relevance as well as for schedule and cost. Due at month 24; CO; R; Lead: NLeSC

D3.2 Report on services in portability and refactoring
Summarise results from porting applications. Discuss how to move forward and advance the codes towards exascale. Due at month 46; PU; R; Lead: BULL

D3.3 Report on services offered on IO, coupling and workflow
A document summarizing the services offered on IO, coupling and workflow to the different modelling groups, analysing the benefits and the long-term sustainability of the improvements achieved during the service period. Due at month 44; PU; R; Lead: CERFACS

D3.4 Report on services offered on weather and climate benchmarks
Provide documentation of new benchmarks and publish benchmark codes (where applicable). Summarise typical performance measures and accuracy levels for different benchmarks on different hardware. Due at month 46; PU; R; Lead: DKRZ

D3.5 To make Europe's Earth system models fit for the exascale
Summarise the work done, identify achievements and describe possible future developments. Discuss the main bottlenecks for Europe's models towards exascale simulations and how to address them. Due at month 48; PU; R; Lead: NLeSC

Work package number	4		**Lead beneficiary**				UREAD			
Work package title	Data systems for scale									
Participant number	1	2	3	8	9	10	11	12	14	19
Short name of participant	DKRZ	CNRS-IPSL	ECMWF	ICHEC	METO	CMCC	UREAD	STFC	SEAGATE	DDN
Person months per participant	15	12	1	6	12	9	48	24	16	20
Start month	1			**End month**				48		

Objectives

WP4 will contribute to both **principal objectives (1) and (2)** of the project and will in particularly meet **specific objective (d)** – to mitigate the effects of the data deluge from high-resolution simulations, this work package specifically addresses ensemble tools and storage middleware. Specifically, WP4 will

1. Support data reduction in ensembles and avoid un-necessary subsequent data manipulations by providing tools to carry out ensemble statistics "in-flight" and compress ensemble members on the way to storage.
2. Provide tools to: a) transparently hide complexity of multiple-storage tiers from applications at run time by developing middleware that lies between the familiar NetCDF interface and storage, and prototype commercially credible storage appliances which can appear at the backend of such middleware, and; b) support manual migration of semantically important content between primary storage on disk, tape, and object stores, including appropriate user-space caching tools (thus allowing some portable data management within weather and climate workflows).

Description of work [Lead: UREAD (B. Lawrence). Co-lead: UREAD (J. Kunkel)]

The primary approach is to extend and exploit existing tools and interfaces which are in common use in the community and/or with which we have previous experience. In doing so, the key philosophy is "Maximum Impact from a Minimum Change Surface" insofar as ESiWACE2 data handling solutions need to

- Maximise their impact on data handling, by minimising the impact of increasing volumes of data from multiple sources, particularly within and between large-scale ensembles,

- have minimal interference with existing working practice and codes, and
- have minimal requirements of the system environment.

The best way to achieve this is by modifying existing tools, developing a minimum of new tools and, where possible, exploiting middleware which can be deployed easily, hiding complexity from end-users. The heart of the middleware approach is to insert software layers in the stack between the model IO code and the files it produces which have traditionally been placed on (parallel) file systems. Two key layers will be involved:

1. Between model ensemble members and the NetCDF library interface
2. Between the HDF library, which itself sits inside the NetCDF stack, and the underlying storage.

The proposed work will build on a number of developments in ESiWACE1, in particular:

1. Semantic Storage Layers (SemSL). "Intelligent knowledge" about the data held in NetCDF and other formats depends on semantics which exploit the format syntax (e.g. variable attributes) but link them in complex ways. The SemSL exploits the NetCDF CF aggregation conventions to fragment NetCDF files into sub-files which can then be managed (using SemSL or other traditional NetCDF tools) across different storage providers whether local or remote.
2. Earth System Data Middleware (ESDM). Existing scientific software formats provide libraries that provide syntactical support for four-dimensional datasets on disk. ESDM targets the widely used HDF library (upon which NetCDF4 depends) to intercept read/writes via HDF to provide more sophisticated, and higher-performance, use of storage.

These two approaches – "above the data format" and "below the data format" – share a common conceptual design philosophy (since they were developed together), but have been developed to address different parts of the problem space. ESDM has been developed to provide performance in simulation and is expected to evolve to provide an "active storage" system which is "weather and climate-aware". SemSL has been developed to assist in data analysis and data management, and is expected to provide "manual control of storage tiering". One of the goals of ESiWACE2 will be to further integrate these parts and further differentiate around the distinctive roles ("smart/performant/active", "simple/performant/manual").

Together, they represent the recognition that the efficient execution of workflows and individual applications within a workflow must be able to harness the different characteristics of a site's storage systems, and that the balance of services between active and manual can and will differ. Both need to address efficiency, performance portability, manageability, data reduction, and robustness (fault tolerance) of workflows.

There is one management task to coordinate and oversee six separate interlocking tasks within this WP. Internal WP milestones will be used to help cross-task control and to synchronise with project-level control points and milestones. All software products will be formally described in D4.3, building on the interim deliverable D4.1 (covering software delivered by tasks 4.2 and 4.3) at month 24. The scope for commercial products will be evaluated by a description of appliance prototypes in D4.2.

Task 4.1: Design and leadership [Lead: UREAD]

Manage the inter-locking tasks, develop an overall architecture, and provide a final report which covers what has been done and provides a roadmap (D4.3) which takes input from other IO and storage projects targeting exascale computing.

Task 4.2: Ensemble services [Lead: UREAD. Partners: CNRS-IPSL, ECMWF]

Develop and deploy support for controlling an ensemble of model instances within one executable feeding data output via an "active" IO server to storage. The IO server will be configurable to support "ensemble diagnostic kernels" which will include simple diagnostics (e.g. maxima, minima, averages) as well as "ensemble compression" – the ability to write out a compressed ensemble (building on work from ESiWACE1). This work will extend an initial prototype developed with UK national funding which exploits the XIOS IO server developed by ESiWACE1 and French national funding, making it more robust and suitable for wider deployment in other models. A key task will be developing support in the ensemble controller to handle failures in any one-ensemble member so that the ensemble system will not crash if one

member fails. The initial software, available at the end of year two (described in D4.1), will be used in WP1 and continually improved thereafter.

Task 4.3: Earth System Data Middleware [Lead: UREAD. Partners: DKRZ, Seagate, CMCC, STFC, DDN, Seagate]

Middleware to interface standard IO libraries (NetCDF/HDF) and storage – whether it be a file system or a combination of burst-buffer, file system, or object store – to support active migration of data within workflows. The effort made in ESiWACE will be more tightly integrated with existing software, hardened and productised, and finally enhanced with several new capabilities that are needed to deal with the growing needs of the scientists: 1) the existing middleware is hardened and performance of the POSIX back-end is further optimised; 2) the current data systems performance model is enhanced to predict back-end behaviour more accurately; 3) we explore a direct NetCDF integration instead of providing a HDF5 VOL driver – we aim to provide a HPC NetCDF back-end driver; 4) we optimise and integrate better compression capabilities as prototyped in the AIMES project into the software stack and make it available in NetCDF; 5) we provide and optimise the storage back-ends which exploit the vendor-specific protocols for Clovis, WOS, IME; 6) we develop and deploy a generic S3 interface to support portability; 7) we provide an optimised ESDM interface for analytics tools using Ophidia as the prototype. A formal software deliverable (D4.1) at the end of year two will be used as input to task 4.7 and WP1 and published for wider distribution – with software support and testing (task 4.6) continuing throughout the remainder of the project.

Task 4.4: Semantic Storage Layer tools (SemSL) [Lead: STFC. Partners: UREAD]

User-space tools suitable for deployment without system administrator interaction so that datasets can be spread across multiple storage tiers (tapes, POSIX disk, object stores) in multiple files that can be accessed through one semantic master file. Includes support for caching and metadata interaction via standard (NetCDF) queries without the need of having underlying data online. The effort made in ESiWACE will be more tightly integrated with existing software, in particular schedulers and workflow managers. Specifically: 1) the performance of the S3NetCDF software will be improved, and it will be more closely integrated into other libraries (including the ESDM itself); 2) the cache subsystem will be improved to support a wider range of tape back-ends with a cleaner management interface. An internal software deliverable at month 30 will make this software available for deployment on tier0 machines, with the resulting experience feeding into D4.3.

Task 4.5: Workflow enhancements [Lead: DKRZ. Partners: UREAD, DDN, METO]

Add explicit support within the SLURM scheduler and the Cylc workflow management software to support the efficient scheduling of the data-intensive workloads with active staging of data products through storage tiers with or without the use of the ESDM and/or Semantic Storage Tools. Specifically: 1) a co-design phase to capture scientific requirements and the data life cycle and to map script-based workflows to directed graphs enriched with IO dependencies, performance needs and information about data lifecycle and user intervention (enabling users to describe the execution of their experiments and the data interaction in detail); 2) enrich the workflow scheduler Cylc that is used to schedule NWP/climate compute workflows with the capabilities to deal with data dependencies and lifecycle information, and to interface with ESDM to query information about data locality and to announce intended data usage to ESDM; 3) to enrich the workflow scheduler SLURM to honour the data dependencies and query information from ESDM about data locality; 4) ESDM is extended with a workflow interface and service that implements the needs of the data lifecycle, e.g., allowing to migrate/copy data between storage back-ends, and to clean out-dated or redundant data. The Cylc enhancements available in month 36 will be made available for use in WP1 and the resulting experience will feed into D4.3.

Task 4.6: Testing [Lead: ICHEC. Partners: All WP partners]

Carry out continuous integration, component level, and end-to-end testing. This will require: 1) setting up

and managing a continuous integration testing environment; 2) component level quality control assessments (for each of XIOS ensemble branch, ESDM, Semantic Storage Tools, SLURM and Cylc branches, and the analytics kernel libraries); 3) scheduler support for the testing environment, and 4) regular end-to-end testing. While all development will involve testing, formal cross-task testing will begin from month 24, and experience with testing will be reported as part of D4.3.

Task 4.7 Industry proof of concept [Lead: SEAGATE. Partners: DDN]

The ESDM software is designed to work with a range of storage environments, but it is anticipated that for wide uptake, many sites may want to purchase an appliance which has vendor support for the storage sub-system. To that end, both Seagate and DDN will create a prototype appliance package which exposes the complete ESDM implementation and includes client libraries that can be deployed by users on the local compute systems. This work will build on the ESDM release at month 24 (D4.1) and result in D4.2.

Deliverables

D4.1 Advanced software stack for Earth system data
Documentation and design description, along with formal code release of the ensemble code and hardened ESDM software and various back-end implementations. Due at month 24; PU; R; Lead: UREAD

D4.2 Report on appliances available for testing
Report on appliance configurations and performance. Due at month 42; PU; R; Lead: SEAGATE

D4.3 Software documentation and roadmap

Formal documentation of software produced, description of any on-going issues discovered during this work, with a forward-looking data handling roadmap. Due at month 48; PU; R; Lead: UREAD

Work package number	5	**Lead beneficiary: CMCC**				
Work package title	Data Post-Processing, Analytics and Visualisation					
Participant number	1	2	5	10	11	16
Short name of participant	DKRZ	CNRS-IPSL	MPIM	CMCC	UREAD	UNIMAN
Person months per participant	15	7	24	15	4	3
Start month	1		**End month**	48		

Objectives

WP5 is directly linked to **principal objectives (1) and (2)** and in particular to **specific objective (d)**.
The main objective of this work package is to provide a consistent view regarding the support for data post-processing, analytics and visualisation at scale in the weather and climate domain by building, on top of the ESDM module developed by WP4, the proper ESDM extensions.
In particular, the tasks in WP5 are to a) design the ESDM interface extensions to support in-flight analytics kernels for post-processing, analysis and visualisation (PAV) needs; b) identify, prioritise, implement, and validate a set of common analytical kernels starting from a set of community-based tools; c) develop a high-performance support to enable ESDM data parallelisation for in-flight analytics; d) validate the ESDM post-processing, analytics and visualisation support on a set of representative case studies regarding community-based weather and climate applications.
All software products will be formally described in D5.1 (ESDM PAV runtime), D5.2 (analytical kernels) and D5.3 (ESDM-enabled PAV applications) building on the architectural report described in M5.1. The final software release of the ESDM-enabled PAV case study application will be documented in M5.2.

Description of work [Lead: CMCC. Co-lead: DKRZ]

Task 5.1 Design of the ESDM interface for data post-processing, analytics and visualisation (PAV) [Lead: CMCC. Partners: DKRZ, UREAD]

This task addresses the design of the ESDM interface API to support data-intensive ***p**ost-processing*, ***a**nalytics*, and ***v**isualisation* (PAV) applications and case studies in the weather and climate domain. Starting from a set of key scientific applications, this task will identify, gather, analyse, and prioritise the key requirements to properly address the design of the ESDM PAV API (M5.1).

Task 5.2 Implementation of the ESDM PAV runtime extensions [Lead: CMCC. Partners: MPIM]

High volume workflows require IO and compute parallelisation in post-processing, analytics and visualisation. There are multiple routes to parallelisation, from on-node support for accelerators, GPUs and threading, single-executable MPI jobs exploiting multiple nodes, to task based parallelisation using simple batch jobs or complex graph-based task scheduling – all could benefit from improved parallel data access exposed via the ESDM and/or the Semantic Layer Tools to perform on-the-fly analytics while data are being transferred from the storage back-ends to the applications built on top of the ESDM module. This task is to support PAV applications, by exploring and prototyping efficient *in-flight analytics* ESDM extensions. Additionally, the support for remote execution via DASK or SLURM for data-driven workflows to reach high degrees of parallelism will be explored. The initial software, available at M18 will be used to test the first analytical kernels in task 2 and will be ready for integration in task 4 and task5 at M24, and continually improved thereafter. The ESDM PAV runtime and related parallelisation strategies will be comprehensively documented in D5.1 and will be ready to be deployed on tier0 with a finale release at M36.

Task 5.3 ESDM PAV analytical kernels [Lead: MPIM. Partners: DKRZ]

In this task, we will develop the core libraries (plug-ins for the EDSM PAV runtime) needed to properly interface and support, from an analytical point of view, the ESDM module to existing community post-processing tools, analytics frameworks, and visualisation applications.
A comprehensive set of key common candidates for analytical kernels, which can exploit semantic metadata, will be established via an iterative evaluate/prioritise/design phase.
More specifically, the supported analytical kernels include, among others, statistical and arithmetic operators, data transformation operators, as well as correlation, regression and interpolation operators, that will benefit from the parallelisation support available from the ESDM runtime. They will be either exposed by the ESDM interface, or made available directly via suitable APIs to applications. To further address performance, the task also targets analytics on massively parallel architectures, by including GPU-based implementation of key analytical kernels. The first software release of the analytical kernels will feed into task 4 and 5 to start with the PAV applications porting to the ESDM at M24; intermediate releases at M30 and M36 will be deployed on tier0 and tested on the prototypes of the ESDM-enabled PAV tools/applications; the final release will be comprehensively documented in D5.2 and delivered at M42.

Task 5.4 Interfacing data post-processing and analytics applications to ESDM [Lead: MPIM. Partners: CMCC, UNIMAN, UREAD]

In this task, we will interface a set of community processing tools and analytics frameworks to the ESDM. As such, a set of key applications (e.g. CDO) will be considered as case studies to test and validate the entire ESDM software stack.
More specifically, work in this task comprises: upgrading the CDO operators via the ESDM PAV interface, also including GPU support; interfacing existing community tools (e.g. CF-Python, Ophidia) to the ESDM PAV; additionally, validation feedback will be provided to, respectively, the output of (i) task 5.3 in terms of functional (e.g. supported analytical kernels) and non-functional requirements (e.g. usability, performance) and (ii) task 5.2 from the ESDM PAV runtime point of view.
An initial release of the ESDM-enabled tools/applications for post-processing and analytics is planned at M30 to be deployed on tier0, whereas the prototypes at M36 will be tested for first integration into tier0

simulations. During the last year, the activity on the tools/application will continue thus leading to the final release at M48, which will be reported in detail in D5.3. The final software implementation of key selected post-processing and analytics tools/applications will be documented in M5.2.

Task 5.5 ESDM-enabled data visualisation case studies [Lead: DKRZ. Partners: CNRS-IPSL]

Depending on the actual size of the data produced by high-resolution weather and climate models, the conventional post-processing and post-visualisation pipeline finally encounters a barrier that can only be overcome with alternative approaches to access, store, analyse and visualise it.
One of which is in-situ visualisation, which analyses and visualises the data alongside the simulation process while the simulation is still running.
In this regard, a choice of different setups exists, that allow a loose or tight coupling between the simulation model and the visualisation software to either perform an on-the-fly data visualisation that shows how the simulation progresses, or to store images, animations and even geometry using a predefined script onto the disk. Other data reduction and transformation techniques, such as compression and decomposition are also possible and may also be explored. Especially interesting is thereby the data decomposition using wavelets with an associated (lossy/lossless) compression. This approach would allow transforming the data into several Level-Of-Detail versions, which can be accessed progressively.
The main goal of this task is to analyse how key approaches, like those mentioned above, can be implemented on top of the ESDM interface for some key visualisation applications (e.g. ICON and DYNAMICO) (M5.2, D5.3).
More specifically the task focuses on: analysing suitable rendering solutions in the light of a pre-exascale HPC system (i.e. GPU vs. CPU); implementing an in-situ visualisation framework using ParaView/Catalyst and exploring the different setups (i.e. in-transit vs. in-situ); exploring the possibilities for an in-situ created image based rendering approach that can be used as preview to the data (i.e. the ParaView Cinema extension); implementing an in-situ wavelet decomposition and compression for a later progressive data access and visualisation using Vapor; exploring lossy data compression and its impact to the data.
As for task 4, an initial release of the ESDM-enabled visualisation applications is planned at M30 to be deployed on tier0; prototypes at M36 will be tested with tier0 simulations. Development will continue during the fourth year. The final release delivered at M48 will be described in D5.3. The final software implementation of selected visualisation applications will be documented in M5.2.

Deliverables

D5.1: Report on the ESDM runtime extensions for parallel in-flight analytics
Formal documentation of the ESDM parallel runtime strategies and software implemented (final release). Due at month 36; PU; R; Lead: CMCC

D5.2: Report on the implementation of the ESDM PAV analytical kernels for post-processing, analysis and visualisation
Formal documentation of the implemented ESDM PAV analytical kernels (final release). Due at month 42; PU; R; MPIM

D5.3: Report on the final implementation of key selected post-processing, analytics and visualisation applications and main outcomes
Report on the main technical porting-to-ESDM aspects, outcomes and guidelines related to key community applications. Due at month 48; PU; R; Lead: MPIM

Work package number	6			**Lead beneficiary**						CNRS-IPSL		
Work package title	Community Engagement and Training											
Participant number	1	2	3	4	7	9	10	11	12	14	15	18	19