EuroHPC-01-2019

IO-SEA

IO - Software for Exascale Architectures

Grant Agreement Number: 955811

D7.4

Collaboration plan with definition of common objectives and activities including milestones

Final

Version: 2.2

Author(s): Ph. Deniel (CEA)

Contributor(s): Maike Gilliot (CEA)

Date: 23.09.2021

Project and Deliverable Information Sheet

IO-SEA Project	**Project Ref. №:** 955811	
	Project Title: IO - Software for Exascale Architectures	
	Project Web Site: http://www.iosea-project.eu	
	Deliverable ID: **D7.4**	
	Deliverable Nature: Report / Other / DEM / DEC	
	Deliverable Level: PU (Public)	**Contractual Date of Delivery**: 30 / September / 2021
		Actual Date of Delivery: 30 / September / 2021
	EC Project Officer: Daniel Opalka	

* - The dissemination levels are indicated as follows: PU = Public, fully open, e.g. web; CO = Confidential, restricted under conditions set out in Model Grant Agreement; CI = Classified, information as referred to in Commission Decision 2001/844/EC.

Document Control Sheet

Document	**Title:** **Collaboration plan with definition of common objectives and activities including milestones**	
	ID: D27	
	Version: 2.	**Status:** Fina
	Available at: http://www.iosea-project.eu	
	Software Tool: Microsoft Word / LibreOffice Writer	
	File(s): IO-SEA_Deliverable_D7.4-Collaboration-Plan-FINAL.docx	
Authorship	**Written by:**	Philippe DENIEL (CEA)
	Contributors:	Maike GILLIOT (CEA)
	Reviewed by:	Artur Podobas (KTH), André Brinkmann (JGU)
	Approved by:	Exec Board/ WP7 Core Group

Document Status Sheet

Version	Date	Status	Comments
0.1	06/July/2021	Draft	Initial version
0.2	21/August/2021	Draft	First corrections
1.0	23/August/2021	Draft	Corrections made
2.0	15/September/2021	Review	First version for review
2.1	20/September/2021	Review	Second version for review
2.2	23/September/2021	Final	Final version

Document Keywords

Keywords:	IO-SEA, HPC, Exascale, Software

Table of Contents

List of Figures

List of Tables

Executive Summary

This document is IO-SEA's deliverable D7.4 / D27, with a focus on "Collaboration plan with definition of common objectives and activities including milestones".

It presents the IO-SEA project's objectives and motivations and the key technological components that build the basements for the solution developed in the scope of the project.

The document then discusses different projects inside the call EuroHPC-01-2019 and identifies possible collaborations, but does as well identify collaborations with others partners outside the scope of the EuroHPC call.

At the end, it summarizes the collaboration axis and the milestones defined.

1 Introduction

This document is deliverable D7.4 for the project IO-SEA. It aims to describe collaborations that the IO-SEA can drive with other projects, in terms of technical collaborations and data sharing.

Collaborations can happen at various levels and it is important to clarify that point before we go further. Each project has its own ambition and objectives, but they share common roots: all of them evolve in the High Performance Computing (HPC) area and all of them share concerns regarding Exascale[1] from the storage systems point of view. Projects can collaborate in various ways:

- they share data;
- they use the same products, the same technologies or they leverage outcomes from the same formerly held project;
- they aim to build something common, for example each is focusing on a facet of a complex software suite;
- they share best practices and methods, and sometimes they share infrastructures;
- they do common dissemination efforts.

Collaborations help projects to proceed together and build consistent solutions, providing an explicit benefit for each project involved in the collaboration. They can be considered as a kind of "meta project" across the ten funded projects. In order to succeed those collaborations must be organised and structured. This structure should include:

- data shared (with reference to the Data Management Plan);
- common technologies and common pieces of software;
- benefits expected from project A to collaborate with project B;
- common deliverables (if this makes sense) and common milestones.

As projects end, results are collected. Those results can be exploited later and lead to new technologies and technical outcomes. If those outcomes can be highlighted, their impact should be evaluated, as far as possible.

This deliverable will follow this approach from the point of view of the IO-SEA project.

[1] Exascale computing means a computing system that can perform at least one exaflop—or one quintillion (a billion billion)—calculations per second. For perspective, it would take 40,000years for one quintillion gallons of water to spill over the Niagara Falls. The Milky Way galaxy is 1 quintillion kilometers wide. From the storage's perspective, massive data (up to tens of Exabrytes,(quitillions of bytes)) will be managed.

2 The IO-SEA project

This section will briefly present the IO-SEA project, depicting the project's approach, its motivations and objectives.

2.1 IO-SEA in a nutshell

IO-SEA aims to provide a novel data management and storage platform for Exascale computing based on hierarchical storage management (HSM) and on-demand provisioning of storage services. The platform will efficiently make use of storage tiers spanning NVMe[2] and NVRAM[3] at the top all the way down to tape-based technologies. System requirements are driven by data intensive use-cases, in a very strict co-design approach.

The concept of ephemeral data nodes and data accessors is introduced. An ephemeral service is associated with a simulation job and provides the IO services for it (for example, it can be a S3 server or a NFS server, exposing the data required for the related job), restricting the exposed data to what the simulation job asked for. It is started automatically with the simulation job and is dedicated to it, in particular it will end at the time this job ends. Ephemeral services run on dedicated nodes, called *data nodes*, as simulations are junning on *compute nodes*. Ephemeral services allow users to flexibly operate the system, using various well-known data access paradigms, such as POSIX namespaces, S3/Swift Interfaces, MPI-IO and other data formats and protocols. These ephemeral resources eliminate the problem of treating storage resources as static and unchanging system components – which is not a tenable proposition for data intensive Exascale environments. The methods and techniques developed in IOSEA are applicable to Exascale class data intensive applications and workflows that need to be deployed in highly heterogeneous computing environments.

Critical aspects of intelligent data placement are considered for extreme volumes of data. This ensures that the right resources among the storage tiers are used and accessed by data nodes as close as possible to compute nodes – optimising performance, cost, and energy at extreme scale.

Advanced IO instrumentation and monitoring features will be developed to leveraging the latest advancements in Artificial Intelligence (AI) and machine learning to analyse systematically the telemetry records to make smart decisions on data placement in order to optimise it. These ideas coupled with in-storage-computation remove unnecessary data movements within the system.

The IO-SEA project (EuroHPC-2019-1 Topic b) has connections to the DEEP-SEA (Topic d) and RED-SEA (Topic c) project. It leverages technologies developed by the SAGE, Sage2 and NextGEN-IO projects, and strengthens the TLR of the developed products and technologies.

[2] Non Volatile memory express

[3] Non Volatime RAM

2.2 IO-SEA leverages elements from other projects

The IO-SEA project takes benefits from formerly held projects, from previous EuroHPC calls. By doing this, IO-SEA will increase the Technology Readiness Level (TLR) of the involved products and technologies.

2.2.1 About SAGE

The SAGE project has developed MERO, an object store platform capable of addressing the requirement for Exascale storage. In particular, the project had the following objectives:

- Provide a next-generation multi-tiered object-based data storage system (hardware and enabling software) supporting future-generation multi-tier persistent storage
- Significantly improve overall scientific output through advancements in systemic data access performance and drastically reduced data movements.
- Provide programming models, access methods and support tools validating their usability, including ‘Big-Data’ access and analysis methods
- Projecting suitability for extreme scaling through simulation .

The MERO object store and the software ecosystem started with SAGE, new features and development came with Sage2.

2.2.2 About Sage2

The IO-SEA project extends the MERO object storage platform to suit the needs of the IO-SEA data management solution. The Object storage API, Clovis, developed in the SAGE/Sage2 project, will be leveraged and extended to suit the needs of IO-SEA software components and applications. To that end, IO-SEA will collaborate with the Sage2 project to utilize the SAGE platform at FZJ that has the different tiers of storage along with MERO/Clovis installed. Any further optimisations on the MERO infrastructure will be fed to each other projects. Concepts of Global Memory Abstraction earlier developed in Sage2 will be further exploited in IO-SEA and will form an area of close collaboration between the projects. IO-SEA hence makes the system realisation of the solution more viable through the use of the SAGE system at FZJ. Sage2 also forms a very strong connection for joint dissemination and exploitation of both projects results.

MERO, originally a commercial product, has been later rebranded as CORTX-MOTR, and it has been released as open source software. IO-SEA will leverage SAGE and Sage2 by intensively using MOTR as a software providing both KVS (key-value store) and object store features.

2.2.3 About MAESTRO

Maestro built a data and memory-aware middleware framework that addresses the ubiquitous problems of data movement in complex memory hierarchies that exist at multiple levels of the HPC software stack.
MAESTRO, the ‘Middleware for memory and data-awareness in workflows’ project, will address the two following major impediments of the modern HPC software: data-awareness (provides data handling functionality to be used in the HPC software stack and the programming models) and memory-awareness (providing more elements about memory

locations, even in multi-tiered architecture models to ease the management of complex data movements).
Maestro proposes a middleware framework that enables memory and data-awareness. Maestro has developed new data-aware abstractions that can be used in any level of software, e.g. compiler, runtime, or application.
Those outcomes will be leveraged in the IO-SEA project for building efficient storage, efficient user interfaces. In particular, benefits are expected on data placement and data movements optimisation.

2.2.4 About LEXIS2020

The LEXIS project will build an advanced engineering platform at the confluence of HPC, Cloud and Big Data which will leverage large-scale geographically distributed resources from existing HPC infrastructure, employ Big Data analytics solutions and augment them with Cloud services (Horizon 2020 project by services developed within the project.) The expected outcomes will be further used to implement efficient data movement between different storage tiers in the IO-SEA.

2.3 Key technologies in IO-SEA

IO-SEA is built on top of several key technologies. Three of them make the basements of the software architecture of IO-SEA:

- Object Stores: the data in IO-SEA is recorded as object inside object stores. Two object stores are used :
  - CORTX-MOTR is an object store developed by Seagate. MOTR, previously known as MERO, was a keystone in the projects SAGE and Sage2. It then was a commercial product, but Seagate made it open source software in late 2019 – early 2020. MOTR can handle any kind of disk devices, from slow hard disk drives (HDD) to fast Solid State Devices (SSD) and even NVRAMs.
  - Phobos is an object store developed by CEA and distributed as open source software. Phobos is capable of dealing with both disks and tapes. Inside IO-SEA, it is responsible of long-term storage, involving tapes located inside tape robots.
- Hierarchical Storage Management (HSM): IO-SEA combining benefits from different hardware. Tapes provide high capacities but are rather slow compared to NVRAM (or NVMe devices) that are quite fast, have a low latency but a reduced capacity (this kind of device has to be attached to NVMe or memory buses and are really expensive). To take advantages from every kind of hardware involved, IO-SEA will used tiers with multiple levels, stacking NVRAM on top of SSD, HDD and finally tapes. This multi-tiered architecture requires data movement between the different levels, both promotion and demotion: this is called Hierarchical Storage Management (or HSM). The HSM feature between different tiers in MERO was developed inside the SAGE project. It will be extended to make MOTR capable of working with Phobos, migrating data in-between the two object stores.
- Ephemeral Services / Service Provisioning: users require different interfaces to operate on data. Some will access it via dedicated middleware Application Programming Interfaces (APIs), other will rely on standards such as S3 to manage objects. Finally, some applications will need a „legacy POSIX“ interface that will expose objects as files within a POSIX tree. As the application starts, it will advertise

the kind of interface that it wants, the IO-SEA software will allocate nodes, known as *data nodes*, to start ephemeral services associated with the run of the application. This temporary service will last as long as the user application runs, and provides services only to the compute nodes involved in the user application (typically, the compute nodes on whicch a simulation code is running). Once the application ends, the associated ephemeral service will flush data to persistent storage before it shutdowns and releases the corresponding datanodes.

Beyond those three key technologies, other technologies are important to IO-SEA:

- Data placement: HSM and the Ephemeral Services allocation require to place the data on the right tier and location. Some pieces of date will need very fast hardware as they will be accessed quite intensively while other may be written once and almost never read and could be quickly placed on slow and inexpensive devices. This leads to the data placement issue: IO-SEA has to build the required mechanisms to identify quickly the taxonomy of the data (potentially involving hints provided by the user himself), making it possible to place it correctly.
- Monitoring and metrology: various kinds of resources are involved in IO-SEA, storage media are such resources but data nodes hosts running the ephemeral services are to be considered as resources. In order to allocate them in the most optimal way, it is critical to have a clear view of the way those resources are used and consumed. Advanced monitoring tools, from system probes to high-level logs and statistics are deployed. The collected information is stored on dedicated systems to make it possible to analyse them and identify patterns and trends in the application and user behaviours.
- Data science and AI technology: tools and algorithms from the ecosystem of he „AI“ (such as Machine Learning methods and Deep Learning methods) will operate on the vast collection of data resulting from the monitoring and metrology tools. Such tools will provide solutions for performing intelligent resource allocation and bring solution to the data placement issue.

3 Collaboration with projects inside the call EuroHPC-01-2019

This section describes the collaborations that will be held with some of the other nine projects inside call H2020-JTI-EuroHPC-2019-1.

3.1 SEA projects

IO-SEA is born inside the scope of the Mousquetaire initiative that led to the submission of this project and the submission of two “brother projects”, DEEP-SEA and RED-SEA. Synergy between the SEA projects is natural since they all belong to the same “meta project” that aims to build tomorrow's software architecture and ecosystem for HPC. Collaboration between those projects is quite natural, and is actually in the DNA of those three projects.

3.1.1 Mousquetaire initiative

The Mousquetaire Initiative is a collaboration that federates four partners from France and Germany: two research institutes (CEA[4] and FZJ[5]) and two industrials (ATOS and Partec).

It started in 2017 as FZJ bought an ATOS supercomputer and chose this architecture as the basement for its future Exascale program. CEA and ATOS already have close connections via active and common Research and Development collaborations. An informal collaboration was established and a Memorandum of Understanding (MoU) was signed in January 2018. The name "Mousquetaire" comes from the MoU acronym and the number of partners (four, just like Alexandre Dumas's famous musketeers).

This action was enhanced by the signature of a more complete and precise agreement in May 2018, built around five axes. Discussions and common workshops started at this point, leading to a shared vision of a future Exascale architecture. In July 2019, a first white paper, depicting this view, was produced and the decision was taken to prepare the proposal for three project, connected together, to answer the challenges of EuroHPC-01-2019 (focused on "Extreme scale computing and data driven technologies").

3.1.2 The SEA projects

The Mousquetaire initiative gave birth to three proposals, DEEP-SEA, IO-SEA and RED-SEA, submitted on January the 14th, 2020. All of the project were accepted by the EU to receive fundings on July the 14th 2020.

The three projects have common basis, including the Modular Supercomputing Architecture (MSA) promoted by FZJ and common goals to build a complete software architecture for a future Exascale system.

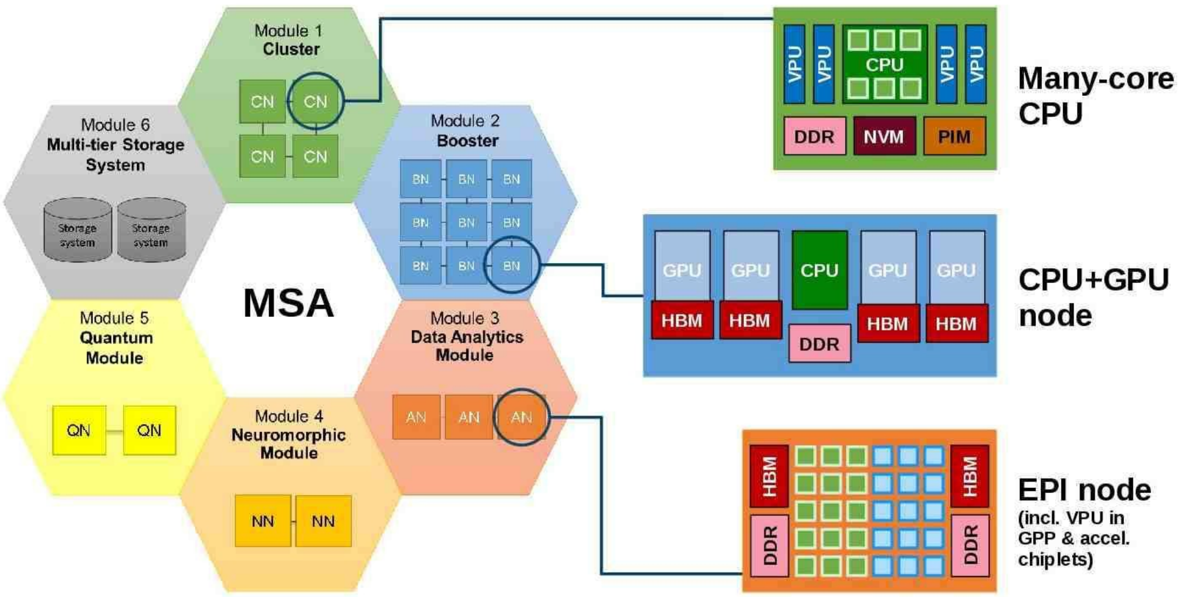

Figure 1: The MSA Architecture

[4] Comissariat à l'Energie Atomique et aux énergies alternatives

[5] Forschnungzentrum Juelich

If we consider the SEA projects more closely, several elements quickly appear. They reveal the links and common points of view shared among the SEA projects, and their native connections:

- the DEEP-SEA software stack is providing a software environment to be exploited by end user on the MSA. IO-SEA software stack will be used by DEEP-SEA for performing IO operations;
- the MSA exposes compute nodes and data nodes that work together, the software managing them, DEEP-SEA and IO-SEA, are naturally associated;
- managing data means moving data and placing it at the right place (via tools provided by IO-SEA), this requires fast networks and optimised APIs and tools, such as what RED-SEA will produce;
- the RED-SEA will expose smart metrics and hints (routing, "distance" between nodes, …) that are key components for optimising the process of the IO-SEA and the DEEP-SEA;
- DEEP-SEA relies on efficient communication libraries (MPI and friends…), RED-SEA will provide them to DEEP-SEA;
- Scheduling programs and services (simulation code, user agent, services spawn to serve as IO servers) is a common problem to all three projects: IO-SEA has ephemeral services to manage, DEEP-SEA submits compute jobs to the supercomputer, RED-SEA exposes elements of the network topology which are essential for optimised scheduling. All three projects must work together to address scheduling the same way;
- Hints: IO-SEA needs hints from the end-user to optimise data process (e.g. for data placement). Such hints come from the apps (they will be using the software environment built by DEEP-SEA), from the system software itself and from the hardware (RED-SEA will expose the "network hints", such as network distance (in hops) between nodes, throughput or latency);
- Metrology: all three projects involve metrology and statistics, common practises should be shared to have common tools across the SEAs. If the collected datasets are big enough and provide enough details, it will be pertinent to use data-science methods (so called „AI" framework with Machine Learning and Deep Learning algorithms);
- Optimisation: all SEA projects will run on the same platform, and they will together to build an Exascale software stack. Therefore they must work together to be as efficient as possible.

3.1.3 Collaboration with SEA projects

Dissemination is the first and the most natural way of collaborating. As an example every project has its own website (http://iosea-project.eu, http://redsea-project.eu and http://deep-projects.eu) used as a gateway for promoting assets, outcomes and achievements of each. A global website exists too (http://sea-projects.eu) as a showcase where common results from all three SEA projects will be exposed.

Common press releases, targeting information websites like HPCWire (http://hpcwire.com) or InsideHPC (http://insidehpc.com) will be written during the lifetime of the projects. At the same time efforts will be done to produce common publications. A way of doing this would be to expose the tools and methods from project A on data gathered by project B, or use tools and methods from project A on a specific use case from project B that reveals an issue and/or a technical point of interest.

Large HPC event, such as SuperComputing or ISC will be opportunities for the SEA projects to have a common booth or for proposing Bird-of-Feather (BoF) sessions to the conference organisers.

Beyond the dissemination efforts, representatives of the other SEA projects will be included in the advisory boards (next to external representatives) of IO-SEA, DEEP-SEA and RED-SEA, so that the different projects's views remain consistent. Those efforts made around technical topics will be associated with workshops and seminars where common points of interests and related solutions will be exposed and discussed.

3.1.4 Milestones

The three SEA projects will have regular joint meetings. As one project organises seminars and / or workshops about its involved technologies, it will invit representatives from other projects when this is pertinent. This will help in disseminating ideas, best practices and the methods between the projects.

Every year, a dedicated workshop will be organised. This one-two days workshop will gather the three SEA projects and will be dedicated to share the common realisations and identify the outcomes from the active collaboration held during the past year. The identified points of interest / collaboration depicted in the current document will be updated, new topics may be added if they are relevant. This meeting will concretise the paths to be followed by the SEA projects during the following year. This yearly event will result in the production of a document, acting as an update / complement / annex of the present document, operating as the collaboration plan for the next years.

Three milestones are defined:

Milestone Name	Description	Date
MS-SEA-Y1	First yearly meeting, defining collaboration directions for Year 2	M12
MS-SEA-Y2	Second yearly meeting, defining collaboration directions for Year 3	M24
MS-SEA-Y3	Final meeting, depicting benefits and outcomes from the related collaboration.	M36

Common Milestones shared by the SEA projects

3.2 ADMIRE

3.2.1 ADMIRE overview

ADMIRE is another project inside the EuroHPC-01-2019 call. ADMIRE's objective is the creation of an active IO stack that dynamically adjusts computation and storage requirements through:

- intelligent global coordination,
- elasticity of computation and IO,
- scheduling of storage resources along all levels of the storage hierarchy, while offering quality-of-service (QoS), energy efficiency, and resilience for accessing extremely large data sets in very heterogeneous computing and storage environments.

The project overview includes:

- creating a European adaptive storage system to boost data-intensive applications;
- a Co-design approach between (malleable) applications and storage using an intelligent global coordination;

- the expected result is software-defined framework including scalable monitoring and control, separated control and data paths, and the orchestration of key system components and applications;
- use cases are weather and air quality, bioinformatics, turbulent flows, artificial intelligence, …

ADMIRE's key technological element is GekkoFS, an efficient parallel and distributed file system, and the concept of "IO Malleability" of applications which enable dynamically IO services (such as burst buffers) to fit the variations of the IO requirements of a running job.

ADMIRE gathers partners from different organisations:

- University Carlos III of Madrid (Coordinator) and Barcelona Supercomputing Center (Spain)
- Johannes Gutenberg University Mainz, TU Darmstadt,
- Forschungszentrum Jülich and Max Planck Computing and Data Facility, Germany
- DDN, Paratools and INRIA (France)
- CINI, CINECA and E4 (Italy)
- Poznan Supercomputing and Networking Center (Poland)
- KTH Royal Institute of Technology in Stockholm (Sweden)

3.2.2 Technical Collaboration

The reader may notice that IO-SEA and ADMIRE share common partners: FZJ, KTH and JGU. Since the FZJ's and KTH's departments involved in those two projects are different and share few connections, strong connections exist between the laboratories involved in both projects at Johannes Gutemberg University. This establishes a natural "bridge" between ADMIRE and IO-SEA.

The coordinators of the two projects met remotely and it quickly appeared that those common assets are points of interest for building a collaboration. The list below exposes the identified points and the related view for the two different projects:

1. Management of data locality / data placement:

    - IO-SEA: the project involves object stores on a multi-tiered storage architecture. Placing the data at the right place is a key point. Objects inside the object stores could be promoted or demoted. A dedicated policy engine, applying placement policies will handle this aspect

    - ADMIRE: The project provides the burst-buffer file system GekkoFS, which places application data on node-local storage systems and therefore ensures data access locality and performance isolation between applications. The number of nodes providing a file system can be adapted based on applications' requirements, e.g., if more performance is required over the runtime of a workflow

2. Storage Aware Scheduling:

    - IO-SEA: the project introduces on-demand services called "ephemeral services". When computation jobs run on compute nodes, ephemeral services are spawned on dedicated data nodes. The scheduler will choose the data nodes to reduce the network distance between compute and data nodes. Scheduling policies (such as karma scheduling) could be used for managing the scheduling of jobs with an attached ephemeral service.

    - ADMIRE: Data is moved between the (backend) parallel file system Lustre and the burst-buffer file system. Stage-in and stage-out are managed by Slurm and through a bandwidth management component, which schedules data flows and instruments the Lustre token bucket filter to fulfill QoS requirements.

3. Instrumentation
   - IO-SEA: it is critical to have a precise view of the system at every level (hardware, system software, application software), this information is collected by a dedicated instrumentation framework. This information will be exploited for establishing optimal data placement policies and scheduling policies.
   - ADMIRE will develop scalable monitoring tools for Exascale systems, which implement gather and reduction mechanisms that will be used for both periodic aggregations and on-demand performance queries. The solution will integrate support for file system monitoring based on the Lustre Intelligent Management Engine (LIME) and system monitoring using Paratools TAU. Offline application profiles will be combined with user hints and dynamic IO behaviour to predict how the IO requirements an application change if the number of nodes it runs on is adjusted.
4. Data Set Management:
   - IO-SEA: data sets exist inside the object store. They will be exposed to the simulation jobs by the ephemeral services, as requested by the jobs themselves. Correct management of data sets implies correct ephemeral service scheduling and optimal data placement.
   - ADMIRE: data sets inside ADMIRE are defined by the set of files which are moved by Slurm during the stage-in process from Lustre to the burst-buffer file system. Furthermore, GekkoFS will be coupled with the Lustre Client Side Caching, so that data can be implicitly moved to GekkoFS.

3.2.2.1 Collaboration with ADMIRE

The collaboration plan with ADMIRE will be similar to the plan established with the other SEA projects, and the same logic will be followed.

Common dissemination will not be as easy and natural, compared to SEA, but common publications and common BoF are interesting paths.

Like the SEA projects, common technical points of interest are identified and will be studied by each project. ADMIRE and IO-SEA should meet at least once a year, during a common workshop where they will expose their achievements, unexpected obstacles and methods. Update about the collaboration plan will be written for defining the road for the next year (the process is the same as the one describe for the SEA projects)

Milestone Name	Description	Date
ADMIRE-IOSEA-Y1	First yearly meeting, defining collaboration directions for Year 2	M12
ADMIRE-IOSEA-Y2	Second yearly meeting, defining collaboration directions for Year 3	M24
ADMIRE-IOSEA-Y3	Final meeting, depicting benefits and outcomes from the related collaboration.	M36

Milestones shared between IO-SEA and ADMIRE

4 Collaboration with the larger HPC ecosystem

As IO-SEA consortium was created, existing connections have been very helpful. HPC is a rather small community, and the community related to mass storage for the HPC is even smaller. People involved in IO-SEA know each other by participating in the same formerly held projects and workgroups. This has another direct consequence: IO-SEA will naturally contribute and collaborate with those entities, just because the same people are working with it.

4.1 ETP4HPC

The European Technology Platform for High Performance Computing (ETP4HPC) is a think-tank led by industrials from the HPC. It accepts and gathers as well people from research centres and academics.

ETP4HPC aims to build a competitive and global European HPC technology value chain. By federating efforts from its members, ETP4HPC will increase the convergence of the involved technologies and so increase Europe's competitiveness in this area. ETP4HPCS is strongly acquainted with the research and innovation efforts of its participants, it designs and updates a Strategic Research Agenda (SRA) in order to provide decision makers with relevant advice and expertise for the log-term development of HPC in Europe, and it facilitates the coordination across actors of the HPC world. By design, ETP4HPC aims to represent the voice of European HPC industry in the worldwide HPC arena.

Most of the IO-SEA partners are members of ETP4HPC and several of the people involved in the project (WP leaders or task leaders) took themselves part of the meeting organised around the different instances of the SRA. Some of the ideas on top on which IO-SEA is built did incubate inside an scientific environment fostered by ETP4HPC. In return, IO-SEA will create new solutions, increase the Technical Level of readiness (TLR) of many projects and allow new concepts and ideas to rise.

5 Conclusion

This document defines paths to be followed to have actual collaboration with other projects, it sets up methods and identifies common points of interest,.

This deliverable therefore describes the start point on the collaboration and the first directions for the beginning of the journey, and this journey has a three year duration. This document is the first of a set of documents, it has to be regularly updated, at least on a yearly time scale, in order to stay aligned with the real progress of the different projects. Those updates will be described in dedicated sections in each yearly deliverable depicting the status of the project.