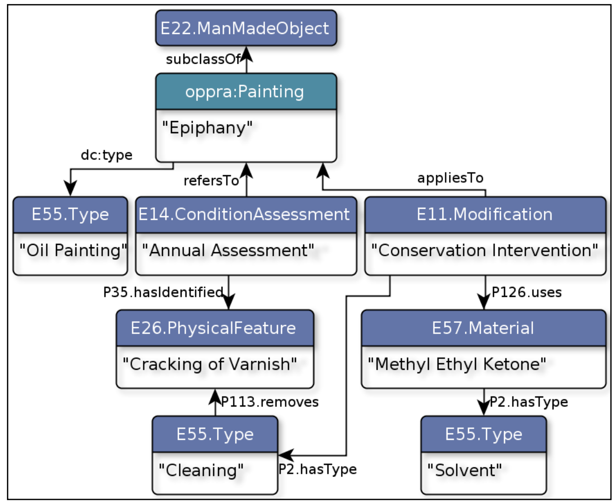

Figure 1: Application of the CIDOC-CRM to painting conservation

Figure 2 below, illustrates how we combine and link the CIDOC-CRM, the oreChem ontologies and the OPPRA extensions, developed to describe paint-specific information.

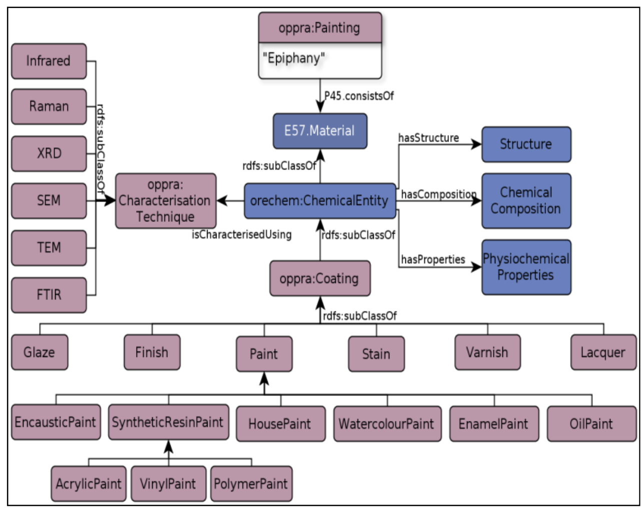

Figure 2: OPPRA Extensions to CIDOC-CRM and oreChem

Due to its size and complexity, it is not possible to illustrate the complete OPPRA ontology in this paper, however the complete ontology is accessible via the Web Portal[1].

Figure 3 also shows a screen dump of the Web interface that allows readers to browse the ontology. This interface was originally developed to enable our collaborators to comment on and provide feedback to our draft ontology – which evolved through an iterative cyclical process. The Web version is dynamically produced using Protégé-OWL 4.1 and the OWLDoc plug-in that exports views of the ontology as HTML. Figure 3 shows the class hierarchy associated with the class *SyntheticResinPaint* – it is a subclass of *Paint* and a superclass of *AcrylicPaint*, *VinylPaint* and *PolymerPaint*.

[1] http://www.20thcpaint.org/oppra-owl/

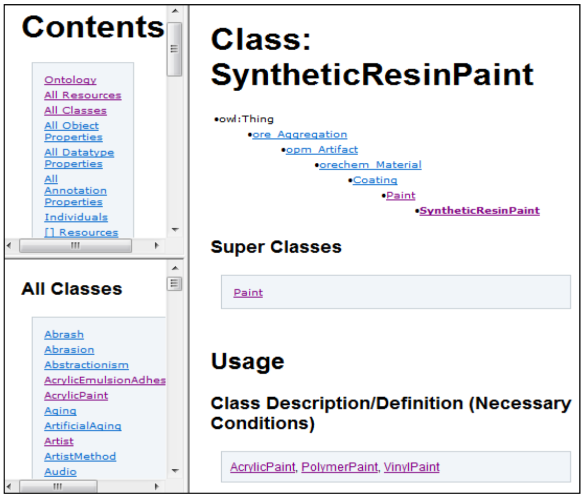

Figure 3: Web Interface to the OPPRA Ontology

VI. POPULATING THE KNOWLEDGE BASE

Given the OPPRA ontology, we are now in a position to generate RDF instances which can be used to populate the knowledge base. Instance data is generated via two methods that are described in detail below:

1. Ingesting experimental data from the local APTCCARN team members/collaborators
2. Harvesting structured semantic information (compliant with the OPPRA ontology) from past publications

A. Ingesting Experimental Data

Consider the case study described in section IV. In order to support the storage and indexing of experimental data being generated by the art conservators on the team, we needed first to provide an interface whereby authenticated users can create a new investigation/set of experiments and record the metadata associated with each experiment. Figure 4 illustrates the metadata and data associated with a typical experiment.

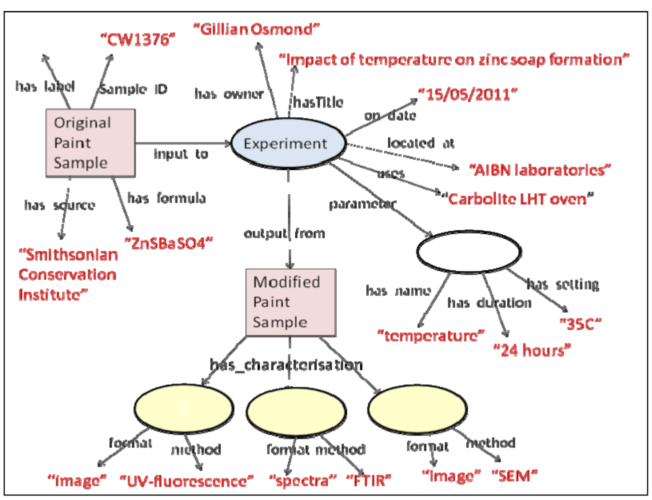

Figure 4: Modelling Experimental Data

Each experiment has associated metadata including: *owner*, *title, date*, *location* and *instruments*. Each experiment also has an *input sample* and set of experimental *parameters*. In the scenario described in Figure 4, the parameter is *"temperature"* and the experiment involves exposing the paint sample to a temperature of 35C for 24 hours. The sample (which was sourced from the Smithsonian Conservation Institute Reference Collection) is then analysed using a variety of different characterization techniques to determine the presence of zinc soaps. Experimental results include the images and spectrographs generated from the post-experimental characterization of the exposed sample. Figure 7 in the next section shows a screen shot of the user interface that enables authenticated users to create and edit experiments and to upload associated characterization data (images and spectra).

B. Harvesting from Past Publications

The aim of this aspect of the work was to identify a corpus of relevant publications about paint conservation and to extract structured semantic metadata from each publication (using the OPPRA ontology), so that it can easily be discovered by the team of conservators with whom we were working. This structured information may also be used to build case studies that describe the results of particular conservation treatments – both positive and negative results.

Altogether we identified 152 publications from 15 different journals. The most relevant journals were: Journal of the American Institute for Conservation, JSTOR Studies in Conservation, Analytical Chemistry and the AICCM Bulletin. Although it is possible to extract structured semantic descriptions from publications using automated techniques (such as supervised machine learning), in the initial phase of this work, we have generated the RDF descriptions manually. For example, consider the publication :Monico, L., G. Van der Snickt, et al. (2011). "Degradation Process of Lead Chromate in Paintings by Vincent van Gogh Studied by Means of Synchrotron X-ray Spectromicroscopy and Related Methods. Part 2. Original Paint Layer Samples." Analytical Chemistry 83(4): 1224-1231.

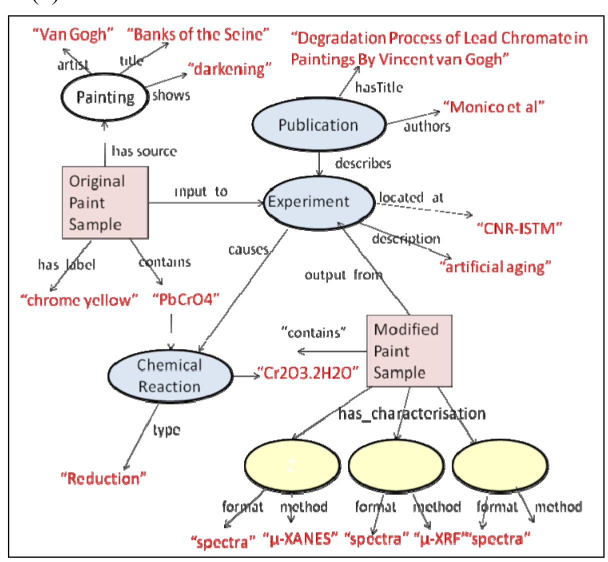

Figure 5: Structured data extracted from example publication

Manual tagging of the textual content within this paper generates RDF instance data corresponding to the structured/modelled information shown in Figure 5.

VII. DESIGN OF THE KNOWLEDGE BASE

The Web Portal [2] provides the online interface to the "Twentieth Century Paint" project. This portal provides both public access to background information associated with the project as well as authenticated access to the members-only knowledge-base developed to satisfy the data storage requirements for the different teams associated with the project:

- Team 1: Art history and conservation
- Team 2: Material developments and deterioration
- Team 3: Scientific tools and techniques
- Team 4: IT tools and techniques

The Web Portal itself and the work described in this paper, has been developed by Team 4. The Web Portal provides a single user interface to the locally deployed web sites and storage. AJAX, JSP, Javascript and CSS are the underlying technologies - chosen to ensure dynamic web interfaces and highly responsive interactivity.

Figure 6 shows the overall architecture of the system. The Web Portal has been developed using a combination of: Apache Tomcat, MediaWiki, MySQL, the Jackrabbit Repository and the Sesame RDF triple store.

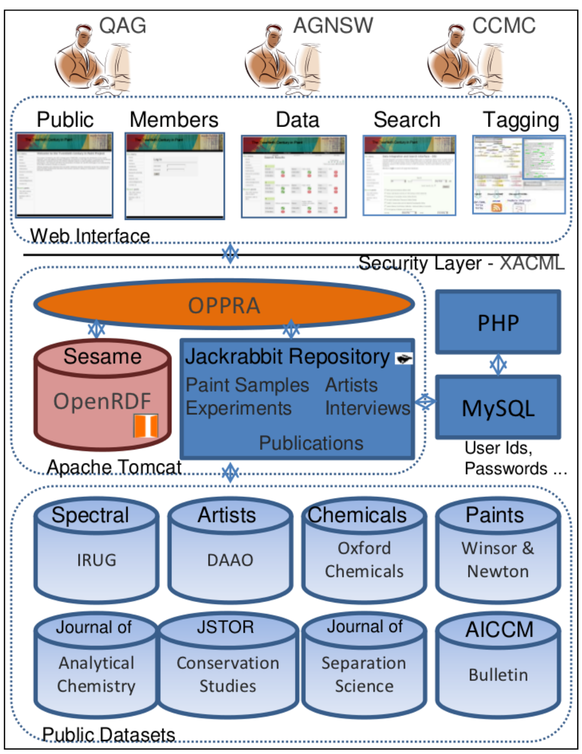

Figure 6: Overview of System Architecture

[2] http://www.20thcpaint.org/

MediaWiki [33] is used to enable informal knowledge sharing among the team members. MySQL is used to store information about users, pages, contents, revisions, files' metadata and security information such as access rights and encryption keys.

Apache Jackrabbit is used as the content repository for storing images and spectra [34]. Apache Jackrabbit offers support for multiple pluggable storage back-ends, fast data modifications , the ability to associate metadata with different file formats, the XPath enabled mechanism to search for files and content, and its security features that are extendable to work with fine-grained access control such as XACML.

The OPPRA-compliant RDF instance data, generated via the mechanisms described in Section VI, are stored in an OpenRDF repository – the Sesame Triple Store [35]. The metadata in the RDF triple store contains links to the image and spectra files stored in the Jackrabbit repository.

The local RDF triple store and associated Jackrabbit repository is only accessible to authenticated project members. After users login, they are able to access the internal Wiki and the local repository and also create, edit and delete experiments. Only the owner of an experiment can edit or delete experiments, but other members of the owners' team can view their experiments. Figure 7 shows the user interface for browsing experiments and viewing microscopy images (TIFF, JPEG files from SEM and TEM) and spectra files (SPA) from FTIR, XRF, Raman analysis – that are associated with particular experiments and paint samples. A RESTful interface enables users to: Create Records, Search and Browse Records, Upload and Delete Files, Add and Update (field-based) Comments, View History and Actors, Restore Deleted Records and Permanently Delete Records.

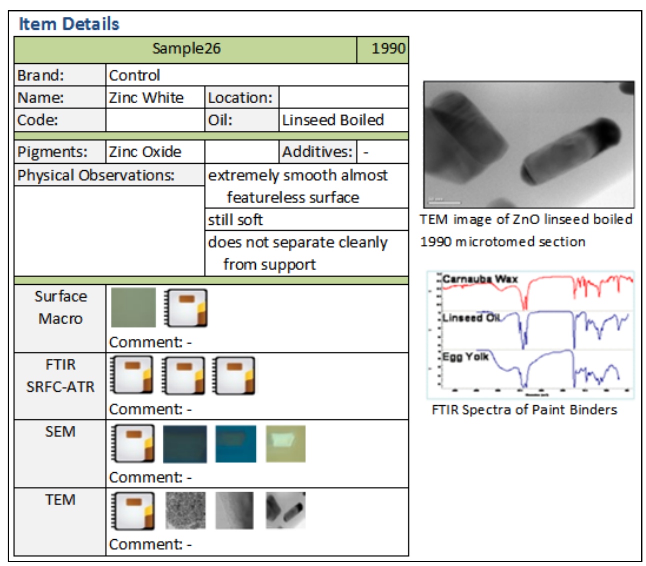

Figure 7: Screen Shot of Experimental Browsing and Upload Interface

A SPARQL query interface over the Sesame triple store has also been implemented – enabling users to search for particular paintings, artists, paints, types of degradation, chemical compounds or characterisation methods. Generating SPARQL queries poses a significant barrier for non-technical users. To overcome this we have developed a user-friendly interface that automatically maps user input into a SPARQL query. Figure 8 shows the interface for the query: “Give me all paintings by Sidney Nolan that show lead soap formation”

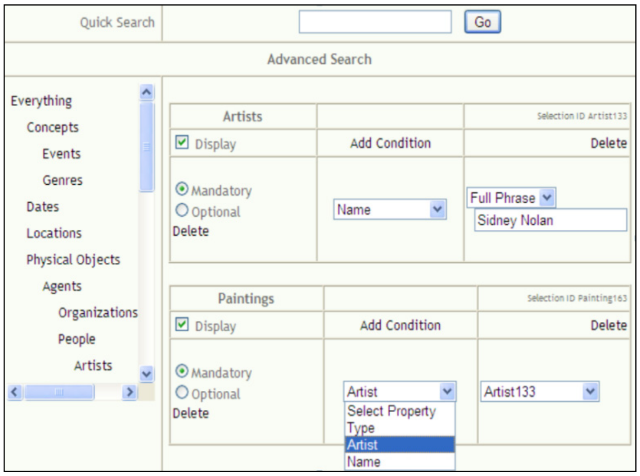

Figure 8: SPARQL Query Interface to the 20th Century Paint Knowledge Base

The Web Portal also provides a single interface to related online collections and databases of relevance to paint preservation[3]. Examples include: art provenance databases (e.g. Dictionary of Australian Artists Online [5], INCCA Database for Artists' Archives [23], Winsor and Newton Archive [22], chemistry databases (e.g. NIST Chemistry WebBook [36], image databases (e.g. IRUG Spectral Database Edition 2000 Search Engine [25]), artistic techniques databases (e.g. The Getty Art and Architecture Thesaurus Online [24]), and publications (e.g. Bibliographic Database of the Conservation Information Network (BCIN) [27]). Future work includes enabling federated searches across both the local database, harvested data and these external databases, through a common metadata store that is based on the OPPRA ontology.

VIII. Discussion

The field of paint conservation is highly multidisciplinary – it demands access to and integration of distributed and heterogeneous information and knowledge from disciplines including art history, chemistry and materials science. As a result, it is an ideal application for semantic web technologies, and in particular, ontologies that enable machine-processable schema mappings, semantic tagging and structured knowledge extraction.

In this paper we firstly describe the OPPRA ontology that we developed (and the process by which it was developed) to support the information and data management requirements of the APTCCARN network. We evaluated and refined the OPPRA ontology via three mechanisms:

- online review and feedback by the APTCCARN team members;
- the ability of the OPPRA ontology to capture local experimental data;
- the ability of the OPPRA ontology to harvest structured data from external publications.

These evaluation processes (combined with the re-use of long-standing and well-tested ontologies such as CIDOC-CRM) have led to the gradual refinement of the ontology and a relatively stable, robust and effective data model. Given the relatively stable OPPRA ontology, we then developed services to support both the capture and upload of local experimental data sets as well as the harvesting and storage of extracted structured RDF instances from publications. The resulting RDF triples have been ingested into a Sesame triple store, linked to an Apache Jackrabbit repository and simple SPARQL queries have been implemented. Metadata input and data upload is simple, fast and efficient. The knowledge base is highly flexible and extensible due to the ontology underpinning it. The ontology-based browse interface is both fast and intuitive and does not require users to have detailed knowledge of terms from a range of disciplines. In the remaining 18 months of the project, we plan to:

- Implement more complex SPARQL queries over the RDF triple store that require the incorporation of a reasoning engine e.g., “What is the best solvent for removing lead oxide crystals in oil paintings?”
- Add support within the Web Portal for uploading artists’ interviews including video and audio files (e.g. AVI, WAV) plus transcripts of interviews (e.g. DOC and PDF). This provides additional content that can potentially be used to extract structured information.
- Investigate the application of semi-automatic techniques (e.g., supervised machine learning) to extract structured RDF statements (compliant with the OPPRA ontology) from publications on painting preservation.
- Undertake a comprehensive user evaluation of the system by working closely with the APTCCARN team members and art conservators more generally.

IX. Conclusions

This paper describes the results of a collaboration between the University of Queensland eResearch Lab and the Asia Pacific Twentieth Century Conservation Art Research Network (APTCCARN), that aims to develop a set of services to enable the extraction, creation and storage of knowledge about paint conservation – in an online semantic knowledge-base, so that it can be discovered, shared, re-used and reasoned across. The outcome is a framework that will hopefully enable paint conservators to improve their understanding of paint degradation processes, and to identify and document new methods for stabilizing, protecting and repairing our valuable but vulnerable paintings.

Acknowledgment

The work described in this paper has been funded through the Australia Research Council, ARC Linkage Grant LP0883309. We would also like to thank our collaborators within the “Twentieth Century in Paint” project – in particular Gillian Osmond, John Drennan, Paula Dredge, Robyn Sloggett and Nicole Tse for their valuable ideas and feedback.

[3] http://www.20thcpaint.org/databases.jsp

REFERENCES

[1] Asia-Pacific Twentieth Century Conservation Art Research Network (APTCCARN) [Online]. Available: http://cultural-conservation.unimelb.edu.au/partners/international/aptccarn/

[2] Museum of Fine Arts (MFA) Boston, CAMEO Materials Database Search [Online]. Available: http://cameo.mfa.org/materials/index.asp

[3] Museum of Fine Arts (MFA), Boston, Forbes Pigment Database [Online]. Available: http://cameo.mfa.org/pigments/search.asp

[4] M. Marincola , "Blink Twice: Making Mistakes in Conservation" AIC 2010 Notes [Online], Available: http://prescan.wordpress.com/2010/05/20/aic-2010-notes-blink-twice-making-mistakes-in-conservation/

[5] Dictionary of Australian Artists Online (DAAO) [Online]. Available: http://www.daao.org.au/main

[6] ICOM/CIDOC CRM Special Interest Group, "CIDOC Conceptual Reference Model, version 5.02", January 2010. Available: http://www.cidoc-crm.org/docs/cidoc_crm_version_5.0.2.pdf

[7] Degtyarenko, K., de Matos, P., Ennis, M., Hastings, J., Zbinden, M., McNaught, A., Alcántara, R., Darsow, M., Guedj, M. and Ashburner, M. (2008) ChEBI: a database and ontology for chemical entities of biological interest. Nucleic Acids Res. 36, D344–D350.

[8] Adams, N., Cannon, E. O., & Murray-Rust, P. (2009). ChemAxiom – An Ontological Framework for Chemistry in Science. International Conference on Biomedical Ontology (Vol. 1, p. 2). Nature Publishing Group.

[9] T.Ashino, Materials Ontology: An Infrastructure for Exchanging Materials Information and Knowledge, Data Science Journal Vol. 9 (2010) pp.54-61

[10] Getty Research Institute, Getty Vocabularies [Online]. Available: http://www.getty.edu/research/tools/vocabularies/index.html

[11] J. van Ossenbruggen, et.al., "Searching and Annotating Virtual Heritage Collections with Semantic-Web Techniques," in Proceedings of Museums and the Web 2007, 2007.

[12] L. Aroyo, et. Al., "CultureSampo: A National Publication System of Cultural Heritage on the Semantic Web 2.0," in The Semantic Web: Research and Applications. vol. 5554: Springer Berlin / Heidelberg, 2009, pp. 851-856.

[13] D. Oldman, "ResearchSpace: Content Management Systems – Evaluation Considerations," The British Museum, 2010 [Online]. Available: http://www.researchspace.org/

[14] ATHENA – Access to Cultural Heritage Networks Across Europe [Online]. Available: http://www.athenaeurope.org/

[15] M. Leona and R. Van Duyne, Chemistry and Materials Research at the Interface between Science nd Art, Report of a Workshop Cosponsored by the NSF and the Andrew W. Mellon Foundation July, 2009 Arlington, Virginia. Available: http://mac.mellon.org/NSF-MellonWorkshop/2009%20NSF%20final%20report%20full%20res.pdf

[16] A. Nevin, "The Master of the Fogg Pietá ~ Maestro di Figline Project," Andrew W. Mellon Foundation, 2009. Available: http://mac.mellon.org/Master%20of%20the%20Fogg%20Pieta%20Abstract.pdf

[17] G. Heydenreich, "The Cranach Digital Archive", the Andrew W. Mellon Foundation, 2009. Available: http://mac.mellon.org/issues-in-conservation-documentation/Cranach%20project%20description%20Nov%202009.pdf

[18] W. Donkersloot, "The Rembrandt Database Inter-institutional Research Resource on Paintings by Rembrandt," Andrew W. Mellon Foundation., 2009. Available: http://mac.mellon.org/issues-in-conservation-documentation/Rembrandt%20Database%202-23-09.pdf

[19] "Mellon Pilot Project on Conservation and Science Documentation," the British Museum, 2007. Available: http://mac.mellon.org/issues-in-conservation-documentation/the_british_museum_pilot.pdf

[20] M. Hofmann, "The National Gallery's Mellon Digital Documentation Project: The Raphael Research Resource," Andrew W. Mellon Foundation, 2009. Available: http://mac.mellon.org/NGL%20Raphael%20Project%20update%203-3-09.pdf

[21] "Mellon Pilot Project: Open-source Solutions for Sharing Data and Images,", the Museum of Fine Arts, Boston 2005. Available: http://mac.mellon.org/SH%20abstract%203-24-09.pdf -.

[22] Winsor & Newton Archive of 19th Century Artists' materials, [Online]. Available: http://www-hki.fitzmuseum.cam.ac.uk/archives/wn/database.php

[23] INCCA Database for Artists' Archives (IDAA) [Online]. Available: http://www.incca.org/idaa

[24] Martin Doerr, D. Iorizzo, The dream of a global knowledge network—A new approach, 2008, ACM Journal on Computing and Cultural Heritage, Vol. 1, No. 1, Article 5, June 2008

[25] IRUG Spectral Database Edition 2000 Search Engine [Online]. Available: http://www.irug.org/ed2k/search.asp

[26] IR-Spectra database [Online]. Available: http://tera.chem.ut.ee/IR_spectra/

[27] Bibliographic Database of the Conservation Information Network (BCIN) [Online]. Available: http://www.bcin.ca/

[28] Conservation Online [Online]. Available: http://cool.conservation-us.org/

[29] L. Na, Z. Leilei, M. Prasenjit, M. Karl, P. Eric, and C. L. Giles, "oreChem ChemXSeer: a semantic digital library for chemistry," in Proceedings of the 10th annual joint conference on Digital libraries Gold Coast, Queensland, Australia: ACM, 2010.

[30] M. Theodoridou, Y. Tzitzikas, M. Doerr, Y. Marketakis, and V. Melessanakis, "Modeling and querying provenance by extending CIDOC CRM," Distributed and Parallel Databases, vol. 27, pp. 169-210, 2010.

[31] Lagoze, Carl (2009) The oreChem Project: Integrating Chemistry Scholarship with the Semantic Web. In: Proceedings of the WebSci'09: Society On-Line, 18-20 March 2009, Athens, Greece.

[32] Osmond, G., K. Keune, et al. (2005). "A study of zinc soap aggregates in a late 19th century painting by R.G. Rivers at the Queensland Art." AICCM Bulletin 29: 37-46.

[33] MediaWiki [Online]. Available: http://www.mediawiki.org/wiki/MediaWiki

[34] Apache. Jackrabbit [Online]. Available: http://jackrabbit.apache.org/

[35] Sesame RDF Triple Store [Online]. Available: http://www.openrdf.org/

[36] NIST Chemistry WebBook [Online]. Available: http://webbook.nist.gov/

[37] ConservationSpace [Online]. Available: http://www.conservationspace.org/Home.html