

Extended Abstract: Opening up Digital Publications - suggestions for making Humanities data available for research

Henriette Roued-Cunliffe^a

^a*Bavarian Academy of Sciences and Humanities / University of Munich*

Introduction

Currently the trend in digital publishing of Humanities data seems to be moving towards openness and interoperability [1]. In this abstract I will examine to what extent and in what way current digital publications are open and accessible. My hypothesis is that while many digital publications are currently made available online and can be searched and viewed by the general public, very few are available to researchers in a meaningful way. By meaningful I mean that external researchers can search and export data for reuse and are possibly even encouraged to define their own search criteria. I believe that this is the true essence of data sharing [2].

Following this, I will propose one approach, using XML and Web Services, to creating a digital publication of Humanities data that would be open to the research community in a meaningful way, as defined above.

1. Background

In 2007 I finished my MSc dissertation titled: "Heritage Portals and Cross-Border Data Interoperability" [2]. The research for this dissertation examined the possibilities of making a heritage portal that could enable a cross-border search of heritage datasets in Europe. However, while conducting this research I came to the conclusion that while enabling a simple search across different heritage data repositories is great, there should be more focus on how to enable external researchers to define new search criteria and re-use results in further research.

One example that I am quite fond of at the moment is the Portable Antiquities Scheme¹, which encourages external researchers to sign up with them and once registered allows exports of different searches as XML and KML formats for further research. Nevertheless, I do not think that this is even going far enough in terms of enabling re-use of data. The data that is exported here is static once it is downloaded onto the recipient's computer. In addition, researchers should be able to use a dynamic

¹ <http://finds.org.uk>

web service, which will update the search results with all new data that is constantly input into the published dataset.

The idea of exposing Humanities data as a web service is not new [3]. Many heritage agencies are developing or re-developing their online data publications as web services in order to share data with each other and create cross-searchable platforms such as ARENA² and Heritage Gateway³. Unfortunately, very few have yet opened up these web services to the public or even to selected/registered researchers. The Danish Agency for Culture⁴ is one of these that I know of, who has a REST web service for their Sites and Monuments database, which is open to selected researchers from Danish Museums [4]. However, they do not have any method for external researchers to register and gain access to the service.

2. Suggestions

For new developments or re-developments of online Humanities data publication I suggest that the following is taken into consideration:

1. That users are divided into the following categories: general public, external researchers and internal researchers.
2. That, for the general public, data is made available from searches in different downloadable formats (at the very least CSV).
3. That, for external researchers, there is a registration scheme, which gives access to data from searches through dynamic web services and possibly also gives the opportunity to suggest new search criteria.
4. That, internal researchers have access to the above and the option of contributing new data to the publication.

As a part of my D.Phil thesis [5] I developed a REST web service, named APPELLO, which enables dynamic searches of TEI formatted XML datasets (such as the online publication, Vindolanda Tablets Online II). The web services is used first and foremost to run the VTO2 website⁵. However, it can also be used by external researchers to re-use for example the list of words found in these ancient documents as a look-up for their own research.

However, VTO2 does not differentiate between external researchers and the general public and this is one of its largest failings. Larger resources would definitely have to take issues such as security and overload of the server into consideration and therefore, I think that the concept of a registration scheme for external researchers, such as the one for PAS, fits this purpose.

The above considerations are ones that I will attempt to implement in future digital publications of Humanities data such as the Virtual Research Environment I am

² <http://ads.ahds.ac.uk/arena/>

³ <http://www.heritagegateway.org.uk/>

⁴ <http://www.kulturstyrelsen.dk/>

⁵ <http://vto2.classics.ox.ac.uk>

developing for the Buddhist Manuscripts of Gandhara project.⁶ I hope to see them implemented in many other projects too.

References

- [1] J. Unsworth et al., Our Cultural Commonwealth: The Report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences. *The ACLS Report*, 2006. Accessed 17.01.2013 at http://www.acls.org/uploadedFiles/Publications/Programs/Our_Cultural_Commonwealth.pdf
- [2] H. Roued Olsen, *Heritage Portals and Cross-border Data Interoperability*. MSc in Archaeological Computing, University of Southampton, 2007.
- [3] S. Waller, Future Connections: The potential of Web service and Portal technologies for the historic environment. *Internet Archaeology*, issue 18, 2005. Accessed 17.01.2013 at <http://intarch.ac.uk/journal/issue18/8/toc.html>
- [4] Kulturstyrelsen, *Webservices*. Accessed 17.01.2013 at <http://www.kulturstyrelsen.dk/kulturarv/databaser/webservices/>
- [5] H. Roued-Cunliffe, Towards an Interpretation Support System for Reading Ancient Documents. *Literary and Linguistics Computing*, 25 (4), 365-379, 2010.

⁶ <http://www.en.gandhara.indologie.uni-muenchen.de/index.html> and <http://gandhara.org>