**HUNTER — Collaborative**

<1 figure>

**A Collaborative Interdisciplinary Knowledge-Base for the Art Conservation Community**

Hunter, J., Odat, S. and Drennan, J.

Due to increasing access to sophisticated analytical techniques, painting conservation has evolved into a highly interdisciplinary research area that requires the integration of knowledge and ‘big data’ about art history (artists, artistic techniques, art provenance), the physical and chemical properties of paint and pigments, the chemical processes associated with treatment and cleaning methods, and advanced high-resolution characterization techniques. The ultimate aim is to determine the precise cause of the degradation or discoloration that is occurring in a particular artwork or across a set of paintings and the optimum methods for treating and preventing it.

The challenge is that the data and information required by art conservators to inform decision making are highly heterogeneous and distributed across databases, scholarly publications, and the Web. Expertise is also distributed across art galleries, conservation centres, and universities around the globe. Although it is possible to find some concentrated authoritative collections of information 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, and Forbes Pigment database), the information is often embedded within databases or is highly unstructured data hidden in textual documents. Relevant information may exist, but it is difficult to find, extract, re-use, interpret, correlate, or compare.

For example, it is very difficult for art conservators to find other past examples of conservation issues that most closely match the problem they are trying to solve. Moreover, the experimental data underpinning publications that describe the long-term effects of different environmental conditions (humidity, temperature, UV light) on different paints and pigments is not accessible, verifiable, or re-usable. Typically, the raw data (FTIR, spectrographs, X-ray diffraction images, SEM/TEM images) associated with the analysis of a particular painting or paint samples are not available even if described within a related publication. Our community of conservators, curators, and scientists want a Web-based search interface that provides answers to questions such as

• What is the best way to treat zinc oxides occurring in paintings by Rover Thomas?

• What are the factors that cause or accelerate the occurrence of lead soaps in paintings by R. Godfrey Rivers?

• What is the best solvent for removing varnish from the painting ‘Epiphany’?

In order to answer such questions, a broad range of provenance information, characterisation data, publications, and experimental results need to be aggregated and synthesized. Hence, a key objective of the Twentieth Century in Paint project1 was to investigate the IT infrastructure and services required to provide decision support tools to an online network of art conservators (APTCAARN—the Asia Pacific Twentieth Century Conservation Art Research Network). An initial analysis of the user requirements of APTCAARN members identified a growing demand for the following:

• An online repository where experiments and experimental results can be described, stored, shared, and discussed.

• A search interface that provides seamless access to publically available databases about paints and pigments.

• Tools to enable the extraction of structured data from publications about art conservation issues.

• A SPARQL query interface that supports querying across all of the above.

This paper will describe the architecture, ontology, repository, and services that constitute the Art Conservation Knowledge-base. It will describe the OPPRA (Ontology for the Preservation of Art) ontology, the experimental data capture interface and archive, the mapping tools (that map external databases to OPPRA), and the GATE pipeline (that integrates a Named Entity Recognition [NER] tool for tagging concepts and named entities within paint conservation publications, with a Relation Extraction classifier for identifying OPPRA-based relations between named entities) to extract structured data from a corpus of key publications on art conservation. We will also give a demo of the ontology-based federated search interface, designed to support the queries described above. Finally, the results of evaluating the system both from a user endpoint and empirically will be presented.

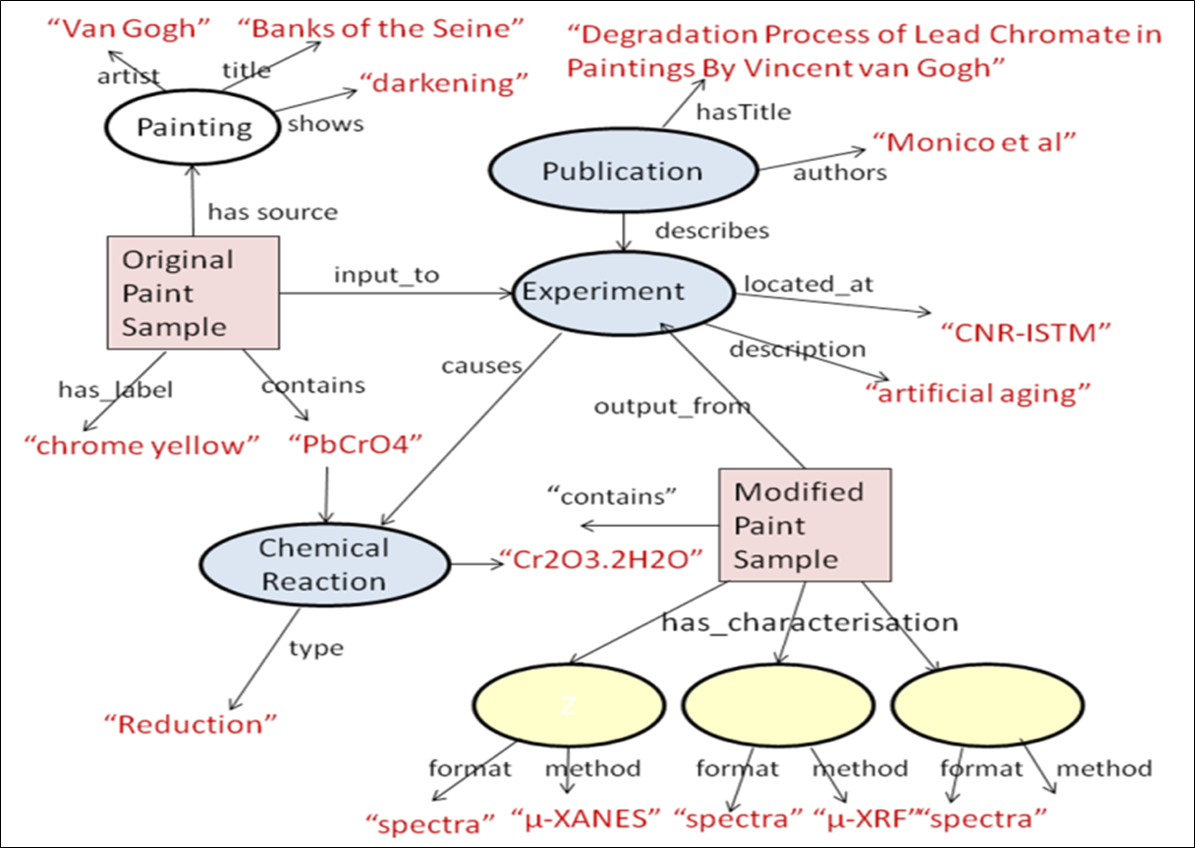
****

Figure 1. Modelling and extracting structured data from a publication about experiments on lead soaps in a painting by Van Gogh.

**Note**

1. http://www.20thcpaint.org/.

**References**

**Hunter, J. and Odat, S.** (2011). Building a Semantic Knowledge-Base for Painting Conservators. In *2011 IEEE 7th International Conference on e-Science*. IEEE, pp. 173–80.

**Odat, S., Groza, T. and Hunter, J.** (2014). Extracting Structured Data from Publications in the Art Conservation Domain. *Literary and Linguistic Computing*, doi:10.1093/llc/fqu002.