



CO₂
Human
Emissions

Data Management Plan

Daniel Thiemert

che-project.eu



Co-ordinated by
 ECMWF



CO₂ Human Emissions

D7.5 Data Management Plan

Dissemination Level:	Public
Author(s):	Daniel Thiemert (ECMWF)
Date:	21/03/2018
Version:	1.0
Contractual Delivery Date:	31/03/2018
Work Package/ Task:	WP7/ T7.4
Document Owner:	ECMWF
Contributors:	WP Leaders
Status:	Final



CO₂ Human Emissions

CHE: CO₂ Human Emissions Project

Coordination and Support Action (CSA)
H2020-EO-3-2017 Preparation for a European
capacity to monitor CO₂ anthropogenic emissions

Project Coordinator: Dr Gianpaolo Balsamo (ECMWF)
Project Start Date: 01/10/2017
Project Duration: 39 months

Published by the CHE Consortium

Contact:

ECMWF, Shinfield Park, Reading, RG2 9AX,
gianpaolo.balsamo@ecmwf.int



The CHE project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 776186.



Table of Contents

1	Executive Summary	5
2	Introduction	5
2.1	Background.....	5
2.2	Scope of this deliverable	5
2.2.1	Objectives of this deliverables.....	5
2.2.2	Work performed in this deliverable	5
2.2.3	Deviations and counter measures	5
3	Open Research Data Objectives	6
3.1	Open Research Data Pilot	6
3.2	CHE Research Data	6
3.3	Data Management Plan Questionnaire	7
4	CHE Data Sets	9
4.1	Work Package 1	9
4.2	Work Package 2	11
4.3	Work Package 3	15
4.4	Work Package 4	17
5	Conclusion	17

Tables

Table 1: CHE Output Datasets.....	7
Table 2: Data Management Plan Questionnaire	7

1 Executive Summary

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

It has already identified data sets for work packages 1 to 4, 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

CHE, as a Coordination and Support Action, is bringing together European expertise and a consolidated approach to building an operational CO₂ emission monitoring capacity. CHE partners are at the forefront of developments in the compilation of emission inventories, the observation of the carbon cycle from ground-based and satellite measurements, the process modelling of the carbon cycle, atmospheric transport modelling, and data assimilation and inversion systems. There will be four main areas of work covering: observations, emission inventories, modelling and inversion systems.

The central questions that CHE will address are:

- What does it take to have a combined bottom-up and top-down estimation system capable of distinguishing the anthropogenic part of the CO₂ budget from the natural fluxes?
- How can we make the first steps towards such a system that can use the high spatial and temporal resolution of satellite observations to monitor anthropogenic emissions at the required time scales?
- And what does it take to transform a research system into a fully operational monitoring capacity?

CHE will support a large community by providing a library of realistic CO₂ simulations from global to city scale to examine the capacity for monitoring future fossil fuel emissions and to adequately dimension space mission requirements.

2.2 Scope of this deliverable

2.2.1 Objectives of this deliverables

D7.5 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 2020¹, 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 projects²”

and applies to data sets that are

“needed to validate the results presented in scientific publications²”.

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 preserved²”.

3.2 CHE Research Data

As per the CHE Description of Action, the products of CHE 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.

All mature data products of CHE will be made publicly available to maximize the uptake by the scientific community. This also answers the requirement of the Call to provide a series of simulation scenarios that could serve to adequately dimension a space mission. It is envisaged to make use of three parallel data portals to ensure full visibility of the datasets. These data portals will be based on the ICOS Carbon portal, the Global Carbon Atlas and the Climate Data Store, which is currently under development by the Copernicus Climate Change Service (C3S). The Technical Annex of the Delegation Agreement between the European Commission and ECMWF regarding C3S explicitly mentions that its Climate Data Store must be designed to allow for the monitoring of climate impacts and climate drivers, including CO₂³. The steps undertaken by CHE towards building a European Support Capacity for Monitoring CO₂ anthropogenic emissions contribute directly to this operational requirement.

Table 1 below presents the envisaged output data sets as per the Description of Action.

¹ http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

² http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm

³ see <http://www.copernicus.eu/main/climate-change> under “technical documents”.

Table 1: CHE Output Datasets

Context	Models	Application	Output Fields
Global	IFS, LMDZ, TM5, TM5+OpenIFS, CCFFDAS	Global scale at spatial resolutions of 10 km or coarser aiming at representing the whole globe with continuous transport models/process models of surface fluxes	CO ₂ fluxes, CO ₂ atmospheric concentrations, other tracers, optimal process parameter values
Regional	CHIMERE, COSMO, FLEXPART, LOTUS-EUROS, WRF-STILT	Regional to continental area at spatial resolution of 5 to 10 km aiming at representing the evolution in limited-area domain with boundary conditions	CO ₂ fluxes, CO ₂ atmospheric concentrations, other tracers
City-scale	CHIMERE, COSMO, WRF-STILT, EULAG	Local targeted areas at spatial resolution of about 1 km or finer aiming at representing detailed emissions	CO ₂ fluxes, CO ₂ atmospheric concentrations, other tracers

3.3 Data Management Plan Questionnaire

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

Table 2: Data Management Plan Questionnaire

<Data set reference and name>	
Data set description	<p><i>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.</i></p> <p><i>Limitations?</i></p> <p><i>Constraints?</i></p>
Standards and metadata	<p><i>Reference to existing suitable standards of the discipline. If these do not exist, an outline on how and what metadata will be created.</i></p> <p><i>Will you generate proper metadata for you data?</i></p>

	<p><i>If yes: how do they look like?</i></p> <p><i>If no: why?</i></p> <p><i>Data format?</i></p> <p><i>Will there be a review process to quality-check the data?</i></p>
<p>Data Sharing</p>	<p><i>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.).</i></p> <p><i>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).</i></p> <p><i>License?</i></p> <p><i>Access URL?</i></p>
<p>Archiving and preservation (including storage and backup)</p>	<p><i>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.</i></p> <p><i>At which Data Center do you want to store your data?</i></p> <p><i>Is there an established workflow for your requested DOI process in place?</i></p> <p><i>According to which standards</i></p>

4 CHE Data Sets

The following sections provide the responses by work packages 1 to 4. Work Packages 5 to 7 do not produce any data sets.

4.1 Work Package 1

OCO-2	
<p>Data set description</p> <p>To be generated:</p> <p>Two new satellite XCO₂ L2 data products from OCO-2 using two complementary approaches.</p> <p>One year of data will be generated (aim: 2015)</p>	<p>Based on UoL algorithm applied to OCO-2 level 2 data, employing machine learning and filtering techniques that select high-quality data that are useful in atmospheric inversions</p> <p>Based on a fast algorithm developed by UoB (Reuter et al., 2017a/b) named FOCAL.</p> <p>These data will be used in the inversions to investigate how source/sink distributions change when satellite data are assimilated.</p>
Standards and metadata	<p>Yes, we will use meta-data. See for example:</p> <p>https://meta.icos-cp.eu/objects/mJbBxyBUWvUxg05G-wlQ2o38</p> <p>Data format: netcdf4</p> <p>Will there be a review process to quality-check the data: yes, users of the data will be asked to provide feedback.</p>
Data Sharing	<p>Data will be shared via existing data repositories like the ICOS-carbon portal</p> <p>https://data.icos-cp.eu/portal/#search, or EUDAT</p>
Archiving and preservation (including storage and backup)	<p>At which Data Center do you want to store your data? EUDAT</p> <p>Is there an established workflow for your requested DOI process in place? Yes</p>
Fossil fuel emissions	
Data set description	<p>These data will be used as input in the inversions.</p>

To be generated: spatiotemporal fossil fuel CO ₂ emission estimates	
Standards and metadata	Meta-data will be used Data format: netcdf4 Will there be a review process to quality-check the data: yes, users of the data will be asked to provide feedback.
Data Sharing	Data will be shared via existing data repositories like the ICOS-carbon portal https://data.icos-cp.eu/portal/#search , or EUDAT
Archiving and preservation (including storage and backup)	At which Data Center do you want to store your data? EUDAT Is there an established workflow for your requested DOI process in place? Yes

Inverted fluxes	
Data set description To be generated: Inversions will produce spatial maps of source sink distributions.	This model output forms the basis for further analysis in WP1
Standards and metadata	The protocol will specify how to use meta-data in describing model output. Data format: netcdf4 Quality checks will be performed
Data Sharing	Data will be shared via existing data repositories like the ICOS-carbon portal https://data.icos-cp.eu/portal/#search , or EUDAT
Archiving and preservation (including storage and backup)	At which Data Center do you want to store your data? EUDAT Is there an established workflow for your requested DOI process in place? Yes

4.2 Work Package 2

IFS CAMS nature runs T1, T2	
Data set description	<p>Two sets (Tier 1 and Tier 2) of high resolution global simulations (9 km resolution) will be performed within WP2 to be used as boundary conditions for the regional simulations. These data sets will also be made available to the external community, e.g. for OSSE studies.</p> <p>The simulations will run for the years 2015 and 2030 (assuming a future emission scenario). The first data set (T1) is based on the current configuration of the operational CAMS CO₂ forecasting system, while T2 will include improved specification of emissions and it will be based on the latest NWP model version. T2 simulations will also be run in ensemble mode at lower resolution to account for uncertainties in the emissions and transport processes. There is also a plan to perform case studies for periods of special interest, e.g. field experiments, where different models could be inter-compared.</p> <p>Note that because this data set is model-based only, it will be affected by systematic errors associated with emissions and model transport.</p>
Standards and metadata	<p>Metadata will be included in the grib header and ECMWF data catalogue.</p> <p>The data will be available in grib format, which can be easily converted to NetCDF using a variety of tools (cdo, ECMWF grib_api software)</p> <p>The simulations will be evaluated using in situ and TCCON observations.</p>
Data Sharing	Data can be made available to partners via ftp and via the Climate Data Store (CDS) to external users.
Archiving and preservation (including storage and backup)	The data will be archived in the CDS.
Global anthropogenic emissions EDGAR v4.2	
Data set description	Global emissions of CO ₂ , CO and other trace gases used as input for Tier 1 global CO ₂ simulation.

	The data is available at 0.1°x 0.1° horizontal resolution as annual means for multiple years up to 2010.
Standards and metadata	The product will be available in netCDF format compliant with CF conventions. Metadata information is included in the global attributes of the netcdf file.
Data Sharing	The data set is publicly available. Access URL: http://edgar.jrc.ec.europa.eu
Archiving and preservation (including storage and backup)	The data set is archived at the European Joint Research Center JRC

Global anthropogenic emissions EDGAR v4.3	
Data set description	Global emissions of CO ₂ , CO and other trace gases used as input for Tier 2 global CO ₂ simulation for present-day and future emissions. The data will be available at 0.1°x 0.1° horizontal resolution as annual means for 2015 and 2030 and information for temporal disaggregation to hourly fields.
Standards and metadata	The product will be available in netCDF format compliant with CF conventions. Metadata information will be included in the global attributes of the netcdf file.
Data Sharing	The data set will be made publicly available as part of Deliverable D2.3 of CHE by Dec 2018.
Archiving and preservation (including storage and backup)	The data set will be archived at the European Joint Research Center JRC

European anthropogenic emissions TNO/CAMS81	
Data set description	European emissions of CO ₂ , CO and other trace gases used as input for European simulations in WP2 of CHE. The data will be available at 1/16°x 1/8° horizontal resolution as annual means for 2015 and 2030 and information for temporal disaggregation to hourly fields.
Standards and metadata	The product will be available in netCDF format compliant with CF conventions.

	Metadata information will be included in the global attributes of the netcdf file.
Data Sharing	The data set will be made publicly available as part of Deliverable D2.3 of CHE by Dec 2018. It will be accessible through the CAMS catalogue.
Archiving and preservation (including storage and backup)	The data set will be archived at ECWMF

European biosphere fluxes from VPRM	
Data set description	European biosphere-atmosphere exchange fluxes of CO ₂ from the Vegetation Photosynthesis and Respiration Model (VPRM) separately for gross photosynthetic production and respiration. The VPRM parameters will be calibrated against CO ₂ flux measurements from the FLUXNET network. The data will be available at approx. 5 km x 5 km horizontal resolution for the year 2015.
Standards and metadata	The product will be available in netCDF format compliant with CF conventions. Metadata information is included in the global attributes of the netcdf file.
Data Sharing	The data set will be made publicly available as part of Deliverable D2.3 of CHE.
Archiving and preservation (including storage and backup)	The data set is archived at ECMWF.

European and regional CO₂ simulations	
Data set description	European simulations of CO ₂ and CO performed with three regional models, COSMO-GHG, WRF-GHG and LOTOS-EUROS. LOTOS-EUROS will also provide output for reactive trace gases and aerosols. COSMO-GHG and WRF-GHG will also provide output of meteorological fields. The data will be available at approximately 5 km x 5 km horizontal and hourly temporal

	<p>resolution for the years 2015 and 2030 (with meteorology from 2015).</p> <p>Regional simulations will be conducted for two different domains, Berlin and Beijing.</p> <p>Models for Berlin: COSMO-GHG, WRF-CHEM, LOTOS-EUROS</p> <p>Models for Beijing: LOTOS-EUROS, WRF-CHEM</p> <p>Output will be at approximately 1 km x 1 km horizontal resolution except for LOTOS-EUROS (approx. 2 km x 2 km).</p> <p>The output from the European and Berlin simulations will be used to generate the synthetic Level-2 satellite observations (Deliverable D2.5).</p>
Standards and metadata	<p>The product will be available as a collection of netCDF files compliant with CF conventions. Metadata information will be included in the global attributes of the netcdf files.</p> <p>Detailed information on the models and simulations is given in Deliverable D2.1 of CHE.</p>
Data Sharing	<p>The model output will be made publicly available as Deliverable D2.4 of the project by July 2018.</p>
Archiving and preservation (including storage and backup)	<p>The data set will be archived at ECMWF</p>

Synthetic Level-2 satellite observations with realistic uncertainties	
Data set description	<p>Based on the model simulations over Europe and Berlin, synthetic Level-2 satellite observations will be generated for a suite of hypothetical satellite orbits accounting for both random and systematic measurement errors.</p> <p>The synthetic satellite data will mimic observations from a constellation of Sentinel-7 satellites with 2 km x 2 km horizontal resolution and a swath of approx. 250 km.</p>

	Two different uncertainty data sets will be generated, one applying a simple error parameterization (for complete year 2015), and a second one based on full retrieval simulations (for selected periods in 2015).
Standards and metadata	The product will be available in netCDF format following the standards for satellite products defined in ESA's GHG-CCI project. For details see Product Specification Document PSDv3 at http://esa-ghg-cci.org
Data Sharing	The data set will be made publicly available as part of Deliverable D2.5 of CHE.
Archiving and preservation (including storage and backup)	The data set is archived at ECMWF.

4.3 Work Package 3

WP3 follows a slightly different version of the questionnaire as details will emerge at a later stage.

Biospheric Net Ecosystem Exchange fluxes	MPI- Jena – Task 3.1
Available: High resolution biospheric NEE fluxes for Europe	Data assimilation of mean seasonal cycles of European flux towers (cross validated), remote sensing and reanalysis data allowed to calculate the NEE fluxes for Europe at 0.1degx0.1deg spatial and hourly temporal resolution. This dataset does not include the direct land use change fluxes.
Available: Low resolution global NEE fluxes	Global NEE fluxes, excluding direct land use change fluxes. This is coarser resolution spatially 0.5degx0.5deg but half-hourly temporally. (Bodesheim et al., 2018, ESSD); doi.org/10.5194/essd-2017-130
Generated: Enhanced European biospheric NEE fluxes	Work in progress for Europe to obtain improved NEE fluxes (making use of recent European flux and eddy covariance data of D. Papale).
Enhanced global NEE fluxes	Work in progress for the global NEE flux to incorporate subdaily weather with hourly ERA-5 reanalysis data.
Generated: Metadata/ Data sharing/ Archiving & preservation	Metadata will be generated. The dataset will be peer reviewed by the scientific journal in which they are published.

	Datasets will be published with DOI number, preferably in the Earth System Science Data journal.
--	--

EDGAR CO2 emission gridmaps	JRC/ ECMWF – Task 3.2
Available: EDGARv4.3.2 basis emission gridmaps for GHG (1970-2012)	Bottom-up global CO ₂ , CH ₄ and N ₂ O emission gridmaps of anthropogenic activities, excluding large scale biomass burning and Savannah fires and Land use, land-use change and forestry. These are annual, sector-specific data for 2012 and the years before till 1970, based on IEA statistics and other international statistics (USGS) and IPCC (2006) emission factors. Gridding was done on the basis of over 300 different spatial proxy datasets (e.g. point sources of EPRT, Open streets map, ...) For 2010 also monthly emission gridmaps are produced. Janssens-Maenhout et al., 2018, ESSD) or doi.org/10.5194/essd-2017-79 and with dataset DOI: https://data.europa.eu/doi/10.2904/JRC_DATASET_EDGAR .
Available: EDGARv.4.3.2FT2016 CO2 emissions dataset (1990-2016)	Global CO ₂ emission time, calculated per country and activity of fossil fuel use and industrial processes (cement production, carbonate use of limestone and dolomite, non-energy use of fuels and other combustion, chemical and metal processes, solvents, agricultural liming and urea, waste and fossil fuel fires). Excluded are: short-cycle biomass burning (such as agricultural waste burning), large-scale biomass burning (such as forest fires) and carbon emissions/removals of land-use, land-use change and forestry (LULUCF). (Janssens-Maenhout et al., 2017, EUR 28766 EN report; http://edgar.jrc.ec.europa.eu/overview.php?v=CO2andGHG1970-2016)
Generated: Updated EDGARv4.3.2 emission gridmaps of CO2 with the fast track dataset EDGARv4.3.2FT2016	Work in progress for updating the EDGARv4.3.2 gridmaps for 2012 to 2016 based on EDGARv4.3.2FT2017 by country- and sector-specific scaling factors for each grid cell.
Generated: EDGARv4.3.2 emission uncertainty gridmaps (ECMWF)	Work started for generating uncertainties of the CO ₂ timeseries per country and sector. This will be translated into uncertainties of the gridmaps with their spatial and temporal distribution assumptions on representativeness. ECMWF can then derive a covariance matrix for the different sectors.
Metadata/ Data sharing/ Archiving & preservation	Metadata are generated. The dataset will be peer reviewed by the scientific journal in which they are published.

	Datasets will be published with DOI number, preferably in the Earth System Science Data journal and also downloadable from the edgar.jrc.ec.europa.eu website.
--	--

4.4 Work Package 4

D4.2 Database of high-resolution scenarios of CO₂ and CO emissions	
Data set description	<p>Database of high-resolution scenarios of CO₂ and CO emissions associated with anthropogenic activities in Europe over a full year including associated uncertainty statistics and documentation.</p> <p>Generated by TNO from the inventory generated in WP2. This one will be further downscaled to 1 x 1 km² for region EU28+NOR+CHE. A family of 10 grids will be constructed by using the uncertainties associated with the emissions of CO₂ and CO (activity data, emission factors, spatial distribution proxies and temporal emission timing) in a Monte Carlo simulation.</p> <p>The dataset will be used by the project (T4.2 and T4.3) and can be also useful for CO₂ and CO transport simulation outside the project. Other products may exist at this scale and for this whole domain, but with less constraint from detailed activity data. It will still be possible to refine it further after the project.</p>
Standards and metadata	This will be decided in the first year of the project.
Data Sharing	This will be decided in the first year of the project.
Archiving and preservation (including storage and backup)	This will be decided in the first year of the project.

5 Conclusion

This initial Data Management Plan has identified a number of data sets for each of the work packages 1 to 4, 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)	07/03/2018	Initial Version for Internal Review
1.0	Daniel Thiemert (ECMWF)	21/03/2018	Final version after internal review

Internal Review History

Internal Reviewers	Date	Comments
Laure Brooker (ADS SAS)	15/03/2018	Approved with comments
Denis Siméoni (TAS)	15/03/2018	Approved with comments

Estimated Effort Contribution per Partner

Partner	Effort
ECMWF	0.1
Total	0.1

This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.