Towards “Linked History”

Richard Light, Independent Consultant

Keywords: Linked Data, cultural heritage information, CIDOC CRM, history

Abstract: Traditionally, museum documentation has tried to meet two competing agendas. On the one hand, it aims to support the care and management of museum objects. On the other, it tries to record the social, technical and historical significance of those objects. In this paper I argue that the collections management agenda, supported by standards frameworks such as SPECTRUM, has dominated proceedings to the point where a radical re-think is required, if museums are to make a meaningful contribution to our collective historical understanding.

Much of a typical museum record consists of management information. Information which relates to the object’s historical context will typically only include production information (who made or created the object, where, and when) and possibly some ownership history. Associative links between an object and historical figures may sometimes be recorded.

Most modern museum databases support the recording of people, places and dates as separate data items. These will often be authority-controlled. The significance of objects will typically only be recorded as a free text description (if it is recorded at all).

In order to share more widely the historical information which museums record, it needs to be expressed in a neutral format. The Linked Data approach currently offers the most practical way of achieving this. The CIDOC CRM is coming into its own as a way of expressing historical information as a set of events. Two more developments are required: shared frameworks of URLs for common information (like people and places) need to be deployed, and museum data needs to be expressed in terms of those shared frameworks. I will describe progress on the development of relevant frameworks (e.g. the Getty vocabularies) and practical techniques for URL-ifying museum data.

It is not clear that existing collections management systems offer a suitable environment for the publication of historical information. Dedicated Linked Data stores, such as the ResearchSpace database, may be a more effective means of pooling historical information from a wide variety of sources.

Paper

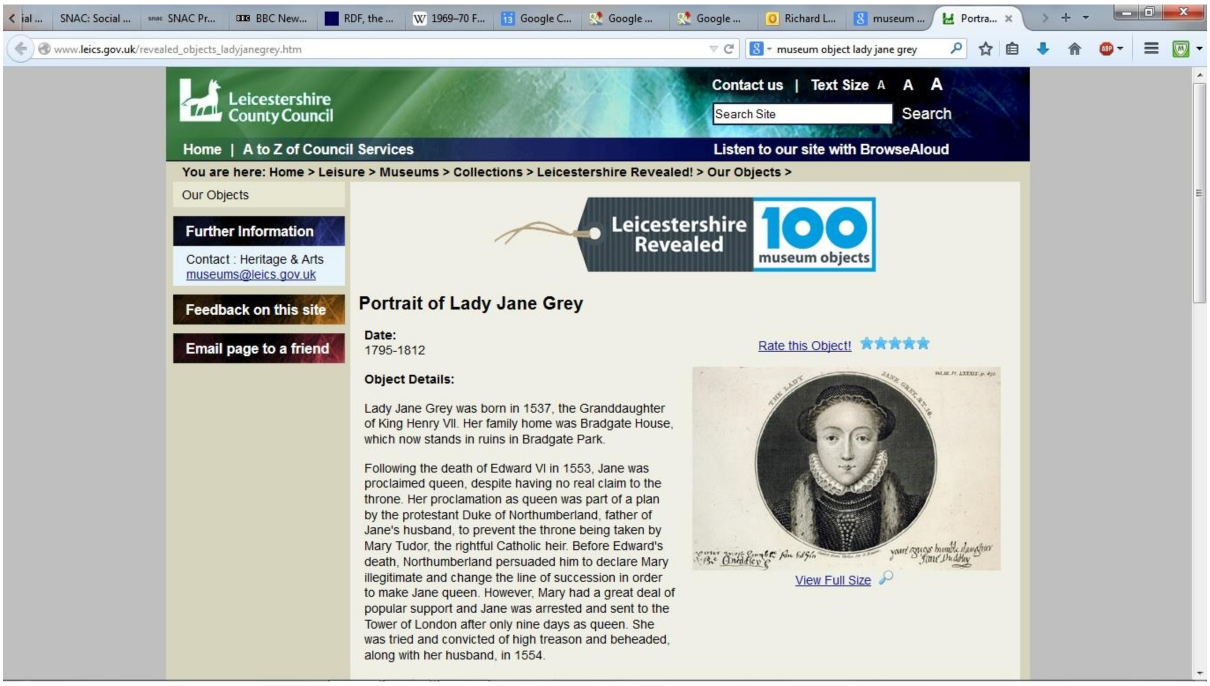
My subject this afternoon is “Museums and History”. I shall be exploring the extent to which museums currently contribute to the study of history, and suggesting how this might be improved.

What do I mean by “history” in the context of museum documentation? Essentially, it is every fact relating to an object in the collection which is public knowledge, i.e. anything which is not internal management information. Thus, the administrative detail of the museum’s acquisition of an object would not count, but its broader ownership history (including a publicly-viewable summary of the acquisition) would be in scope.

Much of the effort that goes into documenting museum collections relates to their management, and so doesn't count as useful for our study of history. What, typically, is left? Essentially, information about events and activities in the "life" of the object before it arrived in the museum: its production, its use, transfers of ownership.

I am not an historian – I expect that few (if any) of us in this room can claim that title. Yet I am, and you are, spending our working lives contributing to the knowledge of humanity's past. We do this by recording (or enabling others to record) assertions about the objects in our museum collections.

So how much history do these records contain? I ran a web search for "museum object Lady Jane Grey", and on the first page it turned up three museum objects: from the British Museum, Leicestershire and York Museums Trust. The results are interestingly varied. The BM and YMT pages give information about the object itself, in a structured form. In the case of the YMT record, this is so focused on the object that no historical context can be inferred. The BM record has links to producers, and a production date. However, this date (1832) merely serves to confirm that this image of Lady Jane Grey is a work of the imagination, created nearly three hundred years after the event it purports to depict. So, you might argue, there is not much historical value there, in terms of Lady Jane Grey herself. By contrast, the Leicestershire record contains (under the heading "Object details") a potted summary of Lady Jane Grey's short life and untimely death. To the historical enquirer, this is interesting information, but it is arguably in the wrong place. It is also presented as free text, so its value as a research information resource is limited.


ial ...
SNAC: Social ...
SNAC Pr...
BBC New...
RDF, the ...
1969–70 F...
Google C...
Google ...
Google ...
Richard L...
museum ...
Portra...
www.leics.gov.uk/revealed_objects_ladyjanegrey.htm
museum object lady jane grey
Leicestershire County Council
Contact us | Text Size A A A
Search Site
Search
Home | A to Z of Council Services
Listen to our site with BrowseAloud
You are here: Home > Leisure > Museums > Collections > Leicestershire Revealed! > Our Objects >
Our Objects
Further Information
Contact : Heritage & Arts
museums@leics.gov.uk
Feedback on this site
Email page to a friend
Leicestershire Revealed
100 museum objects
Rate this Object!
THE LADY JANE GREY, ÆT.16.
View Full Size

Portrait of Lady Jane Grey

Date:
1795-1812

Object Details:

Lady Jane Grey was born in 1537, the Granddaughter of King Henry VII. Her family home was Bradgate House, which now stands in ruins in Bradgate Park.

Following the death of Edward VI in 1553, Jane was proclaimed queen, despite having no real claim to the throne. Her proclamation as queen was part of a plan by the protestant Duke of Northumberland, father of Jane's husband, to prevent the throne being taken by Mary Tudor, the rightful Catholic heir. Before Edward's death, Northumberland persuaded him to declare Mary illegitimate and change the line of succession in order to make Jane queen. However, Mary had a great deal of popular support and Jane was arrested and sent to the Tower of London after only nine days as queen. She was tried and convicted of high treason and beheaded, along with her husband, in 1554.

Let's take the first of those objections: that this historical summary is in the wrong place. What would the "right place" be? You might imagine that there is a central authority file for all, or indeed for all well-known, historical people, which museums can just refer to. Indeed, there should be, but I haven't managed to find one that fits this description. In the U.K., the Dictionary of National Biography (DNB) would serve the purpose to some extent. If the person happens to be an artist, there is ULAN. If they are an author (or someone who is mentioned as the subject of biographies, etc.) then VIAF may be helpful. There is no single resource which has the ambition to potentially include *any* historical (i.e. dead) person who ever lived. The SNAC project is currently attempting to pull together biographical descriptions from a range of archival and bibliographic sources. Maybe that is an initiative which museums can contribute to, or at least emulate.

The second objection to the Leicestershire record is that the data is free text, and so is not in a helpful format. It's great for reading, of course, but one can only read the description if one can find the authority record, and if one is confident that it refers to the actual person you are interested in. In order to be findable, person authority records need to contain structured metadata. In order to establish identity (i.e. to be sure that you're talking about the right person), that metadata needs to be unambiguous and sufficiently detailed. The less ”well-known” the person, the more it is the case that detailed metadata is required in order to disambiguate them from other people, for example from other people with the same name.

Some requirements for shared authority files are that they should be accessible, be freely re-usable, and should provide unique, persistent identifiers for each entity they describe. Thus the BND has a URL for each article, e.g. http://www.oxforddnb.com/view/article/8154 is Lady Jane Grey, but this is not guaranteed to be persistent, and is not freely accessible. The FreeBMD site in the U.K. is a database of births, deaths and marriages from 1837 to the second half of the twentieth century. It has been created by volunteers for the public good, by transcribing publicly available quarterly summaries from registry offices. Yet the resulting site cannot be used as an authority file, partly because it does not provide persistent identifiers for the events it describes, but mainly because users are specifically forbidden to re-use the information which the site provides. In general, the results of genealogical research, such as “single name” web sites, are published in a format which is either just unhelpful to potential re-use, or which actively seeks to prevent it. [see if you can find the CSS-driven “table” on a single-name site]

So, why should museums create authority files for people, if objects are what they are really about? Well, the fact is that they already do, because people are an essential aspect of the historical context into which museum objects need to be placed. However, they do so in isolation even from other museums: take for example the BM person authorities.

And why should museums create *shared* person authority files, when many other agencies are creating information about people? It's a fair question, but one could equally ask “why not”? The museum community would do genealogical researchers and historians a big favour by providing a single comprehensive, open framework for recording historical people. Much potentially useful genealogical data is held behind paywalls, in a form which is helpful to individual family researchers but is not designed to support large-scale querying for historical research. It is certainly not available to museums, either to support person-related research or to enhance the results found. None of the external (non-museum) sources I have found shows any interest in providing unique, persistent identifiers for people.

People are just one axis along which historical enquiry proceeds. As well as “who”, there is also “where” and “when”. In the case of “where”, there is at least one central resource which museums could take advantage of. This is Geonames, a wiki-based framework for recording geographical features. Geonames provides persistent identifiers for each place it describes. If you dereference this identifier, you can get a machine-readable description of the place, which includes useful information such as its coordinates. It is an open system which can be used and re-used freely.

However, Geonames is designed for recording contemporary places, and so additional frameworks such as Pleaides will be needed for recording archaic places.

So, if we take Geonames as a practical example of the type of resource which we would like to have, which allows us to share historical information, the next question is: how do we use it? Our existing records will contain place information, either recorded directly (e.g. as keywords) or using a local place authority file. How do we get the corresponding Geonames identifier for each place recorded? One approach is to set up software which supports "web termlists", so that the live Geonames resource can be searched by a cataloguer for entries which match a particular place name. Selecting one Geonames entry causes its persistent identifier to be inserted into the museum data, alongside the original description of the place. While this might be seen as just additional work, it does bring the benefit that all your records will be geolocated.

If we had our hoped-for person authority, inserting its identifiers into our data would have analogous benefits. The person's date and place of birth and death, their relationship to other people, and their key life events, would all be available for us to use as we see fit.

We can treat dates (a key element of historical enquiry) as simply numbers, or again we can use a system of identifiers, representing say years or decades, such as the Data.Gov identifiers for time intervals: e.g. http://reference.data.gov.uk/doc/year/1677.

So now we have re-expressed our historical data in terms of widely-used identifiers, in what form do we share it with others? I would look no further than our own CIDOC Conceptual Reference Model (the "CIDOC CRM"). This provides a generalized way of describing events and activities, and it provides a set of persistent identifiers with which to do so. While the CIDOC CRM is an abstract model which can be expressed in a number of ways, the most common approach these days is to represent it as Linked Data RDF. Here's an example:

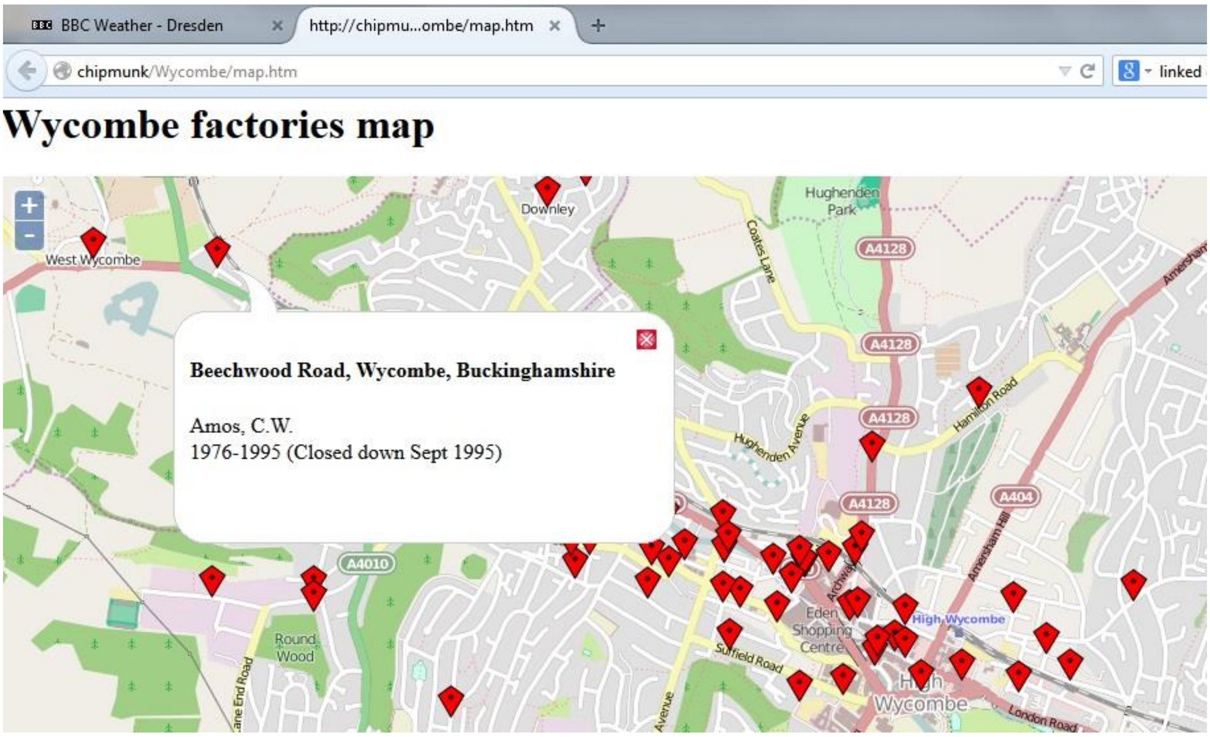
```
<crm:E22_Man-Made_Object
rdf:about="http://collection.britishmuseum.org/id/object/PPA20
6074">
<crm:P45_consists_of
rdf:resource="http://collection.britishmuseum.org/id/thesauri/
x11409"/>
<crm:P62_depicts
rdf:resource="http://collection.britishmuseum.org/id/person-
institution/29806"/>
<rdfs:label>Lady Jane Grey</rdfs:label>
<bmo:PX_has_main_representation
rdf:resource="http://www.britishmuseum.org/collectionimages/AN
01033/AN01033272_001_l.jpg"/>
```

Each object in your collection can be given its own persistent identifier, and this can be associated with an RDF expression of all the information you wish to share with the rest of the world, including the historical data we began by discussing. These identifiers should be URLs, so that they can be resolved on the Web. A helpful pattern to adopt is to present the information as HTML by default if the object's identifier is requested, and to return an RDF version of the information if this is specifically requested. That allows people to easily find out about the object if that is all they want to do, while also providing a machine-processible version of the data for software agents which work with Linked Data.

Where do we put this information so that it can be found by others? It is unlikely that the museum's existing collections database will offer support for storing and querying RDF. One approach is to add a "front end" to your existing system which supports Linked Data delivery. This has the advantage that it can be driven by your existing data: there is no need to set up and maintain a copy database elsewhere. It is helpful to users if some sort of search facility can be provided, to support the discovery of relevant resources. If you use a standard collections management software package, encourage the suppliers of that system to develop a Linked Data front-end, so that all users of the same software can share the benefit.

An alternative approach is to extract the data you want to share as RDF, and store it in a database specifically designed for the purpose: a "triple store" (or "quad store"). This approach has the advantage that it will provide a standard means of querying the data, using the SPARQL query language.

Something else we can do to encourage the advent of Linked History is to encourage the publishers of widely-used authorities to issue a Linked Data version, and then use their identifiers when publishing our own data as RDF. For example, the Getty Research Institute has issued both the Art and Architecture Thesaurus (AAT) and the Thesaurus of Geographic Names (TGN) as Linked Data. Conversely, the widely-used Nomenclature system has yet to be published in this format. Apart from giving you the warm glow which comes from doing the "right thing", using Linked Data identifiers in your data will give you access to additional information in the authority file, in a format you can use programmatically. For example, as noted above, Geonames offers coordinate information, which you can use to generate "pins" on maps.



Wycombe factories map

Beechwood Road, Wycombe, Buckinghamshire

Amos, C.W.
1976-1995 (Closed down Sept 1995)

It’s reasonable to ask if we can afford *not* to work towards Linked History. For example, this year many hundreds of separate projects are busy gathering data to commemorate the start of World War 1. However, there is no framework or mechanism of which I am aware which will enable the results of all this work to be pooled, queried or archived as a whole. I suspect that, in ten years’ time, there will be little or no evidence that all these projects ever took place.