OPEN DATA RESEARCH PILOT AND DATA MANAGEMENT PLAN

11/30/2018	Deliverable 6.5

Version 2

Dissemination level:
Public

Due date of delivery: M5 (31 March 2017)

Actual date of delivery: revised version 30.11.2018

(for actual data repository selection)

Lead beneficiary: Finnish Meteorological Institute

Lead authors: Adriaan Perrels

Version table		
Date	**Name, Party**	**Description**
09.10.2017	Adriaan Perrels	Chapters 1 and 2
08.01.2018	Adriaan Perrels	Chapters 3-5; updates Ch.1 & 2
11.01.2018	Adriaan Perrels	Editing comments
27.11.2018	Jörg Cortekar	Chapter 6 + Annex 2

Internal review table		
Date	**Name, Party**	**Description**
03.11.2017	Maija-Liisa Honkola, Niku Kivekäs, Jaana Palmunoksa (FMI)	Comments; and review meeting in relation to FMI's policy
11.01.2018	Maija-Liisa Honkola, Niku Kivekäs, Jaana Palmunoksa (FMI)	Comments; and review meeting in relation to FMI's policy
17.01.2018	Katja Lamich (HZG); Peter Stegmaier (UT)	Few comments and edits; Approval by email
21.01.2018	Ines Väittinen (ENoLL); Raffaele Giordano (CNR-IRSA)	Approval by email; Others have approved implicitly by not responding during the commenting period

Contributors (Consortium Party, person):

FMI	Adriaan Perrels
HZG-GERICS	Jörg Cortekar, Katja Lamich
CNR-IRSA	
Acclimatise	
CMCC	
U_TUM	
UT	Peter Stegmaier
JR	
ENoLL	

Glossary of Terms

Term	Explanation
climate service	The transformation of climate related data – often together with other relevant information - in to customized information products, offered as such or embedded in consultancy and/or education [condensed version of European Roadmap definition]
climate service: seasonal forecast	A prediction of weather tendencies (often expressed as probabilistic deviations from long term averages typical for the considered period and area) stretching from approx. 1 month to 6 months or more.
climate service: long term forecast	A prediction of climate conditions for a certain area and for typical time units (diurnal to annual) referring to decadal or multi-decadal averages several to many decades ahead
Informed consent	A written statement signed by an interviewee, which states that the interviewee has been sufficiently informed in advance about the nature and purpose of the project and the interviews, and specifies the rights of the interviewee to correct or withdraw statements up to approval of the written interview.
Interview	A question and answer sequence in which a relevant stakeholder – constituting one or more persons, provides answers to designated EU-MACS representatives. Interviews are conducted orally (i.e. not as 'email interview') in most cases as virtual meeting and sometimes face-to-face. Interviews are recorded. The written representation of the interview is sent to the interviewees for inspection and approval.
Meta-data	Description of a data file in terms of its contents, origin, ownership, allowed level of openness, etc.
Stakeholder registry	Contact information of all engaged stakeholders in the various WPs over the course of the project
Survey	Questionnaire based collection of responses of individuals or individuals representing an organisation; part of the responses is coded in numerical scores for ease of comparison and analysis; the rest is text response
Workshop	Gathering of invited experts and other stakeholders to discuss a coherent set of topics relevant for a work package, and to jointly analyse questions and issues with the aim to provide the work package with input helpful for generating solutions and deliverables
Web based exploration	Structured and recurrent sets of inter-related questions answered by invited experts, which help to devise tools and recommendations in WP2-WP4

List of Abbreviations

AEC   Advisory Expert Committee

CA   Consortium Agreement

DoA   Description of Actions

Dn.m   Deliverable no. m from WPn

EB   Executive Board

GA   General Assembly

GAG   Grant Agreement

PMH   Project Management Handbook

WP   Work Package (WPn – Work Package no. n (1 – 7))

List of Contents

Contents

1. INTRODUCTION

The study

To support further product development and effective widespread uptake of climate services, as a means to boost mitigation of and adaptation to climate change as well as capabilities to cope with climate variability, the European Commission has included several actions in its current research programme Horizon 2020 (H2020). Essentially these actions follow from the logic to implement the European Research and Innovation Roadmap for Climate Services (cf. European Commission, 2015)

EU-MACS and its twin project MARCO deal with analysis of the climate services market. In addition demonstration calls were launched on the added value of climate services for supposedly high value added sectors with hitherto little uptake of climate services (SC5-01-2016-2017), while other actions focus more on networking activities interlinking to better connect relevant players, such as the Coordination and Support Action (SC5-05b-2015) called Climateurope. In addition the ERANET for climate services (ERA4CS) is a programme that contains both testing of particular types of climate services in selected sectors and exploration of suitable climate service types for selected sectors.

An extremely important sub-programme in H2020 is the COPERNICUS Climate Change Service (C3S) programme, which aims to generate a very comprehensive coherent and quality assured climate data set meant to support mitigation and adaptation planning, implementation and monitoring. In due course, also coping capabilities of (current) climate variability are addressed.

In this framing, EU-MACS – European Market for Climate Services – will analyse market structures and drivers, obstacles and opportunities from scientific, technical, legal, ethical, governance and socioeconomic vantage points. The analysis is grounded in economics and social sciences, embedding innovation theories on how service markets with public and private features can develop, and how innovations may succeed.

The scope and remit of this report

The project EU-MACS does not produce new climate services itself, neither is it engaged in the processing of climate data or impact data as such. Instead, EU-MACS collects different types of (meta)information regarding the 'field of climate services' and the 'climate services market' (the market being a subset of the field). Apart from literature study, the typical ways of collecting information are surveys, interviews, workshops and internet based interaction formats. In addition, the project assembles stakeholder lists based on snowballing among the consortium partners and some key contacts, contacts during outreach activities, and stakeholder feedback via the website and twitter. Furthermore, purpose designed interaction formats, such as Living Lab embedded workshops and web based explorations on climate service design preferences, are used.

Given the nature of these data not only maximization of access to research data is an issue, but also data protection considerations. As regards data protection, two types of potential limitations on openness play a role, being: (1) personal contact information, and (2) attribution of characteristics or opinions to participating organisations or their representatives. The implication is that for a part of the collected information only meta-information can be provided freely, whereas the actual research files (e.g. interviews) can only be made available under stringent conditions (such as 'view only') and only for the academic purpose of peer review and duplication.

This report is meant to give the consortium partners guidance in how to collect, store, share and describe data with the aim to promote the objective of open access to research data, under the conditions of the

Grant Agreement and the Consortium Agreement, while abiding to the EU regulation on data protection, (EU) 2016/679 [1]). In addition to providing guidelines, the project EU-MACS pursues also an open data pilot in order to tangibly promote the access to research data via EUDAT[2].

This Deliverable is a living document, which may be revised during the project. The first version will be formally submitted as Deliverable to the EC Project Portal.

Structure of the report

Chapter 2 briefly reviews the principal guidelines regarding open data access and data protection for this project. Chapter 3 gives a detailed account of the collected and processed information. Chapter 4 presents the guidelines for properly assembling and storing the project's datasets. Chapter 5 describes the implementation steps planned in order to effectively support open data access by means of EUDAT, while honouring applicable data protection principles. Annex 1 – provides an insight in the description of the metadata.

---

[1] http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.119.01.0001.01.ENG&toc=OJ:L:2016:119:TOC
[2] https://www.eudat.eu/

2. PRINCIPAL GUIDELINES

Open data

- Information generated by the project EU-MACS, either as intermediate or as final product, should as much as possible be made available for reuse by third parties;

- To make availability meaningful for third parties the information should be:
  - Easy to find and identify
  - Described in accompanying meta-information in order to adequately facilitate the judgement of a potential user regarding the usefulness of the considered data; this should include information on the origin, observation technology, and post-processing of the information
  - Clearly organised in datasets with adequate use of object/variable labels and definitions
  - Submitted to a commonly agreed quality control prior to definitive submission
  - Accurate (at the time of its collection)

- Information planned to be made accessible for third parties should be reviewed with respect to data protection regulation *prior to its submission* to facilities meant for retrieval by third parties and should be sufficiently anonymised or conversely sections of the considered datasets should not be submitted to such facilities; in the latter case meta-information disclosing that such information exists and is in possession of one or more consortium partners can be submitted to such facilities;

- All consortium parties should inform the consortium coordinator about the data collected and processed in the EU-MACS project by using a preformatted table provided by the coordinator; this meta-information table will be made available to all consortium partners; the coordinator should see to it that the meta-information table is regularly updated during the project;

- If there are no objections from the point of view of the Grant Agreement and of the Consortium Agreement, and if the considered data are not subject to sharing limitations due to Regulation EU 2016/679, collected and processed data are supposed to be made available for third parties by means of the Open Data Pilot (see below). An exception can be made for those data of which a majority of the consortium is of the opinion that there is no value added for third parties or if the manager of the data repository service is not interested in the submission;

- For data that cannot be shared, meta-information will be made available for third parties, as much as possible;

- After submission of data for the Open Data Pilot, and after quality control has shown that all conditions are met, consortium parties are not obliged to provide further support free of charge with respect to possible maintenance of the submitted data. An exception to this principle applies when preventable inadequacies are detected after submission and quality control.

- Files to be made accessible for third parties will be declared under Creative Commons Licensing system; the recommended options are[3]:
  - Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)
  - Attribution 4.0 International (CC BY 4.0)

---

[3] For definitions see: CC BY 4.0: https://creativecommons.org/licenses/by/4.0/;
CC BY-SA 4.0: https://creativecommons.org/licenses/by-sa/4.0/

Open data pilot

- In the Open Data Pilot data generated in the EU-MACS project will be uploaded to the EUDAT facility (https://www.eudat.eu/) and/or to the Finnish Social Science Data Archive (FSD; http://www.fsd.uta.fi/en/data/index.html) depending on the nature of the files, while an effort will be made to advertise the availability of these data to third parties;

- It is the ambition of the Open Data Pilot to upload all eligible data generated in the EU-MACS project, where eligibility refers to limitations following from the Grant Agreement, the Consortium Agreement, or the Regulation EU 2016/679;

- In order to be able to upload eligible data in a systemized way a set of preparatory actions should be carried out related to selection, quality assurance, data protection review, meta-information, consent from third parties, and formatting of datasets;

- In order to formulate guidelines for selection, quality assurance, data protection review and meta-information the Executive Board will nominate an Open Data Pilot (ODP) committee; the ODP committee will oversee the adequate implementation of the Open Data Pilot, with support from the project's information integrity officer;

- The ODP committee will manage a table of potentially eligible data in the common workspace of the EU-MACS project, and ensure it is regularly updated regarding the declaration of eligible datasets;

- The ODP committee can delegate the actual technical preparations and uploading of datasets to the designated partner(s);

3. TYPES OF DATASETS IN EU-MACS

Survey data

Up to now, one survey has been held by means of an online questionnaire. The questionnaire counts 42 questions. Some questions are multistage and others allow for more than one entry. The questionnaire is addressing both climate service providers and users, for whom partly common and partly distinct questions are presented. Apart from respondent characterization all questions have to do with awareness of, experiences with, and pre-conditions for climate services and their development. The online questionnaire starts with an explanation of the purpose of the questionnaire and the characteristics of the EU-MACS project.

The response types include numerical information, predetermined choice options, and text.

The current number of valid questionnaires is about 150, but we aim to raise this to 200 at least.

The questionnaire results have been collected in an EXCEL file. The respondents are anonymous, and only needed to indicate type of organisation and country from which they operate.

In case of future surveys with identifiable respondents, the publishable file contains a respondent sequence number, while the actual respondents' identities are stored separately.

More information on the survey and its results can be found in Deliverable 1.1, and to some extent in Deliverable 1.2. (http://eu-macs.eu/outputs/# >> Reports)

Interviews

Various consortium partners have conducted interviews, mostly by phone or video connection, sometimes face-to-face. Most interviews are of a semi-structured nature. Interviewees are informed in advance (by email) about the purpose and context of the interview, and are requested to sign and return an informed consent form (scanned pdf). Interview summaries are sent to the interviewees for comments and approval. Interviews have also been recorded (voice).

Neither the recordings nor the individual interview summaries are public research material, but may be made accessible for academic peer review purposes. Syntheses of clusters of interviews can be considered for publication. Also the basic sets of questions for the semi-structured interviews can be published.'

Synthesizing interpretations of the interviews are presented in various WP1 deliverables.

Workshops

The Work Packages 2, 3, and 4 (hereafter WP2, WP3, WP4) organize workshops with representatives from the focus sectors, finance (WP2), tourism (WP3) and urban planning (WP4). The programme and activity descriptions (assignments) of these workshops are available as internal documents. The same applies to the internal detailed reports of the workshops. Summaries (impressions) of the workshops are published in the project's communication channels, such as the Newsletter. The programme, activity descriptions (assignments) and the published summaries are eligible for publication and submission of copies to third parties. As with the interviews the detailed internal reports are only available for academic peer review purposes. Some highlights of the results and their interpretation will be presented in WP2, WP3 and WP4 deliverables.

Web based interactive explorations (bidding games)

For WP2 and WP3 internet based serial questionnaires are used to elicit information on inclinations and tendencies of different types of users with respect to climate service product features and provision modes. The anonymized results will be made available for 3rd party use, provided the respondent numbers are large enough. Summaries of the results and their interpretation will be made presented in WP2 and WP3 deliverables.

Stakeholder registry

The stakeholder registry combines the collected contact information from interviews, workshops, encounters in events, and other contacts, with the exception of the internet based survey responders. The stakeholder registry is only for internal project purposes and is not supposed to be accessible for people outside the EU-MACS consortium. A protected version of the stakeholder registry is available in the protected common work space of the project.

The stakeholder registry will be evicted after the conclusion of the project.

4. PRACTICAL GUIDELINES FOR ASSEMBLY AND STORAGE OF DATASETS

Survey data

Data format

Even though original source data need to be kept, in case errors are found later, the target dataset discussed here is a database format which enables easy use, both as such (e.g. as EXCEL ® or ACCESS ®) and as source file for another data handling programme (SAS®, MATLAB®, etc.), and enables also meta-data specification.

Meta data can in principle contain the following variables:

- Creation date / period
- Creator (name; email)
- Owner(s) (organisation(s) + prime contact)
- Original purpose; project
- Level of openness; Allowed types of re-use (incl. license version of Creative Commons 4.0)
- Number of variables
- Variable list (incl. applied units, resolution, observation method, post-processing steps)
- Observation or measurement period
- Number of observations
- Number of respondents (if not same as number of observations)
- Anonymization method
- Kind of quality check (none; visual; simple checks on allowable ranges and plausible scores; …)
- Handling of missing data (none; …)

Storage

- During the project: in the common workspace
- In the Finnish Social Science Data Archive (FSD) and linked from the EUDAT service at the latest starting at the end of the project

Interviews

Data format

- Audio and audiovisual files of recorded interviews
- Transcripts of interviews (text file)

Meta data

Interview datasets are in principle not public, but their archiving should nevertheless allow for tractability in case of later needs for peer review.

For each audio- and audiovisual file a separate small (text) document has to be made, containing the metadata of the file of interest. In additional a summary meta-data overview for all (audio)visual file and transcripts belong to the same interview collection can be made.

- Creation date / period

- Creator (name; email)
- Owner(s) (organisation(s) + prime contact)
- Purpose; project
- Level of openness; Allowed types of re-use (see Creative Commons 4.0) – if any
- Interviewee(s) – if allowed to disclose
- Length of audio-recording in minutes
- Length of transcript (number of words)
- Used questionnaire (if any)
- Informed consent form

Storage

- During the project: transcript text files in the (protected) common workspace; (audio)visual files offline
- Files should be kept available for at least 5 years after the end of the project in case of peer review requests

Workshops

Data format

- Audio and audiovisual files of recorded interviews
- Transcripts of interviews (text file)

Meta data

Workshop datasets are in principle not public, but their archiving should nevertheless allow for tractability in case of later needs for peer review. See interviews for suggested meta data contents. For workshop files the precise relevant contents is case specific.

Storage

- During the project: in the (protected) common workspace;
- Files should be kept available for at least 5 years after the end of the project in case of peer review requests

Web based interactive explorations (bidding games)

Data format

Easy accessible and usable format, i.e. EXCEL®, ACCESS®, or comparable

Meta data

Same guidelines as for surveys.

Storage

Same guidelines as for surveys.

Stakeholder registry

Data format

Same guidelines as for surveys.

Meta data

None, except indication on openness, i.e. only for project internal use.

Storage

- During the project: in the protected common workspace;
- After the project: dataset should be discarded

5. IMPLEMENTATION STEPS

Organising data sets and their storage during the project

- WP leaders in cooperation with the coordinator and the information integrity officer will review that the relevant datasets are properly uploaded to the common workspace. Where necessary protected subsections and/or keyword protection by file will be applied.
- The established situation will be reviewed by the information integrity officer
- First, the situation for already existing files will be reviewed in January and February 2018. Subsequently the same procedure will applied to newly created files.
- The evolution in the stored datasets will be monitored by the information integrity officer.

Archiving datasets after the project has ended

- Designated eligible files will be transferred to the Finnish Social Science Data Archive (FSD) and announced to EUDAT by the coordinator.
- Designation of eligible files will be prepared by the Open Data Pilot (ODP) committee and approved by the General Assembly of the project. These decisions should be made before the termination of the project (31.12.2018).
- The agreements between FMI, FSD, and EUDAT will be reviewed by the information integrity officer.

6. SUMMARY OF DATA GENERATED DURING THE PROJECT

Brief description of generated data by type of information

In the various EU-MACS work packages, information was collected and evaluated on the basis of a web survey, interviews and workshops.

Web-Surveys

The web survey aimed to collect information from providers and users of climate services on barriers related to the development and uptake / use of climate services in political, economic, social, technological, ethical and legal / regulatory respects. It also addressed topics such as resourcing and quality issues.

The survey was organised in two branches; one provider-specific and one user-specific branch. Each branch itself was organised in three loops. The survey consists of 121 questions in total including open and different forms of closed questions (e.g. multiple choice, rating on 5-Point Likert Scales etc). Participants were routed through the survey based on their responses. None of the participants had to go through all questions.

169 participants took part in the survey, of which 109 considered themselves as a provider (which also includes intermediary providers or so-called purveyors) and 60 as a user. None of the participants had to provide any kind of personalized data. The results of survey are available as an excel file.

In WP2 and WP5 web surveys failed to attract notable numbers of respondents. The modest material from these surveys has been used to provide further support to certain statements and insights In D2.1 and D5.2, but there is no point in uploading these files to a repository.

Interviews

Interviews were mainly conducted in the three case study sectors finance, tourism and urban planning with both, providers and users of climate services. Additional interviews were conducted in WP1 related to quality assurance with climate services providers only and in WP5 to receive additional views on the conclusions drawn.

In many cases the interviews provided a forum to discuss also confidential details, only to be shared in a safe environment and with the knowledge that the shared information will only be used in an aggregated manner. Furthermore, the informed consent form for the interviewees promises confidentiality of the interview. Nevertheless, all interviews have been either recorded (with approved transcripts created afterwards) or trans-scripted.

The information collected during interviews are stored as recorded files and / or (approved) trans-scripted text documents. In only very few cases the documentation only contains of some notes taken during the interviews.

Workshops

Workshops were conducted in two out of three case study sectors, i.e. tourism and urban planning. Participants represented mostly users of climate services but also few climate services providers. Main parts of the workshops were dedicated to Constructive Technology Assessment (CTA). The CTA part of the

workshop offered a set of specific viewpoints to consider scenarios of using climate services, while at the same time giving ample space for discussion of aspects stakeholders find important. Another element of the workshops was a discussion on typical business cases for climate services use in the two sectors.

Even though, proper documentation of the workshops has been done, workshops did not generate material which merits to make generally available that goes beyond the processed products and syntheses as can be found in the respective deliverables of WP3 and WP4.

Proposition for data to submitted to open data repository

Not all information collected in EU-MACS through a variety of interaction formats can be made accessible for the reasons given above. The original files are stored for five years by the organisations that collected the information. Access can be granted upon request to the owner of the information under conditions in accordance with the Consortium Agreement. However, the original files will under no circumstances be made available for commercial purposes.

While the survey does not contain any personalized or confidential information, the interview records and transcripts do. The same applies to a proper workshop documentation. Therefore, the following information on the various interactions and information will be made available:

- The survey will be made available both in terms of a file providing some meta-information as well as the excel file providing all responses from all 169 participants.
- For interviews and workshops, only meta-information will be provided.

The meta-information follow the guidelines and principles presented in chapter 4 and are made available as pdf files (see also annex 2), which will be uploaded to the Finnish Social Science Data Archive and announced to EUDAT by the coordinator.

ANNEX 1. TEMPLATE FOR DATA DECLARATION TABLE

1.Basis and nature of data; subject(s)	2.Size indicators	3.time / period	4.geographical distinctions?	5.source	6.openness	7.data protection issues
Interview (semi-structures; open ; ….)	N interviews; N questions;	period interviews were held	country; region; zone; other;…	collected by partner in EU-MACS; collected by 3rd party in EU-MACS; acquired	dataset available for EU-MACS partners without restrictions; dataset only conditionally available; dataset not available for others	data contain personal or personalizable information*; data contain confidential information of actors
Questionnaire (multiple choice; rating on prefixed scale; open answers; other …)	N respondents; N questior	period survey was held	country; region; zone; other;…	collected by partner in EU-MACS; collected by 3rd party in EU-MACS; acquired	dataset available for EU-MACS partners without restrictions; dataset only conditionally available; dataset not available for others	data contain personal or personalizable information*; data contain confidential information of actors
Economic survey (budgets; turnover; N transactions; N employees; N patents;…)	N actors ; N variables	years or periods for which data are available	country; region; zone; other;…	collected by partner in EU-MACS; collected by 3rd party in EU-MACS; acquired	dataset available for EU-MACS partners without restrictions; dataset only conditionally available; dataset not available for others	data contain personal or personalizable information*; data contain confidential information of actors
Registry data	N actors ; N variables	years or periods for which data are available	country; region; zone; other;…	collected by partner in EU-MACS; collected by 3rd party in EU-MACS; acquired	dataset available for EU-MACS partners without restrictions; dataset only conditionally available; dataset not available for others	data contain personal or personalizable information*; data contain confidential information of actors
Obervations/Measurements (climatic variables; land use ; population;…)	N interviews; N questions;	years or periods for which data are available	country; region; zone; other;…	collected by partner in EU-MACS; collected by 3rd party in EU-MACS; acquired	dataset available for EU-MACS partners without restrictions; dataset only conditionally available; dataset not available for others	data contain personal or personalizable information*; data contain confidential information of actors
Other data sets (……)	N interviews; N questions;	years or periods for which data are available	country; region; zone; other;…	collected by partner in EU-MACS; collected by 3rd party in EU-MACS; acquired	dataset available for EU-MACS partners without restrictions; dataset only conditionally	data contain personal or personalizable information*; data contain confidential information of actors
						*) such as:
						name; address; function; email; (high resolution) GIS coordinates;

ANNEX 2. META-INFORMATION TEMPLATES APPLIED IN EU-MACS

TABLE 1: META-DATA INFOMRATION TEMPLATE FOR SURVEYS

Creation date / period	(Start)
	(End)
Creator	(Name)
	(Email)
Owner(s) (organisation(s) + prime contact)	(Name)
	(Email)
	(Organisation)
Original purpose	Project EU-MACS, funded by the European Commission, Grant Agreement No. 730500
Level of openness	
Anonymization method	
Number of variables	
Variable list	
Observation or measurement period	(Start)
	(End)
Number of observations	
Number of respondents (if not same as number of observations)	
Kind of quality check	
Handling of missing data	

TABLE 2: META-INFORMATION TEMPLATE FOR INTERVIEWS

Creation date / period	(Start)
	(End)
Creator	(Name)
	(Email)
Owner(s) (organisation(s) + prime contact)	(Name)
	(Email)
	(Organisation)
Original purpose	Project EU-MACS, funded by the European Commission, Grant Agreement No. 730500
Level of openness	Allowed types of re-use – if any
Interviewee(s)	If allowed to disclose (maybe generalized info such as Head of … or bla bla bla expert, which urban authority etc.)
Length of audio-recording in minutes	If available
Length of transcript (# of words)	If available
Used questionnaire (if any)	
Informed consent form	