

Assignment 2: Construct a database or dataset of selected archaeological materials/records, provide metadata statements in line with current Archaeology Data Service guidance, and a commentary explaining and justifying its structure and the categorisation/classification of the data

Analysing, interpreting, and contextualising the results of stable isotope analysis is inherently reliant on access to reliable and accurate comparative data (Klein *et al.* 2022). Open access resources such as the Biosphere Isotope Domains GB dataset (BGS 2023a), Stos-Gale and Gale's (2009) Oxford Archaeological Lead Isotope Database (OXALID), and the Global Lead Isotope Database (Westner *et al.* 2021; GlobaLID Core Team 2021) have revolutionised comparative data collection, archival, and visualisation. They are especially useful within isotope provenance studies, with their large datasets and interactive maps showing expected isotope values and variation for different areas at regional and global scales. However, these datasets primarily focus on collating isotope data from environmental sources (such as minerals, raw materials, plants, and water), and as yet, only a few have included data from human samples (see Evans *et al.* 2023). This omission represents a large data gap of useful and accessible reference material. Without comparative human isotope data from contemporaneous or geologically similar burials, skeletal remains are at risk of being analysed and interpreted in isolation or only alongside other remains found at the same site. This may lead to missed opportunities to situate the site and its skeletal population within the wider research context of the associated period and region. For lead (Pb) isotope analysis, a dataset including human data would be particularly invaluable for interpreting residential mobility and level of exposure to pollutant anthropogenic lead sources, by exhibiting how lead concentrations and isotope ratios measured in archaeological populations have varied over time and space, and allowing researchers to compare their results to this (Budd *et al.* 2000; Montgomery *et al.* 2010). The dataset presented here aims to begin to combat this gap, focussing on lead isotopes. Lead isotope ratios and concentrations measured in 307 archaeological human enamel samples from the UK and Ireland have been collated from relevant literature, along with essential contextual information associated with the remains. The dataset's structure, contents, and metadata have been created in line with good practice standards outlined by the Archaeological Data Service (ADS and Digital Antiquity 2013), and

adhere to the FAIR (Findability, Accessibility, Interoperability, and Reuse) principles (Wilkinson *et al.* 2016).

Dataset Structure

As outlined in the file-level metadata statement (see Comparative_Lead_Isotope_Dataset-ADS_Metadata_Statement file), the samples gathered from the reviewed literature have been compiled into a Microsoft Excel spreadsheet. The dataset file is titled “Comparative_Lead_Isotope_Dataset” and is available as an Excel file (.xlsx) and a comma-separated values (.csv) delimited text file in the supplementary data. Whilst the .xlsx file is an appropriate international open standard (ISO/IEC 29500:2008-2016) format for archiving spreadsheets that adheres to ADS deposition guidance (ADS 2023), the inclusion of the dataset in plain text format as a .csv file demonstrates the ease of migrating the data to a more accessible uncompressed format. This file can be used in different software programs, and is more suitable for digital preservation as well as future accessibility and dissemination (ADS and Digital Antiquity 2013; Niven 2023).

The decision to use an Excel spreadsheet to compile and present the dataset, rather than database software such as Microsoft Access, was made for several reasons. Currently, ADS guidelines do not provide a preferred file format for scientific data deposition due to its broad and varied nature (ADS 2023). As a method of scientific investigation, the results of stable isotope analysis are predominantly given and recorded as numerical values. Spreadsheets are particularly adept at storing this type of data, which can be analysed, processed, and organised efficiently using tools such as “sort” and “filter”, and the variety of functions and formulas Excel offers (LOC 2022; Niven 2023). Additionally, this spreadsheet is designed to complement and be used alongside the internal multi-isotope (carbon, nitrogen, oxygen, and strontium) reference Excel dataset used by Durham University’s commercial Archaeological Isotopic and Peptide Laboratory (AIPRL), and the previously mentioned open access datasets, which also store their data using Excel (Stos-Gale and Gale 2009; BGS 2023b). As such, using the same software is considered the most sensible option for data comparison and merging. Moreover, the data gathered is not complex enough to warrant use of a relational database. With the simplicity of its interface and nature

as a widely used program, using Excel will ensure the dataset is as accessible, operable, and reusable as possible for all users (Wilkinson *et al.* 2016).

The spreadsheet is composed of a single worksheet titled “Human_Lead_Isotope_Data” and contains a table with 307 rows of data categorised into 19 filterable columns. The columns can be separated into three categories: contextual information, lead isotope data, and resource references (discussed below). The structure and appearance of the spreadsheet is purposefully simplistic with minimal text formatting, to facilitate data migration and importation into different software programs (ADS and Digital Antiquity 2013; Niven 2023). No colours or styles have been applied to the data cells. This minimal formatting is a conscious choice to avoid assigning meaning to cells (e.g. through colour-coding) that could be lost during deposition and dissemination, as stylistic formatting does not carry over into plain text file formats (ADS and Digital Antiquity 2013). Instead, the data is described to the fullest extent possible within the dataset using multiple related columns containing broad and specific information and terminology, which are also referenced and explained in the metadata statement (MIDAS 2012; Wilkinson *et al.* 2016).

Dataset Content – Categorisation and Classification

Like most data types, scientific data means very little without contextual information. This is especially true for stable isotope analysis, as the ratio calculated from the abundance of each isotope in a sample has no unit of measurement itself, and is an almost meaningless number when viewed independently. Therefore, although collating lead isotope data is the main focus of this dataset, recording as much contextual information as possible for each sample is a priority to enhance the dataset’s quality and contextual integrity (Kansa and Kansa 2022). The contextual information columns can be sub-divided into dating information (two columns: “Period” and “Chronological Date”); locational information (three columns: “Country”, “Region”, and “Site Name”); and sample-specific information (seven columns: “Sample Number”, “Sample Type”, “Tooth Type”, “Fédération Dentaire Internationale (FDI) Tooth Notation”, “Biological Sex Estimation”, “Age Category”, and “Age-at-Death Estimation”). The columns are ordered as they are listed above, and help

break the dataset down into thematic blocks of related information. Following ADS and MIDAS Heritage guidance (ADS and Digital Antiquity 2013; MIDAS 2012), established wordlists and controlled vocabularies are used where necessary to control data entry. These are described in the “Comparative_Lead_Isotope_Dataset-ADS_Metadata_Statement-Supporting_Documentation” file, so future users can understand how and why data has been entered in this particular way, and to ensure any additions to the dataset conform to the chosen standard vocabularies (MIDAS 2012; Wilkinson *et al.* 2016).

To provide broad chronological information, the period associated with each sample is recorded following standardised terminology and dates recognised by Historic England (2023) and FISH (2023). As this vocabulary only applies to England, locational qualifiers are placed before the period names where the chronologies differ significantly between countries (i.e. “Scottish Iron Age”; “Irish Iron Age”). Dates and date ranges are also given where available to provide specific chronological context (MIDAS 2012), and have been derived from radiocarbon dating (differentiated by “cal AD/BC”), grave goods assessments, and nameplates. To standardise data presentation, all dates follow the format of “date/date range” followed by “AD/BC”, rather than using the negative integers suggested by MIDAS Heritage standards (2012) for prehistoric samples. Unlike specific years (integer data), date ranges cannot be mathematically analysed in excel without introducing more complex data entry fields, so wouldn’t benefit from this format.

It is important to remain as true to the consulted literature as possible whilst categorising and classifying the data, so as not to assign new meanings to it. As such, some samples may be recorded as belonging to a specific period to comply with classifications within the literature, despite the date(s) overlapping with other periods. For example, as decapitated individuals from Musselburgh, Scotland have been interpreted as belonging to the Roman period in the literature, they are classified the same in the dataset despite their radiocarbon dates also overlapping with the Scottish Iron Age (Moore *et al.* 2020). This is necessary as consulting reliable comparative data from contemporaneous populations is crucial to situate analysed skeletal remains and their isotope data within the wider chronological context (Montgomery *et al.* 2010). As dating information is arguably the most useful

contextual information included here, the samples are catalogued in chronological order by site name.

Country and region are recorded as it is necessary to be able to filter the dataset to identify relevant comparative data from similar geographical areas. Documenting site names is essential for referring to the comparative data in writing and data visualisation graphs. Unfortunately, no coordinate data is available to plot sites onto a map (e.g. using ArcGIS), although this is an aim for the future. Each sample also has its own unique identification number from the literature, and whilst only enamel samples have been collected so far, it will be possible to filter by sample type when more data is added.

White and Folkens' (1991) tooth notation standard is most prominent in the literature, and is used here as the straightforward format using initials, numbers and superscript/subscript makes visually identifying a sample's tooth type simple and efficient. However, font formatting is often lost in data migration to plain text format, risking loss of information about exact positions of sampled teeth in the dental arcade. To combat this, tooth type is also recorded using FDI notation (Leatherman 1972), which assigns each tooth a specific number, but may be more complex for non-specialists to decipher quickly. Although this is data duplication, both notation styles serve the dataset in unique ways and facilitate preservation of sampled tooth type data. This is especially important as isotope analysis is destructive, and tooth type can impact how measured isotope ratios and concentrations are interpreted (Richards 2004; Mays *et al.* 2013; 2023).

Biological sex and age-at-death estimation information is essential to align isotope data with the osteobiographical context of analysed remains. Sex estimation is recorded following standard terminology (Buikstra and Ubelaker 1994), although as many publications don't distinguish between individuals with intermediate sex assessments and those whose sex could not be estimated, unfortunately the term "undetermined" applies to both groups (Brickley 2004). As most publications use Buikstra and Ubelaker's (1994) age estimation categories for adults, this recognised recording standard (O'Connell 2004) has been adopted for the dataset. For non-adults, more variation is visible in the literature, so Buikstra and Ubelaker's (1994)

categories have been adapted with input from more recent standards (Scheuer and Black 2000). Specific age ranges given in the literature are recorded under “Age at Death Estimation” to provide as much contextual information as is available.

The next six columns contain the lead isotope data. Lead concentrations measured in the enamel samples are reported in parts per million (ppm). Although some earlier publications use equivalent units such as milligrams per kilogram (mg/kg or mg kg⁻¹), as ppm is now more commonly used in the literature for strontium and lead (see for example: Moore *et al.* 2020; Millard *et al.* 2014), this trend is followed in the dataset despite there being no standard reporting method. Five different lead isotope ratios are reported. ²⁰⁶Pb, ²⁰⁷Pb, and ²⁰⁸Pb are ratioed to the non-radiogenic ²⁰⁴Pb isotope, as analysing these ratios together allows full visualisation of the data to identify patterns such as sample clustering, which may indicate cultural focussing (converging) of lead isotopes within lead-using populations (Montgomery *et al.* 2010). This can help identify potential migrants, whose lead isotope ratios would sit outside these clusters. Isotope ratios ²⁰⁷Pb/²⁰⁶Pb and ²⁰⁸Pb/²⁰⁶Pb are also included here, as sometimes ²⁰⁴Pb abundance is too low to be measured during analysis (Budd *et al.* 2000; Montgomery *et al.* 2010). More data is available for lead concentrations than for isotope ratios, although this is not necessarily negative. Concentration data is useful by itself and in combination with isotope ratios to investigate and identify individuals with low enamel lead burdens (e.g. prehistoric or rural individuals), and those exposed to higher levels of anthropogenic lead (Montgomery *et al.* 2010; Budd *et al.* 2004; Moore *et al.* 2020).

The final column contains references to the publications in Harvard’s in-text citation format, avoiding the use of commas to enable migration to a .csv file. The full bibliographical references are documented in the “Comparative_Lead_Isotope_Dataset-ADS_Metadata_Statement-Supporting_Documentation” file. The 307 samples have been collated from well-cited seminal lead isotope publications and influential PhD theses, and represent 50 sites spanning from the Neolithic to the 19th century, thus testifying to the reliability, credibility, and diversity of the chosen data.

Whilst the scope of this dataset is restricted to human enamel samples from the UK and Ireland, its structure and composition demonstrate a concise and uncomplex way of compiling a relatively large amount of lead isotope data and essential contextual information in an accessible format. There is significant potential for future expansion using global human data, animal data, bone samples, and additional contextual information such as pathological evidence. Nevertheless, if deposited within ADS in its current form, this dataset and its metadata will serve to not only complement existing datasets, but enhance comparative lead isotope data availability overall.

Bibliography

Archaeological Data Service, Digital Antiquity, 2013. Caring for Digital Data in Archaeology: A Guide to Good Practice. Oxford: Oxbow Books

Archaeology Data Service, 2023. Help and Guidance: Instructions for Depositors – Data Requirements Table. [Online] Available at: <https://archaeologydataservice.ac.uk/help-guidance/instructions-for-depositors/files-and-metadata/> [Access Dated: 18th December 2023]

Brickley, M., 2004. "Determination of Sex from Archaeological Skeletal Material and Assessment of Parnutrition" In: Brickley, M., McKinley, J.I. (Eds.) *Guidelines to the Standards for Recording Human Remains*. Southampton: BABAO; Reading: IFA, pp. 23-35

British Geological Survey, 2023a. Biosphere Isotope Domains (Great Britain). [Online] Available at: <https://www.bgs.ac.uk/datasets/biosphere-isotope-domains-gb/> [Access Dated: 16th November 2023]

British Geological Survey, 2023b. Biosphere Isotope Domains (Excel Data). [Online] Available at: <https://www.bgs.ac.uk/download/biosphere-isotope-domains-msexcel-data/> [Access Dated: 16th November 2023]

Budd, P., Montgomery, J., Evans, J., Barreiro, B., 2000. Human Tooth Enamel as a Record of the Comparative Lead Exposure of Prehistoric and Modern People. *The Science of the Total Environment*, 263(1-3), pp. 1-10

Budd, P., Montgomery, J., Evans, J., Trickett, M., 2004. Human Lead Exposure in England from Approximately 5500 BP to the 16th Century AD. *The Science of the Total Environment*, 318(1-3), pp. 45-58

Buikstra, J.E., Ubelaker, D.H., 1994. Standards for Data Collection from Human Remains. Fayetteville: Arkansas Archaeological Survey

Evans, J., Mee, K., Chenery, C., Marchant, A., 2023. User Guide for the Biosphere Isotope Domains GB (Version 2) Dataset and Web Portal. *British Geological Survey Open Report*, OR/22/036, pp. 1-35

Forum on Information Standards in Heritage, 2023. Chronology. [Online] Available at: <https://www.heritage-standards.org.uk/chronology/> [Access Dated: 19th December 2023]

GlobalID Core Team, 2021. GlobalID Web Application V. 1.0, database status: 2023-06-18 [Online] Available at: <https://globalid.dmt-lb.de/#> [Access Dated: 16th November 2023]

Historic England, 2023. England's Historic Periods. [Online] Available at: <https://historicensland.org.uk/listing/the-list/historic-periods/> [Access Dated: 19th December 2023]

Kansa, E.C., Kansa, S.W., 2022. Promoting Data Quality and Reuse in Archaeology Through Collaborative Identifier Practices. *PNAS*, 119(43), e2109313118

Klein, S., Rose, T., Westner, K.J., Yiu-Kang, H., 2022. From OXALID to GlobalID: Introducing a Modern and FAIR Lead Isotope Database with an Interactive Application. *Archaeometry*, 64(4), pp. 935-950

Leatherman, G., 1972. Fédération Dentaire Internationale. New Tooth Designation System. *Bulletins and Memoirs of the Society of Anthropology of Paris*, 9(1), pp. 83-85

Library of Congress, 2022. Sustainability of Digital Formats: Planning for Library of Congress Collections – XLSX Transitional (Office Open XML), ISO 29500: 2008-2016, ECMA-376, Editions 1-5. [Online] Available at: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000398.shtml#:~:text=The%20XLSX%20format%20uses%20the,more%20worksheets%20in%20a%20workbook.> [Access Dated: 20th December 2023]

Mays, S., Elders, J., Humphrey, L., White, W., Marshall, P., 2013. Science and the Dead: A Guideline for the Destructive Sampling of Archaeological Human Remains for Scientific Analysis. London: APABE; Swindon: English Heritage

Mays, S., Elders, J., Humphrey., Marshall, P., 2023. Science and the Dead: Destructive Sampling of Archaeological Human Remains for Scientific Analysis (Second Edition). London: APABE; Liverpool: Historic England

MIDAS Heritage, 2012. MIDAS Heritage: The UK Historic Environment Data Standard, v 1.1. London: Historic England

Millard, A., Montgomery, J., Trickett, M., Beaumont, J., Evans, J., Chenery, S., 2014. "Childhood Lead Exposure in the British Isles During the Industrial Revolution" In Zuckerman, M.K. (Eds.) *Modern Environments and Human Health: Revisiting the Second Epidemiologic Transition* (First Edition). Hoboken: John Wiley and Sons, pp. 279-299

Montgomery, J., Evans, K., Chenery, S., Pashley, V., Killgrove, K., 2010. "Gleaming White and Deadly": Using Lead to Track Human Exposure and Geographic Origins in the Roman Period in Britain. *Journal of Roman Archaeology: Supplementary Series*, Suppl. 78(1), pp. 199-226

Moore, J., Rose, A., Anderson, S., Evans, J., Nowell, G., Grocke, D.R., Pashley, V., Kirby, M., Montgomery, J., 2020. A Multi-Isotope (C, N, O, Sr, Pb) Study of Iron Age and Roman Period Skeletons from East Edinburgh, Scotland Exploring the Relationship Between Decapitation Burials and Geographical Origins. *Journal of Archaeological Science: Reports*, 29(1), 102075

Niven, K., 2023. Help and Guidance: Guides to Good Practice – Databases and Spreadsheets. [Online] Available at: <https://archaeologydataservice.ac.uk/help-guidance/guides-to-good-practice/basic-components/databases-and-spreadsheets/introduction-to-databases-and-spreadsheets/what-are-databases-and-spreadsheets/> [Access Dated 18th December 2023]

O'Connell, L., 2004. "Guidance on Recording Age at Death in Adults" In: Brickley, M., McKinley, J.I. (Eds.) *Guidelines to the Standards for Recording Human Remains*. Southampton: BABAO; Reading: IFA, pp. 18-20

Richards, M., 2004. "Sampling Procedures for Bone Chemistry". In: Brickley, M., McKinley, J.I. (Eds.) *Guidelines to the Standards for Recording Human Remains*. Southampton: BABAO; Reading: IFA, pp. 43-45

Scheuer, L., Black, S., 2000. *Developmental Juvenile Osteology*. London: Academic Press

Stos-Gale, Z.A., Gale, N.H., 2009. Metal Provenancing Using Isotopes and the Oxford Archaeological Lead Isotope Database (OXALID). *Archaeological and Anthropological Sciences*, 1(3), pp. 195-213

Westner, K.J., Thomas, R., Klein, S., Yiu-Kang, H., 2021. GlobalID – Global Lead Isotope Database. V. 1.0. *GFZ Data Services*

White, T.D., Folkens, P., 1991. *Human Osteology*. London: Academic Press

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., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltra, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data*, 3(1), 160018