Making Publics 2.0: Humanities data linked through a topical repository

Our previous project, Making Publics (MaPs http://www.makingpublics.org), examined voluntary forms of publics in early modern Europe 1500-1700 [1] and resulted in the creation of a corpus of databases of people, places, and artifacts. These databases catalogue the nearly 2,000 works cited in over 200 articles, papers, books, and essays written by its members. These works contain numerous references to people, places, things, and their interactions on various topics, such as common interests, tastes, and desires in early modern Europe. MaPs allows its users to continue this research in public making through collaboration, collating new references to people, places, things that make up the forms of association around shared interests and topics. Throughout 2012, we re-envisaged how the website might operate to further research as well as present the research findings of the project which ended in 2010. Central to this was consideration of how a web application might serve to both organize and foster ongoing research, and make that research data available to other humanities users in the form of Linked Open Data.

The focus of this panel is to demonstrate our approaches that transform seemingly unorganized humanities research information into data using a topically oriented Resource Description Framework (RDF) repository. Its overall purpose has been to explore what a networked-approach to humanities data might look like using RDF principles on the one hand and to disseminate this humanistic data to academic and non-academic audiences.

The themes of this panel are the following.

Humanities data reshaped to social open data

MaPs’ existing website needed to be changed from merely documenting artifacts and authors into a collaborative research environment to manage an array of materials linked through forms of social interaction. Our approach was to examine how humanities and information science network theories might interoperate, in order to allow us to trace the ways in which objects, their producers and consumers, and sites of interaction might be collated as networks of events or relationships. This wasconcertedly different than a traditional archival or bibliographic model, as it placed emphasis on historical individuals and the ways in which they interacted in places and through cultural objects. Associations were understood not only as topics, shared desires or interests, but interactions such as shared authorship, co-habitation, legal or other forms of social interaction, even citizenship or residency. Consequently, we needed multi-disciplinary humanities data modelling (e.g. historical, prosopographical, geographical, and chronological data etc.) rather than archival metadata models. Even so, the aim of the project was to allow scholars to link existing archival metadata and humanities data through the use of unique identifiers using APIs that offer search and interoperability with resources such as OpenLibrary, DBpedia, OCLC, VIAF, Geonames, and GoogleMaps. Lastly, we attempted to extract the social networks of people, places, things, and interactions that are in the prose works of the MaPs using Natural Language Processing Named Entity Extraction in order to further expand the existing dataset using GATE (General Architecture for Text Engineering). The end product was a data...
model and online tool robust enough to facilitate ongoing humanities research while actively disseminating existing research on historical forms of association to new researchers.

**Building a Topically oriented RDF Repository**

Our repository data is derived from the common research interests of MaPs’ members and not through digitization of an existing archive or data set. It is topically construed and expansible; as new works and data emerge, they are added to the data because of their topical relevance rather than on account of other forms of data collection such as progeny or authorship. The four tasks of the project have been: 1) named entity extraction of people, places, and things from the works; 2) the complete overhaul of the research environment; 3) the creation of a basic RDF schema for the data, including a limited vocabulary for documenting interactions, and 4) the creation of the RDF repository.

RDF is a specification for linking data across websites, which lies at the heart of the emergent semantic web. RDF’s basic unit is the triple, simple subject->predicate->object statements. We found that taxonomies or vocabularies that focus on artifacts are not well suited to documenting interactions. Rather, the description of relationships between humanities data became clearer when we shifted towards quads or named graphs, which add a unique identifier to the standard RDF triple, permitting referencing and nesting of interactions and allowing interactions to be thought of as verbs. We had to map common interaction verbs for person-person relationships, person-thing, thing-thing, place-person, etc. The relationships, which we view as an expanded category of “events,” are shaped by a schema or vocabulary that explicates them for our particular knowledge domain (in the linked data world, RDF data and its vocabularies are shared as web resources, allowing anyone to access dynamic information. A query language, SPARQL, is used to better answer the questions our users may ask.) The vocabulary and repository are encoded in Sesame from the web application’s MySQL database.

**Contribution to the Fields of Digital Humanities and Information Science**

The current MaPs website turned its data into an RDF repository, which can describe historical social networks, places, chronologies, and artifacts. Our repository allows us to offer data in RDF to anyone, can be inverted to harvest data from other RDF sources, turning the research environment into a tool that mediates between existing data and the production of new humanistic data. This dynamic expands the collaborative nature of MaPs, offers them new ways of gathering and presenting information, and third-party users can view and use MaPs data in the same ways.

Therefore, this allows humanities research to be available and accessible to both academic and non-academic communities outside of its traditional prose forms of dissemination. Turning the MaPs corpus into an RDF repository reconceptualises the dissemination and access to humanities research findings, linking ongoing humanities research to other linked open data, making it more dynamic and integrated with other data. Humanities data becomes more collaborative and interdisciplinary, as building an RDF repository required extensive knowledge in several domains (e.g. information science, English, history, literature, music, music technology, geography). Equally as important, however, has been how RDF principles have allowed
humanities researchers to consider humanistic knowledge as networked data. The significance of the project, therefore, cuts both ways: for Information Science, it has suggested a new way of thinking about humanistic research data centered on forms of association, while for humanities researchers it has shown how their findings might operate in the world of Linked Open Data outside and beyond the archival data model. This project therefore not only extends the collaborative and dissemination capability of the MaPs research community by offering its current users access to new research tools as well as an open-access web resource for the world of linked data, it also alters how we think about the fundamental links between humanities research and information management.

This panel consists of the three presenters.

Eun G. Park
Associate professor
School of Information Studies,
McGill University

Qing Zou
PhD candidate
School of Information Studies,
McGill University

Matthew Milner
Academic Associate
Department of History
McGill University

Note
1. It is SSHRC Major Collaborative Research Initiative ‘Making Publics’ at McGill University from 2005-2010, funded by the Social Science Humanities Research Council in Canada.