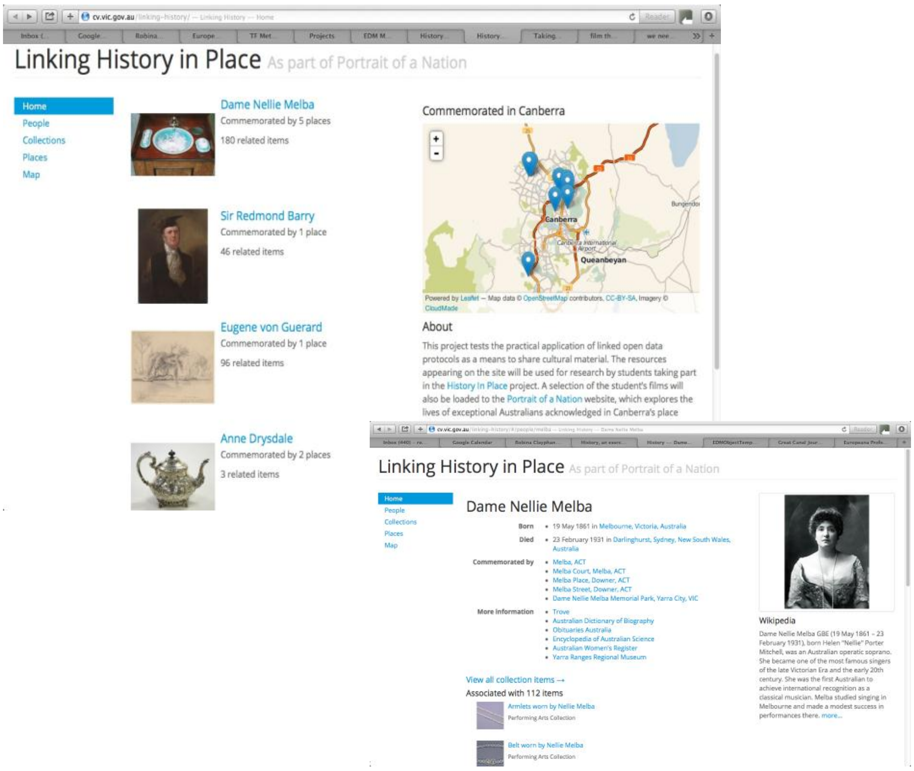
Linking History and EDM
An exercise in linked open data

The **Linking History** site has been created for use as a research tool by students participating in the **History In Place** project. The students used the site to learn more about their subjects and to identify resources for inclusion in short films they were creating about their local history and people featured in the place names of Australia's capital city, Canberra. The resulting films were uploaded to the **Culture Victoria** website and to the **Portrait of a Nation** website (which has subsequently been archived [1]).


cv.vic.gov.au/linking-history/ — Linking History — Home
Home
People
Collections
Places
Map
Canberra
Queanbeyan
Bungendore
Canberra International Airport

Linking History in Place As part of Portrait of a Nation

Dame Nellie Melba
Commemorated by 5 places
180 related items

Sir Redmond Barry
Commemorated by 1 place
46 related items

Eugene von Guerard
Commemorated by 1 place
96 related items

Anne Drysdale
Commemorated by 2 places
3 related items

Commemorated in Canberra

Powered by Leaflet — Map data © OpenStreetMap contributors, CC-BY-SA, Imagery © CloudMade

About

This project tests the practical application of linked open data protocols as a means to share cultural material. The resources appearing on the site will be used for research by students taking part in the History In Place project. A selection of the student's films will also be loaded to the Portrait of a Nation website, which explores the lives of exceptional Australians acknowledged in Canberra's place

Linking History in Place As part of Portrait of a Nation

Dame Nellie Melba

Born
- 19 May 1861 in Melbourne, Victoria, Australia

Died
- 23 February 1931 in Darlinghurst, Sydney, New South Wales, Australia

Commemorated by
- Melba, ACT
- Melba Court, Melba, ACT
- Melba Place, Downer, ACT
- Melba Street, Downer, ACT
- Dame Nellie Melba Memorial Park, Yarra City, VIC

More information
- Trove
- Australian Dictionary of Biography
- Obituaries Australia
- Encyclopedia of Australian Science
- Australian Women's Register
- Yarra Ranges Regional Museum

View all collection items →
Associated with 112 items

Armlets worn by Nellie Melba
Performing Arts Collection

Belt worn by Nellie Melba
Performing Arts Collection

Wikipedia

Dame Nellie Melba GBE (19 May 1861 – 23 February 1931), born Helen "Nellie" Porter Mitchell, was an Australian operatic soprano. She became one of the most famous singers of the late Victorian Era and the early 20th century. She was the first Australian to achieve international recognition as a classical musician. Melba studied singing in Melbourne and made a modest success in performances there. more...

Linking History is an experimental pilot in the practical application of linked open data with the objectives of increasing community engagement with Australia's history and increasing access to archival material. The results are exposed through a web application and as Linked Data.

The project arose out of the collaborative environment and structures of the **Victorian Cultural Network**. It was funded by the **Centenary of Canberra**. The site was built by **Tim Sherratt**, and the following text describing the process is summarised from the blog written by him

Design considerations

There were three elements to be considered in the design of the application: the mechanism for storing and publishing the RDF data, the code to query and retrieve details from the RDF storage, and the way in which these details were presented to users. In addition, the RDF had to be discoverable according to one of the standard

[1] http://pandora.nla.gov.au/pan/134426/20120619-1830/portraitofanation.com.au/index.html

LOD publishing patterns. The decision was to store and deliver static RDF/XML files, and display the details using javascript and html. One advantage of this approach being that provision of an example of how collections, exposed as Linked Open Data, might be integrated into new forms of online publication that are not dependent on particular platforms.

Data structure – Europeana Data Model

Since this project concerns cultural heritage institutions it seemed logical to base the data structure on the model used in Europeana. This model, defined in the Europeana Data Model (EDM)[2], is also one of the foundations used by the Digital Public Library of America.[3]

At the centre of EDM is the idea of an 'aggregation' which brings the cultural heritage object together with any digital representations of that object. Linking History follows the same approach by representing each object with an aggregation that bundles together links to:

- the object itself
- a web page that provides information about that object
- an image of the object if available

A typical object is therefore represented as four interlinked entities: aggregation, object, web page and image. The aggregation is also linked to details of the institution providing the data.

As with EDM, most of the descriptive metadata – such as title, description and date – is attached to the object entity. In addition are a small set of properties to assert the relationships between the object and people or places: these are subject, creator, and edm:hasMet (an EDM property that can be used to connect things, in this case a person and a thing).

Following the model of the EDM contextual classes, People and Places are described as entities in their own right, with their own properties and relationships. People are related to Places through birth or death events, and via the 'named for' property that indicates when a place was named after a particular person. The set of properties and classes used are listed at the end of this paper.

Data processing

The data came from many sources and varied in structure and format. The Open Refine tool[4] (OR) (formerly Google Refine) was used to clean and normalise the data and generate the RDF. It enabled the many different source formats to be reduced to a manageable subset for the interface without a lot of manual intervention. It could also be used to create relationships between objects and people and it supported transformations such as modifying punctuation or combining values into one field.

The more advanced features allow the retrieval and processing of related data. For example, some records did not include direct URLs for images so OR could retrieve and save the html of the web page for each item. Further instructions could then be used to find the image links and save them to a new field by screen scraping.

Enrichment and linking to other data sets

Open Refine's reconciliation services allow you to find links with other data sets. After normalising the artist field from the National Gallery of Victoria data, values were

[2] http://pro.europeana.eu/edm-documentation
[3] http://dp.la
[4] http://openrefine.org

reconciled against DBPedia. The matches were added as additional people to the data set and they could then be related as creators to the original items.[5]

Values can be sent to any number of third party APIs for further processing. Using named entity recognition on the titles from the National Gallery, details of place names were extracted. These were then used against DBPedia and GeoNames to harvest useful metadata, such as coordinates. These places were then related back to the original items as subjects

There was considerable inconsistency from such experiments and quite a lot of manual intervention was still required. The value lies in the way such techniques can create connections to widely-used datasets such as Wikipedia/DBPedia – this is what really puts the 'Linked' into Linked Open Data and opens up new opportunities for discovery.

The five people who were the subjects of the project, and the Canberra places that are named after them, were subject to a considerable amount of manual enrichment. Birth and death details were added, as was information about associated places. Links were created to DBPedia and a range of other biographical sources. In the case of the Canberra places, links were added to the Portrai of a Nation site and to the Government's place name database.

RDF Generation

The RDF extension for Open Refine made it straightforward to design, manage and export Linked Data for consumption by other applications. For each data set an RDF skeleton was defined that set out the basic EDM entities – aggregation, object, web page and image – and mapped values from the data to properties attached to each of the entities. Once the skeleton was defined, the data could be exported in RDF/XML format and added to the interface. The properties are listed below.

As part of the mapping, values can be transformed using Open Refine processing instructions. For example, identifiers for the aggregations and the cultural heritage objects were created by combining the project's namespace, the contributing institution's name and a unique system id.

Interface construction

A web application was constructed giving a series of simple browse lists and an interactive map to provide users with the ability to explore the collections, people and places. A fuller explanation of the interface technology and tools is provided on the blog referred to earlier. This case study focuses on the metadata aspects.

Possibilities and problems

In general, with a clear target model and a good understanding of the source data, the processes of normalisation, mapping and RDF generation were fairly straightforward and the Open Refine application almost made the work enjoyable!

Where there were difficulties they mostly related to the unavailability of data, such as direct image urls. Some web interfaces embedded images within frames or javascript widgets that made it difficult to dig them out automatically.

We are all familiar with the arguments for publishing our web resources using persistent urls. This project demonstrated that it is equally important to manage the

[5] An interesting use of OpenRefine for matching names in the Archive sector can be found at http://archiveshub.ac.uk/blog/2013/08/hub-viaf-namematching/

urls of assets, such as images, that make up such resources. Given an item identifier, it should be possible to retrieve an image of it in a variety of sizes.

The experiments with named entity extraction were encouraging. Another useful feature of Open Refine is that a series of processing steps can be saved and imported into another project. This means that it might be possible to build and share a series of formulas that could be used for enrichment across data sets such as these.

This aim of this project was to aggregate and link resources to promote discovery. Aggregation of resources generally comes at the cost of a simplified data model – the richness of individual descriptive models can be lost. The purposes of aggregation, its costs and its benefits, need to be considered.

Key data considerations for cultural institutions

- Whatever system is used to manage your data there must be a way to get the data out. This seems obvious, but some of the contributors to this project had difficulty exporting to simple formats for data exchange.
- Tools like Open Refine facilitate data clean-up and normalisation, so concerns about data quality or differing vocabularies should not inhibit sharing in projects such as this.
- Good examples of Linked Data based models for resource aggregation already exist in Europeana and the DPLA. It’s worth thinking about how your data might map to such structures.
- The importance of unique identifiers and persistent URLs can’t be stressed enough. Some contributors had trouble supplying these because of limitations in their software. Linked Data needs reliable links to individual resources.

Acknowledgement:
Thanks to Eleanor Whitworth, Linking History Project Manager
eleanor.whitworth@dpc.vic.gov.au