

How FAIR is bioarchaeological data: with a particular emphasis on making archaeological science data reusable

Abstract

Archaeology is witnessing an exponential rise in data generation, and this can be particularly seen in the field of bioarchaeology which encompasses ancient DNA, osteoarchaeology, paleopathology, palaeoproteomics, stable isotopes, and zooarchaeology. This increase in data production is driven by advancements in molecular biology, technology, and innovative publishing practices. These data, derived from finite material sources, hold immense analytical value for unravelling the intricate lives of our ancestors. To ensure the long-term reusability of this invaluable information, embracing the FAIR principles — Findable, Accessible, Interoperable, and Reusable — emerges as a promising avenue.

Despite the growing awareness of bioarchaeological data reuse, the current landscape exhibits a notable lack of standardised data creation and deposition procedures. This results in the deposition of diverse data types across various repositories, with varying levels of Open Access and copyright, even within specialised fields.

Survey findings reveal a diversity of practices among bioarchaeologists: while 43 out of 53 respondents make their data openly accessible, only 24 employ persistent identifiers. Moreover, standardised metadata (16 out of 53) and systematic documentation practices (24 out of 53) remain limited, impeding effective linking and interpretation. Remarkably, ancient DNA performs exceptionally well in FAIR metrics, far surpassing osteoarchaeology and paleopathology.

In light of these challenges, it is crucial to implement standardised data management procedures tailored to each biomolecular subdiscipline. Additionally, introducing a federated search interface would significantly enhance the overall findability, accessibility, interoperability, and reusability of biomolecular archaeology data. A comprehensive approach is essential to optimise the stewardship of this data and ensure equitable and meaningful reuse in the dynamic world of archaeological research.

Keywords: Bioarchaeology, FAIR principles, Archaeological Data Management, Data Stewardship

1. Introduction

The focus of this paper, the field of biomolecular archaeology, is vast and complex, encompassing various sub-disciplines such as ancient DNA (aDNA), osteoarchaeology, paleopathology, palaeoproteomics, stable isotopes, and zooarchaeology (Britton & Richards: 2020: 8; Buikstra & Beck 2016: xvii). Recent advancements in molecular biology and computer science have led to an increase in the volume of data produced by bioarchaeologists (aDNA: Dolle 2020: 2; Green et al. 2017: 180; Hofreiter et al. 2014: 287. Stable isotopes: Katzenberg & Waters-Rist 2018: 470. Palaeoproteomics: Hendy et al. 2018: 16. Overall: Pálsdóttir et al. 2019: 1). Consequently, the data within these sub-disciplines are equally intricate, involving

aspects such as genomic sequencing, isotopic ratios, and prevalence rates, to name a few (Brown & Brown 2011: 3). This data contains a wealth of information that can assist bioarchaeologists in addressing crucial scientific challenges in archaeology, such as understanding human resilience, migration, identity, and human-environment interactions, making it invaluable for archaeological research (Kintigh et al. 2014: 879; Knudson & Stojanowski 2008: 398-399).

Despite the increased ability to analyse bioarchaeological material, the samples themselves are finite (Pálsdóttir et al. 2019: 2). Archaeology, being a destructive activity, destroys the original samples. This is evident in fieldwork, where excavators remove items from their original context (Oakley 2005: 171), as well as in bioarchaeological techniques such as genomic sequencing, palaeoproteomics, ZooMS, and isotopic analysis (Hendy et al. 2018: 791; Matisoo-Smith 2018: 14; Baker and Worley 2014; Doorn 2014: 7998). The destructive nature of research is particularly prominent in the study of zooarchaeological materials, archaeobotanical remains, and human remains (Pálsdóttir et al. 2019: 2; Fossheim 2013: 9). Ethical considerations, such as ownership of materials and the appropriate use of human remains further limit the available material for study (Fox 2020: 411). Therefore, the data generated through these studies must be open for reuse by other researchers while minimising potentially unethical and destructive practices (Ulguim 2018: 71).

To ensure the long-term reusability of this data, one possible approach is to make the data FAIR - Findable, Accessible, Interoperable and Reusable (da Silva Santos et al. 2016: 271; Wilkinson et al. 2016: 6). This research reported in this paper focuses on investigating this need and potential strategies to ensure that the data produced and curated by researchers in biomolecular archaeology is accessible and reusable for academic and non-academic researchers and the general public. Ultimately, this paper aims to determine the extent to which bioarchaeological data is reusable. It is important to emphasise that ethical considerations are carefully addressed throughout this process, following the principle of being “as open as possible, as closed as necessary” (Landi et al. 2020).

1.a. Background

Before delving into the definition of biomolecular archaeology within the scope of this article, it is crucial to define the FAIR data principles. The FAIR data principles were developed to promote the Findability, Accessibility, Interoperability, and Reusability of data and its associated metadata (da Silva Santos et al. 2016: 271). The standard was initially conceived in 2014 and aims to facilitate data reuse and enable greater knowledge discovery (Wilkinson et al. 2016: 6). Numerous elements contribute to achieving FAIRness in data, and a comprehensive list can be found at www.go-fair.org (ibid: 4; Go-FAIR n.d.). The specific variations, adaptations, and interpretations of these principles depend on the researcher. In this article, the principles included shall be defined primarily by the influence of the E-RIHS Data Curation Policy (D.5.3) (Wright & Richards 2020). The key principles are shown in Figure 1. In addition to the FAIR data principles, the addition of the analysis of Data Management Plan usage is also examined as it enables the data producer to reflect on how best to ensure greater reusability for other researchers. The documents encourage the data creator to consider aspects of findability, accessibility, interoperability, and reusability of the data from the beginning of the research project. They prompt thinking about metadata standards, data formats, persistence and availability of identifiers, and options for data sharing and archiving.

Overall, incorporating Data Management Plans along with FAIR principles pushes researchers to maximise the future use and impact of their data by other groups, leading to more rigorous, open, and collaborative science.
















F indable	Persistent IDs/ ORCID/ Citations iD	Metadata schemas 	PIDs in metadata 		
A ccessible	Communication protocols 	Harvestable metadata and Endpoints 	Open Access 	Repositories 	
I nteroperable	Metadata models 	Standardised file formats 	Ontologies 	Controlled vocabulary 	
R eusable	Future ready 	Systematic documentation 	Naming conventions 	Detailed metadata 	Legal requirements 

Figure 1 The FAIR data principles as interpreted for this article (Go-FAIR n.d.).

Data management plans are essential in ensuring effective data management from conception through to deposition and reuse.

Archaeological research has traditionally prioritised data collection, with minimal emphasis on reuse (Huggett 2018). However, the growing recognition of open and reproducible science has transformed this outlook (Kansa et al. 2020). Advocates of open science argue for free, unrestricted access to research data, spurring numerous repositories to house archaeological information (Kansa & Kansa 2013). Digital tools have also expanded analytical possibilities and collaborative potential (Kansa 2012). Yet alongside technology, ethical frameworks like the CARE (Collective benefit, Authority to control, Responsibility, Ethics) and FAIR (Findable, Accessible, Interoperable, Reusable) principles increasingly guide data management per Carroll *et al.* (2020; 2021). As Marwick *et al.* (2020) discuss, modest improvements like uplifting local capacity can offset issues like “over-research” and “ethics dumping” in global archaeology. Ultimately, accessible and equitable data practices serve both scientific insight and social benefit. The conversation now turns to bioarchaeology, assessing how well specimens and datasets currently achieve findability, accessibility, interoperability and reusability. Like wider datasets, bioarchaeological resources call for robust digital infrastructure coupled with ethical orientation to ensure replicable science and collectively shared knowledge from our genetic heritage.

Since the term “bioarchaeology” was coined in 1972, it has undergone a significant revolution (Brown & Brown 2011: 3). However, there is still contention surrounding the definition of bioarchaeology, with different interpretations depending on the source (Buikstra & Beck 2016: xvii). In this article, biomolecular archaeology will be defined within the context of the natural sciences, encompassing the application of techniques and approaches from biology,

chemistry, and physics to archaeological materials, such as ancient DNA, stable isotopes, and ancient proteins (Britton & Richards 2020: 5). The fields of zooarchaeology, osteoarchaeology, archaeobotany and palaeopathology, which are closely intertwined with the biological aspects, are also considered (ibid: 8). Bioarchaeological studies often involve collaborations with biologists and utilise modern samples, as advancements in molecular biology have greatly influenced the field (Navascués et al. 2010: 760; Hunter 2007a: 216). As this article focuses on the long-term preservation and reusability of data, the following sections will discuss the three main specialisms of biomolecular analyses and their impact on data production.

1.b. Ancient DNA

The study of ancient DNA (aDNA) dates back to 1984 when research on an extinct Quagga and an Egyptian mummy was conducted using polymerase chain reaction (PCR) (Willeslev & Cooper 2005: 5; Pääbo et al. 2004: 646). PCR has limitations, including the potential amplification of modern exogenous contaminants, which are better preserved (Letts 2011: 10). However, the introduction of 454 and next-generation sequencing has addressed these issues and led to significant discoveries in bioarchaeology (Hagelberg et al. 2015: 2). These technological advancements have resulted in larger datasets, with a whole genome sequence generating over 100 gigabytes of data (He et al. 2017: 412). As the number of samples being sequenced increases and costs decrease, the need for technical capacities to enable the reuse and preservation of such datasets becomes more crucial (Green et al. 2017: 180).

In the realm of ancient DNA analysis, a strong emphasis is placed on making data fair and sustainable (Orlando et al. 2021). To achieve this, researchers are encouraged to openly share their raw sequence data and alignments in public repositories, ensuring transparency and facilitating future research. This commitment to data fairness promotes research reproducibility. However, there are challenges tied to the long-term viability of data due to destructive sampling methods. Researchers primarily use whole-genome shotgun sequencing or targeted SNP capture, each with its advantages and limitations. Targeted SNP capture is cost-effective but restricts data coverage, while shotgun sequencing offers a more comprehensive view. Additionally, researchers can create immortal DNA libraries for long-term data archiving. To promote data fairness, curators should carefully weigh the material's uniqueness, DNA preservation rates, costs, and long-term archiving plans when authorising destructive sampling. In ancient DNA analysis, the goal is to balance scientific inquiry with responsible data stewardship (ibid).

1.c. Stable isotopes

The first stable isotope analyses were conducted in the 1960s and 1970s, examining $\delta^{13}\text{C}$ values of modern-day plants to distinguish between C3 and C4 photosynthetic carbon dioxide fixation (Chesson et al. 2018: 163; Bender 1971: 1239). Subsequently, the $\delta^{13}\text{C}$ values of human bone collagen were studied in archaeological investigations to understand the importance of maize in prehistoric New York State (Chesson et al. 2018: 163; Vogel & van der Merwe 1977: 239). However, much of the relevant data published in articles remains inaccessible to researchers due to being unpublished or incorporated within articles, hindering access to large datasets (Pauli et al. 2017: 2997). Technological advancements have facilitated the analysis of more samples at a lower cost, leading to increased data production (Katzenberg & Waters-Rist 2018: 470).

In terms of current data practices, the IsoArch initiative, established in 2011, is dedicated to promoting fairness and openness in isotope bioarchaeology data practices (Plomp et al. 2022). Central to this initiative is the IsoArch database, a comprehensive repository of isotopic data from diverse archaeological periods and regions. The IsoArch community includes experts, contributors, users, followers, and sponsors, all working together to shape its mission. Adherents advocate for scientific cooperation and support the database financially. Contributors share isotope data, users access and contribute to the database, followers engage on social media, and sponsors provide essential funding. IsoArch embraces the CARE and FAIR principles, prioritising equity, transparency, and data stewardship. It promotes openness by providing Open Access to data and publications, ensuring equal research participation, and adhering to evolving open science standards. In essence, IsoArch is committed to fostering an open and collaborative culture in isotope bioarchaeology for accessible and equitable scientific progress.

1.d. Palaeoproteomics

In the study of palaeoproteomics, the exploration of recovering amino acids began around 1954 (Demarchi 2020: 3; Abelson 1957: 83). However, significant progress in the field was not achieved until the turn of the 21st century when the sequence of amino acids started to be studied with the development of mass spectrometry technology (Hendy et al. 2020: 35; Ostrom et al. 2000: 1043). Further advancements occurred in 2009 with the introduction of Zooarchaeology by Mass Spectrometry (ZooMS) (Buckley et al. 2009). These developments have led to the generation of larger datasets that can be considered “big data” (Hendy et al. 2018: 16).

It is evident that bioarchaeology is increasingly characterised by the presence of big data, involving large datasets that present challenges in terms of analysis, storage, and visualisation (Pálsdóttir 2019: 2; Sagioglu & Sinanc 2013: 42). One proposed solution to address these challenges is to ensure that the data is FAIR (Findable, Accessible, Interoperable, and Reusable) (see Wright & Richards 2020).

2. Materials and methods

As stated in the introduction, bioarchaeology is generating ever more data sets from a finite amount of research material. As such, a Needs Analysis was essential to establish to what extent these datasets are being managed, made accessible and archived. A comprehensive quantitative Needs Analysis was undertaken, focusing primarily on institutions in the United Kingdom, to delve into the current data practices of bioarchaeologists, specifically focusing on the potential for data reuse. The primary objective was to align these practices with the FAIR Data principles as outlined in the E-RIHS Data Curation Policy document D.5.3 (Wright & Richards 2020). This policy serves as a foundational guide for enhancing the Findability, Accessibility, Interoperability, and Reusability (FAIR) of bioarchaeological data.

The survey questionnaire, a crucial tool in this analysis, was meticulously crafted to incorporate a mix of qualitative, quantitative, and Likert scale questions. It sought to capture insights into various facets of bioarchaeological data management, including motivations for data deposition, preferences regarding data formats and types, storage locations, metadata utilisation, challenges associated with Open Access, and systematic documentation protocols.

By exploring these dimensions, the survey aimed to assess the degree of alignment between current community practices and the FAIR principles set forth by E-RIHS D.5.3.

The questionnaire was strategically distributed to ensure a targeted and thorough representation of the bioarchaeological community. Initially, it was sent via email to all bioarchaeological specialists listed on university websites in the UK, specifically those identified as specialists in bioarchaeology. To extend the reach and encourage broader participation, additional dissemination efforts were employed. Personal emails were utilised to connect with specialists directly, fostering a more personalised engagement. Moreover, the JISC Osteoarchaeology email system was leveraged as another avenue for distribution, tapping into professional networks to enhance the survey's visibility and participation. This dual approach aimed to engage specialists through both institutional and personal channels, fostering a comprehensive and inclusive response from the targeted community. As a result of this methodology, more responses were gathered from UK institutions leading to a possible bias in the results.

The initial dataset garnered a total of 102 responses from bioarchaeologists worldwide. Subsequently, the raw data underwent a meticulous anonymisation process using MS Excel to remove any personal identifiers. Further data refinement occurred through a Python script (see Data, scripts, code, and supplementary information), which systematically excluded responses with completion rates below 25%. This stringent threshold was implemented to ensure the consistency and reliability of the dataset for subsequent analysis.

The final cleaned dataset, comprising 53 responses meeting the $\geq 25\%$ completion criterion, underwent comprehensive statistical analysis. Utilising Python, cluster analysis was employed to discern patterns within bioarchaeological subfields and their adherence to FAIR recommendations. Additionally, descriptive statistics and tests, such as percentages, frequencies, T-tests, Chi-Squared and ANOVA, were calculated to provide a nuanced understanding of key survey areas. The statistics took place using Python scripts of OS, pandas (The pandas development team 2020), seaborn (Waskom 2021) and matplotlib (Hunter 2007b).

The execution of this Needs Analysis adhered strictly to ethical guidelines set forth by the University of York's Archaeology Department (The University of York n.d.). Through this meticulous methodology, the research aimed not only to shed light on the current landscape of data practices in bioarchaeology but also to identify potential areas for improvement, ensuring a more robust foundation for data reuse in the field.

3. Results

The results are supplied as supplementary information (see <https://doi.org/10.5281/zenodo.10551706> and <https://doi.org/10.5281/zenodo.10480295>), alongside this article. The results will only include those that have one specialism to ensure that the data practices align with that individual specialism. The results indicate the following findings in terms of existing reuse, Findability, Accessibility, Interoperability, Reusability and Data Management Plans:

3.a. Existing reuse

Figure 2 demonstrates the widespread reuse of data across specialisms in bioarchaeology. A large number of specialists interact with data from other fields, and Table 4 further shows a high proportion analysing publicly available data. Table 4 also reveals the extent of reuse by highlighting the share of specialists aware their data has been repurposed.

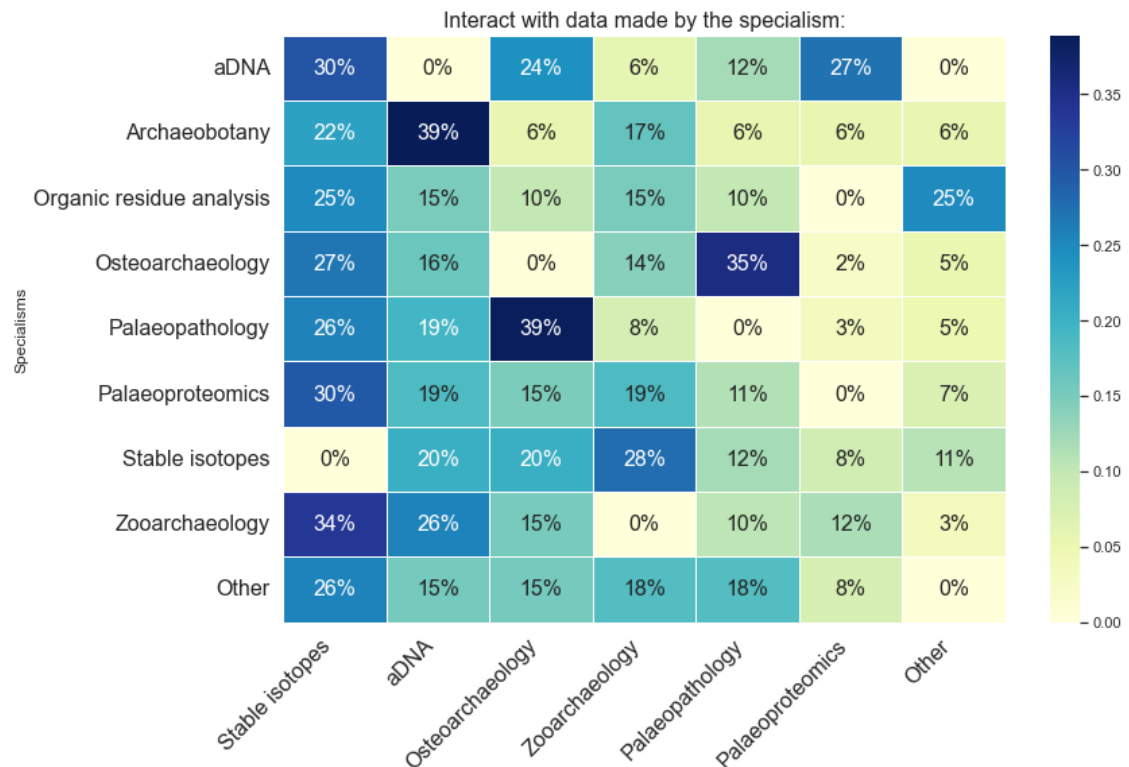


Figure 2. The proportion of specialists who interact with other fields. [Question 10]

3.b. Findability

Persistent identifiers like DOIs provide long-term discoverability by assigning permanent, unique locators. Figure 3 indicates that across specialisms, 24 out of 36 respondents indicated their data has such identifiers. Adoption is high in ancient DNA and palaeoproteomics, exceeding 70%, but usage remains low in osteoarchaeology, palaeopathology, and zooarchaeology. A chi-squared test reveals a statistically significant association between specialism and persistent identifier usage ($\chi^2= 14.76$, $p=0.02$), rejecting the possibility of equal distributions. Specifically, aDNA and palaeoproteomics utilise identifiers more than osteoarchaeology, palaeopathology and zooarchaeology. A two-sample t-test shows palaeoproteomics has significantly higher rates than palaeopathology, a closely related field ($t=3.02$, $p=0.01$).

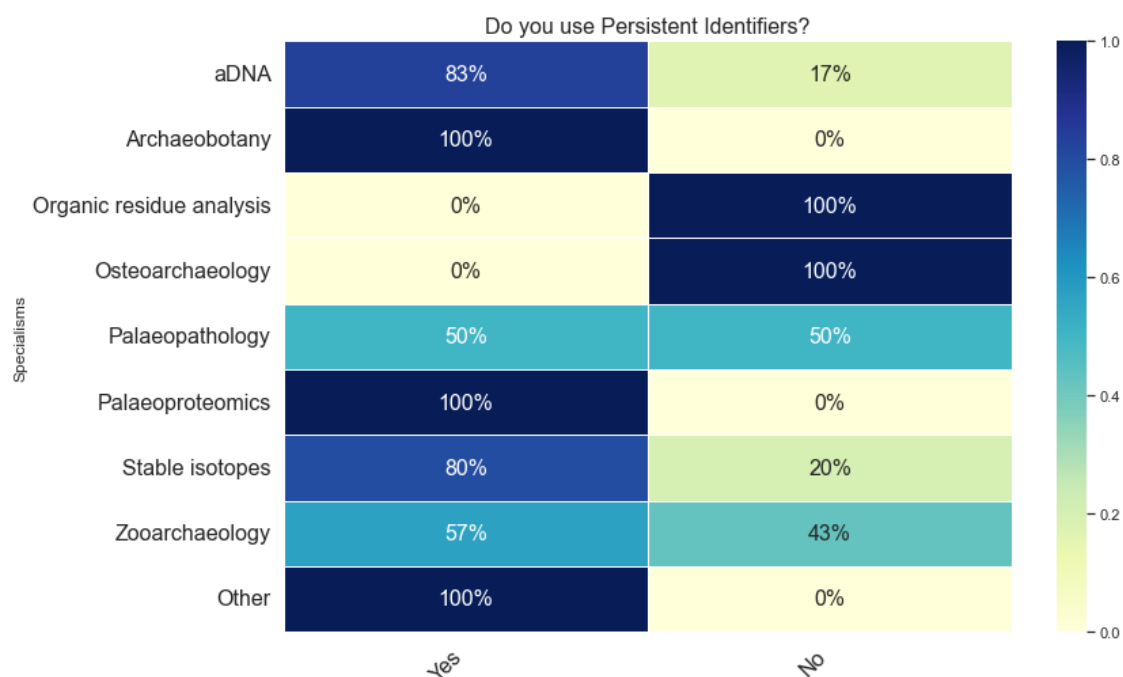


Figure 3. The proportion of specialists who use Persistent Identifiers. [Question 36]

3.c. Accessibility

Figure 4 shows that most participants publish data through reports (49) or specialised repositories (14), with fewer using institutional (4) or general (3) repositories. Chi-squared tests show subspecialism is significantly associated with publication route preferences ($\chi^2=63.42$, $p<0.001$). Figure 7 shows most specialisms make datasets openly accessible (43 yes vs 15 no). A one-sample t-test comparing Open Access rates to 50% is significant ($t=2.77$, $p=0.004$), indicating above-average openness overall. However, an independent t-test between palaeopathology and zooarchaeology finds no Open Access rate difference ($t=0.58$, $p=0.57$), suggesting variable practices between related fields. Responses suggest ethical/legal factors primarily restrict openness rather than researcher preferences (Table 7). Removing these barriers could increase accessibility. Standardised data access request protocols could provide an alternative pathway.

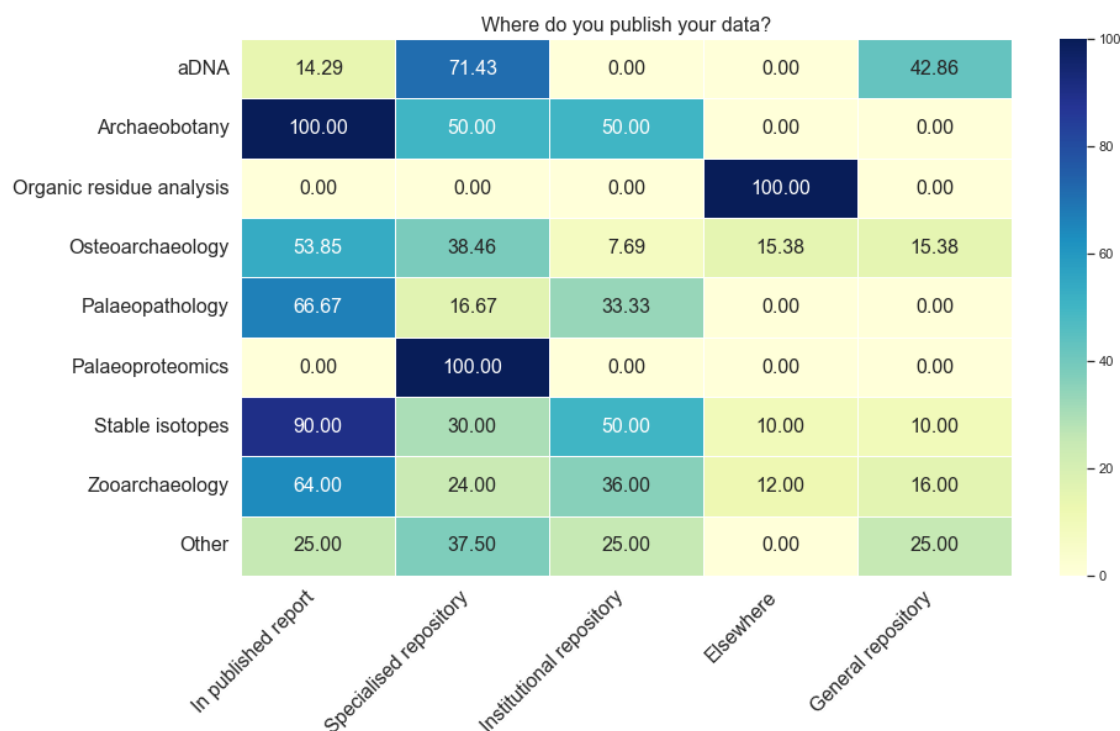


Figure 4. The percentage of locations where data is published. [Question 25]

3.d. Interoperability

Appropriate metadata aids dataset interpretation and integration. However, Table 8 shows under a third of respondents (16/53) confirm an existing schema for their specialism. Specialisms also differ in data processing levels: osteoarchaeology, palaeopathology and zooarchaeology favour fully processed data, while aDNA and palaeoproteomics share raw reads (Figure 9). In terms of file formats this data is published in, PDF is predominant (Figure 5). A chi-squared test shows a significant association between specialism and confirming metadata standards ($\chi^2=28.09$, $p<0.001$). With under a third knowing of schemas, this suggests substantial metadata inconsistencies across fields. ANOVA tests also find significant differences in favoured formats by specialism ($F=3.64$, $p<0.001$).

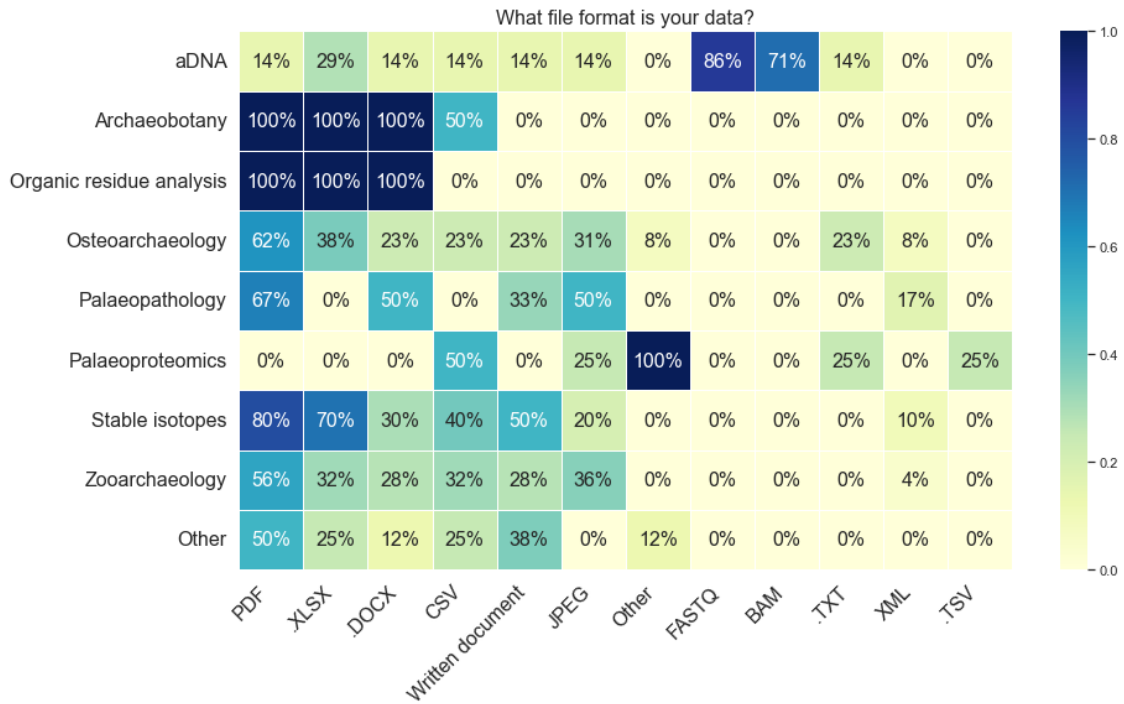


Figure 5. The file formats utilised by each specialism. [Question 24]

3.e. Reusability

Systematic documentation increases potential reuse by explaining variables. Figure 6 shows less than half of respondents (24/53) indicate limited documentation during dissemination, with no significant association with a specialism ($\chi^2=10.25$, $p=0.25$). As shown in Table 11, most participants also do not apply for copyright licenses. A one-sample t-test comparing no license rates to 50% is significant ($t=2.90$, $p=0.002$).

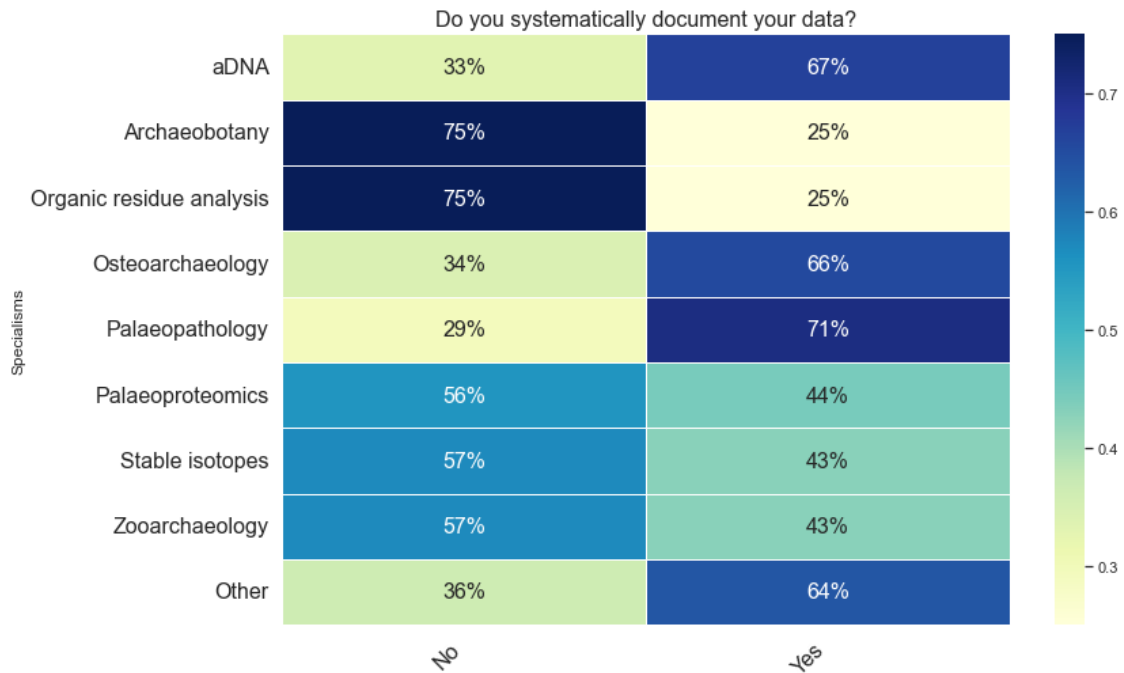


Figure 6. The proportion of specialists who Systematically Document their data. [Question 39]

3.f. Data Management Plans

Table 12 and 7 show a minority of participants (14/53) indicated using data management plans outlining procedures for data handling. A chi-squared test shows no significant association between specialism and plan usage ($\chi^2=6.24$, $p=0.62$). An independent t-test between palaeopathology and zooarchaeology also finds no significant difference in usage rates ($t=0.77$, $p=0.46$).

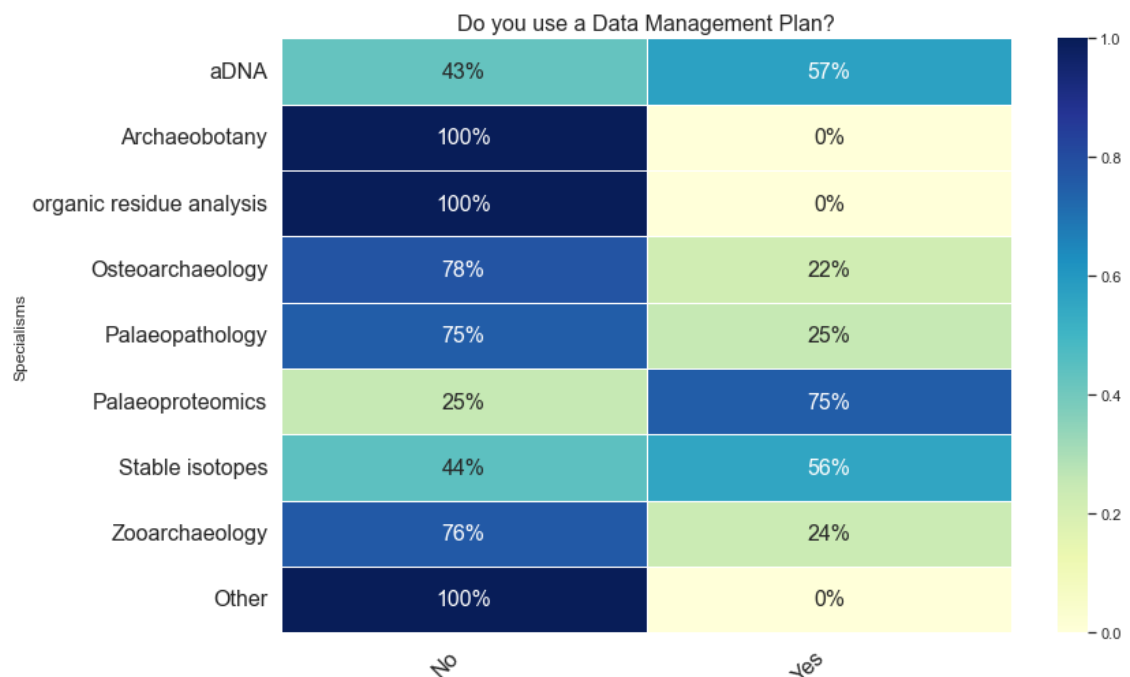


Figure 7. The proportion of specialists who use Data Management Plans. [Question 29]

3.g. Summary of results

Figure 8 summarises the results. When examining individual fields, aDNA meets the most recommendations, followed by palaeoproteomics, stable isotopes, other specialists, osteoarchaeology, zooarchaeology, and palaeopathology meeting the least.

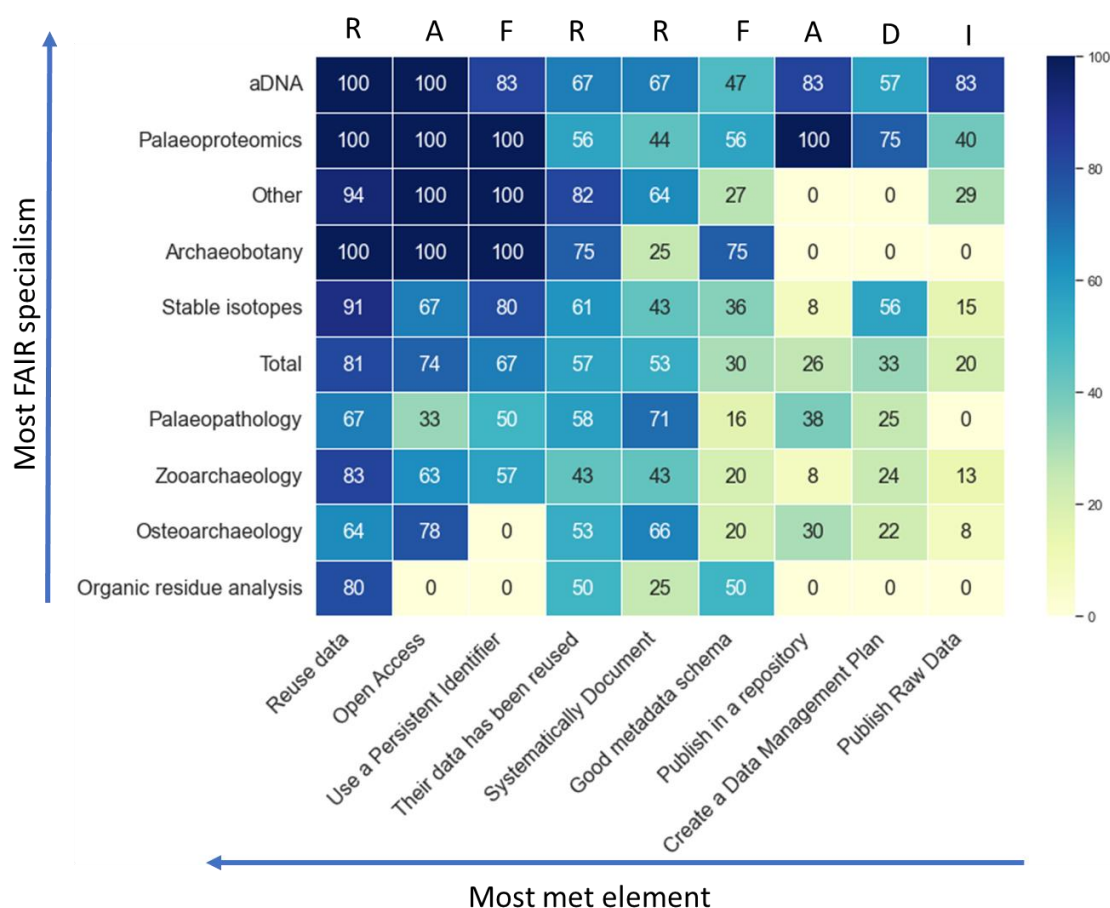


Figure 8. A summary of the results, split into descending order of individual specialisms and aspects of FAIRness. F – Findable, A – Accessible, I – Interoperable, R – Reusable, D – Data Management Plan.

4. Discussion

The objective of this research was to assess the extent to which bioarchaeological data adheres to the FAIR principles (supplementing studies such as Nicholson et al. 2023). The analysis of the results from the Needs Analysis provides valuable insights into the reusability of bioarchaeological data and highlights areas that need improvement to enhance data reusability. One key challenge is data inconsistency both between and within different specialisms. This inconsistency arises from the lack of standardised processes for data creation and deposition. Each specialist approaches data management differently, resulting in heterogeneity even within the same specialisation. This hampers integration and comparability, limiting interdisciplinary research and collaboration potential (Kintigh et al. 2014).

The findings demonstrate that data reuse is considered significant among bioarchaeologists, and there is a high level of interactivity between different specialisms (as such expanding upon general archaeology consideration of the importance of data reuse (e.g. Kansa & Kansa 2013: 89; Wright 2020; Richards et al. 2021; 2022)). The high rates of specialists interacting with external data and analysing publically available datasets demonstrate extensive reuse already occurs in bioarchaeology, aligning with the field's interdisciplinary nature (Plomp et al. 2022). Furthermore, it argues for the need to make data as open to reuse as possible. As such, this

discussion will examine potential preventatives and solutions to increase the reusability of the datasets.

For findability, persistent identifiers are underutilised overall, especially in osteoarchaeology, palaeopathology and zooarchaeology. Implementing identifiers more consistently could increase long-term dataset discoverability, as recommended by Lefebvre et al. (2019) and Król and Zdonek (2019: 21). The significantly higher adoption in ancient DNA and palaeoproteomics highlights field-specific differences in prioritising findability. Targeted training on identifier benefits could promote wider usage.

Regarding accessibility, most data is Open Access, indicating good overall practice (Wright and Richards 2020: 6). However, differences between closely associated fields like paleopathology and zooarchaeology highlight variable openness policies. Ethical and legal factors likely contribute to this variation (Ulguim 2018: 197). Nonetheless, Open Access differs from FAIR, as truly Open Access is obtained when all data is openly accessible, whilst FAIR allows for controlled access (Higman et al. 2019: 3). Furthermore, the results indicate relatively high accessibility in bioarchaeology, with most data shared through published reports. However, raw data accessibility could be improved, enhancing transparency and reanalysis potential. Nonetheless, ethical considerations around human remains are paramount,

To mitigate these challenges, contact details should be provided linking the creator to the dataset and developing standardised protocols for requesting access to closed data could increase the extent to which data could be reused within a controlled manner. As such, it is possible to apply ethical considerations, whilst fostering greater data reuse and following the principle of “as open as possible, as closed as necessary” (Landi et al. 2020: 49), as well as the CARE data principles (Gupta et al. 2023).

There are also substantial inconsistencies in metadata standards between subdisciplines. This likely hinders effective dataset interpretation and integration for reuse as discussed by Kulasekaran et al. (2014: 57-58) and Reiser et al. (2018: 1106). Along with variability in favoured file formats, this signals issues with interoperability. Establishing shared metadata guidelines tailored for each specialism could enhance interoperability (Kintigh 2006: 573), as well as ensure that file formats are “open, international [and] standardised” (Wright & Richards 2020: 7)

Reusability is the primary focus of this paper, and as the results suggest, more must be done to increase the number of datasets open to reuse.

Issues derive from limited systematic documentation during dissemination (Hendy et al. 2018: 798), and ambiguous copyright licensing. Improved documentation and licence specifications are advised to clarify reuse permissions and constraints. These challenges are also indicated by the proportion of specialists aware their data is reused. This trend indicates traceability could be improved. Citations would assist individuals track the use of their datasets, as with increasing web scraping capabilities, routes of data reuse could be identified.

Finally, and concerningly, data management plans are underutilised across all subfields despite being considered best practice (Di Giorgio & Ronzino 2018; Kansa et al. 2020: 45). Plans are crucial for effective management and reusability, and by adopting plans more widely it would promote transparency and accountability in data handling (tDAR 2017). This can lead to making research easier, reducing accusations of fraud or bad science, enabling data sharing

for reuse, gaining credit, meeting funder/institution requirements and preventing data loss (Correia 2019). Limited adoption highlights the need for greater awareness. Targeted FAIR training and infrastructure may help overcome barriers to plan usage.

This research highlights the significance of addressing data reusability in bioarchaeology and emphasises the importance of standardisation and FAIR practices. While promoting consistent data management procedures and encouraging the adoption of all FAIR principles, researchers must also acknowledge and cater to each bioarchaeological specialism's unique needs and characteristics. By doing so, bioarchaeologists can balance consistency and customisation, ensuring the effective implementation of FAIR guidelines while maximising the value of their data for interdisciplinary research and scientific discoveries. Additionally, integrating ethical considerations within data management practices strengthens the responsible and sustainable use of bioarchaeological data to advance knowledge in archaeology and related fields.

Overall, the results reveal variable data management practices both between and within bioarchaeological subdisciplines, despite data reuse being considered significant among bioarchaeologists indicating opportunities to enhance adherence to FAIR principles. To address inconsistency and improve reusability, developing standardised data management protocols encompassing all aspects of bioarchaeology is crucial. Such standardisation would require systematic practice changes, including data collection, storage, and sharing. By implementing consistent protocols, researchers can promote uniformity and facilitate discoverability and accessibility of data for reuse.

5. Conclusion

In conclusion, the study of data practices in bioarchaeology reveals significant disparities in FAIR adherence both between and within subdisciplines. Ancient DNA emerges as the most FAIR-aligned specialism overall, demonstrating high rates of open data deposition, use of identifiers, and sharing of raw data. Conversely, osteoarchaeology, zooarchaeology, and paleopathology exhibit the lowest FAIR alignment presently. Improvements in findability through identifiers and accessibility of raw datasets are particularly needed across multiple fields. Recognising these disparities, targeted measures such as metadata standards, licensing, and training initiatives can foster consistency in data management. Ultimately, striking a balance between customisation and uniformity will be paramount to maximising the reuse potential of these invaluable resources. The study provides critical insights to inform the implementation of FAIR data stewardship in bioarchaeology.

The findings underscore the value placed on data reuse and collaboration within bioarchaeology, yet also highlight significant opportunities for enhancing reusability through wider adoption of FAIR principles and standardised data management practices. Overall, the research emphasises the importance of enhancing the findability, accessibility, interoperability, and reusability of bioarchaeological data. This necessitates a multifaceted approach, including increased use of persistent identifiers and open data platforms, development of specialised metadata standards, encouragement of raw data publication, and integration of ethical considerations into protocols. By promoting FAIR principles while allowing for customisation to suit specific specialisms, we can maximise the potential for data reuse.

Ensuring robust data management through detailed data management plans is essential for responsible stewardship. Cultivating interdisciplinary collaborations through collaborative networks and federated portals can facilitate impactful discoveries. Improved data reusability will accelerate scientific progress in bioarchaeology and related disciplines, thereby advancing the understanding of humanity across time. Achieving this goal demands a collective commitment from the field to elevate data practices. By working collaboratively to make bioarchaeological data FAIR, it will be possible to fully realise the potential of this invaluable resource.

6. Funding:

The research presented in this paper did not receive any specific funding. However, the author would like to disclose that attendance at the Computer Applications in Archaeology 2023 conference was made possible through the support of the Arts and Humanities Research Council's Collaborative Doctoral Partnership Scheme. The author is grateful for this funding opportunity, which facilitated their participation in the conference and provided valuable insights and collaborations for this research. It is important to note that the funding source played no role in the design, execution, analysis, interpretation, or presentation of the research findings in this paper. The research and its conclusions remain independent and solely reflect the author's efforts and expertise in the field of bioarchaeology and data reusability.

7. Conflict of Interest Statement:

The author of this paper declares no conflict of interest related to the content presented in this research. The study, data analysis, and conclusions have been conducted independently and without any financial, personal, or professional affiliations that could potentially influence the objectivity or integrity of the information provided. There are no associations or financial arrangements with any organisations or individuals that could be perceived as having a vested interest in the outcomes or findings of this article.

In the event of any potential conflicts of interest arising in the future, the author will promptly disclose them and take appropriate actions to ensure transparency and maintain the highest level of integrity in the research process. The author is committed to upholding ethical practices in scientific inquiry and reporting to maintain the credibility and credibility of the research findings.

8. Data, scripts, code, and supplementary information availability

To analyse the data presented in this research, a Python script was coded to facilitate quick and standardised analysis of the results. The code, which adheres to best practices in data processing, has been made publicly available at the following DOI: <https://doi.org/10.5281/zenodo.10480289>

The data used in this study is also available at the DOI: <https://doi.org/10.5281/zenodo.10480295>, with the redaction of all personal data to ensure data privacy and protection. By providing access to both the data and the analysis code, we aim to foster transparency and reproducibility in scientific research.

The data tables are supplied as Supplementary material available alongside this paper.

To meet the requirement of providing machine-readable formats, the data and code are deposited in an open repository, Zenodo, with versioning to ensure stability and permanence. In addition, a readme file is included, thoroughly explaining the data and code, ensuring that other researchers can reuse and understand the materials.

We adhere to the principles of open science, and by making the data and code available, we encourage the scientific community to verify and build upon our findings. The provided resources facilitate the replication of the study and contribute to the advancement of knowledge in the field of bioarchaeology and data reusability.

9. Citation for Data, Code and Figures:

Lien-Talks, A. 2024. How FAIR is Bioarchaeological Data – code. Zenodo. <https://doi.org/10.5281/zenodo.10935117>

Lien-Talks, A. 2024. How FAIR is Bioarchaeological Data – data. Zenodo. <https://doi.org/10.5281/zenodo.10480295>

10. References

- Abelson, P. H. 1957. Chapter 5: Organic Constituents of Fossils. In: Ladd, H. S. (ed.) *Treatise on Marine Ecology and Paleoeology*. Boulder: Geological Society of America, pp. 97-101. DOI: <https://doi.org/10.1130/MEM67V2-p87>
- Baker, P. and Worley, F. 2019. *Animal Bones and Archaeology - Recovery to Archive*. Swindon: Historic England.
- Bender, M. M. 1971. Variations in the $^{13}\text{C}/^{12}\text{C}$ ratios of plants in relation to the pathway of photosynthetic carbon dioxide fixation. *Phytochemistry*, 10(6): 1239–1244. DOI: 10.1016/S0031-9422(00)84324-1
- Britton, K. and Richards, M. P. 2020. Introducing Archaeological Science. In: Richards, M. P. and Britton, K. (eds.) *Archaeological Science: An Introduction*. Cambridge: Cambridge University Press: pp. 3–10. DOI: <https://doi.org/10.1017/9781139013826.001>
- Brown, T. A. and Brown, K. 2011. *Biomolecular Archaeology: An Introduction*. Chichester: John Wiley & Sons.
- Buckley, M., Collins, M., Thomas-Oates, J. and Wilson, J. C. 2009. Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid communications in mass spectrometry*, 23(23): 3843-3854. DOI: 10.1002/rcm.4316
- Buikstra, J. E. and Beck, L. A. (eds.) 2016. *Bioarchaeology: The Contextual Analysis of Human Remains*. New York: Routledge. DOI: 10.4324/9781315432939
- Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R. and Sara, R. 2020. The CARE Principles for Indigenous Data Governance. *Data Science Journal*, 19: p. 43. DOI: 10.5334/dsj-2020-043
- Carroll, S. R., Herczog, E., Hudson, M., Russell, K. and Stall, S. 2021. Operationalizing the CARE and FAIR Principles for Indigenous data futures. *Scientific Data*, 8(1): p. 108. DOI: 10.1038/s41597-021-00892

- Chesson, L. A., Tipple, B. J., Youmans, L. V., O'Brien, M. A., and Harmon, M. M. 2018. Forensic Identification of Human Skeletal Remains Using Isotopes. In: Latham K. E., Bartelink E. J. and Finnegan M. (eds). *New Perspectives in Forensic Human Skeletal Identification*. Amsterdam: Elsevier, pp. 157–173. DOI: 10.1016/B978-0-12-805429-1.00014-4
- Correia, A. 2019. Data Management Plans. DESIR winter school, 12th December 2019. Available at <https://content.historicengland.org.uk/images-books/publications/animalbones-and-archaeology/animal-bones-and-archaeology.pdf/> [Last accessed 15 January 2024]
- da Silva Santos, L.B., Wilkinson, M.D., Kuzniar, A., Kaliyaperumal, R., Thompson, M., Dumontier, M. and Burger, K. 2016. FAIR Data Points Supporting Big Data Interoperability. In: Mertins, K., Jardim-Gonçalves, R., Popplewell, K., and Mendonça, J. P. *Enterprise Interoperability VII*. London: Springer International Publishing, pp. 270–279
- Demarchi, B. 2020. *Amino Acids and Proteins in Fossil Biominerals: An Introduction for Archaeologists and Palaeontologists*. Chichester: John Wiley & Sons. DOI: 10.1002/9781119089537
- Di Giorgio, S., and Ronzino, P. 2018. PARTHENOS Data Management Plan template for Open Research in archaeology. In: Addison, A. C. and Thwaites, H. (eds.) *3rd Digital Heritage International Congress (DigitalHERITAGE) held jointly with 2018 24th International Conference on Virtual Systems & Multimedia (VSMM 2018)*. 26 - 30 October. San Francisco: IEEE, pp. 1-4.
- Dolle, D., Fages, A., Mata, X., Schiavinato, S., Tonasso-Calvière, L., Chauvey, L., Wagner, S., Der Sarkissian, C., Fromentier, A., Seguin-Orlando, A. and Orlando, L. 2020. CASCADE: A Custom-Made Archiving System for the Conservation of Ancient DNA Experimental Data. *Frontiers in Ecology and Evolution*, 8(185): 185. DOI: 10.3389/fevo.2020.00185
- Fossheim, H. J. 2013. Introductory remarks. In Fossheim, H. J. (ed.). *More than just bones: ethics and research on human remains*. Oslo: The Norwegian National Research Ethics Committees, pp. 7-10
- Fox, K. 2020. The Illusion of Inclusion - The 'All of Us' Research Program and Indigenous Peoples' DNA. *The New England Journal of Medicine*, 383(5): 411–413. DOI: 10.1056/NEJMp1915987
- Go-FAIR n.d. Go-FAIR. Available at <https://www.go-fair.org/> [Last accessed 06 April 2024].
- Green, E. D., Rubin, E. M., and Olson, M. V. 2017. The future of DNA sequencing. *Nature*, 550(7675): 179–181. DOI: 10.1038/550179a
- Gupta, N., Martindale, A., Supernant, K., and 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. DOI: 10.1017/aap.2022.33
- Hagelberg, E., Hofreiter, M., and Keyser, C. 2015. Ancient DNA: the first three decades. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1660): 20130371. DOI: 10.1098/rstb.2013.0371
- He, K. Y., Ge, D., and He, M. M. 2017. Big Data Analytics for Genomic Medicine. *International Journal of Molecular Sciences*, 18(2): 412. DOI: 10.3390/ijms18020412

- Hendy, J., Welker, F., Demarchi, B., Speller, C., Warinner, C. and Collins, M.J. 2018. A guide to ancient protein studies. *Nature Ecology & Evolution*, 2(5): 791–799. DOI: 10.1038/s41559-018-0510-x
- Hendy, J., van Doorn, N., and Collins, M. 2020. Proteomics. In: Britton, K. and Richards, M. P. *Archaeological Science: An Introduction*. Cambridge: Cambridge University Press, pp. 35–69. DOI: 10.1017/9781139013826.003
- Higman, R., Bangert, D., and Jones, S. 2019. Three camps, one destination: the intersections of research data management, FAIR and Open. *Insights into Imaging*, 32(1): 18. DOI: 10.1629/uksg.468
- Hofreiter, M., Paijmans, J.L., Goodchild, H., Speller, C.F., Barlow, A., Fortes, G.G., Thomas, J.A., Ludwig, A. and Collins, M.J. 2015. The future of ancient DNA: Technical advances and conceptual shifts. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 37(3): 284–293. DOI: 10.1002/bies.201400160
- Huggett, J. 2018. Reuse remix recycle: repurposing archaeological digital data. *Advances in Archaeological Practice*, 6(2): 93-104. DOI: 10.1017/aap.2018.1
- Hunter, P. 2007a. Dig this. Biomolecular archaeology provides new insights into past civilizations, cultures and practices. *EMBO Reports*, 8(3): 215–217. DOI: 10.1038/sj.embor.7400923
- Hunter, J. D. 2007b. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3): 90-95. DOI: 10.1109/MCSE.2007.55
- Kansa, E. 2012. Openness and archaeology’s information ecosystem. *World archaeology*, 44(4): 498-520. DOI: 10.1080/00438243.2012.737575
- Kansa, E. C., and Kansa, S. W. 2013. Open Archaeology: we all know that a 14 is a sheep: data publication and professionalism in archaeological communication. *Journal of Eastern Mediterranean Archaeology & Heritage Studies*, 1(1): 88–97
- Kansa, S. W., Atici, L., Kansa, E. C., and Meadow, R. H. 2020. Archaeological analysis in the information age: guidelines for maximizing the reach, comprehensiveness, and longevity of data. *Advances in Archaeological Practice*, 8(1), 40–52
- Katzenberg, M. A. 2008. Stable Isotope Analysis: A Tool for Studying Past Diet, Demography, and Life History. In: Katzenberg, M. A. and Grauer, A. L. (eds.). *Biological Anthropology of the Human Skeleton*. Hoboken: John Wiley & Sons, Ltd., pp. 411–441. DOI: 10.1002/9780470245842.ch13
- Katzenberg, M. A. and Waters-Rist, A. L. 2018. Stable isotope analysis. In Katzenberg, M. A. and Grauer, A. L. (eds). *Biological Anthropology of the Human Skeleton*. Hoboken: John Wiley & Sons, Inc. pp. 467-504
- Kintigh, K. 2006. The promise and challenge of archaeological data integration. *American Antiquity*, 71(3): 567–578. DOI: 10.2307/40035365
- Kintigh, K. et al. 2014. Grand challenges for archaeology. *American Antiquity*, 79(1): 5-24. DOI: 10.7183/0002-7316.79.1.5
- Knudson, K. and Stojanowski, C. 2008. New Directions in Bioarchaeology: Recent Contributions to the Study of Human Social Identities. *Journal of Archaeological Research*, 16(1): 397–432. DOI: 10.1007/s10814-008-9024-4
- Król, K., and Zdonek, D. 2019. Peculiarity of the bit rot and link rot phenomena. *Global Knowledge, Memory and Communication*, 69(1): 20–37. DOI: 10.1108/GKMC-06-2019-0067

- Kulasekaran, S., Trelogan, J., Esteva, M., and Johnson, M. 2014. Metadata integration for an archaeology collection architecture. In: Moen, W. and Rushing, A. (eds.) *International conference on Dublin Core and metadata applications. 8-11 October*. Austin: DC Papers, pp. 53-6.
- Landi, A., Thompson, M., Giannuzzi, V., Bonifazi, F., Labastida, I., da Silva Santos, L. O. B., and Roos, M. 2020. The “A” of FAIR—as open as possible, as closed as necessary. *Data Intelligence*, 2(1-2): 47–55. DOI: 10.1162/dint_a_00027
- LeFebvre, M. J., Brenskelle, L., Wieczorek, J., Kansa, S. W., Kansa, E. C., Wallis, N. J., King, J. N., Emery, K. F., and Guralnick, R. 2019. ZooArchNet: Connecting zooarchaeological specimens to the biodiversity and archaeology data networks. *PloS One*, 14(4): e0215369. DOI: 10.1371/journal.pone.0215369
- Letts, B. 2011. Using ancient DNA to investigate the influence of environmental change on animal populations. Unpublished thesis (PhD), The Pennsylvania State University.
- Marwick, B., Pham, T. S., and Ko, M. S. 2020. Over-research and ethics dumping in international archaeology. *SPAFA Journal*, 4 (1). DOI: 10.26721/spafajournal.v4i0.625
- Matisoo-Smith, E. 2018. Ancient DNA in Zooarchaeology: New Methods, New Questions and Settling Old Debates in Pacific Commensal Studies. In: Giovas, C. M. and LeFebvre, M. J. (eds.) *Zooarchaeology in Practice: Case Studies in Methodology and Interpretation in Archaeofaunal Analysis*. Cham: Springer International Publishing, pp. 209–225. DOI: 10.1007/978-3-319-64763-0_11
- Navascués, M., Depaulis, F., and Emerson, B. C. 2010. Combining contemporary and ancient DNA in population genetic and phylogeographical studies. *Molecular Ecology Resources*, 10(5): 760–772. DOI: 10.1111/j.1755-0998.2010.02895.x
- 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. DOI: 10.1017/aap.2022.40
- Oakley, K. 2005. Forensic archaeology and anthropology: An Australian perspective. *Forensic Science, Medicine, and Pathology*, 1(3): 169–172. DOI: 10.1385/FSMP:1:3:169
- 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. and Warinner, C. 2021. Ancient DNA analysis. *Nature Reviews Methods Primers*, 1(1):14. DOI: 10.1038/s43586-020-00011-0
- Ostrom, P.H., Schall, M., Gandhi, H., Shen, T.L., Hauschka, P.V., Strahler, J.R. and Gage, D.A. 2000. New strategies for characterizing ancient proteins using matrix-assisted laser desorption ionization mass spectrometry. *Geochimica et Cosmochimica Acta*, 64(6): 1043. DOI: 10.1016/S0016-7037(99)00381-6
- Pálsdóttir, A.H., Bläuer, A., Rannamäe, E., Boessenkool, S. and Hallsson, J.H. 2019. Not a limitless resource: ethics and guidelines for destructive sampling of archaeofaunal remains. *Royal Society Open Science*, 6(10): 191059. DOI: 10.1098/rsos.191059
- Pauli, J.N., Newsome, S.D., Cook, J.A., Harrod, C., Steffan, S.A., Baker, C.J., Ben-David, M., Bloom, D., Bowen, G.J., Cerling, T.E. and Cicero, C. 2017. Why we need a centralized repository for isotopic data. *Proceedings of the National Academy of Sciences*, 114(12): 2997–3001. DOI: 10.1073/pnas.1701742114

- Plomp, E., Stantis, C., James, H. F., Cheung, C., Snoeck, C., Kootker, L., Kharobi, A., Borges, C., Reynaga, D. K. M., Pospieszny, Ł. and Fulminante, F. 2022. The IsoArch Initiative: Working towards an Open and Collaborative Isotope Data Culture in Bioarchaeology. *Data in Brief*, 45(1): 108595. DOI: 10.1016/j.dib.2022.108595
- Pääbo, S., Poinar, H., Serre, D., Jaenicke-Després, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L. and Hofreiter, M. 2004. Genetic analyses from ancient DNA. *Annual Review of Genetics*, 38 (1): 645–679. DOI: 10.1146/annurev.genet.37.110801.143214
- Reiser, L., Harper, L., Freeling, M., Han, B., and Luan, S. 2018. FAIR: a call to make published data more Findable, Accessible, Interoperable, and Reusable. *Molecular Plant*, 11(9): 1105–1108. DOI: 10.1016/j.molp.2018.07.005
- Richards, J. D., Jakobsson, U., Novák, D., Štular, B., and Wright, H. 2021. Digital Archiving in Archaeology: The State of the Art. Introduction. *Internet Archaeology*, 58(1). DOI: 10.11141/ia.58.23
- Richards, J. D., Wright, H. E., Geser, G., and Massara, F. 2022. Data Management Policies and Practices of Digital Archaeological Repositories. *Internet Archaeology*, 59(1). DOI: 10.11141/ia.59.2
- Sagiroglu, S. and Sinanc, D. 2013. Big data: A review. In: *2013 International Conference on Collaboration Technologies and Systems (CTS)*. May 2013. San Diego, CA, USA: IEEE, pp. 42–47. DOI: 10.1109/CTS.2013.6567202
- tDAR. 2017. *Data Management Plans for Archaeological Research*. Available at: <https://www.digitalantiquity.org/wp-uploads/2017/03/2017-DA-Guide-Data-Mgmt-Plan-Template-Final.pdf> [Last accessed 25 January 2024]
- The pandas development team. 2020. *pandas-dev/pandas: Pandas*. Zenodo. DOI: 10.5281/zenodo.3509134
- The University of York. n.d. Research integrity and ethics. Available at <https://www.york.ac.uk/staff/research/governance/research-integrity-and-ethics/> [Last accessed 25 January 2023]
- Ulguim, P. 2018. Models and Metadata: The Ethics of Sharing Bioarchaeological 3D Models Online. *Archaeologies*, 14(2): 189. DOI: 10.1007/s11759-018-9346-x
- van Doorn, N. L. 2014. Zooarchaeology by Mass Spectrometry (ZooMS). In: Smith, C. *Encyclopedia of Global Archaeology*. New York, NY: Springer, pp. 7998–8000. DOI: 10.1007/978-1-4419-0465-2_2418
- Vogel, J. C. and Merwe, N. J. van der. 1977. Isotopic Evidence for Early Maize Cultivation in New York State. *American Antiquity*, 42(2): 238–242. DOI: 10.2307/278984
- Waskom, M. L. 2021. seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60): 3021. DOI: 10.21105/joss.03021
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E. and Bouwman, J. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1): 160018. DOI: 10.1038/sdata.2016.18
- Willerslev, E. and Cooper, A. 2005. Review Paper. Ancient DNA. *Proceedings of the Royal Society B: Biological Sciences*, 272(1558): 3-16. DOI: 10.1098/rspb.2004.2813
- Wright, H. 2020. *Introduction to Saving European Archaeology from the Digital Dark Age*. Zenodo. DOI: 10.5281/zenodo.3842038

CAA 2023 – “50 Years of Synergy” Amsterdam, The Netherlands 2023

Wright, H. and Richards, J. 2020. *D.5.3 Data Curation Policy*. York: E-RIHS. DOI:
10.1080/00934690.2018.15