Deep Data Example: Zbiva, Early Medieval Data Set for the Eastern Alps

Archaeology

Benjamin Štular | ORCID: 0000-0003-1474-7183
Corresponding author
Znanstvenoraziskovalni center Slovenske akademije
znanosti in umetnosti, Ljubljana, Slovenia
benjamin.stular@zrc-sazu.si

Mateja Belak
Znanstvenoraziskovalni center Slovenske akademije
znanosti in umetnosti, Ljubljana, Slovenia
mateja.belak@zrc-sazu.si

Abstract

Zbiva is an open access online research data base for the archaeology of the Eastern Alps in the Early Middle Ages. The data base is the product of four decades of thoughtful digital curation and is continually evolving at the data record level. As such, it is best described by the concept of Deep Data. The authors deposited a subset of the Zbiva data base in a persistent open access repository, Zenodo. This was necessary to ensure stable reference, facilitate the reproducibility of the results, and promote data reuse in their ongoing publication efforts. The deposited data cover the period from 500 to 1000 CE and are spatially restricted to present-day Slovenia, southern Austria, and a small part of north-eastern Italy. The data set is particularly suitable for archaeological GIS analyses.

Keywords

archaeology – digital archaeology – digital humanities – early medieval period – early middle ages – Eastern Alps – Deep Data – Slow Data

–   Related data set "Zbiva" with DOI www.doi.org/10.5281/zenodo.5761811
     in repository "Zenodo"

1.   Introduction

Zbiva is an open access online research data base for the archaeology of the
Eastern Alps in the Early Middle Ages (Pleterski, 2016). Currently it contains
data on 3,833 archaeological sites, 3,428 graves, 15,777 artefacts, and 11,596 asso-
ciated bibliographic units in more than half a million data base fields. As far
as the authors are aware, Zbiva is unmatched in Slavic archaeology. The only
comparable data set for Early Medieval archaeology is OpenAtlas (Filzwieser
& Eichert, 2020) with its affiliate, THANADOS (Eichert, 2021).
     Zbiva has been actively developed since 1987 and its inception was deeply
rooted in the scientific context of the time (e.g., Štular & Pleterski, 2018). It was
conceived for the study of the so-called Carantanian-Köttlach archaeological
culture. This means that its chronological focus was on the period from the
settlement of Slavs (as perceived in the 1980s) in the 6th century CE to the end
of the habitual deposition of grave goods in the 11th century CE. It contained
mainly data from the settlement area of the Alpine Slavs (as perceived in the
1980s), which includes present-day Slovenia, Austria, NW Croatia, and NE Italy.
For comparative purposes, selected relevant sites from neighbouring regions
and from prehistoric times were also included (Pleterski & Belak, 1995).
     The main strength of Zbiva is that a small but dedicated team has contin-
uously curated it since its launch by collecting high-value reference data sets
and fine-tuning them to enhance scholarly output, for example, by regularly
scouring the relevant bibliography. Access to the latter is based on the sys-
tematically built and maintained network for the exchange of publications
between the host institution and all major relevant institutions contributing to
the subject (printed publications remain the main source of new information
on this particular topic). On average, 40 to 50 person days is spent on this task
per year. The existing data set is thus the result of four decades of deliberate
scholarly work and attentive curation, which is indeed a "thoughtful digital
curation" (Kansa, 2016).
     The Zbiva data base consists of four parts: archaeological sites, graves, arte-
facts, and bibliography. The initial Zbiva 1.x, developed in 1987, was a closed
system based on a single PC. Zbiva 2.x was developed in 2000 to facilitate the
transition to the Web. However, due to technical limitations, only sites and
bibliography (under the name Libera) were made available online at that
time. This makes Zbiva one of the oldest open access online archaeological

data bases in the region. In 2016, the front end was migrated to the Zbiva 3.x web application (Pleterski, 2016), which is based on the open source Arches 3.0 platform. It is a full-featured web-GIS application, and all content is searchable via either full text or structured or map-based search. The Zbiva 3.x web application was designed with highly motivated users in mind, who want to search not only for structured information, but also for "hidden" knowledge. The tools were therefore optimised for efficiency rather than ease of use. Zbiva is available in Slovenian, English, and German languages, but free text descriptions are only available in Slovenian (Pleterski, 2016; Štular, 2019, 2021). Zbiva 3.x will be available as long as the system can be technically maintained. Currently (2022), the entire back-end and front-end is being migrated to a customised Zbiva 4.x, which was designed from the beginning as an online and open access system.

Recent discoveries in the field of archaeology of the Slavs (Pavlovič, 2017; Pavlovič et al., 2021) made it necessary to extend the chronological context of Zbiva to include the fifth century AD. Additionally, the data analysis of Zbiva 3.x revealed that the chronology of the archaeological sites is neither consistent nor accurate enough for a regional study (Štular, 2019). Consequently, a concerted effort has been made to expand the content of Zbiva to include fifth-century sites and, more importantly, to enrich the data base with state of the art information on chronology.

The enriched data are currently being analysed, and the results are in the process of being published (Štular et al. [under review]). The subset of data directly relevant to these publications has been deposited at Zenodo, the open research repository of OpenAIRE and CERN (Štular et al., 2021). This was necessary for three reasons. First, to ensure stable referencing in scientific publications; second, to facilitate reproducibility of results; third, we hope that this will encourage the reuse of data.

The focus of this article is on the aforementioned data subset deposited at Zenodo. The data set is described in the "Methods" and "Data" sections. In the "Problem" section, we discuss the nature of the Zbiva data set, which we refer to as Deep Data.

2.   Problem

The FAIR Guiding Principles (Findable, Accessible, Interoperable, Reusable) for the scientific data management and stewardship have become a de facto standard in the research world, including archaeology. The main challenge of data-intensive science targeted by these principles is to improve knowledge

discovery from scientific data and other scholarly digital objects. This is achieved by assisting both humans and their computational agents to discover, access, and analyse these data (Wilkinson et al., 2016). However, implementation is very complex and compliance should not be considered a strict protocol, but rather a desirable goal (Dunning et al., 2017). Furthermore, very little is known about whether data is re-usable and by whom (Wright & Richards, 2018).

The FAIR Data Principles were developed primarily with data-intensive science, and thus Big Data, in mind. However, in archaeology, so-called Deep Data is much more common and is of interest in this article. The term Deep Data describes data that is not very big, but semantically very rich, i.e., it provides contextually rich information intended to provide a rich user experience. The term Deep Data was proposed by analogy with the term Deep Web, which is used to describe the vast amount of information that resides beneath the surface of web pages (Szczuka & Ślęzak, 2013). The term Deep Data has also been used to describe an approach where all the information available in the data is fully exploited to gain knowledge (Belianinov et al., 2015).

A similar concept is Slow Data. Slow Data is primarily seen as a challenge to the evolving narratives about Big Data. The term usually refers to the speed of data transfer, data collection, and the like (which gives it a negative connotation). In the context of the broader "Slow" approach, Slow Data is a resistance to the displacement of knowledge by the pursuit of information in the form of more and more data. At the same time, Slow Data is directed against the belief in data as information rather than as elementary building blocks from which information is derived. Its central concern, therefore, is to emphasise the creative act that is the creation and subsequent manipulation of data (Huggett, 2022). In this sense, and viewed through the prism of DIKW (Data-Information-Knowledge-Wisdom; Ackoff, 1989), one could say that a data base containing Slow Data, a "Slow Data-base", is indeed an "information-base", or to some extent even a "knowledge-base".

We fully endorse Huggett's reasoning. However, for the purposes of this article we consider the terms Deep Data and Slow Data to be interchangeable and we prefer the term Deep Data because of its positive connotation. Also, for the purposes of this article, we define Deep Data as a data set that (i) may not be big but is semantically very rich, (ii) is a result of thoughtful (long-term) digital curation, and (iii) is founded on the belief that data are merely elementary building blocks from which information and possibly knowledge are derived. Deep Data can also be (iv) inherently transient.

Central to this article is that Deep Data are complex, contextual, and often subjective. Consequently, their digital curation is very intricate. This intricacy

is best explained in terms of archaeological chronology. The chronology of the sites is key information for archaeological research. Behind the seemingly simple data record (e.g., from 800 to 900 CE) lies a complex, contextual, and often subjective decision-making process. Typically, an entire scholarly article is devoted to determining the chronology of a single archaeological site or even a single archaeological artefact. Furthermore, the result is transient, as new relevant data or new knowledge may lead to a change.

Let us consider a hypothetical example. With archaeological analysis of artefacts from 20 graves a cemetery (archaeological site X) was dated to period between 800 and 900 CE. This information was entered in Zbiva. However, subsequently 20 additional graves were excavated at the same cemetery, which were dated to the period between 750 CE and 850 CE. Therefore, the chronology of the entire cemetery with 40 graves (archaeological site X) was changed to be from 750 to 900 CE.

Chronology can also be altered by an even more subtle factor, namely new knowledge. For example, new knowledge in the form of scientific publication provides different dates for the artifacts from the existing 20 graves, so the "from" date is changed. In the case of the Zbiva data base, this type of change is the norm.

The described process is different from, say, correcting or deleting corrupted data caused by faulty measuring equipment. In our example, both "from 800" and "from 750" are correct, each based on the data and/or knowledge available at a given time. From a data management perspective, one could say that the process of publishing Deep Data is ongoing and never final, as there are constant changes at the record level.

3. Methods

As mentioned above, new discoveries made it necessary to expand and enrich the chronological data in Zbiva. As part of a dedicated research project (see Acknowledgements) the data enrichment was focused on a geographically limited subset of data: 1,105 archaeological sites located in present day Slovenia, southern Austria (Carinthia, Styria, East Tyrol, parts of Salzburg and Upper Austria) and a small part of northern Italy (the Trieste region) (see Figure 1). Graves and artefacts were not part of this endeavour and are therefore not included in the data subset discussed in this article.

To improve the accuracy of chronology each site was re-examined by an expert using modern typochronologies based on C14 data (Pleterski, 2010a, 2010b, 2013). New chronology is therefore an expert-based knowledge, rather
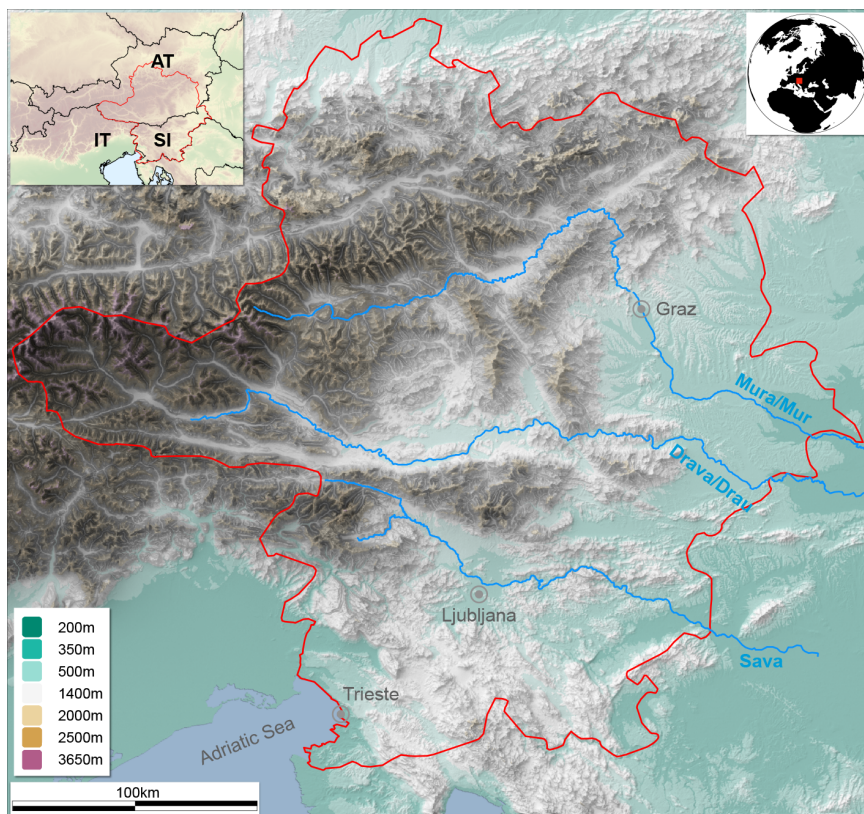
Note: Marked in red; upper left corner Lat. 48.22015, Lon. 12.35667; lower right corner Lat. 45.29785, Lon. 16.41784.

FIGURE 1   Map of the regional extent of the data

than data. The final data set is well suited for the study of the half-millennium between 500 and 1000 CE, which is the stated aim of the Zbiva data base (see Figure 2).

At the same time, the accuracy of location data was improved using maps (historical and modern) and satellite imagery available through freely accessible web GIS applications. Additionally, the data set was enriched with metadata (e.g., the confidence level for chronology and location) and paradata (e.g., sources for dating).

Six archaeologists (domain experts) were involved in this process of knowledge production. In total, they spent about 24 person-months on this task.

Time Series Clustering - Profile (Fourier)
— EMA1   — LA   — EMA2   ⋯⋯ SUM
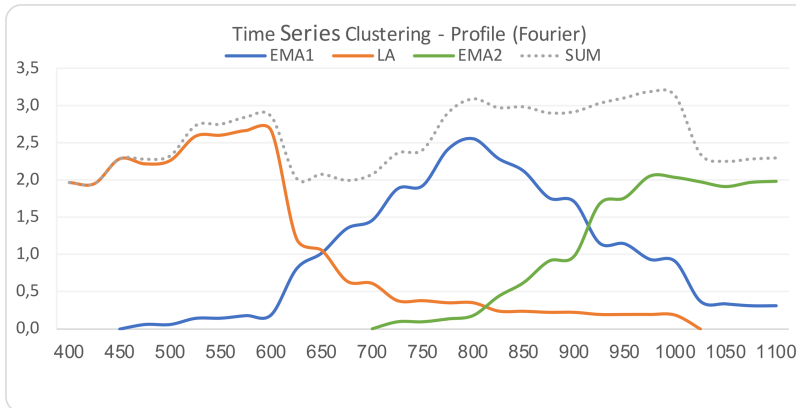
Note: (n = 1,105). Archaeological periodization with time series clustering: LA – Period 1, Late Antiquity; EMA1 – Period 2, Early Middle Ages 1; EMA2 – Period 3, Early Middle Ages 2; SUM – sum of all values.

FIGURE 2   Zbiva data subset

4.   Data

– **Zbiva, Early Medieval Data Set for the Eastern Alps (data sub-set), deposited at Zenodo – doi:**www.doi.org/10.5281/zenodo.5761811
– **Temporal coverage:** 500–1000 CE

Below are the general characteristics of the data set deposited in Zenodo (Štular et al., 2021) in table form (see Table 1). A more detailed description of the data set and technical details, i.e. rich metadata, can be found in the deposited data set documentation.

The dataset consists of categories based on the particular needs of Early Medieval archaeology. Perhaps worth mentioning is the inclusion of confidence metadata (Loconf, Dateconf, Dataconf), which is not common in archaeological data bases. However, given the characteristics of Deep Data, we believe this metadata is extremely important for archaeological analysis.

The data set is deposited as a spreadsheet that can be used directly for many different types of analysis. The format is particularly well suited to GIS analysis. For example, the data can be used for space-time pattern mining following the published GIS protocol (Štular & Lozić, 2022).

5.   Concluding Remarks

We presented an open access online research data base for the archaeology of the Eastern Alps in the Early Middle Ages, Zbiva. The article focuses on the

TABLE 1   Description of data fields that were deposited to Zenodo

Field name	Description	Values
ID	Unique identifier	E.g. 10000052
Lon	Longitude (WGS84), decimal	E.g. 15,866800
Lat	Latitude (WGS84), decimal	E.g. 46,420900
Loconf	Location confidence	1 – location of the nearest settlement; 2 – location in part of settlement or descriptive location; 3 – exact, e.g. geodetic measurements
Name	Published name of the site (in the language of the country of origin, i.e. Slovenian, Italian, German or Croatian)	E.g. Štuki – Marof
Datefrom	The earliest possible date (year CE)	E.g. 640
Dateuntil	The latest possible date (year CE)	E.g. 800
Datetype	Primary source of dating information	0 – no data; 1 – radiometric method, e.g. C14; 2 – typology of artifacts; 3 – after peer reviewed publication
Dateconf	Dating confidence, higher is better	0 – no data; 1 – old publication; 2 – typology by a Zbiva expert, few finds; 3 – full peer reviewed publication of the site according to modern scholarly standards or typology by a Zbiva expert based on sufficient quantity and quality of artefacts, e.g., several grave contexts
Position	Position in the landscape	0 – no data; 1 – plain/flatland; 2 – rise/hill; 3 – terrace/ridge; 4 – in water; 5 – in cave

TABLE 1   Description of data fields that were deposited to Zenodo (*cont.*)

Field name	Description	Values
Sitetype	Type of the site	1 – burial site; 2 – stray find (no context); 3 – cult site/church; 4 – castle/fort; 5 – settlement; 6 – hoard; 7 – other; 8 – urban settlement
Settsize	Settlement size	0 – no data; 1 – small, e.g. several buildings; 2 – medium, e.g. around 10 buildings; 3 – large, e.g. many buildings of different types and functions.
Settype	Settlement type	0 – no data; 1 – predominantly agrarian; 2 – specialised, e.g. iron production; 3 – central settlement, e.g. proto urban with non agrarian functions; 4 – urban settlement
Structyp	Built structure type	0 – no data or unidentifiable; 1 – stone built; 2 – sunken-floor/ pit houses; 3 – ground or above ground level (e.g. wooden cottage)
Cemetsiz	Cemetery size, quantitative values	0 – no data; 1 – up to 10 graves; 4 – 10–60 graves; 8 – 60–150 graves; 20 – over 150 graves
Cemetpos	Position of the cemetery in the archaeological landscape	0 – no data; 1 – adjacent to/in church; 2 – without church; 3 – with non-Christian-church cult structures; 4 – in settlement
Cemetype	Cemetery type	0 – no data; 1 – flat; 2 – burial mound(s); 3 – mixed flat and burial mound(s)
Cemethll	Is cemetery located on a hill?	0 – no data or no; 1 – yes
Burtype	Burial type	0 – no data; 1 – skeletal; 2 – mixed cremation and skeletal; 3 – cremation

TABLE 1   Description of data fields that were deposited to Zenodo (*cont.*)

Field name	Description	Values
Datatype	Type of information	0 – no data; 1 – archaeological record; 2 – written source; 3 – standing building (remains); 4 – oral tradition
Dataconf	Data confidence, higher is better	1 – stray find/no context; 2 – neither 1 nor 3; 3 – peer reviewed publication of the entire site

data subset of Zbiva deposited in the Zenodo repository. Our main goal was to support the publication of our data analysis. We deposited only the data directly relevant to our current research in a spreadsheet format that can be used for many different types of analysis. The entire semantically rich data set is accessible on Zbiva. By data deposition (including rich metadata) we believe we have significantly improved the FAIRness of the data, which we hope will encourage re-use of the data.

However, we believe that this article has broader implications that are relevant not only to archaeology, but also to digital humanities in general. We have introduced the concept of Deep Data. Although this concept is by no means new, it has not been used in archaeology or, to our knowledge, in the digital humanities. By use, we mean that the concept is integrated into all phases of research, from funding and data collection to analysis and publication. Deep Data approach can produce many different rich and dense stories and create narratives about the data itself because it is less intentional and more intuitive (*cf.* Lupi, 2017; Strauss, 2018).

Acknowledgements

Funding

References

Ackoff, R. L. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, *16*(1), 3–9.

Belianinov, A., Vasudevan, R., Strelcov, E., Steed, C., Yang, S. M., Tselev, A., Jesse, S., Biegalski, M., Shipman, G., Symons, C., Borisevich, A., Archibald, R., & Kalinin, S. (2015). Big data and deep data in scanning and electron microscopies: Deriving functionality from multidimensional data sets. *Advanced Structural and Chemical Imaging*, *1*(1). www.doi.org/10.1186/s40679-015-0006-6.

Dunning, A., Smaele, M. D., & Böhmer, J. (2017). Are the FAIR Data Principles fair? *International Journal of Digital Curation*, *12*(2), 177–195. www.doi.org/10.2218/ijdc.v12i2.567.

Eichert, S. (2021). Digital mapping of medieval cemeteries: Case studies from Austria and Czechia. *Journal on Computing and Cultural Heritage*, *14*(1), 1–15. www.doi.org/10.1145/3406535.

Filzwieser, R., & Eichert, S. (2020). Towards an online database for archaeological landscapes. Using the web based, open source software OpenAtlas for the acquisition, analysis and dissemination of archaeological and historical data on a landscape basis. *Heritage*, *3*, 1385–1401. www.doi.org/10.3390/heritage3040077.

Huggett, J. (2022). Is less more? Slow data and datafication in archaeology. In K. Garstki (Ed.), *Critical archaeology in the digital age* (pp. 156–184). Cotsen Institute of Archaeology, Los Angeles, CA.

Kansa, E. C. (2016). Click here to save the past. In E. Averett, J. Gordon, & D. Counts (Eds.), *Mobilizing the past for a digital future: The potential of digital archaeology* (pp. 443–472). Digital Press at the University of North Dakota, Grand Forks ND.

Lupi, G. (2017). Data humanism, the revolution will be visualized. *PrintMag*, *30 January*. www.printmag.com/information-design/data-humanism-future-of-data-visualization/.

Pavlovič, D. (2017). Začetki zgodnjeslovanske poselitve Prekmurja = Beginnings of the Early Slavic settlement in the Prekmurje region, Slovenia. *Arheološki Vestnik*, *68*, 349–386.

Pavlovič, D., Vojakovič, P., & Toškan, B. (2021). Cerklje ob Krki: Novosti v poselitvi Dolenjske v zgodnjem srednjem veku. *Arheološki Vestnik*, *72*, 137–186. www.doi.org/10.3986/av.72.06.

Pleterski, A. (2010a). Datiranje zgodnjesrednjeveške naselbine Lehen pri Mitterkirchnu v Zgornji Avstriji kot kontrola nove datacijske metode s pomočjo referenčne tabele in korelacijske formule ustij loncev. *Vjesnik Arheoloskog Muzeja u Zagrebu* (*Annual Journal of the Museum*), *43*, 309–324.

Pleterski, A. (2010b). *Zgodnjesrednjeveška naselbina na blejski Pristavi: Tafonomija, predmeti in čas*. (Opera Instituti Archaeologici Sloveniae Vol. 19). Založba ZRC, Ljubljana.

Pleterski, A. (2013). Korak v kronologijo zgodnjesrednjeveškega naglavnega nakita vzhodnih Alp = A step towards the chronology of early medieval head ornaments in the Eastern Alps. *Arheološki Vestnik*, *64*, 299–334.

Pleterski, A. (2016). *Zbiva v3.08*. Research Centre of Slovenian Academy of Sciences and Arts, Institute of Archeology. http://zbiva.zrc-sazu.si.

Pleterski, A., & Belak, M. (1995). Zbiva. Cerkve v vzhodnih Alpah od 8. do 10. stoletja (Zbiva. Archäologische Datenbank für den Ostalpenbereich. Die Kirchen in den Ostalpen vom 8. bis 10. Jahrhundert). *Zgodovinski časopis*, *49*(1), 19–43.

Strauss, C. (2018). All in good time. In T. Lijster (Ed.), *The future of the new: Artistic innovation in times of social acceleration* (pp. 55–68). Antennae-Arts in Society No. 26. Valiz.

Štular, B. (2019). The Zbiva web application: a tool for early medieval archaeology of the Eastern Alps. In J. D. Richards & F. Niccolucci (Eds.), *The ARIADNE Impact* (pp. 69–82). Archaeolingua. www.doi.org/10.5281/zenodo.3476712.

Štular, B. (2021). Archiving of archaeological digital datasets in Slovenia: historic context and current practice. *Internet Archaeology*, *58*. www.doi.org/10.11141/ia.58.17.

Štular, B., & Lozić, E. (2022). *GIS protocol for multy-scale emerging hot spot analysis* (*1.0*) [Data set]. Zenodo. www.doi.org/10.5281/zenodo.5813527.

Štular, B., Lozić, E., Belak, M., Rihter, J., Koch, I., Modrijan, Z., Magdič, A., Karl, S., Lehner, M., & Gutjahr, C. (under review). Migration of Alpine Slavs and machine learning: Space-time pattern mining of an early medieval data set from the Eastern Alps. *PLOS ONE*.

Štular, B., & Pleterski, A. (2018). Prologue. Early medieval archaeology in the South eastern Alpine area: Past, present, future. In J. Lux, B. Štular, & Zanier, K. (Eds). *Our heritage: the Slavs*. Zavod za varstvo kulturne dediščine Slovenije.

Štular, B., Pleterski, A., & Belak, M. (2021). *Zbiva, Early medieval data set for the Eastern Alps. Data sub-set v.1.0* (1.0) [Data set]. Zenodo. www.doi.org/10.5281/zenodo.5761811.

Szczuka, M., & Ślęzak, D. (2013). *How deep data becomes big data*. In *2013 Joint IFSA World Congress and NAFIPS Annual Meeting* (*IFSA/NAFIPS*) (pp. 579–584).

Wilkinson, M. D., Dumontier, M., Aalbersberg, IJ. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1). www.doi.org/10.1038/sdata.2016.18.

Wright, H., & Richards, J. D. (2018). Reflections on collaborative archaeology and large-scale online research infrastructures. *Journal of Field Archaeology*, *43*(sup1), S60–S67. www.doi.org/10.1080/00934690.2018.1511960.