Chapter 13
A Data Research Infrastructure for the Arts and Humanities

Tobias Blanke & Mark Hedges

Centre for e-Research, King’s College London, United Kingdom

Abstract

This paper presents elements of a data research infrastructure for Arts and Humanities. This paper wants to offer an answer on how Humanities research data sets can be integrated into new hybrid knowledge networks. These type of objects are particularly challenging, as it is often not clear how deep links between them could look like. The paper will suggest to look at a particular set of data grid technologies to establish such deep links between Humanities data sets. As we do not think that such a task can be addressed independently from the knowledge embedded in particular communities, we look at the data integration challenges in one such community, in classics.

1. Introduction - Changing Data Perspectives in Humanities

Research is changing, not only because the intellectual process is a dynamic activity anyway, but also more specifically due to the changes new digital resources bring about. In the last decade we have witnessed the development of large numbers of digital resources for research in the Arts and Humanities [6]. These require different research means from those that the traditional ‘analogue’ approach required. Alongside these changes in the information retrieval activities for Arts and Humanities, new technologies for networking and digital data capture have emerged [7]. New forms of scholarly communication are enforced by these technologies.

At the same time, digital research objects become ever more connected. The research information unit for Humanities radically changes its features.
The old information unit for research support is a journal article, mostly simply textual. The process of research involves the reading of each others’ research outputs in articles. The new information unit on the contrary is not one single document anymore, but a complete set of linked documents and resources like datasets. In the age of hypermedia technologies, it becomes more and more difficult to exactly say what defines a document [8]. Its boundaries are not clear-cut anymore. Documents of research are not just the final product, e.g. a journal article as the accumulation of a research work. It is as much a process of learning and research that is recorded as part of the final academic output. E.g., the e-mail communications between two researchers working on the same problem could be part of their publication. This is in the interest of research, as it has never been just a result, but always an accumulation of arguments.

According to Sandy Payette, with these developments new networks of research are emerging [8]. Formerly, the actors were disconnected from the research information units, the formal publication. New hybrid networks are emerging, where formal documents stand next to more informal documents (‘grey literature’), and data, services and actors are interlinked. The knowledge cycle of research loses its well-defined stages. Knowledge objects are early registered and preserved. They stem from different sources of knowledge. The researchers will enact in the middle of these objects.

This paper wants to offer an answer on how Humanities research data sets can be integrated into these new hybrid knowledge networks. These type of objects are particularly challenging, as it is often not clear how deep links between them could look like. The paper suggests to look at a particular set of data grid technologies to establish such deep links between Humanities data sets. As we do not think that such a task can be addressed independently from the knowledge embedded in particular communities, we look at the data integration challenges in one such community, in classics.

The paper is organised as follows: In the next section, we discuss what the challenges for creating interoperability between data resources in the Humanities are. We discuss two complementary solutions, where the first one creates a community driven standard everybody uses to create these resources, while the second would be more realistic and uses grid computing to create interoperability through processing several mappings between datasets. In Section 3, we then discuss the background of data grids in the Humanities, before in Section 4 we present our case study. We look at two independently created data sets in Classics. In Section 5, we present our