



D2.3 Data Management Plan Version 1.0

Document Information

Contract Number	801051
Project Website	www.epeec-project.eu
Contractual Deadline	M6, April 2019
Dissemination Level	PU
Type	R
Author	Antonio J. Peña (BSC)
Contributor(s)	Nadia Tonello (BSC)
Reviewer	João Barreto (INESC-ID)
Keywords	Data Management Plan



The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No "801051".

Change Log

Version	Author	Description of Change
V0.1	A. J. Peña	Initial Draft
V0.2	A. J. Peña	Added more details
V1.0	A. J. Peña	Minor fixes

Table of Contents

Executive Summary	4
1. Data Summary	4
1.1 Unique Data Set – Software Management.....	5
1.1.1 Programming language	5
1.1.2 Rules and best coding practices.....	5
1.1.3 <i>Access permissions and license</i> (lifecycle dependency: only specific groups of collaborators, all partners, whole community, ...).....	5
1.1.4 Repository structure (branches, tags, ...).....	5
1.1.5 Backup and Archiving needs	5
1.1.6 <i>Legal/ethical restrictions</i> – privacy and security issues, if any	6
1.1.7 Versioning control and rules/workflows managing	6
1.1.8 Code transfer needs and security.....	6
1.1.9 Long term preservation needs.....	6
1.1.10 Documentation and inline comments rules.....	6
2. FAIR Data	6
2.1.1 Findability: DOIs and how will data be found.....	6
2.1.2 Open access standards: free protocols, long term metadata availability	7
2.1.3 Interoperability: accessible, shared language, referenced to metadata.	7
2.1.4 License of reuse	7
2.1.5 Services and list of datasets which need them.....	7
3. Responsibilities	7
3.1.1 DMP updates plan	7
3.1.2 Storage content (move/remove data), Backups, Permissions control/change...	7
Acronyms and Abbreviations	8

Executive Summary

EPEEC partners elaborate the DMP to comply with EC's objective of making research data findable, accessible, interoperable, and reusable (FAIR). The DMP will be updated during the lifespan of the project. The DMP describes the life cycle for all data sets that will be collected, processed, or generated by the project. It is a document outlining how research data will be handled during a research project, and even after the project is completed, describing what data will be collected, processed, or generated, and following which methodology and standards, whether and how this data will be shared and/or made open, and how it will be curated and preserved.

1. Data Summary

The data generated by the project will be the software developments of the programming environment components and those of the applications involved in the project in the co-design effort. Where open source components are involved, code sources will be collected; otherwise, binaries will be stored.

EPEEC will feature an integrated testing and development platform that will be hosted in a mini-cluster owned by the project, managed by BSC. Centralized git repositories located in this system will be hosting the outcome products of the project. Partners responsible of a piece of software will push their developments into the corresponding repository. Relevant commits will be tagged appropriately, e.g., referencing project deliverable codes.

Documentation will follow the customary format of the different software components. Joint documentation will be developed as part of Task 4.6, being Deliverable D4.6. This document will be stored in the corresponding git repository and released to be publicly available on the project's website.

git's changelog will be used as the official changelog.

Testing procedures are described in Task 4.4.

The different components of the EPEEC framework will be distributed on EPEEC's website (not precluding their separate distributions on the respective owners' website) following their respective licenses.

1.1 Unique Data Set – Software Management

1.1.1 Programming language

The different software components and applications involved in EPEEC use different programming languages. Indeed, multi-programming language support is one of the core features of EPEEC.

Appentra's Parallelware contributions will be stored in binary format, given that their software component is not open source.

1.1.2 Rules and best coding practices

EPEEC, being composed of a heterogeneity of already-existing software components and applications, will not impose any coding practice. Since all software components are mature, they all follow proper coding practices. Moreover, since all software components are designed to work separately as stand-alone pieces, we do not need to enforce common coding practices. A user's guide will be developed as part of Task 4.6, being Deliverable D4.6.

1.1.3 Access permissions and license (lifecycle dependency: only specific groups of collaborators, all partners, whole community, ...)

The git repositories will be private, accessible only by EPEEC partners. Public releases will be available from the project's website, including a changelog and appropriate documentation.

1.1.4 Repository structure (branches, tags, ...)

Different repositories under the project's git server will exist for the different software components. Public releases will be tagged.

External dependencies and inter-dependencies will be described in the accompanying documentation.

1.1.5 Backup and Archiving needs

Daily incremental backups will be set.

1.1.6 Legal/ethical restrictions – privacy and security issues, if any

No partners should release any contents of the git repositories not owned by them in a way that may be reached by any individual not belonging to the EPEEC consortium.

All the members of the consortium will have automatically read permission to all internal git repositories, plus write permission to those where they are actively contributing. The membership application should be addressed by e-mail to the project's Technical Manager.

1.1.7 Versioning control and rules/workflows managing

Different software components are free to follow their versioning conventions. Public releases must be tagged with their version.

A code version will be tagged at the criteria of the owners of the software piece. Relevant events granting a tag include public or internal releases and paper publications.

1.1.8 Code transfer needs and security

There will be per-user access control. The repository will be handled by BSC's personnel from the AccelCom team, under the supervision of the project's Technical Manager.

1.1.9 Long term preservation needs

The contents of the git repositories will be saved at least during the length of the project. We expect partners to incorporate EPEEC's developments into their streamline products. We also expect to have follow-up projects.

1.1.10 Documentation and inline comments rules

Documentation will be generated according to the EPEEC consortium agreement, as described in Task 4.6 (public Deliverable D4.6).

2. FAIR Data

2.1.1 Findability: DOIs and how will data be found

Release versions will be tagged by git's system and made publicly available at EPEEC's website.

2.1.2 Open access standards: free protocols, long term metadata availability

Every software version used at a paper is highly encouraged to be tagged and made publicly available to ensure reproducibility.

2.1.3 Interoperability: accessible, shared language, referenced to metadata.

All software components will interoperate according to EPEEC's objectives.

2.1.4 License of reuse

All software components shall include a file explaining their license terms for the public, which shall be no different from the terms they had prior to EPEEC's developments. All EPEEC partners are allowed to freely use all contents of the git repository, as long as the intended use is in the scope of EPEEC's actions.

2.1.5 Services and list of datasets which need them

EPEEC users should request personal access to EPEEC's repositories to the BSC designate.

3. Responsibilities

3.1.1 DMP updates plan

EPEEC's Technical Manager is responsible for updating this document, gathering feedback from the GA and BSC's data management expert. It will be updated as soon as any addition, refinement, or change to the current conditions occurs.

3.1.2 Storage content (move/remove data), Backups, Permissions control/change

The Technical Manager will delegate this task to members of his team.

Acronyms and Abbreviations

- APPENTRA – Appentra Solutions, S.L.
- BSC – Barcelona Supercomputing Center
- CA – Consortium Agreement
- CERFACS – Centre Europeen de Recherche et de Formation Avancee en Calcul Scientifique
- CINECA – CINECA Consorzio Interuniversitario
- D – Deliverable
- DMP – Data Management Plan
- EPEEC – European joint Effort toward a Highly Productive Programming Environment for Heterogeneous Exascale Computing
- ETA – Eta Scale AB
- FAIR – Findable, Accessible, Interoperable, and Reusable
- Fraunhofer- Fraunhofer Gesellschaft zur Foerderung
- GA – General Assembly / Grant Agreement
- IMEC – Interuniversitair Micro-electronica Centrum
- INESC-ID – Instituto de Engenharia de Sistemas e Computadores, Investigacao e Desenvolvimento em Lisboa
- INRIA – Institut National de Recherche en Informatique et Automatique
- M – Month, Milestone
- UU – Uppsala Universitet