CAA2016

Oslo 30 march-3 april

Oceans of Data

Proceedings of the 44th Conference on Computer Applications and Quantitative Methods in Archaeology

Edited by

Mieko Matsumoto and Espen Uleberg

CAA Computer Applications & Quantitative Methods in Archaeology

Archaeopress Publishing Ltd
Summertown Pavilion
18-24 Middle Way
Summertown
Oxford OX2 7LG

www.archaeopress.com

ISBN 978 1 78491 730 2
ISBN 978 1 78491 731 9 (e-Pdf)

Language Editing: Rebecca Cannell
Cover Design: Bjarte Einar Aarseth

Printed in England by Oxuniprint, Oxford

This book is available direct from Archaeopress or from our website www.archaeopress.com

Contents

Mieko Matsumoto and Espen Uleberg (eds) 2018. *CAA2016: Oceans of Data*
Proceedings of the 44th Conference on Computer Applications and Quantitative Methods in Archaeology.
Oxford: Archaeopress

Oceans of Data: Creating a Safe Haven for Information

Christian-Emil ORE

University of Oslo
c.e.s.ore@iln.uio.no

Abstract
The conference theme of CAA2016 was "Exploring Oceans of Data", hinting at the vast amount of digital data resulting from digitisation projects and from all kind of electronic measuring gadgets used to document excavations and surveys. The digital data are much more fragile than paper and can easily evaporate. The last decade we have been told to avoid information islands and the slogan has been "Open the data silos". Is it easier to find a needle in an enormous haystack than in many small? If we are satisfied with the result lists of the google-type answer, it is a clear yes. If we want to build scientific data sets which may be aggregated into larger data sets, we need common authority systems and ontologies for data integration. Archaeology is neither library nor archival science, but methods for construction, curation and reuse of archaeological data sets must be the main focus. Standardised conceptual data models can ease curation and secure long term reusability and will not impose a straitjacket on research.

Keywords: data preservation, reuse, ontologies, linked data

Introduction

The conference theme of CAA2016 was 'Exploring Oceans of Data', hinting at the vast amount of digital data resulting from digitisation projects and from all kind of electronic measuring gadgets used to document excavations and surveys. A quick look at the CAA2016 book of abstracts will tell you that only a minority of the presentations actually address issues connected to curation, organisation and (re)use of the 'oceans' of data. The majority of the presentations are, as at all CAA meetings, about innovative and experimental use of computers in archaeology and about the application of existing technology to new scientific projects, that is, about activities producing even more data.

This is not unexpected. Academic training is in general focused on how to gain new insights. The most important outcome of a project is considered to be the academic publications. Even in empirical fields like archaeology the main path to success is the number and quality of your academic publications. The future faith of the empirical material and the documentation of it accumulated in an archaeological project are of almost no importance after the paper is published. You will not lose your PhD and your paper will not be rejected after having been published in the case of your field material being later destroyed. The system for academic credits gives little or no award for the preparation of your material for long term preservation and even for the development of research infrastructures to keep such material.

The full title of CAA is Computer Applications and Quantitative Methods in Archaeology. In 2012 CAA celebrated its 40th anniversary. The CAA2012 had a special session called "personal histories" where key members shared their CAA memories. The session was captured on video, can be viewed online and is highly recommended (Personal Histories Project, 2012). Most of the memories are about social events and about the primitive state of computers back then, as it should be. However, there were a few caveats, one by John Wilcock who founded The Research Centre for Computer Archaeology at North Staffordshire Polytechnic in 1970 where a number of central British CAA members got their training. With reference to his participation in the rescue work of the data from the very large BBC Domesday Book project (BBC, 2016), originally published on two laser disks in 1986, Wilcock ended his talk with a comment of the importance of proper archiving preferably on paper (!) and stated 'We can't use the Cloud unless we can read it'.

The flood of digital data and the current situation

Wilcock represents the senior league in our field and many may consider his worries as those of the old man. Today almost all new information is born digital and a majority of information in the world is in a digital format. Paper based data are voluminous and less accessible than digital data but are undeniably much more stable and can eventually find its way to collections and archives. Digital data are fragile and will not usually be readable after years in the attic. Without proper actions, the floods of digital data may evaporate and the oceans of data shrink as an Aral Sea. This may not be of importance for a large number of the billions of instant images in the social media. It will however be a catastrophe for our understanding of the past if the carefully collected documentation of all archaeological excavation since the 1990s disappeared. The problem is twofold. The basic challenge is that the digital data must at least be preserved in the format it was recorded. For example old magnetic tapes and home burned CDs tend physically to deteriorate and PCs with hard disks are recycled because nobody remembers or cares what is on them. This seems to be trivial problem, but may be the most widespread reason for the loss of data. To establish a solution to this problem of 'bit-stream preservation' is at the same time very trivial and very complex. It is trivial because one only needs a permanent organisation responsible for taking proper care. It is complex and even very difficult because such a caretaking body will require permanent funding. Even though it is a prerequisite that the digital data are preserved, they may be of little use if we don't know the format and can interpret the data as meaningful information. The second task is to ensure that the data are also stored in an open, transparent and non-proprietary format. Thus a caretaking body must ensure that the data are stored in such a format. This is not always possible. Measurement data from remote sensing equipment like GPR and LiDAR should be stored as raw data with a sufficiently detailed specification of the format to enable decoding of the data. A parallel is the TIFF image format designed so that a skilled programmer can understand the format and decode the data within two weeks-time without any previous knowledge of the format.

To meet the two challenges described in the above paragraph is the basic task for the long term preservation of digital data in all fields — not only for archaeology. In Europe there are two very good examples of institutions taking care of digital archaeological data: DANS in the Netherlands and ADS in UK. In recent years other initiatives have been established, for example the German IANUS (Heinrich and Schäfer, 2016; or Kolbmann, 2014), the US based tDAR (2015) and Open Context (2016) and others. Unfortunately, many countries do not have such services today. In the ARIADNE project the situation in Slovenia and in Ireland has been studied. According to the ARIADNE booklet (ARIADNE, 2014) the situation is far from ideal. From Slovenia it is reported that 'all digital data from excavations prior to 2013 has been left completely in the hands of the researchers, being either public or private legal bodies'. The only open sources are the written short obligatory excavation reports. According to the booklet, there is a growing understanding for the need of a national depository for archaeological data like the DANS, and some initiatives have been taken. In Ireland the situation is quite similar. The economic boom in the 1990s required a large number of rescue excavation done by private contractors. In the following economic crisis after 2008 many of these firms were closed down or went bankrupt. The fate of the digital data from the excavations is at best unclear. In Ireland as in Slovenia the only available information is what is written in the short obligatory excavation reports.

Based on conversations with colleagues it is my impression that the Irish and Slovenian experience is far from unique. In 2015 the Swedish National Heritage Board (Riksantikvarieämbetet) did a survey of the state of the data from contract excavators, both private companies and regional museums (Törnqvist, 2015). The results of the survey describe a picture quite similar to the Irish and Slovenian with some important differences. The data are stored in many different formats on PCs and servers in several formats. Only the reports, mostly printed on paper, are sent to the Swedish National Heritage Board. The contractors report that they don't have the resources to convert, systematise and transfer the data. On the positive side the survey gives a detailed and more or less complete picture and the data are recoverable given sufficient resources. The Swedish National Heritage Board has established a five year programme, Digital Arkeologisk Prosess (Digital Archaeological Process), 2014–2015, where one of the objectives is to take care of the excavation data.

Requirements from the cultural heritage authorities and the availability of organisations like the Dutch DANS ("Digital Archiving and Networking Services") may solve the Irish-Slovenian-Swedish problem which exists in many other countries as well. There are positive initiatives in Slovenia and Austria, but they have to be followed up by modernising the legislation and archiving requirements in the excavators' contracts.

Three levels of data preservation

One may argue that a digital data archive is simply a giant data silo and the stored data are not directly accessible. A silo is a device for safe storage and an important feature is that one can extract in an unspoiled condition what was originally inserted. The availability of safe data silos for long term, say 100 years, preservation of digital excavation data must be the basic requirement, but such services are not available in many, perhaps most, countries. To ensure that excavation data are stored properly for later use is level 1.

Under the assumption that we manage to create and preserve the data sets, how can the data be utilised? In an ideal world it should be possible for a given area to see a map based view of all sites, monuments, excavations and surveys. It should be possible to zoom in and see the excavation area with structures and finds together with a listing of all data sets, reports and publications documenting the excavation and the researchers' interpretations. This will indeed open the silos.

A data set from a given excavation corresponds to a book in a digital library or a box with documents in a traditional archive. It is a closed, self-confined unit. Data archives like the Dutch DANS or the British ADS store such self-confined units. To find the relevant material, users of libraries and archives are depending on a good catalogue with detailed metadata about each archival unit and books. For an excavation archive this will be detailed information about the excavation, for example: where (coordinates), when, how, what was excavated and who was responsible. In addition to being a finding aid in a given archive, the metadata from all archives should be accessible via APIs and as linked (open) data. Combined with site and monument registries this will create a common index to archaeological excavations and surveys. This will not give full access of the content of the data sets, but it will give open access to the storage units in the silos and make it possible to create maps or other aggregated overviews over known archaeological sites and field research as well as information about where to find the data sets. This is level 2.

In the spirit of the open-the-silos slogan, the content of the data sets should be made available as linked data. This is level 3. In this context a photo, a multimedia object or a LIDAR point cloud will be a singleton member of a data set. If it is analysed into smaller parts then the resulting data will be a data set with links to the original.

One may wonder if it is meaningful to combine detailed excavation data from say the Hellenistic Egypt with data from an excavation of an early Iron Age site in central Norway. The degree of meaningfulness of combining data from a series of excavation is, however, up to each researcher to decide. It can be relevant to compare data from sites with long houses from the Merovingian period in North Germany and Scandinavia. On a very local level, say the remains of the medieval town of Oslo, merging the excavation databases into one will indeed be meaningful.

There is always a snag. A meaningful linking of data (and data sets) requires compatible data models. Integrating databases even just on the level of a common index without a common understanding and harmonisation of the semantic categories and the data model is meaningless. Such a harmonisation may require resources well beyond the limited resources of a small single project. Even today most archaeological projects follow the requirements or recommendations in some manual. For example, one will follow the guidelines when taking samples for dendrochronological analysis. Correspondingly, the overall information architecture of an excavation database should follow some well-defined standard model.

Linkable data, linked data and the web

Internet has existed 40 years and World Wide Web was invented for almost 25 years ago. The idea of common access to all archaeological information and research information in general is of course not new. Besides the traditional archives and libraries, an early example is found in Vannavar Bush's 1945 paper, *As we may think* (Bush, 1945). In his paper Bush describes the Memex (Memory Extension), a machine with indexed and interlinked microfilms. The basic idea is that users may add their own association between images on the films, that is, between entries in data sets. These associations or links can also be annotated. Bush argues that this is the way a human thinks. We follow a series of associations, maybe with side tracks. To store such association, links are important, according to Bush. There are clear similarities between Bush's line of arguments and what we can read in papers about hypertext in the 1980s, see for example Conklin (1987) for a time typical overview. It is also worth noting the many web annotation initiatives that follows the suggestions in Bush's paper. A prominent example now adopted by the W3C is the OpenAnnotation Initiative (Open Annotation Collaboration, 2016). The World Wide Web in itself was originally an implementation of the hypertext idea. Curiously it didn't receive much acceptance in the traditionally academic hypertext scholars in the first few years (Richie, 2011). The inventor of the term 'hypertext', Ted Nelson, found the web and html-encoded texts too simplistic compared with his own Xanadu-system. Around 1990 hypertext and text encoding was to a large extent done by especially interested persons in the fringes of departments for language and literary studies. It was definitely not a topic of great interest among archaeologists. One of the few exceptions must have been the late archaeologist Sebastian Rahtz who later was active in the TEI-community (TEI, 2015). The first very few CAA discussing hypertext and linking of excavation archives was given by the late Nick Ryan at CAA1994, *The Excavation Archive as Hyperdocument?* (Ryan, 1995). The year after, the first paper on extraction of information from XML-encoded archaeological texts was presented at CAA1995 (Holmen and Uleberg, 1996). At CAA1997 the elegant Danish initiative *Gods and Graves* (Hansen, 1999) was presented. This was a web publication combining the Danish sites and monuments registry and the finds database at the Danish National Museum.

Since then web presentations of archaeological information has become the normal. Web based services for archaeologists followed suit. At CAA1996 *ArchWeb* (Wansleeben and van den Dries, 2000) was presented. This was a web based data service for archaeologists in the Netherlands. ArchWEB was a forerunner for the very successful *E-depot Dutch archaeology* (EDNA) at DANS which was launched ten years later, in 2006. As mentioned earlier, a general problem is that in most countries there are no formal obligations to deposit digital excavation data in a common permanent archival system. In many countries (e.g. Ireland, Norway, Slovenia, Sweden) the only requirement is to send a short excavation report to the archaeological authority. The success of DANS is founded on the obligations to deposit the data and the existence of an easy to use deposit system with a formal quality standard the (meta) data must conform to.

Both DANS in the Netherlands and ADS in UK have become successful archives for archaeological data sets. Well-functioning data archives are an absolutely necessary condition for access to data sets. The existence of the data sets is in itself not a sufficient condition for exchange or aggregating data in a meaningful way. The issue has been discussed in many CAA presentations starting with Nick Ryan in 1994 (Ryan, 1995), see also Verhagen, Sueur and Wansleeben (2011) for a practical discussion.

The need of well-defined common conceptual models

In 2001 Berners-Lee, Hendler and Lassila (2001) foresaw a second web, the semantic web, readable for computers and based the RDF-technology. Compared with the traditional web it has not become an undisputable success. Five years later Berners-Lee (2009) suggested a more concrete and practical solution called Linked (Open) Data:

- Use URIs to identify things.
- Use HTTP URIs so that these things can be referred to and looked up by people and user agents.
- Provide useful information about the thing when its URI is dereferenced, using standard formats such as RDF/XML.
- Include links to other, related URIs in the exposed data to improve discovery of other related information on the Web.

The linked data mechanism has become very popular, for example in DBpedia. It is easy to understand, implement and use. In a CAA context especially spatial referential data and type thesauri, are published as Linked Open Data (LOD). In many linked data communities the focus has been on making as much data available as possible under a somewhat post processual device 'everything can be linked':

- Increased amount of data = Increase of amount of information
- Increased interlinking = Increase in information
- Popular view: everything is connected to everything

This is of course not true and may be called 'the principle of entropy fallacy'. Information is generated through exclusion using meaningful distinctions according to a common conceptual model or formal ontology. Organising data using such ontologies and the ontologies themselves can be expressed as RDF triples. Consequentially, Linked Data can function as a medium for generating meaningful statements about data. In other words, to create more than trivial use of linked data in a domain, the linking has to be in compliance with a well-defined ontology for the domain in question.

In Finland a series LOD projects called 'sampos' (after the Finnish mythological object sampo) for Finnish history and culture has been published. The team behind many of these lead by Eero Hyvönen at the Aalto University argues that the well-known 5-star (Bernards-Lee, 2009) model for Linked Open Data should be extended to a 7 star model. The sixth star requires that the schemas (RDFS) used in a LOD data set are explicitly described and published together with the data set if not publicly accessible on the web. The seventh star requires that the "quality of the data set against the given schemas used in it explicated so that the user can evaluate whether the data quality matches her needs" (Hyvönen *et al*., 2014). The most recent of these sampos, called the WarSampo, is about Finland in the Second World War and links a large number of data sets. In WarSampo CIDOC-CRM (CIDOC CRM, 2016.) is used as the harmonising basis for modelling data, with events providing the semantic glue for data linking (Hyvönen *et al*., 2016). This is an elegant example of an advanced LOD application scalable through the use of a common conceptual model designed for data integration. According to Hyvönen the Finnish WarSampo can be extended to larger parts of the history of Second World War by mapping the content of archives and collection to the common conceptual model. There is some distance from the Finnish WarSampo to archaeological excavation data sets. Still the WarSampo illustrates what can be achieved.

Even though an excavation plan may change due to unexpected finds, the documentation methods will usually remain constant. The recorded information will be the result of human interpretation. Raw data are not raw (see Gitelman, 2013). They are a result of both the excavation plan and method and an

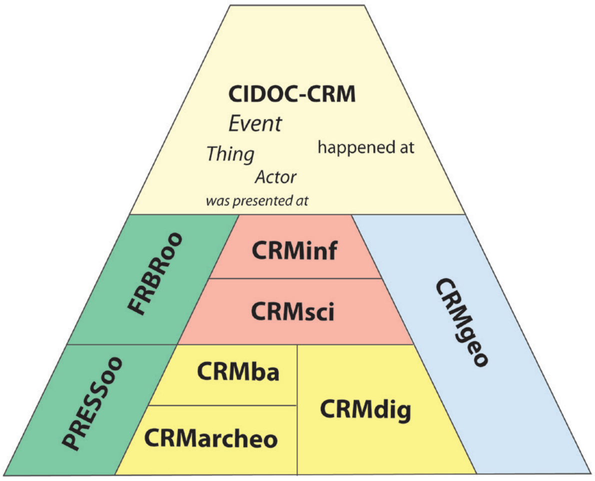

Figure 1. The CIDOC-CRM family of models.

interpretation of what is observed. The sixth and the seventh stars have to be a part of the excavation data set. A data set from an excavation without an explicit data model is meaningless. It is like artefacts without contextual information. To achieve something like an 'ArcheoSampo', the data sets have to be mapped to a common ontology. The original data sets must be kept and the mapping must be formally described. The ARIADNE project is an excellent example of how this can be done by using the family of the CIDOC-CRM ontologies and the mapping specification language X3ML (Marketakis *et al.*, 2016).

A comment on the situation in Norway

In Norway the situation is easier with fewer actors. As a result of two large digitisation and database projects 1992–2006 (see Holmen and Uleberg, 1996; Ore, 1998) there is one common database for finds and one for the site and monuments registry. The overarching data model was inspired by the event oriented model developed at the Danish National museum in 1988-89 (Eaglestone *et al.*, 1996; Rold, 1993), and the data format for texts was based on TEI (Text Encoding Initiative) (TEI, 2015) developed by text philologists from 1987 onwards.

In Norway excavations are done by 7 museums, 19 counties and one semi private foundation. The Swedish GIS based documentation system INTRASIS (see Intrasis, 2016.) for archaeological excavations has become a de facto standard. Even in this tidy situation the backlog of digital excavation data from 1990 and onwards is also a problem in Norway. There is no common database with data sets from excavations and the archival praxis is varying. The Norwegian archaeological institutions must dare to take the small step to publish their data sets in the similar way as is done by DANS and suggested by the ARIADNE project.

Summing up

Archaeology is neither library nor archival science. But a substantial part of archaeological training is how to do sound and accurate documentation of contexts. Methods for construction, curation and reuse of archaeological data sets should be in the central focus as well. Standardised conceptual data models can ease curation and secure long term reusability. Used for these purposes models will not put straitjackets on research.

In the 1980's the hypertext was thought to do the job. The web in the 1990's was an implementation of hypertext on a global scale. Linked data and the semantic web followed without really solving the problem.

The last decade we have been told to avoid information islands and the slogan has been 'Open the data silos'. Is it easier to find a needle in an enormous haystack than in many small? If we are satisfied with the result lists of the google-type answer, the answer is a clear yes. If we want to build scientific data sets which may be aggregated into larger data sets, we need common authority systems and we need to impose some common structure on the data. To do this in a meaningful way, we have to do an ontological analysis of why and how data are produced in our disciplines. That is, we need to understand our data and establish consistent and well-founded data models or ontologies, (Oldman *et al.*, 2016). On the basis of those we can see how our data may be mapped to a common model for integration. Well defined data models are necessary to define standards for storage formats and may help us to write the necessary specification for contract excavators.

In the CAA context the main focus will and should be on innovative ICT applications and good practice. The methodology of common consistent but flexible models for data integration will be a relatively small, but important core activity. The data and the artefacts is all what remains from an excavation. They must be handled with care. We need to create accept among the stakeholders that data are at least as important as the artefacts and need long term curation. This is a task for the entire CAA community as well as for the cultural heritage sector as a whole.

References

ARIADNE (ed.) (2014) *ARIADNE. The way forward to digital archaeology in Europe*. Available at: http://www.ariadne-infrastructure.eu/About (Accessed: 6 March 2017).

BBC (2016) *Domesday Reloaded*. Available at: http://www.bbc.co.uk/history/domesday (Accessed: 19 July 2016).

Berners-Lee, T. (2009) *Linked Data*. Available at: https://www.w3.org/DesignIssues/LinkedData.html (Accessed: 30 July 2016).

Berners-Lee, T., Hendler, J., and Lassila, O. (2001) 'The Semantic Web', *Scientific American*, 284 (May), pp. 34–43. doi: 10.1038/scientificamerican0501-34.

Bush, V. (1945) 'As we may think', *The Atlantic*, July 1945.

CIDOC CRM (2016) Available at: http://www.cidoc-crm.org (Accessed: 31 July 2016).

Conklin, J. (1987) 'Hypertext: an introduction and survey', *Computer*, 20 (9), pp. 17–41. doi: 10.1109/MC.1987.1663693.

Eaglestone, B., Holton, R., and Rold, L. (1996) 'GENREG: A historical data model based on event graphs', in Wagner, R.R. and Thoma, H. (eds) *Database and eExpert systems applications. DEXA 1996*. Berlin, Heidelberg: Springer, pp 254–263. Series: Lecture Notes in Computer Science 1134.

Gitelman, L. (ed.) (2013) *Raw data is an oxymoron*. Cambridge, MA: MIT Press.

Hansen, H.J. (1999) 'Digital Danish archaeology. Gods and graves—an internet publication on the Bronze Age', in Dingwall, L., Exon, S., Gaffney, V., Laffin, S., and van Leusen, M. (eds) *Archaeology in the Age of the Internet. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 25th Anniversary Conference, University of Birmingham, April 1997*. Oxford: Archaeopress, pp. 121–125. Series: BAR International Series 750.

Heinrich, M. and Schäfer, F.F. (eds) (2016) *IANUS—Forschungsdatenzentrum Archäologie & Altertumswissenschaften*. Available at: http://www.ianus-fdz.de/ (Accessed: 21 November 2016).

Holmen, J. and Uleberg, E. (1996) 'The National Documentation Project of Norway—the archaeological sub-project', in Kamermans, H. and Fennema, K. (eds) *Interfacing the past. Computer Applications and Quantitative Methods in Archaeology CAA95 Vol. I*. Leiden: University of Leiden, pp. 43–46. Series: Analecta Praehistorica Leidensia 28.

Hyvönen, E., Heino, E. Leskinen, P., Ikkala, E., Koho, M., Tamper, M., Tuominen, J., and Mäkelä, E. (2016) 'Publishing Second World War history as Linked Data Events on the Semantic Web', in *Digital Humanities 2016, Conference abstracts*. Kraków: Jagiellonian University and the Pedagogical University of Kraków, pp. 571–573. Available at: http://dh2016.adho.org/abstracts/222 (Accessed: 6 March 2017).

Hyvönen, E., Tuominen, J., Alonen, M., and Mäkelä, E. (2014) 'Linked Data Finland: A 7-star model and platform for publishing and re-using linked datasets', in Presutti. V., Blomqvist, E., Troncy, R., Sack, H., Papadakis, I., and Tordai, A. (eds) *The Semantic Web: ESWC 2014 Satellite Events. Anissaras, Crete, Greece, May 25–29, 2014. Revised Selected Papers*. Springer International Publishing Switzerland, pp. 226–230. Series: Lecture Notes in Computer Science 8798. doi: 10.1007/978-3-319-11955-7_24.

Intrasis (2016) Available at: http://www.intrasis.com/ (Accessed: 21 July 2016).

Kolbmann, W. (2014) *A digital registry for archaeological find spots and excavation documentation in IANUS*. Berlin: Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin. Series: Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft 359.

Marketakis, Y., Minadakis, N., Kondylakis, H., Konsolaki, K., Samaritakis, G., Theodoridou, M., Flouris, G., and Doerr, M. (2016) 'X3ML mapping framework for information integration in cultural heritage and beyond', *International Journal on Digital Libraries*, (June), pp. 1–19. Available at: https://link.springer.com/article/10.1007/s00799-016-0179-1/fulltext.html (Accessed: 6 March 2017). doi: 10.1007/s00799-016-0179-1.

Oldman, D., Doerr, M., and Gradmann, S. (2016) 'Zen and the art of Linked Data: new strategies for a Semantic Web of humanist knowledge', in Schreibman, S., Siemens, R., and Unsworth, J. (eds) *A new companion to digital humanities*. Chichester, UK: John Willey & Sons, Ltd, pp. 251–273. doi: 10.1002/9781118680605.ch18.

Open Annotation Collaboration (2016) Available at: http://www.openannotation.org (Accessed: 28 July 2016).

Open Context (2016) Available at: https://opencontext.org/ (Accessed: 21 November 2016).

Ore, C.-E. (1998) 'Making multidisciplinary resources', in Burnard, L., Deegan, M., and Short, H. (eds) *The Digital Demotic. Selected papers from DRH97, Digital Resources for the Humanities conference, St Anne's college, Oxford, September 1997*. London: Office for Humanities Communication, pp. 65–74. Series: Office for Humanities Communication, King's College, London, Publication 10.

Personal Histories Project (2012) *Personal Histories of 40 years Computer Applications and Quantitative Methods in Archaeology (CAA) Conference. 2–4 pm Wednesday 28 March 2012 at CAA 2012 Southampton Avenue Campus, University of Southampton*. Personal Histories Project, University of Cambridge. Available at: http://www.sms.cam.ac.uk/media/1357554 (Accessed: 19 July 2016).

Richie, I. (2011) *The day I turned down Tim Berners-Lee*. Available at: https://www.ted.com/talks/ian_ritchie_the_day_i_turned_down_tim_berners_lee/transcript?language=en (Accessed: 28 July 2016).

Rold, L. (1993) 'Syntheses in object oriented analysis', in Andresen, J., Madsen. T., and Scollar, I. (eds) *Computing the Past. Computer Applications and Quantitative Methods in Archaeology CAA92*. Aarhus: Aarhus University Press, pp. 213–220.

Ryan, N. (1995) 'The excavation archive as hyperdocument?', in Huggett, J. and Ryan, N. (eds) *Computer Applications and Quantitative Methods in Archaeology 1994*. Oxford: Tempus Reparatum, pp. 211–220. Series: BAR International Series 600.

tDAR (The Digital Archaeological Record) (2015) *What can you dig up?* Available at: http://www.tdar.org/ (Accessed: 21 November 2016).

TEI (Text Encording Initiative) (2015) *TEI: P5 Guidelines*. Available at: http://www.tei-c.org/Guidelines/P5/ (Accessed: 19 July 2016).

Törnqvist, O. (2015) *Inventering av data från uppdragsarkeologin. Aktörer, data och förutsättningar för att* återanvända *informationen.* Visby, SW: Riksantikvarieämbetet. Series: Rapport från Riksantikvarieämbetet. Available at: http://samla.raa.se/xmlui/handle/raa/9246 (Accessed: 21 July 2016).

Verhagen, P., Sueur, C., and Wansleeben, M. (2011) 'Setting a standard for the exchange of archaeological data in the Netherlands', in Jerem, E., Redő, F., and Szeverényi, V. (eds) *On the Road to Reconstructing the Past. Computer Applications and Quantitative Methods in Archaeology (CAA). Proceedings of the 36th International Conference. Budapest, April 2–6, 2008.* Budapest: Archaeolingua, pp. 152–155.

Wansleeben, M. and van den Dries, M.H. (2000) ArchWEB: a web site for Dutch archaeologists', in Lockyear, K., Sly, T.J.T., and Mihăilescu-Bîrliba, V. (eds) *CAA96. Computer Applications and Quantitative Methods in Archaeology*. Oxford: Archaeopress, pp. 71–80. Series: BAR International Series 845.