S04 Databases and archives: How do we handle the digital archives?

Evy Berg

As the databases grow old and mature, new versions have to be created and published. Often with different technologies than the previous version to enhance user experience. But what happens to the old versions are they dumped in a digital black hole? For some databases there are rules about how to treat an obsolete version. The National Archives in Norway is responsible for the caretaking of all written materiel, also in the form of digital archives, from the public sector. The deliverance of such data is mandatory. When it comes to a SMR (Sites and Monuments register) the rules are less clear. Also, how should the vast amount of field data from collecting and analysing sites be treated? This includes both written reports, but also all files like GIS, pictures, photogrammetry etc. The file formats keeping the content can be very complex and software specific. How is this materiel archived for long term storage, and what are the chances that the files could ever be opened in 50 or 100 years' time? And who is responsible for taking care of them? The museum/research institution where the files where created? Or do they fall in under the responsibility of a National Archive? If so, what rules exist, or should be made, for taking care of the content regardless of file formats?

S04-01 No one's business: Negotiating the value of archiving archaeology in Sweden

Isto Huvila

Drawing from two empirical studies of the stakeholders of archaeological information in Sweden, this presentation discusses why the archiving of archaeology easily becomes no one's business. Comparing the Swedish situation to international examples of how mandatory depositing of various types of archaeological information has been organised and the interests of the stakeholders of archaeological archives, it is argued that the heterogeneity of archaeological information and the large spectrum of its potential stakeholders is a major, perhaps the most significant impediment of organising the management of archaeological archiving in a comprehensive manner even in cases where the situation is generally considered to be good. In Sweden, the stakeholders of the archaeological information process are focusing on different aspects of information (including various forms of tabular and geographic data, written documents, photographs, illustrations, objects, administrative documents) while no one has a comprehensive responsibility to manage the entire process. At the same time, on a practical level, the understanding of the information, how it is used and what are the responsibilities of other stakeholders varies considerably. An international comparison of archiving policies show that even if the Swedish situation can be characterised as problematic, it is far from being unique.

The presentation is bringing together findings from an interview study of Swedish professionals involved in the process of managing and archiving archaeological documentation and information, a survey study of the stakeholders of archaeological data outside of archaeology and an international comparison of the practices of managing archaeological data outside of Sweden. The work has been conducted as a part of the Swedish Research Council funded research project "Archaeological Information in the Digital Society".

S04-02 From active database to archive: Case of a Sites and Monuments Register

Evy Berg, Stian Finmark, Christer Vinje Gimse

The first version of the Sites and Monuments register "Askeladden" was released in 2004. In 2012 this version was replaced by a new, technologically updated version. Since then the old, and for management concerns, obsolete version has been running as an archive. The software it runs on is getting very old, and to preserve the content it is necessary to move to a platform better suited for long-time storage. The National Archive of Norway has decided this

collection of data is important enough to include in their digital archives as a deposition. The method used to take care of the content is to translate the table content to software designed for the purpose. The presentation will discuss the procedure and the results.

S04-03 Ensuring cultural heritage data longevity with the Arches Heritage Inventory and Management System

Annabel Lee Enriquez, Alison Dalgity, David Myers

A joint project of the Getty Conservation Institute and the World Monuments Fund, the Arches Heritage Inventory and Management System is an open source web-based geospatially enabled platform purpose built to inventory and manage immovable cultural heritage. Developed in direct response to the needs of both cultural heritage organizations worldwide as well as to the complex requirements of cultural heritage data, Arches incorporates concepts and features that help to ensure data quality and longevity. First, Arches integrates two CIDOC standards, the CIDOC-CRM (Conceptual Reference Model [ISO 21127:2006]) and the draft CIDOC International Core Data Standard for Archaeological and Architectural Heritage in order to optimize its data model and structure to be as generic and widely-applicable as possible, as well as generate standards-based self-describing data. By doing these things, Arches promotes data interoperability and portability; in other words, the data housed in the system is structured to actually outlive it. Second, Arches is able to accommodate the myriad ways in which cultural heritage data is recorded. The platform can accommodate multiple attributes and attribute values, grouped temporally, spatially, and thematically with relationships defined between various resources. This flexibility enables all data recorded throughout the lifespan of a cultural heritage resource to be maintained within the system. Third, the project adopted an open source development model using a license for free use with the intention that the Arches open source development community would sustain and improve upon the platform for as long as the system is viable. This allows heritage organizations to implement and develop their own Arches instance as economically as possible and with as much control over their system for as long as needed. In these three ways, combined with an easy to understand yet powerful interface, Arches promotes the longevity and quality of inventory data about cultural heritage places.

S04-04 Row, row, row your boat gently… upstream: A methodological approach to access Portuguese bioarchaeological data using a computer database

Cristina Barroso Cruz, Ana Seabra, Filipa Neto

In accordance with Portuguese Law, which holds the recording and inventory of information as the cornerstone for Heritage Protection and Safeguard, the DGPC - Directorate-General of Cultural Heritage in the capacity of being the state entity responsible for heritage management houses the main archive of Portuguese Archaeology, including fieldwork reports for storage, preservation and consultation. Since 1997 an increase of archaeological fieldwork in Portugal has lead to a consequent increase of the number of reports (grey literature) produced. It is also during this exponential growth that the recording of human remains has evolved from a brief, footnote-like mention to more detailed records, due in part to a stricter demand for accountability concerning this kind of heritage. In an effort to overcome the absence of common standards for the production of anthropological field reports and also to minimize the discrepancies brought on by inter-observer error, in 2014 a legal recommendation was produced addressing the specificities of archaeological human remains and implementing a common core of ad minima procedures to be reported. In addition to the archive of written reports, the DGPC also develops and maintains a database for the recording of a wide spectrum of archaeological data Endovélico, which in recent years has been developing a module dedicated to register bioanthropological data, with the aim to constitute a

complementary record to the written reports, making their information more accessible for research.
In this poster communication we will present our experience with the implementation of the bioanthropological module and offer a reflection regarding the selection of the data to input and the best way to record it with the purpose of enabling access and potentiating research, therefore creating a valuable source of information for the improvement of knowledge.

S04-05 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.

S04-06 Digital archives: More than just a digital skeuomorph?
Emily Nimmo, Peter McKeague
Historic Environment Scotland (HES) continues the important functions of the Royal Commission on the Ancient and Historical Monuments of Scotland as the custodian of the national collection for archaeology and architecture of Scotland. Our Collections and inventory have built up over several decades and we have been collecting digital archive since the early 1990s. We have an established and mature approach to the long term care and dissemination of documentary archive in Scotland.. In this paper we will share our experience and expertise in providing for the long term preservation of digital archaeological archives and how this has to be a collaboration between the data creators and the archives if it is to succeed, not simply an activity at the end of a project or something ‘the archive does'.

However established approaches to digital archiving are essentially developing a skeuomorph of traditional analogue archives. This is the inevitable outcome of project-based deposition and established archival procedures from accessioning and cataloguing to publishing and reporting coupled with often chronic under-resourcing for the long-term preservation of the archive in all its guises. Issues over file formats and preservation only compound the approach.
Digital formats offer new opportunities and insights however we argue that current approaches to cataloguing and making project archives accessible prevent this material from reaching its full potential. Using examples from databases and GIS we question alternative approach to the way we catalogue and present aspects of the digital archive to our users. Digital archiving is not and should not be the continuation of existing archival practices but requires collaboration at an early stage to ensure consistency of approach, to deliver efficiencies and benefits beyond individual projects.

S04-P1 Modelling between digital and humanities: Thinking in practice
Øyvind Eide, Arianna Ciula, Cristina Marras, Patrick Sahle
The proposed poster will present "Modelling between digital and humanities: thinking in practice," a newly started 18 month project with the aim of bringing together scientific and scholarly modelling with creative cognitive practices of coming to know. Modelling is a creative process of reasoning in which meaning is made and negotiated through the creation and manipulation of external representations. Making external representations to reason with has been part of the scholarly Western tradition at least since the Enlightenment; digital humanities extends this practice by actively creating digital artefacts in different media.
To integrate these theories of how humans think through things with a practical dimension, the project will make use of digital humanities as an interdisciplinary departure to study modelling as anchored both to computer science and to the humanities. The project aims to link scholarly modelling as a formal and informal reasoning strategy across disciplinary boundaries, spanning also social, life and techno-sciences and culture heritage modelling, and bridging across modelling in research and in teaching.
Our working hypothesis is that in digital humanities research, implicit and explicit models of cultural phenomena are integrated into external metamodels, e.g. graphical representations, which often embed natural language and are informal. These metamodels can be iteratively translated towards computable implementations via a variety of more or less formal models: models for. The analysis of modelling practices in the areas outlined above will hence aim at gaining new insights in the epistemology of modelling.
While textuality mediates the world we live in, events are central to an epistemological perception and description of the processes shaping this world. In the proposed poster the links between the project and modelling in cultural heritage will be highlighted, with a focus on archaeology and with event modelling as the central case study.

S04-P2 Sustainability = separation: Keeping database structure, domain structure and interface independent
Ian Johnson
The conventional approach to relational databases implements the knowledge domain structure directly as database structure; tables represent entities and specific relationships between them, forms represent views which depend on specific joins, workflows and interfaces are customised through programming. This approach is inherently unsustainable due to dependence on specific software/versions and essential knowledge embedded in the custom programming rather than the database.

A number of archaeological databases, from IDEA in the 1990s through to FAIMS and Heurist today, use a fixed database structure across all applications, and represent the knowledge domain structure as data within the database. They generally incorporate documentation of the knowledge domain structure directly within the database.

In this poster I demonstrate the advantages of such an approach for long-term sustainability, since the only requirement for data accessibility is a database which can manage a set of SQL tables, against which a set of standard SQL queries can be issued to retrieve the knowledge domain structure and content. This requirement should be satisfied as long as SQL servers exist which, despite fierce competition from other models, show no signs of fading away. The poster will also argue that the original user-interface - so often confused with the database itself, due in part to desktop software where the interface is tightly integrated with the data storage infrastructure - is merely an end-user convenience and should be regarded as transitory; its value can be sustained with simple documentation such as an explanation of workflows accompanied by screenshots.

S04-P3 Bioarchaeology module—Loading please hold... Recording human bioarchaeological data from Portuguese archaeological field reports

Ana Seabra, Cristina Barroso-Cruz, Filipa Neto

In accordance with Portuguese Law, which holds the recording and inventory of information as the cornerstone for Heritage Protection and Safeguard, the DGPC – Directorate-General of Cultural Heritage in the capacity of being the state entity responsible for heritage management houses the main archive of Portuguese Archaeology, including fieldwork reports for storage, preservation and consultation.

Since 1997 an increase of archaeological fieldwork in Portugal has lead to a consequent increase of the number of reports (grey literature) produced. It is also during this exponential growth that the recording of human remains has evolved from a brief, footnote-like mention to more detailed records, due in part to a stricter demand for accountability concerning this kind of heritage. In an effort to overcome the absence of common standards for the production of anthropological field reports and also to minimize the discrepancies brought on by inter-observer error, in 2014 a legal recommendation was produced addressing the specificities of archaeological human remains and implementing a common core of ad minima procedures to be reported.

In addition to the archive of written reports, the DGPC also develops and maintains a database for the recording of a wide spectrum of archaeological data – Endovélico, which in recent years has been developing a module dedicated to register bioanthropological data, with the aim to constitute a complementary record to the written reports, making their information more accessible for research.

In this poster communication we will present our experience with the implementation of the bioanthropological module and offer a reflection regarding the selection of the data to input and the best way to record it with the purpose of enabling access and potentiating research, therefore creating a valuable source of information for the improvement of knowledge.