



Horizon 2020 Societal challenge 5:
Climate action, environment, resource
efficiency and raw materials

IMPRES

IMproving PRedictions and management of hydrological EXtremes

641811, Funding type RIA

Deliverable number (relative in WP)	D6.2
Deliverable name:	Data Management Plan (DMP)
WP / WP number:	WP6.2
Delivery due date:	Project month 6 (31/03/2016)
Actual date of submission:	30/03/2016
Dissemination level:	public
Lead beneficiary:	ECMWF
Responsible scientist/administrator:	David Lavers / Florian Pappenberger
Estimated effort (PM):	3 PM
Contributor(s):	Janet Wijngaard (KNMI)
Estimated effort contributor(s) (PM):	--
Internal reviewer:	Florian Pappenberger , Janet Wijngaard

1. Changes with respect to the DoA

(with justification if applicable)

No changes

2. Dissemination and uptake

The deliverable will be used by project partners to communicate the availability of data

3. Short Summary of results (<250 words)

All of the project partners were requested to submit details of their datasets (to be generated) into an online document. These responses were compiled and edited, and after iterating through the responses with project partners, the Data Management Plan for IMPREX was formed.

4. Evidence of accomplishment

(report, manuscript, web-link, other)

See document below, including:

- a) A spreadsheet detailing the datasets to be generated.
- b) Online questionnaire with detailed guidance

Annex: Overview quality control

1. Meet and beat state of the art standards	Name reviewer: Florian Pappenberger (<i>data manager/ WP4 leader</i>)	Date: March 2016
<i>What is the baseline/reference used (e.g. iso) standard, publication, model version, technique, other). How does the deliverable relate to the baseline. Describe & underpin the improvements in relation to the baseline.</i> The DMP follows the guidelines of Horizon2020 on data management.		
2. Fit for purpose	Name reviewer: Janet Wijngaard (<i>project manager</i>)	Date: March 2016
In which field/sector is(are) the user(s) active? How does the user apply the deliverable (enhances process/knowledge/other). Does it contribute to the user's needs, and how will/could it become beneficial to the user? It is a useful and clear overview. It forms a good base to link the data generated by the tooling work packages (WP3-5) with the data needs from the sectoral work packages (WP7-12).		
3. Complete and consistent	Name reviewer: Janet Wijngaard (<i>project manager</i>)	Date: March 2016
<i>Does the deliverable contain all required components. Do the different parts support each other and does the deliverable align to the DoA.</i> The DMP evolves during the lifespan of the project ensuring a complete overview along the course of the project.		

Deliverable 6.2: Data management Plan

1. Aim of the DMP

The purpose of the Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that will be used by the applicants with regard to all the datasets that will be generated by the project. The DMP is not a fixed document, but evolves during the lifespan of the project.

2. Description of DMP

This paragraph gives a brief description of the columns in the DMP overview of datasets (table1).

(The appendix provides details on the questionnaire used to produce this table.)

Column 1: Provisional Data set reference and name

Column 2: Data set description

Description of the data that will be generated or collected, its origin (in case it is collected), nature and scale, and to whom it could be useful.

Column 3: Standards and metadata

Reference to existing suitable standards of the discipline.

Column 4: Data sharing

Description of how data will be shared, including access procedures, outlines of technical mechanisms for dissemination, and definition of whether access will be widely open or restricted to specific groups.

Column 5: Archiving and preservation of the data sets

Identification of the repository where data will be stored is given where known.

Column 6: Name and Institution

The point of contact (and institution) for the data sets.

Column 7: The task that the dataset is associated with

Column 8: The deliverable or Milestone that the dataset is associated with

The data management plan follows the “Guidelines on Data Management in Horizon 2020” (http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf).

D

1. Provisional Data set name	2. Data set description	3. Standards and meta data	4. Data Sharing	5. Archiving and preservation	0. Name and Institution	2.1 Associated task	2.2 Deliverable/Milestone
Seasonal forecasts (ECMWF System 4)	Long range (seasonal) forecasts	GRB (GRidded Binary)	ECMWF MARS archive	ECMWF MARS archive	Louise Arnel, ECMWF	3.2 and 4.2	D3.3, D3.4, D4.3
ECMWF ensemble forecasts	Extended Range forecasts	GRB (GRidded Binary)	ECMWF MARS archive	ECMWF MARS archive	David Laers, ECMWF	3.1	D3.1 and 3.2
GoSEA5	Daily/monthly mean forecast and seasonal ensemble discharge	NetCDF-CF	FTP server	Met Office Archive / Cope Jeff Knight, Met Office	3.2		MSR (month 12) and MS 10 (month 24)
Seasonal hydrological forecasts (SEA)	Seasonal ensemble discharge	PCRaster format.	FTP server	Archiving is done according to Louise Arnel, ECMWF	7.2		D7.3
fd_inetj [date], where [inetj] and [date] correspond to the	The description is separated into two	Flood damage modelling at GFZ is	Contact Heidi Kriebich (heidi.kriebich@gfz.de)	No meta-data standard	Heidi Kriebich, Stefan	5.2; 7.2	D5.2, 5.4; 7.2
RACMO 16 member ensemble	This is a 12 km resolution data set of 15 NetCDF	NetCDF	Data are free& can be provided on request	Data is stored locally on a Geert Lenderink, KNMI	4.4 and 3.3		D4.4 and D3.3
JULES 1.5km climate simulations High resolution nested suite	JULES	NetCDF	Data can be provided on request	Data is stored locally on a Erik van Meijgaard, KNMI	7.3		D7.2
RACMO timeseries	Data to be generated: RACMO-1) bias	NetCDF / comma-separated plain ASCII file	The relevant data will be verified	The processed data will be public soon. *BADC.	Andrea Castelletti,		D3.3, D3.4
Seasonal weather forecast (WP2)	Medium-long term hindcasts	NetCDF / comma-separated plain ASCII file	FTP server	NMME project (monthly h	Virginie Guemas, BSC	3.2	D3.3, D3.4
EC-Earth seasonal predictions	Global 1-degree seasonal predictions	NetCDF	FTP server	http://s2s.cma.cn/index	Virginie Guemas, BSC	3.1	D3.2, D3.4
NMME	Six global seasonal forecast systems.	NetCDF	FTP server	http://apps.ecmwf.int/act	Virginie Guemas, BSC	3.2	D3.3, D3.4
SZS	Sub-seasonal to seasonal climate predict	NetCDF	FTP server	Local server, with regular Johannes Hunink (Future)	11.1 and 11.3		D11.1, D11.2 and D11.4, MS16 and MS19
EUROSIIP	Four global operational seasonal predict	NetCDF	Data can be provided on request	Local server, with regular Johannes Hunink (Future)	8.2		D8.2
Seasonal reservoir inflow hindcasts	SPHY model forced by seasonal hindcast	PCRaster format	The relevant data will be verified	Local server, with regular Andrea Castelletti	8.3		D11.4 and D11.5
TOPKAPETH hydrological model simulation	Simulated streamflow, snow water	NetCDF / comma-separated plain ASCII file	The relevant data will be verified	Local server, with regular Andrea Castelletti	8.2		D4.4 / D11.5
DISTRILAKE reservoir model simulation	Water release, operation rules and	NetCDF / comma-separated plain ASCII file	The relevant data will be verified	Local server, with regular Johannes Hunink (Future)	11.4 and 11.5		D11.2 and D11.3
MAS-IDRAGRA agronomic model simulation	Simulated crop productivity, soil water	NetCDF / comma-separated plain ASCII file	Data can be provided on request	Local server, with regular Johannes Hunink (Future)	11.1 and 11.2		D10.1
Water accounts	Drought projections on climate scale	PCRaster and some tabular format	The data are available to IMPRE	Local data servers (TUC)	11.1 and 11.2		D12.1
Seasonal hydrological forecasts/drought indices	Coupled hydrological/groundwater mode	Time series data format.	Not yet defined in our project	Not yet defined in our project	10.1		D12.1
Water quality forecasts	Lohreat and Segura basins	Not yet defined in our project	Data will be shared upon demand	Data will be part of Water Footprint Network	12.1 and 12.2.		D12.1 and D12.2
Water Footprint of production (WaterStat database)	water footprint of production for different economic sectors (e.g. agriculture and	Data will be stored and available as PCRaster.	FTP server (k.kajja.frieler@potsdam.de)	The data will be archived ; Potsdam Institute for	Fang Zhao, Katja Frieler	12.3	D8.3, D4.3
Water footprint of consumption dataset	water footprint of consumption for Global gridded data on river discharge, flood water depth, flood area, flood	NetCDF	FTP server (Klein@baif.de)	Federal Institute of Hydro	Bastian Klein, Federal	4.3 and 9.3	D8.3
ISIMP-Flood (no reference paper yet)	Reference runoff simulation for the River	NetCDF	FTP server (Klein@baif.de)	Federal Institute of Hydro	Bastian Klein, Federal	9.3	D8.3, D4.3
HBV134 River Rhine runoff simulation	Reference waterlevel / runoff / flow veloci	NetCDF	FTP server (Klein@baif.de)	Federal Institute of Hydro	Bastian Klein, Federal	4.3 and 9.3	D8.3
SOBEK River Rhine waterlevel simulation	Reference runoff simulation for selected	NetCDF	FTP server (Klein@baif.de)	Federal Institute of Hydro	Bastian Klein, Federal	4.1, 4.3, and 9.3	D8.3
LARSIM runoff simulation Central Europe	Short-to-medium-term (1-15d) historic ru	NetCDF	FTP server (Klein@baif.de)	Federal Institute of Hydro	Bastian Klein, Federal	4.1, 4.3, and 9.3	D8.3
Short-to-medium term River Rhine historic runoff hindcast	Short-to-medium-term (1-15d) historic w	NetCDF	FTP server (Klein@baif.de)	Federal Institute of Hydro	Bastian Klein, Federal	4.1, 4.3, and 9.3	D8.3
Short-to-medium term historic River Rhine waterlevel hindcast	Seasonal (1-6 months) historic runoff h	NetCDF	FTP server (Klein@baif.de)	Federal Institute of Hydro	Bastian Klein, Federal	4.4 and 9.3	D8.3
Seasonal historic runoff hindcasts LARSIM-CE Central Europe	Runoff projections for climate time slices	NetCDF	If data is not sensitive to the con	Resources and systems (IRSTEA)		8.2 and 8.3	D4.2, D4.4, D6.1, D6.2
Runoff projections LARSIM-CE Central Europe	We will collect data from WP3 (meteo	NetCDF format. And 'readme' Text files.	he data will be available to the IM	The data will be archived UPV (Abel Solera)		4.4, 5.4, 8.2, 8.3	D4.2, D4.4, D6.1, D6.2
GR_Simulations and ValPrev_Gains	For the case study Jucar, monthly water	NetCDF and text files	HTTP server	MARS archive at NSC	Magnus Lindskog, SMHI	3.1, 4.3, 8.3	D3.2
Monthly water resources variables results	Short range forecasts (couple of months	GRIB and/or NetCDF	HTTP server	MARS archive at NSC	Magnus Lindskog, SMHI	3.1, 4.3, 8.3	D3.2
Harmonie short term forecasts for Scandinavia	Short range forecasts (couple of months	GRIB and/or NetCDF	Data can be provided on request	NSC	Petter Lind, SMHI	3.3	D3.6
Harmonie short term forecasts for southern Europe	Short range climate time slices (weather)	GRIB and/or NetCDF	Data can be provided on request	NSC	Petter Lind, SMHI	3.3	D3.6
Harmonie climate time slices for Scandinavia	Short range climate time slices (weather)	GRIB and/or NetCDF	NSC and SMHI's testing c	illas Pechlivanidis, SMHI	4.1, 4.3, 8.2, 8.3		D4.1 and D4.3
Harmonie climate time slices for southern Europe	Short range climate time slices (weather)	GRIB and/or NetCDF	NSC and SMHI's testing c	illas Pechlivanidis, SMHI	4.1, 4.3, 8.2, 8.3		D4.1 and D4.3
Seasonal hydrological (discharge and inflows) forecasts based hindcasts	NetCDF		FTP server				
Short-term hydrological forecasts based on	NetCDF		FTP server				

Table 1: overview of datasets to be generated

Appendix:

Online questionnaire

The data management plan follows the “Guidelines on Data Management in Horizon 2020” (http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf). It was established following a set of questions:

4.1 Data set reference and name

Identifier for the data set to be produced

4.2 Data set description

Description of the data that will be generated or collected, its origin (in case it is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse.

4.3 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.

4.3.1 Guideline from proposal

All the data will be provided in NetCDF format, unless required otherwise. The data will comply with the CF standards and the metadata documentation protocol developed for the Fifth Coupled Model Intercomparison Project (CMIP5) archive. It will also follow the latest standards defined by the initiatives of the European Network for Earth-System Modelling (ENES) consortium

4.3.2 Guidance on Data Capturing Method

Questions to consider:

How will the data be created?

What standards or methodologies will you use?

How will you structure and name your folders and files?

How will you ensure that different versions of a dataset are easily identifiable?

Outline how the data will be collected/generated and which community data standards (if any) will be used at this stage. Indicate how the data will be organised during the project, mentioning for example naming conventions, version control and folder structures. Consistent, well-ordered research data will be easier for the research team to find, understand and reuse.

4.3.3 Guidance on metadata

Questions to consider:

How will you capture / create the metadata?

Can any of this information be created automatically?

What metadata standards will you use and why?

Metadata should be created to describe the data and aid discovery. Consider how you will capture this information and where it will be recorded e.g. in a database with links to each item, in a ‘readme’ text file, in file headers etc.

Researchers are strongly encouraged to use community standards to describe and structure data, where these are in place. A catalogue of disciplinary metadata standards can be used (e.g. <http://www.dcc.ac.uk/resources/metadata-standards>).

4.4 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).

4.4.1. Guidance on Methods of data sharing

Questions to consider:

How will you make the data available to others?

With whom will you share the data, and under what conditions?

Consider where, how, and to whom the data should be made available. Will you share data via a data repository, handle data requests directly or use another mechanism?

The methods used to share data will be dependent on a number of factors such as the type, size, complexity and sensitivity of data. Mention earlier examples to show a track record of effective data sharing.

4.4.2 Guidance on Restrictions on Sharing

Questions to consider:

Are any restrictions on data sharing required? e.g. limits on who can use the data, when and for what purpose.

What restrictions are needed and why?

What action will you take to overcome or minimise restrictions?

Outline any expected difficulties in data sharing, along with causes and possible measures to overcome these. Restrictions to data sharing may be due to participant confidentiality, consent agreements or IPR. Strategies to limit restrictions may include: anonymising or aggregating data; gaining participant consent for data sharing; gaining copyright permissions; and agreeing a limited embargo period.

4.4.3 Guidance on Data Repository

Questions to consider:

Where (i.e. in which repository) will the data be deposited?

Most research funders recommend the use of established data repositories, community databases and related initiatives to aid data preservation, sharing and reuse.

An international list of data repositories is available via Re3data (<http://www.re3data.org/>).

4.5 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.

4.5.1 Guidance on Preservation Plan

Questions to consider:

What is the long-term preservation plan for the dataset? e.g. deposit in a data repository

Will additional resources be needed to prepare data for deposit or meet charges from data repositories?

Researchers should consider how datasets that have long-term value will be preserved and curated beyond the lifetime of the grant. Also outline the plans for preparing and documenting data for sharing and archiving.

If you do not propose to use an established repository, the data management plan should demonstrate that resources and systems will be in place to enable the data to be curated effectively beyond the lifetime of the grant.

4.5.2 Guidance on Resourcing

Questions to consider:

What additional resources are needed to deliver your plan?

Is additional specialist expertise (or training for existing staff) required?

Do you have sufficient storage and equipment or do you need to cost in more?

Will charges be applied by data repositories?

Have you costed in time and effort to prepare the data for sharing / preservation?

Carefully consider any resources needed to deliver the plan. Where dedicated resources are needed, these should be outlined and justified. Outline any relevant technical expertise, support and training that is likely to be required and how it will be acquired. Provide details and justification for any hardware or software which will be purchased or additional storage and backup costs that may be charged by IT services.

Funding should be included to cover any charges applied by data repositories, for example to handle data of exceptional size or complexity. Also remember to cost in time and effort to prepare data for deposit and ensure it is adequately documented to enable reuse. If you are not depositing in a data repository, ensure you have appropriate resources and systems in place to share and preserve the data.

See UKDS guidance on costing data management

(<https://www.ukdataservice.ac.uk/manage-data/plan/costing>).