EUCALC

Explore sustainable European futures

Data Management Plan

D11.2

08/2017

update: February 2018

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 730459.

Project Acronym and Name	EU Calculator: trade-offs and pathways towards sustainable and low-carbon European Societies - EUCalc
Grant Agreement Number	730459
Document Type	Report
Work Package	11
Document Title	Data management plan for the EU calculator project
Main authors	Costa L., Matton V., Staniaszek D.
Partner in charge	PIK
Contributing partners	CLIMACT, BPIE
Release date	
Distribution	*Public*

Short Description
one paragraph summary of the report

Quality check	
Name of reviewer	**Date**
Miklós Gyalai-Korpos	09-08-2017
Hannes Warmuth	09-08-2017

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Table of Contents

List of Tables

List of Figures

List of abbreviations

DMP – Data management plan

ToU – Terms of use

IPR – Intellectual property rights

VCS – Version control systems

1 Executive Summary

This deliverable outlines the first version of the EU calculator Data Management Plan. The DMP as outlines only deals with process data (not personal data gather during the EU calculator workshops). Accordingly, this document provides:

i) the characterization of the main types of data expected to be collected and produced during the time frame of the project,
ii) the list of licences detailing data re-use policies and intellectual property rights to be adopted,
iii) the list of repositories and strategies to make the data accessible after the termination of the project,
iv) the first version of the metadata standards to be adopted by the EU calculator both for collected and produced data.
v) The first outline of model documentation guidelines.

2 Introduction

2.1 Importance of the plan and objectives

The appropriate management of data is an essential, often forgotten, undertaking of responsible research. At the start of a new research project, it is important to lay down the basic principles relating to data and data management. The creation of a shared Data Management Plan (DMP) at the beginning of the project meets the requirement of informing the involved partners on important procedures and processes regarding data collection, processing, storing and distribution. Furthermore, the DMP is pivotal in guaranteeing that the data collected and processed can be easily and properly shared within the consortium and distributed beyond the project lifetime. A data management plan helps achieve optimal handling, organizing, documenting and enhancing of research data. It is particularly important for facilitating data sharing, ensuring the sustainability and accessibility of data in the long-term, and allowing data to be reused for future research.

For maximum effectiveness, the DMP must start when research is being designed and needs to consider both how data and information will be managed during the research and how they will be shared afterwards. This involves thinking critically of how research data can be shared, what might limit or prohibit data sharing, and whether any steps can be taken to remove such limitations. In the context of the European Calculator project, plans regarding the usage of data and model documentation started during its preparation phase. The time and thoughts devoted to the data management issues were preliminary and hence it is now time to undertake a more concerted effort in elaborating the MP for the European Calculator project. The DMP for a research project in the H2020 program should elaborate on the following aspects (without any particular order of importance).

- 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 and how data will be curated and preserved (including after the end of the project)

These aspects are a mandatory requirement in the Programme Guidelines on FAIR Data Management in Horizon 2020[1]. The guidelines from the European Commission on the requirements for the DMP focus on the issues of data documentation, standards applied and data re-use.

During the lifetime of the European Calculator project the DMP will be updated in D11.7 (Month 35). The DMP described in the sections below is to be taken as a first iteration document to synchronize the knowledge of partners on their data-related responsibilities and to outline the major guiding principles of the EU calculator project in regard to its data policy. I addition, future iterations will also explore potential synergies of the EU calculator DMP with those being developed by INNOPATHS and REINVENT.

2.2 FAIR data principles

The open data policy of the European Union favours the implementation of the G8 Open Data Charter[2] and the FAIR Guiding Principles[3] for scientific data. The latter highlights core principles as enumerated below:

To be Findable:

- Data are assigned a globally unique and persistent identifier
- Data are described with rich metadata
- Metadata clearly and explicitly include the identifier of the data it describes
- Data are registered or indexed in a searchable resource

To be Accessible:

- Data are retrievable by their identifier using a standardized communications protocol
- The protocol is open, free, and universally implementable
- The protocol allows for an authentication and authorization procedure, where necessary
- Metadata are accessible, even when the data are no longer available

To be Interoperable:

- Data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- Data use vocabularies that follow FAIR principles
- Data include qualified references to other data

[1] H2020 Programme Guidelines on FAIR Data Management in Horizon 2020, accessed on the 3-05-2017 13:44, available under https://goo.gl/0D3tUz

[2] http://opendatacharter.net/principles/

[3] https://www.nature.com/articles/sdata201618

To be Reusable:

- Data are richly described with a plurality of accurate and relevant attributes
- Data are released with a clear and accessible data usage license
- Data are associated with detailed provenance
- Data meet domain-relevant community standards

The EU calculator project will adhere to these principles during the elaboration and fulfilment of its DMP.

3 Data origin and type

The European calculator project will make use of a number of heterogeneous data sources that can be broadly classified as external and internal to the project. By external it is meant that datasets are acquired from institutions external to the European calculator consortium. By internal it is meant that the data is owned by the partners or generated within the European Calculator activities. As a rule, external open-datasets that are regularly updated will be favoured although one cannot discard the possibility of the project requiring particular external datasets subject to licensing.

3.1 External data

3.1.1 Open sources

The European calculator project will connect to open data sources that are regularly maintained and updated from European and Global institutions. For example, it is foreseen to substantially rely on Eurostat and OECD data, both as model baseline and as input for statistical relations to quantify some of the model dynamics. External data repositories such the Database on Shared Socioeconomic Pathways hosted at IIASA[4] and climate data from initiatives like CORDEX [5] and the ODYSEE database[6] are also foreseen to be accessed and data retrieved during the running time of the project. Databases from existing and closed EU-funded projects such as Heat Road Maps[7] project will also be scanned for their potential in supplying the European calculator project with data. For example datasets from the TRACCS[8] project or the EU Buildings Database[9] are also expected to be used during the project.

[4] https://tntcat.iiasa.ac.at/SspDb/dsd?Action=htmlpage&page=about

[5] http://www.cordex.org/

[6] http://www.odyssee-mure.eu/

[7] http://www.heatroadmap.eu/

[8] http://traccs.emisia.com/index.php

[9] https://ec.europa.eu/energy/en/eu-buildings-database

3.1.2 Paying sources

At the time of writing there is not yet a clear indication that datasets falling into this category are required. This section will be updated in the next iteration the ECDMP.

3.2 Internal data

In regard to internal datasets, these will refer mostly to the data generated during the lifetime of the project and are foreseen to comprise mostly model outputs and expert opinions gathered during the sectoral expert-consultation workshops. Regarding the latter, there is in place a common procedure to guarantee confidentiality of personal data (see Del. 12.1). The expert consultation workshops will supply the team of the European Calculator with opinions on the modelling approaches undertaken and evidence base support on the choice of particular ambition levers.

To a lesser extent, datasets and model processes owned by the institutions comprising the European Calculator project will also be used. For example, published datasets like the BPIE Building observatory database or dietary patterns and decarbonisation pathways will be used to inform on the development of the model.

3.3 Data types

The following section describes the current data types forecast to be generated during the project. Note that this is a preliminary judgement based on the author's understanding of the EU Calculator model. This section will be updated periodically as part of the ECDMP life cycle.

Identifier	Label	Description	Type	Main responsible	Access
OTS	Observed time series (historical data)	Observations of climate, socio-economic variables and emissions.	Numeric	WP leaders of the respective modules.	Public
FTS	Future time series (generated data)	Projections of climate, socio-economic variables and emissions.	Numeric	WP leaders of the respective modules.	Public
LL	Levels of levers	Levels of technology or lifestyle ambition.	Numeric	WP leaders of the respective modules.	Public
CP	Constant parameters	Time-invariable parameters required for the model.	Numeric	WP leaders of the respective modules.	Public
MC	Model/module code	Source code of modules and model.	Code	WP leaders of the respective modules & CLIMACT.	Public
MD	Model/module documentation	Documentation of model/module code	Text	WP leaders of the respective modules & CLIMACT.	Public

Table 1 – Types of data considered in the European Calculator project.

4 Data gathering guidelines

Within the EU calculator project data gathering is a responsibility of the partners working on the specific WP's and topics. They have the best knowledge within the consortium on the adequacy and goodness data in their respective research fields. Accordingly, the purpose of this section is not to develop guidelines that act on the decision of partners for a given dataset but to provide some general rules partners should attend in order to make the data collection as homogenous across the project as possible. The guidelines mostly refer to OTS data (Table 1), but can be used for reference regarding other types of numerical data.

4.1 General principles

In the choice for data partners in the EU calculator project are advised to follow the following general principles:

- Prioritize freely available databases.
- Extract data from European-specific databases that are curated by established research or institutions.
- Prioritize data that cover all the countries and time frame of focus in the EU calculator.
- If two or more sources of data are available for the same variable of interest, prioritize the most recently updated or the one for which regular updates are done.

The following sections discern on particular cases for which the principles above are challenged, and provide general decision guidelines to the partners.

4.1.1 Spatial completeness

Data types identified as OTS or LL has to be gathered for the 28 member states and Switzerland. In order to guarantee consistency the strategy of data collection should obey to the following hierarchy. First prioritize the use of Europe-specific datasets in which all of the countries of focus are present. This is not always possible, even within datasets like those provided by the Eurostat.

If countries are missing from the European-specific dataset then evaluate if data is available from a global or national database. The decision for which is left to the responsible doing the data collection and should be documented both in the deliverable associated with the database and in the metadata (see section 5). In case no global or national data is found for the particular country, then the partners are asked to approximate the data by the means of meaningful rule. For example, countries with the same GDP and road network have similar transport demand. This option should have the lower priority and only be used in case the option beforehand are no feasible.

The rule developed for data completion should be documented both in the deliverable associated with the database and in the data's metadata.

4.1.2 Time completeness

OTS data (Table 1) used in the EU calculator has to be collected, on a yearly resolution, for the minimum 1990-2015 time span. Longer or sub-annually OTS might be used for the purposes of developing or parameterizing the model. Databases belonging to the OTS type often present gaps for a particular year or

sequence of years. In case gaps exist, a "data filling" routine needs to be used. In case of small data gaps a simple interpolation can be operated. The functional shape of the interpolation (e.g., linear, exponential) is decided by the partner conducting the research and needs to be documented in the associated deliverable and metadata file. It is not advisable that the interpolation is used to cover very large gaps (e.g., more than 10 years as rule of thumb). For the cases where data has to be reconstructed for periods of more than 10 years partners have to make sure they elaborate a reasonable approach identifying main drivers of the variable in question and their time dependencies. The procedure needs to be described accordingly in the respective deliverable or, in case that there is no deliverable attached, in the progress reports.

4.1.3 Contradicting databases

For some cases two or more databases might be available for the same variable of interest. Often the values present in each database can be similar or contradictory; rarely will they be the same. In these cases the responsible partner faces two choices, a) opting for one of the databases or b) combining both databases into a third one (e.g., averaging both or taking lowest/higher values). Partners are asked to prioritize the choice of one database over the option of merging both. The choice of database should reflect the criteria in sections 4.1 and 4.2; namely the best spatial completeness possible and the lower number of missing years (in case of a time series). If both databases are equivalent, then partners are requested to opt for the most recently updated one.

5 Metadata standards and data formats

All the data used within the project will be available using non-proprietary formats and documented accordingly via the use of extensive metadata descriptions and EU-calculator naming conventions. The metadata descriptions will contain the required elements to guarantee that data are easily discovered. Table 2 enumerates and describes the foreseen metadata elements to be used when documenting Observed Time Series (OTS), Future Time Series (FTS), Level of Levers (LL) and Constant Parameter (CP) data in the EU calculator project.

Attribute name	Description
ID	Unique identifier of the dataset.
Title	Dataset title.
Summary	Abstract related to the title attribute.
Variable	Short variable name.
Unit	Unit of the variable.
Activity	Name of the project.
Tags	List of keywords commonly used to describe the subject.
Frequency	Time frequency of the variable.

Period and reference	Time period for which the variable was calculated and respective reference year.
Institution	URL of the home page of the institution compiling or producing the data.
Contact	Email contact of the main responsible for the data as compiled or produced for the EU calculator project.
Contributors and role	Any name of person contributing for the data compiled or produced for the EU calculator project, as well as the respective role.
Methods summary	Brief description of the methodology used to compile/calculate the data.
Data filling	Description of the approach to fill in missing country data.
Source data	Any relevant sources used to compile or produce the data.
Quality control	Description of the quality control process before data publication.
Comment	Miscellaneous information about the data/methods used to derive the dataset.
References	Any additional references.
Date created	Date of data creation (YYYY-MM-DD).
Data type	Type of data
Workpackage and task	Project WP and task from which the data originates.
Version status	Version of the data and its status for usage.

Table 2 – Metadata elements for OTS, PTS, LL and CP data.

Datasets on the Levels of Levers (LL), Observed/Future Time Series (OTS/FTS) and Constant parameters (CP) will be made publically available following the CSV or XLS tabular data standard. OTS refers to historical data collected from sources. The difficulty is to find data for every country and to fill the gaps. This data is used as input of for the EU calculator model. In this respect it should be noticed that 1) only credible sources could be used that have primary/secondary access to the data, i.e. owning/overseeing the system producing/recording the data or legally obliged to collect data (statistical offices) and, 2) if possible, use the data from EU level/international bodies that gather the data from member state level organizations by law, such as Eurostat, the IEA or the World Bank.

Data coded as FTS are generated by the model. This is the "matrix of possibilities" that will be used by the Web application; hence, this data is an output of the model and INPUT of the Web application. LL data are created by each WP in cooperation with the stakeholders. The LL data refers to scenarios built for each lever and are therefore inputs of the model. Finally, CP data refers to constants (physical, geographic, etc...; for example country entity, mass to kcal conversion) that are required and input for the model. The documentation of this data obeys to a template whose first version is described in section 8.

Module Documentation (MD) data will be made available in PDF format. As for the Model Code (MC) this will be made available in the KNIME and Python formats.

- KNIME is an open source data analytics platform (https://www.knime.com/). It integrates various components for machine learning and data mining and possess a graphical user interface to allow

assembly of nodes for data preprocessing (e.g., extraction, transformation, etc), for modeling and data analysis and visualization.

- Python is a high-level programming language for general-purpose programming. It is widely used in the scientific community given its scalability and philosophy emphasizing code readability.

6 Policy of data re-use

6.1 Open source licence

The data produced in the European Calculator project will be stored as a comprehensive database and therefore illegible for intellectual property rights (IPR) and subjected to licencing. When it comes to intellectual property rights and licences for data, the central notion is the database as "a collection of independent works, data or other materials arranged in a systematic or methodological way and individually accessible by electronic or other means"[10]. The database notion is not restricted to a data collection stored in traditional database management systems, it relates also to data stored in a file and organized in a well-structured manner.

The data resulting from the European Calculator will be published under the Creative Commons Attribution International License **CC-BY-4.0** (https://creativecommons.org/licenses/by/4.0/), and the Open Data Commons Attribution Licence **ODC v1.0** (https://opendatacommons.org/licenses/by/). These licences are permissive and do not include a copyleft clause[11]. They allow sharing (copy, redistribute and use the data) as long as the user entity gives appropriate credit.

7 Data storage and preservation

7.1 Model data storage

Data used in the development of the EU calculator model will be stored online using the Amazon S3 (https://aws.amazon.com/en/s3/) solution. Inputs for the model covering the data described in section 3 and documented according the standards described in section 4 will be stored in an EU calculator dependency. The details on how to upload both the data and metadata will be made available in the next iteration of the DMP.

7.2 Long-term storage plan

The data of the European Calculator project will be stored in a research data repository like PANGAEA (www.pangaea.de). PANGAEA is a free publishing repository for environmental data (although some small fees might apply in case

[10] European Community (1996) Directive 96/9/EC of the European Parliament and of the Council on the legal protection of databases.

[11] The inclusion of a *copyleft* clause (also called protective) implies that derived works using European Calculator data will have to licence their products with the same rights. The exclusion of a *copyleft* (also called permissive) clause allows for more freedom for data distribution.

of data from big projects) with a Digital Object Identifier (DOI) service. From this generic storage place for data, the EU calculator team will reach out to other data repositories in order to increase the visibility and secure long-term preservation of our outputs by exploring the following possibilities:

- Linking the data sets produced in the European calculator to OpenEI.org (http://en.openei.org/wiki/Main_Page). The Open Energy Information (OpenEI.org) initiative is a free, open source knowledge-sharing platform created to facilitate access to data, models, tools, and information that accelerate the transition to clean energy systems through informed decisions. OpenEI strives to make energy-related data and information searchable, accessible, useful to both people and machines

- Linking the data sets produced in the European calculator to European's Union Open Data Portal (https://data.europa.eu/euodp/en/data/). The European Union Open Data Portal is the single point of access to a growing range of data from the institutions and other bodies of the European Union. The portal aims to promote their innovative use and unleash their economic potential. It also aims to help foster the transparency and the accountability of the institutions and other bodies of the European Union. The Open Data Portal is managed by the Publications Office of the European Union.

- Upload our datasets to the Open Energy Modelling Initiative (www.openmod-initiative.org/). The Open Energy Modelling Initiative is more in line with the thematic focus of the European Calculator and hence the project outputs could gain more visibility for the community if announced through this channel.

7.3 Model documentation

One of the most important criteria of the EU Calculator is the transparency of the model. This can only be achieved by allowing anyone to understand the model and to find the source of every data that we are using. The data we are collecting are evolving quickly. It is important to keep track of the version of every data we are using. Finally, every assumption/hypothesis that we are taking are influencing the final result of the calculator, we need to track them and to explain them with details.

- For Observed Time Series data that we are collecting as input of the model, we need to document the place where we found them, the date, the owner, as well as the method we are using to fill the missing data.
- For Future Time Series built regarding expert opinion and stakeholder consultations, we need to document the critical hypothesis that we took to build our matrices of scenarios;
- For Levels of Levers data, it is crucial to track the assumption we made to define the lifestyle ambition or the levels of technology.
- For Model Code, everyone should be able to understand the code (Knime or Python) without having to know how to code. This is only possible by documenting the code with a deep Model Documentation (MD) inside the code itself.

7.3.1 Model versioning

The EU Calculator is a large project with several complicated modules/components where multiple developers are working together at the same time. It is simply not feasible to keep track of every modification and to be able to go back to an older version (in needed) without having something called version control.

Multiple Version Control Systems (VCS) are on the market. We propose to use the popular and open source GIT system in the EU Calculator project and to host our code using the Bitbucket solution (https://bitbucket.org/product).

> Git is a free and open source distributed version control system designed to handle everything from small to very large projects with speed and efficiency. It outclasses SCM tools like Subversion, CVS, Perforce, and ClearCase with features like cheap local branching, convenient staging areas, and multiple workflows. (https://git-scm.com/).

Git will allow us to keep track of the changes in the code but it is the task of every developer to document his code and the version he is pushing to the server. The code of the European Calculator can be found at the following address (you need to be registered to access it): https://bitbucket.org/eucalcmodel/

8 Responsibilities and duties

Each work-package leader institution is responsible for supervising the creation, compilation and adequate documentation of data during the life time of the project. This includes making the data available for the rest of the team, as well as guaranteeing that the data compiled or generated adheres to the metadata and format standards described in the ECDMP. PIK will support the partners with doubts regarding the metadata standards. It also falls within each work package the responsibility to keep the module documentation for the EU calculator model up to gate and compliant with the template provided. CLIMACT is responsible for supervising the model documentation, curating the code and making it available in the predefined format.

9 Annex

9.1 Metadata guidelines

This section aims at providing a "user guide" on how to fill the metadata table for a variety of data types identified in section 3.3. Furthermore, it attempts to anticipate some of the inconsistencies that arise when documenting different types of data. These inconsistences will certainly be more visible during the further stages of the project when data starts to be produced. Let us start by filling in the metadata structure for the variable GDP for European member states (see Figure 1).

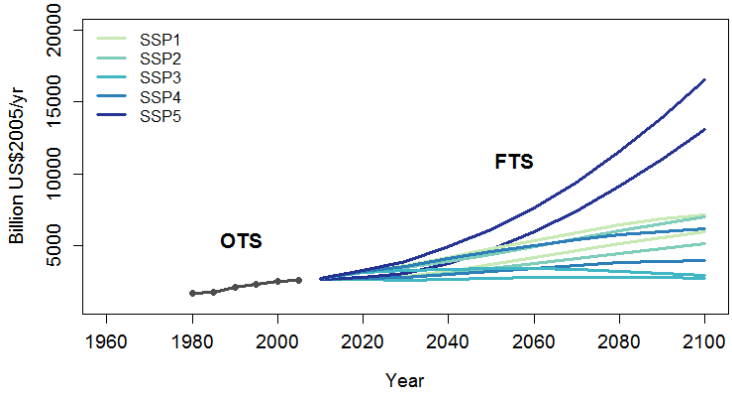

Figure 1-Example of GDP data time series and future trajectories extracted from the IIASA SSP database.

The variable can be retrieved from multiple sources. The figure above shows GDP time series from the World Bank, in black, as well as future trajectories of GDP from two models (OECD-ENV-Growth model[12] and IIASA (Crespo Cuaresma, 2017) implied by the five Shared Socio-economic Pathway narratives in use by the IPCC (O'Neill et al., 2017), blue gradients. The future time series GDP are themselves part of a larger database of socio-economic projection hosted by IIASA[13].

The first step for documenting metadata is to identify the appropriate data type. Accordingly, past observations of GDP fall under Observed Time Series (OTS) data while the projections fall under Future Time Series (FTS). This allows for the **ID** field to be filled (see Table 2) according to the following standard (all in small case):

<data type>_<module trigram>_<variable name>

[12] https://www.oecd.org/environment/indicators-modelling-outlooks/Flyer%20ENV-Growth%20model%20-%20version%2025%20Sept%202013.pdf

[13] https://tntcat.iiasa.ac.at/SspDb/dsd?Action=htmlpage&page=about

"Data type" refers to the type of data as referred in Table 1. The "module trigram" refers to the module abbreviation in for which the data is produced. These abbreviations are given in the table below.

Module	Trigram
Lifestyles	lfs
Transport	tra
Buildings	bld
Materials	mat
Energy	erg
Land, Water and biodiversity	lwb
Social Impacts	sip
Transboundary	trb
Climate	clm

At the time of writing, it is not yet clear the final number of modules in the EU calculator model. Accordingly, in this deliverable we provide the abbreviations for the modules currently foreseen. The second update of the DMP will present the final list of modules.

By "variable name" it is meant a short name (abbreviation) reflecting the full name of the variable being documented. For the particular case of GDP time series provided by the World Bank. The GDP variable has been compiled under WP1 as part of Task 1.3. For short name let us document the dataset as **gdp** or **pop** in case of population.

As good practice the naming of the variable should be kept short but without hindering understandability. It is also advisable to avoid spaces when documenting the variable name.

Accordingly, the ***ID*** attribute for this particular dataset would translate to: **ots_lfs_gdp**. The ***ID*** attribute for the GDP compiled from the IIASA database would translate to: **fts_lfs_gdp**.

Attribute name	Description
ID	ots_lfs_gdp ***or*** fts_lfs_pop

The attributes **Title**, **Summary**, **Tags**, **Variable** and **Unit** are rather self-explanatory. The important thing to keep in mind is to be, at the same time, complete in your descriptions but concise enough so that metadata can be explored quickly. Below some examples on how to fill the highlighted attributes.

Attribute	Description

name	
Title	Gross Domestic Product
Summary	This dataset contains Gross Domestic Product (GDP) numbers observed for European Union member states + Switzerland between 1970 and 2005. ***or*** This dataset contains projected Gross Domestic Product (GDP) numbers for European Union member states + Switzerland between 2010 and 2100.
Tags	GDP, Europe, Economy, Country
Variable	GDP
Unit	billion US$2005/yr

Under the **Activity** attribute it should be identified the project, initiative, etc… under which the data was created. This field might not be always applicable or can be more relevant for some datasets than others. For example; the mandate for the World Bank to collect GDP data does not necessarily fall under a specific "activity" but it is an intrinsic part of the institution's activities. Hence, in this case simply add **–** (minus) in the respective field. For cases in which the data is derived from other European projects, and specially the EU calculator, then filling the attribute becomes mandatory. There are no specific guidelines on how to do this. Try to use the project acronym in case you retrieve data from an existing project, e.g., CMIP5[14], TRACCS[15]. In case the dataset is the product of the EU Calculator project then code the attribute as **EU calculator**. The attribute **Source data** refers to any sources used to produce the dataset being documented. In case the data stems from the **EU calculator** model then reference the Module documentation or Deliverable in which the data methods can be found. For our particular example the **Activity** and **Source data** attributes are filled as follows:

Attribute name	Description
Activity	- or Fifth Assessment Report
Source data	http://data.worldbank.org/indicator/NY.GDP.MKTP.CD or https://tntcat.iiasa.ac.at/SspDb/dsd?Action=htmlpage&page=about

Below you find examples on how to document the **Frequency** and **Period and reference** attributes for the GDP data. These attributes might not be always logic in regard to all datasets being produced. In cases for which the attributes are not logic please code the metadata field with – (minus). In regard to data type Level of Levers it is mandatory to fill in the abovementioned attributes. In case the data does not have a reference year document only the respective time period.

[14] http://cmip-pcmdi.llnl.gov/cmip5/

[15] http://traccs.emisia.com/index.php

Attribute name	Description
Frequency	5 years ***or*** 10 years
Period and reference	1970-2010 (2005) ***or*** 2015-2100 (2005)

Below we provide examples of how to fill the metadata attributes **Institution, Contact, Contributors and role, Methods summary, Data filing,** and **References.** These attributes are rather self-explanatory and provide additional information to track down the responsible person for the data and methodological details.

Attribute name	Description
Institution	www.pik-potsdam.de
Contact	carvalho@pik-potsdam.de
Contributors and role	Luis Costa extracted the data from its original source and compiled if for the purposes of the EUcalculator project.
Methods summary	The data was compiled for all EU member states + Switzerland from the documented sources. No spatial or temporal aggregation was conducted to the source data and original units were kept unchanged.
Data filling	Due to the completeness of the original data, no procedure to fill in data gaps was required.
References	A more in depth description of the projected GDP data compiled, in particular the logic and narratives of the different scenarios, please consults the following publications: Dellink, Rob, et al. "Long-term economic growth projections in the Shared Socioeconomic Pathways." *Global Environmental Change* (2015). Leimbach, Marian, et al. "Future growth patterns of world regions–A GDP scenario approach." *Global Environmental Change* (2015). Cuaresma, Jesús Crespo. "Income projections for climate change research: A framework based on human capital dynamics." *Global Environmental Change* (2015).

The remaining metadata attributes provide entry points to track down data progress and data status. Under **Quality control** a brief description of the control mechanisms to assured data quality should be described. This is particular relevant when the data is generated by the EU calculator team. **Data created** should obey to the following standard YYYY-MM-DD. **Data type** refers to the type of data and **Workpackage and task** identifies the respective WP and task from which the data emerges. **Version status** should highlight the maturation stage of the data and should be coded as one of the following

standards: “draft/internal use only”, “preliminary/internal use only”, “near final/internal use only” and “final/free to use”.

Attribute name	Description
Quality control	The data did not undergo in quality control as it is taken without modification from its original source. The EUcalculator team assumes that there has been adequate quality control of the original dataset by the responsible institutions.
Date created	2017-03-22
Data type	fts
Workpackage and task	Workpackage 1 task 1.3
Version status	Preliminary/internal use only

Each data file collected and produced for the EU calculator model needs to have a corresponding metadata file. The metadata file itself should be saved in **.xls** format and named as in **ID**, followed by the suffix “**_md**” standing for “metadata”: Accordingly, **fts_lfs_gdp_md.xls** The **.xls** template of the metadata attributes is shown below and available for download here. Both the data and metadata files have to be uploaded to the EU calculator AWS dependency once it is online (see section 6.1).

	A	B
1	*EU calculator metadata template version 1.0*	
2	*This file documents the metadata atributes for data collected and produce in the EU Calculator project.*	
3	*Each file with numerical data required to running the EU calculator model needs to be have a correspoding metadata file.*	
4	*Please refer to deliverable 11.2 Annex 8 on how to fill the metadata attributes.*	
5	Attribute name	Description
6	ID	*Unique identifier of the dataset.*
7	Title	*Dataset title.*
8	Summary	*Abstract related to the title attribute.*
9	Variable	*Short variable name.*
10	Unit	*Unit of the variable.*
11	Activity	*Name of the project.*
12	Tags	*List of keywords commonly used to describe the subject.*
13	Frequency	*Time frequency of the variable.*
14	Period and reference	*Time period for which the variable was calculated and respective reference year.*
15	Institution	*URL of the home page of the institution compiling or producing the data.*
16	Contact	*Email contact of the main responsible for the data as compiled or produced for the EU calculator project.*
17	Contributors and role	*Any name of person contributing for the data compiled or produced for the EU calculator project, as well as the respective role.*
18	Methods summary	*Brief description of the methodology used to compile/calculate the data.*
19	Data filling	*Description of the approach to fill in missing country data.*
20	Source data	*Any relevant sources used to compile or produce the data.*
21	Quality control	*Description of the quality control process before data publication.*
22	Comment	*Miscellaneous information about the data/methods used to derive the dataset.*
23	References	*Any additional references.*
24	Date created	*Date of data creation (YYYY-MM-DD).*
25	Data type	*Type of data*
26	Workpackage and task	*Project WP and task from which the data originates.*

10 References

Crespo Cuaresma, J., 2017. Income projections for climate change research: A framework based on human capital dynamics. Glob. Environ. Change 42, 226–236. doi:10.1016/j.gloenvcha.2015.02.012

O'Neill, B.C., Kriegler, E., Ebi, K.L., Kemp-Benedict, E., Riahi, K., Rothman, D.S., van Ruijven, B.J., van Vuuren, D.P., Birkmann, J., Kok, K., Levy, M., Solecki, W., 2017. The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century. Glob. Environ. Change 42, 169–180. doi:10.1016/j.gloenvcha.2015.01.004