



Dokumentation av arkeologiska arbetsprocesser underlättar framtida återanvändning av forsknings- och undersökningsdata: CAPTURE-projektet

Av Isto Huvila, Amalia Juneström, Jessica Kaiser, Olle Sköld, Institutionen för ABM, Uppsala universitet

Återanvändning av arkeologiska forsknings- och undersökningsdata förutsätter inte bara dataläskunnighet utan också en omfattande förståelse av hur data kommer till (t.ex. Voss, 2012; Faniel och Yakel, 2017).

Forskningsprojektet CAPTURE, finansierat av det europeiska forskningsrådet, undersöker den tidigare förhållandevis utforskade frågan om exakt vilken information om skapande, hantering och användning av forskningsdata som är nödvändig för att data skall vara återanvändbar i framtiden. Parallellt undersöker CAPTURE hur denna information kan fångas på ett både effektivt och tillräckligt omfattande sätt för att stödja dataåteranvändning. Data om dataskapande och manipulationsprocesser artikuleras i projektet som *paradata* (Börjesson et al., 2020). Begreppet är känt i det arkeologiska sammanhanget främst inom kulturarvsvisualisering genom London- och Sevilla-principerna (Carrillo & Gea, 2013), som stipulerar grundvalar för dokumentation av visuella (re)presentationer i arkeo-

logiska och kulturarvsrelaterade sammanhang.

Syftet med CAPTURE är att skapa ny kunskap och utveckla förståelse för hur paradata skapas och används idag samt att utveckla och testa metoder för att fånga paradata. Utifrån resultaten bidrar CAPTURE därmed till att skapa standarder och verktyg för paradata och framsteg i data-intensiva forskningsområden som använder heterogena forskningsdata med olika ursprung.

Dokumentera tillräckligt, inte för mycket

En utmaning för att fånga och bevara paradata är att olika dataanvändare har olika behov i olika situationer. Utan tillräcklig dokumentation av hur forskare och andra dataskapare och databrukare har skapat, förstått och tolkat data riskerar vi att hamna i en digital mörk tidsålder (Bollacker, 2010) där svåråtkomlig och svårfunnen ”mörk data” dominerar (Geser och Niccolucci, 2016). Bristande kontext kring skapandet av arkeologisk forskningsdata - både digital och

analog - kan i värsta fall leda till undermåliga och svårtolkade dataset som kanske inte stödjer framtida forskning och skapande av ny arkeologisk kunskap i tillräckligt stor utsträckning.

Den praktiska målsättningen i CAPTURE är att utveckla förståelsen för paradata och processdokumentation och därmed stödja implementeringen av nationella, europeiska och globala policyer för datahantering och öppen data (Beck och Neylon, 2012; DCC, 2017; Kansa och Kansa, 2011). Parallellt bidrar projektet till effektiv delning och återanvändning av data i vetenskapsdisciplinspecifika, tematiska och interdisciplinära kunskapsekosystem och repositorer (jfr Bruseker et al., 2017). Projektet utvecklar en kritisk förståelse för de sociala kontexter och användningen av infrastrukturer som betonats i nyare forskningsagendor (t.ex. ARIADNE Aloia et al., 2017 och Lambourne et al., 2014) och empirisk forskning (t.ex. Mayernik et al., 2017). Det skapar också ny kunskap om vad dataskapare och -användare finner viktigt att dokumentera om data-relaterade processer, vilka explicita och implicita behov av dokumentation det finns och hur de kan tillfredsställas i praktiken.

Det är praktiskt omöjligt att dokumentera allt. Mångfalden av behov och svårigheten att förutse vilka behov som finns gör det komplicerat att dokumentera datarelaterade processer. Ett stort problem är att avgöra hur man dokumenterar och bevarar precis tillräckligt. I likhet med all data om data (Mayernik och Acker, 2018) kommer även paradata att bli ofullständig. Som en konsekvens är det viktigt att fokusera på att hitta en rimlig balans mellan vad som kan fångas automatiskt och vad som bör dokumenteras manuellt (t.ex. Stamatogiannakis et al., 2015). Det är alltså viktigt att undersöka vilken information som redan finns inbäddad i själva datan (Huggett, 2012; Gant och Reilly, 2017) och vad som kan överlåtas till framtida användare att själva ta reda på med hjälp av olika typer av icke-bokstavliga ”arkeologiska” eller

”forensiska” post-hoc metoder (Kirschenbaum et al., 2010) om att ”gräva ut” befintliga data. Hittills har en hel del forskning undersökt vart och ett av dessa tillvägagångssätt men det har saknats forskning som täcker hela paradatafenomenet och hur det kan användas till att stödja återanvändning av forsknings- och undersökningsdata.

Pågående studier om arkeologers dataskapande och återanvändning

CAPTURE-projektet använder flera olika metoder för att undersöka intellektuella processer som ligger till grund för skapandet och användningen av forskningsdata inom och utanför arkeologin samt för att föreslå och utveckla metoder för att fånga dem. Metodpaletten består av dokument- och dokumentationsstudier (t. ex. Huvila et al., 2021; Huvila et al., 2022a; Börjesson et al., 2022a), begrepps- (Sköld et al., 2022) och citeringsanalys (Huvila et al., 2022b), etnografi, granskning och testning av tidigare föreslagna och nyutvecklade metoder för att dokumentera paradata samt intervjuer (t. ex. Börjesson et al., 2022b) och fokusgruppdiskussioner med nyckelintressenter.

Projektet och dess tillvägagångssätt är mångvetenskapligt. Medverkande forskares bakgrund ligger bland annat i arkivvetenskap, informationsvetenskap, arkeologi och vetenskapsstudier. Mångfalden i projektgruppen möjliggör forskning om paradata och processdokumentation både inifrån och utifrån arkeologin.

Mycket paradata dokumenteras redan i nuläget

Resultaten från intervju- och enkätstudier, forskningspublikationer och data visar att mycket paradata redan finns tillgänglig på olika platser. Inom arkeologi är undersökningsrapporter en viktig källa för paradata i och med att de förvän-

tas dokumentera både resultat och undersökningsprocess. Förutom i regelrätta arbetsbeskrivningar finns kunskap om arbetsprocesser, till exempel i resultatbeskrivningar och i information om medverkande aktörer. En viktig källa är också fotografier, särskilt sådana som visar pågående arbete och omständigheter på fyndplatser (Huvila et al., 2021). Även närläsning av dataset kan bidra med information om bakomliggande processer. Ordval, beskrivningar och tidsstämplar är bara några exempel på element i databaser som innehåller processkunskap (Börjesson et al., 2022a).

Att mycket paradata redan finns i befintlig dokumentation betyder att den primära utmaningen med processdokumentationen inte nödvändigtvis är att utöka dess mängd eller omfång. Ett av problemen är att paradata finns utspridd och att det kan vara svårt att få ett helhetsgrepp på vilken paradata som finns att tillgå. Viktigare är att hitta paradata, förstå vad som saknas och komplettera den med nödvändig information.

Att dokumentera rätt

Paradata är inte heller alltid tillgänglig och det är inte heller alltid så att tillgänglig paradata motsvarar användarnas behov. Särskilt information om informationshantering, standarder och strukturering av data är sällan dokumenterad i detalj. Därtill visar resultaten att dataskapare och -användare ofta har olika syn på vilken paradata som behövs (Börjesson et al., 2022b). När paradata dokumenteras är det naturligt att fokusera på element som är uppenbara för dataskaparen och följer dennes föreställningar om vad som är centralt i dataskapandet samt att de är lätta att dokumentera. Dataanvändare däremot förväntar sig och har behov av paradata som hjälper dem att förstå data utifrån sina behov.

Utmaningen med det uppenbara glappet mellan vad dataskapare och -användare anser är viktigt är att ändå kunna skapa och tillhand-

hålla sådan paradata som är meningsfull för båda parter. Användarna behöver insikter i hur dataskapandet har gått till och kompetens att förstå hur dataskapandeprocessen fungerar. Samtidigt är det viktigt att dataskapare förstår hur användare tänker, vilken paradata som förmodligen kan vara behjälplig och att de har dataanvändning i åtanke när de skapar och dokumenterar data.

Paradata som data och perspektiv

CAPTURE-projektet har också undersökt hur begreppet paradata används och förstås i olika domäner med syfte att bidra till dess utveckling som ett teoretiskt och praktiskt koncept (Sköld et al., 2022). Det är uppenbart att ”paradata” används på olika sätt i olika kontexter. Därför är det viktigt att tydliggöra vad som konkret åsyftas när begreppet används. I en konkret bemärkelse kan paradata förstås som en särskild typ av data som skiljer sig från både data och metadata. Begreppet kan också förstås i en teoretisk mening som ett sätt att begreppsliggöra olika former av data eller information som kan bidra med kunskap om olika typer av processer och praktiker.

Det är sannolikt att en viss mångfald kommer att bestå även i fortsättningen. Därför är det viktigt att tydliggöra vad som avses med paradata när begreppet används i både teoretiska och praktiska sammanhang. Lika nyttigt kan vara att behålla paradata som ett samtidigt teoretiskt och praktiskt begrepp som å ena sidan hänvisar till en viss konkret typ av data eller dokumentation och å andra sidan innebär ett perspektiv som är behjälplig i konceptualiseringen av paradata, paradatabehov och processförståelse som fenomen.

Håll dig uppdaterad om CAPTURE

Uppdateringar om CAPTURE-projektets aktiviteter, inklusive workshops, online-CAPTURE-seminarier med inbjudna talare och forskningsresultat finns på CAPTURE:s webbplats (<http://www.uu.se/en/research/capture>) och X (Twitter) (@CAPTURE_ERC).

Acknowledgements

Capturing Paradata för att dokumentera dataskapande och användning för framtidens forskningsprojekt (CAPTURE) har fått finansiering från Europeiska forskningsrådet (ERC).



Litteratur

- Aloia, N., Debole, F., Felicetti, A., Galluccio, I., & Theodoridou, M. (2017). Mapping the ARIADNE catalog data model to CIDOC CRM: Bridging resource discovery and item-level access. *Ricerca Scientifica e Tecnologie Dell'Informazione*, 7(1), 1–8. <https://doi.org/10.2423/i22394303v7n1p1>
- Beck, A., & Neylon, C. (2012). A vision for Open Archaeology. *World Archaeology*, 44(4), 479–497. <https://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. <https://doi.org/10.1511/2010.83.106>
- Börjesson, L., Sköld, O., & Huvila, I. (2020). The politics of paradata in documentation standards and recommendations for digital archaeological visualisations. *Digital Culture and Society*, 6(2), 191–220. <https://doi.org/10.14361/dcs-2020-0210>
- Börjesson, L., Sköld, O., Friberg, Z., Löwenborg, D., Pålsson, G., & Huvila, I. (2022a). Re-purposing Excavation Database Content as Paradata: An Explorative Analysis of Paradata Identification Challenges and Opportunities. *KULA: Knowledge Creation, Dissemination, and Preservation Studies*, 6(3), Article 3. <https://doi.org/10.18357/kula.221>
- Börjesson, L., Huvila, I., & Sköld, O. (2022b). Information needs on research data creation. *Information Research*, 27(Special Issue), isic2208. <https://doi.org/10.47989/irisic2208>
- Carrillo Gea, J. M., Toval, A., Alemán, J. L. F., Nicolás, J., & Flores, M. (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. <https://doi.org/10.4995/var.2013.4275>
- Beck, A., & Neylon, C. (2012). A vision for Open Archaeology. *World Archaeology*, 44(4), 479–497. <https://doi.org/10.1080/00438243.2012.737581>
- DCC. (2017). *An analysis of open data and open science policies in Europe, May 2017*. SPARC Europe & DCC.
- Faniel, I. M., & Yakel, E. (2017). Practices do not make perfect: Disciplinary data sharing and reuse practices and their implications for repository data curation. In L. R. Johnston (Ed.), *Cu-*

rating research data, volume one: Practical strategies for your digital repository (pp. 103–126). ACRL.

Gant, S., & Reilly, P. (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. <https://doi.org/10.1080/14702029.2017.1384974>

Geser, G., & Niccolucci, F. (2016). *D2.4: Final innovation agenda and action plan*. ARIADNE.

Huggett, J. (2012). Promise and paradox: Accessing open data in archaeology. In C. Mills, M. Pidd, & E. Ward (Eds.), *Proceedings of the digital humanities congress 2012, Sheffield, 6–8th September 2012*. Humanities Research Institute.

Huvila, I., Sköld, O., & Börjesson, L. (2021). Documenting information making in archaeological field reports. *Journal of Documentation*, 77(5), 1107–1127. <https://doi.org/10.1108/JD-11-2020-0188>

Huvila, I., Börjesson, L., & Sköld, O. (2022a). Archaeological information-making activities according to field reports. *Library & Information Science Research*, 44(3), 101171. <https://doi.org/10.1016/j.lisr.2022.101171>

Huvila, I., Andersson, L., & Sköld, O. (2022b). Citing methods literature: Citations to field manuals as paradata on archaeological fieldwork. *Information Research*, 27(3), paper941. <https://doi.org/10.47989/irpaper941>

Kansa, E., & Kansa, S. W. (2011). Toward a do-it-yourself cyberinfrastructure: Open data, incentives, and reducing costs and complexities of data sharing. In E. C. Kansa, S. W. Kansa, & E. Watrall (Eds.), *Archaeology 2.0: New Approaches to Communication and Collaboration* (pp. 57–91). Cotsen Institute of Archaeology, UC

Los Angeles.

Kirschenbaum, M. G., Ovenden, R., & Redwine, G. (2010). *Digital Forensics and Born-Digital Content in Cultural Heritage Collections*. CLIR.

Lambourne, G., Stoakes, L., Cassar, M., Balen, K. V., Rhisiart, M., Thomas, M., Miller, R., & Burnell, L. (2014). *Strategic Research Agenda*. JPI Cultural Heritage and Global Change. <http://www.jpi-culturalheritage.eu/wp-content/uploads/SRA-2014-06.pdf>

Mayernik, M. S., Hart, D. L., Maull, K. E., & Weber, N. M. (2017). Assessing and tracing the outcomes and impact of research infrastructures. *JASIST*, 68(6), 1341–1359.

Mayernik, M. S., & Acker, A. (2018). Tracing the traces: The critical role of metadata within networked communications. *JASIST*, 69(1), 177–180. <https://doi.org/10.1002/asi.23927>

Sköld, O., Börjesson, L., & Huvila, I. (2022). Interrogating paradata. *Information Research. Proceedings of the 11th International Conference on Conceptions of Library and Information Science, Oslo Metropolitan University, May 29 - June 1, 2022*, 27(Special Issue), paper colis2206. <https://doi.org/10.47989/ircolis2206>

Stamatogiannakis, M., Groth, P., & Bos, H. (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). Springer. https://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. <https://doi.org/10.1017/s1380203812000219>