Chapter 40
ArchaeoGRID Science Gateways for Easy Access to Distributed Computing Infrastructure for Large Data Storage and Analysis in Archaeology and History

Giuliano Pelfer
University of Florence, Italy

ABSTRACT

This article describes how archaeological and historical research grew as a multidisciplinary and interdisciplinary activity due to availability of larger amount of data within the reconstruction of global historical and archaeological contexts at a global spatio-temporal scale. The increased information, also integrated with data from the Earth Sciences, has had an effect on the exponential increase of complex sets of data and of refined methods of analysis. For such purposes, this article discusses the ArchaeoGRID Science Gateway paradigm for accessing ArchaeoGRID Cyberinfrastructure (CI), a Distributed Computing Infrastructure (DCI), that can supply storage and computing resources for managing and analyzing large amount of archaeological and historical data. In fact, ArchaeoGRID Science Gateway is emerging as high-level web environment that makes easier the access, in a transparent way, to DCI, as local high-performance computing, Grids and Clouds, from no specialized Virtual Research Communities (VRC) of archaeologists and historians.

INTRODUCTION

The project of the ArchaeoGRID Cyberinfrastructure (CI) is an international initiative that, starting from the first proposals (Pelfer, 2004, 2006; Pelfer, Barceló, & McDonnell, 2005), aimed to provide the possibility to exploit advanced Distributed Computing Infrastructure (DCI) technologies for acquisition and analysis, for storage in large databases, for management and dissemination of archaeological and

DOI: 10.4018/978-1-5225-7033-2.ch040

historical data. Main ArchaeoGRID CI goals are addressed to curation and analysis of big amount of historical and archaeological data coming from written sources, remote sensing, field surveys, excavations and laboratories, with a view to research, knowledge, heritage management, information exploitation and preservation. ArchaeoGRID CI also gives the possibility to understand the historical and archaeological knowledge as commons in the era of digital revolution and allows the development and test of data simulation models with many variables and multiple space-time information.

In fact, challenges and problems affecting current historical and archaeological research deal with the analysis of historical processes in large geographical areas and for long time periods. So that archaeologists and historians increasingly view mass data sets as the means by which it can be analyzed multivariate phenomena such as the human-environmental dynamics, the development of social complexity and the integration of ancient economies. Mass data sets have been produced through collaborative field projects, regionally integrated databases held by state and federal agencies, technologies such as LiDAR, compilations of heritage data and new forms of data storage and retrieval. Terrestrial and Satellite Remote Sensing applications in archaeology and in history have become increasingly frequent.

Applications depend mostly on the rising interest of the scientific community in modern methods for surveying geographic data, which have become increasingly powerful, automatic and reliable. Remote Sensing, with its various techniques, offers the rapid acquisition of a huge quantity of metric and qualitative data from large contiguous territories, making it possible to study settlement process in a framework of larger spatial structures. For an appropriate and widespread use of these data, it is still necessary to have recourse to GIS techniques and simulation methods; as a matter of fact, only the combined use of such methodologies provides a full exploitation of their potential for an in-depth understanding and an effective utilization of data.

Archaeological and historical knowledge building is a collective work performed by the entire community of researchers, and this fits very well with the DCI paradigm of Virtual Research Communities (VRC) using large and distributed computing and storage resources. Easy access to, and management of large data sets and complex software packages, allowed by Cyberinfrastructures, is hence an asset for the production of relevant archaeological and historical knowledge. In fact, archaeological and historical data useful for analysis are growing and getting not only larger and larger, but also their complexity and heterogeneity are increasing, implying that the extraction of meaningful knowledge requires more and more computing and storage resources. Data include contributions from scientific fields as different as physics, chemistry, Earth sciences, biology, geography, anthropology and social sciences, as well as techniques developed according to the archaeological methods and theories. Modern archaeological science depends on large collections of diverse, mundane objects -such as potsherds, stone tools and debris, animal and plant remains - rather than small collections of treasures. Sites are unique, non-renewable resources easily destroyed by erosion or modern land use and data curation aspects are increasingly important. Old collections, original field notes and reports of previous work have enduring research value. Archaeological and historical data also lose much of their meaning when they are taken out of the original space-time context in which they were discovered. This spatio-temporal context of modern archaeological research spans from the dig site itself to much larger regions and long-time periods. One of the most important goals of ArchaeoGRID CI is then the development of digital repositories of archaeological and historical data containing space-time information.

Archaeological data significance and therefore databases organization changes depending on new discoveries, on the evolution of analysis methods and on the theoretical developments. One of the major