6th International Digital Curation Conference
December 2010

# Open Access, Reuse and Preservation of Palæoclimate Data

Gregory Tourte, University of Bristol
Emma Tonkin, UKOLN
Paul Valdes, University of Bristol

## Abstract

Data preservation for high performance applications is a challenge. Using the BRIDGE palæoclimate research material as a case study, we explore the data management issues surrounding palæoclimate research data, and discuss the issues arising from involvement of ensuring data availability for a wider interdisciplinary research environment. Using interview data as a starting point we use an automated log file analysis mechanism to explore the different types of dataset, from those with purely short-term usefulness to educational reuse, and demonstrate that the different 'types' have characteristic signatures.

The BRIDGE research group was set up in 2003, and aims to improve the understanding of natural climate and environmental variability and to use this knowledge to predict future changes more accurately and assess its impact on all aspects of human society. Its work inherits from prior developments in the same area, including a software and service basis designed to support a primary activity of the research group - palæoclimate modelling via software simulation around sampled historical data. Palæoclimate simulations generate a great deal of data that is frequently reused by researchers in the sciences and humanities, across the world. However, as a high-performance computing application, the quantities of data created are extremely large, meaning that participants have historically had to develop and apply de-facto preservation and data compression or summarisation policies.

The BRIDGE website has several roles and allows scientists to work on and with their data throughout their workflow and the life span of the data. For the climate modellers within the research group, all operations to the data after its generation on HPC facilities (institutional or national), can be done using the web interface. As the site is used throughout experiment/analysis workflows, the term "data" can describe several type of objects. The underlying motivation for curating or preserving data varies according to the discipline, the type of data being considered, use scenarios, and the wider research environment. All data are different and different disciplines and sub-disciplines will have different requirements to consider when formulating curation strategies.

Overall access to the BRIDGE dataset since 2008 shows sustained access over that period, with specific 'peaks'. Looking more closely at that data enables us to identify some of the features in the data and their causes. The proportion of internal and external data reuse is approximately equal. Educational reuse occurs periodically, as might be expected from an activity closely tied to specific course activities within a term-based structure, but there is significantly less educational reuse than there is research-oriented reuse,

The data shows us that each dataset beats with a different pulse, and that, the key to gaining a better understanding of these patterns of use is to begin to model them according to the forces that drive each underlying process. To provide a detailed model for each form of reuse is perhaps beyond practical reach, since some are more amenable to modelling than others.

The 6th International Digital Curation Conference will be taking place 6-8 December 2010 in Chicago. Please ensure you use the guidance in this template to produce your paper. Please submit your paper in one of the following formats: Microsoft Word (.doc), Open Office (.odt) or Rich Text (.rtf).