D1.4

Data Management Plan for pilot on Open Research Data

Workpackage:	1	Management	
Author(s):	Denis Dutoit		CEA
Authorized by	Denis Dutoit		SCAPOS
Reviewer	Guy Lonsdale		SCAPOS
Reviewer	Paul Carpenter		BSC
Reviewer	Elyes Zekri		ATOS
Dissemination Level	Public		

Date	Author	Comments	Version	Status
2017-02-24	Denis Dutoit	Initial draft	V0.0	Draft
2017-03-09	WP leaders	Updates from WP1 Updates from WP2 Added datasets from WP3 Added datasets from WP5	V0.1	Draft
2017-03-27	Denis Dutoit	The comments of the reviewers are addressed	V1.0	Final

Executive Summary

A Data Management Plan (DMP) describes in detail how reports and data are handled within the ExaNoDe project. It is critical to develop a DMP that outlines the practices for collecting, organizing, backing-up, and storing the data that are generated throughout the project. Within this deliverable D1.4, the data management plan for the Pilot on Open Research Data is defined, while it will be updated during the lifespan of the project so as to present the status of the project's reflections on data management.

Table of Contents

Table of Figures

Table of Tables

1 Introduction

This Data Management Plan (DMP) describes the data management life cycle for all datasets that will be collected, processed or generated by the ExaNoDe project. This document outlines how research data will be handled during the project and after the project completion.

The DMP is not a fixed document, but it evolves during the lifespan of the ExaNoDe project. This is the first version of the DMP which has been aligned with the amendment reference No AMD-671578-13. The DMP will be updated according to project needs.

Several categories of datasets are identified within the ExaNoDe project:

- **Reporting material** such as Consortium Agreement, Grant Agreement, deliverables or any other kind of material exchanged with the European Commission.
- **Scientific publications, presentations and dissemination material** that describe the research work within ExaNoDe.
- **Technical datasets** including the technical work such as source code of tools, libraries, RTL codes, netlist, design scripts, etc.
- **Evaluation datasets** that accompany the scientific publications and/or deliverables and usually provide more information than the one included in the publications.

This DMP addresses the points below on a dataset by dataset basis and reflects the current status of reflection within the consortium about the data that will be produced:

- **Data set reference and name:** identifier for the data set to be produced.
- **Data set description:** description of the data that will be generated or collected, its origin (in case it is collected), nature and scale and to whom it could be useful.
- **Standards and metadata:** reference to existing suitable standards of the discipline.
- **Data sharing:** description of how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups. Identification of the repository where data will be stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.).
- **Archiving and preservation (including storage and backup):** description of the procedures that will be put in place for long-term preservation of the data. Indication of how long the data should be preserved.

The following sections provide the current status of reflection within the consortium about the data that will be produced for the various identified categories.

2 Reporting material

Several non-technical documents for reporting the project progress with the European Commission will be produced within ExaNoDe, such as the Consortium Agreement, Grant Agreement, and meeting minutes.

A list of these datasets is provided in Table 1.

Type	Reference and name	Description	Standards and metadata	Data sharing	Archiving and preservation
Document	Grant Agreement number: 671578	ExaNoDe Grant Agreement electronically signed by all consortium members	pdf file	Codendi repository[1] for ExaNoDe consortium members, EC repository[2] for Commission services.	Dataset will be available on Codendi at least two years after the end of the project.
Document	Consortium Agreement ref. 22773	ExaNoDe Consortium Agreement signed by all consortium legal entities	Paper copy	All partners have the same hard copy of the CA. An electronic copy is shared on Codendi repository[1].	Hard copy must be maintained and preserved by each partner even after the end of the project (at least 6 years).
Document	Amendment Reference No AMD-671578-13	ExaNoDe amendment electronicaly signed by the Commission and the Coordinator	pdf file	Data sharing on: Codendi repository[1] for ExaNoDe consortium members, EC repository[2] for Commission services.	Dataset will be available on Codendi at least two years after the end of the project.
Guide	Project manual	ExaNoDe project manual	Microsoft Office documents	Data sharing on Codendi repository[1] for ExaNoDe consortium members.	Dataset will be available on Codendi at least two years after the end of the project.
Minutes	Meeting minutes	ExaNoDe meeting minutes	pdf files	Data sharing on Codendi repository[1] for ExaNoDe consortium members.	Dataset will be available on Codendi at least two years after the end of the project.
Document	Periodic Progress Report	ExaNoDe Periodic Progress Reports (for check meetings with the Commission and formal reviews)	pdf files	Data sharing on Codendi repository[1] for ExaNoDe consortium members. An electronic copy is provided to the reviewers and the Project Officer before each check meeting or review.	Dataset will be available on Codendi at least two years after the end of the project.
Presentations	Project meeting, check meeting and formal review presentations	ExaNoDe presentations made during the project meetings, the check meetings and the formal reviews	pdf files	Data sharing on Codendi repository[1] for ExaNoDe consortium members. An electronic copy is provided to the reviewers and the Project Officer before each check meeting or review.	Dataset will be available on Codendi at least two years after the end of the project.
Document	Memorandum-of-Understanding	Memorandum-of-Understanding signed between the coordinators of the projects: ExaNoDe, ExaNeSt and EcoScale.	pdf file	Data sharing on Codendi repository[1] for ExaNoDe consortium members.	Dataset will be available on Codendi at least two years after the end of the project.

Table 1: Reporting datasets

Notes from table:

(1) Codendi is a project management environment (https://minalogic.net/account/login.php) used for ExaNoDe, which offers an easy, secured and consortium-limited access to ExaNoDe project datasets. Figure 1 shows the document menu which can be used to upload and download reporting materials including technical deliverables.
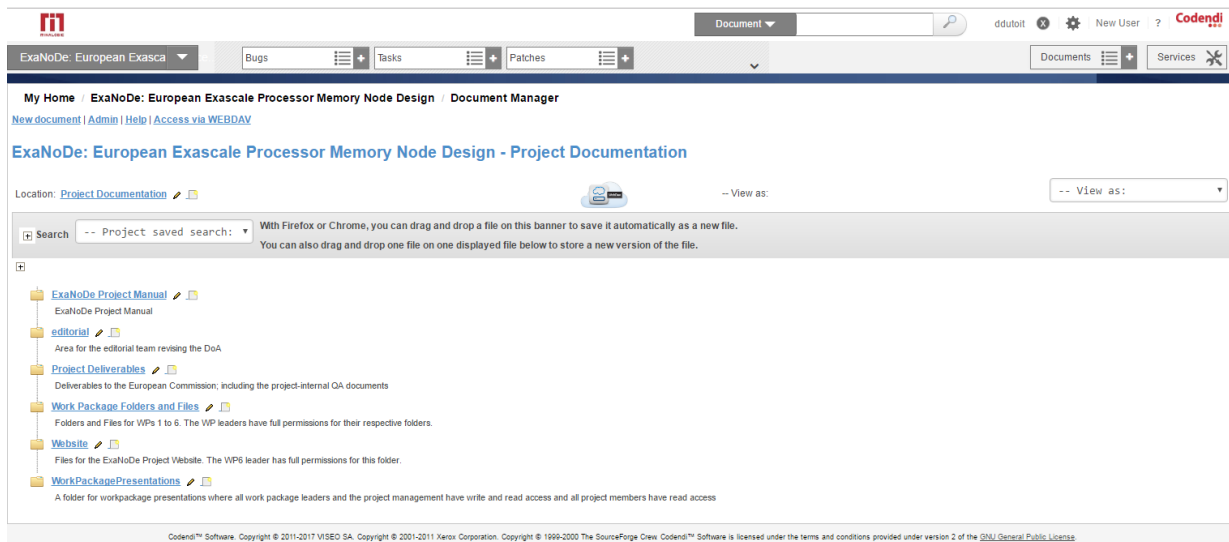
Figure 1: Codendi document repository of the ExaNoDe project

(2) Research Participant Portal to manage the ExaNoDe project and communicate with the European Commission throughout the project's life cycle :
http://ec.europa.eu/research/participants/portal/desktop/en/home.html

Apart from the above non-technical reporting dataset, ExaNoDe deliverables are considered to be scientific reports of the work within the project. Several deliverables will be publicly available and posted on the ExaNoDe website in PDF format. Deliverables will be stored in the Codendi repository in Microsoft Word format. The table below lists the deliverables based on their type and dissemination level.

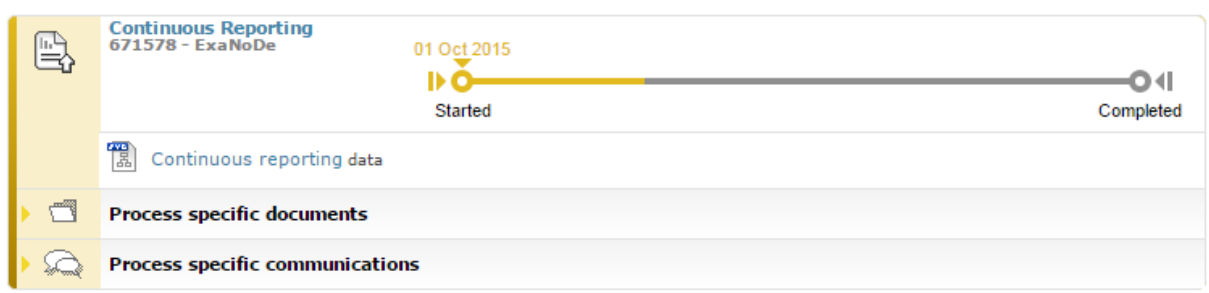| Type	Reference and name	Dissemination level	Data sharing	Archiving and preservation
Reports	D1.1, D1.3, D3.4, D3.9, D4.1, D4.2, D4.7, D6.4, D6.7	Confidential	Codendi repository for ExaNoDe consortium members, EC continuous reporting repository[3] for Commission services.	Dataset will be available on Codendi at least two years after the end of the project.
Others	D3.5, D4.3, D4.5, D5.1, D5.3, D5.4	Confidential	Codendi repository for ExaNoDe consortium members, EC continuous reporting repository[3] for Commission services.	Dataset will be available on Codendi at least two years after the end of the project.
Demonstrators	D4.4, D4.6	Confidential	Codendi repository for ExaNoDe consortium members, EC continuous reporting repository[3] for Commission services.	Dataset will be available on Codendi at least two years after the end of the project.

ORDP: Open Research Data Pilot	D1.4	Public	Codendi repository for ExaNoDe consortium members, EC continuous reporting repository[3] for Commission services, ExaNoDe website[4] for public access.	Dataset will be available at least two years after the end of the project.
Reports	D2.1, D2.2, D2.3, D2.4, D2.5, D2.6, D2.7, D3.1, D3.2, D3.6, D3.7, D5.2, D5.5, D6.2, D6.5	Public	Codendi repository for ExaNoDe consortium members, EC continuous reporting repository[3] for Commission services, ExaNoDe website[4] for public access.	Dataset will be available at least two years after the end of the project.
Demonstrators	D3.3, D3.8	Public	Codendi repository for ExaNoDe consortium members, EC continuous reporting repository[3] for Commission services, ExaNoDe website[4] for public access.	Dataset will be available at least two years after the end of the project.
Others	D6.1, D6.3, D6.6	Public	Codendi repository for ExaNoDe consortium members, EC continuous reporting repository[3] for Commission services, ExaNoDe website[4] for public access.	Dataset will be available at least two years after the end of the project.

Table 2: Deliverables

Notes from table:

(3) Figure 2 shows the continuous reporting menu which is used for reporting the ExaNoDe documentation with the European Commission.

Figure 2: Continuous reporting repository of the ExaNoDe project:

http://ec.europa.eu/research/participants/portal/desktop/en/home.html

(4) http://exanode.eu/

3 Scientific publications, presentations and dissemination materials

There is a need to maximise the impact of European collaborative projects, and this is one of the primary goals of the European Commission's funding schemes, aiming at promoting Europe's strategic position in target technical fields. The ExaNoDe consortium considers dissemination activities to be as important as the technical work carried on within each task, to maximize the impact of the project and get feedback from outside the project environment to drive the work performed in a successful manner.

In the ExaNoDe project, the partners plan to give great importance to dissemination, by presenting posters or publishing scientific papers to international conferences or journals, by participating in events such as workshops organized by the European community and by providing dissemination materials with project website, flyers, and so on. For this purpose, a large amount of dissemination data will be generated throughout the project.

For more information on the planned dissemination activities, deliverable D6.2 "Dissemination strategy Document" presents the plan for the dissemination of the ExaNoDe project outcomes, and the project manual [1] sets the basic rules for publications, presentations and copyright usage. This section of the DMP, describes in detail how dissemination data will be handled within the ExaNoDe project.

The following table provides the available scientific publications, presentations and dissemination materials, as at M18 of the ExaNoDe project:

Type	Reference and name	Description	Data sharing	Archiving and preservation
Article	ExaNoDe @ HPC Wire 2016	HPC Wire article on "EU Projects Unite on Heterogeneous ARM-based Exascale Prototype" published on Feb 24th 2016.	HPC Wire repository	
Poster	ExaNoDe @ DATE 2016	Poster presented at DATE 2016 in Dresden by CEA	Codendi repository for ExaNoDe consortium members; ExaNoDe website for public access.	Dataset will be available at least two years after the end of the project.
Poster	ExaNoDe @ HiPEAC 2017	Poster presented at HiPEAC 2017 in Stockholm by CEA	Codendi repository for ExaNoDe consortium members; ExaNoDe website for public access.	Dataset will be available at least two years after the end of the project.
Presentation	ExaNoDe @ ISC 2016	Presentation at workshop on International Cooperation at ISC 2016 in Frankfurt by BSC.	Codendi repository for ExaNoDe consortium members.	Dataset will be available at least two years after the end of the project.
Presentation	ExaNoDe @ Mont-Blanc workshop 2017	Presentation at Mont-Blanc project workshop in Barcelona by CEA.	Codendi repository for ExaNoDe consortium members.	Dataset will be available at least two years after the end of the project.
Publication	ExaNoDe @ ISC 2017	Position paper submitted at ISC	Codendi repository for ExaNoDe consortium members.	Dataset will be available at least two years after the end of the project.

		2017 in Frankfurt by VOSYS.		
Publication	ExaNoDe @ MEMSYS 2016	Publication by BSC at MEMSYS 2016	Codendi repository for ExaNoDe consortium members.	Dataset will be available at least two years after the end of the project.
Publication	ExaNoDe @ IWOMP 2016	Publication by UOM at IWOMP 2016	Codendi repository for ExaNoDe consortium members.	Dataset will be available at least two years after the end of the project.
Publication	ExaNoDe @ ISPASS 2016	Publication by UOM at ISPASS 2016	Codendi repository for ExaNoDe consortium members.	Dataset will be available at least two years after the end of the project.
Publication	ExaNoDe @ PACT 2016	Publication by UOM at PACT 2016	Codendi repository for ExaNoDe consortium members.	Dataset will be available at least two years after the end of the project.
Poster	ExaNoDe @ ACACES 2016	Poster presented at ACACES 2016 in Fiuggi by FORTH	Codendi repository for ExaNoDe consortium members.	Dataset will be available at least two years after the end of the project.
Website	ExaNoDe project website	exanode.eu	Website subcontractor premises	Dataset will be available at least two years after the end of the project.
Flyer	ExaNoDe flyer	Public information about the project, its objectives and future achievements	Website for public access.	Dataset will be available at least two years after the end of the project.

Table 3: Scientific publications, presentations and dissemination materials

Table 3 will be amended throughout the project, in order to include newly generated dissemination data.

The publications and related research data will be publicly provided by research data repositories respecting the policies and rules set out by the publishers (journals or conferences). The partners will use an open access repository, connected to the tools proposed by the European Commission (e.g. OpenAIRE), to grant access to the publications and to a bibliographic metadata in a standard format including information requested by the European Commission. Moreover, some posters shown in Table 3 will be posted on the ExaNoDe website.

All scientific publications involving BSC will be uploaded to the UPCommons open access repository of *Universitat Politècnica de Catalunya* (UPC).

4 Technical Datasets

The technical datasets generated by the ExaNoDe project include:
- Mini-application and performance-critical kernel codes (from WP2 "Co-Design for Exa-scale HPC systems");
- Firmware, OS, virtual machine, parallel programming models and runtime libraries codes (from WP3 "Enablement of Software Compute Node");
- System-on-Chip design databases: RTL code, hard-macro design, gate netlist, design scripts, design environment, GDS2 file (from WP4 "Compute node design and manufacture");
- Board design datasets (from WP5 "System Integration & Evaluation").

The following technical datasets are foreseen for ExaNoDe project:

Type	Reference and name	Description	Standards and metadata	Data sharing	Archiving and preservation
Code	Mini-apps or kernels of performance-critical algorithms	Source or binary code of the mini-applications or performance critical kernels.	Binary files or plain text files.	No such datasets have been generated at M18 of the ExaNoDe project. Data sharing, and archiving policies will be described in an updated DMP.	
Binary code	Virtual machine enhanced checkpoint with post-copy	Source code of the virtual machine memory snapshot based on post-copy.	plain text files (e.g., .h, .c)	The code will be released in the form of diff patches to the open source Qemu and Linux communities.	Source code available in various mailing lists archives, and in QEMU code tree once upstream.
Binary code	Virtual machine incremental checkpoint	Binary release of the Virtual Machine incremental checkpointing feature.	Binary file	Binary will be released to the ExaNoDe consortium for integration and final project prototype demonstration.	Binary released to ExaNoDe consortium for use only within the project life-time.
Binary code	ExaNoDe firmware	Realization of UNIMEM support on experimental prototypes of the ExaNoDe project (hardware design for FPGAs together with Linux device drivers)	Binary files	The firmware, in the form of binary code (FPGA bitstreams, Linux device drivers and kernel configuration), will be installed in the shared testbed, hosted at FORTH premises, available for use by partners.	Dataset will be available at least two years after the end of the project.
Binary code	ExaNoDE operating system	Enhancements and additions to the Linux kernel (together with API libraries) for exposing Unimem platform functionality to programming models and end-user applications.	Binary files	The operating system, in the form of binary code (Linux with Unimem support and low-level API libraries), will be installed in the shared testbed, hosted at FORTH premises, available for use by partners.	Dataset will be available at least two years after the end of the project.
Source code	UNIMEM-optimized MPI library	MPI library that has been optimized for use with UNIMEM memory scheme	Implementation of the MPI standard consisting of plain text files (e.g., .h, .c)	The UNIMEM MPI source code will be maintained by BSC and FORTH and available from a git or SVN repository hosted at BSC. The code will be freely downloadable with an open source licence.	The UNIMEM MPI source code will be available at least three years after the end of the project.

Source code	GPI	PGAS based distributed one-sided and asynchronous programming model.	GPI implements the GASPI standard to be found at: www.gaspi.de	The GPI code will be maintained at Fraunhofer's premises and released with an open source licence.	Dataset will be available at least five years after the end of the project.
Source code	Mercurium	OmpSs compiler	plain text files (e.g., .h, .c)	The Mercurium compiler source code is maintained by BSC and freely downloadable from a git repository hosted at BSC, with the LGPL licence.	The Mercurium source code will be available at least 3 years after the end of the project.
Source code	Nanos 6	OmpSs runtime system	plain text files (e.g., .h, .c)	The Nanos 6 runtime system source code is maintained by BSC in a git repository hosted at BSC. Before the end of the project the code will be freely downloadable with an open source licence.	The Nanos 6 source code will be available at least 3 years after the end of the project.
Source code	OpenStream	OpenStream compiler and runtime system	Plain text files (e.g., .h, .c)	The OpenStream source code is maintained by UoM and publicly available through a dedicated portal (www.openstream.info) and git repository hosted at UoM. The runtime system code is freely available under GPLv2 license, the compiler is based on the GNU C Compiler and inherits its licenses (mostly GPLv2 and 3).	Dataset will be available at least two years after the end of the project.
Source code	Thermal management	Power capping and thermal management	plain text files (e.g., .h, .c, .py)	The thermal management runtime code is maintained by ETHZ in a git repository hosted at ETHZ. Before the end of the project the code will be freely downloadable with an open source licence.	The thermal management source code will be available at least 3 years after the end of the project.
Chiplet RTL code	ExaNode_Chiplet_RTL	RTL code of chiplet design	VHDL, Verilog… files	The RTL code will be maintained at CEA's premises.	Dataset will be available at least two years after the end of the project.
ExaConv RTL code	ExaNode_ExaConv_RTL	RTL code of the Convolution Hardware operator	VHDL, Verilog… files	The RTL code will be maintained at ETHZ's premises.	Dataset will be available at least two years after the end of the project.
UoM RTL code + Rx/Tx cells	ExaNode_UoM_Macros	Rx/Tx Hard macro	VHDL, Verilog… files	The RTL code + Hard Macro will be	Dataset will be available at least

				maintained at UoM's premises.	two years after the end of the project.
Chiplet netlist	ExaNode_Chiplet_Netlist	Mapping of the RTL code of the chiplet design onto 28FDSOI technology	Verilog netlist	The RTL code will be maintained at CEA's premises.	Dataset will be available at least two years after the end of the project. Snapshot backup policy is in place + copy in another room on a daily basis with 6 months retention of the data
Design scripts	ExaNode_Design_Scripts	Scripts and constraints used to generate both netlist and GDS starting from RTL code	sh, csh, tcl, sdc …. files	The design scripts will be maintained at CEA's premises.	
GDS2 file	ExaNode_GDS2	Output of the design flow, file sent to the foundry	GDS2	The GDS2 code will be maintained at CEA's premises.	
Verification environment	ExaNode_Verif_Env	Testbench and associated files used to validate the RTL code both functionally and in test mode	sh, csh, tcl, C …. files	The verification environment plaform will be maintained at CEA's premises.	
FPGA RTL code	ExaNode_Chiplet_FPGA_RTL	RTL code for FPGAs chiplet programming	VHDL, Verilog files	The RTL code will be maintained at FORTH's premises.	Dataset will be available at least two years after the end of the project.
FPGA scripts	ExaNode_Chiplet_FPGA_Scripts	Script for program load onto the FPGAs chiplet	sh, csh, tcl, C files	The script bunch will be maintained at FORTH's premises.	Dataset will be available at least two years after the end of the project.
FPGA Design	ExaNode_CFPGA1_RTL	RTL code and Vivado files for MCM Compute FPGA1 programming	VHDL, Verilog, Vivado files	The RTL code and Vivado files will be maintained at FORTH's premises.	Dataset will be available at least two years after the end of the project.
FPGA Design	ExaNode_CFPGA2_RTL	RTL code and Vivado files for MCM Compute FPGA2 programming	VHDL, Verilog, Vivado files	The RTL code and Vivado files will be maintained at FORTH's premises.	Dataset will be available at least two years after the end of the project.
FPGA scripts	ExaNode_CFPGA_Scripts	Script for program load onto the MCM FPGAs	sh, csh, tcl, C files	The script bunch will be maintained at FORTH's premises.	Dataset will be available at least two years after the end of the project.
Schematics	ExaNodeDemonstration board schematics	Schematics in ORCAD format	.opj	The schematics will be maintained at FORTH's premises.	Dataset will be available at least two years after the end of the project
PCB Layout	ExaNodeDemonstration board layout	Layout in Allegro format	.brd and gerber files	The PCB layout files will be maintained at FORTH's premises.	Dataset will be available at least two years after the end of the project

Table 4: Technical Datasets within ExaNoDe

Some of the technical datasets will be maintained at the partners' premises and made available for ExaNoDe partners for prototype integration. This could also be done through the CVS repository of the Codendi web-based project management environment for ExaNoDe. It offers an easy, secured and consortium limited access to ExaNoDe project datasets.

Moreover several of the technical datasets will be provided under open-source licenses and are publicly available for download from a git or SVN repository hosted at the partner's premises (see Table 4 for related datasets).

Aiming to reach the widest audience possible, and accomplishing dissemination and communication strategy and with the aim of getting sustainability at the end of the project, the ExaNoDe website can provide descriptions and links to the various individual sites hosting the source code licensed under any kind of open-source license. This kind of decisions will be taken along the project lifecycle and thus, the DMP will be updated accordingly.

5 Evaluation Datasets

The evaluation datasets accompany the scientific publications and deliverables. These datasets include evaluation and performance measurements of the ExaNoDe architecture and prototype. We foresee the following evaluation datasets:

Type	Reference and name	Description	Standards and metadata	Data sharing	Archiving and preservation
Simulation results	Chiplet simulation results	Simulation results from the verification testbench	Completed verification plan	The simulation results will be available at CEA's premises.	Dataset will be available at least two years after the end of the project. Snapshot backup policy is in place + copy in another room on a daily basis with 6 months retention of the data
Measurements	Critical kernels performances	Performance measurements of mini-applications or critical kernels	Measurement result files	The results will be available from the Codendi SVN repository for project internal sharing.	Dataset will be available at least two years after the end of the project.
Measurements	Synthetic benchmark performances	Performance measurements of synthetic benchmarks	Measurement result files	The results will be available from the Codendi SVN repository for project internal sharing.	Dataset will be available at least two years after the end of the project.
Measurements	Micro-benchmarks for UNIMEM mechanisms	Performance measurements of synthetic benchmarks using the UNIMEM firmware on ExaNoDe testbench	Measurement result files	The results will be available from the Codendi SVN repository for project internal sharing.	Dataset will be available at least two years after the end of the project.
Measurements	benchmark performances	Performance measurements of communication benchmarks	Measurement result files	The results will be available from the Codendi SVN repository for project internal sharing.	Dataset will be available at least two years after the end of the project.

Table 5: Evaluation datasets

6 Concluding Remarks

This deliverable presents the Data Management Plan (DMP) of the ExaNoDe project aligned with the amendment reference No AMD-671578-13. It describes the data management life cycle for all datasets that will be collected, processed or generated by ExaNoDe project. Several categories of datasets have been identified. Categories by categories, this DMP addresses data sharing and archiving and reflects the current status of reflection within the consortium about the data that will be produced.

The DMP evolves during the lifespan of the ExaNoDe project and it will be updated according to project needs.

7 References and Applicable Documents

[1]   Wray F. ExaNoDe Project Manual, V1.3 Dec. 2016