

Building an API is not enough! Investigating Reuse of Cultural Heritage Data

 blogs.lse.ac.uk/impactofsocialsciences/2014/09/08/investigating-reuse-of-cultural-heritage-data-europeana/

9/8/2014

*The Europeana cultural heritage archive has a wealth of digital content that can be used for a variety of purposes, both by researchers and practitioners in the community. **Vicky Garnett** and **Jennifer Edmond** chart the progression of research into how this content is being used and accessed and what technical requirements would improve the digital archive's development. For example, is an API the answer? How big a part do web-services actually play in their overall research? One of the most common problems participants have reported encountering is the quality of the metadata in the content they are accessing. If the metadata can't be relied on, neither can the results.*

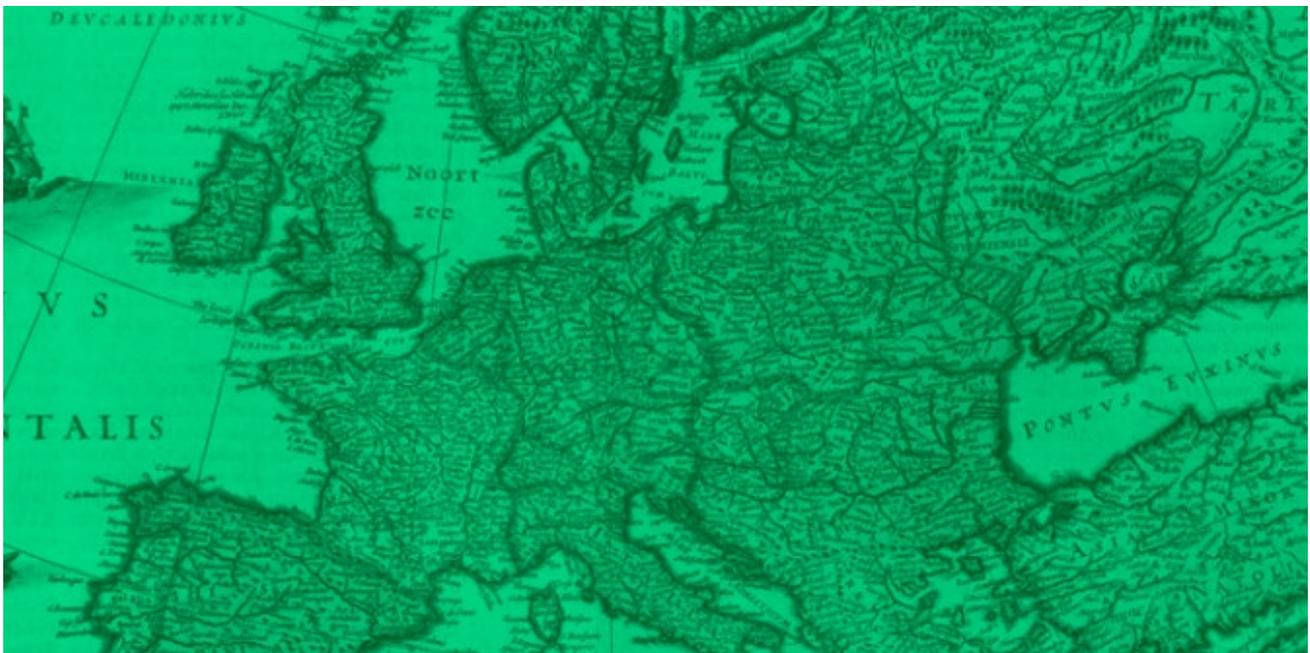


What is Europeana?

Europeana is a single access point to millions of articles, photos, census records, sound files, manuscripts and anything else that can be digitised from cultural heritage institutes and archives across Europe. The items that can be access from Europeana can be used to reveal much about the social, economic and cultural history of nations within Europe, and the data available through the Europeana Portal can be used by both researchers and the general public alike.



The **Europeana Cloud** project is aiming to enhance the data and access for researchers to do just that. This project is currently underway- having started for a 3 year duration in February 2013 – and is looking extensively at how users of cultural data work, and what they need (and indeed want) to make it easier for them to access and use the content. The end-result of this exploration into user requirements will be Europeana Research, a platform that will enable users from all different backgrounds and motivations to get hold of content and metadata, and make use of what they find, be it for social history research looking at impacts of war on a small community, economic research mapping population movement across Europe, or software developers looking to develop a new app.



Web-services in qualitative research

One of the innovations of Europeana Research will be to offer enhanced access to the Europeana collections via customised web-services such as Application Programming Interfaces (APIs), or 'pipelines'. APIs are in use in the background of many web-services for all sorts of tasks, and many websites now provide access to one of their APIs to allow users to access certain items from their data collections. This could be through a simple one-off search function to extract items from a collection without having to take the whole lot, or it might be to make a constant and continual call to extract items that are created daily (for example, making use of Twitter or Instagram APIs to get tweets or photos with a particular hashtag delivered to a website for display).

APIs, however, are not the only way of accessing this information, as RSS feeds, pipelines, data downloads and other web-services are also available. In fact many users of such services are claiming that an API is often not necessary, but it seems to be the 'accessory' of the moment for archives and collections that might not actually be the right tool for the job.

This is what our current research aims to find out. We have begun investigating the different routes and means that researchers might use to get their hands on big data collections for research purposes, and what they are doing with it once they have it. Is an API the answer? How big a part do web-services actually play in their overall research, and what barriers have they encountered to accessing that information?

To find this out we first conducted some research to get a general feel for the current state of digital research techniques within the humanities and social sciences. We then identified people working in various fields that might make use of a web service or API to gather large sets of data from digital collections.

Data Creation, Data Access and Data Reuse

One of the most common problems our participants have reported encountering is the quality of the metadata in the content they are accessing. Metadata gives basic information about an item, such as who created it, what date it was created, what date it might have been digitised, what country it originates from, etc. This is all very important if someone is looking into statistics for, say, artists active in the Netherlands in the 19th Century, or fluctuations in participation in census recordings in Germany throughout the 20th century (perhaps). Problems are encountered, however, if metadata fields are left blank, or not filled in with the correct format (or worse, inconsistent formatting). Trusting that the information is even accurate is another problem. If the metadata can't be relied on, neither can the results.

But a more fundamental problem encountered when trying to access and obtain items via a web-service like an API is that it relies on the metadata being correct in order to include, or more importantly reject an item from its output. Items that could be vital to answer a research question could be missed purely because a date was missing, or a title was misspelled. The metadata is therefore of great importance, but the researchers' ability to access the metadata in the first place is also important. We have found that many researchers don't necessarily have the know-how or even confidence to try to use a web-service to access data, and moreover don't know what to do with it once they've found it.

This isn't simply a matter of technical know-how. Technical know-how is of course important for working with digital tools, but enthusiasm and an understanding of the potential gain in doing so is equally vital.

Revealing new questions in the Social Sciences

These are lessons that are all being taken on-board by the Europeana Cloud project, and we are working on ways to ensure that not only is the metadata useable, but is also easily accessible to the 'techy' and the 'non-techy' researchers. But what does this mean for social scientists?

We've already mentioned the abundance of data available through Europeana, most of which is of huge interest to

social historians and social scientists. The census data from different countries, images such as maps, and photographs of daily life throughout the 19th and 20th Century across Europe are undoubtedly going to be useful. Access to that data could open up potential new questions for social scientists, and reveal new facets to economic history that might not necessarily have been investigated previously.

Europeana Cloud has still got some way to go in ensuring that Europeana Research will enable access to the best quality metadata – indeed trying to combine multiple metadata standards from multiple institutions raises its own problems which can't be tackled easily, and certainly won't be fixed over-night. Further user-research is underway to identify workflows, which we hope to feed into the final product.

Participants needed...

In the meantime, however, the work into web-services continues. Our interview stage is nearly complete, but we are hosting a workshop, currently planned for early November in The Hague, for which we will be seeking social scientists who already make use of web services to access large sets of digital data in their research. If you are interested in taking part, please contact either Vicky Garnett (garnetv@tcd.ie) or Dr. Jennifer Edmond (edmondj@tcd.ie).

Note: This article gives the views of the author, and not the position of the Impact of Social Science blog, nor of the London School of Economics. Please review our [Comments Policy](#) if you have any concerns on posting a comment below.

About the Authors

Dr Jennifer Edmond is Coordinator of or Principle Investigator on several European projects, including Europeana Cloud, and is based in the Trinity Long Room Hub. She has broad experience as a technology implementation adviser for the arts and humanities, and has led the development and implementation of the strategy for Digital Humanities at Trinity College Dublin. She manages a number of large scale funding grants as part of her role as Director of Strategic Projects for the Faculty of Arts, Humanities and Social Sciences in Trinity College Dublin.

Vicky Garnett is a Researcher on the Europeana Cloud project. She has worked in Digital Humanities at the Trinity Long Room Hub since 2012. Prior to this, her work mainly centred around Higher Education policy and development. Her academic background lies in Linguistics, and she is also currently undertaking a PhD in Sociolinguistics looking at population movement and language change.

- Copyright © The Author (or The Authors) - Unless otherwise stated, this work is licensed under a Creative Commons Attribution Unported 3.0 License.