Data

Management

Plan

Deliverable D9.9

climatestorylines.eu
RECEIPT_eu

RECEIPT project information	
Project full title	Remote Climate Effects and their Impact on European sustainability, Policy and Trade
Project acronym	RECEIPT
Grant agreement number	820712
Start date and duration	01/09/2019, 48 months
Website	www.climatestorylines.eu

Deliverable information	
Work package number	WP 9
Work package title	Communication and Visualisation
Deliverable number	D9.9
Deliverable title	Data Management Plan
Description	This report summarizes RECEIPT's Data Management Plan. The plan incorporates measures to mitigate ethics and privacy concerns, in particular those related to personal data and data coming from third countries.
Lead beneficiary	Deltares
Author (s)	Maarten van Meersbergen, Dana Stuparu, M. Umit Taner

Dissemination level of the document

X	PU	Public
	PP	Restricted to other programme participants
	RE	Restricted to a group specified by the consortium
	CO	Confidential, only for members of the consortium

Versioning and contribution history			
Revision	Date	Modification	Author
V.0.0.1	10 Feb 2020	Initial version of DMP	Maarten van Meersbergen, Dana Stuparu, M. Umit Taner, Bart van den Hurk
V.1.0.0	25 Feb 2020	Review	Bart van den Hurk

Approvals			
	Name	Organisation	Date
Coordinator	Bart van den Hurk	Deltares	25 Feb 2020
WP Leaders			

Contents

Executive Summary

A Data Management Plan (DMP) is especially helpful in a large group of independent institutes like the RECEIPT consortium because it makes it possible to synchronize data streams. It forces the group from an early date to think about sharing and storing data and about making it findable and interoperable for (in the first place) the group members of the consortium institutes themselves.

The data gathered and generated by the RECEIPT project will be primarily stored in Zenodo, a general-purpose open-access repository developed under the European OpenAIRE program and operated by CERN. Within the consortium we make the agreement that (re)distributing data can be done in two ways. Data can either be directly opened to the public in the Zenodo repository or be placed there with an embargo, waiting on a specified publication event.

Once all reports have been written, the full datasets gathered and analysis generated within the scope of the project, data will be opened for further use by any interested party. This allows for further re-use of datasets and knowledge. It also allows for verification of the RECEIPT projects results.

Some of the data gathered in the RECEIPT project will never be directly opened due to security and privacy reasons. As of writing this document, we expect this to only apply to stakeholder contact information and interview transcript data.

This is the first version of the DMP for the RECEIPT project. It is intended to be a living document, which will be updated regularly when new insights are gained.

1. Introduction

The goal for the RECEIPT project is to shed light on the causal relations of climate change impacts elsewhere in the world, and how those affect Europe. Data will be gathered and generated within the scope of this project in support of this goal. Furthermore, storylines will be constructed to highlight these causal relations and their variations, consistent with Shared Socioeconomic Pathways (SSPs) and Climate Scenarios. To make these storylines as convincing, realistic and appealing as possible, additional data will be required for their construction.

In short, data within the RECEIPT project shall be gathered and generated to:

- Assess global climate *hotspots* for investigation in terms of their potential impacts on Europe
- Assess supply chain, financial, agricultural, migration/displacement and coastal infrastructure risks based on selected climate hotspots and associated *connections* to Europe
- Model the climate change impacts on selected global climate hotspots
- Model the direct and indirect connections on Europe of said climate change impacts.
- Assess the *variability* of impact on connections based on "Paris Scenarios" and "Shared Socio-economic Pathways"
- *Analyse risks* to European countries, institutions and individuals associated with these connections
- Develop *storylines* based on climate hotspots, connections, variability and risk analyses
- Document the data and assumptions on which the storylines are based.
- Visualize the finalized storylines

To achieve these goals, the RECEIPT project shall re-use as much data as is possible from its consortium members and societal partners, as well as any pertinent data to be found with permissive licences. Where additional data is needed to support the conceptual storylines, the consortium members will generate that data based on scientific models. It is conceptually possible that data will be gathered or generated through other means like questionnaires. Once the need for that is determined, this DMP will be updated to reflect these data types.

2. Data Summary

From "Guidelines on FAIR Data Management in Horizon 2020":

Data Management Plans (DMPs) are a key element of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project. As part of making research data findable, accessible, interoperable and re-usable (FAIR), a DMP should include information on:

- the handling of research data during and after the end of the project
- what data will be collected, processed and/or generated
- which methodology and standards will be applied,
- whether data will be shared/made open access
- how data will be curated and preserved (including after the end of the project)

The storylines that will be produced by the project are based on data originating from observational records, model simulations and qualitative indices supporting scenario assumptions.

With respect to data ownership and data dissemination level within the project, we distinguish between two categories of research data:

1) Data developed and manipulated by project partners during the preliminary research activities before construction of the final projects' storylines;
2) Data that represents finalized elements of the storylines as produced by the project and will be displayed on the website in the storyline builder/visualizer.

We consider the first category of data of intermediate nature, aimed to support the analysis and research activities as a collaborative effort by project partners organized in the work packages. This intermediate data will only become public (and therefore fall under this DMP) if this is required to support the argument for choosing specific climate hotspots. Other intermediate data will not be disseminated by the RECEIPT project and is therefore not within scope of this Data Management Plan. This applies for instance for native (climate) model output data, input data and high resolution (satellite) observations. The second category consists of data that is directly used to illustrate components of the storylines on the RECEIPT platform and websites and will be fully within the scope of this Data Management Plan.

3. Structure of the DMP

This DMP includes this information in the next chapters. *Types and formats* will outline the data formats and standards that will be used for certain types of data. *Metadata* and *FAIR* expand on that by adding comprehensive information on the origin, legislation and findability of that data.

Open and Accessible data outlines the expected data domains and expected open-ness of these datasets. Finally, the chapter

Allocation of Resources and Data Security handles questions on data handling and data retention.

The RECEIPT project uses data from widely different scientific domains (Finance/Insurance, Agriculture, Climate Science, Manufacturing/Supply Chains and Coastal Infrastructure), and therefore data size will vary widely on the various types of data. Volumes of geospatial raster data on a global scale could measure terabytes, while tabular data for financial and supply chain connections could measure as little as kilobytes.

Data generated in the RECEIPT project will be of use to other scientists exploring the implications of an intricately connected planet, especially scientists interested in ties between global climate hotspots and Europe. Finally, the software created in the project will be of interest to anyone seeking to visualise scientific and narrative data or further research data visualisation.

4. Data flow

Data in this project will in most cases flow from a (public or private) research environment to the Zenodo RECEIPT community, either for sharing with consortium partners in the process of research, for use in the storyline building tools, or for long term storage after the project is finished.

Reports and deliverables will follow a slightly different path. They will be shared in the intermediate stage on Microsoft Sharepoint for ease of collaboration and will only be added to the Zenodo RECEIPT repository when finished. We store the final reports in Zenodo to ensure long term storage, even after the project is completed.

Once in the Zenodo repository, data stored will be provided with Metadata as indicated in this document.

5. Types and formats

The data to be gathered and generated in RECEIPT shall, wherever possible, follow the following recommendations for formats:

Table 1: Data Types

Geospatial data	CF- and CMIP6-compliant NetCDF
Tabular data	CSV
Textual data	Commonmark
Image data	PNG or SVG
Audio data	Lossless Audio Codec (FLAC) or MPEG-1 Audio Layer 3 (.mp3) if originally created in this format
Video data	MPEG-4
Documentation	MS Word or PDF
Papers & Articles	LaTeX, MS Word
Storyline Definitions	JSON

These recommendations stem from the list of official recommendations for datasets as set by the UK data service (https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats) and recommended for EU projects.

This list of acceptable types is intentionally kept short, as any additional type would mean a significant amount of additional work would go into writing software for compatibility. For datasets that will be accessible via web services, Open Geospatial Consortium web-services such as WMS, WFS and WCS, and OPeNDAP are appropriate. This could potentially be done on ESGF nodes. Further investigation within the consortium is needed to determine the need and availability of this type of data storage. When existing datasets are used or new datasets are generated in the context of the RECEIPT project, they will be added to the RECEIPT Zenodo community (https://zenodo.org/communities/receipt) if the licence allows it.

If or when following these recommendations is not possible, the party responsible for the addition of the dataset will provide a motivation. Possible motivations may include being unable to modify the data from its initial format for licencing reasons or being forced to use specific software that cannot handle the recommended format. Having to spend additional effort to change the format is not a valid reason for deviating from the recommendations.

6. Metadata

Metadata will be added to any dataset to specify (at least) the following:

Table 2: Metadata

Data Name	Provide a (unique, short) name for the dataset to be used as a reference to this dataset in the RECEIPT project.
Digital Object Identifier (DOI)	A persistent identifier or handle used to identify objects uniquely, standardized by the International Organization for Standardization (ISO)
Owner (ORCID)	The ORCID of the responsible party for including this dataset into the RECEIPT project.
Project Partner	The name of the project partner institute that is responsible for this dataset.
Description	Description of the dataset as a whole and variables in the dataset.
Intended use	Indicate the intended use for the dataset.
Type	The type of data as indicated in Table 1: Data Types
Format	Clearly note what format(s) your data will be in, e.g., plain text (.txt), comma-separated values (.csv), geo-referenced TIFF (.tif, .tfw).
Size	Note what volume of data you will create in MB/GB/TB. Indicate the proportions of raw data, processed data, and other secondary outputs (e.g., reports).
Origin	Please indicate the origin of the dataset.
Provenance	For datasets that are *created* in the context of the RECEIPT project, W3C Prov standards will be used to record provenance.
License	Will said data be open access, only accessible for project purposes, or only accessible by the origin institute. In case an existing dataset is used, please indicate the license of that dataset. In case the dataset is created, please choose a licence. In both cases, if possible, please use the following website to determine the license code of your dataset.
Usage restrictions	If the data contains privacy sensitive information or if there are special requirements for use (notification to the origin, encryption, etc.), please indicate so here.

An example metadata description is provided here, for reference:

Table 3: Metadata example

Data Name	Goods_StatCan
Digital Object Identifier (DOI)	https://doi.org/10.25318/1210007101-eng
Owner (ORCID)	https://orcid.org/0000-0002-9585-4142
Project Partner	Netherlands eScience Center
Description	Trade in goods by importer characteristics, by enterprise employment size and number of partner countries
Intended use	Example dataset to demonstrate DMP
Type	Tabular data with metadata
Format	CSV
Size	6KB
Origin	Statistics Canada https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1210007101
License	https://www.statcan.gc.ca/eng/reference/licence
Usage restrictions	Acknowledgment of Source (a) You shall include and maintain the following notice on all licensed rights of the Information: Source: Statistics Canada, name of product, reference date. Reproduced and distributed on an "as is" basis with the permission of Statistics Canada. (b) Where any Information is contained within a Value-added Product, you shall include on such Value-added Product the following notice: Adapted from Statistics Canada, name of product, reference date. This does not constitute an endorsement by Statistics Canada of this product.

7. FAIR

Data gathered and generated within the RECEIPT project shall follow the FAIR (findable, accessible, interoperable and re-usable) guidelines.

7.1. Findable & Accessible

Data used and created in the RECEIPT project will be made open and accessible via the Zenodo Community (https://zenodo.org/deposit/new?c=receipt). Software to be created for the RECEIPT project will be Open Source and this (as well as links to any such software used from other sources) will be gathered in the GitHub organization created for the project (https://github.com/RECEIPT-H2020)

Naming conventions for data files and folders are yet to be established. Consistent keywords for searching will also be defined later. Clear and consistent versioning of documents, datasets and software releases will be used based on the SemVer (Semantic Versioning) standard.

*For software, g*iven a version number MAJOR.MINOR.PATCH, increment the:

- MAJOR version when you make incompatible API changes,
- MINOR version when you add functionality in a backwards compatible manner, and
- PATCH version when you make backwards compatible bug fixes.

Additional labels for pre-release and build metadata are available as extensions to the MAJOR.MINOR.PATCH format.

*For documents and datasets, g*iven a version number MAJOR.MINOR.PATCH, increment the:

- MAJOR version when you make major changes.
- MINOR version when you make minor changes or only add additional files to a dataset.
- PATCH version when you fix minor mistakes or update fonts etc.

7.2. Interoperable & Re-Usable

The proper use of the data type standards mentioned in Table 1: Data Types as well as clear and consistent licencing and metadata guidelines as stated in Table 2: Metadata will ensure the Interoperability and re-usability of the datasets used and created in RECEIPT.

Specific ontologies for variable names etc. will be determined later, after a survey among consortium members.

8. Open and Accessible data

#	Data type	Data openly available?	Justification & Alternative solution
1	Stakeholder contacts collection	No, statistics only	Although the contacts of the collection are professionals' contacts that may be publicly available, the consortium cannot publish them due to GDPR considerations. Furthermore, the contact information will only be used in the context of the RECEIPT project and will be destroyed at most 1 year after the project ends. The statistical information on the stakeholder data (such as how many, from which countries, which professions etc.) will be integrated in the public report.
2	Expert interview data	No	The data from expert interviews (recordings, protocols and transcriptions) will not be published as primary data due to privacy and security concerns and will be destroyed at most 1 year after the project ends. Anonymization is not considered as an alternative, the sample size allows drawing conclusions on the respondents.
3	Hotspot selection criteria	Yes	Climate hotspot selection criteria should be open access as soon as possible to allow for meaningful discussion with project stakeholders.
4	Geospatial data	Yes	Data shall be gathered in [zenodo RECEIPT community](#)
5	Data on storylines, connections between hotspots and Europe	Yes	Data shall be gathered in [zenodo RECEIPT community](#)
6	Software created in the RECEIPT project	Yes	Software and links to software shall be gathered under the GitHub Organization https://github.com/RECEIPT-H2020
7	Background information	If license allows.	Background is defined as "data, know-how or information (…) that is needed to implement the action or exploit the results". The partners have identified and agreed on the Background for the project and have also, where relevant, informed each other that access to specific Background is subject to legal restrictions or limits. This information is available in Annex 1 of the Consortium Agreement.

8	Statistics on visitor, downloads, likes etc. on the various public channels and websites.	Yes	Anonymized
9	Reports and deliverables	Yes	Intermediate (work in progress) documents will be on the Sharepoint maintained by Arctik, final reports will on Zenodo.
10	Publications	Yes	Open access to research publications is an *obligation* in Horizon 2020

9. Allocation of Resources and Data Security

The Zenodo RECEIPT community (https://zenodo.org/deposit/new?c=receipt) will be used for the collection and sharing of all (finalized) datasets and documents in the RECEIPT project. Zenodo offers a data retention for the lifetime of the repository, which is stated to be guaranteed for at least 20 years.

Zenodo's policies will be adhered to by all members of the RECEIPT project. For clarity and reference, a copy of version 1.0 is provided here:

Content

- **Scope**: All fields of research. All types of research artifacts. Content must not violate privacy or copyright, or breach confidentiality or non-disclosure agreements for data collected from human subjects.

- **Status of research data**: Any status is accepted, from any stage of the research lifecycle.

- **Eligible depositors**: Anyone may register as user of Zenodo. All users are allowed to deposit content for which they possess the appropriate rights.

- **Ownership**: By uploading content, no change of ownership is implied and no property rights are transferred to CERN. All uploaded content remains the property of the parties prior to submission.

- **Data file formats**: All formats are allowed - even preservation unfriendly. We are working on guidelines and features that will help people deposit in preservation friendly formats.

- **Volume and size limitations**: Total files size limit per record is 50GB. Higher quotas can be requested and granted on a case-by-case basis.

- **Data quality**: All information is provided "as-is", and the user shall hold Zenodo and information providers supplying data to Zenodo free and harmless in connection with the use of such information.

- **Metadata types and sources**: All metadata is stored internally in JSON-format according to a defined JSON schema. Metadata is exported in several standard formats such as MARCXML, Dublin Core, and DataCite Metadata Schema (according to the OpenAIRE Guidelines).

- **Language**: For textual items, English is preferred but all languages are accepted.

- **Licenses**: Users must specify a license for all publicly available files. Licenses for closed access files may be specified in the description field.

Access and Reuse

- **Access to data objects**: Files may be deposited under closed, open, or embargoed access. Files deposited under closed access are protected against unauthorized access at all levels. Access to metadata and data files is provided over standard protocols such as HTTP and OAI-PMH.

- **Use and re-use of data objects**: Use and re-use is subject to the license under which the data objects were deposited.

- **Embargo status**: Users may deposit content under an embargo status and provide and end date for the embargo. The repository will restrict access to the data until the end of the embargo period; at which time, the content will become publicly available automatically.

Removal

- **Revocation**: Content not considered to fall under the scope of the repository will be removed and associated DOIs issued by Zenodo revoked. Please signal promptly, ideally no later than 24 hours from upload, any suspected policy violation. Alternatively, content found to already have an external DOI will have the Zenodo DOI invalidated and the record updated to indicate the original external DOI. User access may be revoked on violation of Terms of Use.

- **Withdrawal**: If the uploaded research object must later be withdrawn, the reason for the withdrawal will be indicated on a tombstone page, which will henceforth be served in its place. Withdrawal is considered an exceptional action, which normally should be requested and fully justified by the original uploader. In any other circumstance reasonable attempts will be made to contact the original uploader to obtain consent. The DOI and the URL of the original object are retained.

Longevity

- **Versions**: Data files are versioned. Records are not versioned. The uploaded data is archived as a Submission Information Package. Derivatives of data files are generated, but original content is never modified. Records can be retracted from public view; however, the data files and record are preserved.

- **Replicas**: All data files are stored in CERN Data Centres, primarily Geneva, with replicas in Budapest. Data files are kept in multiple replicas in a distributed file system, which is backed up to tape on a nightly basis.

- **Retention period**: Items will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least.

- **Functional preservation**: Zenodo makes no promises of usability and understandability of deposited objects over time.

- **File preservation**: Data files and metadata are backed up nightly and replicated into multiple copies in the online system.

- **Fixity and authenticity**: All data files are stored along with a MD5 checksum of the file content. Files are regularly checked against their checksums to assure that file content remains constant.

- **Succession plans**: In case of closure of the repository, best efforts will be made to integrate all content into suitable alternative institutional and/or subject based repositories.

10. Ethical Aspects

Good research practices are based on fundamental principles of research integrity. They guide researchers in their work as well as in their engagement with the practical, ethical and intellectual challenges inherent in research. These principles will be adhered to by all RECEIPT project partners who are expected to be familiar with the European Code of Conduct (European Science Foundation, & All European Academies, 2011).

These principles are:

- Reliability in ensuring the quality of research, reflected in the design, the methodology, the analysis and the use of resources.
- Honesty in developing, undertaking, reviewing, reporting and communicating research in a transparent, fair, full and unbiased way.
- Respect for colleagues, research participants, society, ecosystems, cultural heritage and the environment.
- Accountability for the research from idea to publication, for its management and organization, for training, supervision and mentoring, and for its wider impacts.

For the ethical aspects of stakeholder interactions, we refer to the RECEIPT stakeholder interaction protocol, Deliverable 10.2 of WP10: POPD - H - Requirement N0.2. This document includes (among other items) forms for confirmation of written and verbal informed consent.

11. Responsibility and accountability

The responsibility for adhering to the guidelines set out by this data management plan lies with the individual researchers and additional partners in the consortium.

To make sure the guidelines in this document are clear and sufficient, and that they are followed, the consortium data steward will be in frequent consultation with Work Package leaders through a Data Protocol Panel. This panel will be set up by the work package leaders and consist of members from each work package. The members of this panel will each ensure that the guidelines are followed for their individual work packages, and will be educated at the general assembly to able to answer questions about the data management plan.

12. Outlook towards next DMP

As mentioned in the introduction, this document is a living document. Updates will happen as the RECEIPT project progresses. The next version of this DMP will be due in month 18 of the project. This new version will likely include additional findings on data types, formats and datasets that cannot currently be foreseen.

13. References

Digital Curation Center (DMP online template https://dmponline.dcc.ac.uk/)

European Commission. (2016). H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020.

European Science Foundation, & All European Academies. (2011). The European code of conduct for research integrity. European Science Foundation. (https://allea.org/wpcontent/uploads/2017/05/ALLEA-European-Code-of-Conduct-for-Research-Integrity-2017.pdf)

Florian Huber, SYNYO et al. 2017. Data Management Plan of TAKEDOWN: Providing Resources and Support for preventing and countering extremism and crime, Horizon 2020 – FCT-16-2015, 2017

Appendix A: Practical guide to storing documents and datasets

The workflow for storing (confidential) files on Zenodo is as follows:

1. Go to https://zenodo.org/communities/receipt
2. Log in (preferably using ORCID)
3. Click "New Upload"
4. Choose Files **(and Click Start Upload)**
5. Ensure RECEIPT community is selected (is by default when following this tutorial)
6. Choose Upload type, in this case either "Video / Audio" or "Publication / Report?"
7. Basic info:
    a. If a Digital Object Identifier (DOI) was already pre-generated, please use it. It will otherwise be generated.
    b. Fill in date, Title, authors (ORCIDs) and Affiliations, description etc.
    c. Use version numbers (F.E. version 1.2.1):
        i. Given a version number MAJOR.MINOR.PATCH, increment the:
            1. MAJOR version when you make major changes.
            2. MINOR version when you make minor changes or only add additional files to a dataset.
            3. PATCH version when you fix minor mistakes or update fonts etc.

8. Licence:
    a. Access Rights, use
        i. Closed Access for files that will only need archiving.
        ii. Restricted Access for files that other consortium members may request access to.
        iii. Embargoed Access for files that will become Open Access after publication or end-of-project.
        iv. Open Access for everything else.
9. Funding:
    a. I've requested EU / RECEIPT to be added, it has not come through yet. Will update when it has.
10. Optionally add any additional information that is relevant
    a. Basic Information -> Keywords
    b. References
    c. Subjects
    d. etc.
11. Click **SAVE** and **PUBLISH**

RECEIPT