‘A puzzle in 4D’: integrating and archiving the resources of a long-term excavation project

The ‘A puzzle in 4D’ project is dealing with the problem of integration of heterogeneous and incomplete digital and non-digital records of archaeological long-term excavations to prepare them for long-term archiving based on national and international standards and open-access online publication for specialists and the general public.

Since 1966, Austrian excavations took place at Tell el-Daba (TD) in Egypt, an archaeological site revealing evidence from a society with contacts to many parts of the eastern Mediterranean dating to the 12th to 18th dynasties (early second millennium BC). Since the beginning of the excavations, the archaeological discipline has seen major changes, most notably developments in information technology have caused a shift from analogue to digitally-born data. As a result, the TD archive at the Institute for Oriental and European Archaeology OREA contains a huge and heterogeneous resource of digital and non-digital photographs, plans, drawings and written documentation. The ‘A Puzzle in 4D’ project aims to provide long-term preservation for the rich archaeological resources of this Austrian long-term excavation project. The project is also a case study for the development of a repository for archaeological data at the Austrian Academy of Sciences.

In this presentation we will describe the challenges we are facing when preparing these resources for archiving and introduce the approach we have chosen. We will describe our data model and the requirements for the system architecture. We use the CIDOC CRM ontology with extensions as a reference model to structure our data and apply semantic web technologies to perform the actual integration of the data. This presentation will show where the project stands midterm after approximately 15000 analog resources like field drawings, photos and find inventory reports have been digitized and metadata for these resources have been created. We will also present the technology stack being designed and developed for preservation and publication of this rich heterogeneous resource, comprising digital asset management system (repository), semantic mapping tool, triple store, and a python/Django based integrative web application, that also has the potential to enable the access to the data through a map interface based on Web GIS technologies.