FAIR data in bioarchaeology - where are we at?

Claudia Speciale ⓘD *based on peer reviews by **Jan Kolar**, **Emma Karoune** and 2 anonymous reviewers*

The increasing reliance on digital and big data in archaeology is pushing the scientific community more and more to reconsider their storing and use [1, 2]. Furthermore, the openness and findability in the way these data are shared represent a key matter for the growth of the discipline, especially in the case of bioarchaeology and archaeological sciences [3].

In this paper, [4] the author presents the result of a survey targeted on UK bioarchaeologists and then extended worldwide. The paper maintains the structure of a report as it was intended for the conference it was part of (CAA 2023, Amsterdam) but it represents the first public outcome of an inquiry on the bioarchaeological scientific community. A reflection on ourselves and our own practices. Are all the disciplines adhering to the same policies? Do any bioarchaeologist use the same protocols and formats? Are there any differences in between the domains? Is the Needs Analysis fulfilling the questions?

The results, obtained through an accurate screening to avoid distortions, are creating an intriguing picture on the current state of "fairness" and highlighting how Institutions' rules and policies can and should indicate the correct workflow to follow. In the end, the wide application of the FAIR principles will contribute significantly to the growth of the disciplines and to create an environment where the users are not just contributors, but primary beneficiaries of the system.

[1] Huggett j. (2020). Is Big Digital Data Different? Towards a New Archaeological Paradigm, Journal of Field Archaeology, 45:sup1, S8-S17. `https://doi.org/10.1080/00934690.2020.1713281`

[2] Nicholson C., Kansa S., Gupta N. and Fernandez R. (2023). Will It Ever Be FAIR?: Making Archaeological Data Findable, Accessible, Interoperable, and Reusable. Advances in Archaeological Practice 11 (1): 63-75. `https://doi.org/10.1017/aap.2022.40`

[3] Plomp E., Stantis C., James H.F., Cheung C., Snoeck C., Kootker L., Kharobi A., Borges C., Reynaga D.K.M., Pospieszny Ł., Fulminante, F., Stevens, R., Alaica, A. K., Becker, A., de Rochefort, X. and Salesse, K. (2022). The IsoArcH initiative: Working towards an open and collaborative isotope data culture in bioarchaeology. Data in brief, 45, p.108595. https://doi.org/10.1016/j.dib.2022.108595

[4] Lien-Talks, A. (2024). How FAIR is Bioarchaeological Data: with a particular emphasis on making archaeological science data Reusable. Zenodo, 8139910, ver. 6 peer-reviewed and recommended by Peer Community in Archaeology. https://doi.org/10.5281/zenodo.8139910

Reviews

Evaluation round #2

DOI or URL of the preprint: https://doi.org/10.5281/zenodo.10480296
Version of the preprint: 3

Authors' reply, 22 January 2024

Dear Claudia,

I appreciate your swift response and the constructive feedback you provided. I would like to assure you that the recommended amendments have been duly implemented in response to the reviewers' comments. Following your suggestion, the tables featuring absolute numbers (figg. 2-13) have been relocated to Supplementary Materials, given their redundancy with the visually represented figures 14 and percentages.

Furthermore, in line with your advice, I have included the most noteworthy survey results in the conclusions section, aiming to encapsulate the key findings and their broader implications effectively.

I remain dedicated to ensuring that the revised manuscript aligns seamlessly with your expectations and elevates the overall quality of the work. If there are any additional recommendations or specific areas you would like me to focus on during the revision process, please do not hesitate to let me know.

Thank you once again for your valuable guidance and collaboration.

Best regards,

Alphaeus

Decision by Claudia Speciale , posted 19 January 2024, validated 21 January 2024

Almost there!

Dear Alphaeus,

I thank you for the big efforts you made in addressing most of the changes proposed by the reviewers. I just recommend you to put the tables with the absolute numbers (figg. 2-13) as Supplementary Materials, because they reproduce the same results that are well visually showed with the use of figures 14- and percentages.

I would also include in the conclusions the most significant results of the survey.

Evaluation round #1

DOI or URL of the preprint: https://doi.org/10.5281/zenodo.8139910
Version of the preprint: 1

Authors' reply, 10 January 2024

Here is my attempt to improve the letter and make it more British English:

Dear recommender

I sincerely appreciate your thoughtful and thorough review of my manuscript. Your insightful feedback has greatly assisted in refining the paper. Below I have grouped my responses to your comments into key themes for clarity:

Abstract and Introduction:

1. I have incorporated numerical results and percentages into the abstract, providing a more comprehensive summary.

2. The introduction now clearly establishes a focus on biomolecular archaeology and bioarchaeology from the outset.

Results and Discussion:

3. Figures are now directly referenced in the text, improving the overall flow of information.

4. Results and graphs have been more seamlessly integrated into the Discussion section.

5. Claims regarding data accessibility have been revised to accurately reflect the dataset, addressing concerns about overstatement.

Methods Section:

6. The Methods section has undergone significant expansion, offering detailed descriptions of data collection, processing, and analysis procedures.

7. A questionnaire document has been included in the supplementary materials.

References and Citations:

8. Relevant references relating to general archaeological data management have been incorporated into the text.

9. The use of persistent identifiers, including examples such as DOI, has been clarified.

Dataset and Data Management:

10. The raw dataset, data cleaning steps, and analysis specifics have been added to the supplementary materials.

11. A paragraph explaining data management plans and their connection to FAIR data principles has been added.

Specific Requests and Clarity Improvements:

12. Various specific requests, such as changing terminology, clarifying headings, and addressing particular line references, have been implemented per your suggestions.

I trust these revisions address your concerns and enhance the overall quality of the manuscript. I am committed to ensuring the final version meets the highest standards of scholarly research. Thank you once again for your time and effort in reviewing my work.

Yours sincerely,

Alphaeus Lien-Talks

Decision by Claudia Speciale , posted 02 October 2023, validated 02 October 2023

Very interesting but minor revisions are needed

The paper is of great interest. I enjoyed very much that the results are the collection of researchers' opinions. I think this is the most significant part of the paper, despite some areas are missing - for example, there are no archaeobotanists in the list. I suggest to follow precisely the indications of the reviewers to increase the value and the significance of the manuscript. Especially if the author's aim is to transform it into a paper to publish in a journal.

Reviewed by Jan Kolar, 12 September 2023

Review of How fair is bioarchaeological data: with particular emphasis on making archaeological science data reusable

For Peer Community in Archaeology, by Jan Kolar (UCL Institute of Archaeology)

The paper aims to evaluate bioarchaeology from the perspective of data accessibility and reusability. The author conducted a needs analysis for identifying existing practices and needs of the bioarchaeological community. His main conclusions are:

- Bioarchaeology significantly reuses data from other papers.
- There is no standard process of data deposition and back-up across the subdisciplines of bioarchaeology.
- The bioarchaeological community perceives the importance to store and share the data transparently, but only part of it is doing so.

The paper certainly aims to discuss important issues of current scientific practice – fairness and open access. The author decided to analyse that through an online survey and then discussing the results. The intention behind this endeavour is good, the paper fails to provide an up-to-date and broad analysis. Firstly, there are no references dated after 2020 and even from that year there are only couple of them. The author does not even refer to an extensive review of data sharing in archaeology (Marwick and Birch, 2018). There are no mentions to open access aDNA databases and databases of isotopic datasets (see my comments below), so it seems that the author is not aware of their existence. Secondly, the paper is highly over-visualised (31 figures!) which makes the text very fragmented and not easy to follow. Thirdly, there are no clear conclusion in the Conclusion section.

This paper needs a lot of work and I cannot recommend its publication in the current state. However, I believe that after substantial amendments it can be published in a form of a short report with much lower number of figures. That of course depends on the place of submission.

Major Issues:

Through the paper – The author refers to the paper as the "thesis". Is it because it was originally part of a thesis? Confusing, please amend.

Line 79 and further: Background: This section does not belong to Materials and methods as it explains the background concepts of the paper and at the end it presents the main aims of the paper. The whole section should be moved to the introduction, just after the existing text there. Then it would have a logical order.

Lines 125-127: Here the author finally claims that the paper will deal with the biomolecular archaeology. The focus of the paper was so far slightly confusing, as the author was all the time speaking about bioarchaeology as whole, but still not mentioning e.g. archaeobotany. The focus should be clear from the beginning, so please narrow your subject of interest much earlier, or keep it wide for the whole paper. Be consistent.

Line 128 and further: Ancient DNA section: Here the author somehow does mention only the need to store the raw data of aDNA without mentioning that it is actually happening. I recommend checking Orlando et al. (2021), which has references to open access aDNA databases.

Line 142 and further: Stable isotopes section: here the author again presents what are stable isotope analyses but does not provide any references to open access repositories of these analyses (e.g. Plomp et al., 2022). I believe that is missing as open access repositories are the main topic of this paper. Please amend.

Results: The author uses throughout this section pie charts for visualisation of the results of the survey. Although pie charts are used relatively frequently in archaeology (I am also guilty of that crime on data visualisation), pie chart is not the best method for such visualisations. Bar charts and histograms are much better. The author can find the details and instructions for better visualisations on the links below:
https://scc.ms.unimelb.edu.au/resources/data-visualisation-and-exploration/no_pie-charts

https://medium.com/analytics-vidhya/dont-use-pie-charts-in-data-analysis-6c005723e657
https://www.data-to-viz.com/caveat/pie.html

line 211, Findability: Here the author discusses use of doi and ORCID. In case of doi he does not mention them, and speaks only generally about persistent identifiers. It is not clear why. To me it is also not clear why does he speaks in this paper about ORCID, which is used to identify authors and link them properly with their published papers. How is this related to the use of persistent identifiers for datasets is not clear from the text. Please clarify.

Results: Overall, this section is too fragmented and many sections have just few sentences and then many figures. This paper seems to be "overvisualised". It is certainly desirable to have good diagrams in our papers, but 31 figures in this paper is just too much. I don't know where this paper will be published, but the usual limit in scientific journal varies between 5 and 10 figures per article. I strongly recommend cutting the number of figures to max 10 (the most important ones) and build the paper around them. This is anyway the maximum a reader can comprehend.

Line 225: here the author claims that 40% is majority. Majority of a whole is more than 50%, 40% in this case is the largest proportion.

Figures 15 and 28: These figures are totally ineffective and not understandable. I recommend another visualisation which would be easily comprehensible.

Figure 23: The barplot is useless, the table under is sufficiently showing the data.

Figure 24: Too many categories, too complicated, not clear what is the actual difference.

Line 288: "Possibly one of the most restricting elements is the lack of Data Management Plans, with over three-fifths of the respondents stating they do not create them." Restricting in what sense? Unclear. Please clarify.

Line 296 and further: Analysis. Structurally, analysis should be under the results section. Please omit this heading and simply add this summarizing part at the end of your results section.

Lines 314-318: As I outlined already above, the author mixed up doi (for datasets) and ORCID (for authors). This is confusing and the use of ORCID by the authors of datasets cannot be used as an argument for findability of datasets as ORCID is not used for identifying datasets.

Lines 319-322: The statements regarding datasets accessibility in this section are contradictory. Please clarify.

Figure 31: This figure (technically a table) is not fully clear. What does DMP stand for? Percent of what? Please clarify.

Line 363-378: 7.b Again the already mentioned issue with mixing up doi and ORCID. Please clarify.

Minor Issues:

Line 33-34: "understanding communities" – It is not clear what is meant be that. Please clarify.

Line 38-39: "damaging the raw material" – Does the author mean by that destroying the original samples? I think "raw material" is not appropriate in this context.

Line 57: "Methods and materials" – This section should be rather called "Materials and methods" as is more usual in research papers.

Line 58: "Methods and Materials" – duplicated heading within the paragraph. Please delete.

Lines 323-237: Would be interesting to analyse also demographic of the responders? DNA specialists tend to be younger, more familiar with databases etc, tend to receive funding which requires open data and tend to publish in journals which require repositories.

Line 339: Is this a dissertation or a paper? Please clarify and refer correctly.

Line 429: Is this a thesis or a paper? Please clarify and refer correctly.

References:

Marwick, B., Birch, S.E.P., 2018. A Standard for the Scholarly Citation of Archaeological Data as an Incentive to Data Sharing. Adv. Archaeol. Pract. 6, 125–143. https://doi.org/10.1017/aap.2018.3

Orlando, L., Allaby, R., Skoglund, P., Der Sarkissian, C., Stockhammer, P.W., Ávila-Arcos, M.C., Fu, Q., Krause, J., Willerslev, E., Stone, A.C., Warinner, C., 2021. Ancient DNA analysis. Nat. Rev. Methods Primer 1, 14. https://doi.org/10.1038/s43586-020-00011-0

Plomp, E., Stantis, C., James, H.F., Cheung, C., Snoeck, C., Kootker, L., Kharobi, A., Borges, C., Moreiras Reynaga, D.K., Pospieszny, Ł., Fulminante, F., Stevens, R., Alaica, A.K., Becker, A., De Rochefort, X., Salesse, K., 2022. The IsoArcH initiative: Working towards an open and collaborative isotope data culture in bioarchaeology. Data Brief 45, 108595. https://doi.org/10.1016/j.dib.2022.108595

Reviewed by anonymous reviewer 2, 08 September 2023

The manuscript submitted has an interesting approach regarding open access on bioarchaeological data and their opportunities to reuse and maximize the benefits derived from archaeological research. It presents a coherent structure with an introduction to the main disciplines that study bioarchaeological material (with the exception of archaeobotany, which is not mentioned in the manuscript). Secondly, it presents methods and materials, where the authors used A Need Analysis based on questionnaires answered by email. Although it is spread in different countries and institutions, it comes to the attention that only includes the UK, some of the Institutes in the United States, and some institutions from other parts of Europe. Although it is a wide spectrum, I considered that some institutions are not represented (e.g. Latin American, which in some cases has a strong position regarding open access archaeological data).

As for the results presented in the manuscript, they are coherent with the methodology planned and the research questions. Discussion and interpretation of these results allows the authors to attend to the different parts of the preserving, sharing and reusability of the data.

Reviewed by Emma Karoune, 11 September 2023

I fully support the overall message of this article, which is advocating for the use of the FAIR data principles in Bioarchaeology, however, I think it needs considerable reworking and revisions to make it of publishable standard.

- The paper would benefit from being restructured – you have background sections after the materials and methods section. The background sections on FAIR and the intros to the different types of bioarchaeological data should be part of the introduction. Analysis section should be part of the discussion section.

- Methods section needs much more development - your methods of data collection, data processing and analysis need to be described in much more detail and your article would benefit from additional documentation around the methods such as a questionnaire document that showed all the questions.

- You are lacking many references relating to general archaeological, zooarchaeological and archaeobotanical data management.

- I also don't think that your data does support your analysis – I think you have overstated how well bioarchaeologists are making their data reusable.

Download the review

Reviewed by anonymous reviewer 1, 04 September 2023

Download the review