REFLEX – Analysis of the European energy system
under the aspects of flexibility and technological progress

Call: H2020-LCE-21-2015

Grant Agreement Number: 691685

Deliverable

D2.4 Updated Data Management Plan (DMP)

Deliverable type:	ORDP: Open Research Data Pilot
WP number and title:	WP2: Data Warehouse and Model Coupling
Dissemination level:	Public
Due date:	Month 28 – 31.08.2018
Lead beneficiary:	ESA²
Lead author(s):	**Robert Kunze**
Reviewers:	Claudia Hawke and Steffi Schreiber (TUD) Artur Wyrwa (AGH)

This project REFLEX has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 691685.

DOCUMENT HISTORY

Version	Date	Author/Editor	Description
First DMP (D2.1)			
0.1	20.10.2016	Robert Kunze	Initial version, sent to reviewers
0.2	27.10.2016	Robert Kunze	Revised initial version, sent to reviewers
1.0	01.11.2016	Robert Kunze	Published version with reviewer comments incorporated
Updated DMP (D2.4)			
1.1	22.08.2018	Sophia Wolter Robert Kunze	Initial version of updated DMP, sent to reviewers • Review/minor changes in: scope of the document; administrative details; abbreviations; references • Major adaptations and inclusions in: data summary, FAIR data • New sub-chapters on licensing in general, choice of licenses for REFLEX data and potential legal conflicts
1.2	31.08.2018	Artur Wyrwa Steffi Schreiber Claudia Hawke	Reviewed version sent back to authors
2.0	31.8.2018	Robert Kunze	Published version with reviewer comments incorporated

TABLE OF CONTENTS

ABBREVIATIONS

CA	Consortium agreement
DMP	Data management plan
DSM	Demand side management
DWH	Data warehouse
EC	European Commission
EIM	Exploitation and innovation manager
EMS	Energy models system
EU	European Union
FAIR	Findable, accessible, interoperable and reusable
GA	Grant agreement
LCA	Life cycle assessment
NDA	Non-disclosure agreement
RES	Renewable energy sources
SET-Plan	Strategic energy technology plan

LIST OF FIGURES

LIST OF TABLES

SCOPE OF THE DOCUMENT

This document provides the updated version of the Data Management Plan (DMP) for the REFLEX project according to the Open Research Data Pilot (ORD Pilot) under Horizon 2020. The purpose of this DMP is to support the data management life cycle of all data that are collected, processed or generated in the course of the project.

Research data shall be findable, accessible, interoperable and reusable (FAIR data management). We, therefore, aim to publish modelling input data as open as possible, in order to support a transparent research environment that allows third parties to validate results presented. We, second, aim to publish final project modelling results in a format that makes it possible for third parties to reuse (i.e. access, mine and exploit) the data. In this context, a high-quality data management has a significant relevance.

The document structure and contents are based on the Guidelines on FAIR Data Management in Horizon 2020 (EC, 2016, Version 3.0) and on the Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020 (EC, 2016b, Version 2.1). The structure of the document was generated by using the Digital Curation Centre's DMP online tool.

The following sections outline the purpose and procedures of data collection and generation as well as relevant data flows in the REFLEX project; the types and formats of collected and generated data; the data utility for third parties; how these data will be exploited and made accessible for verification and reuse respecting data ownership; the application of guidelines on FAIR data management and the choice of license options for open data; and finally how data will be curated and preserved upon closure of the project.

ADMINISTRATIVE DETAILS

Project Name: REFLEX (Horizon 2020 DMP)
Grant Agreement No.: 691685
Principal Investigator / Researcher: Prof. Dominik Möst, Technical University of Dresden

Project Description:

The future energy system is challenged by the intermittent nature of renewable energy sources (RES) and requires therefore an increased use of flexibility options, on both supply- and demand side. Still, the interaction between different flexibility options, the optimal portfolio of generation technologies and the impact on environment and society are unknown. It is thus the core objective of REFLEX to analyse and evaluate the development towards a low-carbon energy system with the focus on flexibility options (e.g. Demand Side Management) in the EU to support the implementation of the EU's Strategic Energy Technologies Plan (SET-Plan).

The analysis is based on a modelling environment that considers the full extent to which current and future energy technologies and policies interfere and how they affect the environment and society while considering technological learning of low-carbon and flexibility technologies. For this purpose, REFLEX brings together the comprehensive expertise and competences of known European experts coming from nine institutions of six different countries. Each partner focusses on one of the research fields of techno-economic learning, fundamental energy system modelling, or environmental and social life cycle assessment.

To link and apply these three research fields in a compatible way, an innovative and comprehensive "Energy Models System" (EMS) is developed. The model pool of REFLEX partners contains bottom-up simulation tools and fundamental system optimization models on national and European level as well as approaches for life cycle assessment. The EMS couples the models and tools. It is based on a common database and a common scenario framework.

Results from applying the EMS will help to understand the complex links, interactions and interdependencies between different actors, available technologies and impact of the different interventions on all levels from the individual to the whole energy system. In this way, the knowledge base for decision-making concerning feasibility, effectiveness, costs and impacts of different policy measures will be strengthened, which will assist policy makers and support the implementation of the SET-Plan.

Stakeholders will be actively involved during the entire project from the definition of REFLEX scenario storylines to the dissemination and exploitation of results via workshops, publications and the project website.

Nature:

Research and innovation actions based on energy system modelling.

Research questions:

1) How do current and future energy technologies and policies interfere?
2) What will be an optimal combination of different flexibility options to cope with the future flexibility needs?
3) How do these technologies and policy measures affect the environment, economy and society?

Purpose:

Support for the implementation of the SET-Plan:

- Analysing the impact of technological development and innovation on the energy system and its dynamics;
- Comparative assessment of the impacts and the sustainability performance of all relevant energy technologies;
- Assessing the related impacts on the environment, society and economy, analysis of technology policy measures in the framework of the SET-Plan;
- Understanding the complex links/interactions/interdependencies between the different actors, the available technologies and the impact of the different interventions on all levels from the individual to the whole energy system;
- Providing model-based decision support tools for the different actors in the energy system in order to facilitate handling the complex system.

Funder:

European Commission (Horizon 2020).

1 DATA SUMMARY

At the core of the REFLEX project is the development of the comprehensive "Energy Models System" (EMS), coupling different models and tools from the REFLEX partners. A good understanding of data types and data flows is therefore indispensable. In what follows, we discuss the purpose of data collection and generation, the types and format of data, data utility for third parties as well as the REFLEX data protection and exploitation strategy.

1.1 PURPOSE OF DATA COLLECTION AND GENERATION

As briefly introduced above, the main objective of the REFLEX project is to analyse and evaluate the development towards a low-carbon energy system in the EU up to the year 2050. The focus is laying on the evaluation of several flexibility options to support the system integration of increasing generation from intermittent RES (see Deliverable 4.1: Overview of techno-economic characteristics of different options for system flexibility provision; and Deliverable 4.3: Report on cost optimal energy technology portfolios for system flexibility in the sectors heat, electricity and mobility). Thereby, the analysis and assessment are based on a modelling environment that considers the full extent to which current and future energy technologies and policies interfere and how they affect the environment and society while considering technological learning of low-carbon and flexible technologies.

For analysing and answering the given research questions (see Administrative details), the different models and approaches are coupled to the EMS (see Deliverable 2.3: Report on modelling coupling framework). The *purpose for data collection* and their preparation within REFLEX is to provide the needed input data for the applied mathematical energy system models. The *purpose of data generation* is to provide the quantitative background which will help to understand and investigate the complex links, interactions and interdependencies between the different actors and technologies within the energy system as well as their impact on society and environment.

The following sections give an overview on data flows in the REFLEX project and introduce the project's data structure and different data types.

1.1.1 DATA FLOWS IN THE REFLEX PROJECT

The **model pool** of the REFLEX partners contains bottom-up simulation tools and fundamental system optimisation models on national and European level as well as approaches for Life Cycle Assessment (LCA). Typically, one model cannot cover all aspects of an energy system or the implications of specific policies. Each of these different models focuses on a specific sector or aspect (e.g. heat, electricity, mobility, environmental / social impacts etc.) of the energy system. The models applied in REFLEX can be grouped into four fields:

- Energy supply (ELTRAMOD, TIMES-Heat),
- Energy demand/usage (TE3, ASTRA, FORECAST, eLOAD),
- Energy market design (PowerACE), and
- Impacts on environment, society and economy (eLCA, sLCA, πESA).

For analysing and answering the given research questions, the different models and approaches are **coupled to the integrated Energy Models System (EMS)**. Applying the EMS allows to perform an in-depth and at the same time holistic assessment of the system transformation and shall contribute to the scientific underpinning of the EU's SET-Plan. Final modelling results shall help to understand and investigate the complex links, interactions and interdependencies between the different actors and technologies within the energy system as well as their impact on society and environment.

All models used within the project have already been used as stand-alone applications. Thus, each model has its own database with already existing data. In the course of the project, **common input data** as well as **model-specific input data** have been defined. Moreover, through model coupling, essential exogenous parameters of the models become endogenous variables of the EMS, i.e. relevant output data of one model serve as input data of another model. It is necessary that collective input variables are harmonized. Therefore, a **common REFLEX database** with a scenario storyline has been developed. This allows for alternative assumptions regarding e.g. the development of macroeconomic parameters, or the impact of fuel prices. Figure 1 illustrates REFLEX data flows.

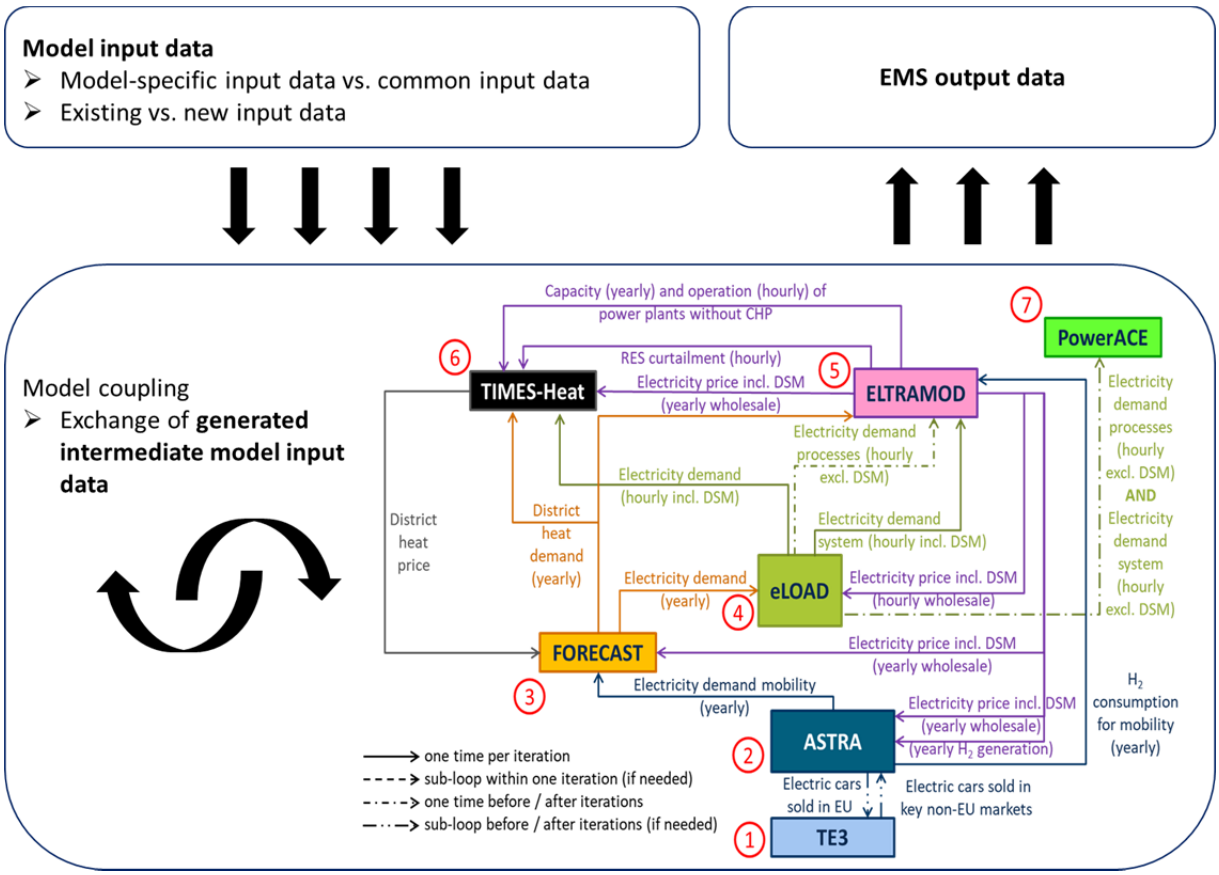

Figure 1: REFLEX data flows

1.1.2 DATA STRUCTURE IN THE REFLEX PROJECT

The results of a model-based analysis depend not only on the chosen methodology, but also on the quality of the data used. For a consistent analysis within the EMS in REFLEX, a common database with harmonised datasets was implemented in a Data Warehouse (DWH).[1]

The database of the REFLEX project contains **four groups of data**, i.e. (1) existing model input data, or so-called background data, (2) collected and generated new model input data, or so-called foreground data, (3) generated intermediate model output data for exchange between the models during the iteration process, and (4) generated final result data of the EMS. Table 1 gives a brief overview on these four groups, which will be described in more detail in the following sections.

Table 1: Groups of data within the database of the REFLEX project

Data group	Description / content
Existing model input data	- Data collected, generated or purchased from commercial providers by a project partner before the start of the REFLEX project - Input data collected, generated or purchased from commercial providers by a project partner in the context of projects on behalf of other clients run in parallel to the REFLEX project
Collected and generated new model input data	- Data collected from publicly available sources or purchased from commercial providers by a project partner in the context of REFLEX - Data collected through surveys conducted by a project partner in the context of REFLEX - Data generated based on existing, new collected or new purchased data by a project partner in the context of REFLEX
Generated intermediate model output data	- Intermediate results of the model applications for data exchange between the models or for further assessments within the project
Generated final result data of the EMS	- Final results of REFLEX generated by model applications, e.g. CO_2-emissions, energy demand, technology impact evaluation, etc.

Regarding model input data we thus can differentiate among existing and new data, as well as among input data collected from publicly available sources, input data generated by project partners in the context of REFLEX, purchased data coming from commercial providers, and confidential data that have been provided to REFLEX project partners after the signature of non-disclosure agreements (see Figure 2). These different types of data sources will also have an impact on our ability to (re-)publish respective datasets in an openly accessible research data repository (see below).

[1] The DWH contains the needed data for the project from several heterogeneous sources in a consolidated/harmonized form and provides tools for data processing optimized for analysis purposes. The DWH is managed by and hosted at a server of the project partner ESA² (Germany).

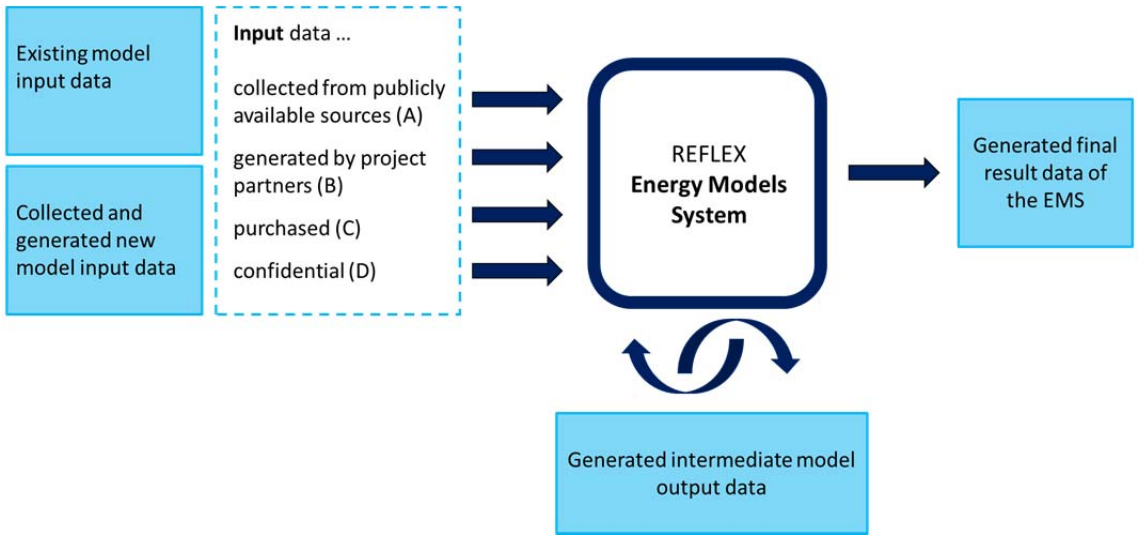

Figure 2: REFLEX data structure

1.2 TYPES AND FORMATS OF DATA COLLECTED AND GENERATED

The following sections describe the four groups of REFLEX data regarding the specific types and formats of datasets collected and generated.

1.2.1 EXISTING MODEL INPUT DATA

Each of models applied in REFLEX has already been used as a stand-alone application, and thus each model has its own database with already existing data input. These datasets originate from own previous work and own assumptions of the project partners as well as from literature and have been developed over many years. Most of these input data are rather model-specific and an unconditional application over several models is limited. Some of the existing data, however, will be reused in the EMS, given that they are up to date, or that no better data are available.

The **REFLEX data warehouse** thereby is the central element of common data storage, use and exchange. It includes

- all data needed for more than one model within the EMS and which therefore have to be harmonized; and/or
- all data needed to validate the results presented in project reports and publications (so called "underlying data").

The harmonization of input data is necessary to ensure a consistent analysis within the EMS. For the same information (e.g. power plants' start-up costs, or efficiency factors) the same dataset (values) have to be used in all models. The consortium has decided which of the existing datasets are applied. These are included in the project's database and provided

to all models before initializing the EMS runs. Table 2 gives an overview on the essential existing model input datasets reused in REFLEX.

Table 2: Essential existing model input datasets reused in REFLEX

Dataset	Time period covered	Spatial scope	Sources
Power plants' availability	2010-2050	EU28	DIW 2013
Power plants' efficiency	2010-2050	EU28	DIW 2013
Power plants' emission factor	2010-2050	EU28	UBA 2014
Power plants' interest rate	2010-2050	EU28	IEA 2010
Power plants' lifetime of investment	2010-2050	EU28	IEA 2010
Power plants' load change costs (depreciation)	2010-2050	EU28	DIW 2013, Traber & Kemfert 2011, own assumptions
Power plants' load change costs (fuel factor)	2010-2050	EU28	DIW 2013, Traber & Kemfert 2011
Power plants' operation management costs (fixed)	2010-2050	EU28	DIW 2013, VGB 2011a
Power plants' operation management costs (variable)	2010-2050	EU28	DIW 2013, Traber & Kemfert 2011
Power plants' specific investment	2010-2050	EU28	DIW 2013
Power plants' start-up costs (depreciation)	2010-2050	EU28	Traber & Kemfert 2011
Power plants' start-up costs (fuel factor)	2010-2050	EU28	Traber & Kemfert 2011
Vehicles CO_2 standard	2010-2050	EU28	EU regulation, own assumptions
Vehicles fuel consumption factors	2010-2050	EU28	GHG-TransPoRD project, ASSIST project

1.2.2 COLLECTED AND GENERATED NEW MODEL INPUT DATA

Some of the needed input data for the EMS have been updated or newly defined according to the research questions and the focus of analysis within the REFLEX project. Furthermore, publicly and commercially available data are used. Additionally, required input data not available in the literature, in existing data repositories or on the market, have been generated by the project partners via empirical surveys and/or appropriate assumptions. These data are included in the project's database, too, and in that way provided as harmonized datasets for all models within the EMS.

The group of collected and generated new model input data includes:

- data for the REFLEX scenario framework;
- data for demand side management as one source of flexibility; and
- data for experience curves to allow for an endogenous modelling of technological learning.

DATA FOR THE REFLEX SCENARIO FRAMEWORK

Data for the REFLEX scenario storylines describe the overall framework for the model-based analysis and include the main macro-economic and societal drivers as well as techno-economic parameters and regulatory conditions of the political environment.

Two main scenarios are distinguished (for a detailed description see Deliverable 1.1: Scenario Description). First, a reference scenario ("Mod-RES") based on observed trends and most recent projections is in line with the PRIMES 2016 Reference Case (Capros et al., 2016). Second, for a more ambitious policy scenario ("High-RES"), framework conditions are similar to those of Mod-RES in terms of population and economic growth, while both fuel- and CO_2 prices are assumed to be higher. More ambitious climate policies are considered. In order to capture the different possible stances on a future energy system, a centralized and a decentralized version are distinguished.

The defined scenario storylines are translated into quantitative model input parameters until the year 2050, which is the defined horizon for the analysis. The data may be aggregated for whole Europe, or disaggregated on a national, sectoral or technological level. The macro-economic trends and the societal drivers are based upon official projections provided by the European Commission. All political assumptions have been elaborated considering current and past policy implementations and are discussed with the European Commission and various stakeholders. Table 3 gives an overview on the data for the REFLEX scenario framework.

Table 3: Data for the REFLEX scenario framework

Dataset	Time period covered	Spatial scope	Sources
Gross domestic product	2000-2050	EU28+NO+CH (NUTS 0 level)	Capros et al. 2016 (assumptions for NO+CH based on other Horizon 2020 projects)
Population	2000-2050	EU28+NO+CH (NUTS 0 level)	Capros et al. 2016 (assumptions for NO+CH based on other Horizon 2020 projects)
Price electricity (initial average costs of gross electricity generation)	2000-2050	EU28+NO+CH (NUTS 0 level)	Capros et al. 2016
Price fossil energy carriers	2015-2050	EU28	Capros et al. 2016
CO_2-price ETS	2015-2050	EU28	Capros et al. 2016, own assumptions

DATA FOR DEMAND SIDE MANAGEMENT

Relevant data for investigating system flexibility in the form of demand side management (DSM) are rarely available from public or commercial sources. In particular, the available database for the tertiary sector with regard to DSM is highly incomplete. Therefore, an empirical survey on DSM in the tertiary sector has been conducted with the aim to improve the model input data and to fill data gaps (see Deliverable 2.2 Report on Survey Findings). Based on the analysis of the collected specific empirical data, existing datasets have been extended as well as new datasets generated.

The design of the survey has been established by the REFLEX partners. The survey itself has been conducted for 4 European countries by an international market research institute. Empirical data for one further country (Germany) was provided in-kind by REFLEX partner Fraunhofer ISI. DSM data for further countries are deduced from the survey results. Relevant model input parameters for modelling DSM options which should be deduced from the empirically ascertained data are given in Table 4.

Table 4: Data for modelling DSM options

Dataset	Time period covered	Spatial scope	Sources
DSM potential (share of flexible load per energy usage process)	2015-2050	EU28+NO+CH (NUTS 0 level)	empirical survey, public and commercial sources
DSM costs (activation costs per energy usage process)	2015-2050	EU28+NO+CH (NUTS 0 level)	empirical survey, public and commercial sources
DSM time of interfere (maximum load reduction time)	2015-2050	EU28+NO+CH (NUTS 0 level)	empirical survey, public and commercial sources
DSM number of interventions (frequency of DSM measures)	2015-2050	EU28+NO+CH (NUTS 0 level)	empirical survey, public and commercial sources
DSM shifting time (allowed points of time or time periods/frames for DSM)	2015-2050	EU28+NO+CH (NUTS 0 level)	empirical survey, public and commercial sources

DATA FOR EXPERIENCE CURVES

In order to enable an endogenous modelling of technological developments and resulting production cost reductions, experience curves for the most relevant technologies for each sector (electricity including storage and power-to-X, industry, mobility, heating/cooling, energy end-use) have been developed in a first step (see Deliverable D3.2 Comprehensive Report on Experience Curves). In a second step, they have been implemented in the sectoral models of the EMS.

Special attention is given to the determination of uncertainty ranges of progress ratios (i.e. the slopes of the experience curves), as these can have a major impact on modelling results, especially for long-term modelling up to the year 2050. Especially for technologies that depend strongly on either the available geographical potential (e.g. wind onshore, offshore) or on raw material prices, a decomposition of the experience curve using a multi-level

experience curve is performed. This allows the determination of the most important factors behind cost development, such as variations in steel or oil prices, as well as scale effects.

The needed empirical data for defining experience curves has been collected by means of interviewing industry experts, conducting specific survey methods and analysing detailed statistics (e. g. construction, production and consumer price indices as well as installed capacities and cost developments in the electricity, heat and mobility sectors). Table 5 gives an overview on the technologies for which experience curves have been developed.

Table 5: Technologies by sector for which experience curves are developed

Category	Technology
Electricity generation	CCS (membrane, oxyfuel, pre-combustion, post-combustion)
	Photovoltaics: modules (mono/poly, CdTe), BOS, systems
	Wind onshore
	Wind offshore
	Fuel cell micro-CHP
Electricity storage	Battery: Lithium-ion (utility, residential)
	Battery: Redox-flow
Heating/cooling	Heat pump (air/water)
Industry	Industrial heat pumps (large scale)
	Industrial heat/steam (Industrial CCS)
Mobility	Battery electric vehicles (EV Lithium-ion battery)
	Fuel cell vehicles (EV fuel cell stack)
	Hybrid electric car (HEV NiMH battery)
Power-to-X	Power-to-hydrogen (alkaline electrolysis)

To estimate the potential of alternative fuel technologies in Europe, the global automotive market (especially including North America and Asia) is considered. The analysis is focused on major driving patterns. The reason for analysing global passenger car markets is to identify the global market penetration of electric vehicles, which in turn will influence the demand for Li-ion batteries substantially and thus will have a crucial impact on technological learning. This information helps to assess the future prices of batteries and fuel cells based on the learning curve theory.

1.2.3 GENERATED INTERMEDIATE MODEL OUTPUT DATA

By coupling the different approaches of the REFLEX partners, the system boundaries of each stand-alone model are partly disbanded and most **exogenous parameters of each model become endogenous variables of the EMS**. This is done by using the relevant output data of one model as input data of another model.

The REFLEX database, which was developed in the course of the project, facilitates the data exchange between the models and stores relevant input and output datasets. Technical routines are thereby able to overcome several compatibility issues, i.e. models may use different levels of detail or aggregation, time intervals of the same variables may differ, and finally, different models use different identifiers for the same datasets. Therefore, when importing the intermediate results into the database, a mapping and any necessary value

aggregations or data splits take place at the same time. For this purpose, each data table contains the identifier structure used for both, the model from which these intermediate results originate and for other models using these intermediate data as own input data. To achieve a stable final state of the EMS within each REFLEX scenario storyline, several iterations with all models are performed.

Table 6 gives an overview on main input and output variables of the different models coupled to the EMS. It shows which input is exogenous if the respective model is used as a stand-alone application but becomes an endogenous input dataset in the EMS.

Table 6: Main inputs and outputs of the different models

Model	Needed input i. a. (exogenous if stand-alone application)	Endogenous in EMS? (If yes, provided by)	Provided output i. a.
ELTRAMOD	Electricity demand	Yes (eLOAD)	Electricity prices Capacity and operation of power plants RES curtailment
	Techno-economic data for power plants	No	
	H2-demand for mobility	Yes (ASTRA)	
	Fuel prices	No	
	Heat demand	Yes (FORECAST)	
TIMES-Heat	Heat demand	Yes (FORECAST)	Capacity and operation of heat plants
	Capacity and operation of power plants	Yes (ELTRAMOD)	
	Electricity wholesale prices	Yes (ELTRAMOD)	
	Techno-economic data for power plants	No	
TE3	Macro-economic framework data	No	Used transport technologies in global key markets
	Techno-economic data for vehicles	No	
	Fuel prices	No	
ASTRA	Electricity prices	Yes (ELTRAMOD)	Used transport technologies and energy demand for mobility in EU H2-demand for Mobility
	Macro-economic framework data	No	
	Techno-economic data for vehicles	No	
	Fuel prices	No	
FORECAST	Electricity wholesale prices	Yes (ELTRAMOD)	Yearly energy demand by sector
	Electricity demand mobility	Yes (ASTRA)	
	Macro-economic framework data	No	
	Techno-economic data for demand side technologies	No	
	Fuel prices	No	
eLOAD	Electricity wholesale prices	Yes (ELTRAMOD)	Electricity demand structure (load profiles)
	Electricity demand (yearly)	Yes (FORECAST)	

PowerACE	Electricity demand	Yes (eLOAD)	Capacity and operation of power plants under different market designs
	Framework data for energy markets	No	
	Techno-economic data for power plants	No	
	Fuel prices	No	
eLCA/ sLCA/ πESA	Capacity, operation and emissions of power plants	Yes (ELTRAMOD)	Emissions Impacts on humans and environment
	Capacity and operation of heat plants	Yes (TIMES-Heat)	
	Energy / electricity demand	Yes (FORECAST, eLOAD, ASTRA, TE3)	

Table 7 summarizes datasets of generated intermediate model output data for the data exchange within the EMS.

Table 7: Generated intermediate model output data for data exchange within the EMS

Dataset	Time period covered (10-year steps)	Spatial scope
Price electricity (hourly)	2015-2050	EU28+NO+CH (NUTS 0 level)
Demand electricity (hourly)	2015-2050	EU28+NO+CH (NUTS 0 level)
Demand electricity for mobility (yearly)	2015-2050	EU28+NO+CH (NUTS 0 level)
Demand district heating (yearly)	2015-2050	EU28+NO+CH (NUTS 0 level)
Power plants installed capacity and operating (yearly and hourly)	2015-2050	EU28+NO+CH (NUTS 0 level)
Power plants emissions (yearly)	2015-2050	EU28+NO+CH (NUTS 0 level)
Power plants demand energy (yearly)	2015-2050	EU28+NO+CH (NUTS 0 level)
Mobility demand energy (yearly)	2015-2050	EU28+NO+CH (NUTS 0 level)
Mobility emissions (yearly)	2015-2050	EU28+NO+CH (NUTS 0 level)

1.2.4 GENERATED FINAL RESULT DATA OF THE EMS

After achieving a stable state based on several iterations with all models within the EMS for each REFLEX scenario storyline, the result data of the different models are collected and combined within the project's database to the final result data of the EMS. These data are analysed to derive the key findings on the impacts of technological development and innovation on the energy system and on the environment, society and economy and are the basis for answering the research questions of the REFLEX project (see Administrative details). Table 8 gives an overview of the major generated final result data of the EMS.

Table 8: Major generated final result data of the EMS

Dataset	Time period covered (10-year steps)	Spatial scope
Price electricity average yearly	2015-2050	EU28+NO+CH (NUTS 0 level)
Demand electricity	2015-2050	EU28+NO+CH (NUTS 0 level)
Demand district heating	2015-2050	EU28+NO+CH (NUTS 0 level)
Power plants installed capacity	2015-2050	EU28+NO+CH (NUTS 0 level)
Power plants operation	2015-2050	EU28+NO+CH (NUTS 0 level)
Power plants emissions	2015-2050	EU28+NO+CH (NUTS 0 level)
Mobility demand energy	2015-2050	EU28+NO+CH (NUTS 0 level)
Mobility emissions	2015-2050	EU28+NO+CH (NUTS 0 level)
Life cycle environmental and resource impacts	2015-2050	EU28+NO+CH (NUTS 0 level)
Life cycle human health (damage / toxicity)	2015-2050	EU28+NO+CH (NUTS 0 level)
Life cycle societal impacts (risk level)	2015-2050	EU28+NO+CH (NUTS 0 level)
Costs external	2015-2050	EU28+NO+CH (NUTS 0 level)

1.3 DATA UTILITY

As elaborated above, REFLEX collects and generates a certain amount of research data. On the one hand, these data are necessary to meet the objectives of the project and to answer the research questions. On the other hand, most of the collected and generated data will be useful for further research by the project partners themselves or by third parties, as well as for stakeholders in the energy industry and for policy makers.

1.3.1 EXISTING MODEL INPUT DATA

During the last years, comprehensive databases have been created by the project partners in order to answer a wide range of research questions based on their individual models in stand-alone applications. These data have been collected from a variety of sources, including institutional reports (e.g. from the EC or IEA), scientific publications (e.g. published by Elsevier) and data documentations from third modelling works (e.g. from the DIW Berlin). Moreover, a part of the data has been generated based on previous own modelling exercises as part of previous research projects and own assumptions.

As these existing model input data are essential underlying data for the REFLEX model runs, their publication ensures the transparency of the generated results and allows third parties to compare project outcomes to other studies with a similar scope. Moreover, the collection of comprehensive datasets then re-published in one single place and in one single re-usable format, e.g. related to the variety of power plant parameters, facilitates future research as searching the various data coming from various sources and having been published in various formats (pdf, data-files, etc.) will not be necessary again.

1.3.2 COLLECTED AND GENERATED NEW MODEL INPUT DATA

In what follows, we distinguish among the three categories for collected and generated model input data introduced above, i.e. data for the REFLEX scenario framework, data for DSM and data for experience curves.

DATA FOR THE REFLEX SCENARIO FRAMEWORK

The collected and prepared data for the scenario framework regarding macro-economic and societal drivers, techno-economic parameters and regulatory conditions are tailored to the aim and scope as well as the specific research questions of the REFLEX project and the applied models to answering them. They are thus primarily useful as underlying data to ensure the transparency of the generated results and the comparability of the project outcomes to other (existing and future) studies with similar analysing scope.

Their publication as open research data will also allow a critical discussion of the impact of scenario framework assumptions on model outcomes. Whereas the forecasts of the development of GDP and population have been relatively stable over time, this has not been the case for fuel prices, and even less for the level of costs related to CO_2 emissions. The data will be used for updating the existing databases of the different models and for further model-based research of the project partners outside of REFLEX.

DATA FOR DEMAND SIDE MANAGEMENT

Parameters related to different DSM measures have been generated in the course of the REFLEX project based on elaborate empirical surveys. With the collected specific empirical data, the database for investigating system flexibility by DSM has been improved substantially because relevant data – especially for the tertiary sector – have been rarely available from public and commercial sources. Thus, existing, highly restricted datasets have been extended and at the same time new datasets have been created.

These data have a high potential for re-use after the end of the REFLEX project. The survey data will not only be a valuable resource for future research activities in general, but it will also allow the identification of promising energy applications and DSM potentials in the selected sector in different European countries. To date, this is a very topical issue as European electricity systems are evolving towards a generation mix that is more decentralized and less predictable with additional flexibility being expected to be provided by the demand side. This implies that – also small-scale – consumers must be shifted from the current ‘passive’ role to providing ‘active’ demand response, and thus, new business models valorizing flexibility provided by the tertiary sector and adaptations in market design and regulation are required. REFLEX research data on DSM are therefore useful not only for the research community, but also for stakeholders from industry and policy making.

DATA FOR EXPERIENCE CURVES

Endogenizing technological learning through experience curves allows for an enhanced assessment of the evaluation of impacts from policy measures or alternative incentive schemes on realizable future cost reduction. In addition, in view of current rapid and necessary changes in energy systems (driven partially by policies and partially by markets) and the ensuing need for flexibility, the endogenous modelling of cost development of existing and new innovative energy-related technologies in bottom-up models will become even more important.

However, the data and experience curves required to do so have not been available when the REFLEX project started. A review of certain energy supply and demand technologies has been published by Junginger et al. (2010), but these required an updating. More recent studies have been published for some individual technologies (e.g. Bolinger and Wiser 2012; Candelise et al. 2013 or Chen et al. 2012). However, a comprehensive and up-to-date overview was missing. Especially with regard to technologies needed for increasing the flexibility in energy systems (such as storage technologies or DSM-devices) little or no experience curves had been published. Thus, to advance the energy models included in REFLEX beyond the state-of-the-art, data collection was required to devise or update experience curves for existing technologies and to estimate experience curves for new ones.

Furthermore, it will require smart and innovative incorporation and interlinkage of these experience curves in various sectoral energy models to comprehensively assess the effects of technological learning and the demand for increased flexibility in energy systems.

The collected data have a high potential for re-use after the end of the REFLEX project. The outcome – a comprehensive state-of-the-art and up-to-date overview of experience curves and underlying databases – will be of high value for other energy models outside the project (developed in the EU as well as worldwide) in order to meet the challenges of modelling our changing energy systems with increasing penetrations of innovative (improved and new) technological solutions for the coming decades.

1.3.3 GENERATED INTERMEDIATE MODEL OUTPUT DATA

These data are only intermediate results of the EMS and are transferred between the models during the iteration processes. A relevant benefit of these data for further applications outside the framework of REFLEX is not expected.

1.3.4 GENERATED FINAL RESULT DATA OF THE EMS

The overall objective of REFLEX is to support the EU's SET-Plan by strengthening the knowledge base for transition paths towards a low-carbon energy system based on a cross-sectoral analysis for the entire energy system of the EU. Due to the complexity of this system, it is obvious that the implementation of the SET-Plan requires in-depth knowledge on the interrelationship between the different sectors (electricity, heat and mobility), energy

technologies but also on the interdependencies between energy and non-energy industries, environment (beyond greenhouse gas emissions) and society.

The result data of the EMS within REFLEX help to understand and investigate the complex links, interactions and interdependencies between the different actors and technologies within the energy system as well as their impact on society and environment. Based on the EMS result data, recommendations for effective strategies for a transition of the European energy system to a low-carbon system can be derived. Policy makers at EU level as well as at regional level can use these findings when developing policy measures. Furthermore, the generated final result data of the REFLEX project can be used as a reference or starting point for further research work by the project partners or third parties on the future design of the energy system of the European Union.

1.4 DATA PROTECTION AND EXPLOITATION STRATEGY

In order to ensure efficient dissemination and exploitation activities, free of any legal conflicts, the REFLEX project partners signed a Consortium Agreement (CA).[2] The CA is, among other things, dealing with the details on the partners' background data and on the rights to, the protection and the exploitation of pre-existing datasets and results generated solely and/or jointly during the lifetime of the project. Moreover, the CA sets up specific rules on how to deal with dissemination activities (see also Deliverables D7.3 and D7.4 Dissemination and Communication Plan) and to ensure open access to peer-reviewed scientific publications.

The following basic CA rules regarding data protection and exploitation apply:

- All partners define their individual existing background data required for their successful participation in the project. Background data are own and/or commercial model input data generated/purchased before the start of the project. The rights to these data remain with the respective partner, but royalty-free access to the others is granted if not restricted by third parties, and if it is required to enable the research activities in the context of the project.
- Data that are acquired by individual partners during the project without using REFLEX funds, e.g. in the context of projects on behalf of other clients run in parallel, will be also treated as pre-existing background data.
- The property rights to data collected and datasets/results generated during the project by using REFLEX funds (foreground data) belong to those partners involved in the collection and generation processes. When more than one consortium member is involved, the dataset will be jointly owned by the respective consortium members.

[2] For the REFLEX consortium agreement, the DESCA template version 1.0 (www.desca-2020.eu) has been applied.

Dissemination and exploitation of data and results is executed in accordance with EU laws and with respect to specific laws in the participating countries. Before any dissemination activity takes place, respective legal aspects are examined and clarified. This is particularly the case for data from the DSM-survey and for experience curve data as well as data purchased from commercial providers. The possibility for protection of generated results within the project (consortium) is also examined before publication. All participants have departments specifically devoted to managing intellectual property. These departments manage the relevant protection processes.

Within REFLEX, the dissemination and exploitation of results not only in terms of knowledge and insights, but also in terms of data, is coordinated by the Exploitation and Innovation Manager (EIM) regarding knowledge management and innovation activities. Thus, and as specified in the Grant Agreement, the EIM is responsible for:

- maintaining a registry of relevant background data,
- maintaining a registry of foreground data gathered and generated in the work packages during the project,
- assessing the opportunities for exploitation, for example by following political events in the energy sector or searches of other scientific databases for similar developments, and
- proposing specific exploitation measures, e.g. policy briefs and events.

During the project, periodic analyses of transfer opportunities to adjust the exploitation strategies take place. Thereby, the EIM can identify synergies to ensure the best and suitable use and exploitation of results. All consortium partners contribute to the exploitation plan of the project throughout its life span. The EIM is in close contact and regularly informed about the exploitation plans of the partners and regularly advises the consortium and individual partners about possible strategies.

The exploitation strategy related to REFLEX research data is outlined below:

- First, it is decided whether to (re-)publish a dataset in the REFLEX data repository or not, and if yes in what way (data format, metadata provided, open or restricted access).
- Participants inform the EIM and other consortium members if they wish to publish or disseminate any datasets, whether in a direct way or indirectly.
- Before any dissemination activity takes place, the participants examined the possibility of protecting generated results, for instance with regard to potential reuse for commercial purposes.
- In case of collected input data: It is examined whether the rights of third parties are affected and, if necessary, their consent to the re-publication of these data is obtained.
- Upon (affirmative) dissemination decision, the dataset is made available (regarding different dissemination types see Sections 2.2 and 2.4).

2 FAIR DATA

Collected and generated data relevant for EMS model runs are implemented in the developed REFLEX database in a standardized way. The database is implemented in postgreeSQL. For managing the database, the management-tool "pgADMIN" and a proprietary developed interface tool is used, which provides several data-preparing and identifier-mapping functions.

A selection of existing data as well as data collected and generated during the project will be made available to interested research groups and stakeholders from policy and industry. The following sections outline how the data are exploited and made accessible for verification and re-use and how data will be curated and preserved upon closure of the project.

2.1 MAKING DATA FINDABLE – REFLEX DATA REPOSITORY

For making data findable, a data catalogue – the so-called **REFLEX data repository** – has been prepared. The consortium will continue to provide the data via the REFLEX project website for a limited period of time after the end of the REFLEX project (<18 months). The project website will be maintained during this period (only ensuring online accessibility).

For long-term data provision, the following approaches (or a combination of them) are currently evaluated:

i) The data remain in the DWH of the project partner ESA². The data provision is transferred to the website of the ESA² Company (http://www.esa2.eu). There a reference to the project will be created and access to the data will be enabled.

ii) The data are published in the ZENODO repository (https://zenodo.org/). This is an online, free of charge storage created through the European Commission's OpenAIREplus project and is hosted at CERN, Switzerland. It encourages open access deposition of any data format, but also allows deposits of content under restricted or embargoed access.

iii) The data are published in the inter-disciplinary OpARA repository (https://tu-dresden.de/zih/forschung/projekte/opara). This is an online, long-term repository for research data, hosted by Technische Universität Dresden.

iv) The data are published in the "OpenEnergy Platform" (https://wiki.openmod-initiative.org/wiki/Proposal_for_the_OpenEnergy_Platform). This platform is still under development and aims to expand the existing "OpenMod" online presence by, amongst others, offering a place to store and exchange data (raw data and processed data), which are needed for modelling works.

Approach i) is preferred. For the case that a permanent provision via the website of ESA[2] should no longer be possible at any time, option iv) is preferred. The structure of the REFLEX data has already been designed for this option. Additionally, this platform offers the possibility to implement information about the REFLEX project. In any case the data will be available for third parties after the end of the project with free access to metadata and the open access data contents. The length of time for which the data will remain re-usable is not restricted.

The **REFLEX data catalogue** is currently implemented on the REFLEX project website (http://reflex-project.eu/public/data-publication/) with a simplified preliminary frontend version which has a format similar to Table 9 below. The catalogue thus gives a comprehensive overview on all datasets available in the data repository and at the same time allows users to access metadata sheets with detailed information on the content and scope of a specific dataset directly. Moreover, a short-link for downloading the individual dataset in a user-friendly format is provided.

Table 9: Data catalogue in the REFLEX data repository

Dataset	ID	Version	Description	Download
[dataset name and link to metadata sheet]	[dataset identifier]	[version number, i.e. year_month_day]	[brief description]	[direct link for download]
Dataset 1				
Dataset 2				
Dataset 3				
Dataset 4				
…				

The frontend-concept of the final version is shown in Figure 3. This is currently under construction and will give a user-friendly overview with a content related hierarchic (tree) structure about the available data. By clicking on a data set, the associated metadata and a data preview are displayed and the data set can be downloaded.

The scope and design of the metadata has been oriented on the metadata structure of the "Open Power System Data" platform (http://www.data.open-power-system-data.org/). Table 10 shows the **REFLEX metadata sheet**. It includes not only general descriptive issues (e.g. name, description, version number, etc.), but also detailed information on its scope (time, sectoral, spatial) as well as administrative matters (e.g. recommended text for attribution or the licensing of the dataset). Each dataset can be unambiguously identified via the combination of dataset name and the version label. Both will be included in the unique dataset ID.

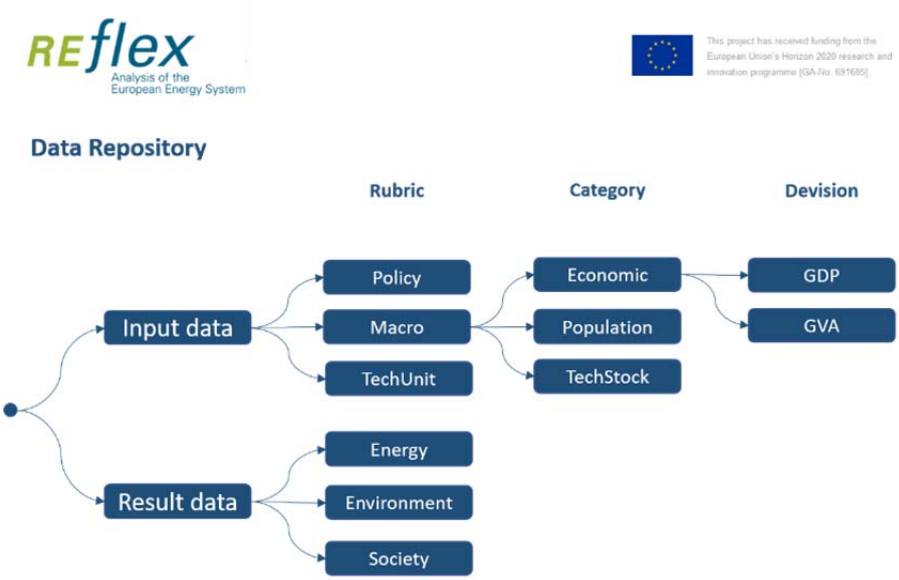

Figure 3: Concept of the frontend for the REFLEX data repository

Table 10: Scope and contents of the metadata for a provided dataset

Category	Content
Name	Name of the dataset (a concise one, short but informative)
ID	Dataset identifier (unique dataset ID including the dataset name and version number)
Version	Year_month_day (e.g. 2015_04_21)
Keywords	List of keywords
Description	Short description of the dataset
Remarks	Specific remarks (e.g. restrictions, data gaps)
Timescale	Time period covered (e.g. 2005 to 2030) and timesteps (e.g. hourly, yearly, 5-year steps)
Spatial scope	Countries/regions covered with scope and level of differentiation or aggregation (e.g. EU-28, national data, list of countries)
Sectoral scope	Sectors covered (e.g. households, industry, traffic) as well as sub-categories if available (e.g. road traffic, rail traffic, aviation)
Sources and input data	Used sources to prepare/provide the dataset, if possible with links to the primary data
Attribution	Recommended text for attribution
Contact	Contact information for questions/remarks
Access	Terms of data access/usage, ideally a standard license
Download	Here link to the dataset for download, e.g. CAP_installed.xls
Field documentation	Link to extra file/page [see Table 11]

The field documentation as last item of the metadata sheet and being fully displayed on a separate page/window includes the complete list of variables within the dataset with the subcategories field name, -type, -unit, and -description (see Table 11).

Table 11: Field documentation

Field name	e.g. CAP_inst	…	…	…	…
Field type	e.g. Number	…	…	…	…
Field unit	e.g. MW	…	…	…	…
Field description	e.g. Installed capacity end of the year	…	…	…	…

With regard to the publication and transparency of results of the project work (e.g. in scientific journals, book chapters, conference proceedings or policy briefs), the preparation of tailored dataset packages for the different publications is planned. Such a package contains, together with a short content description, a compilation of (1) the metadata of the published results, and (2) the metadata of the relevant datasets required to verify the results, as long as these can be made available.

2.2 MAKING DATA OPENLY ACCESSIBLE – REFLEX OPEN DATA

At the beginning of the REFLEX project, three different possibilities for data dissemination have been considered by the project partners, i.e.

a) *Open access data publication*: In this case, data owners grant a royalty-free access of a meaningful selection of generated results to other participants and to the public, possibly restricted by appropriate embargo periods and/or respecting restrictions from editors of scientific journals and organizers of conferences.

b) *Commercial data exploitation*: In this case, data suitable for commercial exploitation (e.g. for a commercial re-use by consulting companies) will be managed by the project partner ESA[2], with the explicit purpose of exploiting research results (including research data) related to (coupled) energy systems modelling.

c) *Indirect data publication*: In this case, parts of the generated data are disseminated only indirectly as part of intermediate or final results of models and/or as qualitative outcome based on post-analysis of results.

The following discussions concentrate on direct data publication. Both, input data collected from publicly available sources (existing and new model input data) as well as data generated via the project partners' application of the different simulation models are made available as open data as far as possible according to the guidelines of the EU. Thereby, any legal conflicts have to be avoided (see Section 2.5 for details). The same holds for generated final results of the EMS (see Table 12).

The project partners are in discussion to make the results of the DSM-survey and the generated different DSM-datasets openly available for research purposes with added value for society. However, the commercial use by third parties shall be subject to explicit permission and an adequate royalty payment.

A similar discussion is taking place for data on experience curves. The required data for technological learning curves have been collected by means of interviewing industry experts, conducting specific survey methods and analysing detailed statistics (e. g. construction, production and consumer price indices as well as installed capacities and cost developments in the electricity, heat and mobility sectors).

Model input data purchased by the project partners, in contrast, cannot be re-published as open data due to the data usage conditions of the commercial providers. Finally, the same holds for confidential data on fuel price developments provided by the European Commission (Capros et al. 2016), for which non-disclosure agreements have been signed.

Table 12: REFLEX data published in the data repository (the right to make changes is reserve)

Data group	Data source	(Re-)publication in the REFLEX data repository?
Existing model input data	Data collected from publicly available sources	**YES** – if no legal conflicts
	Data generated by project partners	Model outcomes: **YES**
	Purchased data	**NO**
Collected and generated new model input data	Data collected from publicly available sources	**YES** – if no legal conflicts
	Data generated by project partners	Model outcomes: **YES** Data for DSM: **YES, but** restricted access[1] Data for experience curves: **YES, but** restricted access[1]
	Purchased data	**NO**
	Confidential data	**NO** – NDAs have been signed
Generated intermediate model output data	EMS models	**NO** – transfer between models during iteration process, no relevant benefit for further applications
Generated final result data of the EMS	EMS models	**YES**

[1] Open and royalty-free access restricted to non-commercial use

In the case of an *Open Access Data Publication*, the dataset can be easily downloaded in a standard format. The planned standard format for data (re-)publication is ".csv". The download links for different formats are given within the metadata sheet in the category "Download".

In case of *Commercial Exploitation* of a dataset, a registration procedure for all those interested in such datasets is an option. This includes the opportunity to differ the conditions for access depending on the type of the inquirer or planned re-usage (e.g. dataset being free of charge for public scientific institutions, but subject to a charge in case of commercial reuse by a company). After registration of the request of a dataset, a time-limited download link is provided via e-mail to the registered contact together with the terms of usage and, as the case may be, with the invoice. The requisition could be implemented in the respective metadata sheets, which are available free of charge in any case, in the category “Access”.

The datasets are made available to third parties as soon as they are generated, prepared and reviewed for publication/commercial exploitation and when the conditions of dissemination are decided and possible protections of datasets are clarified within the consortium. However, additional restrictions by setting appropriate embargo periods and/or respecting restrictions from editors of scientific journals and organizers of conferences are also possible. A generally valid statement regarding the embargo periods is not possible at the moment. It can differ from case to case.

2.3 MAKING DATA INTEROPERABLE

To increase the interoperability of provided data, commonly used vocabularies for the metadata contents as well as for the identifiers and the contents of the identifiers within the datasets will be applied. These include standardized name conventions and codes used in official statistics (e.g. for countries, regions etc.). Furthermore, specific energy system topics related name conventions will be orientated on the “Open Power System Data” platform (http://www.data.open-power-system-data.org/). An additional mapping procedure or the provision of mapping tools for data users is not envisaged.

2.4 INCREASE DATA REUSE – LICENSING OF REFLEX DATA

Intellectual property rights on databases are subject to different pieces of legislation. First, copyright applies to a wide range of creative, intellectual and artistic works – including data. Second, in 1996, the EU adopted the Database Directive (96/9/EG), which had to be implemented into national law by the Member States. Moreover, rights and duties resulting from data privacy law or existing licenses on datasets have to be respected.

In order to increase data re-use for REFLEX research data and avoid any legal ambiguities for data users, the (re-)publication of REFLEX data goes hand in hand with a clear licensing procedure. In what follows, we discuss different types of licenses available (i.e. standard vs. custom; open vs. restricted), give a more detailed overview on standardized licenses, describe how the data are licensed in order to permit the widest reuse possible, and elaborate on data quality assurance.

2.4.1 TYPES OF LICENSES

A license is a contract by which the licensor (in our case the data owner) allows the licensee (the data user) to use otherwise protected material. Thus, it clarifies the conditions under which data can be reused. In contrast, in the absence of a license, the data owner still retains proprietary copyright (see also Section 2.5 for further elaborations thereon).

A first distinction can be made among customized and standardized licenses. Drafting a new **customized license**, allows tailoring it to the individual needs and desires of the partners. However, it might leave issues open, and thus might in the end leave data users with default copyright (i.e. no open data). Moreover, it might be unclear or ambiguous on certain issues. An established, **standardized license**, in contrast, has the advantages of content and scope being well understood, and the legal text being rigorously written and suitable for different legal systems. The REFLEX project partners, therefore, agreed on the use of standardized licenses for datasets published in the REFLEX data repository.

A second distinction can be made among open data and non-open data licenses. Thereby, "**open data** […] can be freely used, modified, and shared by anyone for any purpose" (see also https://opendefinition.org/).[3] More specifically, data to be classified as open data must be "available under an open license; available as a whole and at no more than a reasonable one-time reproduction cost, preferably downloadable via the internet; and it must be provided in a convenient and modifiable form, machine-readable, available in bulk and provided in an open format" (see also https://open-power-system-data.org/legal).[4]

Three main types of **open data licenses** – with increasing obligations for users – exist. These are public domain licenses under which users are completely free to use, modify and share the data; attribution licenses which demand users to appropriately credit the dataset creator; and finally share-alike licenses which in addition to the ones above also condition that any derivative work is made available under the same license.

Licenses for non-open data – with increasing restrictions for users – include no-derivatives licenses where it is only allowed to copy, distribute and display original copies of the work and any modification is subject to explicit permission; as well as non-commercial use only licenses and different combinations thereof. Figure 4 illustrates the different types of licenses distinguishing among open and non-open data.

[3] Open Definition is an initiative that sets out principles that define "openness" in relation to data and content.

[4] Open Power System Data is an initiative that did set up an open platform for data required by energy system models.

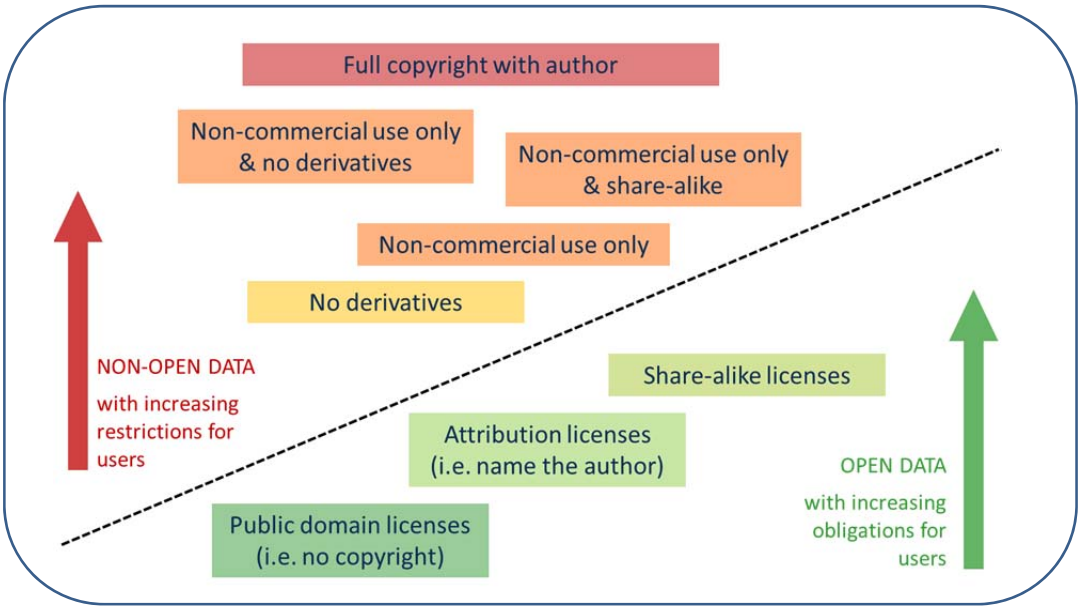

Figure 4: Different types of licenses – open vs. non-open data

Several initiatives developed sets of standardized licenses for specific works such as artwork in general (e.g. Creative Commons), software (e.g. MIT), or even explicitly for data publication (Open Data Commons). In what follows, we give an overview on the different standardized licenses available including examples.

- **Public domain licenses** – ZERO licenses (i.e. no copyright): Users are completely free to use, modify and share the data, even for commercial purposes and all without asking permission.

  Examples: Creative Commons Zero (CC0 1.0); Open Data Commons Public Domain Dedication and Licence (PDDL 1.0); Data licence Germany - Zero 2.0 (dl-de-2-0).

- **Attribution licenses** – BY licenses (i.e. name the author): Users are completely free to use, modify and share the data, even for commercial purposes and all without asking permission. The only requirement is the adequate attribution, i.e. name of the data provider and references to the license and dataset.

  Examples: Creative Commons Attribution (CC-BY 4.0); Open Data Commons Attribution (ODC-BY 1.0); Data licence Germany – attribution – 2.0 (dl-de/by-2-0).

- **Share-alike licenses** – SA licenses: Users are completely free to use, modify and share the data, even for commercial purposes and all without asking permission. The only requirements are an adequate attribution (see above) plus that in case a database is amended or modified, it shall be published under the same license as the original one.

  Examples: Creative Commons Attribution-ShareAlike 4.0 (CC BY-SA 4.0); Open Data Commons Open Database License (ODbL 1.0).

- **No derivatives license** – ND licenses: It is allowed to copy, distribute and display only original copies of the work. Any modification is subject to explicit permission.

  Example: Creative Commons Attribution-NoDerivs (CC BY-ND 4.0).

- **Non-commercial use only license** – NC licenses: It is allowed to use, copy, distribute, modify and display the work for non-commercial purpose only. Any commercial use is subject to explicit permission.

  Example: Creative Commons Attribution-NonCommercial (CC BY-NC 4.0).

- **Different combinations**: Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND 4.0); Creative Commons Attribution-NonCommercial-ShareAlike (CC BY-NC-SA 4.0).

2.4.2 LICENSING REFLEX DATA

The REFLEX project partners decided to apply the widely used **Creative Commons licenses** for the (re-)publication of REFLEX datasets (see Table 13). For data collected from publicly available sources (existing and new) as well as for input data generated by the project partners' modelling works, the open CC-BY license has been chosen, providing open access to any user for any use on the single condition of an adequate attribution to the original dataset. The same holds for generated final project result data of the EMS.

For generated data for demand side management as well as experience curves the open CC-BY-NC license has been chosen. It restricts open access to non-commercial use only. Any commercial use will be subject to explicit permission and the payment of royalty fees to be specified in the respective agreement between the REFLEX representative and the commercial party.

The REFLEX project partners refrained from using share-alike licenses as their application to a certain extent restricts future users of a dataset in the re-publication of processed data. Especially if he/she might rely on several share-alike licenses, these might be incompatible and leave the user with a potential conflict.

Table 13: Draft of REFLEX data licensing (the right to make changes is reserved)

Data group	Data source	General license type	Standard license chosen
Existing model input data	Data collected from publicly available sources	**Open data** with attribution	Creative Commons Attribution **(CC-BY 4.0)**
	Data generated by project partners	Model outcomes: **Open data** with attribution	Creative Commons Attribution **(CC-BY 4.0)**
	Purchased data	-	-
Collected and generated new model input data	Data collected from publicly available sources	**Open data** with attribution	Creative Commons Attribution **(CC-BY 4.0)**
	Data generated by project partners	*Model outcomes:* **Open data** with attribution *Data for DSM:* **Restricted access** for non-commercial use only[1] and with attribution *Data for experience curves:* **Restricted access** for non-commercial use only[1] and with attribution	*Model outcomes:* Creative Commons Attribution **(CC-BY 4.0)** *Data for DSM:* Creative Commons Attribution-NonCommercial **(CC BY-NC 4.0)** *Data for experience curves:* Creative Commons Attribution-NonCommercial **(CC BY-NC 4.0)**
	Purchased data	-	-
	Confidential data	-	-
Generated intermediate model output data	EMS models	-	-
Generated final result data of the EMS	EMS models	**Open data** with attribution	Creative Commons Attribution **(CC-BY 4.0)**

[1] Any commercial use subject to explicit permission.

In theory, the **compatibility of different licenses** could be an issue (see e.g. www.github.com). When processing data and publishing results, the license(s) of input data need(s) to be respected. Typical questions that might arise include: Can data published under license A be merged with other data published under license B? What license could be applied to such a derived or aggregated dataset? Are there any provisions associated with the license of an input dataset that constrain the creation and publication of derivations? Precisely, a data user might face the conflict of two input datasets being licensed under two different, incompatible share-alike licenses. For the REFLEX project no issues related to incompatibility of licenses occurred.

The REFLEX project partner **ESA[2] is responsible for data management and (re-) publication.** For datasets collected from project-external sources, an agreement on the use and republication of the respective data will be signed between the data owner and ESA[2] (see Section 2.5 and Figure 6 below for a detailed discussion of the proceeding). For data generated by REFLEX project partners themselves – whether outside (background data) or inside the project (foreground data) – the CA says that “results shall be vested in the party that has generated them” and that “where results are generated from work carried out jointly

by two or more parties […], they shall have joint ownership." To allow ESA^2 to take care of data publication and licensing, and to also manage commercial exploitation where envisaged, a contract between the concerned partner(s) and ESA^2 will be signed (see Figure 5). Thereby, data ownership remains with the party who has collected/generated the dataset.

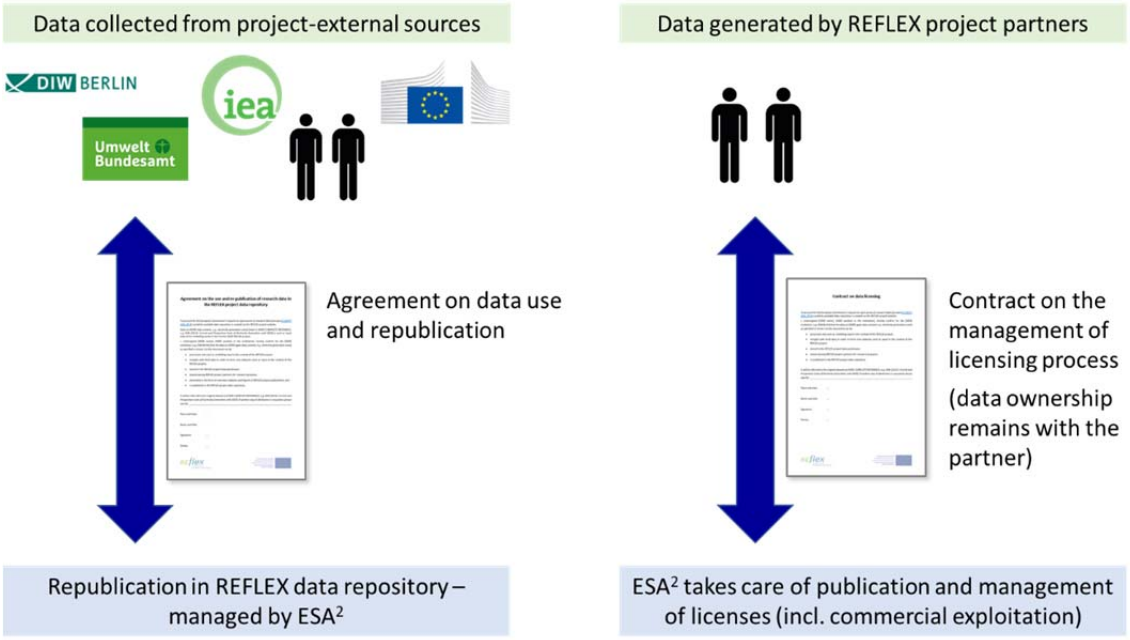

Figure 5: Procedure for giving ESA^2 the right to republish data and manage licensing

2.4.3 DATA QUALITY ASSURANCE

Open data obviously are only of use for future research or stakeholders from industry and policy, if they are up-to-date, as complete as possible and free of any human-made mistakes in data collection, storage and reporting. Moreover, coupled models – as is the case for the EMS – should use harmonized input data and be based on common scenario storylines. Within the REFLEX project, quality assurance processes as described in Table 14 therefore have been implemented for the different data groups.

Table 14: Processes of data quality assurance

Data group	Processes
Existing model input data	- Harmonization of model input data to ensure a consistent analysis within the EMS and regarding the defined REFLEX scenario storylines. For the same information, the same dataset (values) have been used in all models (consortium decision). - Harmonized data being provided to all models before initializing the EMS runs.
Collected and generated new model input data	- A minimum of two internal reviews of the generated new model input data took place. - Additionally, external peer-reviews in case of publication of selected modelling results and/or quantitative research in scientific journals. - Harmonization of model input data to ensure a consistent analysis within the EMS and regarding the defined REFLEX scenario storylines. For the same information have to be used the same dataset (values) in all models (consortium decision). - Harmonized data being provided to all models before initializing the EMS runs.
Generated intermediate model output data	- Plausibility check for intermediate output data of a model during EMS runs by the responsible modeller before implementing the data in the DHW for data transfer to another model.
Generated final result data of the EMS	- A minimum of two internal reviews of the generated final result data took place. - Additionally, external peer-reviews in case of publication of selected modelling results and/or quantitative research in scientific journals.

2.5 REFLEX ANSWERS TO POTENTIAL LEGAL CONFLICTS

Data management in the REFLEX project on the one hand refers to the management of data flows of the different models forming the EMS. Issues related to data harmonization, data storage, data exchange, and data up-/download have been dealt with by implementing the REFLEX data warehouse (see also Deliverable 2.3: Report on modelling coupling framework). Legal aspects associated to the ownership of input and result data, the rights to use them and possible confidentiality restrictions are defined in the Consortium Agreement. It provides details on the partners' background data and on the rights to, the protection and the exploitation of datasets/results generated solely and/or jointly during the lifetime of the project. Thus, no potential legal conflicts did arise here.

On the other hand, data management in the REFLEX project refers to the **(re-)publication** of input data in order to support a transparent research environment, and to the publication of final project modelling results in a format that makes it possible for third parties to access, mine and exploit the data. In general, potential legal conflicts might relate to the following questions:

- How to avoid any copyright infringements when using input data from project-external sources for modelling works and re-publishing those in the REFLEX data repository?
- How to avoid any legal conflicts when using these data for modelling works and publishing EMS modelling results in an open data format?
- How to avoid any legal ambiguities for third parties who wish to use REFLEX final result data?

Copyright may apply to a wide range of creative, intellectual or artistic works, including data. One thereby has to be clear in the use of terminology. Whereas a **single date is not protected** under copyright, structured or organized data which have been collocated to a database via a substantial investment (e.g. time, manpower) might be. In this context, the EC's Database Directive (96/9/EG, Art. 1.2) defines a **database** as "a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means."

Besides copyright law and the EU Database Directive, different other legal concepts may govern data ownership and rights for data use and publication. These include amongst others data privacy laws, national legislation, licenses, or terms-of-use clauses.

When searching for power system data, one will find that many data collections are available online, free of any charge. However, this does not imply that one is allowed to use such data freely. In the absence of any license agreement, the default is that the copyright holder reserves, or holds for his/her own use, all the rights provided by copyright and related law. This includes that already the use of such data – without an explicit consent of the owner – can be a copyright infringement. Discussions with operators of energy data platforms, who are dealing with these issues since several years, confirmed the concerns of the REFLEX partners. It has come to light that long and comprehensive debates are ongoing regarding the question if datasets can be (re-)published and what implications such a publication has on the rules for access, usage etc. This seems still a grey zone in which many researchers and platform operators work.

Having these issues in mind, potential legal conflicts that did arise in the course of the REFLEX project were related to

(i) the use of datasets – i.e.

- Are we allowed to use and process the dataset for our modelling work? If "reuse is authorized provided the source is acknowledged" what is meant by "reuse"?
- Are we allowed to copy, present, share, print, or transmit the dataset?
- Are we allowed to modify or merge with own data and with the data of others to form new datasets?

(ii) the (re-)publication of datasets – i.e.

- Are we allowed to re-publish the dataset in public and non-public electronic networks, and in our project data repository?

and finally

(iii) implications of non-open-access input data on output data publication – i.e.

- Are there any legal conflicts when using data with unclear licensing (or purchased/ confidential data) as input and then publishing our modelling results as open data?

Point (iii) could be neglected, as EMS modelling results are completely new created datasets, not mirroring input data. No conclusion to the values of input datasets can be drawn. Points (i) and (ii), in contrast, are more complex and will be discussed in-depth below. Table 15 summarizes potential legal conflicts for REFLEX input data, which are provided under a number of different legal regimes, i.e.:

- © Public institutions & "All rights reserved.";
- © European Union & "Reuse is authorised provided the source is acknowledged.";
- © Publisher of academic journal;
- data published in online reports without any specification on licensing/reuse;
- data purchased;
- confidential data coming from EC; or
- data generated by project partners (model outcomes, data from surveys, etc.)

with respect to the use of a dataset in the project context and its (re-)publication in the REFLEX data repository.

Table 15: Potential legal conflicts for REFLEX data related to use and (re-)publication

Legal rules on REFLEX input data	Potential legal conflicts regarding the use of the datasets?	Potential legal conflicts regarding the (re-) publication of the dataset?
© Public institutions & "All rights reserved."	**YES** – Full copyright with the author	**YES** – Full copyright with the author
© European Union & "Reuse is authorised provided the source is acknowledged."	**Maybe YES** – What is meant by "reuse"? Only processing? Or also modification, combination with another database? Etc.	**Maybe YES** – No specification regarding re-publication
© Publisher of academic journal	**Maybe YES** – Depends on specific rules	**Maybe YES** – Depends on specific rules; possibly re-publication not allowed for certain time period
Data published in online reports without any specification on licensing/reuse	**YES** – Default is full copyright with the author	**YES** – Default is full copyright with the author
Data purchased	**NO** – explicitly purchased for use in the modelling works	Not relevant here, re-publication forbidden
Confidential data coming from EC	**NO** – explicitly provided for use in the modelling works	Not relevant here, re-publication forbidden
Data generated by project partners (model outcomes, data from surveys, etc.)	**NO**	**NO**

Copyright law only becomes relevant in case of the use of “substantial parts” of a database. As a rule of thumb, one can say that the use of less than 5% will not cause any legal conflicts; for the use of more than 15%, however, full copyright law applies. As REFLEX partners use parts of the respective databases to an extent above this last threshold, copyright law cannot solve the conflicts identified in Table 15. Thus, well-defined contracts are needed for the different datasets at hand in order to avoid any copyright infringements with the use and (re-) publication of REFLEX data.

Within REFLEX, the project partner ESA[2], being responsible for data management and the research data repository, therefore prepared an **“Agreement on the use and re-publication of research data in the REFLEX project data repository”**, including an annex with the respective metadata.

One thereby had to consider that the **author is not always the copyright holder**, but only the copyright holder her-/himself has the right to assign a license. Thereby, copyright ownership is determined by a number of (national) laws depending on the type of employment (see RUI (2018) for further details). For instance, under private employment which is the case for a researcher at a research institute, in Germany all works created during working hours belong to the employer according to §43 UrhG and §69b UrhG. This, for instance, is the case for all REFLEX input datasets coming from DIW (2013).

In the first step, consequently, the copyright holder has been identified. Second, a contact person being endowed with the right to sign such an agreement letter had to be found. Third, the agreement letter, including the annex and an e-mail explaining the context has been sent out, asking the respective contact person for completion and signature. Signed agreement letters finally have been archived by ESA[2].

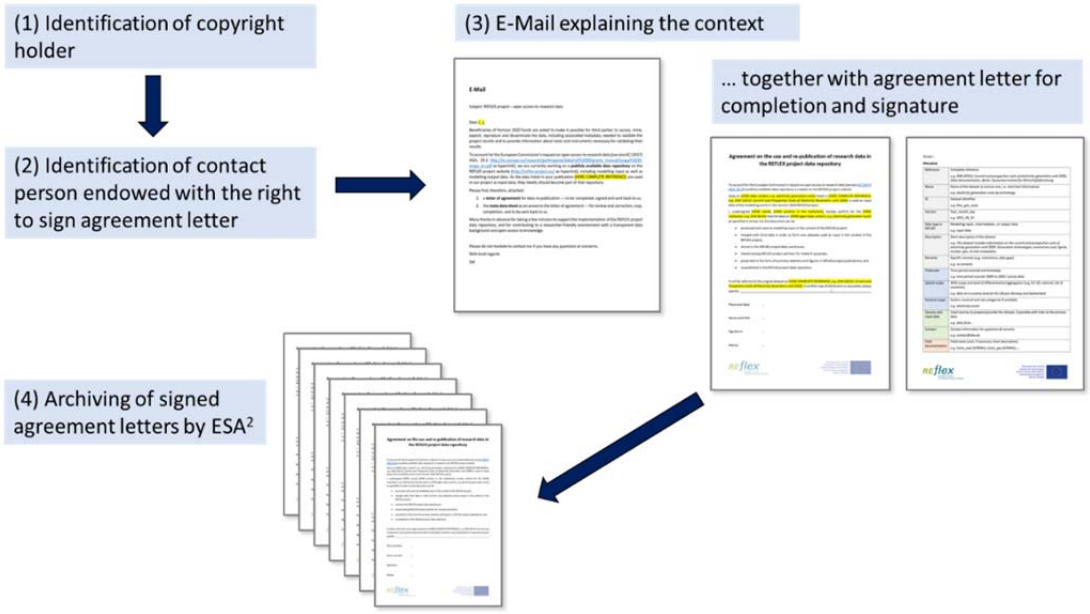

Figure 6: REFLEX process for gathering agreements on use and re-publication of data

Besides copyright law, severe legal conflicts can originate from **data privacy law**, which becomes relevant if data refer to particular individuals. The publication of anonymous data, in contrast, is unproblematic from the data privacy law perspective. In the case of REFLEX research data, no related issue could be identified:

- *DSM parameters* (flexible load, shifting time, etc.) – absence of any reference to particular households or individuals in the respective datasets summarizing the DSM parameters. Representative and average values resulting from the empirical survey conducted for ten EU Member States have been used. DSM data for further countries have been deduced from the survey results.
- *Power plant parameters* (availability, efficiency, cost factors, etc.) – absence of any reference to particular installations. Partly average values for the different generation technologies have been used.
- *Electricity demand* – absence of any reference to particular consumers. Data aggregated over regions and/or sectors has been used.
- *Heating demand* – absence of any reference to particular consumers. Data aggregated over regions and time has been used.

3 ALLOCATION OF RESOURCES

The EIM is responsible for data management within the REFLEX project (see Section 1.4). Estimated costs for making REFLEX data FAIR are 150.000 Euro (over all partners).

These costs include:

- the clarification of data protection issues and licences available,
- the final preparation of data by each project partner for publishing (without effort/costs for data collection/purchasing/generation etc.),
- the processes for assurance of data quality,
- the development and implementation of the data catalogue in the project website,
- the implementation of the registration procedure for access to commercially exploited datasets,
- the data hosting and backup for security,
- the data updating and maintenance of the data and of the data provision

during the project lifespan. These costs are covered by the project funds, mainly by the budgeted personnel costs.

The costs for long-term preservation after the end of the project are difficult to estimate. They depend mainly on the chosen platform, the effort/cost for maintaining but also on the scope and size of the collected and generated datasets. The permanent costs of preserving datasets on the ESA² DWH provided via ESA² Website are estimated with 800 EUR per year (under current conditions for the v-server and without personnel costs for maintenance). The OpenEnergy Platform and the ZENODO repository would be free of charge as long as the single dataset storage is no greater than the maximum 2GB of data. The permanent costs of preserving datasets on the OpARA repository are planned to be free of charge for TUD members, but the final decision on costs has not been taken yet. The costs for long term preservation shall be covered by the collected charges from the commercial exploitation of datasets during the project lifespan and after.

4 DATA SECURITY

Most of the data handled in the REFLEX project are not sensitive regarding the laws governing data protection and data security. An exception represents the data from the DSM survey. A provision/publication of these data is only possible in an anonymous form.

The DWH as well as the data provision via websites will be implemented on servers with regular backup and data recovery procedures.

5 ETHICAL ASPECTS

Data collection, data storage, data usage, data generation and data dissemination in this project do not raise ethical issues.

6 OTHER

No other national-, funder-, sectorial-, or departmental procedures for data management will be used.

REFERENCES

Bolinger, M., Wiser, R. (2012): Understanding wind turbine price trends in the U.S. over the past decade. Energy Policy; 42, 628-641.

Candelise, C., Winskel, M., Gross, R.J.K. (2013): The dynamics of solar PV costs and prices as a challenge for technology forecasting. Renewable and Sustainable Energy Reviews; 26, 96-107.

Capros P. et al. (2016): EU Reference Scenario 2016, Energy, transport and GHG emissions trends to 2050. Luxembourg: Publications Office of the European Union. Download: https://ec.europa.eu/energy/en/data-analysis/energy-modelling

Chen, X., Khanna, M., Yeh, S. (2012): Stimulating learning-by-doing in advanced biofuels: Effectiveness of alternative policies. Environmental Research Letters; 7, 045907.

DIW (edt.) – Schröder, A., Kunz, F., Meiss, J., Mendelevitch, R., Hirschhausen, C.v. (2013): Current and prospective costs of electricity generation until 2050, Berlin: Deutsches Institut für Wirtschaftsforschung.

EC (2016): Guidelines on FAIR Data Management in Horizon 2020 (Version 3.0). Download: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

EC (2016b): Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020 (Version 2.1). Download: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

IEA, NEA & OECD (2010): Projected costs of generating electricity, Paris: International Energy Agency, Nuclear Energy Agency, Organisation for Economic Cooperation and Development.

Junginger, M., van Sark, W., Faaij, A. (eds.) (2010): Technological learning in the energy sector. Lessons for policy, industry and science. Edward Elgar Publishing Limited. Cheltenham, UK, ISBN 978 2 84844 834 6, p. 332.

RUI (2018): Open data licensing, presentation by Ludwig Hülk, © Reiner Lemoine Institut.

Traber, T., Kemfert, C. (2011): Gone with the wind? Electricity market prices and incentives to invest in thermal power plants under increasing wind energy supply. Energy Economics; 33, 249-256.

UBA (edt.), Ichta, P. (2014): Entwicklung der spezifischen Kohlendioxid-Emissionen des deutschen Strommix in den Jahren 1990 bis 2013, Dessau-Roßlau: Umweltbundesamt.

VGB PowerTech (2011): Investment and operation cost figures – generation portfolio, Essen: VGB PowerTech.