**2nd Call for H.F.R.I.'s Research Projects to Support Faculty Members & Researchers**

**Proposal ID 03684**

**PERSIST: How to Compute in Persistent Memory Systems**



**Deliverable D1.1**

**Data Management Plan**

[D1.1 Summary 1](#_Toc2028908665)

[D1.2 Communication Tools 1](#_Toc1141146082)

[D1.3 File Sharing 1](#_Toc493985021)

[D1.4 PERSIST Web Page and Social Networks 1](#_Toc1238380506)

[D1.5 Data Management in PERSIST 1](#_Toc1101729157)

[D1.5.1 Evaluation Datasets 1](#_Toc53456926)

[D1.5.2 Scientific Publications, posters, presentations, keynotes 8](#_Toc828263106)

[D1.5.3 Other Material 13](#_Toc1030415533)

[D1.6 Open Access Pilot 14](#_Toc785684214)

[D1.7 References 14](#_Toc276865207)

*Edited by*

*Myron Tsatsarakis*

October 2023

# D1.1 Summary

Next generation computer systems will rely on emerging memory technologies, such as Non-Volatile Memory (NVM), to address the high computational demands of modern applications and provide persistence. PERSIST aspires to have significant impact in the proper understanding of the functioning of current and future NVM approaches and the capability to formally assess them, and in harnessing of the performance challenges of NVM-based recoverable computing and the better conceptualization of its performance characteristics and boundaries.

This deliverable provides the PERSIST Data Management Plan (DMP). It describes how the reports and data are handled and outlines the practices for organizing, backing up, and storing the data that will be generated and collected throughout the project. It also presents a set of tools that have been established in the context of the project for facilitating external communication and the wide dissemination of the project. This DMP complies with the guidelines on Implementation of Open Access to Scientific Publications and Research Data [ECD-OA, OAIRE].

# D1.2 Communication Tools

The official means for communication between the members of PERSIST are the Zoom conference tool and Skype instant messaging platforms.

PERSIST members participate in teleconferences which are performed using Zoom[[1]](#footnote-1). Zoom offers free web conferencing, supporting screen and desktop sharing. Specifically, Zoom supports high-definition video with integrated audio, real-time content sharing (e.g. documents, agendas, notes), recordings, the possibility of participation to meetings through mobile devices (e.g. cellular phones), and security and strict access control policies.

Sometimes, instant messaging tools may be used as the fastest way to communicate for discussing specific topics. Skype[[2]](#footnote-2) is probably one of the most widely used applications that provides instant messaging facilities, combined with voice and video support. Skype also provides file sharing and screen sharing. PERSIST members may use Skype or other instant messaging applications to communicate with each other on specific topics.

# D1.3 File Sharing

In terms of file sharing, services such as Github and Microsoft OneDrive are utilized.

Shared or public source code will be stored in Github[[3]](#footnote-3). Github is a version control system, so it provides features to access and examine previous versions of the material uploaded on it, update to previous versions of files and directories as well as of the metadata that accompanies them. With a simple interface, the Github repository can be checked out and existing working copies can be restored exactly as they were at any date in the past.

Presentation slides, documents, images and shared source code will be stored in the cloud using the Microsoft OneDrive[[4]](#footnote-2187) service. The FORTH institute provides its members with cloud storage space using the OneDrive service under the Office365 tool suite. Authorization and Authentication of the team members using the service is accomplished through the Authentication and Authorization Infrastructure (AAI) of GRNET's Delos Federation

# D1.4 PERSIST Web Page and Social Networks

An Internet web page has been developed for PERSIST. Its main goal is to diffuse the PERSIST objectives and results as wide as possible, throughout interesting stakeholders and beyond. This web site can be found at: <http://persist-project.gr/> .

All deliverables of type PU that will be produced in the context of the project will be uploaded on the web site. Additionally, the web site provides access to press releases, presentations, posters, and other material related to the project.

PERSIST has a Facebook page <https://www.facebook.com/profile.php?id=61552692594031> .The Facebook page of PERSIST is displayed in Figure 2.

PERSIST has a LinkedIn group <https://www.linkedin.com/groups/9235182/> . The LinkedIn group of PERSIST is displayed in Figure 3.

# D1.5 Data Management in PERSIST

The following categories of datasets are identified within the PERSIST project:

* **Evaluation datasets:** Collections of data series used for the evaluation of the implementations and for the validation of the results described in the publications.
* **Scientific publications:** Papers that describe the research work within PERSIST.
* **Other material:** Leaflets, documentation, dissemination material, etc.

## D1.5.1 Evaluation Datasets

In this section, we discuss a set of datasets (synthetic and real) that could be used in the PERSIST project.

|  |  |
| --- | --- |
| **Data Set Name** | PERSIST\_Random\_Dataset |
| **Data Set Type** | Collection of randomly chosen data series. |
| **Data Set Description** | This is a set of synthetic datasets with sizes from 50 million to 200 million data series composed of random walks of length 256. Each data point in the data series is produced as xi + 1 = N(xi,1), where N (0,1) is a standard normal distribution. |
| **Reuse** | The synthetic data generator code is included in the source code we will make available. Thus, it will be possible to be used by individuals for verification, further analysis and publications. |
| **Intellectual Property** | Not applicable |

|  |  |
| --- | --- |
| **Data Set Name** | PERSIST\_Seismic\_Dataset |
| **Data Set Type** | Publicly available data series collection describing seismic activity. |
| **Data Set Description** | The seismic dataset, Seismic, was obtained from the IRIS Seismic Data Access archive [Seismic]. It contains seismic instrument recordings from thousands of stations worldwide and consists of 100 million data series of size 256. |
| **Reuse** | The data are currently (publicly) available at:  <https://ln5.sync.com/dl/0b8135230/39vxx8su-tkfi7t2s-dgsvh8rp-k8ixcs8p?sync_id=12175200430004> |
| **Intellectual Property** | Not applicable |

|  |  |
| --- | --- |
| **Data Set Name** | PERSIST\_Astro\_Dataset |
| **Data Set Type** | Publicly available data series collection representing celestial objects. |
| **Data Set Description** | The astronomy dataset, Astro, represents celestial objects and was obtained from [Astro]. The dataset consists of 100 million data series of size 256. |
| **Reuse** | The data are not yet (publicly) available |
| **Intellectual Property** | Not applicable |

## D1.5.2 Scientific Publications, posters, presentations, keynotes

During the period of time that PERSIST is running, 13 papers have been produced. We anticipate the publication of more papers during the lifetime of PERSIST. A poster has also been created for the project. Table 1 provides information about these papers and posters.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Type** | **Name** | **Description** | **Sharing** | **Archiving and Preservation** |
| Scientific Publication | PERSIST\_paper\_SRDS\_2023 | Paper to be submitted to SRDS 2023 | Available on PERSIST website for public access, after publication. | <https://doi.org/10.48550/arXiv.2310.11602> |
| Scientific Publication | PERSIST\_paper\_PPoPP\_2023 | Paper to be submitted to PPoPP 2023 | Available on PERSIST website for public access, after publication. | <https://doi.org/10.48550/arXiv.2212.13557> |
| Scientific Publication | PERSIST\_paper\_ PODC\_2022 | Paper to be submitted to PODC 2022 | Available on PERSIST website for public access, after publication. | <https://doi.org/10.48550/arXiv.2205.14213> |
| Scientific Publication | PERSIST\_paper\_MEDES\_2023 | Will appear in Proceedings of the 15th International Conference on management of Digital Ecosystems, May 2023 | Available on PERSIST website for public access, after publication. | <https://doi.org/10.48550/arXiv.2310.14101> |
| Scientific Publication | PERSIST\_paper\_DISC\_2021\_I | Paper presented at the 35th International Symposium on Distributed Computing (DISC), Freiburg, Germany, October 4-8, 2021. | Available on PERSIST website for public access, after publication. | <https://doi.org/10.48550/arXiv.2108.02775> |
| Scientific Publication | PERSIST\_paper\_VLDB\_2022 | Paper in Proceedings of the VLDB Endowment | |  | | --- | | Available on PERSIST website for public access, after publication. | | <https://doi.org/10.48550/arXiv.2212.13297> |
| Poster | PERSIST\_poster\_2023 | PERSIST Poster. | Available on PERSIST website for public access. | <https://persist-project.gr/_docs/PERSIST_OCT_2023_poster.pdf> |

Table 1: Scientific Publications and posters

Table 2 summarizes information regarding presentations given in the context of PERSIST.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Type** | **Name** | **Description** | **Sharing** | **Archiving and Preservation** |
| Slides | PERSIST\_presentation\_OPODIS\_2022 | Invited Talk at 26th International Conference on Principles of Distributed Systems (OPODIS 2022) | Available at PERSIST website for public access, after publication. | <https://persist-project.gr/_docs/OPODIS2022-faturu.pdf> |
| Slides | PERSIST\_presentation\_SPTDC\_2023 | Invited Talk, School on the Practice and Theory of Distributed Computing, October-November 2023 | Available on PERSIST website for public access, after publication. | <https://persist-project.gr/_docs/MultiversionDataStructures-final-toprint.pdf> |
| Slides | PERSIST\_presentation\_SPTDC\_2023 | Invited Tutorial, International Conference on management of Digital Ecosytems, May 2023, | Available on PERSIST website for public access, after publication. | <https://persist-project.gr/_docs/MEDES-faturu-talk-May2023-toPrint.pdf> |
| Slides | PERSIST\_presentation\_PPoPP\_2023 | Paper presented at PPoPP 2023 | Available at PERSIST website for public access. | <https://persist-project.gr/_docs/PPoPP2023Presentation.pdf> |
| Video | PERSIST\_video\_ DISC\_2021 | Video of DISC 2021 paper “Space and Time Bounded Multiversion Garbage Collection”. | Available at PERSIST website for public access. | <https://www.youtube.com/watch?v=mU3JZdvapBc> |
| Slides | PERSIST\_presentation\_WP3\_emulate\_2023 | WIP slides related to WP3, Development of new techniques for conducting experimental analyses in NVM settings and for emulating restarts and other types of failures. | Available at PERSIST website for public access. | <https://persist-project.gr/_docs/Failure_Framework_Simulation_Design_v0.1.pdf> |

Table 2: Events at which a presentation about PERSIST was given

The above two tables will be updated during the project’s lifetime in order to include new publications, additional posters that will be created, and new presentations that will be given about the project. The publications and related research data will be publicly provided not only on the web site of the project, but also on repositories respecting the policies and rules set out by the publishers. The publications will be archived at arXiv[[5]](#footnote-6).

Apart from the above material, PERSIST deliverables are considered to be scientific reports of the work performed during the project. Several deliverables will be publicly available and posted on the PERSIST website. Table 3 lists the deliverables based on their dissemination level.

|  |  |  |  |
| --- | --- | --- | --- |
| **Type** | **Deliverable** | **Dissemination Level** | **Sharing** |
| Deliverable | D1.1, D1.2, D1.3, D1.4, D2.1, D2.2, D3.1, D3.2, D4.1, D4.2, D4.3, D4.4, D4.5 | PU | Available at PERSIST website for public access. |

Table 3: List of deliverables based on their dissemination level

## D1.5.3 Other Material

In the context of PERSIST, several other documents will be produced, including documentation, press releases, videos, and others. A list of these datasets is provided in Table 4.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Type** | **Name** | **Description** | **Sharing** | **Archiving and Preservation** |
| Article | PERSIST\_article | Article in magazine HiPEAC info, No 69, Innovation Europe, page 29, July 2023 | Website for public access. | <https://persist-project.gr/_docs/PERSIST_JULY_2023_article.pdf> |
| Press Release | PERSIST\_press\_release | PERSIST 1st general Press Release, November 2022 | Website for public access. | <https://persist-project.gr/_docs/PERSIST-gr-Nov2022-PressRelease-final.pdf> |

Table 4: Other data that will be generated within PERSIST

All dissemination material is planned to be made available through the website and other dissemination channels, such as YouTube[[6]](#footnote-7), a very popular platform for hosting videos.

This project does not cope with sensitive data and therefore there is no need to cope with data protection issues.

# D1.6 Open Access Pilot

PERSIST will ensure open access to all scientific information produced during the project. In this way, PERSIST will maximize the potential of sharing the knowledge created throughout the project lifetime. Specifically, in addition to providing access to publications via the project’s website, the papers will be stored in the arXiv online repository.

**Software**: To make the software easily available to the EU Industry and Academy, PERSIST will use (when possible) well-known widely used public R&D repositories. Those repositories are used also to disseminate project results.

**Publications at conferences and journals:** Most of the conferences and journals where we expect to publish PERSIST results are published by IEEE and ACM, which are world-leading providers of publishing services for research papers. Both of them support flexible author rights policies. Most of them allow authors who publish in their journals or conferences to share their research by posting a free draft copy of their article to non-commercial repositories and/or websites. Non-commercial repositories are repositories owned by non-profit organizations that do not charge a fee for accessing deposited articles and that do not sell advertising or otherwise profit from serving articles. When publishing in a repository is not allowed, the papers will be published in gold open access mode.

All publications [BG+21-I, FKK21-I] produced in the context of PERSIST will acknowledge project funding both in the publication and in the metadata. Specifically, each paper will include in its metadata the terms *European Union (EU)* and *Horizon 2020*, the name of the action, the acronym and the grant number, the publication date, the length of embargo period if applicable, and a persistent identifier (e.g. DOI), to comply with the EC open access instructions [OAIRE].

To enable third parties to better access and disseminate the published data, we will consider attaching Creative Common Licenses[[7]](#footnote-8), e.g. CC BY[[8]](#footnote-9), or CC0[[9]](#footnote-10), to the data deposited.

# D1.7 References

|  |  |
| --- | --- |
| [Astro] | S. Soldi, V. Beckmann, W. Baumgartner, G. Ponti, C. R. Shrader, P. Lubi ́nski, H. Krimm, F. Mattana, and J. Tueller. Long-term variability of agn at hard x-rays. Astronomy & Astrophysics, 563:A57, 2014. |
| [ECD-OA] | European Commission, Directorate - General for Research & Innovation, “Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020”, <https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf> |
| [OAIRE] | OpenAIRE, <https://www.openaire.eu/> |
| [Seismic] | I. R. I. for Seismology with Artificial Intelligence. Seis-mic Data Access. <http://ds.iris.edu/data/access/>, 2018. |
| PERSIST\_paper\_SRDS\_2023 | Panagiota Fatourou, Eleftherios Kosmas, Themis Palpanas, and George Paterakis, ``FreSh: Lock-Free Index for Sequence Similarity Search,” |
| PERSIST\_paper\_PPoPP\_2023 | Yuanhao Wei, Guy E. Blelloch, Panagiota Fatourou, Eric Ruppert:Practically and Theoretically Efficient Garbage Collection for Multiversioning. PPoPP 2023: 66-78 |
| PERSIST\_paper\_PODC\_2022 | Carole Delporte-Gallet, Panagiota Fatourou, Hugues Fauconnier, and Eric Ruppert,``When is Recoverable Consensus Harder Than Consensus?,” PODC 2022: pp. 198-208. |
| PERSIST\_paper\_MEDES\_2023 | Panagiota Fatourou, ``Parallel and Distributed Data Series Processing on Modern and Emerging Hardware”, In Proceedings of the15th International Conference on management of Digital Ecosytems, May 2023 |
| PERSIST\_paper\_DISC\_2021\_I | N. Ben-David, G. E. Blelloch, P. Fatourou, E. Ruppert, Y. Sun, Y. Wei, "Space and Time Bounded Multiversion Garbage Collection", In Proceedings of the 35th International Symposium on Distributed Computing, Freiburg, Germany, October 2021. |
| PERSIST\_paper\_VLDB\_2022 | K. Echihabi, P. Fatourou, K. Zoumpatianos, T. Palpanas, and H. Benbrahim, "Hercules against data series similarity search", Proc. VLDB Endow. 15:10, pp. 2005-2018, 2022. |
| PERSIST\_poster\_2023 | <https://persist-project.gr/_docs/PERSIST_OCT_2023_poster.pdf> |
| PERSIST\_presentation\_OPODIS\_2022 | Panagiota Fatourou, ``Recoverable Computing (Invited Talk)”, OPODIS 2022: 2:1-2:2  <https://persist-project.gr/_docs/OPODIS2022-faturu.pdf> |
| PERSIST\_presentation\_SPTDC\_2023 | <https://persist-project.gr/_docs/MultiversionDataStructures-final-toprint.pdf> |
| PERSIST\_presentation\_PPoPP\_2023 | <https://persist-project.gr/_docs/OPODIS2022-faturu.pdf> |
| PERSIST\_video\_ DISC\_2021 | <https://www.youtube.com/watch?v=mU3JZdvapBc> |
| PERSIST\_video\_ PPoPP\_2022 | <https://www.youtube.com/watch?v=pfeh5iStYAM> |
| PERSIST\_presentation\_WP3\_emulate\_2023 | <https://persist-project.gr/_docs/Failure_Framework_Simulation_Design_v0.1.pdf> |
| PERSIST\_Article | <https://persist-project.gr/_docs/PERSIST_JULY_2023_article.pdf> |
| PERSIST\_Press\_Release | <https://persist-project.gr/_docs/PERSIST-gr-Nov2022-PressRelease-final.pdf> |

1. <https://zoom.us/> [↑](#footnote-ref-1)
2. <https://www.skype.com/> [↑](#footnote-ref-2)
3. <https://github.com/> [↑](#footnote-ref-3)
4. <https://www.microsoft.com/el-gr/microsoft-365/onedrive/online-cloud-storage> [↑](#footnote-ref-2187)
5. <https://arxiv.org/> [↑](#footnote-ref-6)
6. <https://www.youtube.com/> [↑](#footnote-ref-7)
7. <https://creativecommons.org/> [↑](#footnote-ref-8)
8. <http://creativecommons.org/licenses/> [↑](#footnote-ref-9)
9. [https://creativecommons.org/abou t/cc0/](https://creativecommons.org/abou%20t/cc0/) [↑](#footnote-ref-10)