

# ArchLabour Data Management Plan

## ArchLabour Research Data

### Description

In the realm of architecture, the focus on the ‘designing elite’ has inadvertently eclipsed a critical facet – the invisible labour force shaping colonial public works. Overlooking the identities, recruitment, and experiences of these workers raises pressing questions about the legacy of spatialised architectural plans. In this context, the ERC-funded ArchLabour project will introduce a new theoretical framework, delving into the untold stories of labourers in African countries with Portuguese as an official language. By bridging history, science, and post-colonial studies, ArchLabour aims to unveil the complexities of mass labour, shedding light on a crucial yet overlooked chapter in architectural heritage. Overall, the project will foster a deeper understanding of the complex legacy left by colonial architecture and labour.

### Researchers

Ana Vaz Milheiro (orcid:0000-0001-7213-0478), Beatriz Serrazina (orcid:0000-0003-4679-606X), Filipa Fiúza (orcid:0000-0002-3456-2234), Francesca Vita (orcid:0000-0001-6455-6583), Filipe Carmo (orcid:0000-0002-7036-6417), Patricia Noormahomed (orcid:0000-0003-3407-8124), Sónia Henrique (orcid:0000-0001-8374-8919), Inês Lima Rodrigues (orcid:0000-0002-0859-9367)

### Funder

ERC, ArchLabour, 1101096606

### Grant

Architecture, Colonialism and Labour. The role and legacy of mass labour in the design, planning and construction of Public Works in former African territories under Portuguese colonial rule/  
No 101096606

## Description

### 1.1 Brief description of the described research output

#### 1.1.1 What kind of research output are you describing?

Research Data

#### 1.1.2 Is it physical or digital?

Digital

#### 1.1.3 Are you generating or re-using it?

Both

ArchLabour uses new data, which is going to be collected through fieldtrips; oral histories: interviews; and workshops to build upon 20 case studies that cover the multiple typologies, timeframes and scales which composed the imperial ground. Testimonies of privileged agents will then be collected. These will contribute to an oral history that may confront and counter the prevalent Eurocentric narratives in the history of colonial architecture. ArchLabour data will validate results in scientific publications. The data package will contain: the processed data and a 'read\_me.txt' describing the data.

Archlabour project intends to use existing data which is openly available through ISCTE trusted repository. Existing data has been collected during other projects coordinated by Ana Vaz Milheiro (e.g., Coast to Coast and ArchWar projects, FCT-funded) will be re-used where appropriate for achieving the aims of Archlabour. Any existing data, which have been collected for previous projects and might be re-used during the Archlabour project, will be handled according to the terms of use defined by ISCTE and the ethical approvals granted to the original projects. The data package will also contain: the processed data and a 'read\_me.txt' describing the data.

#### 1.1.4 What is the type of the described dataset?

Observational

The selection of the case studies will determine the data processing and the arrangement of the geographical databases which in turn will enable the profiling of mass labour.

Oral histories: interviews. Groups such as women and children will be considered, as well as issues of race and subsequent conflicts, continuing in the transition period up to independence. These interviews will include not only the testimonies of survivors, but also the testimonies of their families and communities, treated as privileged informants (holding key roles within the community structure), allowing the reconstruction of their journeys and the elaboration of narratives that

challenge Eurocentric history inside architectural culture. A standard questionnaire will be prepared, covering: a) basic data, including dates, places, family ties, professional category and training; and b) additional data, from memories/testimonies about labour and societal experiences to inequalities resulting from ethnic background, professional hierarchisation or gender. Interviews will be recorded and transcribed, to be used later in podcasts, short films or peer review publications.

Video recordings: in open file format is .MP4

Each interview has a maximum of 45 minutes to 1 hour. The size of each interview with the following features: 720p, 1280\*720, HD, has 900MB-1GB. We need over 1 TB for interviews. We used the Seagate calculator for the storage estimation.

The audio stream (speech): The audio will be transcribed using Scribie (transcription software). The text resulting from the audio transcription will be in Portuguese using the plain text (.txt) format.

The datasets will be accompanied by a README file with a description of the data.

The audio records format will be MP3, which allows the interoperability of the files. The interview transcripts are going to be saved in Plain Text (txt). The analysis files will be saved in tabular form in CSV format.

Workshops. This task will take 27 months, from the quarter after the start of the collection of oral testimonies. There will be a selection of participants, chosen on a voluntary basis and from among the key informants. Workshops will be video recorded and transcribed. A selection of the audio-visual material will be used in the exhibition, podcasts, short films or peer-reviewed publications.

Formats and procedures will be as described above.

Documentation. Apart from textual records graphic documents will be confronted to collect data, namely: cartography, iconography, photographs and other visual forms. The material is expected to range from: 1) administrative and legislative documentation related to CPW departments; 2) design documentation; 3) rural and military settlement operations; 4) bibliographic sources; and 5) periodical publications. This data will provide information on mass labour teams' structure, the number of workers involved, their ethnic and gender background, skills, wages, recruitment, and labour legislation.

#### 1.1.5 What is its format?

Data will be acquired, processed and stored using one of these open formats:

Text, Documentation, Scripts: XML, PDF/A, HTML, Plain Text.

Statistical data may be stored as SPSS (\*.sav) or STATA file formats.

Still Image: TIFF, JPEG 2000, PNG, JPEG/JFIF, DNG (digital negative), BMP, GIF.

Geospatial: Shapefile (SHP, DBF, SHX), GeoTIFF, NetCDF.

Video: MOV, MPEG-4, AVI, MXF.

Database: XML, CSV, TAB.

Where possible, files will be stored in open archival formats e.g. Word files converted to PDF-A or simple text files encoded in UTF-8 and Excel files converted to CSV. In case this is not possible, information on the software used and its version number will be included.

#### 1.1.6 What is its expected size?

Over 1000GB

#### 1.1.7 Why are you collecting/generating or re-using it?

To obtain information

ArchLabour will develop a theoretical framework for assessing mass labour in Colonial Public Works (CPW) to shine a spotlight on these invisible workers, thus drawing a connection between historical subalternity and the inequality that still haunts communities inheriting this past. Through the study of the diverse colonial experiences of the African countries that have Portuguese as one of their official languages (Cape Verde, Guinea-Bissau, São Tomé and Príncipe, Angola and Mozambique), and covering a wide period from the modern colonization that begins after the Berlin Conference through the industrial capitalism's exploitation praxis up to the years immediately following African independence, the project will cross the history of colonial architecture and the subject of labour, with the history of Science applied to construction and post-colonial studies in architecture.

#### 1.1.8 What is its origin / provenance?

All data to be used in the proposed study will be obtained from Portuguese, African and other international archives. Archives in Portugal, Cape Verde, São Tomé and Príncipe, Guinea-Bissau, Angola and Mozambique will be the main sources to conduct archival and historical research along with bibliographic searches (ArchLabour WP1: month 1-month 54) focusing on archival Colonial Public Works (CPW) archival records. The data collected through WP1's (month 1-month 54) comparative approach will provide the historical and contextual background as the origin of the collected data is related to four types of archives.

i) National archives, providing data to describe the main typologies of Public Works.

ii) Corporative archives (not held or deposited at National Archives) including the main companies responsible for the design, coordination and execution of large-scale works involving mass labour, namely seaports, railway, and airport companies; large majestic resource exploration companies; construction companies; etc.

iii) Private archives, covering private documentation collected by professionals engaged in mass work teams. The location of personal archives will depend heavily on personal networks fostered during fieldwork.

iv) International Organizations archives, namely the United Nations archives and those with documentation on the former European colonial empires (favouring the cases of Great Britain; France, Belgium; Germany and Italy, with possibilities of extension to other former empires).

#### 1.1.9 To whom might it be useful ('data utility')?

- Researchers
- Education
- The public

ArchLabour will organize an international congress and an exhibition, bringing together researchers and experts with theoretical and historical perspectives on Architecture, Colonialism and Labour). Additionally, the team members will participate in conferences, seminars and scientific meetings (e.g. DocoMomo; SAH, EAHN; ENHR, IASTE); whilst will produce double-blind peer-reviewed and indexed articles in their area of expertise (JSAH, Journal of Architecture, African Geographic Review, Planning Perspectives, International Labour and Working-Class History). The dissemination of data through scientific events is beneficial to researchers and the education sector. However, the exhibit is also effective for the general public, in addition to experts.

### 2.1 Publications

#### 2.1.1 Does the described output support any scientific publication?

No

#### 2.1.2 Is there a data availability statement provided along with the publication?

No

### 2.3 Software

#### 2.3.1 Does the described output use or support any software?

No

#### 3.1.1 Making data findable, including provisions for metadata

##### 3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset / output?

- Data identifiers

- Researchers identifiers
- Organizations identifiers
- Projects identifiers

DOI

ORCID

ROR

Cordis

Data identifiers (DOI)

Researchers identifiers (ORCID ID)

Organizations identifiers (Iscte ROR is available at: <https://ror.org/014837179>)

Cordis: <https://cordis.europa.eu/project/id/101096606/es>

#### 3.1.1.2 Will you provide metadata for the described dataset / output?

Yes

DataCite Metadata Schema

For the metadata we will use DataCite Metadata Schema v4. This metadata standard is supported by Zenodo. Some advantages of this metadata schema: data objects will be assigned with a globally unique identifier - DOI; ORCID identifier of creator; ROR for organisation; access right; license; version of the resource; etc. (<https://about.zenodo.org/principles/>).

#### 3.1.1.3 What type(s) of metadata?

Descriptive

The datasets will be accompanied by a README file with a description of the data.

#### 3.1.1.4 Do the metadata use standardised vocabularies?

No

#### 3.1.1.6 Are the metadata searchable?

Yes

#### 3.1.1.7 How are searchable metadata provided?

## Metadata repository

The data will be deposited in Zenodo Repository. Zenodo's metadata is compliant with DataCite's Metadata Schema minimum and recommended terms, with a few additional enrichment. License is one of the mandatory terms in Zenodo's metadata. DataCite's Metadata Schema meets one of the broadest cross-domain standards available.

### 3.1.1.8 Are keywords provided in the metadata?

Yes

Mass labour, colonial architecture, Portuguese colonialism

It will be included, where possible: subject, keyword, classification code, or key phrase describing the resource.

### 3.1.1.9 Are metadata harvestable?

Yes. Zenodo Metadata for individual records as well as record collections are harvestable using the OAI-PMH protocol by the record identifier and the collection name.

## 3.2.1 Repository

### 3.2.1.1 In which repository will the dataset / output be deposited?

Zenodo

<https://zenodo.org/communities/iscte/>

The data produced will be deposited in Zenodo, where a DOI is issued for every published record, making the data easy to find. The Zenodo community 'Iscte - Instituto Universitário de Lisboa' allows the sharing and publication of research data produced by the scientific community.

### 3.2.1.2 Is the selected repository a trusted source?

Yes

- Follows repository standards
- Details terms of use

### 3.2.1.4 Add appropriate arrangements made with the repository(ies) where the described dataset will be deposited

- Has an open access content policy
- Supports retention
- Supports withdrawal

- Supports back up
- Provides Open Access content (free at the point of use)
- Assigns PIDs
- Follows metadata standards
- Uses non-proprietary formats
- Supports mid- and long-term preservation
- Follows curation processes
- Supports authentication and authorization of users
- Has data security mechanisms in place

#### 3.2.1.5 Does the repository(ies) assign datasets / outputs with persistent identifiers?

Yes

#### 3.2.1.6 Does the repository(ies) resolve the identifiers to a digital object?

A DOI is issued to every published record on Zenodo. The DOI is a top-level and a mandatory field in the metadata of each record.

#### 3.2.1.7 Does the repository support versioning?

Yes

### 3.2.2 Data

#### 3.2.2.1 What is the described dataset / output title?

ArchLabour

#### 3.2.2.2 How is the dataset / output shared?

Open

#### 3.2.2.3 What is the reason of limiting access to the dataset / output?

All parts of the dataset can be shared.

#### 3.2.2.5 Are there any methods or tools required to access the dataset / output?

No

#### 3.2.2.8 Is the described dataset / output supported by a data access committee?



No

#### 3.2.2.9 Please specify how the dataset / output will be accessed during and after the project ends

Access data will follow the principle as open as possible as closed as necessary. If certain datasets need to be shared under restrictions, will be clearly explained the reasons. The data needed to validate results in scientific publications will be available in the Iscte community in Zenodo.

#### 3.2.2.10 Please specify how long after the project has ended the dataset / output will be made accessible for

Most data resulting from the project will be preserved for at least 10 years. The only data that will not be preserved are the interview audio files. At the point of transcription, the data will be anonymized and once the transcript is approved by the research participants, the audio file will be destroyed.

### 3.2.3 Metadata

#### 3.2.3.1 Will you provide metadata even if the described dataset / output can not be openly shared?

Yes. Zenodo supports the DataCite Metadata Schema v4. The following additional article level fields are supported: journal title/volume/issue/pages, conference title/acronym/dates/place/website, book publisher/place/ISBN/title/pages, alternate persistent identifiers (see all fields).

#### 3.2.3.2 Under which license will metadata be provided?

Creative Commons Zero (CC0)

#### 3.2.3.3 Do metadata provide information about how to access the described dataset / output?

Yes

The deposited research data will be in line with the FAIR principles (Findable, Accessible, Interoperable and Re-usable), in particular machine-actionable. It will be provided open access to research data under the principle 'as open as possible, as closed as necessary'.

#### 3.2.3.4 Will metadata remain available after the dataset / output is no longer available?

Yes. All ArchLabour submitted data in Zenodo will have open and accessible metadata.

### 3.3 Making data and other outputs interoperable

#### 3.3.1 Does your (meta)data use a controlled vocabulary?

No, it uses a Lexicon of Learning

3.3.2 If you created the vocabulary, where can it be found?

3.3.3 Have you applied a standard schema for your (meta)data?

Yes. DataCite Metadata Schema

3.3.5 What is the methodology followed?

Regarding metadata properties Mandatory (M) properties must be provided, Recommended (R ) properties are optional, but strongly recommended.

3.3.6 What community-endorsed interoperability best practices are followed?

### 3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset / output?

Creative Commons Attribution 4.0

3.4.2 What reusability and / or reproducibility methods are followed?

Readme files

3.4.3 Will you provide the described dataset / output in the public domain?

Yes

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

Yes. The Archlabour data will be archived in the Iscte repository.

### 4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

ArchLabour has estimated the cost of preparing and collating materials for deposit at eight to ten weeks, once the project has committed to deposit all data. In this sense, all data-related activities and resources for the entire data lifecycle - from data creation, through processing, analysis and storage, to sharing and long-term preservation - need to be priced to calculate the total cost of all data creation, data sharing, data access and preservation activities

This task will be carried out by the ArchLabour Project Manager, hired in July 2024.

#### 4.1.2 How will this cost be covered?

Use of institution infrastructure

### 5.1 Data Security

#### 5.1.1 What security measures are followed?

Encryption

Good practices available at: <https://ukdataservice.ac.uk/learning-hub/research-data-management/store-your-data/encryption/>

#### 5.1.2 What conditions do the security measures meet?

- Data storage
- Data sharing

Examples:

The digital files will be stored in the secured institutional infrastructures of Iscte.

Informed consents will be stored safely in a locked cabinet in a locked room at Iscte.

Backups:

All research data will be stored on Iscte networked drives from which backups are made automatically by the SIIC division.

During data collection, automatic backups will be made in the Iscte Office365.

#### 5.1.3 How will you preserve the described dataset / output in the long term?

Following the Iscte requirements, data and documentation needed to reproduce findings from this project will be stored for at least 10 years after the completion of the project.

### 6.1 Ethical aspects

#### 6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset / output?

Yes

The Ethics Council of ISCTE, in its statement numbered 130/2023, underlines the fact that data collection outside the EU requires explicit consent, and the study should clarify its intentions regarding sensitive information and potential risks to participants. Transparency is crucial, including informing participants of any plans to disseminate material and the limitations of confidentiality.

(segment summarised and translated) As such, “The consent states that the data will be treated confidentially, destroyed or anonymised at the end of the project. However, the study proposes to disseminate diverse material, with voice or audio recordings (see method, for example in podcasts and films). In accordance with the principle of transparency, participants should be informed of this disclosure as a limitation of confidentiality, and the moments of anonymisation and/or destruction clarified”. (segment translated)

#### 6.1.2 Does the described dataset / output contain sensitive information?

No

#### 6.1.3 Does the described dataset / output contain personal data?

No

### 7.1 Other

#### 7.1.1 Do you make use of other procedures for data management?

No

*Powered by*

