DOCUMENTING ARCHAEOLOGICAL WORK PROCESSES FOR ENABLING FUTURE REUSE OF DATA: THE CAPTURE PROJECT

by Isto Huvila (Department of ALM, Uppsala University/Sweden, isto.huvila@abm.uu.se)

Reusing archaeological data requires a comprehensive understanding what the data is about and as increasingly acknowledged, also of how the data came about (e.g. Voss, 2012; Faniel and Yakel, 2017). The European Research Council funded research project CAPTURE investigates the latter less studied issues: what information about the creation and use of research data is needed and how to capture enough of that information to make the data reusable in the future. The data about data creation and manipulation processes is conceptualized in the project in terms of paradata (Börjesson et al., 2020), a concept that is best known in the archaeological community in the context of heritage visualisation and the London and Seville Charters (2013) on their documentation.

The aim of CAPTURE is to develop an in-depth understanding of how paradata is being created and used today, to elicit and test methods for capturing paradata, and to synthesize the findings to inform the capturing of paradata and enabling data-intensive research using heterogeneous research data stemming from diverse origins.

Documenting enough, not documenting too much

The major problem of solving the issue of capturing and preserving paradata is that different data users have different needs in different contexts and situations. However, without a good-enough documentation of the human processes of creating, understanding and interpretation of data, there is a risk of what has been described as a digital dark age (Bollacker, 2010) and the proliferation of difficult to find and access "dark data" (Geser and Niccolucci, 2016). Even if data would not turn entirely unusable, collections of digital and non-digital archaeological data might remain too difficult to use to be really usable, incapable of supporting future research and cration of new archaeological knowledge and in some cases, even worse, leading researchers to conduct research on erroneous premises and drawing false conclusions on data that has been created under incompatible premises. By developing the understanding of paradata and process documentation, CAPTURE supports the implementation of European and global open data policies (Beck and Neylon, 2012; DCC, 2017; Kansa and Kansa, 2011), and effective sharing and reuse of data in disciplinary and cross-disciplinary knowledge ecosystems (cf. Bruseker et al., 2017) by developing critical knowledge of

the social context and use of infrastructures stressed in recent research agendas (e.g. ARIADNE Aloia et al., 2017 and JPI Lambourne et al., 2014) and empirical research (e.g. Mayernik et al., 2017) alike.x

The impossibility to document everything, diversity of needs and difficulty to anticipate them makes the documentation of data-related processes intricate. A major problem is how to determine how to document just enough. Similarly to all data on data (Mayernik and Acker, 2018), also this type of documentation is bound to be incomplete and as a consequence, it is important to focus on finding a reasonable balance between what can be captured automatically, should be documented manually (e.g. Stamatogiannakis et al., 2015), what is already embedded in the data itself (Huggett, 2012; Gant and Reilly, 2017), and what can be left to future users to figure out by themselves using post-hoc, forensic methods (Kirschenbaum et al., 2010) of 'excavating' existing data (Figure 1). So far a fair amount of research has investigated each of these approaches but it has tended to happen in relative isolation.
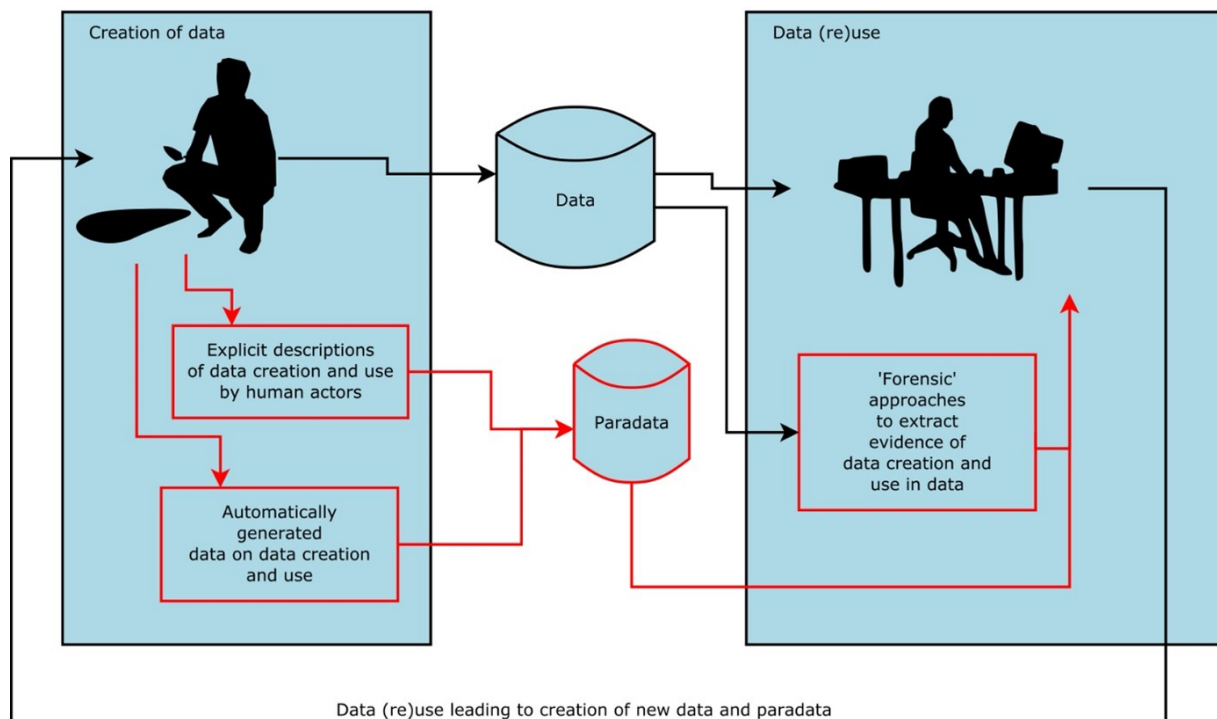

Figure 1: Sources of paradata in data creation and (re)use processes (data flow in black, paradata flow in red colour).

Survey on the views on archaeologists' data creation and reuse The CAPTURE project uses several different methods to investigate intellectual processes underpinning the creation and use of research data in archaeology and beyond and to propose and develop means to capture them. The methodological pallette consists of document and documentation studies, review and testing of earlier proposed and newly developed approaches for documenting paradata, and interviews with key stakeholders. At the moment, CAPTURE is also running a survey in collaboration with recently completed COST Action ARKWORK that is open

for everyone who has created or used archaeological research data in their work. The survey can be found in https://sunet.artologik.net/uu/Survey/271

Keeping updated on CAPTURE

Updates on CAPTURE project activities, including workshops, online CAPTURE talks with invited speakers, and outputs can be found on CAPTURE website at http://www.uu.se/en/research/capture and on Twitter at @CAPTURE_ERC.

Acknowledgements

References

- Aloia, N., , C. Binding, S. Cuy, M. Doerr, B. Fanini, A. Felicetti, J. Fihn, D. Gavrilis, G. Geser, H. Hollander, C. Meghini, F. Niccolucci, F. Nurra, C. Papatheodorou, J. Richards, P. Ronzino, R. Scopigno, M. Theodoridou, D. Tudhope, A. Vlachidis, and H. Wright. 2017. Enabling european archaeological research: The ARIADNE e-infrastructure. Internet Archaeology (43). URL http://dx.doi.org/10.11141/ia.43.11.
- Beck, A. and C. Neylon. 2012. A vision for Open Archaeology. World Archaeology 44 (4):479–497. URL http://dx.doi.org/10.1080/00438243.2012.737581.
- Bollacker, K. D. 2010. Avoiding a digital dark age: Data longevity depends on both the storage medium and the ability to decipher the information. American Scientist 98 (3):106–110. URL http://dx.doi.org/10.1511/2010.83.106.
- Börjesson, L., O. Sköld, and I. Huvila. 2020. The politics of paradata in documentation standards and recommendations for digital archaeological visualisations. Digital Culture and Society 6 (2):191–220. URL http://dx.doi.org/10.14361/dcs-2020-0210.
- Bruseker, G., M. Doerr, and M. Theodoridou. 2017. D5.1. Report on the Common Semantic Framework. PARTHENOS.
- Carrillo Gea, J. M., A. Toval, J. L. F. Alemán, J. Nicolás, and M. Flores. 2013. The London charter and the Seville principles as sources of requirements for e-archaeology systems development purposes. Virtual Archaeology Review 4 (9):205–211. URL http://dx.doi.org/10.4995/var.2013.4275.
- DCC. 2017. An Analysis of Open Data and Open Science Policies in Europe, May 2017. Apeldoorn: SPARC Europe & DCC.
- Faniel, I. M. and E. Yakel. 2017. Practices do not make perfect: Disciplinary data sharing and reuse practices and their implications

for repository data curation. In Curating Research Data, Volume One: Practical Strategies for Your Digital Repository, ed. L. R. Johnston, pp. 103–126. Chicago: ACRL.

- Gant, S. and P. Reilly. 2017. Different expressions of the same mode: a recent dialogue between archaeological and contemporary drawing practices. Journal of Visual Art Practice 17 (1):100–120. URL http://dx.doi.org/10.1080/14702029.2017.1384974.
- Geser, G. and F. Niccolucci. 2016. D2.4: Final Innovation Agenda and Action Plan. ARIADNE.
- Huggett, J. 2012. Promise and paradox: Accessing open data in archaeology. In Proceedings of the Digital Humanities Congress 2012, eds. C. Mills, M. Pidd, and E. Ward. Sheffield: Humanities Research Institute.
- Kansa, E. and S. Kansa. 2011. Toward a do-it-yourself cyberinfrastructure: Open data, incentives, and reducing costs and complexities of data sharing. In Archaeology 2.0: New Approaches to Communication and Collaboration, eds. E. C. Kansa, S. W. Kansa, and E. Watrall, pp. 57–91. Los Angeles, CA: Cotsen Institute of Archaeology, UC Los Angeles.
- Kirschenbaum, M. G., R. Ovenden, and G. Redwine. 2010. Digital Forensics and Born-Digital Content in Cultural Heritage Collections. Tech. rep., CLIR, Washington, D.C.
- Lambourne, G., L. Stoakes, M. Cassar, K. V. Balen, M. Rhisiart, M. Thomas, R. Miller, and L. Burnell. 2014. Strategic Research Agenda. Tech. rep., JPI Cultural Heritage and Global Change, Rome. URL http://www.jpi-culturalheritage.eu/wp-content/uploads/SRA-2014-06.pdf.
- Mayernik, M. S. and A. Acker. 2018. Tracing the traces: The critical role of metadata within networked communications. JASIST 69 (1):177–180. URL http://dx.doi.org/10.1002/asi.23927.
- Mayernik, M. S., D. L. Hart, K. E. Maull, and N. M. Weber. 2017. Assessing and tracing the outcomes and impact of research infrastructures. JASIST 68 (6):1341–1359. URL http://dx.doi.org/10.1002/asi.23721.
- Stamatogiannakis, M., P. Groth, and H. Bos. 2015. Looking inside the black-box: Capturing data provenance using dynamic instrumentation. In Provenance and Annotation of Data and Processes: 5th International Provenance and Annotation Workshop, IPAW 2014, Cologne, Germany, June 9-13, 2014. Revised Selected Papers, pp. 155–167. Cham: Springer. URL http://dx.doi.org/10.1007/978-3-319-16462-5_12.
- Voss, B. L. 2012. Curation as research. a case study in orphaned and underreported archaeological collections. Archaeological Dialogues 19 (2):145–169. URL http://dx.doi.org/10.1017/s1380203812000219.