

Article

Not peer-reviewed version

The Future of Bioarchaeological Data: Why FAIR, CARE, and Machine Learning Are Essential for Sustainable Research

[Alphaeus Lien-Talks](#) *

Posted Date: 1 January 2026

doi: 10.20944/preprints202601.0016.v1

Keywords: bioarchaeology; FAIR data; CARE principles; machine learning; data management; grey literature; osteoarchaeology; ancient DNA; stable isotopes; data sustainability



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

The Future of Bioarchaeological Data: Why FAIR, CARE, and Machine Learning Are Essential for Sustainable Research

Alphaeus Lien-Talks

University of York, Historic England, Historic Royal Palaces and Archaeology Data Service;
alfietalks@live.co.uk

Abstract

Bioarchaeological materials represent finite and irreplaceable resources, with many analytical techniques requiring consumptive sampling that permanently limits future research opportunities. This challenge is particularly acute for Modern era contexts (1492–1945 CE), where industrial and colonial period collections offer crucial evidence of health transitions and disease emergence yet remain under-utilised due to data accessibility challenges. Evidence from 145 bioarchaeology specialists across 23 countries demonstrates that whilst 97% recognise data reuse as critical, fewer than half consistently implement basic measures such as persistent identifiers. Only ancient DNA research consistently meets FAIR (Findable, Accessible, Interoperable, Reusable) standards. Meanwhile, data volumes expand exponentially through technological advances. The situation is unsustainable and ethically problematic. This perspective argues three integrated commitments are essential: universal adoption of FAIR principles with appropriate infrastructure, implementation of CARE (Collective benefit, Authority, Responsibility, Ethics) principles ensuring ethical treatment of human remains, and strategic development of artificial intelligence and machine learning tools for knowledge extraction. The finite nature of bioarchaeological materials makes transformation urgent. Every sample destroyed without proper data preservation represents irreversible loss of knowledge about human heritage.

Keywords: bioarchaeology; FAIR data; CARE principles; machine learning; data management; grey literature; osteoarchaeology; ancient DNA; stable isotopes; data sustainability

1. Introduction

Introduction: The Finite Materials Imperative

Bioarchaeological materials represent direct connections to past lives, offering insights into health, diet, mobility, social structures, and human–environment interactions. For the Modern era (1492–1945 CE), these materials are particularly valuable for understanding profound transformations brought by globalisation, industrialisation, colonialism, and emerging infectious diseases [1]. Yet these irreplaceable materials face a fundamental constraint: they exist in finite quantities whilst analytical techniques increasingly require their consumption.

Archaeology is inherently destructive, with excavators removing items from situ [2]. Bioarchaeological analytical techniques compound this issue, as genomic sequencing, palaeoproteomics, ZooMS, and isotopic analysis require consumptive sampling [3–6]. This destructive nature is particularly acute for zooarchaeological materials, archaeobotanical remains, and human remains [7,8]. Ethical considerations regarding appropriate use of human remains and questions of material ownership further limit availability for study [9,10]. Consequently, data created through these studies must be open to reuse by other researchers whilst limiting the potentially unethical nature of repeated destructive investigations [10].

Yet current practices fail to meet this imperative. Analysis of data management across bioarchaeological subdisciplines reveals that most data is not Findable (fewer than 50% of specialists use persistent identifiers in most subdisciplines), not fully Accessible (raw data often unavailable, deposition scattered across journals), not Interoperable (fewer than 50% believe appropriate metadata schemas exist), and consequently not Reusable (systematic documentation lacking in most fields). This situation is unsustainable given the finite nature of materials and the accelerating pace of analytical consumption.

There are increasing amounts of data at stake. The volume and complexity of bioarchaeological data have increased exponentially. High-throughput ancient DNA sequencing can generate over 100 gigabytes per genome [11], stable isotope databases continue to proliferate [12], and proteomic analyses accelerate [13]. This expansion constitutes 'Big Data'. This brings additional associated difficulties in storing, analysing, and visualising information [14,15].

2. The Current State: A Crisis of Data Stewardship

2.1. Evidence from Bioarchaeological Practice

A comprehensive needs analysis of 145 bioarchaeology specialists across 23 countries provides clear evidence of inadequate data stewardship practices. Whilst 97% of respondents consider data reuse 'very important', actual implementation falls substantially short of enabling effective reuse. The analysis reveals dramatic variation across subdisciplines, with ancient DNA research consistently meeting FAIR criteria whilst other specialisms show mixed or poor adoption [16].

For Findability, persistent identifier (PID) adoption varies dramatically. Palaeoproteomics shows 100% adoption amongst single-specialism practitioners, and ancient DNA shows high adoption. However, zooarchaeology shows only 40% adoption, whilst osteoarchaeology shows 0% adoption amongst single-specialism practitioners. Critically, high proportions of specialists report uncertainty about whether their data possesses PIDs, particularly in osteoarchaeology (71.43%), suggesting insufficient training or awareness. Without PIDs, data becomes vulnerable to 'link rot' as URLs break and project websites disappear [17]. The average lifespan of project-based online resources is under five years [18].

ORCID (Open Researcher and Contributor ID) adoption shows similar patterns. All palaeoproteomics specialists, all single-specialism ancient DNA specialists, and all single-specialism stable isotope specialists use ORCIDs when depositing data. However, only 33.33% of osteoarchaeology single-specialism practitioners use ORCIDs. Moreover, 41.67% of osteoarchaeology specialists and 32.00% of zooarchaeology specialists deposit data without ORCIDs, limiting ability to identify data creators and facilitating reuse through direct contact [19,20].

For Accessibility, data deposition patterns create substantial barriers. The most common deposition location across subdisciplines is published reports: palaeopathology (100%), stable isotopes (90.00%), zooarchaeology (80.00%), and osteoarchaeology (77.78%). Only ancient DNA (71.43%) and palaeoproteomics (100%) primarily deposit in specialised repositories. This fragmentation across hundreds of journals, each with different access models and supplementary data formats, directly contradicts findability principles. Journal supplementary materials typically provide PDF tables difficult to extract computationally, minimal metadata beyond article content, no standardised identifiers beyond article DOIs, and limited long-term preservation guarantees [21,22].

Regarding Interoperability, metadata schema availability represents perhaps the most fundamental barrier. Fewer than 50% of specialists in most subdisciplines believe appropriate metadata schemas exist for their data. Palaeopathology shows 100% 'not sure' amongst single-specialism practitioners, and osteoarchaeology shows 57.14% 'not sure'. Only palaeoproteomics (75.00%) and ancient DNA (57.14%) report belief in appropriate schema existence. Without metadata schemas, computational discovery becomes impossible, data cannot be linked across projects, and long-term data intelligibility is compromised [23,24].

Data type analysis reveals PDF as the most common format overall, followed by XLSX, JPEG, and DOCX. Whilst PDF facilitates human reading, it resists computational analysis and data extraction [25]. Processing level analysis shows the majority of data is published fully processed rather than raw or partly processed, potentially limiting reusability by obscuring processing methodologies and preventing alternative analytical approaches [26,27].

For Reusability, systematic documentation rates are below 50% for most specialisms (except ancient DNA and osteoarchaeology). Without documentation of collection methods, analytical parameters, quality control measures, file relationships, and reference standards, data becomes interpretable only by original creators [28]. Copyright clarity also presents challenges. The highest proportion across specialisms report no copyright, but the second-highest category is 'did not understand these codes', suggesting confusion about licensing rather than conscious choice for openness.

2.2. *The Grey Literature Challenge*

Compounding inadequate data stewardship is the vast archive of 'grey literature'; unpublished fieldwork reports created for commercial or institutional purposes that lack peer review and have limited discoverability [29]. Within archaeology, this literature primarily comprises site excavation reports, typically available only as PDF documents in repositories such as the UK's Archaeology Data Service (ADS). Whilst these reports contain valuable bioarchaeological information, they remain largely inaccessible for systematic research due to limited searchability [30].

The Crossrail project (2012–2018) exemplifies both scale and challenge. Developer-funded rescue archaeology associated with building the railway line proved revolutionary for understanding London's past, with over 100 archaeologists discovering hundreds of thousands of artefacts spanning 55 million years [31]. This generated 381 reports archived at the ADS, constituting an archive classifiable as 'Big Data': large datasets with complex structures creating difficulties in storage, analysis, and visualisation [14,30]. Without effective tools for navigation and information extraction, bioarchaeological knowledge within these reports remains effectively inaccessible except through manual review; a task requiring hundreds of hours for even single projects like Crossrail, becoming impossible at the scale of thousands of projects archived nationally and globally.

For Modern era bioarchaeology, grey literature represents particularly rich but under-utilised resources. Industrial period excavations, post-medieval cemetery investigations, colonial context analyses frequently generate extensive reports archived locally but rarely synthesised. Understanding health transitions during industrialisation, disease emergence in colonial contexts, or biological impacts of globalisation requires access to information currently trapped in grey literature.

2.3. *Ancient DNA as Proof of Concept*

Ancient DNA research demonstrates that comprehensive FAIR implementation is achievable. This subdiscipline consistently meets FAIR criteria across all assessed dimensions: high PID and ORCID adoption, deposition in specialised repositories (NCBI Sequence Read Archive) with standardised metadata, Open Access data, systematic documentation, and high Data Management Plan creation rates. This success reflects several factors: established connections to molecular biology bringing mature data management practices, specialised repositories providing clear deposition pathways with standardised schemas, large data volumes necessitating systematic management, and journal requirements increasingly mandating data deposition [32].

Palaeoproteomics shows similar patterns, benefiting from biological sciences connections and specialised repositories (ProteomeXchange) [33]. These fields demonstrate that appropriate infrastructure, community norms, training, and institutional support can achieve full FAIR implementation. The challenge is scaling this success across all bioarchaeological subdisciplines whilst accommodating their unique requirements, particularly ethical constraints surrounding human remains research.

3. CARE Principles: Ethical Imperatives for Human Remains Research

FAIR principles alone prove insufficient for bioarchaeological data governance. The CARE principles - Collective benefit, Authority to control, Responsibility, Ethics - provide essential ethical framework, particularly for research involving human remains and materials from Indigenous and descendant communities [34].

3.1. Ethical Concerns in Current Practice

The needs analysis revealed substantial ethical concerns amongst practitioners. Participants emphasised: 'ethics are of primary concern', 'need to consider the ethical impacts of...data drawn from human remains from whom we do not have direct consent...consider WHO these biological data belong to', and 'as a US practitioner, I would find a database of US archaeological remains to be distasteful at best, and illegal at worst'. These concerns reflect legitimate considerations about consent, respect, cultural sensitivity, and legal frameworks.

Palaeopathology's 0% Open Access adoption amongst single-specialism practitioners directly reflects these ethical considerations. However, this raises questions about whether all-or-nothing approaches optimally balance competing interests. Many paleopathological findings, disease prevalence rates, demographic patterns, and general health trends, could potentially be shared without exposing individual-level data that might be considered disrespectful or culturally inappropriate.

The Human Tissue Act 2005 provides one framework through its 100-year threshold for human remains [35], but cultural contexts vary globally. Colonial legacies in Africa create specific concerns about material and data repatriation [36]. Indigenous rights in North America [37] and Oceania [38] demand respect for community authority over ancestral remains. Diverse cultural attitudes toward death, ancestors, and appropriate treatment of remains require flexible, contextually sensitive approaches [39].

3.2. CARE Principles in Practice

CARE principles complement FAIR by ensuring data governance respects Indigenous data sovereignty and community authority [40]. The four principles translate to specific practices. Collective Benefit requires that data use should benefit communities, not solely researchers, through consultation before data sharing decisions, ensuring research questions address community interests, sharing results with communities in accessible formats, and supporting community capacity for data interpretation and use [34]. Authority to Control means communities maintain authority over data derived from their ancestors, including rights to restrict access, require consultation before use, specify appropriate research questions, and review publications before release, extending beyond legal ownership to cultural and ethical rights [40]. Responsibility places burden on researchers for ensuring respectful, appropriate data use, including monitoring for misuse (such as racist interpretations), correcting misrepresentations, maintaining ongoing community relationships, and ensuring data governance protocols are followed [37]. Ethics requires that ethical considerations, including privacy, cultural sensitivity, and respect for the dead, override pure openness principles, recognising that not all data should be Open Access, that some research questions may be inappropriate regardless of scientific interest, and that community wishes take precedence over researcher convenience [10].

3.3. Balancing Openness with Protection

The formulation 'as open as possible, as closed as necessary' captures the required balance [41]. Several approaches can balance research needs with ethical requirements. Tiered Access Systems enable metadata (site, date, general findings) to remain fully Open Access enabling discovery, whilst detailed data (measurements, DNA sequences, pathology descriptions) requires authentication and justification, allowing researchers to identify relevant datasets without exposing sensitive

information [42,43]. Community Consultation Protocols ensure that for materials from descendant communities, consultation and consent should precede data sharing decisions, potentially including community representatives on data governance boards, required consultation before data release, and community veto power over inappropriate uses [38]. Temporal Restrictions follow models like the 100-year threshold, though flexibility may be appropriate based on community wishes and research contexts [35]. Metadata-Only Records establish existence without exposure for materials where data sharing is inappropriate, preventing redundant sampling requests whilst respecting ethical constraints [16]. Progressive Data Release allows initial publications to present only aggregate statistics, with more detailed data available upon application with clear research justification, balancing immediate protection with potential future research value [10].

For Modern era bioarchaeology, these considerations prove particularly acute. Colonial period collections raise questions about consent, repatriation, and appropriate use [36]. Industrial period skeletal collections from marginalised communities (pauper burials, workhouse cemeteries, asylum grounds) require sensitivity to social inequality and dignity. Post-medieval collections may include individuals with living descendants, complicating consent and privacy considerations.

Implementing CARE alongside FAIR is not merely ethical nicety but practical necessity. Failure to respect communities erodes trust, potentially foreclosing future research opportunities entirely. Conversely, genuine partnership approaches can enhance research through community knowledge, improve interpretation through cultural context, and ensure research benefits extend beyond academic publications to communities whose ancestors generated the data [44].

4. Machine Learning and AI: Unlocking Archives Without Further Destruction

Artificial intelligence and machine learning offer transformative potential for maximising value from existing data without requiring additional destructive sampling. The finite nature of bioarchaeological materials makes this particularly crucial; every computational advance enabling knowledge extraction from existing data reduces pressure on finite physical specimens.

4.1. Natural Language Processing for Grey Literature

Natural Language Processing (NLP) and Named Entity Recognition (NER) provide solutions for extracting structured information from unstructured text, including grey literature [45]. The Archaeology Data Service has successfully deployed these technologies in multiple projects demonstrating feasibility and value.

Archaeotools (2007) was the first UK archaeological project using faceted classification and NLP to index ADS databases by monument types, analysing unpublished fieldwork reports to create enhanced searchability [46,47]. STAR (Semantic Technologies for Archaeological Resources) enhanced scholarly access through mapping to CIDOC CRM-EH ontology, establishing metadata standards enabling cross-search across documents [48]. STELLAR built upon this by creating Linked Data mapping tools enabling non-specialists to contribute data without requiring technical expertise [48]. SENESCHAL introduced standardised UK controlled vocabularies (Monument Types Thesaurus, Event Types Thesaurus, MIDAS Archaeological Periods List) transformed into Linked Open Data, increasing ability to use controlled terminology ensuring uniform annotation [49,50]. ARIADNE created international infrastructure aggregating 2,000,000 datasets from 11,000 archaeologists across Europe using rule-based text mining and machine learning [51–53].

These projects demonstrate that computational approaches can enhance grey literature accessibility whilst maintaining distributed archives, creating searchable interfaces, and preserving institutional autonomy.

4.2. Osteoarchaeological Entity Search: Current Capabilities and Limitations

Building on zooarchaeological precedents [54], the Osteoarchaeological and Palaeopathological Entity Search (OPES) project extended NLP/NER approaches to human remains data [30]. Using the

U.S. National Library of Medicine Medical Subject Headings (MESH) as controlled vocabulary, the project employed three-round annotation processes involving researchers, domain experts, and super-annotators to establish gold standards. Five Crossrail reports on osteoarchaeology, dental calculus, and stable isotopes provided training data, annotated using GATE Developer platform to create machine-readable XML files with MESH unique identifiers [30].

The resulting tool was evaluated by 83 participants (26 experts, 42 students, 15 non-archaeologists) using 7-point Likert scales. Results revealed both promise and limitations. Accessibility scored 5.69/7 with modal response of 7 (43% of respondents), indicating non-specialists found interfaces approachable. Perceived utility scored 5.21/7, suggesting imperfect tools still provide value. Students scored likelihood of future use at 5.17/7, and technical feasibility was demonstrated: NLP/NER can extract osteoarchaeological terminology from PDFs [30].

However, reliability scored only 4.79/7 with modal response of 5, reflecting substantial concerns. False positives identified terms in wrong contexts (e.g., 'fractures' identifying pottery fractures rather than bone fractures). Boolean logic errors occurred where multiple search terms broadened rather than narrowed results. Missing terms meant searches returned documents supposedly containing terms that were not actually present. Vocabulary gaps arose where MESH uses medical terminology with American spelling, missing archaeological conventions [30].

4.3. Addressing Limitations and Future Directions

These limitations, whilst significant, are addressable through systematic improvements. The OPES project used only 5 of 381 available Crossrail reports for training, severely limiting contextual examples for NER learning. Expanding training corpora to use all available reports would provide much more diverse contexts, particularly for rare terms that appeared only once or twice in initial training [30]. Vocabulary development represents another clear priority. Developing bioarchaeology-specific controlled vocabularies incorporating British English terminology, archaeological conventions, and comprehensive coverage would dramatically improve term recognition and contextual accuracy [30]. Enhanced contextual analysis represents the third major improvement area. Moving beyond simple term recognition to analyse sentence structure, identify modifying terms (e.g., 'pottery' preceding 'fracture'), and assess semantic relationships would dramatically reduce false positives [55]. Iterative refinement through user feedback mechanisms would enable ongoing improvement, with systems where researchers can flag errors creating training data for continuous enhancement, gradually improving accuracy over time [56].

Critically, NLP/NER tools need not achieve perfection to provide enormous value. If computational search reduces documents requiring manual review from 381 to 50, even with 20% false positive rates, researchers save hundreds of hours [30]. These tools serve as discovery mechanisms, filtering and flagging potentially relevant documents for expert review rather than replacing human judgement.

Beyond information extraction, machine learning offers additional applications. Predictive Modelling could identify high-potential samples for destructive analysis, maximising information gain whilst minimising material consumption, with models trained on existing datasets predicting which samples are most likely to yield well-preserved DNA, sufficient enamel for isotopic analysis, or high pathology probability [57]. Computer Vision through deep learning applied to osteological images could extract measurements, identify pathologies, assess age and sex indicators, and quantify taphonomic changes without requiring physical specimen access [58]. Automated Metadata Generation using natural language generation could automatically create standardised metadata from methodological descriptions in reports and publications, reducing documentation burden whilst ensuring completeness [59]. Data Integration and Synthesis through machine learning approaches could link disparate datasets, connecting isotopic analyses to skeletal material to excavation contexts to environmental reconstructions to historical records, creating knowledge networks where discoveries in one dataset illuminate others [60]. Quality Control through automated

systems could flag potential data quality issues (outliers, inconsistencies, unlikely values) for expert review, improving dataset reliability whilst reducing manual validation burden [61].

The key insight remains consistent: every computational advance enabling knowledge extraction from existing data reduces pressure on finite physical materials. This is not merely convenient but ethically necessary given material scarcity and the irreversible nature of destructive analyses.

5. Integration: A Roadmap for Transformation

Achieving sustainable bioarchaeological research requires integrating FAIR principles, CARE ethics, and AI/ML capabilities through coordinated action across multiple stakeholder groups. Individual researchers form the foundation of data management practice and should create ORCID accounts and use them consistently in all publications and data deposits [19], apply appropriate Creative Commons licences (CC-BY as default unless restrictions are ethically necessary) to all deposited data [62], deposit data in recognised repositories with persistent identifiers (DOIs) rather than solely in journal supplementary materials [63], document methodologies comprehensively including equipment specifications, parameters, calibration procedures, and quality control measures [16], and create Data Management Plans for all new projects using standardised templates such as PARTHENOS [16,64]. Ongoing practice should preserve both raw and processed data with clear documentation linking processing steps [26,27], use standardised file naming conventions and systematic organisation enabling future intelligibility [65,66], maintain contemporaneous documentation rather than retrospective reconstruction [28], create 3D digital models before destructive sampling where technically and financially feasible [67,68], link publications to datasets via DOIs in formal data availability statements [62,69], and consult with descendant communities before depositing or publishing data from culturally sensitive contexts [40].

Professional organisations including the British Association for Biological Anthropology and Osteoarchaeology (BABAO), Institute of Field Archaeologists (IFA), and equivalent international bodies play crucial coordination roles. For standards development, organisations should establish community-endorsed metadata schemas for bioarchaeological subdisciplines through working groups and consensus processes [16], create controlled vocabularies using British English terminology and archaeological conventions rather than medical terminology [30,70], develop ethical guidelines balancing openness with respect for human remains and descendant communities [71,72], and define minimum acceptable documentation standards as professional obligations [16]. Training and support should include offering regular professional development workshops on data management, FAIR principles, copyright and licensing, and ethical data governance [16], providing template DMPs, metadata forms, and documentation guides tailored to bioarchaeological contexts [64], recognising exemplary data practices through awards, conference sessions, and professional highlighting [61], and fostering community norms explicitly valuing data sharing and reuse as professional virtues [73]. Advocacy efforts should engage with research funders to support appropriate data management resources in grant allocations [74], work with publishers to establish consistent, feasible data policies that balance requirements with practical implementation [75], promote data management as core professional competency equivalent to analytical skills [76], and support infrastructure development including repositories, search tools, and controlled vocabulary maintenance [16].

Repositories including the Archaeology Data Service (ADS), The Digital Archaeological Record (tDAR), Open Context, and institutional repositories must enhance services and coordination. Enhanced services should provide comprehensive documentation, training materials, and examples specific to bioarchaeological data types [63], offer active data deposit support services beyond self-service portals, including consultation on formats, metadata, and documentation [16], develop bioarchaeology-specific metadata schemas beyond generic archaeology, incorporating subdiscipline requirements [16], and implement user-friendly interfaces requiring minimal technical expertise whilst maintaining sophisticated backend functionality [77]. Federation and coordination requires

registering endpoints centrally (e.g., E-RIHS DIGILAB) with clearly documented access protocols [16,78], implementing OAI-PMH or similar harvesting protocols enabling metadata aggregation [79,80], collaborating on unified search interfaces enabling discovery across multiple repositories [81,82], and developing Linked Open Data capabilities using CIDOC CRM ontologies enabling dataset interlinking [83,84]. Tool development should advance NLP/NER tools for information extraction from grey literature with community input on vocabulary and priorities [30,54], implement user feedback mechanisms enabling continuous improvement and error correction [56], provide confidence scores and transparent limitations for computational tools rather than false precision [85], and enable community contributions to controlled vocabularies and training data through accessible interfaces [61].

Research funders including Research Councils UK, European Research Council, National Science Foundation, and equivalent bodies internationally must enforce requirements and provide resources. Mandate requirements should require comprehensive DMPs as standard components of all grant applications, with evaluation as peer review criteria [74,76], allocate specific percentages of grant budgets (e.g., 5–10%) explicitly for data management activities [16], require repository deposition with PIDs as condition for final grant payments, with verification before closure [63], and audit compliance with data management requirements and consider in future funding decisions [69]. Provision of support should fund repository development, maintenance, and enhancement to ensure sustainable infrastructure [86], support creation and maintenance of disciplinary standards, controlled vocabularies, and metadata schemas [16], enable retrospective curation of legacy data through targeted funding programmes [87], and fund training initiatives in data management at graduate and professional development levels [76]. Recognition efforts should evaluate deposited datasets as research outputs in grant assessments and impact case studies [62], highlight data sharing as merit criterion in proposal review processes [69], support data papers describing significant datasets as legitimate publication types [88], and recognise infrastructure contributions (repository development, standards creation, tool building) as fundable research activities [86].

Academic publishers including MDPI (Heritage journal), Cambridge University Press, Wiley, Elsevier, and others must establish and enforce appropriate policies. Policy requirements should require data availability statements in all bioarchaeology articles specifying deposition location, access conditions, and persistent identifiers [62,69], mandate data deposition in recognised repositories before publication acceptance, not merely at proof stage [75], verify data accessibility during peer review as standard editorial procedure [69], and provide clear, consistent guidance on acceptable repositories, formats, and access conditions [16]. Infrastructure support should assign persistent identifiers (DOIs) to supplementary datasets enabling formal citation [89], enable data citations contributing to author impact metrics and institutional assessments [62], publish data papers describing significant datasets as distinct article type [88], and consider waiving or reducing article processing charges for papers documenting valuable datasets without new analyses [75].

Universities and research institutes must provide infrastructure, training, and recognition. Infrastructure provision requires providing repository options (institutional and/or disciplinary) with long-term preservation commitments [90,91], offering data management support through libraries, IT services, or dedicated data steward positions [76], ensuring adequate storage and preservation capabilities for large bioarchaeological datasets [92], and supporting computational infrastructure for analysis, curation, and tool development [16]. Training integration should embed data management in graduate curricula as core competency requirements, not optional extras [76], require DMPs for theses and dissertations with assessment as examination criteria [74], provide workshops and resources for faculty, researchers, and students at all career stages [16], and include data management in research integrity training as ethical obligation [9]. Career recognition demands counting datasets explicitly in hiring, promotion, and tenure decisions using defined evaluation criteria [62], allocating workload credit for data curation activities including legacy data improvement [76], highlighting datasets in institutional research showcases, press releases, and

impact communications [91], and supporting data publication with equivalent prestige to traditional article publication [88].

Commercial units conducting developer-funded archaeology must adopt standardised practices within business constraints. Standards adoption requires implementing minimum metadata and documentation standards as standard operating procedures [16], using standardised recording systems enabling data aggregation across projects and organisations [63], applying persistent identifiers and clear licences to all deposited data [62], and including data management costs explicitly in project budgets and client contracts [74]. Data deposition should deposit data in recognised repositories (e.g., ADS) beyond delivery to immediate clients [86], ensure grey literature is indexed and discoverable through appropriate metadata [30], link specialist bioarchaeological reports to field data through persistent identifiers [16], and make data accessible within ethical and contractual constraints using tiered access where necessary [41]. Staff development requires ensuring staff training in data management best practices as professional development priority [76], providing dedicated time and resources for proper documentation rather than treating as overhead [74], recognising data quality in performance evaluations and career progression [61], and fostering organisational culture valuing long-term preservation over merely project delivery [63].

6. Long-Term Vision and Future Directions

Imagining bioarchaeological research in 2035 provides concrete targets for transformation. In this envisioned future, every sample extracted receives a persistent identifier at collection, creating an unbroken digital thread linking subsequent analyses. Comprehensive Data Management Plans guide research from conception through archiving, with templates tailored to bioarchaeological contexts addressing ethical considerations, descendant community consultation, and technical requirements. Data deposits occur in federated repositories with rich metadata following community-endorsed schemas, with bioarchaeologists choosing appropriate repositories based on project needs, with all endpoints registered centrally enabling unified search.

Artificial intelligence tools monitor new deposits, automatically extracting and linking information. Researchers query federated search portals discovering relevant data regardless of deposition location. Natural Language Processing systems analyse grey literature as reports are deposited, creating searchable metadata linking osteoarchaeological, zooarchaeological, archaeobotanical, and other specialist findings to fieldwork contexts. Machine learning models suggest relevant datasets based on research questions, predict sample potential before destructive analysis, and flag potential quality issues for expert review.

Linked Open Data networks connect disparate datasets using CIDOC CRM ontologies and controlled vocabularies. Isotopic analyses link to skeletal material, which links to excavation context, which links to zooarchaeological data, which links to environmental reconstructions, which links to historical records. This creates knowledge networks where discoveries in one dataset illuminate others, multiplying research value and enabling synthetic analyses impossible with isolated datasets.

Crucially, ethical frameworks balance openness with respect. CARE principles ensure Indigenous data sovereignty with community representatives on governance boards for culturally sensitive collections. Human remains data receives appropriate protection through tiered access systems enabling discovery through Open Access metadata whilst restricting detailed data to authenticated researchers with clear justification. Descendant community consultation precedes data sharing decisions for post-medieval and colonial period collections. International coordination accommodates diverse legal and cultural contexts through flexible frameworks rather than imposed uniformity.

For Modern era bioarchaeology specifically, this transformation enables synthetic analyses previously impossible. Understanding health transitions during industrialisation requires aggregating data from hundreds of cemetery excavations currently siloed in grey literature. Tracing disease emergence in colonial contexts requires linking skeletal evidence, historical records, environmental data, and zooarchaeological findings. Assessing biological impacts of globalisation

requires international dataset integration currently prevented by incompatible formats and inadequate metadata. The envisioned infrastructure, standards, and tools make these synthetic analyses feasible.

This vision is achievable. Ancient DNA research demonstrates comprehensive FAIR implementation success through appropriate infrastructure, training, community norms, and institutional support. The challenge involves scaling this across all bioarchaeological subdisciplines whilst accommodating unique requirements. Palaeoproteomics is following this trajectory. Stable isotope research shows promising developments through IsoMemo and similar initiatives [93]. Osteoarchaeology, zooarchaeology, palaeopathology, and other specialisms require dedicated effort but face no insurmountable barriers.

7. Conclusions

Bioarchaeological research stands at a critical juncture. Materials under study are finite and irreplaceable, with every analysis consuming portions of this precious resource. Meanwhile, data volumes expand exponentially through technological advances, yet data management practices lag dangerously behind. This situation is unsustainable and ethically problematic.

Current practices fail on multiple dimensions. Most data is not Findable (fewer than 50% use persistent identifiers), not fully Accessible (scattered across journals rather than repositories), not Interoperable (fewer than 50% believe appropriate metadata schemas exist), and consequently not Reusable (systematic documentation lacking). Only ancient DNA research consistently meets FAIR criteria, benefiting from molecular biology's established data culture. Grey literature containing invaluable information remains computationally inaccessible, requiring manual review that proves impossible at scale. Ethical frameworks for human remains data lack consistency, with some specialisms showing 0% Open Access whilst appropriate balancing mechanisms remain underdeveloped.

Three integrated commitments are essential for sustainable bioarchaeological research. First, universal adoption of FAIR principles with appropriate infrastructure, training, and incentives, requiring persistent identifiers for all data, deposition in recognised repositories rather than journal supplements, community-endorsed metadata schemas enabling computational discovery, controlled vocabularies using appropriate terminology, clear licensing specifying usage conditions, systematic documentation of methodologies and provenance, and Data Management Plans guiding research from conception through archiving. Ancient DNA and palaeoproteomics demonstrate this is achievable; other subdisciplines must follow.

Second, implementation of CARE principles ensuring ethical data governance, requiring recognising collective benefit to descendant communities, respecting community authority to control data from ancestors, accepting researcher responsibility for appropriate use, and prioritising ethics over pure openness. Practical implementations include tiered access systems, community consultation protocols, metadata-only records for sensitive materials, and temporal restrictions where appropriate. The formulation 'as open as possible, as closed as necessary' captures this essential balance.

Third, strategic development of artificial intelligence and machine learning tools for knowledge extraction. Natural Language Processing and Named Entity Recognition can unlock grey literature archives, though current implementations show only moderate success requiring improvements in training data, vocabularies, and contextual analysis. Beyond information extraction, machine learning enables predictive modelling for sample selection, computer vision for non-invasive analysis, automated metadata generation, and data integration creating knowledge networks. Every computational advance reduces pressure on finite physical materials.

These three commitments are interdependent. FAIR without CARE risks ethical violations alienating communities and researchers. CARE without FAIR limits research value from ethically obtained data. AI/ML without FAIR cannot function (requiring standardised formats and metadata),

whilst FAIR without AI/ML leaves masses of data computationally inaccessible. Integration is essential.

The roadmap for transformation requires coordinated action from researchers, professional organisations, repositories, funders, publishers, institutions, and commercial organisations. No single stakeholder can succeed alone; collective action is necessary. The finite nature of bioarchaeological materials makes this transformation urgent rather than merely desirable. Every sample destroyed without proper data preservation represents irreversible loss of knowledge about human heritage. Every dataset published without appropriate documentation limits future reanalysis when new questions and methods emerge. Every grey literature report left computationally inaccessible contains information that may never be rediscovered.

For Modern era bioarchaeology specifically (1492–1945 CE), transformation opens unprecedented opportunities. Understanding health transitions during industrialisation, disease emergence in colonial contexts, biological impacts of globalisation, and biocultural adaptations to rapid change requires synthetic analyses currently prevented by data inaccessibility. Industrial period skeletal collections, post-medieval cemetery excavations, colonial context investigations - these represent rich but under-utilised resources that proper data management could unlock.

The bioarchaeological community has demonstrated commitment, with 97% recognising data reuse as critically important. The challenge involves translating commitment into practice through infrastructure, standards, training, incentives, and tools. Ancient DNA research proves comprehensive implementation is achievable. The task ahead involves scaling this success across all subdisciplines whilst accommodating unique requirements and ethical considerations. The materials studied are finite. The knowledge they contain need not be. Through excellent data stewardship integrating FAIR principles, CARE ethics, and AI/ML capabilities, research can ensure every consumed sample generates maximum lasting value, existing knowledge remains accessible for future generations asking unanticipated questions, and practice honours the significance of irreplaceable materials through responsible data governance.

Funding: This work was supported by the Arts and Humanities Research [grant number AH/W002469/1].

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

aDNA	Ancient DNA
ADS	Archaeology Data Service
AI	Artificial Intelligence
ARIADNE	Advanced Research Infrastructure for Archaeological Dataset Networking in Europe
BABAO	British Association for Biological Anthropology and Osteoarchaeology
CARE	Collective benefit, Authority, Responsibility, Ethics
CC-BY	Creative Commons Attribution
CE	Common Era
CIDOC CRM	Comité International pour la Documentation Conceptual Reference Model
CIDOC CRM-EH	CIDOC Conceptual Reference Model English Heritage Extension
DMP	Data Management Plan
DNA	Deoxyribonucleic Acid
DOCX	Microsoft Word Document Format
DOI	Digital Object Identifier
E-RIHS	European Research Infrastructure for Heritage Science
FAIR	Findable, Accessible, Interoperable, Reusable
GATE	General Architecture for Text Engineering
IFA	Institute of Field Archaeologists
IT	Information Technology
JPEG	Joint Photographic Experts Group
MeSH	Medical Subject Headings

ML	Machine Learning
NCBI	National Center for Biotechnology Information
NER	Named Entity Recognition
NLP	Natural Language Processing
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OPES	Osteoarchaeological and Palaeopathological Entity Search
ORCID	Open Researcher and Contributor ID
PDF	Portable Document Format
PID	Persistent Identifier
SENECHAL	Semantic ENrichment Enabling Sustainability of arCHAEological Links
STAR	Semantic Technologies for Archaeological Resources
STELLAR	Semantic Technologies Enhancing Links and Linked data for Archaeological Resources
tDAR	The Digital Archaeological Record
UK	United Kingdom
US/USA	United States of America
XLSX	Microsoft Excel Spreadsheet Format
XML	Extensible Markup Language
ZooMS	Zooarchaeology by Mass Spectrometry

References

1. Brown, T.A.; Brown, K. *Biomolecular Archaeology: An Introduction*; Wiley-Blackwell: Chichester, UK, 2011.
2. Oakley, K. Forensic archaeology and anthropology: An Australian perspective. *Forensic Sci. Med. Pathol.* **2015**, *1*, 169–172. <https://doi.org/10.1385/fsmp:1:3:169>
3. Hendy, J.; Welker, F.; Demarchi, B.; Speller, C.; Warinner, C.; Collins, M.J. A guide to ancient protein studies. *Nat. Ecol. Evol.* **2018**, *2*, 791–799. <https://doi.org/10.1038/s41559-018-0510-x>
4. Matisoo-Smith, E. Ancient DNA and the human settlement of the Pacific: A review. *J. Hum. Evol.* **2015**, *79*, 93–104. <https://doi.org/10.1016/j.jhevol.2014.10.017>
5. Baker, O.; Worley, F. *Animal Bones and Archaeology: Guidelines for Best Practice*; English Heritage: Swindon, UK, 2014.
6. Doorn, N.L. Zooarchaeology by mass spectrometry (ZooMS). In *Encyclopedia of Global Archaeology*; Smith, C., Ed.; Springer: New York, NY, USA, 2015; pp. 7998–8000.
7. Pálsdóttir, A.H.; Bläuer, A.; Rannamäe, E.; Boessenkool, S.; Hallsson, J.H. Not a limitless resource: ethics and guidelines for destructive sampling of archaeofaunal remains. *R. Soc. Open Sci.* **2019**, *6*, 191059. <https://doi.org/10.1098/rsos.191059>
8. Fosshem, H.J. Introductory remarks. In *More than Just Bones: Ethics and Research on Human Remains*; Fosshem, H.J., Ed.; The Norwegian National Research Ethics Committees: Oslo, Norway, 2013; pp. 7–10.
9. Fox, K. The illusion of inclusion—The 'All of Us' research program and indigenous peoples' DNA. *N. Engl. J. Med.* **2020**, *383*, 411–413. <https://doi.org/10.1056/NEJMp191598>
10. Ulguim, P. Models and metadata: the ethics of sharing bioarchaeological 3D models online. *Archaeologies* **2018**, *14*, 189–228. <https://doi.org/10.1007/s11759-018-9346-x>
11. He, K.Y.; Ge, D.; He, M.M. Big data analytics for genomic medicine. *Int. J. Mol. Sci.* **2017**, *18*, 142. <https://doi.org/10.3390/ijms18020412>
12. Katzenberg, M.A.; Waters-Rist, A.L. Stable isotope analysis. In *Biological Anthropology of the Human Skeleton*; Katzenberg, M.A., Grauer, A.L., Eds.; John Wiley & Sons: Hoboken, NJ, USA, 2018; pp. 467–504.
13. Hendy, J.; Welker, F.; Demarchi, B.; Speller, C.; Warinner, C.; Collins, M.J. A guide to ancient protein studies. *Nat. Ecol. Evol.* **2018**, *2*, 791–799. <https://doi.org/10.1038/s41559-018-0510-x>
14. Sagioglu, S.; Sinanc, D. Big data: a review. In *Proceedings of the International Conference on Collaboration Technologies and Systems (CTS)*; Smari, W.W., Fox, G.C., Eds.; IEEE: San Diego, CA, USA, 2013; pp. 42–47.
15. Pálsdóttir, A.H.; Bläuer, A.; Rannamäe, E.; Boessenkool, S.; Hallsson, J.H. Not a limitless resource: ethics and guidelines for destructive sampling of archaeofaunal remains. *R. Soc. Open Sci.* **2019**, *6*, 191059. <https://doi.org/10.1098/rsos.191059>
16. Wright, H.; Richards, J. *D.5.3 Data Curation Policy*; E-RIHS: York, UK, 2020. <https://doi.org/10.1080/00934690.2018.1511960>

17. Król, K.; Zdonek, D. Peculiarity of the bit rot and link rot phenomena. *Glob. Knowl. Mem. Commun.* **2019**, *69*, 20–37. <https://doi.org/10.1108/GKMC-06-2019-0067>
18. Wright, H.; Richards, J.D. Reflections on collaborative archaeology and large-scale online research infrastructures. *J. Field Archaeol.* **2018**, *43*, S60–S67. <https://doi.org/10.1080/00934690.2018.1511960>
19. Fenner, M.; Gómez, C.G.; Thorisson, G.A. Collective Action for the Open Researcher & Contributor ID (ORCID). *Serials* **2011**, *24*, 277–279. <https://doi.org/10.1629/24277>
20. Baessa, M.; Lery, T.; Grenz, D.; Vijayakumar, J.K. Connecting the pieces: Using ORCIDs to improve research impact and repositories. *F1000Research* **2015**, *4*. <https://doi.org/10.12688/f1000research.6502.1>
21. Kansa, S.W.; Atici, L.; Kansa, E.C.; Meadow, R.H. Archaeological analysis in the information age: guidelines for maximizing the reach, comprehensiveness, and longevity of data. *Adv. Archaeol. Pract.* **2020**, *8*, 40–52. <http://dx.doi.org/10.1017/aap.2019.36>
22. Sobotkova, A. Sociotechnical obstacles to archaeological data reuse. *Adv. Archaeol. Pract.* **2018**, *6*, 117–124. <https://doi.org/10.1017/aap.2017.37>
23. Kulasekaran, S.; Trelogan, J.; Esteva, M.; Johnson, M. Metadata integration for an archaeology collection architecture. In *International Conference on Dublin Core and Metadata Applications*; Moen, W., Rushing, A., Eds.; Austin, TX, USA, 2014; pp. 53–63.
24. Kintigh, K.W.; Altschul, J.H.; Beaudry, M.C.; Drennan, R.D.; Kinzig, A.P.; Kohler, T.A.; Limp, W.F.; Maschner, H.D.; Michener, W.K.; Pauketat, T.R.; et al. Grand challenges for archaeology. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 879–880. <https://doi.org/10.1073/pnas.1324000111>
25. Evans, T.N.; Moore, R.H. The use of PDF/A in digital archives: a case study from archaeology. *Int. J. Digit. Curation* **2014**, *9*, 123–138. <https://doi.org/10.2218/ijdc.v9i2.267>
26. Huggett, J. Reuse remix recycle: repurposing archaeological digital data. *Adv. Archaeol. Pract.* **2018**, *6*, 93–104. <https://doi.org/10.1017/aap.2018.1>
27. Hart, E.M.; Barmby, P.; LeBauer, D.; Michonneau, F.; Mount, S.; Mulrooney, P.; Poisot, T.; Woo, K.H.; Zimmerman, N.B.; Hollister, J.W. Ten Simple Rules for Digital Data Storage. *PLoS Comput. Biol.* **2016**, *12*, e1005097. <https://doi.org/10.1371/journal.pcbi.1005097>
28. Reiser, L.; Harper, L.; Freeling, M.; Han, B.; Luan, S. FAIR: a call to make published data more Findable, Accessible, Interoperable, and Reusable. *Mol. Plant* **2018**, *11*, 1105–1108. <https://doi.org/10.1016/j.molp.2018.07.005>
29. Tillett, S.; Newbold, E. Grey literature at The British Library: revealing a hidden resource. *Interlend. Doc. Supply* **2006**, *34*, 70–73. <https://doi.org/10.1108/02641610610669769>
30. Talks, S.A. Osteoarchaeological and Palaeopathological Entity Search (OPES): The Use of Natural Language Processing and Named Entity Recognition for Bioarchaeology. MA Thesis, University of York, York, UK, 2019.
31. Paris, R.; Myatt, C.; de Silva, M. Crossrail project: environmental management during delivery of London's Elizabeth line. In *Civil Engineering, Proceedings of the Institution of Civil Engineers*; Linnell, E., Ed.; ICE Publishing: London, UK, 2017; pp. 49–55. <https://doi.org/10.1680/jcien.17.00021>
32. Green, E.D.; Rubin, E.M.; Olson, M.V. The future of DNA sequencing. *Nature* **2017**, *550*, 179. <https://doi.org/10.1038/550179a>
33. Vizcaíno, J.A.; Deutsch, E.W.; Wang, R.; Csordas, A.; Reisinger, F.; Rios, D.; Dianes, J.A.; Sun, Z.; Farrah, T.; Bandeira, N.; et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **2014**, *32*, 223–226. <https://doi.org/10.1038/nbt.2839>
34. Gupta, N.; Martindale, A.; Supernant, K.; Elvidge, M. The CARE principles and the reuse, sharing, and curation of indigenous data in Canadian archaeology. *Adv. Archaeol. Pract.* **2023**, *11*, 76–89. <https://doi.org/10.1017/aap.2022.33>
35. Payne, S. Archaeology and human remains: Handle with care! Recent English experiences. In *More than Just Bones: Ethics and Research on Human Remains*; Fossheim, H.J., Ed.; The Norwegian National Research Ethics Committees: Oslo, Norway, 2013; pp. 49–64.
36. Prendergast, M.E.; Sawchuk, E. Boots on the ground in Africa's ancient DNA 'revolution': archaeological perspectives on ethics and best practices. *Antiquity* **2018**, *92*, 803–815. <https://doi.org/10.15184/aqy.2018.70>

37. Bardill, J. Native American DNA: Ethical, legal, and social implications of an evolving concept. *Annu. Rev. Anthropol.* **2014**, *43*, 155–166. <https://dx.doi.org/10.1146/annurev-anthro-092412-155504>
38. Matisoo-Smith, E. Working with indigenous communities in genomic research: a Pacific perspective. *Soc. Am. Archaeol. Archaeol. Rec.* **2019**, *19*, 14–19.
39. Lynnerup, N. The ethics of destructive bone analyses (with examples from Denmark and Greenland). In *More than Just Bones: Ethics and Research on Human Remains*; Fossheim, H.J., Ed.; The Norwegian National Research Ethics Committees: Oslo, Norway, 2013; pp. 81–94.
40. Carroll, S.; Garba, I.; Figueroa-Rodríguez, O.; Holbrook, J.; Lovett, R.; Materechera, S.; Parsons, M.; Raseroka, K.; Rodriguez-Lonebear, D.; Rowe, R.; et al. The CARE principles for indigenous data governance. *Data Sci. J.* **2020**, *19*, 43. <https://doi.org/10.5334/dsj-2020-043>
41. Landi, A.; Thompson, M.; Giannuzzi, V.; Bonifazi, F.; Labastida, I.; da Silva Santos, L.O.B.; Roos, M. The "A" of FAIR—as open as possible, as closed as necessary. *Data Intell.* **2020**, *2*, 47–55. https://doi.org/10.1162/dint_a_00027
42. Browning, S.; Guédon, J.C.; Kaplan, L. Metadata and Open Access: reliably finding content and finding reliable content. In *Proceedings of the Charleston Library Conference*; Bernhardt, B.R., Hinds, L.H., Strauch, K.P., Eds.; Charleston, SC, USA, 2014. <http://dx.doi.org/10.5703/1288284315314>
43. Cheby, L.E. Open Access metadata for journals in directory of Open Access journals: who, how, and what scheme? *Sch. Inf. Stud. Res. J.* **2016**, *6*, 4. Available online: <https://scholarworks.sjsu.edu/cgi/viewcontent.cgi?article=1247&context=ischoolsrj> (accessed on 1 December 2025).
44. Atalay, S. Indigenous Archaeology as Decolonizing Practice. *Am. Indian Q.* **2006**, *30*, 280–310. <http://www.jstor.org/stable/4139016>
45. Richards, J.; Jeffrey, S.; Waller, S.; Ciravegna, F.; Chapman, S.; Zhang, Z. The Archaeology Data Service and the Archaeotools Project: Faceted Classification and Natural Language Processing. In *Archaeology 2.0 New Approaches to Communication & Collaboration*; Kansa, E., Whitcher Kansa, S., Watrall, E., Eds.; Cotsen Digital Archaeology: Los Angeles, CA, USA, 2011; pp. 31–56.
46. Richards, J.; Jeffrey, S.; Waller, S.; Ciravegna, F.; Chapman, S.; Zhang, Z. The Archaeology Data Service and the Archaeotools Project: Faceted Classification and Natural Language Processing. In *Archaeology 2.0 New Approaches to Communication & Collaboration*; Kansa, E., Whitcher Kansa, S., Watrall, E., Eds.; Cotsen Digital Archaeology: Los Angeles, CA, USA, 2011; pp. 31–56.
47. Jeffrey, S.; Richards, J.; Ciravegna, F.; Waller, S.; Chapman, S.; Zhang, Z. The Archaeotools project: faceted classification and natural language processing in an archaeological context. *Philos. Trans. R. Soc. A* **2009**, *367*, 2507–2519. <https://www.jstor.org/stable/40485597>
48. Tudhope, D.; Binding, C.; Jeffrey, S.; May, K.; Vlachidis, A. A STELLAR role for knowledge organization systems in digital archaeology. *Bull. Assoc. Inf. Sci. Technol.* **2011**, *37*, 15–18.
49. Binding, C.; Tudhope, D. Improving interoperability using vocabulary linked data. *Int. J. Digit. Libr.* **2016**, *17*, 5–21. <https://doi.org/10.1007/s00799-015-0166-y>
50. May, K.; Binding, C.; Tudhope, D. Barriers and opportunities for linked open data use in archaeology and cultural heritage. *Archäol. Inf.* **2015**, *38*, 173–184.
51. Niccolucci, F. The ARIADNEplus integration of archaeological datasets. In *Cultural Heritage and New Technologies 24*; Vienna, Austria, 2019.
52. Meghini, C.; Scopigno, R.; Richards, J.; Wright, H.; Geser, G.; Cuy, S.; Fihn, J.; Fanini, B.; Hollander, H.; Niccolucci, F.; et al. ARIADNE: A Research Infrastructure for Archaeology. *J. Comput. Cult. Herit.* **2017**, *10*, 1–27.
53. Aloia, N.; Binding, C.; Cuy, S.; Doerr, M.; Fanini, B.; Felicetti, A.; Fihn, J.; Gavriliis, D.; Geser, G.; Hollander, H.; et al. Enabling European Archaeological Research: The ARIADNE E-infrastructure. *Internet Archaeol.* **2017**, *43*. <https://doi.org/10.11141/ia.43.11>
54. Talboom, L. Improving the Discoverability of Zooarchaeological Data with the Help of Natural Language Processing. MSc Thesis, University of York, York, UK, 2017.

55. Richards, J.D.; Tudhope, D.; Vlachidis, A. Text Mining in Archaeology: Extracting Information from Archaeological Reports. In *Mathematics in Archaeology*; Barcelo, J., Bogdanovic, I., Eds.; CRC Press-Taylor & Francis Group: Boca Raton, FL, USA, 2015; pp. 240–254. <https://doi.org/10.1201/b18530-15>
56. Kansa, E.C.; Whitcher Kansa, S.; Arbuckle, B. Publishing and pushing: mixing models for communicating research data in archaeology. *Int. J. Digit. Curation* **2014**, *9*, 57–70. <https://doi.org/10.2218/ijdc.v9i1.301>
57. Dolle, D.; Fages, A.; Mata, X.; Schiavinato, S.; Tonasso-Calvière, L.; Chauvey, L.; Wagner, S.; Der Sarkissian, C.; Fromentier, A.; Seguin-Orlando, A.; et al. CASCADE: A Custom-Made Archiving System for the Conservation of Ancient DNA Experimental Data. *Front. Ecol. Evol.* **2020**, *8*, 185. <https://doi.org/10.3389/fevo.2020.00185>
58. Nelson, A.J.; Wade, A.D. IMPACT: development of a radiological mummy database. *Anat. Rec.* **2015**, *298*, 941–948. <https://doi.org/10.1002/ar.23130>
59. Raval, P.; Bhaidasna, H. A Review of Extracting Metadata from Scholarly Articles using Natural Language Processing (NLP). In *2024 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS)*; IEEE: Tirunelveli, India, 2024; pp. 1355–1359. <https://doi.org/10.1109/ICACRS62842.2024.10841656>
60. Marchetti, N.; Angelini, I.; Artioli, G.; Benati, G.; Bitelli, G.; Curci, A.; Marfia, G.; Roccetti, M. NEARCHOS. Networked archaeological open science: advances in archaeology through field analytics and scientific community sharing. *J. Archaeol. Res.* **2018**, *26*, 447–469. <https://doi.org/10.1007/s10814-017-9112-4>
61. Costello, M.J.; Michener, W.K.; Gahegan, M.; Zhang, Z.Q.; Bourne, P.E. Biodiversity data should be published, cited, and peer reviewed. *Trends Ecol. Evol.* **2013**, *28*, 454–461. <https://doi.org/10.1016/j.tree.2013.05.002>
62. Kansa, E. Openness and archaeology's information ecosystem. *World Archaeol.* **2012**, *44*, 498–520. <https://doi.org/10.1080/00438243.2012.737575>
63. Richards, J.; Hardman, C. *Collections Policy*; Archaeology Data Service: York, UK, 2019. Available online: <https://archaeologydataservice.ac.uk/advice/collectionsPolicy.xhtml> (accessed on 1 December 2025).
64. Di Giorgio, S.; Ronzino, P. PARTHENOS Data Management Plan template for Open Research in archaeology. In Proceedings of the 3rd Digital Heritage International Congress (DigitalHERITAGE) held jointly with 2018 24th International Conference on Virtual Systems & Multimedia (VSM 2018); Addison, A.C., Thwaites, H., Eds.; IEEE: San Francisco, CA, USA, 2018. <https://doi.org/10.1109/digitalheritage.2018.8810039>
65. Han, Y. The road to digital: building unique Afghanistan collections. *OCLC Syst. Serv.* **2010**, *26*, 46–57. <https://doi.org/10.1108/10650751011018509>
66. Marwick, B.; Birch, S.E.P. A standard for the scholarly citation of archaeological data as an incentive to data sharing. *Adv. Archaeol. Pract.* **2018**, *6*, 125–143. <https://doi.org/10.1017/aap.2018.3>
67. Evin, A.; Lebrun, R.; Durocher, M.; Ameen, C.; Larson, G.; Sykes, N. Building three-dimensional models before destructive sampling of bioarchaeological remains: a comment to Pálsdóttir et al. (2019). *R. Soc. Open Sci.* **2020**, *7*, 192034. <https://doi.org/10.1098/rsos.192034>
68. Ulguim, P.F. Recording in situ human remains in three dimensions: applying digital image-based modeling. In *Human Remains: Another Dimension*; Erickson, D., Thompson, T., Eds.; Academic Press, Elsevier: London, UK, 2017; pp. 71–92. <http://dx.doi.org/10.1016/B978-0-12-804602-9.00007-2>
69. Poole, A.H. How has your science data grown? Digital curation and the human factor: a critical literature review. *Arch. Sci.* **2015**, *15*, 101–139.
70. Labrador, A.M. Ontologies of the future and interfaces for all: Archaeological databases for the twenty-first century. *Archaeologies* **2012**, *8*, 236–249. <https://doi.org/10.1007/s11759-012-9203-2>
71. Mays, S.; Elders, J.; Humphrey, L.; White, W.; Marshall, P. *Science and the Dead: A Guideline for the Destructive Sampling of Archaeological Human Remains for Scientific Analysis*; English Heritage and Advisory Panel on the Archaeology of Burials in England: Swindon, UK, 2013.
72. Hagelberg, E. Analysis of DNA from bone: Benefits versus losses. In *More than Just Bones: Ethics and Research on Human Remains*; Fossheim, H.J., Ed.; The Norwegian National Research Ethics Committees: Oslo, Norway, 2013; pp. 95–112.

73. Kansa, E.C.; Kansa, S.W. Open Archaeology: we all know that a 14 is a sheep: data publication and professionalism in archaeological communication. *J. East. Mediterr. Archaeol. Herit. Stud.* **2013**, *1*, 88–97.
74. Digital Preservation Coalition. *DPC Technology Watch Publications*. Available online: <https://www.dpconline.org/digipres/discover-good-practice/tech-watch-reports> (accessed on 1 December 2025).
75. Morgan, C.; Eve, S. DIY and digital archaeology: what are you doing to participate? *World Archaeol.* **2012**, *44*, 521–537. <http://www.jstor.org/stable/42003547>
76. Michener, W.K. Ten simple rules for creating a good data management plan. *PLoS Comput. Biol.* **2015**, *11*, e1004525. <https://doi.org/10.1371/journal.pcbi.1004525>
77. Faniel, I.M.; Barrera-Gomez, J.; Kriesberg, A.; Yakel, E. A comparative study of data reuse among quantitative social scientists and archaeologists. In *iConference 2013 Proceedings*; Schamber, L., Ed.; Fort Worth, TX, USA, 2013. <https://doi.org/10.9776/13391>
78. Pezzati, L.; Felicetti, A. DIGILAB: A new infrastructure for heritage science. In *ERCIM News. Special Theme Digital Humanities*; Bruseker, G., Kovács, L., Niccolucci, F., Eds.; **2017**, *111*. Available online: <https://ercim-news.ercim.eu/images/stories/EN111/EN111-web.pdf> (accessed on 1 December 2025).
79. Haslhofer, B.; Schandl, B. Interweaving OAI-PMH data sources with the Linked Data Cloud. *Int. J. Metadata Semant. Ontol.* **2010**, *5*, 17–31.
80. Devarakonda, R.; Palanisamy, G.; Green, J.M.; Wilson, B.E. Data sharing and retrieval using OAI-PMH. *Earth Sci. Inform.* **2011**, *4*, 1–5. <https://doi.org/10.1007/s12145-010-0073-0>
81. Klerkx, J.; Vandeputte, B.; Parra, G.; Santos, J.L.; Van Assche, F.; Duval, E. How to share and reuse learning resources: the ARIADNE experience. In *European Conference on Technology Enhanced Learning*; Goodyear, P., Retalis, S., Eds.; Springer: Berlin/Heidelberg, Germany, 2010. https://doi.org/10.1007/978-3-642-16020-2_13
82. Ternier, S.; Verbert, K.; Parra, G.; Vandeputte, B.; Klerkx, J.; Duval, E.; Ochoa, X. The Ariadne infrastructure for managing and storing metadata. *IEEE Internet Comput.* **2009**, *13*, 18–25. https://ui.adsabs.harvard.edu/link_gateway/2009IIC....13d..18T/doi:10.1109/MIC.2009.90
83. Isaksen, L. *Archaeology and the Semantic Web*. PhD Thesis, University of Southampton, Southampton, UK, 2011.
84. Binding, C.; May, K.; Tudhope, D. Semantic interoperability in archaeological datasets: Data mapping and extraction via the CIDOC CRM. In *International Conference on Theory and Practice of Digital Libraries*; Christensen-Dalsgaard, B., Castelli, D., Ammitzbøll Jurik, B., Lippincott, J., Eds.; Aarhus, Denmark, 2008.
85. Haklay, M. *Citizen Science and Policy: A European Perspective*; Science and Technology Innovation Program, Woodrow Wilson International Center for Scholars: Washington, DC, USA, 2015. Available online: https://www.wilsoncenter.org/sites/default/files/media/documents/publication/Citizen_Science_Policy_European_Perspective_Haklay.pdf (accessed on 1 December 2025).
86. Richards, J.D. Twenty years preserving data: a view from the United Kingdom. *Adv. Archaeol. Pract.* **2017**, *5*, 227–237. <https://doi.org/10.1017/aap.2017.11>
87. Clarke, R. Big Data, Big Risks. *Inf. Syst. J.* **2016**, *26*, 77–90. <https://doi.org/10.1111/isj.12088>
88. Chavan, V.; Penev, L. The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC Bioinform.* **2011**, *12* (Suppl. 15), S2. <https://doi.org/10.1186/1471-2105-12-S15-S2>
89. Paskin, N. Digital object identifiers for scientific data. *Data Sci. J.* **2005**, *4*, 12–20. <https://doi.org/10.2481/dsj.4.12>
90. Proudfoot, R. The White Rose Consortium ePrints Repository: creating a shared institutional repository for the Universities of Leeds, Sheffield and York. *Aliss Q.* **2005**, *1*, 19–23.
91. Covey, D.T. Recruiting content for the institutional repository: the barriers exceed the benefits. *J. Digit. Inf.* **2011**, *12*. <https://doi.org/10.1184/R1/6472595>

92. Lau, H.; Kansa, S.W. Zooarchaeology in the era of big data: Contending with interanalyst variation and best practices for contextualizing data for informed reuse. *J. Archaeol. Sci.* **2018**, *95*, 33–39. <https://doi.org/10.1016/j.jas.2018.03.011>
93. Schneider, L.; Haberle, S. Global markers of the anthropocene—workshop report. *Quat. Australas.* **2019**, *36*, 29.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.