

discovery project review

OpenART

<http://yorkdl.wordpress.com/category/openart>

OpenART was funded by JISC to improve access to collections that support research and education. This document is part of a series that describes the lessons from 8 JISC projects funded under the Discovery programme in 2011 to explore open metadata for libraries, museums and archives. More information about the projects can be found at: <http://www.jisc.ac.uk/whatwedo/programmes/inf11/infrastructureforresourcediscovery.aspx>. The other documents in the series can be found at: <http://discovery.ac.uk>

Background

"The primary aim of OpenART is to expose open metadata and Linked Data about resources and collections from an exemplary research project funded by the AHRC, and to provide for and support services that will consume and visualise the data. The secondary aim is to produce analysis of the process and lessons learnt from OpenART to enable others to benefit from the project and re-use its approaches."

Institution	The University of York
Responsible group	OpenART is managed by the Digital Library team within York Library & Archives, and also draws upon expertise from the university's History of Art Department and the Tate, with technical input from Acuity Unlimited.
Capacity	OpenART is loosely associated with a larger ongoing Arts & Humanities Research Council (AHRC) project between York and the Tate; Court, Country, City ¹ . Although concerned with exposing the same data, this larger project does not currently have spare effort to support an expansion of the Linked Data exploration undertaken by OpenART.
Data Scope	OpenART is working with data relating to the sale of works of art on the London art market in the late 17th and early 18th centuries, tracking people, places and sources related to each work and its sale.
Data Scale	Initially concerned with 2 or 3 sales events, each comprising several hundred works. The underlying resource includes several thousand sales events, and draws upon 20,000 source documents.

Mechanics

Formats

OpenART data has been extracted from a pre-existing Excel spreadsheet and is stored within York University's Fedora repository as RDF/XML. RDFa may also be embedded within individual pages on the AHRC project website when it launches.

A rich OWL ontology² describes the relationship between people, places and sales events within the data, and development of this ontology required significant effort during the early stages of the project.

Technologies

York University has an existing commitment to Fedora³, which powers the York Digital Library⁴. Some time was spent considering other options, including a database-driven website and direct use of Fedora's underlying RDF triple store, Mulgara⁵. In the end, familiarity with and availability of Fedora resulted in this being used for the project.

¹ <http://www.york.ac.uk/history-of-art/court-country-city>

² <http://yorkdl.wordpress.com/2011/07/26/getting-to-grips-with-the-openart-ontology>

³ <http://fedora-commons.org>

⁴ <http://www.york.ac.uk/library/electroniclibrary/yorkdigitalibraryyodl>

⁵ <http://www.mulgara.org>

Enhancement

OpenART makes use of the Getty's Union List of Artist Names (ULAN) to identify and describe artists. As ULAN is not available in linked form online, OpenART manually assigns more easily accessible VIAF⁷ identifiers and inserts a sameAs assertion⁸ within each record to refer to ULAN data.

Reusability

Data is freely available for download as RDF/XML from the York Digital Library.

Impact

Licensing

OpenART is planning⁹ to utilise Open Data Commons licensing for the project's data. A final decision has not yet been made, but it is likely that the Open Data Commons Open Database License¹⁰ will be selected, as "The OpenART dataset is the result of several years work and has involved considerable intellectual efforts. Asking its contributors to cede copyright and attribution is quite a leap." This may be a rather extreme interpretation of the differences between the available licenses, as the 'Community Norms'¹¹ in the more generous Public Domain Dedication & Licence¹² would also enable the contributors to request appropriate attribution.

Benefits

The project anticipates benefits in terms of avoiding data duplication (linking to third party sources such as ULAN, and being linked to by other researchers), but these have yet to be demonstrated.

Outcome

Aspects of OpenART live on within the larger AHRC project. There may also be opportunities to link data from the wider York Digital Library to OpenART, although it is currently unclear whether meaningful links will be found within the available collections.

Lessons Learned

- An ontology is key to meaningfully expressing the relationships between data elements, although the project possibly spent too long trying to achieve 'perfection,' rather than settling for 'good enough.'
- Key pieces of terminology infrastructure (such as the Getty thesauri) already exist, but it is far harder than it should be to link through to authoritative online versions of these resources. It is far easier to store local copies, or to create local terminologies from scratch, and this is counterproductive.

⁵ <http://www.getty.edu/research/tools/vocabularies/ulan/>

⁶ <http://viaf.org>

⁷ <http://www.w3.org/TR/owl-ref/#sameAs-def>

⁸ <http://www.w3.org/TR/owl-ref/#sameAs-def>

⁹ <http://yorkdl.wordpress.com/2011/07/15/openart-and-open-licensing>

¹⁰ <http://opendatacommons.org/licenses/odbl/summary>

¹¹ <http://opendatacommons.org/norms>

¹² <http://opendatacommons.org/licenses/pddl/summary>

See Also

- **Union List of Artist Names (ULAN)** - <http://www.getty.edu/research/tools/vocabularies/ulan>
One of several thesauri from the Getty Research Institute. ULAN provides authoritative naming and biographical data for artists
- **Virtual International Authority File (VIAF)** - <http://viaf.org/>
“A joint project of several national libraries plus selected regional and trans-national library agencies. The project’s goal is to lower the cost and increase the utility of library authority files by matching and linking widely-used authority files and making that information available on the Web.” VIAF is hosted by OCLC, and is exposed as Linked Data¹³.

¹³ <http://outgoing.typepad.com/outgoing/2010/05/viafs-new-linked-data.html>