European Commission - Horizon 2020

PACHAMAMA - 101062179 — HE-MSCA-PF-2021

FAIR Data Management Plan

Name of fellow: Luis Nicanor Pezo Lanfranco.
Website: https://www.researchgate.net/profile/Luis-Lanfranco
Orcid: https://orcid.org/0000-0003-3163-5984
CV Lattes: http://lattes.cnpq.br/0949460440547327

Project: Paleodietary analyses of the first Andean cities: high-resolution assessment to macronutrients using a multiproxy approach - PACHAMAMA - 101062179 — HE-MSCA-PF-2021.

Affiliation: Institute of Environmental Science and Technology (ICTA), Universitat Autònoma de Barcelona.

Name of Supervisor: André Carlo Colonese.

Project start date: 01/10/2022
Project end date: 30/09/2024
DMP Last modification date: 25/01/2023

As part of PACHAMAMA Project (101062179 — HE-MSCA-PF-2021) according to the requirements and recommendations of the Horizon Europe Funding Program, through the Regulation (EU) 2021/695 of the European Parliament and the Annex 5-Specific Rules about 17th article of Horizon Europe Grant Agreement, that emphasize in the importance of responsible data management and open access to scientific production and research data, my supervisory team and I have designed this FAIR Data Management Plan (F-DMP).

This F-DMP is based in the Guidelines on FAIR Data Management and the Guidelines to the Rules on Open Access to Scientific Publications and Research Data from the Horizon 2020 Programme and follows the FAIR principles: findability, accessibility, interoperability, and "reusability" for the openness of research data, resumed in the slogan *"as open as possible, as closed as necessary."*

The current version (PACHAMAMA_F-DMP_v.1) describes in detail the process of research data management and the activities to be implemented by the fellow and the host institution (ICTA-UAB) for disclosure research data. In this sense, the F-DMP covers some critical aspects about the research data (i.e., type of data produced, quality assurance, organization of data), and how we expect to deal with it during and after the Project (i.e., long-term storage, open-access, and re-use).

Brief overview of Research Project and major accomplishments expected

The origin of the first urban centres and their associated social complexity are among the most debated topics in prehistoric archaeology. In the Central Andes, these processes have their antecedents in coastal settlements with impressive signs of monumentality, dating from at least the third millennium BC. According to a seminal 1975 publication, these societies would have developed complex systems of socio-political organization, which eventually led to States and Empires, sustained by the exploitation of the rich shoals of endemic marine species. On the other hand, newer archaeological evidence suggests that the adoption of plants in the region precedes in at least 3 millennia the rise of monumental architecture. Further, archaeological research conducted in the Peruvian North-Central coast (PNCC) for the last two decades has drastically changed our knowledge about the trajectory of early urban developments. At least 35 planned sites integrated into intra-valley hierarchical systems of settlements in the PNCC (i.e., Huaura, Supe, Fortaleza, and Pativilca valleys), dating between 4100-1800 BCE, came to reinforce the hypothesis that early civilization, based on powerful agricultural development, arose in the middle valleys rather than on the coastline. However, several major questions are still unanswered: 1) What food sources characterized the diets of the earliest urban populations of the Central Andes during the Initial Formative period (3000-1800 BCE)? 2) What were the roles of plant and marine resources in the diet of coastal and inland emergent urban sites? 3) How did these diets change over time? With the support of the

Marie Skłodowska-Curie Actions program, the PACHAMAMA project will analyse paleodiet using a multiproxy approach (stable isotopes, Bayesian mixing models, and palaeoproteomic analyses of dental calculus) aiming to identify the main macronutrients of the diet, to unlock diet composition among early urban groups in coastal Peru. To identify dietary food molecules, we will use stable carbon and nitrogen isotope analyses of bulk collagen from bone and teeth using gas chromatography combustion isotope ratio mass spectrometry (GC-C-IRMS), the analysis of stable carbon and nitrogen isotopes on single amino acids (AA) from bone collagen using liquid chromatograph-IRMS (LC-IRMS). Palaeoproteomics will be applied to identify specific food molecules and diet-correlated pathogens in the oral microbiome. The integration of palaeodietary isotopic reconstruction, palaeoproteomics, and other bioarchaeological (i.e., oral pathology) data is the most innovative component of PACHAMAMA.

1. Data

1.1. Type of data to be produced

The information that we expect to produce into the frame of PACHAMAMA Project consist of two kinds of data. The first consists of isotopic measurements of archaeological interest ($\delta^{13}C$, $\delta^{15}N$, $\delta^{34}S$, $\delta^{18}O$, and $^{87}Sr/^{86}Sr$) generated by mass spectrometry from tissues (i.e., collagen and bioapatite) extracted from bones and teeth of human individuals exhumated in archaeological sites of the Peruvian North-Central Coast. The second consists of proteins contained in dental calculus from those individuals. These two kinds of data are the primary goals of the Project and are essential to reach the Project objectives.

Other additional data consist of isotopic and proteomic markers of modern and archaeological fauna and flora specimens, recovered from the studied archaeological sites and region. Fauna and flora data will provide comparative baselines to make inferences about diet and mobility in humans. Because the diverse chronological origin (spanning a period between 3000 and 500 BC), we expect also obtain an still undetermined number of radiocarbon measurements to refining the chronology of the contexts that will be requested to specialized laboratories.

In stable isotopes analyses the quality controls and sample's validation is made during laboratory procedures following well defined quantitative criteria (DeNiro 1985; vanKlinken et al. 1999; DeFrance et al. 2020)[1]. Regular routine data-check protocols will be conducted by the fellow and the laboratory team to ensure the data quality. Isotopic measures in the mass spectrometers will be replicated twice. For control of systematic

[1] DeNiro, M. J., & Epstein, S. (1978). Influence of diet on distribution of carbon isotopes in animals. Geochimica et Cosmochimica Acta, 42, 495–506.
Van Klinken, G. J. (1999). Bone collagen quality indicators for palaeodietary and radiocarbon measurements. Journal of Archaeological Science, 26, 687–695.
France, C. A. M., Sugiyama, N., & Aguayo, E. (2020). Establishing a preservation index for bone, dentin, and enamel bioapatite mineral using ATR-FTIR. Journal of Archaeological Science: Reports, 33, 102551.

distortion, the standards and instruments will be calibrated before every procedure of data collection by the lead technicians of the ICTA laboratories. For consistency, adequate language will be adopted for pilot and formal analyses reports. In this case we will use the recommendations of Szpak et al. (2017)[2].

The mass spectrometers provide the results in compatible numerical formats. Thus, Excel data sheets containing "raw" isotope values will be available in .xlsx format immediately after the analytical procedures. Depending on the statistical packages to be used for modeling these raw data will be also reformatted to .xlsx, for compatibility.

In sum, the new data that will be produced by PACHAMAMA Project can be divided into two categories:

1. *Raw laboratory data*

a) Isotope values and laboratory parameters: consisting of numerical data accessible with Excel, and SPSS v.28 (Microsoft), and RStudio packages, stored in files of .csv, .xlsx, .spo, and .spv formats, and free-access .r formats.
b) Photos of processed samples and laboratory procedures: imaging data, accessible and processed with Photoshop and Paint packages, in imaging formats (i.e., TIFF, PNG, JPG).

2. *Statistical and modeling data*

a) Statistical and modeling tests results: numerical and text data, processed with Excel, SPSS v.28, FRUITS v.3.1, ReSources 22.11.4, and R Packages, in .xlsx, .spv, .spo, and .r formats, correspondingly.

For comparative procedures is expected the re-use of published data produced by the fellow in past projects. There are any constraints to re-using these existing data, that will be added to the new Database generated by PACHAMAMA Project.

The fellow, Luis Nicanor Pezo Lanfranco, is responsible for data collection, data production, data curation and data management. Data quality control is responsibility of the fellow, the Supervisor Prof. André Colonese, and the Senior Technician Krista McGrath (ICTA-UAB). The data produced by PACHAMAMA Project may be useful for future research based on stable isotopes and proteomics and should be well received by archaeologists, anthropologists and scholars working with biomolecules.

1.2. Size of the dataset

All the raw data and statistical and modeling results to be shared with the scientific community can be easily stored in text and numeric formats. The strings and algorithms for statistical tests and graphics generation will be stored in the original format

[2] Szpak, P., Metcalfe, J. Z. & Macdonald R. A. (2017). Best Practices for Calibrating and Reporting Stable Isotope Measurements in Archaeology. Journal of Archaeological Science: Reports, 13, 609-616.

files retrieved by the packages (mainly .spv, .spo, and .r files), but also in Excel (.xlsx) format, that is a widely distributed software. Those files usually are light files.

Individual's bioanthropological reports (text files: .txt and .pdf), photos of the samples and laboratory procedures (image files: .png, .tiff, .jpg), and other relevant information of archaeological contexts, such as maps, field sketches produced with other tools (Corel Draw .cdr, AutoCad .dwg, ArcGis .gpx, .kml, kmz and QGis GeoPackage etc.), may vary from light to very heavy files.

However, we prevent that the data produced and collected by PACHAMAMA Project will be stored in no more than 500 GB – 1 TB capacity storage devices.

1.3. Data storage during the development of the project

Although the data should be immediately stored after produced in the HD of a personal computer of the Project, we prevent the use of two external HDs of 1 TB each for data storage during the Project's development. The first one, will be labeled as "Working-HD" and destined to preserve the raw data, intermediary analytical tests, and final results. The second one, labeled as "Buck up-HD", should be updated immediately after the acquisition of new sets of final results. Photos of the samples and laboratory procedures and additional data will also be stored in the same devices. Data storage and backup during the execution of the Project is a shared responsibility of the fellow and the supervisor (Prof. André Colonese).

The results will also be shared with Ruth Shady, Director of the Caral Project (ZAC-UE-003, MinCu Perú), to be integrated with other archaeological information, eventually published in scientific reports, and translated to information for big audiences and dissemination. ZAC partners will contribute to the intellectual discussion in eventual publications and take part in the dissemination of PACHAMAMA results. Thus, data publishing is responsibility of Luis Nicanor Pezo Lanfranco, André Colonese, Krista McGrath (ICTA-UAB), Ruth Shady Solis (ZAC-MC Perú) and her team.

2. FAIR Data

2.1. Findability

The data will be stored in virtual folders and physical folders in printed archives organized after every relevant research stage. The files will be easily accessible using both Windows® system and equivalent free-access software. When necessary, we will include the necessary software (i.e., links or reference), including the open-source code.

Our virtual directory structure is based on the two main research fronts or PACHAMAMA Project, Stable Isotopes and Proteomics, organized in a hierarchical system of five levels, from the root folder to the 4th category folders, as showed in the Appendix Table 1.

For isotopic data, the final dataset is included in a Excel format file with three datasheets, the first for humans, the second for animals and the third for plants data, that have a predefined metadata structure similar to a previous database developed by the Pezo-Lanfranco, Colonese and colleagues in 2021 called SAAID (South American Archaeological Isotopes Database), which includes all the available isotopic data of archaeological interest produced in South America. The metadata structure appears in the Appendix Table 2.

Every data folder includes an Excel file with a brief description of the data contained in it. This Excel will register the name of the file, the type of data, author information, date of data collection, date of final results, institution of analysis, and Reference if it is a published information, as well as the archaeological site of provenience of the data, the number of entries, and the corresponding correlative numbers in the main PACHAMAMA Database (Appendix Table 3).

For naming the files we will use a simple convention that includes the name of the file (referred to the topic evaluated), the version, and the date in days, month, year (DD-MM-YYYY). The file versioning system is also simple and uses correlative numbers preceded by the letter v. Thus, the files will appear with the following nomenclature:

File name_v.#_DD-MM-YYYY

Particular files and project's documentation will also be uploaded to a Google Drive folder called PACHAMAMA_Files to be shared with all of the partners of the Project.

2.2. Availability and Open access

According to the EU's open access policies, all the methodological developments, datasets, and research outputs generated by PACHAMAMA will be shared with academic audiences via scientific reports and open access repositories. Open access to the research data will be made possible immediately after the data analysis and curation, without delays, and published data will be shared immediately. All the PACHAMAMA's data, as the current F-DMP, will be published and made available for re-use with a persistent identifier (Digital Object Identifier) and Creative Commons By license.

Following the UAB Open Access Policy for research data and the Open Publication of research data from UAB research projects from the Library Service, Computing Service and Research Management Area, our results will be deposited in the *UAB Digital Documents Dipòsit* (DDD - https://ddd.uab.cat), which is the tool that stores, manages, and disseminates the scientific production of the UAB, and is linked to the *CORA - Repositori de Dades de Recerca* (CORA RDR https://dataverse.csuc.cat), a data repository that centralizes research production of Catalan centers and publish research datasets in a FAIR way following the guidelines of the European Open Science Cloud (EOSC). These repositories assign persistent identifiers to scientific reports and datasets. For UAB members, DDD and CORA RDR warrant storage of research datasets until 25GB for free.

The Raw Data will be published as soon as possible after organization and curation. We prevent the submission of Data Updates in the UAB's Repository with a periodicity of 6 months, starting in the second half of the first year of the PACHAMAMA action.

In accordance with the Horizon Europe prescriptions, the final versions of research articles, should be published immediately, without embargo in Gold Access Journals (with Article Processing Charges – APC, imputable to the Project) or Hybrid Journals (with APC not imputable to the Project) or Subscription Journals that allow publishing of Accepted Versions without embargo with CC license. Considering that eventual publications will be signed by the fellowship, as first author, and the Supervisor as senior author, the publishing of research results as scientific articles in this vehicles should be funding by PACHAMAMA project and the UAB, according to UAB's Open-Science rules (https://www.uab.cat/web/investigar/open-access-uab/financament-per-publicar-1345841696025.html). Accepted and revised versions of scientific articles should be published in Open Research Europe (https://open-research-europe.ec.europa.eu) and other trusted repositories.

Peer Review articles will be made available in a folder called "Published Results_Scientific Reports" containing the .pdf article file, the Supplemental Information files and the information related to the Reviewing Process like cover letters, history of reviews, Reviewers Comments, and Responses to Reviewers. Authorizations and permissions for sample exportation, destructive analyses, and data publishing, will also be reported. We will also make available other kind of scientific reports and papers in construction following the same prescriptions.

A copy of the final Database will be attached, as a DOI file, to the South American Archeological Isotopes Dataset (SAAID, Pezo-Lanfanco et al. upcoming), hosted in the Pandora Platform (https://pandoradata.earth/dataset/), as well as other certificate repositories that support open-access, such as, Zenodo (https://zenodo.org), that allows the storage of both publications and data, and Mendeley Databases (https://data.mendeley.com), which provides space for databases storage.

The following metadata, based on a modified version of the Dublin Core metadata standard (https://www.dublincore.org), should be included in an attached text file to describe our uploaded database:

1. Contributors and affiliations, responsible by data generation
2. Project Name, Acronym, and number
3. Place and chronology of the data
4. Date and place of data generation
5. Description of data
6. File format and dimension of the file
7. Persistent identifier (DOI, in this case)
8. Language of the resource

9. License and terms of use (Creative Commons License, CC By in this case, https://creativecommons.org).
10. Title and subject

There is no sensitive data (living persons data, ethics, ownership, commercial interests, public security, IP rights) to restrict or limit which data access and wide publication. There are no intellectual property (IP) rights associated with the produced data. During the period of the Project, Luis Pezo-Lanfranco (the fellow) and André Colonese (the Supervisor) have shared control access over the data.

2.3. Interoperability

PACHAMAMA data is interoperable and allows data exchange and re-use between researchers under open-access modality. Basic statistics can be performed with SPSS, R Package, or any other free-access software. The specific softwares used for modeling are also open-access, warranting reproducibility. The metadata structure of our raw- and final-datasets is friendly and easy to use, facilitating integrations and re-combinations with different datasets from different origins, especially with the SAAID, produced by the signatories of this F-DMP.

The language used for headings of datasets follows the recommendations of Szpak et al. (2017), a guideline on the adequate use of terms in isotopic studies, making our results interoperable. The use of standard vocabulary in every possible case allows inter-disciplinary interoperability.

2.4. Reusability

Any research output and analytical tools (i.e., R Package strings, used software and/or algorithms, and references to the used instruments) needed to reuse or validate the data will also be available in the repository. These data will be licensed with CC BY license. There are no special selection criteria or restrictions to reuse of the data after the Project ends.

All data collected and produced by PACHAMAMA, including raw data, will be preserved for reuse, and archived internally (i.e., ICTA - ERC-TRADITION network drive) for a minimum of 10 years. After the end of the Project, the data will remain available for reuse, with no access restrictions, in the UAB DDD and CORA platforms, that should host the data for 10 years.

3. F-DMP costs

The fellow will spend around two hours by week in the actions proposed in this F-DMP, especially in organization and curation of data for upload in free-access repositories. The economic resources invested in data storage will be the coast of the storage devices, in this case two External Hard Drives of 1 TB each, with an approximate price of 200 Euros. There are no additional costs related to preparing the PACHAMAMA data for reuse, nor costs to making it available in the selected repositories, that are for free. The publication of results in Gold Open Access or Hybrid journals will be covered by the MSCA - Horizon 2020 grant, according to Grant Agreement conditions. Alternatives to raise funding for long-term preservation, if necessary, will be discussed with the Project Supervisor.

4. Data security

The PACHAMAMA Project does not comprehend the production of sensitive data. The *Digital Documents Dipòsit* of the UAB is a secure and certified repository that ensure safe storage for a long-term period, easy access, and safe transfer of data. In addition, the personal computer provided by the host institution have enough storage capacity, technical support, and installed antivirus of last generation to prevent eventual data loss.

5. Ethical Aspects

About ethical issues linked to the collecting, analyzing, and publishing of PACHAMAMA data, as mentioned in the original Project, it is planned to import human tissues such as bone and teeth, animal and plant remains from archeological origin, and modern native flora samples, from Peru to the ICTA-UAB (Spain) for destructive analyses (proteomics and stable isotopes). These activities, however, do not raise potential ethics issues. PACHAMAMA Project has the formal authorizations of the Peruvian Ministry of Culture (MinCu Perú) to perform the analyses and has a strong collaborative relationship with ZAC, the curator of the samples.

The PACHAMAMA Project data does not involve data requiring protection against unauthorized disclosure, nor involve any other restriction.

6. F-DMP further support

This comprehensive data management plan will be published in January 2023 in the European Commission Portal linked to the PACHAMAMA Project (https://ec.europa.eu/info/funding-tenders/opportunities/portal/) and the official repository of the host institution (DDD-https://ddd.uab.cat).

Access links to the F-DMP will also be disponible at the social networks of the PACHAMAMA Project (see the Communication and Dissemination Plan).

The F-DMP was built attending the directions of the eiNa DMP: Pla de Gestió de Dades de Recerca developed by the *Consorci de Serveis Universitaris de Catalunya* (available at https://dmp.csuc.cat), and will be updated at the middle (September 2023) and at the end (September 2024) of the project.

Bellaterra, January 25th, 2023.

LUIS NICANOR PEZO LANFRANCO - DNI 221006185 Firmado digitalmente por LUIS NICANOR PEZO LANFRANCO - DNI 221006185 Fecha: 2023.01.26 05:56:49 +01'00'	ANDRE CARLO COLONESE - DNI X9981221A Digitally signed by ANDRE CARLO COLONESE - DNI X9981221A Date: 2023.01.26 10:08:13 +01'00'
Signature of fellow	Signature of supervisor

APPENDIX

Table 1. PACHAMAMA Project: Hierarchical structure of the folders storage system

Root folder	1st category folder	2nd category folder	3rd category folder	4th category files formats
PACHAMAMA	Inventory of folders			Excel
	Isotopic_Data	Raw_Data	Bulk_collagen_isotopes	Excel, SPSS, R files
			Apatite_isotopes	Excel, SPSS, R files
			Amino_Acids_isotopes	Excel, SPSS, R files
		Intermediary_results	Statistical_tests	Excel, SPSS, R files
			Modeling_data	SPSS, R files
		Final_results	Curated_Datasets	Excel, SPSS, R files
			Curated_tables	Excel, SPSS, R files
			Curated_figures	Excel, SPSS, R files
	Proteomic_Data	Raw_Data	Proteomic_DentalCalclus	
		Intermediary_results	Statistical_tests	Excel, SPSS, R files
		Final_results	Curated_Datasets	Excel, SPSS, R files
	Scientific_Reports	Scientific_texts_isotopes	Versions_text	word files
			References_isotopes	Pdf files
		Scientific_texts_proteomics	Versions_text	word files
			References_proteomics	Pdf files
		Imagings_for publication	Curated_imagins	TIFF, PNG, JPG files

Table 2. Fields of PACHAMAMA Isotopic Database

Section	Variable	Data sheets	Definition	Categories	Type of data (unit)
Identification of the entry	Entry	All	Correlative number of the record.	Correlative numbers.	Numeric
	Country	All	South American modern country in which the sample was collected.	Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Guyana, Paraguay, Peru, Suriname, Uruguay, and Venezuela.	Text-Categorical
	Site Name	All	Name or label of the archaeological site where the sample was collected.	Name of the site.	Text, Numbers-Categorical
	Sample Id	All	Identification number or name of the individual/burial/sample analyzed, as reported in original publication.	Id of the sample or specimen.	Text, Numbers, Categorical
Data of the Archaeological Site	Biome 1	All	Biome in which sample was collected, according to the generic classification of Olson et al. (2001), updated by the Center for International Earth Science Information Network (CIESIN, 2012).	Tropical and subtropical moist broadleaf forests. Tropical and subtropical dry broadleaf forests. Tropical and subtropical coniferous forests. Tropical and subtropical grasslands, savannas and shrublands. Flooded grasslands and savannas. Temperate broadleaf and mixed forests. Temperate coniferous forests. Temperate grasslands, savannas and shrublands. Boreal forests/Taiga	Text-Categorical

				Montane grasslands and shrublands. Rock and Ice. Tundra. Deserts and xeric shrublands. Mediterranean forests, woodlands, and scrub. Lakes Mangroves.	
	Biome 2	All	Biome in which sample was collected, according to local or national classification system or notation.	Name of the biome.	Text-Categorical
	Region	All	Archaeological or geographic region in which sample was collected, according to local classification system or notation.	Name of the region.	Text-Categorical
	Valley/locality/closest town/political jurisdiction	All	Name of valley and/or locality/village/city and/or political jurisdiction in the vicinity of the collection site. These data aim an easy identification of site's location.	Name of valley and/or locality/village/city and/or political jurisdiction.	Text-Categorical
	Latitude (WGS84)	All	Latitude in decimal degrees, using the reference coordinate system WGS84.	Latitude (WGS84)	Numeric
	Longitude (WGS84)	All	Longitude in decimal degrees, using the reference coordinate system WGS84.	Longitude (WGS84)	Numeric
	Altitude (masl)	All	Altitude in meters above the sea level.	Altitude (masl)	Numeric
	Exact coordinates?	All	Specification of whether exact collection location is known.	Exact Approximate	Text-Categorical
	Site location radius (km)	All	In the case of "Approximate" location, the estimated radius in kilometers of the circular area in which the collection site is located using as midpoint of the	Site location radius (km).	Numeric

			circle the provided coordinates or the exact location of the site.		
	Ref coordinates	All	Name(s) of the author(s) and year of publication of the reference that provide the site coordinates.	Reference of coordinates.	Text-Categorical
	Archaeological site type	All	Extended description of collection site by type. Combinations are permitted using plus sign (+).	Site Type: Open site, Cave, Rock-shelter, Shell midden, Mound, Monumental architecture, Public architecture, Domestic architecture, Fortress, Battlefield, Funerary contexts, or combinations.	Text-Categorical
	Archaeological site function	All	Extended description of the function of the collection site. Combinations are permitted using plus sign (+). Three or more functions configure a “Complex” site.	Site function: Camp, Residence, Funerary, "Town, Urban Center, "Ceremonial Center, combinations of two categories, and Complex (three or more categories).	Text-Categorical
	Archaeological culture	All	Probable cultural affiliation, polity, or regional group.	Name of the archaeological culture. For instance: Inca, Sambaqui, Tupinamba, Muisca, Chinchorro, Tiwanaku, Tiwanakota, Moche, Mochica etc.	Text-Categorical
Chronology	Chronology Method	All	Method used to estimate time of sample death or deposition of archaeological specimen.	Relative Absolute Relative and Absolute	Text-Categorical
	Relative chronology (Period/phase)	All	Cultural and/or historical name of time interval (Period/Phase) that includes death or deposition of sample, according to country or region.	Name of the archaeological period or phase.	Text-Categorical
	Relative date from	All	Initial date of the period in years.	Age in years according to Age System.	Numeric

	Relative date to	All	Final date of the period in years.	Age in years according to Age System.	Numeric
	Age System	All	The chronological system in which the age of sample is reported.	BCE/CE (use the sign minus (-) for BCE and blank for CE) BP cal BP	Text-Categorical
	Absolute chronology method	All	The type of absolute chronological method employed to estimate age of sample.	14C 14C-AMS OSL (SAR) TL	Text-Categorical
	Lab Code	All	Laboratory-assigned identification code associated with a radiocarbon or other dating measurement.	Laboratory identification code.	Text, Numbers, Categorical
	Conventional 14C date (BP)	All	Uncalibrated radiocarbon date (conventional radiocarbon age).	Date in years BP.	Numeric
	Error (±1σ)	All	Standard deviation associated with conventional radiocarbon age. Expressed as ±1σ	Standard deviation of the date.	Numeric
	Calibrated/modelled 14C (95%) from	All	Initial calibrated date of the specimen in years according to Age System.	Initial calibrated date at 95% (2σ) for the specimen/site	Numeric
	Calibrated/modelled 14C (95%) to	All	Final calibrated date of the specimen in years according to Age System.	Final calibrated date at 95% (2σ) for the specimen/site	Numeric
	Material dated	All	Material subject to dating from the sample or associated with the sample.	Material used to date. For instance, "bone collagen", "dentine collagen", "charcoal", “associated material” etc.	Text-Categorical
Individual's Data	Lab id	All	Laboratory number associated with analyzed material as reported in original publication. If not available, the Id of the individual should be reported instead.	Lab Id of the sample or specimen.	Text, Numbers, Categorical

	Type of source	Animals Plants	The origin of the source. For animals and plants only.	For Animals: Terrestrial fauna; Freshwater fauna; Marine fauna. For Plants: Terrestrial flora; Freshwater flora; Marine flora.	Text-Categorical
	Taxon/Local name	Animals Plants	The common name of the reported specimen in English and/or its local (vernacular) name in other language. This field was filled mostly with the information provided by the authors, however, when the species was reported, we complemented the information of taxon or local name.	The common name of the specimen.	Text-Categorical
	Genus/Species	Animals Plants	The scientific name of the specimen at Genus or Species level, following current taxonomic nomenclature.	The scientific name of the specimen using taxonomic notation.	Text-Categorical
	Sex indiv	Humans Animals	Biological sex of sampled individual. This field was designed for humans, but also applies to animals whenever available. We used only males, females and undetermined. Other classifications were adjusted (i.e., PM= possible male, was classified as M).	M = Male F = Female und = undetermined	Text-Categorical
	Age category	Humans Animals*	Age-at-death category of sampled individual. In most cases we recorded the age-at-death reported in the publication if it meets our criteria of wide age-categories. If not, classifications were adjusted accordingly. If authors provide the category but not the age-range, we attribute the age-range following our	Fetal: not born Infant-1: 0-1 year Infant-2: >1-5 years Child-1: 6-8 years Child-2: 9-11 years Adolescent-1: 12-15 years Adolescent-2: 16-20 years Young Adult: 21-35 years Middle Adult: 36-50 years	Text-Categorical

			categories. If authors provide the age-range but not the category we use it to classify the individual into wide age-categories. *This field was designed for humans, but also applies to animals whenever available. In such case, we follow the authors' classification.	Old Adult: >50 years Subadult: Subadult with unknown age Adult: Adult with unknown age und: Individual with unknown age	
	Min age	All	Minimum estimated age-at-death of individual, in years. Whenever available, we follow the age-range suggested by the author. If the author attributes an exact age for the individual, for example, 13 years, we adjusted the initial age to 12 years and the final age to 14 years, considering the age-range variation suggested by Buikstra and Ubelaker (1994) for juveniles.	Minimum estimated age in years. For instance, 0.5, 12.	Numeric
	Max age	All	Maximum estimated age-at-death of individual, in years.	Maximum estimated age in years. For instance, 1.5, 14.	Numeric
	Photosynthetic Pathway	Plants	The Photosynthetic Pathway of the reported plant.	C3 (Calvin pathway) C4 (Hatch-Slack pathway) CAM (Crassulacean Acid Metabolism) blank if the pathway was not reported or is unknown.	Text-Categorical
	Plant domesticated	Plants	The status of domestication of the reported plant.	If the plant is domesticated: Yes If the plant is wild: No If the plant is wild but managed: Managed	Text-Categorical

					If there is not available data about the domestication status of the plant: Unknown	
		Sample type	All	Type of organ sampled for stable isotopes.	For humans: bone, tooth, hair, nail, muscle, skin etc. For animals: bone, tooth, hair, fur, nail, hoof, muscle, skin etc. For plants: depending on the plant type, root, tuber, bulb, seed, pod, stem, fruit, husk, flower, leaf, cob etc.	Text-Categorical
Isotope Values	In Collagen/Bulk plant	Tissue	Humans Animals	Type of tissue source of isotopic values.	For humans: bone collagen, dentin collagen, enamel, bone apatite, hair keratin, nail keratin, skin collagen etc. For animals: bone collagen, dentin collagen, enamel, bone apatite, fur keratin, hoof keratin, skin collagen, soft tissue etc.	Text-Categorical
		Element	Humans Animals	The element sampled (i.e., bone type in anatomic notation, teeth type in anthropological notation, plant part in botanical notation etc.) to extract the tissue source of isotopic values. The bone element reported in the original source was simplified to be reported in SAAID (for instance: right radius was reported as radius). For human teeth we use uppercase letter for permanent dentition. For	For humans: cranial bone, humerus, phalange, long bone, Max LM1 (maxillary left first molar), Mand Ldm1 (mandibular left first deciduous molar). For animals: cranial bone, operculum, coracoid, metapodial etc.	Text-Categorical

				deciduous dentition we use lowercase letters and added the letter “d” in front of the letter that describes the tooth to ensure an easy recognition and eliminate ambiguity. I, i = incisor, C, c = canine, Pm= premolar, M, m = molar. The number denotes tooth’s position, M1 first molar, M3 third molar (Hillson, 1996). The classic anthropological nomenclature was preserved but we replaced the terms Upper (U) and Lower (L) for Maxilla (Max) and Mandible (Mand) in front of each tooth. The letters to denominate laterality, L (Left) and R (Right), were preserved.		
		Tissue age	Humans Animals	Age-category for the tissue or segment sampled, in years. For bones, due to bone turnover, it is impossible to calculate a reliable age of tissue formation. However, considering that it occurs in a time relatively close to the death period, we classify tissue-age as the same age-at-death category for subadults, and as “Adult” for any adult category. For teeth or dentine or dental slices sequences, we follow the chronology suggested by The London Atlas of Human Tooth Development and Eruption (AlQahtani et al. 2010). An approximate age-range of formation is provided for each tooth, or each section.	Age-range of tissue formation in years for bones. For instance: Infant-1 → Tissue-age: Infant-1 Child-1 → Tissue-age: Child-1 Adolescent-1 → Tissue-age: Adolescent-1 Young Adult → Tissue-age: Adult Old Adult → Tissue-age: Adult Age-range of tissue formation in years for teeth. For instance: Max Ldm2 (maxillary left first molar) → Tissue age: Fetal/Infant-1: -0.7 – 0.9 years	Text, Numbers, Categorical

		Min age	Humans Animals	Minimum estimated age of individual at the time of tissue formation, in years. The field is specifically designed for teeth, following AlQahtani et al. (2010).	Minimum estimated age in years. For instance, 0.5, 12.	Numeric
		Max age	Humans Animals	Maximum estimated age of individual at the time of tissue formation, in years. The field is specifically designed for teeth, following AlQahtani et al. (2010).	Maximum estimated age in years. For instance, 1.5, 14.	Numeric
		%Yield	Humans Animals	Collagen yield	Mass of collagen recovered from the analyzed tissue, expressed as percent abundance %.	Numeric
		wt%C	All	Weight of Carbon in collagen sample or in the bulk sample in the case of plants.	Mass of Carbon relative to the mass of recovered collagen, expressed as percent abundance %.	Numeric
		wt%N	All	Weight of Nitrogen in collagen sample or in the bulk sample in the case of plants.	Mass of Nitrogen relative to the mass of recovered collagen, expressed as percent abundance %.	Numeric
		C:N	All	Carbon-Nitrogen Ratio.	Carbon-Nitrogen ratio or weight percent carbon and nitrogen in collagen sample, with one decimal.	Numeric
		$\delta^{13}C$	All	Stable Carbon isotope value from collagen or the bulk sample in the case of plants.	Delta $^{13}C/^{12}C$ value with one decimal, relative to VPDB standard.	Numeric
		$\delta^{15}N$	All	Stable Nitrogen isotope value from collagen or the bulk sample in the case of plants.	Delta $^{15}N/^{14}N$ value with one decimal, relative to AIR standard.	Numeric
		wt%S	All	Weight of Sulphur in collagen sample.	Mass of Sulphur relative to the mass of recovered collagen, expressed as percent abundance %	Numeric

		$\delta^{34}S$	All	Stable Sulphur isotope value from collagen.	Delta $^{34}S/^{32}S$ value with one decimal, relative to VCDT standard.	Numeric
		Other isotope or analytical value (two fields)	All	Any relevant analytical value obtained from collagen provided by the reference, i.e., other isotopic values, other preservation criteria values, other values relevant to the research.	Expressed in numeric notation, with a commentary in the cell of the first record that indicates the type or analytical value.	Numeric
	In Bioapatite/Bulk plant	Tissue	Humans Animals	Type of tissue source of isotopic values.	For humans: bone apatite, enamel, dentin apatite etc. For animals: bone apatite, enamel, dentin apatite, shell carbonate etc.	Text-Categorical
		Element	Humans Animals	The element sampled (i.e., bone type in anatomic notation, teeth type in anthropological notation, plant part in botanical notation etc.) to extract the tissue source of isotopic values.	For humans: cranial bone, humerus, phalange, long bone, Max LM1 (maxillary left first molar) etc. For animals: cranial bone, operculum, coracoid, metapodial etc.	Text-Categorical
		Tissue age	Humans Animals	Age-category for the tissue or segment sampled, in years. For bones, due to bone turnover, it is impossible to calculate a reliable age of tissue formation. However, considering that it occurs in a time relatively close to the death period, we classify tissue-age as the same age-at-death category for subadults, and as “Adult” for any adult category. For teeth or dentine or dental slices sequences, we follow the chronology suggested by The London Atlas of Human Tooth Development and Eruption (AlQahtani et al. 2010). An	Age-range of tissue formation in years for bones. For instance: Infant-1 → Tissue-age: Infant-1 Child-1 → Tissue-age: Child-1 Adolescent-1 → Tissue-age: Adolescent-1 Young Adult → Tissue-age: Adult Old Adult → Tissue-age: Adult Age-range of tissue formation in years for teeth.	Text, Numbers, Categorical

				approximate age-range of formation is provided for each tooth, or each section.	For instance: Max Ldm2 (maxillary left first molar) → Tissue age: Fetal/Infant-1: -0.7 – 0.9 years	
		Min age	Humans Animals	Minimum estimated age of individual at the time of tissue formation, in years. The field is specifically designed for teeth, following AlQahtani et al. (2010).	Minimum estimated age in years. For instance, 0.5, 12.	Numeric
		Max age	Humans Animals	Maximum estimated age of individual at the time of tissue formation, in years. The field is specifically designed for teeth, following AlQahtani et al. (2010).	Maximum estimated age in years. For instance, 1.5, 14.	Numeric
		$\delta^{13}C_{carbonate}$	All	Stable Carbon isotope value from bioapatite carbonate.	Delta $^{13}C/^{12}C$ value from carbonate with one decimal, relative to VPDB standard.	Numeric
		$\delta^{18}O_{carbonate}$	All	Stable Oxygen isotope value from bioapatite carbonate.	Delta $^{18}O/^{16}O$ value from carbonate with one decimal.	Numeric
		$\delta^{18}O$ standard	All	Standard used by the reference to show Oxygen values.	VPDB: Vienna PeeDee Belemnite SMOW: Standard Mean Ocean Water VSMOW: Vienna Standard Mean Ocean Water	Numeric
		$\delta^{18}O_{phosphate}$	All	Stable Oxygen isotope value from bioapatite phosphate.	Delta $^{18}O/^{16}O$ value from phosphate with one decimal.	Numeric
		$^{87}Sr/^{86}Sr$	All	Strontium isotope value	Value of the isotopic relation $^{87}Sr/^{86}Sr$ with 5 or 6 decimals.	Numeric
		$^{206}Pb/^{204}Pb$	All	Lead isotope value	Value of the isotopic relation $^{206}Pb/^{204}Pb$ with 2 decimals.	Numeric
		$^{207}Pb/^{204}Pb$	All	Lead isotope value	Value of the isotopic relation $^{207}Pb/^{204}Pb$ with 2 decimals.	Numeric
		$^{208}Pb/^{204}Pb$	All	Lead isotope value	Value of the isotopic relation $^{208}Pb/^{204}Pb$ with 2 decimals.	Numeric

		Other isotope or analytical value (two fields)	All	Any relevant analytical value obtained from bioapatite provided by the reference, i.e., other isotopic values, other preservation criteria values, other values relevant to the research.	Expressed in numeric notation, with a commentary in the cell of the first record that indicates the type or analytical value.	Numeric
Extra Information and Comments		Extra information	All	Other information referred to the individuals' features, the specific burial context, or any other data relevant to the individual context knowledge.	Expressed in numeric notation or as text, with a commentary in the cell of the first record that indicates the type or analytical value.	Text, Numbers, Categorical
		Additional info source	All	Source or reference (in APA style) that contains other data relevant to the contextualization of individuals and archaeological sites (i.e., archaeological data from the sites, complementary contextual data, draws, inventories, location maps etc.).	Reference, DOI etc.	Text
		Preservation criteria	All	Classification of the Preservation Criteria availability.	1: The reference provides all the expected Preservation Criteria. 2: The reference provides only the basic Preservation Criteria values (i.e., C:N rates for collagen). 3: The reference provides only the raw data and no Preservation Criteria values.	Numeric
		Commentaries	All	Any comment relevant to understand the article-source of data, or individual records.		Text
Referenc e		Original Reference	All	Author(s) and year of publication, for better identification of the source.	Author(s) (year).	Text
		Full Reference	All	Complete reference of the Original Reference of the source in the citation	Samec, C., Yacobaccio, H.D., & Panarello, H.O. (2017). Carbon	Text

			style of the American Psychological Association (APA).	and nitrogen isotope composition of natural pastures in the dry Puna of Argentina: a baseline for the study of prehistoric herd management strategies. *Archaeological and Anthropological Sciences*, 9, 153–163.	
	Compilation Reference	All	Author(s) and year of publication, for better identification of the source.	Author(s) (year).	Text
	Full Reference	All	Complete reference of the Compilation Reference in the citation style of the American Psychological Association (APA).	Scaffidi, B., Tung, T., & Knudson, K. (2021). Seasonality or short-term mobility among trophy head victims and villagers? Understanding late-life dietary change in the pre-Hispanic Andes through stable isotope analysis (δ13C/δ15N) of archaeological hair keratin and bone collagen. *Journal of Archaeological Science: Reports,* 39(2),103152. DOI: 10.1016/j.jasrep.2021.103398	Text
	Submitter Id	All	The co-author who extracted, edited, and upload the data in SAAID.	Name of the submitter.	Text
	Link to source	All	Electronic link to the original source or reference that contains the data on the internet (when available).	DOI, URL, or link to other kind of data repository.	Text