D1.2 Data Management Plan
Version 1.0

Document Information

Contract Number	101143931
Project Website	www.pop-coe.eu
Contractual Deadline	M6, June 2024
Dissemination Level	PU
Nature	R
Author	Marta Garcia (BSC), Elena Markocic (BSC)
Contributor(s)	Bernd Mohr (JSC), Samir Ben Chaabane (TERATEC)
Reviewer	Christian Terboven (RWTH)
Keywords	Data management, analysis reports, website, co-design activities, FAIR data, GDPR

This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 101143931. The JU receives support from the European Union's Horizon Europe research and innovation programme and Spain, Germany, France, Portugal, and the Czech Republic.

Change Log

Version	Author	Description of Change
V0.1	Marta Garcia	Initial Draft
V0.2	Marta Garcia	Revised content after internal review
V1.0	Marta Garcia	Final version for submission

Table of Content

Executive Summary

This deliverable states how the data will be treated during the lifetime of the project, POP3 “Performance Optimization and Productivity”. It includes the main elements of the data management policy regarding all the datasets collected, processed and/or generated during the project period and afterward. Therefore, it describes which data will be collected or generated, according to which methodology and standards, whether and how this data will be shared and/or made openly accessible, and finally, how it will be curated and preserved.
According to FAIR principles for H2020 projects[1], this document outlines the data management life cycle for all datasets to make research data findable, accessible, interoperable, and reusable (FAIR).
The Data Management Plan is a living document. POP management team is responsible for updating this document, collecting feedback from the GA members and the BSC data management experts. The document will be updated as soon as any relevant modification is implemented.

1. Introduction

The POP3 data management plan continues with the same guidelines as were implemented in the POP2 project. The basic principles are documented at the POP project website at https://pop-coe.eu/contact/privacy-policy. This document is structured as follows. First, it presents how the project classifies the different kind of data that is generated during the project in the different work packages (WP) and tasks (T). Then, it explains how each kind of data is treated. Finally, it explains how the FAIR data principles are achieved.

2. Data classification and analysis

The data managed by POP3 can be classified into these types:

1. Data from POP3 customers (T2.3 and T2.4)
2. Data generated for marketing and dissemination activities (T2.1, T2.2 and T2.3)
3. Performance data gathered from POP3 assessments and second level services (T3.1 and T3.2)
4. Code changes or mock up codes developed within a second level service (T3.2)
5. Data generated by the co-design activities (T4.1 and T4.2)
6. Data generated by the methodology task (T4.4)
7. Code developed in T4.3

1 European Commission. Guidelines on FAIR Data Management in Horizon 2020, Version 3.0, 26 July 2016.

2.1 Data from POP3 customers

Data from POP3 customers is provided by the customers through the different questionnaires that they fill. In particular, there are two kinds of questionnaires: the ones filled before the assessment and the so-called “customer advocacy survey” after the completion of an assessment.

The data gathered through the forms before the assessment is stored in the TRAC system. This data is only accessible within the project and will not be shared in any case. We will use statistics and aggregated data to obtain an overview of the status of codes, the scale at which they operate and their field of science. In particular, this data is necessary to report the KPIs.

The collected data for the evaluation of the customer satisfaction is detailed in POP3 Deliverable D2.1 Customer feedback methodology and the data exploitation and management comply with Terms and Conditions[2] and the Data Privacy Policy[3] of the POP project as explained in section 2.2.2 “Participation Terms and Data Privacy Policy“ of D2.1.

2.2 Data collected and generated for marketing and dissemination activities (T2.1, T2.2 and T2.3)

The POP website (https://pop-coe.eu/) uses the "Google Analytics" service offered by Google Inc. (1600 Amphitheater Parkway Mountain View, CA 94043, USA). Google Analytics is based on cookies. The information collected by the cookies is usually sent to a Google server in the US and stored there. Visitors have the option to prevent the storage of cookies on their device by making the appropriate settings in their browser.

Furthermore, they can use a browser plug-in to prevent the information collected by cookies (including your IP address) from being sent to Google Inc. and used by Google Inc. The following link leads them to the corresponding plugin: https://tools.google.com/dlpage/gaoptout.

Google LLC complies with European data protection law and is certified under the Privacy Shield Agreement.

The recorded IP address of the users of this website is shortened within the member states of the EU and the European Economic Area and in the other contracting states of the agreement. Only in individual cases, the IP address is initially transferred unabridged to the United States to a Google server and shortened there. This reduction eliminates the personal reference of your IP address. The user's IP address provided by the browser will not be combined with other data stored by Google.

2 https://pop-coe.eu/sites/default/files/public/popw1/POP2_Terms_and_Conditions.pdf
3 https://pop-coe.eu/contact/privacy-policy

The data collected by Google on our behalf will be used to evaluate the use of our online offering by individual users, such as to generate activity reports on the website to improve our online offering.

POP also maintains a quarterly newsletter service. Interested persons need to signup (opt-in) for the newsletter by providing an email address. No further information is collected. The list of emails is stored as part of the POP webserver hosted at BSC. People can un-subscribe at any time.

On the POP website, blog posts about project news, results and success stories are posted regularly. We only use pictures captured by POP project members and collect the publication permission of all persons visible on these pictures.

A minimal amount of data (typically name, organization, email address, and gender) is collected when persons register for online (e.g. webinars) or face-to-face (e.g. training) events organized by POP. For this purpose, we use an instance of the event management system indico3, hosted by the German Helmholtz association, the umbrella organization of the (POP project member JUELICH). The data related to POP events is only accessible to the POP dissemination leads, and will only be stored as long it is necessary for the successful execution of the project.

2.3 Performance data gathered from POP3 assessments and second level services (T3.1 and T3.2)

Under this category, we foresee three different kind of data: the data provided by the customer to be able to perform the assessment (code, binary and/or input set), the “raw” performance data and the performance report produced as an outcome of the assessment or the second level service.

The data provided by the customer will be stored in the machines used to perform the assessment only while it is necessary and will belong to the customer at all times.

The raw performance data obtained by POP3 will be stored in the machine used by the POP partner to conduct the audit. The only requirement is to keep the data while it is foreseen to have further activities with the same user. The performance data can be shared between partners, unless the study required an NDA. In any case, the data will not be distributed outside the consortium without previous consent from the user.

As output of the analysis, we produce a report (slides or document) with the results and observations. These documents are stored in the TRAC and wiki. If the user grants us explicit permission, they will also be published on the website.

2.4 Code changes or mock up codes developed within a second level service (T3.2)

As an outcome of the second level service, a code change or a mock-up code can be produced in addition to the report containing all the information regarding the service that is already covered in the previous section.

This code will be stored in internal repositories or clusters used by the POP partner providing the second level services. Once the service is completed, the code will be provided to the customer that will be able to make it public according to their license.

2.5 Data generated by the co-design activities (T4.1 and T4.2)

Data gathered and generated by the co-design activities will be stored in the co-design repository. This repository is a centralized database for source code and metrics associated with it. Its main goal is to provide the HPC community with a set of kernels which potentially could be used for training, dissemination, and system/software design activities.

The co-design repository was established during POP2 and is now publicly accessible: https://co-design.pop-coe.eu/

During POP3 we will continue to extend this repository. The kernels will be managed through the GitLab Community Edition, which is open source software. More information on the structure the repository and how it is maintained can be found in POP2 deliverable D7.1[4].

Private access during analysis development guarantees accesses only for POP3 team members. Once the reporting accomplishes the minimum requirements, public access will be provided on the project website.

The final goal is to have a permanent repository to be queried beyond the project duration.

2.6 Data generated by the methodology task (T4.4)

The data generated by the methodology task will consist in new metrics or advances in the methodology. For this reason, it will be produced in the form of technical papers or articles, and it can also be published on the project web page. These artifacts will not contain and user data unless with explicit permission.

4 https://pop-coe.eu/sites/default/files/pop_files/pop2_d7.1.pdf

2.7 Code developed in T4.3

All the software developed under T4.3 is open source and will be accessible through the corresponding repository.

The access to download the tools is provided in the dedicated section of the project website: https://pop-coe.eu/partners/tools

3. FAIR Data

Relevant results of the project with respect to the POP methodology will be published on the project website, the co-design repository, other institutional repositories, Zenodo (https://zenodo.org/) or in open access scientific publications.

Main documents (performance assessment and second level services reports) are described in Section 2, but only those with previous customer consent will be finally published on the website. These documents follow an internal naming convention that remains when published on the website:

- Performance assessment: POP3_AR_nnn_xxx
  - Where “nnn” is a unique identifier among all assessment reports
  - “xxx” is the name of the code assessed.

- Second level services: POP3_YYY_nnn_xxx
  - Where YYY encodes the kind of second level service:
    - PoC: Proof-of-concept
    - CC: Correctness Check
    - EE: Energy Efficiency
    - AS: Advisory Study
  - “nnn” is a unique identifier among all second level service reports.
  - And “xxx” is the name of the code.

A subgroup of these documents will also be published under the website section “success stories” (https://pop-coe.eu/target-customers/success-stories) where the POP service activity will be detailed. Success stories will also link to the original corresponding report.

The co-design repository may also publish results derived from a performance assessment or proof-of-concept public reports. In these cases, the co-design website will also link the original document utilizing its URL.

All the kernels included in the co-design repository will contain a license file describing the terms of use. The project will encourage that all of them are based on GPL or LGPL licenses to avoid any kind of restriction on terms of use and distribution.

Another option is the resulting (scientific-technical) papers, where lessons learned from multiple similar performance assessments will be summarized and documented for outside readers.

Acronyms and Abbreviations

- CoE – Center of Excellence
- D – Deliverable
- DMP – Data Management Plan
- EC – European Commission
- GA – General Assembly
- GDPR - General Data Protection Regulation
- GitLab – Open source application for the entire software development lifecycle
- GPL - General Public License
- HPC – High Performance Computing
- LGPL - Lesser General Public License
- NDA – Non Disclosure Agreement
- POP – Performance Optimization and Productivity
- TRAC – Wiki and issue tracking software
- USTUTT-HLRS – University of Stuttgart - High Performance Computing Center Stuttgart
- URL – Universal Resource Locator
- WP – Work Package

POP3 Beneficiaries

- BSC: BARCELONA SUPERCOMPUTING CENTER - CENTRO NACIONAL DE SUPERCOMPUTACION
- FZJ: FORSCHUNGSZENTRUM JÜLICH GMBH
- RWTH: RHEINISCH-WESTFAELISCHE TECHNISCHE HOCHSCHULE AACHEN
- IT4I@VSB: TECHNICAL UNIVERSITY OF OSTRAVA
- INESC ID: INSTITUTO DE ENGENHARIA DE SISTEMAS E COMPUTADORES, INVESTIGACAO E DESENVOLVIMENTO EM LISBOA
- TERATEC: TERATEC
- UVSQ: UNIVERSITE DE VERSAILLES SAINT-QUENTIN-EN-YVELINES
- USTUTT: UNIVERSITY OF STUTTGART FOR ITS HIGH PERFORMANCE COMPUTING CENTER STUTTGART