S11 Supporting researchers in the use and re-use of archaeological data: Continuing the ARIADNE thread

Julian Richards, Franco Niccolucci, Holly Wright, Kate Fernie

Following on from the successful conversation begun at CAA Siena in 2015, this session seeks to further expand dialogue in this critical area. Ever-increasing amounts of data are available within data repositories in individual institutions, national infrastructures and international services. The EC Infrastructures funded ARIADNE project is working to bring together archaeological research data from across Europe, for use and re-use in new research. There are challenges, such as raising awareness about the available data, integrating datasets produced by very different projects using differing methodologies and various technologies. There are GIS, databases, 3D data, scientific datasets and more, all produced in a variety of languages. ARIADNE is building vital infrastructure to bring together, manage and provide access to these datasets. The project is embracing Linked Open Data, Natural Language Processing, deploying Web Services and new tools to provide enhanced access to researchers. ARIADNE is also offering training and opportunities for archaeologists to access the research infrastructure, and to share knowledge and expertise.

The aim of this session is to stimulate discussion between researchers and data specialists, and to:

- Showcase best practices and relevant work supporting access and use of digital archaeology from ARIADNE and other services
- Present case studies demonstrating innovative re-use of archaeological datasets
- Develop an understanding of the challenges in providing access to research data and the opportunities offered by ARIADNE and other services
- Discuss how these challenges can be addressed and how the opportunities can be maximized
- Generate ideas for future training, access and research

[Session themes]

The focus is on access, discovery and research reuse of archaeological datasets, and contributions are invited on the following (and related) topics:

- 3D and Visualization
- Remote Sensing and Spatial Data
- Excavation and Monument Data
- Scientific Datasets
- Grey Literature
- Linked Data
- Design of Archaeological Datasets
- Conversion of Legacy Datasets

S11-02 Methodological tips for mappings to CIDOC CRM

Maria Theodoridou, George Bruseker, Maria Daskalaki, Martin Doerr

The CIDOC Conceptual Reference Model (ISO 21127:2006) has been chosen as the core model for use in several Cultural Heritage projects including ARIADNE , ITN-DCH , PARTHENOS , and ResearchSpace . A foundational activity of these projects then has been the effort to convert the data stored in existing schemata to an expression in CIDOC-CRM and its extensions. The goal of the conversion process is to enable information exchange and integration between heterogeneous sources of cultural heritage information. In order to support the scalability of these activities, which entail careful analytic work by someone familiar both with the domain and with the ontology, a series of training events were initiated aimed at professionals who have an understanding of the CIDOC CRM and need to develop skills in data mapping techniques.

This paper will present some modelling principles and methodological good practices that we have empirically derived from the above exercises in the sys-tematic mapping of diverse cultural data sets to CIDOC CRM and its extensions. This experience is specifically derived from mapping activities using the 3M (Mapping Memory Manager) tool in the above mentioned projects. In the paper we will address a number of fundamental issues. First, we look at the question of how to determine at the beginning of a mapping - the sufficiency of CIDOC-CRM and/or its extensions, for covering a given data set and when it would prove necessary to extend the model and introduce a new class, and how. We suggest practices related to the handling of identifiers that are local in the original source data set. We also treat some special issues of how to model the roles of people and organizations, including accidental roles, and present a methodology for introducing implicit contextual information. Finally we discuss issues related to the modeling of nationality, country and imaginary places.

S11-03 An essay of mapping archaeological land-record system used by Inrap with CIDOC─CRM and CIDOC─CRMarchaeo extension using 3M online tool
Christophe Tuffery, Achille Felicetti, Patrick Jard, Nicolas Holzem, Thomas Guillemard
During a summer school on the CIDOC-CRM organized by the PIN of 21 to 25 July 2015 in Prato for the ARIADNE program(1), we have had the opportunity to use the CIDOC-CRMarchaeo extension (2). The aim of the summer school was to work effectively with a set of our own data field and try to match fields between archaeological land registration systems used by Inrap and those of CIDOC- CRM. We had the opportunity to use the 3M application (Memory Mapping Manager), an online tool developed by ICS (3), which controls whether archaeological data can be matched with the CIDOC CRM model. We tested the matching process with two fields of archaeological land recording systems used by Inrap (one based on Access and another on FileMaker Pro). Then we did the same with a prototype application on development by Inrap. This is an interface using Google Chrome, SQLite, JavaScript and HTML5, and witch aims to allow data exchange with the two previous land registration systems mentioned above. Working with the on-line application 3M (4) allowed assessing whether two of the main archaeological entities (the stratigraphic units and archaeological facts) can be matched with the CIDOC CRM model and CIDOC-CRMarcheo extension. The work demonstrated the 3M on-line tool meets the needs of matching fields of archaeological recording systems tested with classes Model CIDOC-CRM and especially its CIDOC-CRMarchaeo extension. This matching procedure has demonstrated the ability to assess whether an archaeological land recording system may or may not be considered as matching with CIDOC-CRM models and how to adapt it to conform if it doesn't initially. Therefore new tests will soon be conducted with other land archaeological recording systems used by Inrap.
(1) http://vast-lab.org/bando-summer-school-ariadne-2015/ (2) http://www.ics.forth.gr/isl/index_main.php?l=e&c=711 (3) Institute of Computer Science (http://www.ics.forth.gr), Forth Foundation, Greece (4) http://139.91.183.3/3M/

S11-04 Formalisation and reuse of methodological knowledge on archaeology across European organisations
Cesar Gonzalez-Perez, Patricia Martín-Rodilla, Elena Epure
Archaeological projects vary greatly in size, complexity, object of study, timescale and other aspects. Finding the most suitable methodology for a project is often difficult, and an inadequate choice can ruin many months' worth of fieldwork, bias data interpretation, and slow down or impede cross-project comparison of results. An archaeological methodology should be as adjusted as possible to the project needs, take into consideration techniques and approaches successfully applied in the past, and clearly expressed for better understanding

and sharing among the involved agents. These goals are usually pursued informally through the application of tacit knowledge that exists within archaeology organisations, leading to situations where: 1) it is difficult to convey what is expected to be done, especially to new team members or external collaborators; 2) methodological knowledge is underutilised and rarely reused, especially across organisations; and 3) the improvement of methodologies over time is difficult since no explicit knowledge about them exists. As we have previously proposed [Gonzalez-Perez and Hug 2012, "Crafting Archaeological Methodologies"], situational method engineering (SME) can be used to mitigate these problems. SME does not conceive a methodology as a monolithic black box, but as an assembly of pre-existing components that are selected from a repository and composed together. Each component encapsulates a proven, reusable and self-contained "atom" of knowledge that can be reused, recombined in different situations, and improved over time. In the context of the FP7 ARIADNE project, we have applied an SME approach by which the informal methodological knowledge of seven European archaeological organisations (including university departments, research centres and museums) was formalised as discrete components, stored into a database, and linked to other components. Natural language processing techniques have been used to assist in the information extraction and formalisation process. The resulting repository has allowed us to obtain variations of established methodologies to cater for different project situations; combine different methodologies for collaborations and other hybrid scenarios; and carry out a comparative analysis of commonalities and differences between the archaeological practices of the selected organisations.

S11-05 Semantic database applications at the Samtavro Cemetery, Georgia
David Bader, Aleksandra Michalewicz, Oded Green, Jessie Birkett-Rees, Jason Riedy, James Fairbanks, Anita Zakrzewska
In 2013 a paper was offered to the CAA concerning archaeological legacy data and semantic database applications, with some preliminary results for a study conducted into the Samtavro cemetery, situated in the South Caucasus in the modern republic of Georgia. The present paper presents further research outcomes of data mining the Samtavro material. Over four thousand graves were excavated at this site, used most intensively during the Late Bronze and Iron Ages, and later in the Roman and Late Antique periods. The current project focuses on the latter period'and the legacy of Soviet and post-Soviet excavations'in a collaborative effort between computer scientists based at the Georgia Institute of Technology, USA, and archaeologists at the University of Melbourne and Monash University, Australia.
Data for 1075 tombs, 1249 individuals, and 5842 grave accoutrements were collected across 74 data fields, resulting in the identification of 9 tomb types, 37 artefact types and 320 artefact subtypes. Methods tested against the Samtavro material culture included the application of clustering techniques to understand associations of related items based on patterns of co-occurrence, using traditional data mining (hierarchical link clustering) and spectral graph theory'focusing on tomb types in relation to artefact types. The other method calculated the probability of each event occurring and comparing this to what we would expect if these were truly random'focusing on artefact types in relation to biological sex and age brackets.
In some instances, our work confirmed previously established relationships, but it likewise revealed new results concerning particular entities. The project demonstrates that although sites for which comprehensive archival records exist can benefit from these types of approaches, often the greatest limitation in taking a ‘big data' approach is the relative scarcity of archaeological data.

S11-06 A catalog for archaelogical resources

Franca Debole, Nicola Aloia, Christos Papatheodorou, Dimitris Gavrilis, Carlo Meghini

The European funded project Ariadne (http://www.ariadne-infrastructure.eu/) aims to develop an infrastructure to aggregate, enrich, integrate and make available the data and services so far developed by the international archaeology research communities. The project enriches and integrates data resources such as descriptions of datasets, collections, metadata schemas, vocabularies, etc. - and services in order to create a universally accessible shared knowledge base for the archaeology domain.

In the context of Ariadne a crucial concept to integrate and manage different resources is the catalog, or registry. The catalog of Ariadne lists and describes what is available from the project partners, and more generally the whole community of archaeologists, to identify, through refined search mechanisms, the candidate resources for integration. Data registries is in effect a well-known data organization and management approach that provides an environment in which datasets, collections, metadata schemas and vocabularies along with their mappings would be hosted and described by a common schema. Actually, the data registries enhance the accessibility and re-usability of the (research) data.

This paper presents the data model of the Ariadne catalog named Ariadne Catalog Data Model (ACDM) that extends the existing data registry standards. The central notion of the model is the class ArchaeologicalResource, specialized in the classes: (i) DataResource, whose instances represent the various types of data containers (e.g. collections, GIS, datasets) owned by the ARIADNE partners and lent to the project for integration; (ii) LanguageResource, having as instances vocabularies, metadata schemas, gazetteers and mappings (between language resources); (iii) Services, whose instances represent the services owned by the Ariadne partners and lent to the project for integration. The paper presents the aggregation service that is based on the ACDM model and enables the partners to upload huge volumes of metadata to the Catalog as well as the main functionalities of the Ariadne portal (http://ariadne-portal.dcu.gr/).

S11-07 Using semantic technologies for the deep integration of research items in ARIADNE

Philipp Gerth, Wolfgang Schmidle, Sebastian Cuy

One important goal of the EU-funded ARIADNE project is to integrate data originating in a variety of different disciplines in the archaeologies and connected subjects in order to facilitate access to heterogeneous data sources. This integration on the one hand happens on a large scale by incorporating descriptions of vast amounts of research resources into the ARIADNE catalog. On the other hand experiments on the tight integration of the detailed descriptions of single item of research investigates workflows and use-cases for semantically integrated data.

In this paper we will describe a practice-orientated approach on dealing with this problem with the help of Semantic Web Technologies. A specific use case on integrating finds from various databases will be presented. We will highlight the difficulties in integrating databases with differing genesis (museum catalog, object database, excavation database), therefore different terminology, focus and languages.

The integrated datasets will be accessible via a unified programming interface, which allows rich querying possibilities. This interface lays the groundwork for a user interface, which facilitates the intuitive formulation of queries for accessing the integrated data.

S11-08 Fasti surveys

Elizabeth Fentress, Michael Johnson, Florence Laino, Stuart Eve

Fasti Surveys The Fasti platform has proved so useful for the quick registration and retrieval of excavations in participating countries that we have decided to clone it for other types of data, creating a trinity of Fasti applications. Of these, Fasti Archaeological Conservation has been easily fit onto the original model, substituting conservation projects for objects and sites for the excavations. Fasti Survey presents other challenges, however, including the serving of polygons for regional surveys, crowd sourcing, and very unequal projects. We have thus devised a two-level site, in which the first, like the excavations, simply provides an overview of each project, complete with its bibliography. We have an enormoujs head start in the sharing of the data from 320 projects already compiled by the project MAGIS (Mediterranean Archaeology GIS, http://www.iosa.it/2007/05/17/magis-mediterranean- archaeology-gis/) generously shared with us by Pedar Foss and Rebecca Schindler. The second level will serve data from participating projects.

Cancelled **S11-09 The advantages of integrated 3D photogrammetry and Reflectance Transformation Imaging empirical acquisition**

Mark Mudge, Carla Schroer

Today, there is a convergence of recent advances in computer vision, computer graphics and computational photography technologies such as Reflectance Transformation Imaging (RTI) and 3D photogrammetry. New photogrammetric technologies using Structure from Motion (SfM), along with new understanding of the geometry of multiple photographs underlying Multi-viewpoint Stereo (MVS) are driving widespread adoption of these technologies. The software produced by this convergence, run on ever more powerful computing platforms, make it possible to derive densely sampled, low uncertainty, high quality 3D measurements. These advances have greatly improved both photogrammetric measurement quality and ease of use.

These photogrammetric advancements have enabled new synergies from the combined use of RTI and photogrammetry. This talk will offer an overview of the methodology used for this joint computational photography empirical acquisition. Using the procedures in the talk, RTI representations can be captured and spatially registered to the associated 3D geometry built with photogrammetry. This spatial registration dramatically increases the information available to RTI representations.

The photo sequence used to build the RTI can use the camera calibration(s) derived from the photogrammetric Structure from Motion algorithms. This permits the complete distortion correction of the RTI photo sequence. In turn, these undistorted photos can produce optical-distortion-free RTIs.

Using advanced photogrammetry software, the registration between the RTI and photogrammetry geometry can enable depth map generation of for each photo used in the photogrammetry project. This means that each registered RTI can have corresponding depth information. This undistorted RTI and depth map combination greatly increases the informational robustness of RTI imaging.

The combined use of the normal field and material reflectance information acquired by the RTI can also be used to improve the photogrammetry-based textured 3D model. Future work will exploit this relationship to produce material representations that more closely approximate the surface properties of the imaging subject.

S11-10 Best practices to re-use remote sensing data coming from marine geophysical surveys for the 3D reconstruction of underwater archaeological deep-sites
Manuela Ritondale, Gaia Pavoni, Roberto Scopigno, Marco Callieri, Matteo Dellepiane
Since several decades, underwater archaeology takes advantage of the tools available for marine geophysics to map, document and monitoring the seabed. Impressive levels of details have been reached in underwater photogrammetry. Nonetheless, a huge amount of data has been collected in the past with several different purposes and with rather different technologies. The aforementioned materials, particularly those coming from commercial archaeology, is underexplored and might be re- used and optimized by using digital technologies to enhance the accessibility of underwater sites through virtual reconstructions. Our goal is, on one hand, to investigate the technical and the legal/bureaucratic constraints that prevent the re-use of remote sensing data coming from geophysical marine surveys. On the other side, we propose a solution to reuse stills images and video data acquired with ROVs-based surveys of archaeological sites (characterized by non-homogeneous quality and resolution), aimed at obtaining 3D image-based models. Our approach uses standard photogrammetric solutions, but is based on specific pre-processing and enhancements of the input raw data. In the first pre-processing phase, we propose an automatic frame extraction algorithm working on video streams, able to reject the damaged or non-useful frames and to select the more proper for 3D reconstruction, taking into account all the best practices for an optimal photogrammetric reconstruction. In the second phase, we apply underwater-specific image enhancement filters to either images or video frames to sharpen useful details and to correct undesirable aspects for a good reconstruction like color absorption and blur. Practical examples and first results will be discussed.

S11-11 Digging into and re-using image data for archaeology
Christopher Power, Andrew Lewis, Helen Petrie, Julian Richards, Katie Green, Mark Eramian, Ekta Bhullar, Brittany Chan, Isaac Sijaranamual, Maarten de Rijke
Each year thousands of archaeological field studies are undertaken. One of the resources with the largest potential to help archaeologists in their work is the hundreds of photographs that are taken during field studies. These photos, which can now be labelled with captions and uploaded to repositories direct from the field, could open new possibilities for re-use in many research tasks. Unfortunately, the potential of these resources has not yet been realised. Due to time pressure and lack of personnel, most photos do not have appropriate content-related metadata associated with them. While it is possible to identify what collection an image is from, or where it was taken using GPS coordinates, knowing what is actually in the image is often impossible. Even if tools were available to provide such content-related metadata, it is unlikely that the person power would be available to provide this metadata for the thousands of photos that already exist in digital form. The DADAISM Project is addressing this issue by using a mixed-initiative approach, where the deep domain knowledge of the archaeologist can be used to identify a number of key features in an image, and then automated processing can identify images that are similar, digging into the image data and extracting relevant information from the content. These identified images can then be re-used by archaeologists for their research, or even automatically labelled with appropriate content-related metadata. The new data created from this labelling can then be published to improve the robustness of searches by other archaeologists during their research. This paper will present preliminary results from the DADAISM Project on the identification of images in two specific archaeological domains, flint tools and Anglo-Scandinavian brooches, and will also present the interactive system to enable archaeologists to work with the DADAISM image identification system.

S11-12 A data integration infrastructure for archaeology

Dimitris Gavrilis, Eleni Afiontzi, Johan Fihn, Olof Olsson, Sebastian Cuy, Achille Felicetti, Franco Niccolucci

Most infrastructure projects, both recent and ongoing, involve a data aggregation task in order to bring together the heterogeneous information one expects to see in a typical EU landscape. The main reason for this is the plethora of technologies, standards, languages and practices that is found in the EU. Data aggregation typically includes the homogenization of heterogenous data through some kind of process that includes: ingestion, normalization, transformation and validation processes. The European funded project Ariadne (http://www.ariadne-infrastructure.eu/) aims at true integration of data by modelling the underlying domain and provding the technical framework for automatic integration of heterogeneous resources.

This infrastructure, comprises of a set of heterogeneous technologies such as: a metadata aggregator, including a set of enrichment and data integration micro-services, an RDF store with reasoning capabilities (through SPARQL), and a powerful indexing mechanism. The output of this process is published to a portal which can provide useful information to a variety of potential users ranging from simple visitors to domain researchers.

The data integration services can mine for links among resources, link them together and against language resources such as vocabularies. Complex records can be split into their individual components, represented, enriched and stored separately while maintaing their identity using semantic linking. These individual components are represented in the underlying model (ACDM) and include agents, language resources, datasets, collections, reports, databases, etc. Each integrated resource is assigned a URI and is published in RDF. This practice enables knowledge mining, semantic queries and reasoning engines which are provided within the project (e.g. SPARQL engine and Jena).

The technical infrastructure has been developed using various programming languages such as Java, PHP, Javascript, it is distributed spanning multiple virtual machines and brings together different established technologies and components. The portal is based on the Laravel PHP framework and uses ElasticSearch search engine to collect and browse through the data. Both the technical infrastructure and the portal will be presented and demonstrated in more detail.

POSTER

S11-P1 Combining analytical and digital data in archaeology: Towards a multidisciplinary ontological solution. The Salamis terracottas case study

Sorin Hermon, Valentina Vassallo, Giusi Sorrentino, Uros Damnjanovic

Multidisciplinary research produces heterogeneous data types, such as graphs, spectra, numbers and so on. Usually, what is presented to the research community is the interpretation of that scientific data. In a perspective of data transparency documentation, the auspicable solution would be to put at disposal all data and paradata that brought to that specific interpretation. How to make available and traceable these different kind of information? How to combine scientific, technological and archaeological data? This paper focuses on the ongoing multidisciplinary research, carried out within the frame of the EU funded projects ARIADNE and GRAVITATE, based on the integration of such data (e.g. archaeological, digital, chemical) and on the trace of the reasoning in scientific data documentation, analysis and interpretation. The organization of the information according to a rich and cross-domain metadata and to a standard conceptual reference model (CIDOC-CRMsci) will help towards the establishment of a multi-disciplinary research infrastructure. The research is applied within the terracotta figurines from Salamis-Toumba (Cyprus) project. The archaeological site

was excavated in the 19th century and the artefacts are currently stored in different museums. In the past the collection has been studied and the terracottas have been partially published along with their traditional stylistic description. Recently, within the GRAVITATE project, a further study of the collection has been undertaken. The project wants to identify and virtually reconstruct and re-unify parts of shattered or broken cultural objectsw ith a multidisciplinary approach and integration of different digital analyses (e.g. 2D and 3D digital data acquisition, non-invasive and non-destructive chemical/physical analysis, 3D geometrical analysis).

Cancelled **S11-P2 New tools for Digital Lab Notebook creation for use in Reflection Transformation Imaging**

Mark Mudge, Carla Schroer

Scientific digital documentation of cultural heritage and natural science subjects can be a powerful tool for e-science and citizen scholarship. For centuries, the scientific method has required the recording of all empirical data's collection contexts and processes in a lab notebook, which provides informational transparency and enables informed reuse.

This talk will introduce two metadata and knowledge management software tools called Digital Lab Notebook:Capture Context (DLN:CC) and Digital Lab Notebook: Inspector (DLN:Inspector), These packages take the form of user-friendly toolkits that record the contexts in which the original photographic sets of empirical information were acquired and inspect these photographic datasets for successful processing.

This methodology is designed for digital representations that are built with computational photography technologies. While this software's first iteration is optimized for the computational photography technique Reflectance Transformation Imaging (RTI), the software is designed for easy adaptation to other computational photography technologies.

The near-automatic nature of computational photography has advantages for the creation of scientific digital surrogates. A digital surrogate is a "stand-in" for "real world" subjects. They can be used for subsequent scientific or scholarly research.

Here's how it works. First, the DLN:CC harvests the capture context and process metadata associated with the empirical data and automatically maps this metadata to the CIDOC/CRMdig ontology. Next the processed photosets are inspected by DLN:Inspector to see if they will or will not successfully generate an RTI digital surrogate. When successful, the metadata is sent to the DLN. The metadata information in the DLN is then published as both XML and Research Description Framework (RDF) Linked Open Data files.

These DLN tools enable future evaluation of surrogate reliability and aids long-term archiving. When applied across the field of computational photography, the results of this strategic approach will be to enhance the digital data and knowledge sustainability of humankind's legacy.