Archeologia e Calcolatori
35.2, 2024, 47-54
doi 10.19282/ac.35.2.2024.06

DOES NOT COMPUTE! ACCOUNTING FOR SOCIAL AND SITUATED EXPERIENCES WHILE INTEGRATING AND ANALYZING OBSIDIAN SOURCING DATA

1. Data sharing in practical context

Open data platforms, which host published datasets on the web, boast several potential benefits, including the ability to extract untapped value from older 'legacy' data, to integrate data collected at various archaeological sites, to validate analytical claims more effectively, and to ensure that the general publics who fund archaeological research are able to access freely and make sense of the fruits of their investments (Kansa 2014; Kintigh *et al.* 2014). However, as the open data movement matures and data-sharing platforms become integral parts of the publishing process, archaeologists are becoming more aware of their limitations. For instance, J. Huggett (2018) reports on how open data platforms have generally failed to encourage reuse, and R. Opitz *et al.* (2021) demonstrate how effective data integration and synthesis is primarily derived from preexisting collaborative relationships. Furthermore, Atici *et al.* (2013), Faniel *et al.* (2013) and Kansa *et al.* (2013) show that archaeologists who re-use data seek out additional information that could otherwise only be obtained through close collaborative relationships. These concerns, raised in response to the initial implementation of open data norms and practices, hint at a growing recognition that data are not asocial, apolitical, and disembodied entities, and that effective data re-use involves reconciliation with the prior decisions and circumstances that informed the original constitution of a dataset (Dallas 2015; Huggett 2022; Batist 2024).

While open data hold enormous potential to re-shape how archaeologists work, it has become clear that the systems we have constructed to support data-sharing at scale generally fail to account for how they might fit into or alter existing value regimes and forms of professional interaction. In order to more fully unlock this potential, we need to broaden our imagination of what data-sharing actually entails, including the social and communicative processes that underlie these acts. This paper highlights some of these challenges through the case of the Database of Obsidian Sourcing Studies (DObsiSS), an integrated dataset compiled for the completion of the Author's Master's Thesis in 2015 (Batist 2015a). Specifically, this paper draws attention to past archaeologists' inability to anticipate future use cases for their data which renders them incomparable, and the struggle to reconcile past circumstances in a context of data re-use.

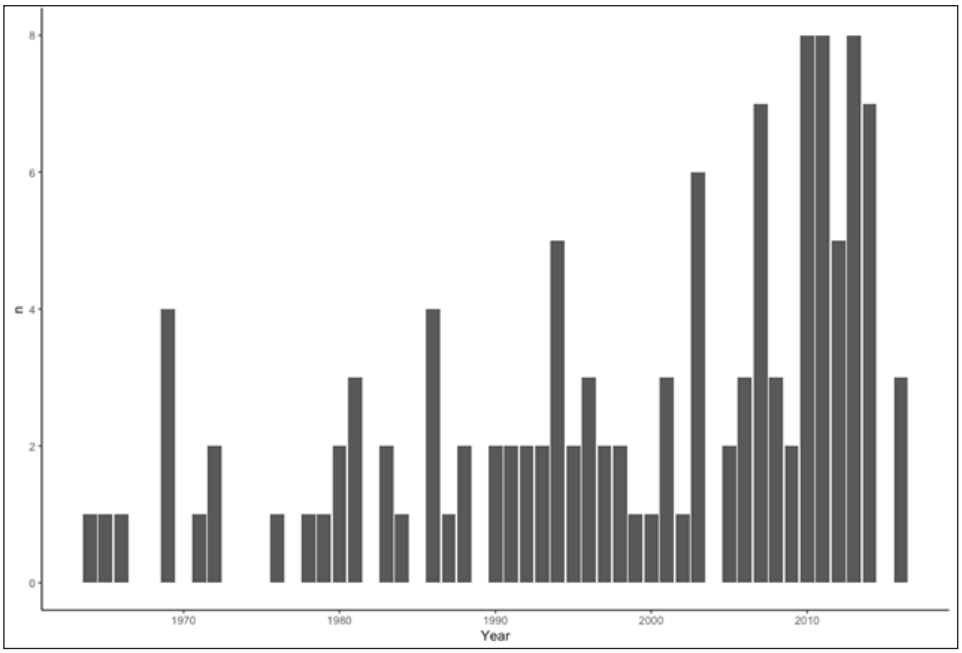

Fig. 1 – Distribution of obsidian sourcing studies referenced in DObsiSS.

2. Assembling the Database of Obsidian Sourcing Studies

DObsiSS is a combined dataset assembled from dozens of published reports, and served as the basis for a Master's Thesis that entailed comparative and diachronic analysis of chemically characterized obsidian artefacts found across southwest Asia and Anatolia between 12,000-5700 BP (Batist 2015a). Assemblng DObsiSS involved integrating the findings derived from obsidian sourcing studies published from 1964-2014 (Fig. 1) into a single spreadsheet, which was then used to compare the composition of archaeological assemblages over broad time spans and across a vast geographic range. The similarity of obsidian assemblages, measured in terms of the proportions of raw material originating from different geological sources, was used as a vector for tracing cultural contacts.

2.1 *Obsidian sourcing studies in historical context*

Integrating the myriad data sources forced the Author to reckon with significant practical and epistemic challenges that impacted the combined dataset's value, and which ultimately influenced what could be gleaned from the totality of knowledge produced over the prior 50 years. For instance,

during the 1960s and 1970s, only handfuls of obsidian artefacts from each site were subjected to chemical characterization, whereas more recently it has become the norm to analyze entire obsidian assemblages. This development was in part caused by rapid technological advances, especially the advent of portable X-Ray Fluorescence (pXRF), which enables archaeologists to collect raw data concerning artefacts' chemical compositions with great efficiency in fieldwork and museum settings (Forster, Grave 2012; Frahm 2014; Milić 2014). This helps mitigate against the time and logistical expense necessary to ship samples to a dedicated lab.

Moreover, the capability to generate data on entire assemblages reflects a keen desire to perform comparative analyses. Earlier on, archaeologists were content to know simply that long-distance interactions occurred and that their site was engaged in these relations that facilitated the movement of these materials (Freund 2013; Kuzmin *et al.* 2020). But over the past ten years, obsidian sourcing has been heralded as a key vector for understanding regional interaction, primarily through the use of network and geospatial analysis techniques applied on these data (Golitko 2023).

However, much of the data that informs these analyses may be more problematic than they would initially appear to be. Archaeologists have rapidly discovered and defined new obsidian sources based on more intensive geophysical survey, and have significantly refined their sample preparation protocols, calibration curves, and analytical techniques (Chataigner *et al.* 1998; Gratuze *et al.* 2001; Frahm 2023). Due to the gradual and cumulative nature of these advancements, the findings produced by older studies are generally less precise, less accurate, and less complete than their newer counterparts (Frahm, Carolus 2022, 9-10).

2.2 *Re-using the data*

Given the historical circumstances, much of the combined dataset was completely incomparable to the rest. This was problematic because the methods that defined the Thesis depended upon having a large volume of data. Either a large portion of older data could be removed, which would have reduced the potential impact of the work, or analysis could proceed using the whole dataset – with all its flaws – accompanied by several caveats that would have significantly dulled the findings. The latter approach was followed.

The end product deriving from analysis of the combined dataset simply re-stated what everyone in the field already knew, but substituting hard-earned experience for fancy figures and code as the sources through which the findings were legitimized. Moreover, in ignoring important contextual information – like typological or technical characteristics, or cultural transitions that are evident on a local level – the Thesis examined less variables than what closer readings of the material would have accounted for.

As a student who was never involved in the community of practice surrounding obsidian sourcing at that point, the Author had no experience analyzing obsidian artefacts, was disconnected from the specialist community who deals with these materials and methods, and had little knowledge of the deep, yet extremely interpersonal history of this field. However, it was clear

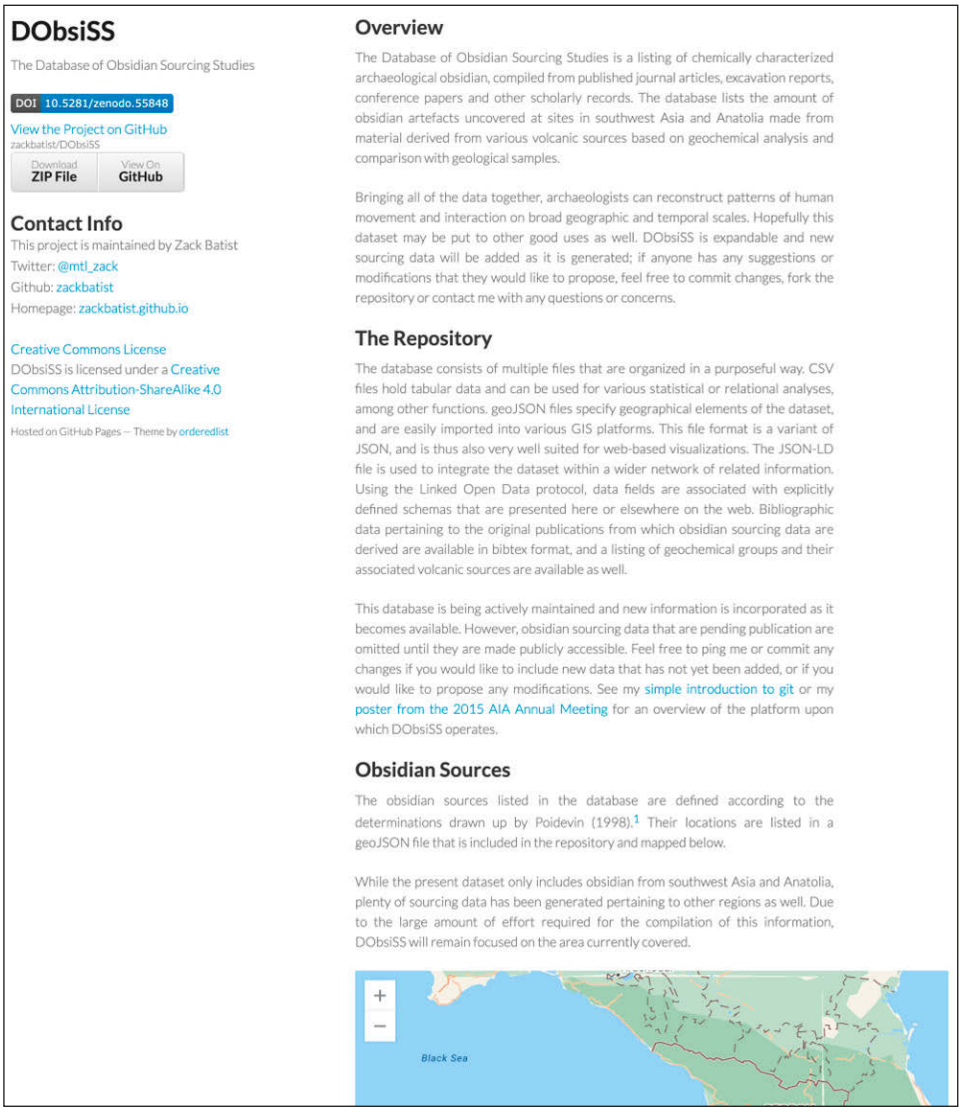

Fig. 2 – The DObsiSS website (https://zackbatist.info/DObsiSS).

that crucial contextual information hidden between the lines of published reports was crucial for making sense of the inconsistencies that were apparent in the primary sources. For instance, in addition to obtaining access to all the journal articles and published reports, it was necessary to account for obsidian artefacts published multiple times in separate reports, the physical and chemical processes that differentiate obsidian characterization methods, the intricacies of various calibration methods, and nuanced debates concerning the chronological schemes and regional divisions that were applied to index the dataset.

Since the Thesis was driven by network analysis methods – and by extension, by the numbers alone – the situated and historical context pertaining to each prior study was effectively reduced to metadata without fully understanding that each one, at various points in time, was engaging with and was limited by the body of work that existed up to that point. This contributed to the Author experiencing intense epistemic anxiety (as described by Wylie 2017; Lucas 2019, 55-57; Huggett 2022, 274-278) owing to a desire to do justice to all prior work while also acknowledging that this was near impossible given the Author's lack of understanding and community support.

Moreover, the combined dataset was a product of the Author's distinct goals and biases and assembling it involved making many decisions and trade-offs that suited the needs of the specific application. Nevertheless, the dataset was made available on the web (zackbatist.info/DobsiSS, Fig. 2), and others were encouraged to participate in its continued development. However, no one actually expressed any interest in helping to maintain it (Batist 2015b). There was one aborted attempt to repeat this work independently, and DObsiSS itself was a naive re-hash of a prior effort (cf. Varoutsikos, Chataigner 2010).

The Master's Thesis was the only comprehensive study deriving from these efforts, which testifies to their lack of value. However, numerous similar, scaled-down studies have yielded solid insights (Freund, Batist 2014; Carter *et al.* 2017; Frahm, Carolus 2022). These differed in that their work, including data collection and integration, was driven by intent to address specific research questions, and that their work was conducted with full awareness of the nuances and challenges pertaining to the defined scope. In other words, these projects incorporated epistemic context, whose understanding is derived from social and professional experience, into their design.

3. Conclusion

In thinking about data-sharing as means of bridging series of situated experiences, a world of under-appreciated social and epistemic implications begins to unfold. When creating or re-using data, an archaeologist engages in a collaborative commitment by making their experiences accessible beyond the

moment of archaeological encounter and by assuring themselves and others that these records are reliable (Dallas 2015; Batist 2023). Archaeological epistemic culture imbues a sense of professional solidarity which facilitates this aspect of work by establishing norms and expectations regarding how to best work with data in order to instill trust in a dataset's legitimacy. This is not typically formally encoded in transmitted data files, but is shared through subtextual cues and alternative forms of personal communication that surround the formal document (Chapman, Wylie 2016, 207; Batist 2024).

As such, simply making data available on the web, perhaps supported by formal documentation, is not enough to breathe new life into old data. Archaeologists apply a great deal of work to make data work for new purposes, which typically entails reconciling differences between disparate datasets and between each dataset and the new objectives that they were never originally meant to address. The productive integration of perspectives and practices is in fact supported by social structures, which play significant roles in coordinating labour and information produced thereof, and which extend beyond what the infrastructures that 'open archaeology' – often presented as a primarily tech-driven solution – is capable of addressing.

Zachary Batist
Faculty of Information
University of Toronto
z.batist@mail.utoronto.ca

Acknowledgements

I appreciate the support of Tristan Carter and Andy Roddick, who supervised the Master's Thesis that this paper is about. Thanks also to Ellery Fraham, whose writing on the quality of obsidian sourcing data inspired me to write this paper. I owe an additional debt of gratitude to Anna Maria Marras, Alessio Palmisano and the rest of the ArcheoFOSS 17 Organizing Committee for including me in the conference.

REFERENCES

Atici L., Whitcher Kansa S., Let-Tov J., Kansa E.C. 2013, *Other people's data: A demonstration of the imperative of publishing primary data*, «Journal of Archaeological Method and Theory», 20, 663-681 (https://doi.org/10.1007/s10816-012-9132-9).

Batist Z. 2015a, *Obsidian Circulation Networks in Southwest Asia and Anatolia (12,000-5700 B.P.): A Comparative Approach*, Master's Thesis, Hamilton, Ontario, McMaster University (https://macsphere.mcmaster.ca/handle/11375/16528).

Batist Z. 2015b, *The Database of Obsidian Sourcing Studies (DObsiSS): Using Git for collaborative management of archaeological data*, Poster presented at the American Archaeological Institute 116th Annual Meeting (New Orleans 2015) (https://raw.githubusercontent.com/zackbatist/DObsiSS/master/AIA2015_DObsiSS_Poster.pdf).

Batist Z. 2023, *Archaeological Data Work as Continuous and Collaborative Practice*, PhD Thesis, Toronto, Ontario, University of Toronto (https://tspace.library.utoronto.ca/handle/1807/130306).

Batist Z. 2024, *On the value of informal communication in archaeological data work*, «Open Archaeology» (https://doi.org/10.1515/opar-2024-0014).

Carter T., Batist Z., Campeau K., Garfinkel Y., Streit K. 2017, *Investigating pottery Neolithic socio-economic 'regression' in the Southern Levant: Characterizing obsidian consumption at Sha'ar Hagolan (N. Israel)*, «Journal of Archaeological Science: Reports», 15, 305-317 (https://doi.org/10.1016/j.jasrep.2017.08.016).

Chapman R., Wylie A. 2016, *Evidential Reasoning in Archaeology*, London, Bloomsbury Academic (https://books.google.com?id=tMohDQAAQBAJ).

Chataigner C., Poidevin J.L., Arnaud N.O. 1998, *Turkish occurrences of obsidian and use by prehistoric peoples in the Near East from 14,000 to 6000 BP*, «Journal of Volcanology and Geothermal Research», 85, 1, 517-537 (https://doi.org/10.1016/S0377-0273(98)00069-9).

Cook K. 2018, *Open data as public archaeology: The Monumental Archive Project*, «AP: Online Journal in Public Archaeology», 8, 2, 177-194 (https://doi.org/10.23914/ap.v8i2.152).

Cook K., Çakirlar C.,Goddard T., DeMuth R.C., Wells J. 2018, *Teaching Open Science: Published data and digital literacy in archaeology classrooms*, «Advances in Archaeological Practice», 6, 2, 144-156 (https://doi.org/10.1017/aap.2018.5).

Dallas C. 2015, *Curating archaeological knowledge in the digital continuum: From practice to infrastructure*, «Open Archaeology», 1, 176-207 (https://doi.org/10.1515/opar-2015-0011).

Faniel I., Kansa E., Whitcher Kansa S., Barrera-Gomez J., Yakel E. 2013, *The challenges of digging data: A study of context in archaeological data reuse*, in J.S. Downie, R.H. McDonald (eds.), *JDCL '13: Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital libraries (2013)*, 295-304 (https://doi.org/10.1145/2467696.2467712).

Fisher M., Fradley M., Flohr P., Rouhani B., Simi F. 2021, *Ethical considerations for remote sensing and open data in relation to the endangered archaeology in the Middle East and North Africa project*, «Archaeological Prospection», 28, 3, 279-292 (https://doi.org/10.1002/arp.1816).

Forster N., Grave P. 2012, *Non-destructive PXRF analysis of museum-curated obsidian from the Near East*, «Journal of Archaeological Science», 39, 3, 728-736 (https://doi.org/10.1016/j.jas.2011.11.004).

Frahm E. 2014, *Characterizing obsidian sources with portable XRF: Accuracy, reproducibility, and field relationships in a case study from Armenia*, «Journal of Archaeological Science», 49, 105-125 (https://doi.org/10.1016/j.jas.2014.05.003).

Frahm E. 2023, *The obsidian sources of eastern Turkey and the Caucasus: Geochemistry, geology, and geochronology*, «Journal of Archaeological Science: Reports», 49, 104011 (https://doi.org/10.1016/j.jasrep.2023.104011).

Frahm E., Carolus C.M. 2022, *Identifying the origins of obsidian artifacts in the Deh Luran Plain (Southwestern Iran) highlights community connections in the Neolithic Zagros*, «Proceedings of the National Academy of Sciences», 119, 43, e2109321119 (https://doi.org/10.1073/pnas.2109321119).

Freund K.P. 2013, *An assessment of the current applications and future directions of obsidian sourcing studies in archaeological research*, «Archaeometry», 55, 5, 779-793 (https://doi.org/10.1111/j.1475-4754.2012.00708.x).

Freund K.P., Batist Z. 2014, *Sardinian obsidian circulation and early maritime navigation in the Neolithic as shown through Social Network Analysis*, «The Journal of Island and Coastal Archaeology», 9, 3, 364-380 (https://doi.org/10.1080/15564894.2014.881937).

Golitko M. 2023, *Geochemical networks*, in T. Brughmans, B.J. Mills, J. Munson, M.A. Peeples (eds.), *The Oxford Handbook of Archaeological Network Research*, Oxford, University Press, 132-148 (https://doi.org/10.1093/oxfordhb/9780198854265.013.8).

Gratuze B., Blet-Lemarquand M., Barrandon J.-N. 2001, *Mass spectrometry with laser sampling: A new tool to characterize archaeological materials*, «Journal of Radioanalytical and Nuclear Chemistry», 247, 3, 645-656 (https://doi.org/10.1023/A:1010623703423).

Gupta N., Martindale A., Supernant K., Elvidge M. 2023, *The CARE principles and the reuse, sharing, and curation of indigenous data in Canadian archaeology*, «Advances in Archaeological Practice», 11, 1, 76-89 (https://doi.org/10.1017/aap.2022.33).

Huggett J. 2018, *Reuse remix recycle: Repurposing archaeological digital data*, «Advances in Archaeological Practice», 6, 2, 93-104 (https://doi.org/10.1017/aap.2018.1).

Huggett J. 2022, *Data legacies, epistemic anxieties, and digital imaginaries in archaeology*, «Digital», 2, 2, 267-295 (https://doi.org/10.3390/digital2020016).

Kansa E.C. 2014, *The need to humanize Open Science*, in S.A. Moore (ed.), *Issues in Open Research Data*, London, Ubiquity Press, 31-58 (http://www.ubiquitypress.com/site/chapters/10.5334/ban.c/download/63/).

Kansa E.C., Whitcher Kansa S. 2013, *We all know that a 14 is a sheep: Data publication and professionalism in archaeological communication*, «Journal of Eastern Mediterranean Archaeology and Heritage Studies», 1, 1, 88-97 (https://doi.org/10.5325/jeasmedarcherstu.1.1.0088).

Kintigh K.W., Altschul J.H., Beaudry M.C., Drennan R.D., Kinzig A.P., Kohler T.A., Limp W.F., Maschner H.D.G., Michener W.K., Pauketat T.R., Peregrine P., Sabloff J.A., Wilkinson T.J., Wright H.T., Zeder M.A. 2014, *Grand challenges for archaeology*, «American Antiquity», 79, 1, 5-24 (https://doi.org/10.7183/0002-7316.79.1.5).

Kuzmin Y.V., Oppenheimer C., Renfrew C. 2020, *Global perspectives on obsidian studies in archaeology*, «Quaternary International», 542, 41-53 (https://doi.org/10.1016/j.quaint.2020.02.036).

Lucas G. 2019, *Writing the Past: Knowledge and Literary Production in Archaeology*, New York, Routledge.

Milić M. 2014, *PXRF characterisation of obsidian from central Anatolia, the Aegean and central Europe*, «Journal of Archaeological Science», 41, 285-296 (https://doi.org/10.1016/j.jas.2013.08.002).

Opitz R., Strawhacker C., Buckland P., Cothren J., Dawson T., Dugmore A., Hambrecht G., Koster W., Lethbridge E., Mainland I., McGovern T., Newton A., Palsson G., Ryan T., Streeter R., Stade E., Szabo V., Thompson P. 2021, *A Lockpick's guide to dataARC: Designing infrastructures and building communities to enable transdisciplinary research*, «Internet Archaeology», 56 (https://doi.org/10.11141/ia.56.15).

Varoutsikos B., Chataigner C. 2010, *Obsidatabase: Collecter et organiser les données relatives à l'obsidienne préhistorique au Proche-Orient et en Transcaucasie*, in O. Henry (ed.), *Archéologies et espaces parcourus*, Instanbul, Institut Français d'Études Anatoliennes, 11-21 (https://shs.hal.science/halshs-00718893).

Wylie A. 2017, *How archaeological evidence bites back: Strategies for putting old data to work in new ways*, «Science, Technology, & Human Values», 42, 2, 203-225 (https://doi.org/10.1177/0162243916671200).

ABSTRACT

The Database of Obsidian Sourcing Studies (DObsiSS; https://zackbatist.info/DObsiSS) is an openly accessible integrated dataset comprising the results of dozens of obsidian sourcing studies published over a 50-year span. While the history of technological and professional progress pertaining to obsidian sourcing methods has yielded immense positive returns, this also produced extremely inconsistent data that were very difficult to integrate. This contributed to a sense of epistemic anxiety deriving from the struggle to reconcile the myriad outlooks that informed each data point -- with each other, and with a new objective of analyzing the dataset as an integrated whole. These challenges were further deepened by the Author's position as a student and relatively uninitiated novice at the time when DObsiSS was being compiled. This paper reflects on various social, technical and epistemic challenges that the Author faced while assembling DObsiSS and while attempting to use this integrated dataset for a failed network analysis study, and relates this experience to a broader commentary on practices relating to data re-use.