

# Data Management Plan

Deliverable 5.4



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 675191

## About this document

**Work package in charge:** WP5 Management & Dissemination

**Actual delivery for this deliverable:** 28 February 2016

**Dissemination level:** PU (for public use)

**Lead author:** German Climate Computing Center (DKRZ), Project office, Kerstin Fieg, Chiara Bearzotti

**Other contributing partners:**

German Climate Computing Center (DKRZ), Julian Kunkel

European Centre for Medium-Range Weather Forecasts (ECMWF), Daniel Thiemert, Peter Bauer

Centre National de Recherche Scientifique - Institut Pierre-Simon Laplace (CNRS-IPSL), Sylvie Joussaume

Barcelona Supercomputing Center (BSC), Oriol Mula-Valls, Kim Serradell

Max Planck Institute for Meteorology (MPI-M), Reinhard Budich

Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique (CERFACS), Sophie Valcke

Science and Technology Facilities Council (STFC), Martin Juckes

Sveriges Meteorologiska och Hydrologiska Institut (SMHI), Uwe Fladrich

**Contacts:** [esiwace@dkrz.de](mailto:esiwace@dkrz.de)

**Visit us on:** [www.esiwace.eu](http://www.esiwace.eu)

**Follow us on Twitter:** [@esiwace](https://twitter.com/esiwace)

Disclaimer: This material reflects only the authors view and the Commission is not responsible for any use that may be made of the information it contains.

## Index

1. Abstract / publishable summary .....	4
2. Conclusion & Results .....	4
3. Project objectives .....	4
4. Detailed report on the deliverable.....	5
4.1 Why is a Data Management Plan needed?.....	5
4.2 Who is responsible for the implementation of the DMP?.....	6
4.3 What kind of data will be affected by DMP? .....	6
5. References ( <i>Bibliography</i> ).....	7
6. Dissemination and uptake.....	7
6.1 Dissemination .....	7
6.2 Uptake by the targeted audience .....	7
7. The delivery is delayed: <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No.....	7
8. Changes made and/or difficulties encountered, if any .....	7
9. Efforts for this deliverable.....	7
10. Sustainability.....	8
10.1. Lessons learnt: both positive and negative that can be drawn from the experiences of the work to date.....	8
10.2 Links built with other deliverables, WPs, and synergies created with other projects .....	8
11. Dissemination activities.....	8

## 1. Abstract / publishable summary

As a research infrastructure project, ESIWACE is part of the Horizon 2020 Open Research Data Pilot, the Pilot project of the European Commission which aims to improve and maximise access to and reuse of research data generated by projects. The focus of the Pilot is on encouraging good data management as an essential element of research best practice.

This Data Management Plan (DMP) of ESIWACE describes the life cycle of all modelling and observation data collected and processed in ESIWACE.

## 2. Conclusion & Results

The DMP has to be seen as starting point for the discussion with the community about the ESIWACE data management strategy and reflects the procedures planned by the work packages at the beginning of the project. After asking the work package leaders to answer a questionnaire<sup>1</sup> on which data they were expecting to produce and collect, it has become clear that currently only work package 2 (Scalability) plans to generate or collect data that can be classified as data relevant according to the definition of the European Commission<sup>2</sup>.

Nonetheless, it can be the case that this situation evolves during the lifespan of the project. Thus the DMP will be updated twice during project lifetime with the Project Periodic Reports.

The current version of the DMP is attached as **Annex** to this document.

## 3. Project objectives

This deliverable contributes directly and indirectly to the achievement of all the macro-objectives and specific goals indicated in section 1.1 of the Description of the Action [DoA]:

Macro-objectives	Contribution of this deliverable?
Improve the efficiency and productivity of numerical weather and climate simulation on high-performance computing platforms	Yes
Support the end-to-end workflow of global Earth system modelling for weather and climate simulation in high performance computing environments	Yes
The European weather and climate science community will drive the governance structure that defines the services to be provided by ESIWACE	Yes
Foster the interaction between industry and the weather and climate community on the exploitation of high-end computing systems, application codes and services.	Yes
Increase competitiveness and growth of the European HPC industry	Yes

<sup>1</sup> The questions the WP leaders had to reply were the following:

- What types of data will the project generate/collect?
- What standards will be used?
- How will this data be exploited and/or shared/made accessible for verification and re-use? If data cannot be made available, explain why.
- How will this data be curated and preserved?

<sup>2</sup> The definition can be found in Annex 1

Specific goals in the workplan	Contribution of this deliverable?
Provide <b>services</b> to the user community that will impact beyond the lifetime of the project.	Yes
Improve <b>scalability</b> and shorten the time-to-solution for climate and operational weather forecasts at increased resolution and complexity to be run on future extreme-scale HPC systems.	Yes
Foster <b>usability</b> of the available tools, software, computing and data handling infrastructures.	Yes
Pursue <b>exploitability</b> of climate and weather model results.	Yes
Establish governance of common software management to avoid unnecessary and redundant development and to deliver the best available solutions to the user community.	Yes
Provide <b>open access</b> to research results and <b>open source</b> software at international level.	Yes
Exploit <b>synergies</b> with other relevant activities and projects and also with the global weather and climate community	Yes

## 4. Detailed report on the deliverable

### 4.1 Why is a Data Management Plan needed?

The partners of ESIWACE participate in the **Open Access Pilot for Research Data**. The Data Management plan specifies the implementation of the pilot, in particular with regard to the data generated and collected, the standards in use, the workflow to make the data accessible for use, reuse and verification by the community and define the strategy of curation and preservation of the data.

Thus, we refer to the **ESIWACE Grant Agreement (GA), Article 29.3** about **“Open Access to research data”**:

“Regarding the digital research data generated in the action (‘data’), the beneficiaries must:

(a) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:

(i) the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;

(ii) other data, including associated metadata, as specified and within the deadlines laid down in the 'data management plan' (see Annex 1);

(b) provide information — via the repository — about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and — where possible — provide the tools and instruments themselves).”

Moreover in the **ESIWACE GA Part B, Section 2.2.2** about **“Data/software policy and management of intellectual property rights (IPR)”** clarifies the purpose and sustainability of the document:

“The strategy for the knowledge management, protection, dissemination and for the exploitation of results will be defined in Dissemination and Exploitation Plan [D5.5]. The strategic document will be regularly updated during the entire project. Updates will be submitted to the European Commission as an integral part of the Project Periodic Reports. A final document, a Strategy for intellectual property exploitation [D5.6] will also be made available at the very end of the project. ESIWACE results will be exploited at European and international level by weather and climate modelling groups (research institutions, weather forecast services) relying on HPC resources”

These requirements are summarized in the **Description of Actions (DoA): Task 5.3 Innovation and IPR management, exploitation of results [Lead: DKRZ. Participants: ECMWF]**

“ESiWACE [...] is voluntarily taking part in the European Commission **Open Access Data Pilot for Research Data** (see Section 2.2.2): we have included a **Data Management Plan** as a deliverable for project-month 6 **[D5.4]** to be drafted in compliance with the guidelines given on data management in the Horizon 2020 Online Manual. This deliverable will evolve during the lifetime of the project and represent faithfully the status of the project reflections on data management. Updates of the data management plan are thus planned and will be submitted to the EC as an integral part of the Project Periodic Reports.”

#### **4.2 Who is responsible for the implementation of the DMP?**

Lead for this task will be with DKRZ, co-lead with CNRS-IPSL and ECMWF, though all partners are involved in the compliance of the DMP.

The partners agreed to deliver datasets and metadata produced or collected in ESIWACE according to the rules described in the DMP (Annex 1). The project office and in particular the Scientific Officer are also central players in the implementation of the DMP and will track the compliance of the rules agreed.

#### **4.3 What kind of data will be affected by DMP?**

The Open Research Data Pilot applies to two types of data<sup>3</sup>:

- 1) the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible.
- 2) other data, including associated metadata, as specified and within the deadlines laid down in the data management plan—that is, according to the individual judgement by each project.

According to the “Guidelines on Data Management in Horizon 2020” (2015) the DMP describes the handling of *numerical datasets* processed or collected during ESIWACE lifetime. The DMP include clear descriptions and rationale for the access regimes that are foreseen for collected data sets.

Thus the DMP leaves explicitly open the handling, use and curation of products like tools, software and written documents, which could also be subsumed under the generic term “data”; we restrict the focus of our DMP to numerical data products like produced model data or observation data.

---

<sup>3</sup> Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, Version 30 October 2015.

## 5. References (*Bibliography*)

Guidelines on Data Management in Horizon 2020, Version 2.0, 30 October 2015:

[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)

Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, Version 2.0, 30 October 2015:

[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)

Annotated GA version 30 October 2015, p. 218

Fact Sheet: Open Access in Horizon 2020:

[http://www.nks-swg.de/media/content/FactSheet\\_Open\\_Access.pdf](http://www.nks-swg.de/media/content/FactSheet_Open_Access.pdf)

Webpage of European Commission regarding Open Access:

[http://ec.europa.eu/research/science-society/open\\_access](http://ec.europa.eu/research/science-society/open_access)

## 6. Dissemination and uptake

### 6.1 Dissemination

Not applicable.

### 6.2 Uptake by the targeted audience

As indicated in the Description of the Action, the audience for this deliverable is

X	The general public (PU)
	The project partners, including the Commission services (PP)
	A group specified by the consortium, including the Commission services (RE)
	This reports is confidential, only for members of the consortium, including the Commission services (CO)

**This is how we are going to ensure the uptake of the deliverables by the targeted audience:**

The Data Management Plan will be made available to the public on the website [www.esiwace.eu](http://www.esiwace.eu) and to the beneficiaries in the project intranet.

**7. The delivery is delayed:**  Yes  No

## 8. Changes made and/or difficulties encountered, if any

Not applicable to this specific deliverable.

## 9. Efforts for this deliverable

Person-months spent on this deliverable:

Beneficiary	Person-months	Period covered	Names of scientists involved and their gender (f/m)
DKRZ	0.5	01.11.15 – 29.02.16	Kerstin Fieg (F) Chiara Bearzotti (F) Julian Kunkel (M)
ECMWF	0.15	01.11.15 – 29.02.16	Daniel Thiemert (M) Peter Bauer (M)
CNRS-IPSL	0	01.11.15 – 29.02.16	Sylvie Jousaume (F)
MPI-M	0	01.11.15 – 29.02.16	Reinhard Budich (M)
CERFACS	0	01.11.15 – 29.02.16	Sophie Valcke (F)
BSC	0	01.11.15 – 29.02.16	Oriol Mula-Valls (M) Kim Serradell (M)
STFC	0	01.11.15 – 29.02.16	Martin Juckes (M)
SMHI	0	01.11.15 – 29.02.16	Uwe Fladrich (M)
<b>Total</b>	<b>0.65</b>		

## 10. Sustainability

### 10.1. Lessons learnt: both positive and negative that can be drawn from the experiences of the work to date

Our lesson learnt is that we have to keep the group aware of the project participation in the Pilot, and to monitor closely with the WP leaders if and how the project will collect data during its lifespan; that is why updates of the DMP are planned in correspondence with the reporting periods to the EC.

For this reason, we have planned to send out the original questionnaire to the work package leaders at the reporting deadlines, in order to get an overview on numerical modelling or observation data sets the work packages will produce or collect.

### 10.2 Links built with other deliverables, WPs, and synergies created with other projects

Not applicable to this specific deliverable.

## 11. Dissemination activities

Not applicable to this specific deliverable.



# **Annex: Data Management Plan**

Deliverable 5.4

**Version 1.0, Feb. 2016**



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 675191

## About this document

**Work package in charge:** WP5 Management & Dissemination

**Actual delivery for this deliverable:** 28 February 2016

**Dissemination level:** PU (for public use)

**Lead author:** German Climate Computing Center (DKRZ), Project office, Kerstin Fieg, Chiara Bearzotti

**Other contributing partners:**

German Climate Computing Center (DKRZ), Julian Kunkel

European Centre for Medium-Range Weather Forecasts (ECMWF), Daniel Thiemert, Peter Bauer

Centre National de Recherche Scientifique - Institut Pierre-Simon Laplace (CNRS-IPSL), Sylvie Joussaume

Barcelona Supercomputing Center (BSC), Oriol Mula-Valls, Kim Serradell

Max Planck Institute for Meteorology (MPI-M), Reinhard Budich

Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique (CERFACS), Sophie Valcke

Science and Technology Facilities Council (STFC), Martin Juckes

Sveriges Meteorologiska och Hydrologiska Institut (SMHI), Uwe Fladrich

**Contacts:** [esiwace@dkrz.de](mailto:esiwace@dkrz.de)

**Visit us on:** [www.esiwace.eu](http://www.esiwace.eu)

**Follow us on Twitter:** [@esiwace](https://twitter.com/esiwace)

Disclaimer: This material reflects only the authors view and the Commission is not responsible for any use that may be made of the information it contains.

## Index

1. Executive Summary.....	4
2. Introduction .....	4
3. Register on numerical data sets generated or collected in ESiWACE .....	5
3.1 Datasets collected within WP1 .....	5
3.2 Datasets collected within WP2 .....	5
3.3 Datasets collected within WP3 .....	8
3.4 Datasets collected within WP4 .....	8
3.5 Datasets collected within WP5 .....	9
4. References ( <i>Bibliography</i> ).....	9
5. Glossary.....	9

## 1. Executive Summary

The Data Management Plan (DMP) of ESIWACE gives an overview of available research data, access and the data management and terms of use. The DMP reflects the current state of the discussions, plans and ambitions of the ESIWACE partners, and will be updated as work progresses.

## 2. Introduction

### Why a Data Management Plan (DMP)?

It is a well-known phenomenon that the amount of data is increasing while the use and re-use of data to derive new scientific findings is more or less stable. This does not imply, that the data currently unused are useless - they can be of great value in future. The prerequisite for meaningful use, re-use or recombination of data is that they are well documented according to accepted and trusted standards. Those standards form a key pillar of science because they enable the recognition of suitable data.

To ensure this, agreements on standards, quality level and sharing practices have to be negotiated. Strategies have to be fixed to preserve and store the data over a defined period of time in order to ensure their availability and re-usability after the end of ESIWACE

### What kind of data are considered in the DMP?

The main purpose of a Data Management Plan (DMP) is to describe *Research Data* with the metadata attached to make them *discoverable, accessible, assessable, usable beyond the original purpose and exchangeable* between researchers.

According to the "Guidelines on Open Access to Scientific Publication and Research Data in Horizon 2020" (2015):

*"Research data refers to information, in particular facts or numbers, collected to be examined and considered and as a basis for reasoning, discussion, or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form."*

However, the overall objective of ESIWACE is to improve efficiency and productivity of numerical weather and climate simulations on HPC systems by enhancing the scalability of numerical models, foster the usability of community wide used tools and pursue the exploitability of model output.

Thus ESIWACE focuses more on the production process and tools than on production of research or observation data and so the amount of *Research Data* which ESIWACE intend to produce is limited, at least at this stage of the project.

### What can be expected from ESIWACE DMP?

In the following we will describe the lifecycle, responsibilities and review processes and data management policies of research data, produced in ESIWACE. The DMP reflects the current status of discussion within the consortium about the data that will be produced. It is not a fixed document, but evolves during the lifespan of the project.

The target audience of the DMP is all project members and research institutions using the data and data produced.

### 3. Register on numerical data sets generated or collected in ESiWACE

The register has to be understood as living document, which will be updated regularly during project lifetime. The intention of the DMP is to describe numerical model or observation datasets collected or created by ESiWACE during the runtime of the project.

The information listed below reflects the conception and design of the individual work packages at the beginning of the project. Because the operational phase of the project started in January 2016, there is no dataset generated or collected until delivery date of this DMP.

The data register will deliver information according to Annex 1 of the Horizon 2020 guidelines (2015) (*in italics*):

- **Data set reference and name:** *Identifier for the data set to be produced.*
- **Data set description:** *Descriptions 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, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse.*
- **Standards and metadata:** *Reference to existing suitable standards of the discipline. If these do not exist, an outline on how and what metadata will be created.*
- **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.).*  
*In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).*
- **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, what is its approximated end volume, what the associated costs are and how these are planned to be covered*

#### 3.1 Datasets collected within WP1

<b>WP1 Governance, Engagement and long-term sustainability</b>	
<b>What types of data will the project generate/collect?</b>	WP1 is not going to generate numerical data sets.

#### 3.2 Datasets collected within WP2

<b>WP2 Scalability</b>	
<b>Data set reference and name</b>	EC-Earth model output and performance data
<b>Data set description</b>	EC-Earth high-resolution model output will be generated for test runs. Furthermore, performance data will be collected.

	Constraints: IFS data may not be used for commercial purpose.
<b>Standards and metadata</b>	<p>Model output will be in NetCDF and GRIB.</p> <p>No metadata is automatically generated by the model. CMIP6-compliant metadata generation may become available during the course of the project.</p> <p>No quality check is applied automatically. If necessary, CMIP6 compliant quality checking may be applied.</p>
<b>Data Sharing</b>	<p>EC-Earth model data and performance data will be shared (if useful):</p> <ul style="list-style-type: none"> <li>- Within the ESIWACE project, particularly WP2</li> <li>- Within the EC-Earth consortium</li> <li>- Within the ENES community, particularly the IS-ENES2 project</li> </ul> <p>Data sharing will generally be through access to the HPC systems or data transfer to shared platforms.</p> <p>If common experiments are run in the context of other projects (e.g. PRIMAVERA, CMIP6), data publication may be through ESGF.</p>
<b>Archiving and preservation (including storage and backup)</b>	<p>Long-term data storage will most likely not be needed for the data created in this project, the exception being potential common experiments with other projects. In the latter case, data storage will be provide by the respective projects.</p>
<b>Reported by</b>	Uwe Fladrich (uwe.fladrich@smhi.se)

<b>Data set reference and name</b>	BSC Performance Analysis
<b>Data set description</b>	<p>In WP2, BSC will carry on performance analysis and modifications to the source code of the earth system models to run in others programming models (like OmpSs).</p> <p>While the modified model code is no data to be described here, the performance analysis will produce trace outputs that contain the information of an execution of the model. In this case, the size can be a constraint. On many-core systems, the traces generated by a complex model can have a very big size (more than hundreds of gigabytes) so this can be a problem to share this information between partners. The integration and the reuse of this information would not be a problem if the different actors take a first decision in the tools to be used in these performance analyses.</p>
<b>Standards and metadata</b>	<p>All the tools to trace executions provide information about the format of the outputs and how to read them. Moreover, some of these tools can convert formats to improve the compatibility.</p> <p>Data can be in a raw binary or text format. In this last case, CSV or XML are usual formats to deal with the information.</p> <p>In the case of Paraver tool, in each trace there is a file describing which events are in the trace. This file usually contains a code and a text description for each event.</p>

<b>Data Sharing</b>	For the traces, a repository allowing the distribution of big files must be implemented. If the distribution is individual and sporadic, a solution like an FTP can fit to the requirement. If we want to setup a repository with all the traces for further analyses, another solution must be deployed. The solution will have to classify data among the model run, the platform, the configuration. This can lead to a big number of different combinations.
<b>Archiving and preservation (including storage and backup)</b>	Codes will be stored in the gitlab, during the time that the partners consider it convenient, but for the traces, due to the high volume of the data generated, another strategy has to be designed. Long term storage solution (like tapes) could be a good solution. Traces are usually a collection of big files suited to be stored in tape solution archive.
<b>Reported by</b>	Kim Serradell (kim.serradell@bsc.es)

<b>Data set reference and name</b>	IFS and OpenIFS model output.
<b>Data set description</b>	IFS and OpenIFS model integrations will be run and standard meteorological and computing performance data output will be generated. Both will be run at ECMWF, and only performance data will be made available to the public. The meteorological output will be archived in MARS, as it is standard research experiment output. The data will be used for establishing research and test code developments, and will enter project reports and generally accessible publications. The IFS will not be made available, OpenIFS is available through a dedicated license.
<b>Standards and metadata</b>	IFS meteorological output (incl. metadata) and format follows WMO standards. Compute performance (benchmark) output will be stored and documented separately. Data will be in ASCII and maintained locally. The output will be reviewed internally, and the ECMWF facilities allow reproduction of this output if necessary.
<b>Data Sharing</b>	All output can be shared within the ESIWACE consortium, and is primarily located in the ECMWF archiving system MARS. Data provision to the public is limited for meteorological output, and it adheres to the ECMWF data policy. Access can be granted in individual cases. Computing performance output can be made publicly available. This output can be managed by the ESIWACE website.
<b>Archiving and preservation (including storage and backup)</b>	As no large quantities of data will be produced, there are no requirements for long-term data management. The experiment output is stored in MARS that is backed up regularly. Volumes and cost are negligible.
<b>Reported by:</b>	Peter Bauer (peter.bauer@ecmwf.int)

<b>Data set reference and name</b>	
<b>Data set description</b>	WP2 will extend the benchmark suite fro coupling technologies

	<p>currently developed in IS-ENES2 to target new platforms with <math>O(10K-100K)</math> cores accessible during the ESiWACE longer timeframe. OASIS, OpenPALM, ESMF, XIOS and YAC will be considered.</p> <p>Benchmark suites for I/O libraries and servers will have to be built from scratch. The inter- comparison will include XIOS, ESMF and CDI-pio.</p> <p>A subset of the results of these benchmarks for specific technologies on specific computing platforms will be collected and made available as a reference.</p>
<b>Standards and metadata</b>	<p>The data per se will be just text files containing numbers (e.g. the communication time for a specific coupling exchange as a function of the number of cores used to run the coupled components) and will not adhere to any specific standard.</p> <p>The metadata attached to the data will contain the revision number of the benchmark sources that will be managed under SVN or GIT and a description of the parameters tested for a specific set of results (e.g. number of cores, number of coupling fields, etc.). The metadata will appear also as a text file (in the form of a Readme file) available in the data directory.</p> <p>The results of the benchmarks will be reviewed by the participating IS-ENES2 partners and reported in ESiWACE D2.1</p>
<b>Data Sharing</b>	<p>The benchmark sources (managed under SVN or GIT) and subset of results will be freely accessible to all. The description on how to access the sources and results will be available on ESiWACE web site.</p>
<b>Archiving and preservation (including storage and backup)</b>	<p>The subset of benchmark results and associated metadata will be uploaded to a data centre (e.g. DKRZ) and attached with a standard data DOI. Specific subset of results data will curated and preserved as a reference to compare with for the people who would want to run the benchmark themselves for <math>O(10)</math> years and will be regularly replaced by new subsets of new tests for new platforms.</p>
<b>Reported by:</b>	Sophie Valcke (sophie.valcke@cerfacs.fr)

### 3.3 Datasets collected within WP3

<b>WP3 Usability</b>	
<b>What types of data will the project generate/collect?</b>	WP3 is not going to generate typical numerical data sets, WP3 is going to produce papers and reports, and to some extent software code.

### 3.4 Datasets collected within WP4

<b>WP4 (Exploitability)</b>	
-----------------------------	--

<b>What types of data will the project generate/collect?</b>	WP4 (Task 4.3) will generate semantic mappings between metadata standards. The mappings will be made available through a SPARQL server and curated at STFC and ECMWF
--	--

### 3.5 Datasets collected within WP5

<b>WP5 Management and Dissemination</b>	
<b>What types of data will the project generate/collect?</b>	WP5 is not going to generate numerical data sets

## 4. References (*Bibliography*)

Guidelines on Data Management in Horizon 2020, Version 2.0, 30 October 2015:

[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)

Guidelines on Open Access to Scientific Publication and Research Data in Horizon 2020, Version 2.0, 30 October 2015

[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)

## 5. Glossary

DOI	Digital Object Identifier
DMP	Data Management Plan
EC	European Commission
GRIB	GRIdded Binary format, WHO
H2020	Horizon 2020, EU funding Strategy for 2014 - 2020
pdf	Portable Document Format
NetCDF	NETwork Common Data Format
ppt	Power Point
RDF	Resource Description Framework
SPARQL	SPARQL Protocol and RDF Query Language