Structured Data Resources Integration

Tobias Blanke

King's College London, United Kingdom

For many years humanities researchers have been producing digital outputs of various forms, a significant proportion of them being databases or corpora of marked-up texts. This presentation will offer two new perspectives on interoperability between digital classics resources. We present an approach using grid computing. In our particular case, we are using a particular example of grid software, OGSA-DAI (http://www.ogsadai.org.uk/) to integrate diverse structured data resources. OGSA-DAI supports virtualisation of data by adding standardised workflows of related processing to it. We can therefore provide "on-the-fly" common virtual interfaces to data. The OGSA-DAI (Open Grid Service Architecture–Data Access and Integration) project supports the exposure of data resources, such as relational or XML databases, on to grids. Various interfaces are provided and many database management systems are supported, with a particular view to querying, transforming and delivering data in different ways via a simple toolkit for developing client applications. OGSA-DAI is designed to be extensible, so users can provide their own additional functionality. The LaQuAT (Linking and Querying Ancient Texts) project\footnote{http://www.laquat.cerch.kcl.ac.uk} is investigating the use of the OGSA-DAI technology to provide integrated SQL based views of structured humanities data resources, including epigraphy and papyrology. The project has two complementary aims: (1) Develop an architecture for the effective integration of structure data resources in the Humanities and (2) verify this architecture using a set of three Classics resources: (a) The Heidelberger Gesamtverzeichnis (HGV) (http://www.rzuser.uni-heidelberg.de/~gv0/) der griechischen Papyrusurkunden Aegyptens is a collection of metadata records (largely bibliographic, geographical, and dating) for 65,000 Greek papyri from Egypt, stored in a large Filemaker Pro database. (b) The Project Volterra (http://www.ucl.ac.uk/history/volterra/) is a database of legal texts from the Roman empire, currently in the low tens of thousands but very much in progress, stored in a series of themed tables in MS Access. (c) The Inscriptions of Aphrodisias (IAph) is a corpus of just under 2,000 ancient Greek inscriptions from a single city in Asia Minor, published in TEI XML. Once humanities scholars are persuaded of the feasibility of this approach, there are many other datasets, in France, Italy, Germany and the US, among others, which could be exploited in such a way. The data-silo mentality could be gently undermined once scholars can see their own construct as remaining identifiable, while at the same time greatly enriched. The LaQUaT infrastructure will be sustained initially by King’s College London and the UK National Grid Service (NGS), and subsequently as part

of the European infrastructure being developed by the DARIAH project funded by the EU FP7 programme.