

implementation of Digital Humanities degrees, we do not intend to rely primarily on a call for papers.

We will not rely on a call for papers primarily. During April 2012 a set of documents will be sent to all institutions in Europe, and many beyond, which are known to organize a degree course in the Digital Humanities or a closely connected field, with an invitation to join the workshop. We hope for results leading to direct and *practical* cooperation within existing frameworks. So the *primary* target group of this workshop are the European academic institutions offering or planning degree courses in the Digital Humanities. This said, of course we also invite the wider community to join the conceptual discussions.

Participation of institutions we are not aware of, particularly from those which are currently only in the planning stages of Digital Humanities' degree courses, is very much hoped for. Please direct enquiries to manfred.thaller@uni-koeln.de to receive additional material from the preparatory round of discussions and supporting material before the start of the workshop.

The workshop will run for a full day. The following program is tentative, to be adapted to accommodate for proposals and explicit wishes from participants during the preparatory stage.

09:00 – 10:30 Setting the agenda – reports on existing degree courses.

11:00 – 12:30 What do we have in common? I: Parallels and differences in the scope of individual degree courses.

14:00 – 15:30 What do we have in common? II: Parallels and differences in the concept of 'Digital Humanities' underlying the individual courses.

16:00 – 17:30 Are there synergies? Creating a work program to facilitate discussions of exchange facilities and curricular coordination across national boundaries.

Free your metadata: a practical approach towards metadata cleaning and vocabulary reconciliation

van Hooland, Seth

svhoolan@ulb.ac.be

Université Libre de Bruxelles, Belgium

Verborgh, Ruben

ruben.verborgh@ugent.be

Ghent University, Belgium

De Wilde, Max

madewild@ulb.ac.be

Université Libre de Bruxelles, Belgium

1. Tutorial content and its relevance to the DH community

The early-to-mid 2000s economic downturn in the US and Europe forced Digital Humanities projects to adopt a more pragmatic stance towards metadata creation and to deliver short-term results towards grant providers. It is precisely in this context that the concept of Linked and Open Data (LOD) has gained momentum. In this tutorial, we want to focus on metadata cleaning and reconciliation, two elementary steps to bring cultural heritage collections into the Linked Data cloud. After an initial cleaning process, involving for example the detection of duplicates and the unifying of encoding formats, metadata are reconciled by mapping a domain specific and/or local vocabulary to another (more commonly used) vocabulary that is already a part of the Semantic Web. We believe that the integration of heterogeneous collections can be managed by using subject vocabularies for cross linking between collections, since major classifications and thesauri (e.g. LCSH, DDC, RAMEAU, etc.) have been made available following Linked Data Principles.

Re-using these established terms for indexing cultural heritage resources represents a big potential of Linked Data for Digital Humanities projects, but there is a common belief that the application of LOD publishing still requires expert knowledge of Semantic Web technologies. This tutorial will therefore demonstrate how Semantic Web novices can start experimenting on their own with non-expert software such as Google Refine. Participants of the tutorial can bring an export (or a subset) of metadata from their own projects or organizations. All necessary operations to reconcile metadata with

controlled vocabularies which are already a part of the Linked Data cloud will be presented in detail, after which participants will be given time to perform these actions on their own metadata, under assistance of the tutorial organizers. Previous tutorials have mainly relied on the use of the Library of Congress Subject Headings (LCSH), but for the DH2012 conference we will test out beforehand SPARQL endpoints of controlled vocabularies in German (available for example on <http://wiss-ki.eu/authorities/gnd/>), allowing local participants to experiment with metadata in German.

This tutorial proposal is a part of the Free your Metadata research project.¹ The website offers a variety of video's, screencasts and documentation on how to use Google Refine to clean and reconcile metadata with controlled vocabularies already connected to the Linked Data cloud. The website also offers an overview of previous presentations. Google Refine currently offers one of the best possible solutions on the market to clean and reconcile metadata. The open-source character of the software makes it also an excellent choice for training and educational purposes. Both researchers and practitioners from the Digital Humanities are within cultural heritage projects inevitably confronted with issues of bad quality metadata and the interconnecting with external metadata and controlled vocabularies. This tutorial will therefore provide both practical hands-on information and an opportunity to reflect on the complex theme of metadata quality.

2. Outline of the tutorial

During this half day tutorial, the organizers will present each essential step of the metadata cleaning and reconciliation process, before focusing on a hands-on session during which each participant will be asked to work on his or her own metadata set (but default metadata sets will also be provided). The overview of the different features will approximately take 60 minutes:

- Introduction: Outline regarding the importance of metadata quality and the concrete possibilities offered by Linked Data for cultural heritage collections
- Metadata cleaning: Insight into the features of Google Refine and how to apply filters and facets to tackle metadata quality issues.
- Metadata reconciliation: Use of the RDF extension which can be installed to extend Google Refine's reconciliation capabilities. Overview of SPARQL endpoints with interesting vocabularies available for Digital Humanists, in different languages.

After a break, the participants will have the opportunity to work individually or in group on their own metadata and to experiment with the different operations showcased during the first half of the tutorial. The tutorial organizers will guide and assist the different groups during this process. Participants will be given 60 minutes for their own experimenting and during a 45 minutes wrap-up, participants will be asked to share their the outcomes of the experimentation process. This tutorial will also explicitly try to bring together Digital Humanists with similar interests in Linked Data and in this way stimulate future collaborations between institutions and projects.

3. Target audience

The target audience consists both of practitioners and researchers from the Digital Humanities field who focus on the management of cultural heritage resources.

4. Special requests/equipment needs

Participants should preferably bring their own laptop and, if possible, have installed Google Refine. Intermediate knowledge of metadata creation and management is required.

Notes

1. See the projects website on <http://freeyourmetadata.org>.