



**European Research Council**  
Executive Agency

Established by the European Commission



**Horizon 2020**  
**European Union Funding**  
**for Research & Innovation**

## **European Research Council (ERC)**

### **ERC Data Management Plan**

### **Template**

**Version 1.0**  
**21 April 2017**



# ERC OPEN RESEARCH DATA MANAGEMENT PLAN (DMP)

European Research Council  
Established by the European Commission

Project Acronym	Project Number
NEOGENE	772390

***Template for the ERC Open Research Data Management Plan (DMP)<sup>1</sup>. The following sections should describe how you plan to make the project data Findable, Accessible, Interoperable and Reusable (FAIR). Each of the following five issues should be addressed with a level of detail appropriate to the project.***

**SUMMARY** (dataset<sup>2</sup> reference and name; origin and expected size of the data generated/collected; data types and formats)

Three types of data are being generated in this project:

1- Genomic data: Illumina high throughput (Hiseq / Novaseq) genome sequencing data from archaeological human or sheep bones / teeth.

This will include a total of ~400 individuals (with >0.01X genome coverage), which will be analysed within the project.

The final total size of the raw FASTQ and aligned BAM files (sequencing data mapped to human or sheep reference genomes) will be on the order of 10TB.

2- Carbon14 dates: Radiocarbon dates obtained from ~200 archaeological human or sheep bones. Total size <1 GB.

3- Material culture datasets: Digitalised compilations of material culture features collected from archaeological literature.

<sup>1</sup> Based on '[Guidelines on FAIR Data Management in H2020](#)', version 3.0. 26.07.2016, Annex1

<sup>2</sup> Several datasets may be included into a single DMP.

## ERC OPEN RESEARCH DATA MANAGEMET PLAN (DMP)

### **1. MAKING DATA FINDABLE** (*dataset description: metadata, persistent and unique identifiers e.g., DOI*)

1- Genomic data: Raw and aligned genomic data will be deposited to the European Nucleotide Archive (ENA, <https://www.ebi.ac.uk/ena/>) in FASTQ and BAM formats.

The ENA identifiers (e.g. <https://www.ebi.ac.uk/ena/data/view/ERS1232261>) will be listed in the related publications, which is the standard in the field.

2- Carbon14 dates: Radiocarbon dates will be deposited to Open Context (<https://opencontext.org/>), and also published in the related publications.

3- Material culture datasets: This is not new data but compilations from the literature. The datasets will be published as supplementary material in the related publications.

### **2. MAKING DATA OPENLY ACCESSIBLE** (*which data will be made openly available and if some datasets remain closed, the reasons for not giving access; where the data and associated metadata, documentation and code are deposited (repository?); how the data can be accessed (are relevant software tools/methods provided?)*)

All data and related documentation will be open to access (through databases and/or directly within publications) simultaneously with publications. Thus, all data produced within the project will become open by the end of the project.

Processing the data does not require additional software.

## ERC OPEN RESEARCH DATA MANAGEMET PLAN (DMP)

### **3. MAKING DATA INTEROPERABLE** *(which standard or field-specific data and metadata vocabularies and methods will be used)*

No specific vocabulary will be employed.

### **4. INCREASE DATA RE-USE** *(what data will remain re-usable and for how long, is embargo foreseen; how the data is licensed; data quality assurance procedures)*

No embargo or licensing will be applied. Genomic databases are under the ODC Open Database License (ODbL).

## ERC OPEN RESEARCH DATA MANAGEMET PLAN (DMP)

**5. ALLOCATION OF RESOURCES and DATA SECURITY** *(estimated costs for making the project data open access and potential value of long-term data preservation; procedures for data backup and recovery; transfer of sensitive data and secure storage in repositories for long term preservation and curation)*

The ENA is a well-established database funded by EMBL and provides free service.  
The C14 data will be published in Open Context for an estimated ~1000 USD.  
We will be keeping backup copies of all the data at our server at the host university (METU).

**DISCLAIMER.** Please note that the ERC Data Management Plan is not a part of the Ethics Review. It is the responsibility of the Principal Investigator to inform the ERCEA Ethics Team of any ethics issues/concerns regarding the collection, processing, sharing and storage of data in relation to the project. The Principal investigator can also be asked to submit an Ethics Data Management Plan (Ethics DMP).