

Project Report: Building Bridges to the Future of a Distributed Network: From DiRT Categories to TaDiRAH, a Methods Taxonomy for Digital Humanities

Jody Perkins
Miami University
perkintj@miamioh.edu

Quinn Dombrowski
University of California
quinnd@berkeley.edu

Luise Borek
Technical University of Darmstadt
borek@linglit.tu-darmstadt.de

Christof Schöch
University of Würzburg
christof.schoech@uni-wuerzburg.de

Keywords: digital humanities; taxonomies; research methods; ontologies; RDF; Resource Description Framework; LOD; Linked Open Data

1. Project Background

This poster presentation traces the development and application of ‘TaDiRAH’ (Taxonomy of Digital Research Activities in the Humanities), a shared taxonomy of digital humanities research goals and methods (e.g. capture, enrichment, analysis), objects (e.g. data, images, manuscripts), and techniques (e.g. cluster analysis, encoding, topic modeling) created for the purpose of bridging the divide between related digital humanities hubs.

Earlier efforts to establish centralized hubs of information relevant to digital humanities (DH) have proven unsustainable over the long term. These comprehensive hubs (such as *arts-humanities.net*, a European initiative which previously aggregated information about events, jobs, news, projects and tools) are currently being re-designed with a smaller scope and more focused curation. However, this smaller scope comes with the risk of decontextualization—a digital humanities project is best understood through the intersection of its subject matter, methodologies and applications, not all of which are captured by any single site.

An example of a focused directory is the DiRT (Digital Research Tools) Directory, an established, well-regarded source of information about tools available to support scholarship in the humanities. DiRT is currently undergoing a new phase of development, with the goal of making information about digital tools available outside the DiRT directory itself using RDF and APIs.¹ However, the ad-hoc set of categories that have been used to organize tools on DiRT since its inception are of no utility outside DiRT itself. Adopting a shared taxonomy would provide a means to connect DiRT’s tool data with related information provided by other sites.

2. Development Process

Early in 2013, as part of an effort to improve usability of the site, members of the DiRT Steering Committee/Curatorial Board conducted an analysis of DiRT’s categories and free-form tags. Shortly thereafter we began a series of discussions with the DARIAH-DE (Digital Research Infrastructure for the Arts and Humanities-Germany) team that was developing a taxonomy for their ‘Doing Digital Humanities’ Zotero bibliography. Recognizing our common goal, we formed a transatlantic collaboration around the task of developing a shared taxonomy.

In the process of developing TaDiRAH we drew from three primary sources: 1) the *arts-humanities.net taxonomy* for DH projects, tools, centers, and other resources, especially as it has

¹ <http://dirtdirectory.org/development>

been expanded by digital.humanities@oxford in the UK and DRAPier (Digital Research and Projects in Ireland); 2) the *DiRT categories* for digital research tools, re-launched under Project Bamboo in the US but now continuing on after the end of that project; and 3) the *scheme used by the DARIAH 'Doing Digital Humanities' Zotero bibliography* to organize literature on all facets of DH. These resources were mapped, analyzed and distilled into their essential parts, producing a simplified taxonomy of two levels: eight top-level “goals” that are broadly based on the steps of the scholarly research process and a number of lower-level “methods” associated with each goal. In addition, there are two separate open ended lists of digital humanities research “objects” and “techniques” that can be freely associated with higher level methods.

In September 2013, and again in January 2014, we opened a draft version of the taxonomy for public comment and received a tremendous amount of feedback from the DH community. The response shows the ongoing relevance of a task that has been under discussion in digital humanities circles since John Unsworth introduced his concept of 'scholarly primitives' in 2000. We hope that one outcome of this presentation will be to extend the conversation beyond the boundaries of the DH community.

3. Challenges and Future Work

This presentation will also cover some of the challenges encountered during TaDiRAH's development, including: selection of terms that facilitate consistent application vs. terms that represent entities in a more precise manner², avoiding conflation of concepts, reconciling terms against existing taxonomies, minimizing redundancy, balancing theoretical “correctness” on one hand against the necessity of adopting commonly used terms to ensure findability on the other (e.g. visualization + geospatial coordinates object vs. “mapping”), and responding to thorough (and sometimes conflicting) feedback from the digital humanities community.

We will also present several use cases based on the shared taxonomy, demonstrating how it will work to serve both task and user-oriented endeavors. Applying TaDiRAH to actual directories will provide an opportunity to assess the degree to which it can accommodate real-world data. In the coming months we will conduct a comprehensive review of all DiRT tool entries, adding terms from the TaDiRAH taxonomy. DHCommons will also add TaDiRAH terms to project profiles based on existing free-form metadata. Information from DiRT and DHCommons will be exposed using RDF, making the content available as linked open data, as well as through APIs that are currently under development.

The “Doing Digital Humanities” bibliography curated by DARIAH-DE has already implemented the TaDiRAH taxonomy. The Zotero-based bibliography is using “collections” (similar to subfolders) for the seven broad goals, and the tags for the research activities, objects and techniques. Each entry is tagged with at least one activity and one object to enable a faceted browsing of the bibliography, starting with either research activities or objects. Most recently TaDiRAH has been adopted by two additional DARIAH initiatives: the Digital Humanities Course Registry and the Training Materials Collection (Schulungsmaterial-Sammlung).

DARIAH-EU has committed to using this taxonomy as a basis for their development of a more complex ontology of digital scholarly methods, and we are also engaged in ongoing dialog with other ontology initiatives, including NeDiMAH's (Network for Digital Methods in the Arts and Humanities) work around scholarly methods. Our goal is to share at least high-level categories with NeDiMAH's ontology, so that objects (projects, tools, articles, etc.) classified using our taxonomy can be automatically “mapped” to some level of the NeDiMAH ontology, and vice versa.

² While the use of specific terms supports precision, the use of more broadly defined terms tends to provide better support for consistent application, collocation and recall. In the context of search, precision and recall are often inversely related.

The projects and collections that adopt TaDiRAH will also inform its evolution. TaDiRAH can be found online at GitHub³, where we will be using the issue tracker to collect further feedback to be incorporated into future revisions. A SKOS version soon to be available on the GitHub site and a SPARQL endpoint through a TemaTres instance are currently in development. We expect that TaDiRAH will continue to evolve as a relatively flexible scheme of associated scholarly methods, techniques and object types that can be applied to a variety of DH resources.

Acknowledgements

The authors would like to recognize the work of Matt Munson, who was a member of the TaDiRAH Coordinating Committee during its initial development phase.

References

- Digital Humanities Course Registry, DARIAH-DE Cologne / CLARIAH-NL Rotterdam. (2014). Retrieved from <http://dhcoursereg.hki.uni-koeln.de/>
- DARIAH-DE (Digital Research Infrastructure for the Arts and Humanities-Germany). (2014). Retrieved from <https://de.dariah.eu/>
- DHCommons. (2014). Retrieved from <http://dhcommons.org/>
- DiRT (Digital Research Tools Directory). (2014). Retrieved from <http://dirtdirectory.org/>
- Doing Digital Humanities - A DARIAH Bibliography. (2014). Retrieved from https://www.zotero.org/groups/doing_digital_humanities_-_a_dariah_bibliography/items/order/creator/sort/asc
- NeDiMAH (Network for Digital Methods in the Arts and Humanities). (2014). Retrieved from <http://www.nedimah.eu/>
- Schulungsmaterial-Sammlung, DARIAH-DE Würzburg/Cologne (2014). Retrieved from <https://de.dariah.eu/schulungsmaterial-sammlung>
- TaDiRAH (Taxonomy for Digital Research Activities in the Humanities). (2014). Retrieved from <https://github.com/dhtaxonomy/TaDiRAH> and <http://tadirah.dariah.eu>.
- Unsworth, John. (2000). *Scholarly Primitives: What Methods Do Humanities Researchers Have in Common, and How Might Our Tools Reflect This?* London: King's College London. Retrieved May 16, 2014 from <http://people.brandeis.edu/~unsworth/Kings.5-00/primitives.html>.

³ <http://github.com/dhtaxonomy/TaDiRAH>