



**European Research Council**  
Executive Agency

Established by the European Commission



**European Research Council (ERC)**

**ERC Data Management Plan**

**Template**



European Research Council  
Established by the European Commission

## ERC OPEN RESEARCH DATA MANAGEMENT PLAN (DMP)

Project Acronym	Project Number
MoSD-Exam	-

**Template for the ERC Open Research Data Management Plan (DMP). 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>1</sup> reference and name; origin and expected size of the data generated/collected; data types and formats*)

- **Datasets:**
  - Data on "14-day notification rate of new COVID-19 cases and deaths" (ECDC): Weekly national data on cases, deaths, and 14-day rates (ca. 12,600 entries, ~18 MB original).
  - Data on "testing for COVID-19 by week and country" (ECDC): Weekly testing data including volume and positivity rates (ca. 6,100 entries, ~10 MB original).
- **Origin:** Primary from The European Surveillance System (TESSy) reported by EU/EEA Member States; secondary from public online sources (web-scraped daily by ECDC).
- **Expected end size:** ~2.5x original after merging/cleaning (ca. 18,700 entries total in processed form).
- **Data types:** Integer (e.g., cases\_weekly, tests\_done), Float (e.g., rate\_14\_day, testing\_rate), Strings (e.g., country, year\_week).
- **Dataformats:** Original: CSV, JSON, XML, XLSX; Processed: CSV (per country).

<sup>1</sup> Several datasets may be included into a single DMP.

# ERC OPEN RESEARCH DATA MANAGEMENT PLAN (DMP)

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

- **Identifier:** No formal DOI/PID yet; stable URLs via GitHub repository
- **Plan:** Upload to Zenodo for DOI assignment to ensure long-term persistence.
- **Description:** Datasets described with rich metadata in README.md and metadata.md (e.g., column descriptions: 'year\_week' as YYYY-WW string for record date, 'rate\_14\_day' as float for cases per 100k over 14 days)
- **sources:** TESSy/public scraping
- **limitations:** Europe-focused, potential inconsistencies from multi-sources
- **keywords:** COVID-19, testing volume, cases, ECDC). Metadata human-readable (Markdown) and machine-readable (structured in JSON for provenance). Searchable via GitHub index/Google.

## 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 openly available:** Original raw datasets, processed CSVs (merged/cleaned per country), metadata, code (Jupyter Notebooks for quality control/integration/analysis). No closed datasets (aggregated public health data, no sensitive/personal info).
- **Repository:** GitHub – free, versioned, public. Metadata/documentation in README/AsciiDoc/PDF; code in .ipynb files. Backup: Local copies + potential Zenodo upload for archiving.
- **Access:** Via HTTPS (no login required); tools provided: Python (Pandas for cleaning, Seaborn/Matplotlib for plots) in Notebooks; methods: Download CSVs directly, run notebooks with venv (dependencies listed in README).

## ERC OPEN RESEARCH DATA MANAGEMENT PLAN (DMP)

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

- **Standards:** CSV format (tabular, interoperable with Excel/Pandas/R); metadata in Markdown/AsciiDoc (readable by humans/machines). Field-specific: ECDC column names/vocabularies (e.g., 'year\_week' ISO-like, 'country' as ISO codes where possible).
- **Methods:** Standardized column names post-cleaning (e.g., 'country', 'year\_week'); qualified references to ECDC originals (URLs in metadata); data merge on common keys ('country', 'year\_week'). Interoperable with open tools (Python/R).
- **Metadata:** Human and machine readable as PDF and in its source format (markdown)
- **Plots:** Reproducible Plots which can be saved as .PNG image files

### 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)*

- **Re-usable:** All data indefinitely (no embargo); suitable for further COVID analyses (e.g., vaccine impact, regional comparisons).
- **License:** MIT License (open for reuse, modification; acknowledge ECDC as source).
- **Quality Assurance:** In data\_quality.ipynb Uniqueness/Timeliness/Validity checked via Pandas;
- **Consistency:** Date formats unified, redundant columns removed). Cleaning: NaN rows dropped, merge on 'year\_week'.
- **Provenance:** Git commits/log.json track changes.
- **Re-use Potential:** Living document (update via Git)

## ERC OPEN RESEARCH DATA MANAGEMENT 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)*

- **Costs:** zero (GitHub free, open-source tools: Python 3.12, Pandas 2.2.2, etc.; no additional hosting).
- **Value:** Long-term preservation enables future research on pandemics (e.g., testing strategies' impact); historical value for policy analysis.
- **Backup/Recovery:** GitHub + local backups; version control via Git for recovery. No sensitive data (aggregated, no personal info). Secure Storage: Public GitHub; plan Zenodo for curation/DOI.

**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.