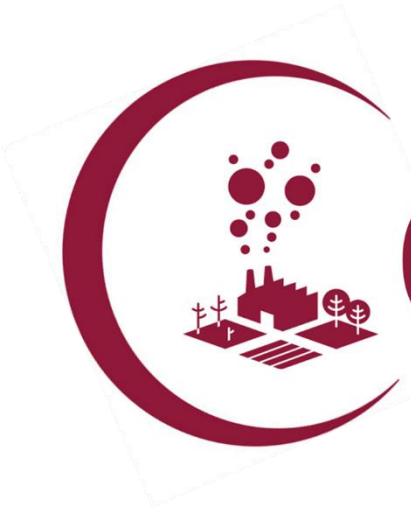

CONFESS

Consistent representation of temporal variations of boundary forcings in reanalyses and seasonal forecasts

Data Management Plan

Daniel Thiemert

www.confess-h2020.eu

Co-ordinated by ECMWF

D4.5 Data Management Plan

Author(s): Daniel Thiemert (ECMWF)

Dissemination Level: Public

Date: 27/04/2021
Version: 1.0
Contractual Delivery Date: 30/04/2021
Work Package/ Task: WP4/ T4.4
Document Owner: ECMWF
Contributors: All Partners
Status: Final

CONFESS

Consistent representation of temporal variations of boundary forcings in reanalyses and seasonal forecasts

Research and Innovation Action (RIA)
H2020- LC-SPACE-18-EO-2020 Copernicus evolution: Research activities in support of the evolution of the Copernicus services - Copernicus Climate Change Service (C3S

Project Coordinator: Dr Magdalena Alonso Balmaseda (ECMWF)
Project Start Date: 01/11/2020
Project Duration: 36 months

Published by the CONFESS Consortium

Contact:
ECMWF, Shinfield Park, Reading, RG2 9AX, United Kingdom
Magdalena.Balmaseda@ecmwf.int

The CONFESS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004156.

Contents

Tables

1 Executive Summary

The CONFESS Data Management Plan responds to the requirements of the H2020 Open Research Data Pilot to document which research data is being produced by the CONFESS project, in which format, and how it will be made available.

It has already identified data sets for work packages 1 to 3, but is only to be seen as an initial version which requires periodic updates to provide the necessary detail as it emerges.

2 Introduction

2.1 Background

A climate resilient society requires reliable monitoring and forecasting information of the climate trends, patterns and disturbances, both at global and regional scales. Through consistent representation of temporal variations of boundary forcings in reanalyses and Seasonal forecasts, CONFESS will contribute to the emerging societal need for an enhanced Copernicus Climate Change Service (C3S) that can support adaptation and mitigation strategies facing increased frequency and intensity of climate extremes.

The aim of CONFESS is to improve the reliability and usability of C3S information in the land-atmosphere coupled system by exploiting new and improved Earth Observations data records of land-cover/use, vegetation states and surface-emitted aerosols delivered across different Copernicus Services. CONFESS developments will be integrated consistently for use in future C3S systems, enhancing the service's accuracy by representing annual changes of land use, and adding satellite-derived and prognostic vegetation states along with aerosols emissions due to hazardous/extreme events such as volcanic eruptions and large-scale biomass burning (e.g. wildfires).

The added capacity to represent temporal variations and trends of these variables and the occurrence of hazardous/extreme events will be supported by a rapid uptake of new Earth Observations. The impact on the Earth system will be evaluated by assessing the quality of global reanalysis as well as seasonal forecasts using state-of-the-art modelling systems.

The infrastructure and knowledge developed within CONFESS will contribute to improve the C3S capabilities for reliable monitoring and forecasting with particular focus on extremes.

2.2 Scope of this deliverable

2.2.1 Objectives of this deliverable

D4.5 Data Management Plan provides the initial outline of the data management plan including information on which data sets will be created in the project and how they will be made available. This document represents only the initial version where details may not be available yet, and it will be further developed over the course of the project.

2.2.2 Work performed in this deliverable

The work performed included, as per the DoA, the collection of the available descriptions of data sets to be produced by the project, through a questionnaire.

2.2.3 Deviations and counter measures

No deviations have been encountered.

3 Open Research Data Objectives

3.1 Open Research Data Pilot

As per the Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020[1], Research Data

> "Refers to information, in particular facts or numbers, collected to be examined and considered 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."

The Open Research Data Pilot

> "aims to improve and maximise access to and re-use of research data generated by Horizon 2020 projects[2]"

and applies to data sets that are

> "needed to validate the results presented in scientific publications[2]".

The Data Management Plan is expected to

> "specify what data will be open: detailing what data the project will generate, whether and how "it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved[2]".

3.2 CONFESS Research Data

As per the CONFESS Description of Action, the products of CONFESS will comprise reports, graphical displays, datasets and improved methods, algorithms and code. The datasets will target a wide user community to support them with parallel or alternative studies. CONFESS will generate and use model output from Weather & Climate (W&C) models, weather observations, etc. The datasets will be curated by ECMWF as they are meant to become a valuable asset of the W&C community.

Overall, CONFESS will produce:

- Mapping of harmonized multi-year records of vegetation (Leaf Area Index - LAI) and land cover ready for ingestion by the CHTESSEL, EC-Earth HTESSEL-LPJGuess and SURFEX models, suitable for their use in reanalyses and seasonal forecasts. This should improve the temporal variations of land properties, with expected impacts on regional climate and hydrology.
- A harmonized multi-year data record of tropospheric aerosol by merging CMIP6 and CAMS information. This should result in an improved and more accurate representation of the time

[1] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
[2] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm

variation of tropospheric sulphate aerosols, available for use in future reanalyses and seasonal forecasts, with potential for improving global trends and low frequency large scale climate variability. It will be the first attempt to create a seamless temporal record of tropospheric aerosols suitable for a range of applications: reanalysis, seasonal forecasts, near-term predictions and multi-year climate integrations.

3.3 Data Management Plan Questionnaire

The following questionnaire has been provided to CONFESS work packages to gather the information for this first version of the Data Management Plan.

Table 1: Data Management Plan Questionnaire

<Data set reference and name>	
Data set description	*Description of the data that will be generated or collected (or is already available to the project), 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.* *Limitations?* *Constraints?*
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.* *Will you generate proper metadata for you data?* *If yes: how do they look like?* *If no: why?* *Data format?* *Will there be a review process to quality- check the data?*
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). *License?* *Access URL?*
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.* *At which Data Center do you want to store your data?* *Is there an established workflow for your requested DOI process in place?* *According to which standards*

4 CONFESS Data Sets

The following sections provide the responses by work packages 1 to 3. Work Package 4 does not produce any data sets.

4.1 WP1

Leaf Area Index (LAI)	
Data set description	The Leaf Area Index is defined as half the total area of green elements of the canopy per unit horizontal ground area. The satellite-derived value corresponds to the total green LAI of all the canopy layers, including the understory which may represent a very significant contribution, particularly for forests. Practically, the LAI quantifies the thickness of the vegetation cover. https://land.copernicus.eu/global/products/lai - Leaf Area Index observational data set based on SPOT-VEGETATION + PROBA-V (LAI v2) Origin: Copernicus Global Land Services (LAI v2) Time span : 1999-2019 Time resolution : 10-daily Horizontal resolution: 1 km Reference: Verger et al., 2014 + - Leaf Area Index observational data set based on AVHRR Origin: C3S (LAI v1) Time span : 1981-2005 Time resolution : 10-daily Horizontal resolution: 5 km ==> To be Produced in CONFESS: Merged LAI data at 1km resolution 1993-2019

	10-daily
Standards and metadata	- Data format : NetCDF4 -Metadata according to the Climate and Forecast (CF) conventions (v1.6) as described by the Copernicus Global Land Service that provide the data - Review process carried out by the data supplier (see publicly available Validation reports) Merged LAI data will be checked for temporal consistency
Data Sharing	Data sharing between partners through ftp temporary repositories without embargo. No dedicated public repository. Data set publicly available from the dedicated site https://land.copernicus.eu/global/products/lai Copernicus data and information are available on a full, open and free of charge basis, subject to limitations concerning registration, dissemination formats, and access restrictions.
Archiving and preservation (including storage and backup)	Data set stored and publicly available from the dedicated space in the Copernicus Global Land Service Data Center: https://land.copernicus.eu/global/products/lai Each partner will use its own archive system for long-term data storage

Land Surface Model Simulations	
Data set description	Generated data: - Land surface model output time series with and without improved vegetation and land-cover Time span : 1993-2019 Time resolution: daily Horizontal resolution: SURFEX-ISBA-CTRIP model : 1.5° CHTESSEL model: Tco199 (64 km) with a subset on Tco319 (36km) EC-Earth HTESSEL-LPJGuess model : T255 (~ 0.7°x0.7°).
Standards and metadata	- Data format and metadata: NetCDF SURFEX-ISBA-CTRIP model, EC-Earth HTESSEL-LPJGuess model & CHTESSEL model: - No specific review process besides basic checks usually done for any numerical model experiment
Data Sharing	Data sharing between partners through ftp temporary repositories, no embargo. No dedicated public repository. Data can be shared publicly by partners on-demand.

Archiving and preservation (including storage and backup)	Each partner will use its own archive system for long-term data storage

Land Cover	
Data set description	In the framework of the Climate Change Initiative (CCI) project, a "Land Cover" Essential Climate Variable was defined. Land Cover can be defined as the physical material at the surface of the earth. It include grass, asphalt, trees, bare ground, water, etc... http://www.esa-landcover-cci.org/ - Land cover observational data set : Origin: ESA-CCI Land Cover Time span: 1993-2019 Time resolution : yearly Horizontal resolution : 1km Reference: ESA. Land Cover CCI product User Guide Version 2. tech. Rep. (2017)
Standards and metadata	- Data format and metadata: GRIB and converted to NetCDF4 - No proper metadata: original metadata will be kept - Review process carried out by the data supplier (see publicly available Validation report: http://www.esa-landcover-cci.org/sites/default/files/documents/documents/Deliverables/Land_Cover_CCI_PVP_1.3.pdf)
Data Sharing	Data sharing between partners through ftp temporary repositories without embargo. No dedicated public repository. Data set publicly available from the dedicated site : http://www.esa-landcover-cci.org/?q=node/164
Archiving and preservation (including storage and backup)	Each partner will use its own archive system for long-term data storage

Land Use Harmonisation (LUH2)	
Data set description	The land-use harmonization strategy estimates the fractional land-use patterns, underlying land-use transitions, and key agricultural management information, annually for the time period 850-2100 https://luh.umd.edu/ Origin: Land Use Harmonization (LUH2) Time Span : 1993-2019 Time resolution: yearly Horizontal resolution : 0.25° Reference :Hurtt et al. 2020
Standards and metadata	- Data format and metadata: NetCDF4 - No proper metadata: original metadata will be kept

	- Review process carried out by the data supplier (see the dedicated scientific article by Hurtt et al. 2020)
Data Sharing	Data sharing between partners through ftp temporary repositories without embargo. No dedicated public repository. Data set publicly available from the dedicated site: https://luh.umd.edu/
Archiving and preservation (including storage and backup)	Each partner will use its own archive system for long-term data storage

Fraction of Vegetation Cover (FCover)	
Data set description	The Fraction of Vegetation Cover (FCover) corresponds to the fraction of ground covered by green vegetation. Practically, it quantifies the spatial extent of the vegetation (ranges from 0-1). Because it is independent from the illumination direction and it is sensitive to the vegetation amount, FCover is well suited for the replacement of classical vegetation indices for the monitoring of ecosystems. https://land.copernicus.eu/global/products/fcover Origin: Copernicus Global Land Services Time span : 1999-2019 Time resolution : 10-daily Horizontal resolution: 1 km References: Baret et al., 2013; Verger et al., 2014
Standards and metadata	- Data format: NetCDF4 -Metadata according to the Climate and Forecast (CF) conventions (v1.6) as described by the Copernicus Global Land Service that provide the data; see details at distribution page: https://land.copernicus.eu/global/products/fcover - Quality assessment carried out by the data supplier (Camacho et al., 2013; see publicly available reports at distribution page)
Data Sharing	Data sharing between partners through ftp temporary repositories without embargo. No dedicated public repository. Data set publicly available from the dedicated site https://land.copernicus.eu/global/products/fcover Copernicus data and information are available on a full, open and free of charge basis, subject to limitations concerning registration, dissemination formats, and access restrictions.

Archiving and preservation (including storage and backup)	Data set stored and publicly available from the dedicated space in the Copernicus Global Land Service Data Center: https://land.copernicus.eu/global/products/fcover Each partner will use its own archive system for long-term storage and backup.

4.2 WP2

Harmonised CMIP6/CAMS time-varying climatology of tropospheric aerosols	
Data set description	CMIP6 provides methods and data for generating time-evolving tropospheric aerosol loads, but the spatial detail is limited and the values are not well constrained by observations. ECWMF have developed a CAMS-based climatology of tropospheric aerosol which is well-constrained but only represents a fixed recent period of time. These two sources of data will be used to derive a new, time-varying climatology which keeps the best characteristics of each source. It will cover the time period from the mid-twentieth century to projected future values in the next few decades. It is intended to be of interest and use both for future reanalysis projects and for seasonal and perhaps decadal prediction systems. Although the methodology could be applied more generally, the dataset generated will be specific to the aerosol species represented in the IFS.
Standards and metadata	At present, aerosol climatologies are provided to the IFS as a NetCDF file with internal metadata specifying the type of aerosol, bin size (if appropriate), units, and descriptive metadata on how the data were created. This NetCDF format will be re-used and extended to include an additional time dimension. The existing metadata standards will be carried over and enhanced where necessary, taking account of CF standards. The NetCDF files will be visualized to allow basic visual quality control. They will also be tested in model runs. The datasets and number of fields are sufficiently small that manual checking is envisaged as being sufficient to guarantee the integrity of the data.
Data Sharing	The dataset will be shared on request. Although it is straightforward for project scientists to make the data available via an ftp server, ECMWF usually requires that data and software requests are forwarded to and handled by their Data Services team, to ensure appropriate record

	keeping, handling of any license conditions and the quality of the data transfer process. Since the dataset is in NetCDF format, no special software is needed to read it. The license terms for the data will be those usually applied at ECMWF for datasets whose production was funded by H2020, taking account of any attribution requirements coming from the input data.
Archiving and preservation (including storage and backup)	The dataset has low spatial and temporal resolution, and as such the data volume is anticipated to be small. If the existing file sizes are duplicated to cover a 100 year period, total file size would be less than 0.4 Gbytes, although the final approach taken may lead to a different value. The main working copies will be kept on backed-up and duplicated HPC file systems at ECMWF. Copies of the input data, processing software and final data product will also be retained and stored separately. The dataset will be kept for at least as long as it is in use by forecasting or re-analysis systems, and for some years afterwards. It is expected that it will be superseded by new improved datasets in the years ahead, so it will become obsolete in time. Copies may be kept indefinitely by users of the data for purposes of traceability, but that will not be the responsibility of this project. No DOI is planned at present, but it is not ruled out.

4.3 WP3

Seasonal re-forecasts	
Data set description	This data set comprises ensemble seasonal re-forecasts generated with - the CNRM-CM coupled climate model - the ECMWF/IFS model to investigate the impact of land surface and aerosols on predictability. Data will be global, cover the 1993-2016 period for several initialization dates. Data will be useful to scientists involved in CONFESS WP3.
Standards and metadata	Data will be provided in NetCDF4 or GRIB format. Metadata for NetCDF4 will follow the Copernicus Climate Change Services standards.

Data Sharing	Data will be shared among partners via ECMWF storage system. Dissemination of data through existing repositories (e.g. CHFP database) is currently under investigation.
Archiving and preservation (including storage and backup)	CNRM will ensure local archiving of re-forecasts with CNRM-CM produced in the framework of H2020-CONFESS.

Multi-year integrations	
Data set description	This data set comprises multi-year integrations with - the CNRM-CM coupled climate model - the EC-Earth coupled climate model to investigate the impact of vegetation and volcanic forcings. Data will be global and the integrations will cover a multi-decadal time period. Data will be useful to scientists involved in CONFESS WP3.
Standards and metadata	Data will be provided in NetCDF4 format. Metadata for NetCDF4 will follow CMIP6 (DCPP) standards.
Data Sharing	Data will be shared among partners for analysis via external ftp or another platform to be agreed upon. Dissemination of data through existing repositories is currently under investigation.
Archiving and preservation (including storage and backup)	CNRM will ensure local archiving of re-forecasts with CNRM-CM produced in the framework of H2020-CONFESS. The BSC and ISAC-CNR will do the same with EC-Earth re-forecasts.

5 Conclusion

This initial Data Management Plan has identified a number of data sets for each of the work packages 1 to 3, identifying the required details (where possible) on what data will be open, how it will be made accessible, and how it will be curated.

The Data Management Plan is to be seen as a living document and will be reviewed and revised periodically to ensure that information contained therein is up-to-date and correct.

Document History

Version	Author(s)	Date	Changes
0.1	Daniel Thiemert (ECMWF)	15/04/2021	initial version with WP input
1.0	Daniel Thiemert (ECMWF)	27/04/2021	final version

Internal Review History

Internal Reviewers	Date	Comments
Lauriane Batté (MF)	19/04/2021	Approved with comments
Andrea Alessandri (CNR-ISAC)	26/04/2021	Approved with comments

Estimated Effort Contribution per Partner

Partner	Effort
ECMWF	0.25
Total	**0.25**