Report No ECHOES 2.1 – D2.1 Report of existing data and databases as additional to new data collection and formulation

Report

Report of existing data and databases as additional to new data collection and formulation

Picture under creative common licence. © FreePhotosART at pixabay.com

Report of existing data and databases as additional to new data collection and formulation

KEYWORDS:
SSH, social science, humanities, low carbon, Set-Plan, electric mobility, buildings, smart energy

VERSION 03

DATE 31.08.2017

AUTHOR(S)
Lassi Similä, Tiina Koljonen, VTT Technical Research Centre of Finland Ltd

Quality ensurance:
Christian A. Klöckner, Jens Olgard Dalseth Røyrvik, NTNU

PROJECT NO. 727470 (H2020)

NUMBER of PAGES/APPENDICES: 54/1

ABSTRACT
The overarching objective of ECHOES (Energy CHOices supporting the Energy union and the Set-plan) is to unlock the policy potential of an integrated social science perspective bounded by central socio-cultural, socio-economic, socio-political, and gender issues that influence individual and collective energy choices and social acceptance of the energy transition in Europe. The internal working document D2.1 of Workpackage 2 aimes at introducing existing databases on social science and humanities (SSH) that are relevant for low carbon energy transition. The working document D2.1 supports the development of an open access ECHOES database, that in turn collects and synthesises existing data relevant for the focus of ECHOES in addition to the new data that will be generated during the ECHOES project.

According to analysis presented in this report, the nature of SSH data can be described as fragmented, dispersed, consisting of varying technical implementations, varying languages, as well as of varying formats. Furthermore, providers of data portray a large variety. Particularly, national statistical agencies are mentioned frequently while ministries are another typical source. The spectrum of other data providers ranges from research institutes to private companies to NGOs and industry organisations. The reliability of classified data sources is not self-evidently homogenous between all different sources. This issue should be considered in the development of the database of ECHOES project

REPORT NO. D2.1

ISBN NA

CLASSIFICATION public

CLASSIFICATION THIS PAGE public

Document history

VERSION	DATE	VERSION DESCRIPTION
1	11.8.2017	1st draft report for internal comments of ECHOES partners
2	25.8.2017	2nd draft report for preliminary copy editing
3	31.8.2017	Final version

Table of contents

FIGURES

TABLES

LIST OF ABBREVIATIONS

ACEEE	American Council for an Energy-Efficient Economy
ALLBUS	Allgemeine Bevölkerungsumfrage, the German General Social Survey
AR	Assessment Report
ATEsT	Analysing Transition Planning and Systemic Energy Planning Tools for the implementation of the Energy Technology Information System
BP	British Petroleum
CCS	Carbon Capture and Storage
CEPII	A French research center in international economics
D2.1	Deliverable 2.1
ECHOES	Energy CHOices supporting the Energy union and the SET-plan
EIA	U.S. Energy Information Administration
ENVERDER	Energy Efficiency Association
ETSAP	Energy Technology System Analysis Programme (of the International Energy Agency, IEA)
EU	European Union
EUDAT	European Data Infrastructure
EV	Electric Vehicle
FAIR	Findable, Accessible, Interoperable, and Re-usable
GDP	Gross Domestic Product
GfK	Gesellschaft für Konsumforschung, Association for Consumer Research
GHG	Greenhouse gas
GTAP	Global Trade Analysis Project
ICT	Information and Communications Technology
IEA	International Energy Agency
INE	El Instituto Nacional de Estadística, National statistical office of Spain
IPCC	Intergovernmental Panel on Climate Change
IPR	Intellectual Property Rights
ISSP	International Social Survey Programme
JRC	Joint Research Centre (of the European Commission)
NGO	Non-Governmental Organisation
NTNU	Norges Teknisk-Naturvitenskaplige Universitet, Norwegian University of Science and Technology
OECD	Organisation for Economic Co-operation and Development
PV	Photovoltaic
SESAME	Securing the European Electricity Supply Against Malicious and accidental thrEats
SET-Plan	European Strategic Energy Technology Plan

SOEP	Socio-Economic Panel
SSH	Social Sciences and Humanities
SSPs	Shared socio-economic pathways
TEVEM	Turkey Energy Efficiency Assembly
UN	United Nations
VTT	VTT Technical Research Centre of Finland Ltd
WP	Work Package

1 INTRODUCTION

The overarching objective of ECHOES (Energy CHOices supporting the Energy union and the Set-plan) is to unlock the policy potential of an integrated social science perspective bounded by central socio-cultural, socio-economic, socio-political, and gender issues that influence individual and collective energy choices and social acceptance of the energy transition in Europe. ECHOES will therefore foster the implementation of the European Strategic Energy Technology Plan (SET-Plan[1]) and advance the Energy Transition, in addition to the decarbonization of the EU's future energy system. ***This report is an internal working document of Work Package 2 aimed at introducing existing databases on social science and humanities (SSH) that are relevant for low carbon energy transition. The working document D2.1 supports the development of open access ECHOES database, that in turn collects and synthesises existing data relevant for the focus of ECHOES in addition to the new data that will be generated during the ECHOES project.***

Central to all research activities in ECHOES are the technological topics of a) smart energy technology, b) electric mobility, and c) buildings. ECHOES addresses the challenges by employing the innovative theoretical concept of "energy collectives" which covers determinants of energy choices from the perspective of (1) individual decision-making as part of collectives (micro), (2) collectives constituting energy cultures and life-styles (meso), and (3) formal social units (macro).

Smart energy technologies are at the core of an integrated roadmap towards realizing an energy revolution as explicated by the SET-Plan[2]. This includes distributed, small-scale renewable energy production technologies (typically rooftop photovoltaic (PV), solar thermal and micro wind, heat pumps and bioenergy), in addition to a range of technologies for the traditional "demand side" (e.g. in-home displays, home automation, smart home appliances, etc.) and energy storage. Thus, our energy system is undergoing important transformations with the deployment of smart meters, smart controls, smart appliances, and their integration in home networks.

The SET-Plan identifies *electric mobility* as one of the core technologies to be implemented and further developed to enhance road transport efficiency. Both passenger and goods transportation account for a substantial amount of the environmental impact of the member states of the EU (Hertwich & Peters, 2009).

The last technology focus is *buildings* - which include construction activities, insulation, energy efficiency upgrading, heating, cooling, illuminating, and energy use behaviour in buildings. A reasonable use of the territory, resulting in compact urban structures, was outlined among the four main aspects of key importance for urban sustainability (Leipzig Charter on European Sustainable Cities, 2009[3]).

1.1 Energy policy targets of the European Union

European Commission has set a long-term target for achieving the low carbon economy by 2050 (EC 2011a, b). The goal of the European Union and its Member States is to promote a transition to a consumer-driven sustainainable energy system. Communication "A policy framework for climate and energy in the period from 2020 up to 2030" (EC 2014) included the first proposal of the climate and policy framework up to 2030, which was discussed further in the "Clean Energy for All Europeans[4]" legislative proposals, which were published in November 2016. The proposal has three main goals: putting energy efficiency first, achieving global leadership in renewable

[1] see http://ec.europa.eu/energy/en/topics/technology-and-innovation/strategic-energy-technology-plan

[2] see https://ec.europa.eu/energy/sites/ener/files/documents/set-plan_progress_2016.pdf

[3] Informal Ministerial Meeting on Urban development and territorial cohesion. Leipzig Charter on Sustainable European Cities, 2007; www.eufunds.bg/document/355

[4] see http://ec.europa.eu/energy/en/news/commission-proposes-new-rules-consumer-centred-clean-energy-transition

energies and providing a fair deal for consumers. Therefore, the central focus and aim of this so-called Winter Package are active consumers, who are the central players on the energy markets of the future. Consumers of the future across the EU should, on the one hand, have a better choice of supply, access to reliable energy price comparison tools and the possibility to produce and sell their own electricity. On the other hand, it is necessary to mitigate the societal impact of the clean energy transition and the package also recognizes the risks of "energy poverty" and thereby includes measures to protect the most vulnerable consumers against rising energy bills.

1.2 SSH in low carbon energy transition

Mitigating greenhouse gas (GHG) emissions and stabilizing GHG concentrations to a safe level will require large-scale transformations in human societies, for instance, the way and amount we produce and consume energy, goods and food, and how we use and protect land surfaces and other natural resources. There are numerous technical solutions for GHG mitigation, but several of the GHG mitigation options are value driven and are dependent on our behaviour, such as the choice of diets or ways of moving. Moreover, there are several barriers, for instance public acceptance of certain technologies (e.g. nuclear, CCS, or even wind) or lock-ins to conventional technologies due to their long life-time (e.g. energy and industrial plants, buildings, and vehicle stock).

There are numerous analyses on a sustainable transition to low carbon energy systems, including policy-impact analysis and market assessments. However, the quantitative analysis, in particular, have primarily focused on technological portfolios, viability and pathways while the social changes and acceptance, political feasibility, and institutional aspects have largely been missing from the analysis (Koljonen *et al*. 2016). The content analysis by Sovacool (2014) of 4,444 research articles in three leading energy journals from 1999 to 2013 indicates that electricity clearly prevailed with almost one-quarter of articles and with quantitative tools being the most used methodological approach. Behaviour was listed among the least favoured topics in this sample together with land use and research and development. The research carried out in the earlier EU ATEsT[5] (Analysing Transition Planning and Systemic Energy Planning Tools for the implementation of the Energy Technology Information System) project of the 7th Framework Programme revealed a similar story: Tools and methods focused on analysing the effectiveness of RD&D policies, consumer and/or investor behaviour and institutional factors are mostly lacking. Furthermore, energy system modelling scarcely takes these issues into account (Koljonen *et al.* 2012).

Taking into consideration the energy and climate policy targets of the European Union and its Member States, it is evident that there is a considerable gap between current research methods and practices, which would consider human behaviour and social aspects on different levels in the transition to sustainable low carbon society. Generally, the bottleneck in analysis is insufficient or missing data, or if relevant data exists, it could translate to being costly, non-transparent, and/or difficult to use for certain purposes (for example, the data is unharmonized between countries or sectors, infrequently updated, or requires specific technical knowledge on databases). However, several studies are available that focus on acceptability of certain clean energy technologies or energy and climate policy targets, which can provide useful data on SSH in low carbon energy transition. Thus, D3.1 presents the results from reviewing literature relevant for the scope of ECHOES and therefore this literature is not presented in D2.1.

Data on Social Sciences and Humanities (SSH) generally deals with non-technical issues such as individual and collective energy choices and behaviour. SSH relates to both quantitative and qualitative data on, along others, energy cultures, consumer and investor behaviour, acceptance, policies and regulation. With regards to low-carbon energy system research, data sets can be utilized, for example, for analysing:

1. techno-economical and socio-economical potential of changes in technological systems and behaviour;
2. the social and institutional barriers of the transition;
3. policies and regulation required to reach the required targets.

[5] see http://www.cres.gr/atest/index.htm

Of significant importance is the inclusion of the SSH dimension in both the *ex-post* and *ex-ante* analysis to understand the impacts of human behaviour on realised and expected developments. This enables risk avoidance and the identification of best practice governance strategies for different types of formal social organizations, the targeting of specific technologies and desired practice/behaviour changes.

The fast developments of ICT technologies is viewed as raising an interest and activating potential towards open data approaches in both low-carbon energy research as well as in SSH fields, therein setting the scene for new opportunities. Efficient utilization of data has several benefits. It could diminish overlapping research, increase transparency of research results and enable the achievement of greater benefits of the data through increased re-use and to promote knowledge sharing and innovations.

1.3 Scope of the report

The goal of the multidisciplinary ECHOES project Work Package 2 includes collecting, synthesizing, and curating relevant data, in addition to formulating indicators related to individual and collective energy choices and behaviour. As a first step in database creation, Tasks 2.1 and 2.2 of the ECHOES project commenced with mapping and categorizing the state-of the art by reviewing existing quantitative and qualitative SSH data and databases. The result of this state-of the art SSH data analysis is reported in this first deliverable of the WP2, which includes:

(i) identifying and listing the key characteristics of databases, sources and gaps on SSH data;
(ii) drawing a proximate picture of the data need and availability profiles between different topics and areas;
(iii) drawing on conclusions based on key characteristics for the database developed in the project.

The results presented in this deliverable D2.1 significantly rely on public literature and a survey conducted within and outside the ECHOES consortium.

The results of WP2.1 and WP2.2 will serve in forming a knowledge base for later steps of the ECHOES project, whereby an open access database of SSH related data will be created. The database will collect and formulate SSH related quantitative and qualitative data from existing national, regional, and EU level databases and based on the results from the other WPs of the project.

This deliverable D2.1 is an internal working document within ECHOES, which provides the needed background information for the next steps of WP2, for instance, the creation of SSH indicators (Task 2.3) for sustainable energy system transition, the WP2 expert workshop (Task 2.4) and finally preparation of the open access ECHOES database. Hence, the empirical basis for the analysis of existing databases will be strengthened by the WP2 workshop results and possibly expert interviews postliminary to the workshop, and D2.1 will be updated after these results are available.

The working report D2.1 aims at answering the following questions, which will be assessed in the next tasks of WP2:

(1) What is the definition of a “SSH database” for the ECHOES project? What are the key characteristics the ECHOES database will possess?
(2) What kind of SSH data is available and where are the greatest bottlenecks?
(3) How the three technology foci of ECHOES, i.e. buildings, electric mobility, and smart energy technology, should be considered in WP2 analysis and database?

This report is structured as follows:

- Chapter 2 reviews characteristics of databases, SSH data and their significance for the ECHOES project and this report
- Chapter 3 presents results of the surveys based on all the responses received
- Chapter 4 takes an in-depth look at databases selected most relevant for this report
- Chapter 5 maps directions for ECHOES database development based on requirements and suggested knowledge gaps
- Chapter 6 concludes and outlines the next steps

2 OVERVIEW OF SSH DATABASES IN ECHOES

2.1 Background of the SSH data used in energy system transition analysis

Social science studies human society and the relationships among individuals within a society. A number of scientific disciplines are included in social sciences, such as sociology, psychology, anthropology, economics, political science, and history. Furthermore, the transition to low carbon energy systems and societies is a multidisciplinary challenge, which combines research on technologies, natural ecosystems, demographics, governance, economy, and human behaviour.

There are well-established research methodologies and methods that are used in social science and, moreover, in analysing energy systems and transition to low carbon societies. Examples of popular research designs and in social science include laboratory experiments, field experiments, field surveys, secondary data analysis, case research, and focus groups. Data may be collected using survey questionnaires, interviews, observational techniques, analysis of internal or external documents, moderated discussions, and different combinations of the aforementioned tools. Data (or values of attributes) may be quantitative (numeric) or qualitative (non-numeric), and unidimensional (contain a single underlying dimension) or multidimensional (consist of two or more underlying dimensions) (Bhattacherjee 2012). In contrast, common research designs on low carbon energy and its transition are largely based on quantitative and qualitative scenario and case analysis using secondary data given that data creation based on laboratory and pilot experiments for certain technologies or physical systems are exceedingly costly to develop and a lifetime of experimental work might take years or even decades.

The analysis on sustainable energy system transition primarily focuses on quantitative modelling of future energy systems and therein largely neglecting non-technical issues, like behaviour. On the contrary, many non-technical data, like numeric values for population, its structure, and demographics, is used in quantitative analysis to study transition to low carbon energy systems. In addition to technological portfolios, the quantitative analysis of energy transition focuses on calculation of future energy demands, greenhouse gas and other emissions, land use and its change, gross domestic product (GDP), employment, energy and other commodity prices, or even health impacts. All the aforementioned have linkages to behaviour and acceptance: how and how much we produce and consume energy and what are the consequences of transition on our economic welfare, health, or energy cultures and energy memories of societies.

Research on sustainable future energy systems has primarily focused on the mitigation of greenhouse gases to a safe level. In the 5th Assessment Report (AR) by the Intergovernmental Panel on Climate Change (IPCC 2014) shared socio-economic pathways (SSPs) and narratives were used to more effectively analyse climate change impacts, adaptation and vulnerabilities and for integrated assessment of climate response strategies (Kriegler et al. 2012). In consideration of this, Scheweitzer and O'Neill (2014) identified 13 determinants of climate change mitigation and adaptation challenges (e.g. key variables or determinants) at a globally aggregated scale (e.g. all countries and regions are aggregated). They surveyed experts that were working with SSPs and according to the survey, the most relevant factor for challenges to both mitigation and adaptation was income per capita (84 % of responses) thereby reflecting the high challenge of low income countries. In terms of mitigation, the following factors were identified as having the highest impact on the growth of GHG emissions in addition to income per capita, and/or GHG mitigation and adaptation challenge during this century (percentage of responses):

- Global population (68%)
- Carbon intensity of global economies (44%)
- Energy-related technological change (44%)

- Agricultural productivity (e.g. the higher the productivity the lesser land use and other GHG emissions) (40%)
- Energy intensity of global economies (36%)

Correspondingly, for adaptation the following factors were identified (percentage of respondes):

- Extreme poverty (76%)
- Quality of governance (52%)
- Water scarcity (46%)
- Proportion of populaton on coasts (44%)
- Innovation capacity (36%)
- Urbanisation (32%)
- Educational attainment (24%).

Interestingly, the experts that worked with mitigation mainly named technical and economic factors as significant while experts on adaptation primarily named factors that were more related to societies and humanities. This presents itself as a good example of how important it is to combine different research disciplines, for instance SSH in energy system transition, in order to ensure social learning and knowledge creation in solving multidisciplinary problems, such as the transition to sustainable energy systems.

2.2 Definition and description of SSH data in ECHOES

Evidently there is a great challenge to narrow down the type of SSH data, which should be included and studied in the ECHOES project. The results presented in this report are the first attempt to "paint the big picture". During the project's lifetime, more progress towards indepth details will be possible promptly after data is delivered from the other WPs and after organising the WP2 workshop, whereby the SSH data requirements will be further discussed.

To more effectively answer the question "What is SSH data in ECHOES" we should take into account the 1) general objectives of the project; 2) potential usage of data; and 3) data created during the project's life time. However, the definition should be discussed during the entire project lifetime in accordance with new knowledge creation. For this report, the following topics have been considered:

- Objectives: SSH data relevant for EU's energy system transition and the three technology foci (smart energy technologies, electric mobility, buildings).
- Potential use of data: quantitative and qualitative ex-post and ex-ante analysis of existing policies, regulation, and effective governance of energy transition. SSH in energy related decision-making (micro, meta, and macro levels).
- ECHOES data based on, along with others, interviews, literature and other surveys, meta-data analysis and formulation of indicators.

2.3 What do we mean by databases and data sources?

To analyse existing databases, we first need to briefly discuss the term "database" and outline the different classifications of databases. Noticeably, the term *database* is utilized with several meanings and in different contexts in public media and scientific publications. There exist, for instance, different types of content (e.g. numbers, interviews, and bibliographical information), user interfaces, technological solutions and usage purposes for databases. Wikipedia defines a database as follows: *"A database is an organized collection of data. It is the collection of schemas, tables, queries, reports, views, and other objects."* In the following subsections, the classifications of the characteristics of databases relevant for ECHOES are presented.

2.3.1 Classification by content

Database size can range from very general and wide databases maintained by statistical officials such as "Eurostat", to more specific, small datasets developed for a single study. The size of a single database could be registered by indicators such as storage space required (Mb), or the number of sectors/variables included in the database. Size is a relevant characteristic for the ECHOES database development through, for instance, the limits set by data depository providers. **Data type** indicates what kind of data is stored in the database. As an example, as presented on the website of Oregon State University[6], the following classification is provided with regards to data **types:**

Observational

- Captured in situ
- Cannot be recaptured, recreated or replaced
- **Examples**: Sensor readings, sensory (human) observations, survey results

Experimental

- Data collected under controlled conditions, in situ or laboratory-based
- Should be reproducable, but can be expensive
- **Examples**: gene sequences, chromatograms, spectroscopy, microscopy

Derived or compiled

- Reproducible, but can be very expensive
- **Examples**: text and data mining, derived variables, compiled database, 3D models

Simulation

- Results from using a model to study the behavior and performance of an actual or theoretical system
- Models and metadata, where the input can be more important than output data
- **Examples**: climate models, economic models, biogeochemical models

Reference or canonical

- Static or organic collection [peer-reviewed] datasets, most probably published and/or curated
- **Examples**: gene sequence databanks, chemical structures, census data, spatial data portals

[6] http://guides.library.oregonstate.edu/research-data-services/data-management-types-formats

Within the above-mentioned categories, the content of databases under investigation can be divided further into quantitative and qualitative data. The difference of quantitative and qualitative data is illustrated by examples extracted from the ECHOES project presented in Table 1. Further specifications can also be made regarding the three foci areas of ECHOES project: smart energy technology, electric mobility, and buildings.

Table 1. Quantitative and qualitative data

Task 2.1: Quantitative data	Task 2.2: Qualitative data
Compilation of national energy and emission taxes and supports (incl. consumer and prosumer supports), consumer and prosumer energy prices, size of energy bills for different income groups, population structure, share of urban/rural population, mobility and building related data, etc.	Compilation of data on acceptance of new energy technologies and services, vested interests, level of awareness as well as cultural, and religious aspects, level of democracy/transparency of policy making processes (government autonomy), the design of administrative and legal approval processes (strength of the bureaucracy), geopolitical issues, national implementations of relevant EU Directives, national implementations of building codes, interest group strength, alliance structures (interest groups-politicians-bureaucracy), etc.

However, there are also contrasting understandings of quantitative and qualitative data. Classifications could also be constructed between "hard" and "soft" data, for instance, between physically measurable variables and consumer perceptions or attitudes.

2.3.2 Classification by maintaining/administrating organisation

The organisations maintaining or having created the databases under investigation vary widely - from national or international statistical offices, international organisations, ministries, public and private organisations, NGOs, research institutes and universities, to name a few. These pieces of information are useful in a preliminary assessment of the possibility for openness, transparency and reliability of data.

2.3.3 Classification by technical implementation

As a third layer in database classification, there are several possibilities in data formats and technical implementations. These questions become important in developing the project's own database. One important dimension is the requirement imposed on data to be stored in databases such as metadata definitions. There are several initiatives by the European Union and other scientific organisations for open protocols in this area. Examples include the EU-originated Zenodo, EUDAT, B2SHARE, B2SAFE, as well as Figshare (by UK based Digital Science & Research Ltd)).

2.3.4 Classification by frequency of updating

Finally, a clear need to distinguish data sources by frequency of updating is recognized. On the one hand, some data sources appeared as static documents that were dated, at times, several years ago. On the other hand, data maintained by national statistical organisation are typically updated on a regular basis. This has implications for the re-usability of data.

2.4 Methodological approach and structure of this report

The methodological approach of this report is highly based on surveys conducted internal and external for the ECHOES consortium. The survey inside the ECHOES consortium was conducted in January 2017. A similar questionnaire was sent to a specific international modelling community of the IEA ETSAP (Energy Technology System Analysis Programme) in May 2017. Table 2 summarizes the phases and methods applied in the study.

For the purpose of studying the state-of the art of ECHOES relevant data, no strict definitions for data were applied in the initial surveys. Therefore, to broadly collect different views from different countries, the respondents were allowed to express their thoughts relatively freely. As a consequence, a substantial part of the suggestions consisted of individual research articles or policy papers, organisations, and generic websites as data sources. These types of suggestions do not necessarily fit in with the narrow meaning of the term database. In this report, a more general term of *data sources* is used to highlight this particular difference from the term *databases* which reflects a more organised configuration.

Based on theoretical discussions (Chapter 2), a classification for suggested data sources, is developed (Chapter 3). The responses obtained from the surveys are first superficially investigated, classified and then listed according to the developed classification (Chapter 3). Based on the different classes identified, the *"Databases"* class is considered most relevant for the design of the ECHOES project database. The justifications are provided in Chapter 4. The data sources under this category are further investigated in Chapter 4. An overview of a small selection of databases allowed us to make preliminary conclusions on issues that require acknowledgement in database development of the ECHOES project.

Chapter 5, to a greater extent, takes the development needs and actions to a concrete level. For example, a question of what kind of technical data depositories are provided by the EU and other research infrastructures is addressed. The more concrete the research phase was, the more it was supported by expert discussions with knowledge service and database experts at VTT (see Table 2).

Table 2. Phases and methods used of the study.

Phase	Method	Result	Reference in this report
1 Survey for the ECHOES consortium	Survey implemented by Webropol software	Identifying and listing the key characteristics of databases, sources and gaps on SSH data Providing a proximate picture of the data need and availability profiles between different topics and areas;	Chapter 3
2 Survey for the ETSAP community	Survey implemented by Webropol software	Identifying and listing the key characteristics of databases, sources and gaps on SSH data Providing a proximate picture of the data need and availability profiles between different topics and areas;	Chapter 3
3 Analysis and classification of results	Desk research	Identifying the most relevant databases for the ECHOES database design. Identifying potential barriers of use.	Chapter 4

4 Orientating development needs for the ECHOES database	Desk research, utilizing knowledge service expertise at VTT	Identifying existing potential data infrastructures, discussing the substance of ECHOES database	Chapter 5
5 Concretizing the results for the ECHOES project	Desk research, discussions with database expert at VTT	Concluding on key characteristics for the database to be developed in the project. Suggestions for next steps for the ECHOES database development in the project	Chapter 5, Chapter 6

3 RESULTS OF THE INITIAL SURVEY ON SSH DATA

The mapping of existing databases and their characterization with respect to the aforementioned criteria, as an initial step, commenced with an initial survey. The survey was conducted inside the ECHOES consortium in January 2017 and received 13 key responses from around Europe. That is, the respondents were located in Norway, Austria, Germany, Turkey, Bulgaria, Finland, Spain, and Italy. A similar questionnaire was sent to a specific international modelling community of the IEA ETSAP (Energy Technology System Analysis Programme) in May 2017. Six responses were received. Many of the members of the IEA ETSAP work for their governments to model and analyse the impacts of energy and climate policies, which motivated the selection of this target group for the mapping outside of the ECHOES consortium. However, due to the relatively small number of respondents, a detailed presentation of the results is not provided here. The observations were in any case; exceedingly similar to the ones received from ECHOES partners. Most importantly, according to the inquiry, the reasons for not answering were predominantly related to lack of knowledge of the SSH field, despite it being recognized as very important. Accordingly, many of the ETSAP partners support their governments in formulating energy and climate strategies as well as analysing the impacts of policies on national energy systems and economy. This in turn, provides some indication of current methods and their applicability in the consideration of SSH dimension in energy system transition.

Notably, all the respondents (e.g. both ECHOES and ETSAP communities) represented research organisations. This highlights the need to attain more insights from other stakeholder groups in the following phases of the WP2. Furthermore, the need for high-level mapping of data sources and their characteristics, primarily for internal purposes of the project, could be served. The questionnaire was designed to identify and list the key databases, sources and gaps on SSH data, and researchers professionally working in the field are considered to be an appropriate starting point. Utilizing the survey, a proximate picture of the data need and availability profiles between different topics and areas could be delineated. As the sample size of the questionnaire is small, the results are most usable in characterizing the nature of SSH data and generating hyphotheses. Consequently, the results should not be interpreted as general observations but rather be used for orientation and as input for forthcoming phases of the research.

After an intial review of the content of the answers, a general conclusion drawn was related to the problem of the vaguely defined concept of “SSH” data. The answers reflected highly varying contents of data. In addition to being classified as qualitative or quantitative, it emerged apparent that very diverse phenomena in society are seen eligible for inclusion under “SSH” category, which was previously discussed in the chapter 2.2.

3.1 General SSH data

In the initial survey, the respondents were first initiated to assess the SSH data availability in the areas depicted in Figure 1 in general, that is, with no tailored approach for any of the foci areas. The assessment was approached by introducing the aims of the WP2 as formulating and collecting both quantitative and qualitative SSH data related to, among other areas, energy cultures, consumer behaviour, acceptance, policies, and regulation, which should be taken into consideration in the transition to the sustainable low carbon energy systems. The assessment was briefed through the following question: *Consider the question of transition to sustainable low carbon energy systems. Please assess the data availability in this respect on a scale 1..5, where 1=No data available, ... 5=Abundance of data*. According to the results, *population structure and share of rural and urban populations* are suggested to be the areas with highest availability of data. On the other hand, *energy cultures*, *gender issues*, and *acceptance* scored the lowest averages. The assessment results, based on 13 answers, are illustrated in Figure 1. Additionally, the results of the ETSAP community survey are presented in comparison.

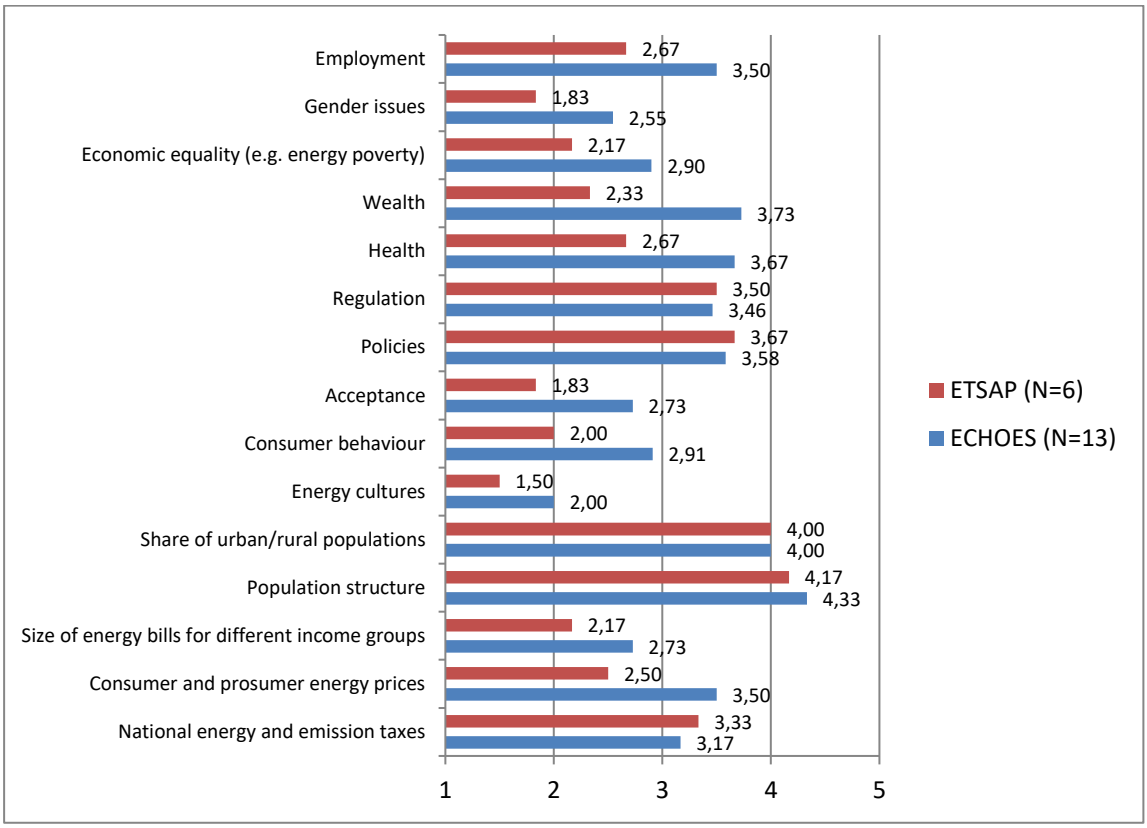

Figure 1. The ECHOES consortium and ETSAP researchers' assessment on data availability in different areas in general (overall mean of the ECHOES group = 3.26; overall mean of the ETSAP group = 2.69)

The respondents were requested to specify the SSH data sources in these areas by way of free-format answers (see summary in the appendix). The data sources were provided using links to websites or databases, listing the data providing organisations, as well as scientific articles. Conclusively, the results of the questionnaire suggested a large selection of data sources on the question of transition to sustainable low carbon energy systems.

Due to the different nature in data source suggestions, it was deemed not reasonable to support explicitly listing the data sources by sizes or technical implementation means (section 2.3). However, this dimension is further discussed in a more focused analysis of selected data sources (Chapter 4).The suggestions were categorised using the following, more abstract categorization: (1) *Websites* include links to many forms and sources of data; (2) *Documents* group consists of, as examples, policy and strategy reports, scientific articles, market reports by consultancy organisations. (3) *Databases* that present the most "genuine" characteristics with respect to database definition. That is, they for example allow for making queries, or for downloading data in an organised form.

A demonstration of the different natures of the SSH data sources are provided by a reviewed website of SESAME project (Securing the European Electricity Supply Against Malicious and accidental thrEats). According to its website, the study contains an in-depth survey of 250 households in every of the 27 member states (totaling 6,750- the largest survey of its kind ever to be conducted). The survey deals with electricity consumption habits and preferences regarding reliability of electricity supply. Despite the availability of the results purely as a pdf document, on the website, it is reasonable to believe that data, in some way, has been stored in a more organised form.

Based on the answers, it can explicitly be concluded that size of the databases varies extensively. This dimension ranges from very general data sources such as “Eurostat”, “IEA”, or national statistical offices to a specific dataset developed for a single study focusing on a highly specified research question. As an example of very specific classification of the data sources, as a part of the ECHOES WP3 literature review, a preliminary list received from one respondent includes approximately 160 papers, as well as a simplified list of an estimated 60 papers, exclusively including papers related to energy consumption, transport, and policy support. The studies are classified methodologically to experiment, field and questionnaires, dealing with dependent variable, predictors and sample sizes. As another general observation, databases must be differentiated with respect to those related to one survey without constant updates to more dynamic databases such as those maintained by national statistical offices updated for instance annually.

The questionnaire revealed several different governing, owner, or maintaining organisations for the data sources. These organisations include various international or national organisations in research, public and private sectors. These observations highlight the need for specifying and classifying databases accurately as they are designed for the ECHOES needs. While the complete list of suggested databases in the internal inquiry is presented in the appendix, the maintenance of the databases can be classified under the following national organisation types:

Types of organisations

- National statistical offices
- National policy/strategy documents (typically run by ministries)
- National public organisations (energy regulators, consumer research offices, energy efficiency offices…)
- National private organisations (e.g. banking institutes, engineer organisations…)
- Scientific projects/databases/publications – both national and EU projects.

National statistical offices, operating on a national mandate to produce official statistics, were mentioned in practically every answer. Commonly, ministries and reports or databases maintained by them were mentioned as sources in several issues. This can be viewed as a natural activity belonging to their respective sectors (policies, regulation, health, wealth, employment). Overall, conceivably in line with ministerial mandates and responsibilities to implement national policy measures within these areas, the data availability specific to these areas was assessed as relatively well-covered. In contrast, there were no such generic constantly updated databases listed among the weakest covered areas - *energy cultures*, *gender issues*, and *acceptance*. The profile of the ETSAP community is largely in line with these, as they additionally scored the second lowest in the ESTAP survey, followed by *consumer behavior*. If at all mentioned, these areas are often covered by single studies or questionnaires.

In addition to national sources, the respondents mentioned **international/European** data sources/organisations on general SSH data (Table 3). These data sources can broadly be classified as (I) generic, to (II) more specialized in social sciences, and (III) more emphasis on energy data.

Table 3. International data sources mentioned in the study classified.

Data sources	Description
Generic data sources	The European Commission, Eurostat, ScienceDirect, Google Scholar, Elsevier, ResearchGate, U.S. Government's open data (data.gov), UK Government's open data (data.gov.uk), World Bank, United Nations, CEPII (a French research center in international economics), EU open data portal, JRC, OECD
Data sources with a SSH data emphasis	International Social Sorvey Programme, European Values Study, World values survey, Gesis Sowiport, World Input-Output Database
Data sources with an energy data emphasis	International Energy Agency (IEA), BP, EIA, ACEEE American Council for an Energy-Efficient Economy

An overview of the data sources suggested raised an issue of unstandardized, diverse formats making the interoperability of data challenging. With this understanding, it was also observed that different languages used in the sources can be seen as problematic. Some of the data sources presented originated from a decade ago or even older. Given our awareness of fast technological progress of, for instance, renewable energy technology (especially photovoltaic and wind) and electric mobility, in addition to the political and economic environment under turbulence as a common global trend during the last decade, the validity of studies on socio-economic issues cannot be viewed as self-evident.

The reliability of classified data sources is not self-evidently homogenous between all the suggestions. Despite the limitation of the inability to measure data source by data source, with a reasonable workload for all the sources presented in this report, the wide variability of sources originating from different types of authors and methods suggests that there may be differences regarding this aspect.

3.2 Mobility related data

As a second part of the questionnaire, the respondents were requested to assess the availability of data in the mobility sector to serve one of the technology foci in ECHOES, that being - electric mobility. Notably, the question was not restricted to e-mobility in order to additionally get a broad picture of the data availability on competing technologies. The focus area specific section followed the structure of the more generic section. In other words, the respondents were first requested to assess the data availability and then to specify the sources.

Figure 2 presents the respondents' views on availability of mobility related data. No large differences can be established between the areas. *Prices in the mobility sector* and *Regulation* are seen as areas with the best data availability, while *Acceptance of mobility technologies* was indicated as the section with the weakest data availability. The overall mean of "mobility" is 3.18, which is in the same magnitude that has been reported for general SSH data.

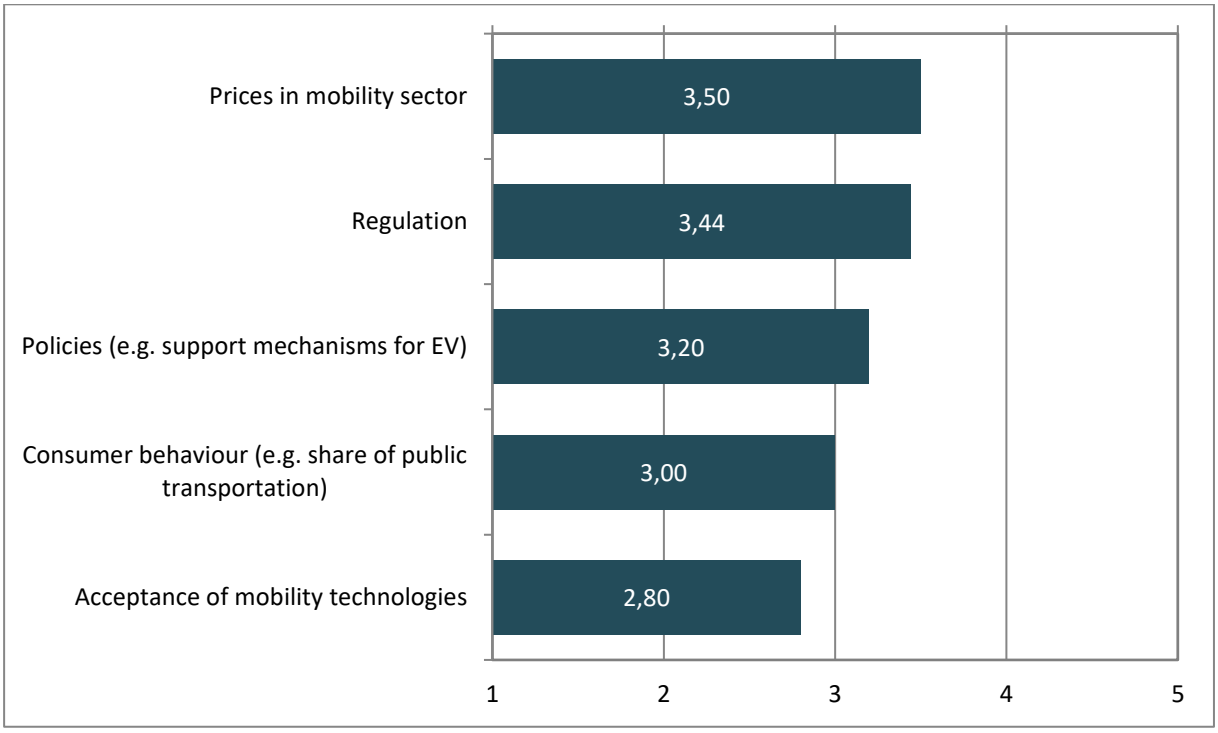

Figure 2. Data availability: Mobility related data (overall mean = 3.18)

An overview of the specific data sources named in the mobility sector (see Table 4) illustrates a somewhat different profile compared to the section on general SSH data. That is, the data appears to be more scattered, emerging from single, national studies or strategy papers instead of large data sets. Moreover, the emphasis on 'static' documents and general websites is eye-catching compared to "genuine" databases showing queries and search parameter qualities, among others. Furthermore, it was observed that a considerable amount of the data sources were documented in national languages according to the governing organisation. In line with this, the lack of sources in English was, in particular, raised as a concern by one respondent. The following comment given by another respondent summarizes this discrepancy: '*no specific database known. I search via google.scholar and on elsevier, sciencedirect, researchgate etc.'* One potential reason for the observation is that as the EVs have recently commenced to penetrate in most of the national markets during the last few years, there are no established monitoring systems for e-mobility, except for the number of electric and hybrid vehicles that were sold. However, regarding traditional modes of transportation the situation appears to be generally better as reflected by data provided by national statistical offices.

Table 4. Data sources on mobility suggested.

Data source	Type	Qualitative Quantitative	Managed by/ Author	National/ International	Published
Gesellschaft für Konsumforschung (Association for Consumer Research) www.gfk.com/de	Website	Including various information	GfK, market research company	International: company with offices in different nations	Constantly updated
"Regierungsprogramm Elektromobilität" (Government program on electromobility) https://www.bmbf.de/files/programm_elektromobilitaet.pdf	Document: National plan	Qualitative	Government of Germany	National: Germany	2011
Electric Vehicles – Mobility in Turkey https://www.rvo.nl/sites/default/files/2015/06/Electric_vehicles_report.pdf	Document: market report	Mostly qualitative with some numbers	Netherlands-Turkey trade collaboration	National: Turkey	2015
The Tenth Development Plan (2014-2018) http://www.mod.gov.tr/Lists/RecentPublications/Attachments/75/The%20Tenth%20Development%20Plan%20(2014-2018).pdf	Document, policy plan	Qualitative with some numbers	Republic of Turkey Ministry of Developments	National: Turkey	2014
TÜRKİYE SANAYİ STRATEJİSİ BELGESİ 2015-2018 (Turkish Industry Strategy Document 2015-2018) http://sanayi.gov.tr/DokumanGetHandler.ashx?dokumanId=e9f6e3f2-f8ab-4fd1-9d65-22d553867dc1	Document, policy plan	Qualitative with some numbers	Ministy of Science, Industry and Technology	National: Turkey	2015
Transportøkonomisk institutt, TØI (The Institute of Transport Economics) https://www.toi.no/?lang=no_NO	Website	Including various information	The Institute of Transport Economics (Transportøkonomisk institutt, TØI), national institution	National: Norway	Constantly updated
SIFO http://www.sifo.no/	Website	Including various information	Consumer research institute	International: company with offices in different nations	Constantly updated
Norsk elbilforening (Norwegian electric vehicle association) Purchase of electric vehicles: http://elbil.no/elbil-2/elbilstatistikk/	Database	Quantitative	NGO (Norwegian electric vehicle association)	National (Norway)	Constantly updated

Statistik Austria (Statistics Austria) http://www.statistik.at	Website	Including various information	National Statistical Agency	National: Austria	Constantly updated
Statistik Austria Motor Vehicle Stock http://tinyurl.com/z6bucce	Database	Quantitative	National Statistical Agency	National: Austria	Constantly updated
VCÖ: Consumer behaviour http://tinyurl.com/zastdcp	Website	Quantitative	NGO working on environmentally friendly transport	National: Austria	Constantly updated
e-control: prices for gasoline and diesel http://tinyurl.com/h82ksaz	Database	Quantitative	Energy regulation agency	National: Austria	Constantly updated
Personalized routing for multitudes in smart Cities https://doi.org/10.1140/epjds/s13688-015-0038-0	Document: scientific article	Based on quantitative analysis	Scientific author: De Domenico et al.	City of Milan, Italy	2015
INE: Time use for daily movements http://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176815&menu=resultados&idp=1254735976608	Database	Quantitative	National Statistical Agency	National: Spain	(not updated presently, 2011)
Estadística de movilidad de residentes (outdated) (Resident mobility statistics) http://www.fomento.gob.es/MFOM/LANG_CASTELLANO/ATENCION_CIUDADANO/INFORMACION_ESTADISTICA/Movilidad/Movilia2006_2007/default.htm	(unspecified)	(unspecified)	Ministry (Ministerio de Fomento)		2006-2007?
Regional and local statistics(Multiple). Example: Basque Country and Barcelona http://www.eustat.eus/estadisticas/tema_505/opt_0/ti_Movilidad/temas.html; http://www.amb.cat/es/web/area-metropolitana/dades-estadistiques/mobilitat-i-transport	Multiple	Multiple	Multiple	National: Spain	Various
Observatorio de la Movilidad Metropolitana (OMM Urban mobility) http://www.observatoriomovilidad.es/es/publicaciones/informes.html	Document: report	Technical report containing a lot of quantitative information	Centro de Investigación del Transporte Universidad Politécnica de Madrid, various public contributors	National: Spain	Annual report (2014)

Estudio sobre la movilidad eléctrica entre los universitarios (Electric mobility) http://www.transyt.upm.es/index.php/es/	Website	Including various information	Centro de Investigación del Transporte (TRANSyT)	National: Spain	Constantly updated
El transporte en las ciudades (Environmental impact) http://www.greenpeace.org/espana/Global/espana/2016/report/cambio-climatico/AF-movilidad.pdf	Document: report	Technical report containing a lot of quantitative information	Greenpeace (NGO)	National: Spain	App. 2015
AUTOMOCIÓN Y SOSTENIBILIDAD. COMPORTAMIENTO Y ACTITUDES DE LOS ESPAÑOLES (AUTOMOTION AND SUSTAINABILITY BEHAVIOR AND ATTITUDES OF THE SPANISH) http://www.clubsostenibilidad.org/f_publicaciones/sostenibilidadyautomocion%282%29.pdf	Document: report	Research report based on questionnaire containing a lot of quantitative information	Nielsen, research company	National: Spain	2015
Ciudad y movilidadla regulación de la movilidad urbana sostenible (Regulation) https://www.academia.edu/8897967/Ciudad_y_movilidad._La_regulaci%C3%B3n_de_la_movilidad_urbana_sostenible	Document	Research report	Universitat de Valencia	National: Spain	2014
Europe New Mobility Survey GBD Germany http://survey.greenbusinessdevelopment.de/	Documents	Report, quantitative information based on expert survey	Green Business Development GmbH, consulting company	International: Europe (by German-based organisation)	2017
Growing the Digital Business: Accenture Mobility Research 2015 https://www.accenture.com/us-en/insight-growing-digital-business	Document: Presentation	Report, quantitative information based on survey of executives	Accenture, private consulting company	International: global	2015

3.3 Buildings related data

In the third part of the questionnaire, the respondents were requested to assess the data availability on energy use in buildings to serve the second technological focus. Similarly, the focus-area-specific section followed a similar structure as presented earlier in this chapter. That is, the respondents were first requested to assess the data availability and then to specify the sources.

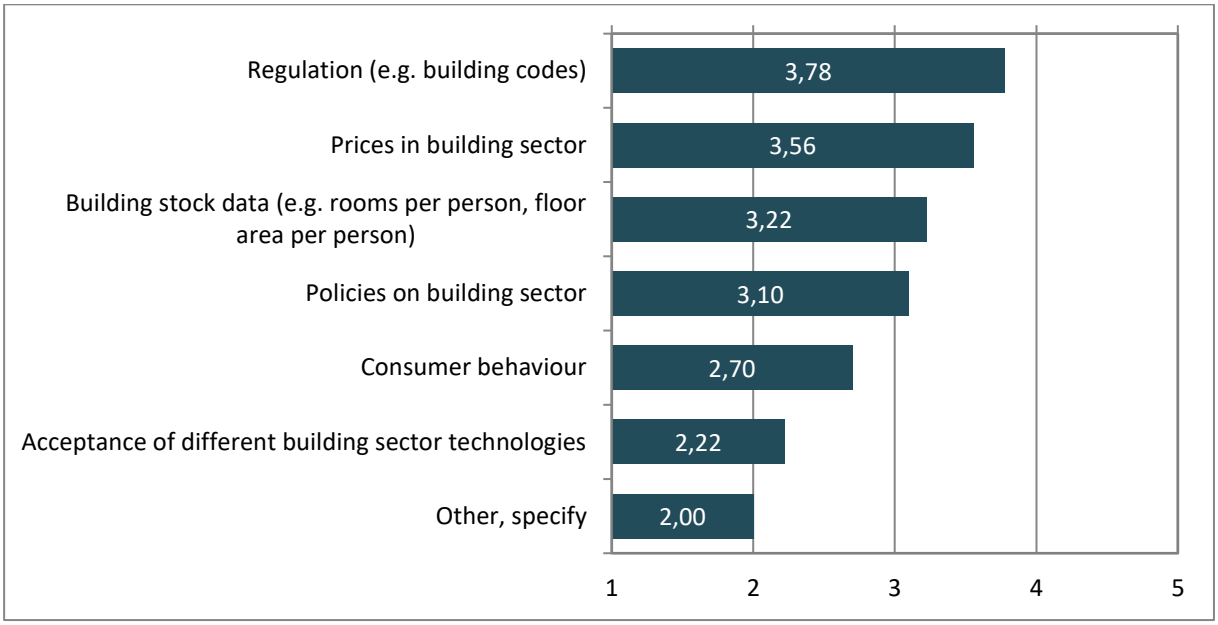

Figure 3. Data availability: Buildings related data (overall mean = 3.07)

Figure 3 presents the respondents' views on availability of data related to energy use in buildings. Compared to mobility, differences detected between the different aspects are higher. Similar to the mobility sector, *Regulation* and *Prices* regarding energy use in the building sector are viewed as areas with the best data availability. However, their order is different with regulation assessed best covered by mean. *Acceptance of different building sector technologies* implicates the weakest data availability, a result similar to the mobility sector. The overall mean of the focus area "energy use in buildings" is 3.07, lower than that of the general SSH data, although not necessarily significant in a statistical sense. There was also one free field answer registered, specified as "*citizen behaviour*".

An overview of the data sources suggested with a focus on energy use in buildings is presented in Table 5. Markedly, there are several data sources identified with functional database properties. This may reflect the fact that buildings are under surveillance of authorities and there exist data-needs for taxation, regulational responsibilities for building safety and construction permits. This might have led to a long-established infrastructure for data collection, apparently often collected by public authorities. Furthermore, as housing prices have a central influence in the development of national economies, and that they do constitute a major share of a typical citizen's budget, and additionally from a consumer's point of view, there is a high interest in price and other data in this area.

Table 5. Data sources on buildings suggested.

Data source	Type	Qualitative Quantitative	Managed by/ Author	National/ International	Published
Turkey Report for Energy and Energy Efficiency "Transition to a Green Economy" http://www.enver.org.tr/UserFiles/CKUpload/Upload/TEVEM_Report_English.pdf	Document	Including various information	Turkey Energy Efficiency Assembly - TEVEM and Energy Efficiency Association - ENVERDER.	National (Turkey)	2010

ENOVA SF https://www.enova.no/	Website	Including various information	Ministry-owned energy efficiency and GHG mitigation office	National (Norway)	Constantly updated
Statistisk sentralbyrå (Statistic Norway) https://www.ssb.no/bygg-bolig-og-eiendom	Database	Quantitative	National Statistical Agency	National (Norway)	Constantly updated
Statistik Austria (Statistics Austria) http://tinyurl.com/zolqzws	Database	Quantitative	National Statistical Agency	National (Austria)	Constantly updated
Austrian Institute of Construction Engineering (Building codes) https://www.oib.or.at/	Database	Qualitative	Industry organisation	National (Austria)	Constantly updated
Immopreisatlas.at (Real estate prices) http://www.immopreisatlas.at/	Database	Quantitative	Banking institute	National (Austria)	Constantly updated
National Program for Renovation of Residential Buildings http://strategy.bg/StrategicDocuments/View.aspx?lang=bg-BG&Id=432	Policy document	Qualitative with some numbers	Ministry/ Governmental	National (Bulgaria)	2005
National programme to improve the living conditions in Bulgaria http://strategy.bg/News/View.aspx?lang=bg-BG&Id=252	Policy document referred in website	Qualitative with some numbers	Ministry/ Governmental	National (Bulgaria)	2009
National plan for buildings with near-zero energy 2015-2020. http://strategy.bg/StrategicDocuments/View.aspx?lang=bg-BG&Id=978	Policy document	Qualitative with some numbers	Ministry/ Governmental	National (Bulgaria)	2015
National Statistical Institute, Republic of Bulgaria Housing types, ownership http://www.nsi.bg/bg/content/3087 http://censusresults.nsi.bg/Census/Reports/1/1/R1.aspx http://censusresults.nsi.bg/Census/Reports/1/1/R2.aspx	Database	Quantitative	National Statistical Agency	National (Bulgaria)	Constantly updated (last 2011)
A literature survey on smart cities, Activity Recognition for the Smart Hospital, CrowdSC Building Smart Cities with Large-Scale Citizen Participation,	Document: scientific article (ScienceDirect)	Scientific articles based on search on ScienceDirect database	Scientific Authors: Yin et al 15; Sanchez et al 08; Karim et al 13; Duca 14	Various	From 2008 to 2015

From energy-efficient buildings to energy-efficient users and back, Ergonomic issues in intelligent buildings design, Home smart home, A danish energy-positive home designed with daylight www.sciencedirect.com			Hansen et al 13		
Estudios del sector de la Construcción - Construction Observatory (until 2001) General studies in the building sector http://www.oect.es/portal/site/Observatorio/menuitem.1a9b11e0bf717527e0f945100bd061ca/?vgnextoid=c548ffbb4b056210VgnVCM1000000705350aRCRD&vgnextchannel=cb322f5af8dc6410VgnVCM1000008130110aRCRD	Website	Links to various studies containing quantitative and qualitative information	Ministry/ Governmental Ministerio del Empleo	National (Spain)	Studies published in different years
Building census 2011 http://www.ine.es/censos2011_datos/cen11_datos_resultados.htm#	Database	Quantitative	National statistical institute INE Instituto Nacional de Estadistica.	National (Spain)	2011
Coyuntura construcción (Monthly construction report) http://www.minetad.gob.es/es-ES/IndicadoresyEstadisticas/Industria/EncuestaCoyuntura/Encuesta%20Coyuntura/dossier%20construccion.pdf	Document	Technical report containing quantitative information	Ministry/Governmental Ministry of Energy, Tourism and the Digital Agenda	National (Spain)	2017 (latest)
Estadística de ejecuciones hipotecarias http://www.ine.es/inebmenu/mnu_financie.htm	Database	Quantitative	National statistical institute INE Instituto Nacional de Estadistica.	National (Spain)	Constantly updated
Ceramics in construction (building materials) http://www.observatoriomercado.es/	Website	Including various information	Market Observatory	National (Spain)	Constantly updated
Grupo InvestitaciónTecnología de la Construcción y Ciencia de los Materiales para la edificación y la obra Civil (Construction technology and materials science for the building and civil engineering) http://www.upm.es/observatorio/vi/index.jsp?pageac=grupo.jsp&idGrupo=188	Website	Including various information	Research centre within UPM company	National (Spain)	Constantly updated

Observatorio tecnológico de la Construcción http://www.otc.construccion.as/cac/portal/home	Website	Including various information	ASPROCON (Construction Confederation) Confederación Asturiana de la Construcción	National (Spain)	Constantly updated
Real estate prices https://www.idealista.com/news/estadisticas/indicevivienda	Database	Quantitative	El idealista	National (Spain)	Constantly updated
Índice de Precios de Vivienda (IPV) (Real estate price index) http://www.ine.es/jaxiT3/Tabla.htm?t=2184	Database	Quantitative	National statistical institute (INE)	National (Spain)	Constantly updated

3.4 Smart energy technology related data

As the fourth part of the questionnaire, the respondents were requested to assess the data availability regarding smart energy technology related data, this being the final technology focus in ECHOES. This focus area includes distributed, small-scale renewable energy production technologies (typically rooftop solar thermal and PV, micro wind, heat pumps and biomass), also in addition to a range of technologies for the traditional "demand side" (e.g. In-home displays, home automation, smart home appliances, new tariffs etc.) and energy storage. This final focus area specific section followed a similar structure as those presented earlier in this chapter. That is, the respondents were first requested to assess the data availability and then to specify the sources.

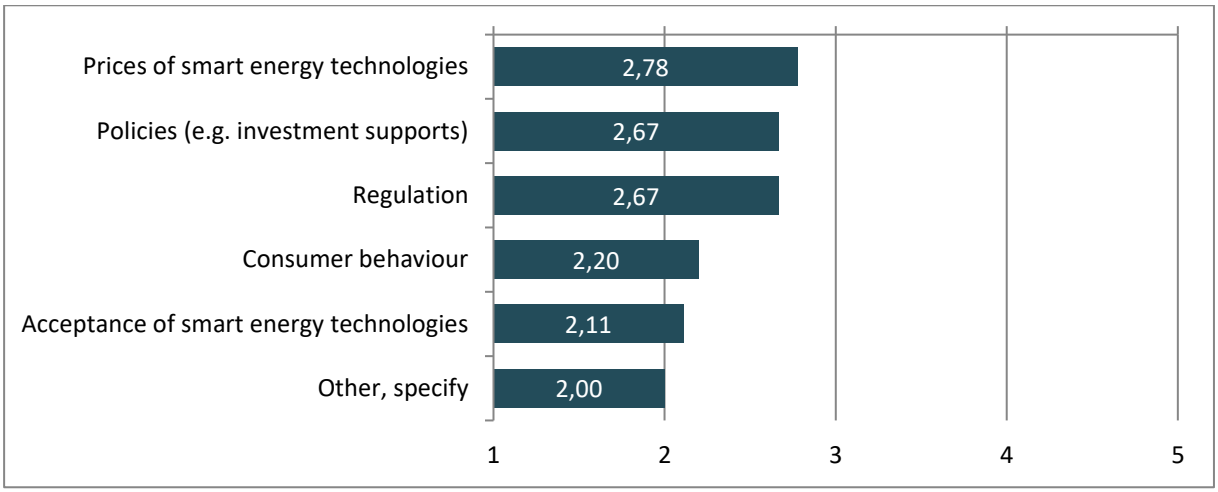

Figure 4. Data availability: Smart energy technology related data (overall mean = 2.47)

Figure 4 presents the respondents' views on smart energy technology related data sources. Immediately, the low assessed value of data availability can be observed. That is, the overall mean is 2.47. Furthermore, the *Acceptance of smart energy technologies* and *Consumer behaviour* in this technology focus are assessed the weakest-covered areas. However, the differences between different areas are not very large. The aspects *Prices, Regulation,* and *Policies* were perceived as having the best data coverage, though still on the negative side of the answering scale. There was also one free field answer registered, specified as "Savings, simulating possiblities".

Table 6 presents an overview of the data sources suggested for smart energy technology. The suggestions can be characterized as scarce. The situation is described by the following free field answer: "*No data on smart energy*

technologies, but various studies and reports on single technologies. [A respondent] has been involved in a number of projects and can come up with cost/price data for different technologies". This, in essence, sums up the nature of smart energy technology related data.

A general speculation, and while applying common knowledge about the reasons behind the low experienced data availability in smart energy technology, is that regarding demand-side technologies, such as household appliances, the technology lifetime is generally small, even only a couple of years (see e.g. Philibert & Pershing (2002), according to IEA (2008)). Therefore, the appliance stock renews quickly and building a constantly up-to-data database appropriately would demand substantial effort. Furthermore, the presence of a substantial variety of manufacturers and models is also worth noting. Typically, however, appliances constitute only a minor impact in typical households' expenditures, and there is a lack of readily available infrastructure qualified to make measurements. In the case of small-scale energy production technologies, the challenge of capturing the rapid development should also be recognized. Given that the penetration of PVs, as an example, in the markets has speeded up basically during the last few years, the proper implementation of them in energy statistics, whose development dates back decades ago, may not have been fast enough in all the cases.

Another key disadvantage for the build-up of databases in the area of smart energy technology is related to the fact that clear definitions are not established. Consequently, very diverse technologies in energy supply chains are used under this term. This may obscure, on the one hand, a respondent's ability to assess the data availability in this area, or, contrarily, make it difficult for the operator of a database to decide what kind of data should be included. Thus, there appear to be several factors that can explain the low availability of data in the area of smart energy technology.

Table 6. Data sources on smart energy technology suggested.

Data source	Type	Qualitative Quantitative	Managed by/ Author	National/ International	Published
Scientific papers: Smart grid opportunities and applications in Turkey http://www.sciencedirect.com/science/article/pii/S1364032114001208 A demand side management strategy based on forecasting of residential renewable sources: A smart home system in Turkey http://www.sciencedirect.com/science/article/pii/S0378778814004460	Documents: scientific articles	Including various information	Ilhami Colak Ramazan Bayindir, Gianluca Fulli, Ibrahim Tekin, KenanD emirtas, Catalin-Felix Covrig A.Tascikaraoglu, A.R. Boynuegri, M. Uzunoglu	National (Turkey)	2014, 2014
SIFO http://www.sifo.no/index.php?factory=index	Website	Including various information	Consumer research institute	International company with offices in different nations	Constantly updated
TopTen.eu – Best Products in Europe www.TopTen.eu	Database, website listing e.g. energy efficiency characteristics for products	Quantitative	European Commission, WWF, European Climate Foundation, VDE Institute	International (Europe)	Constantly updated

e-control: (Energy regulation Agency, Mostly regulation) www.e-control.at	Database	Quantitative	Energy regulation agency	National (Austria)	Constantly updated
Actors and factors in land-use simulation, The challenge of urban shrinkage; Potential energy savings by radiative cooling system for a building http://www.sciencedirect.com	Document: scientific article (ScienceDirect)	Scientific articles based on search on ScienceDirect database	Haase et al 12 Hanif et al 14		2012, 2014
Young prosumers - niños y adolescentes agentes del cambio social a través de la creación de contenidos digitales https://idus.us.es/xmlui/handle/11441/39092	Document: doctoral thesis	Scientific thesis containing quantitative information	Paula Herrero Diz Universidad de Sevilla	National (Spain)	2016
ENERFICIENCY project (2011 - 2014) (Energy efficiency) http://exit.udg.edu/enerficiency/	Website	Including various information	University Universitat de Girona /research consortium	National (Spain)	2011-2014
Spanish Technological Platform of Electrical Grids FUTURED (Smart energy technologies) http://www.futured.es/que-es-futured/	Website	Including various information	Platform to integrate all of the agents involved in the electricity sector	National (Spain)	Constantly updated
Centro de Servicios Inteligentes para la Gestión Energética. Center for Intelligent Services for Energy Management. (Smart energy technologies) http://www.csige.es/portfolio/foro-energia-inteligente/	Website	Including various information	Sevilla, industry	National (Spain)	Constantly updated

4 IN-DEPTH ANALYSIS OF SELECTED DATA SOURCES

According to the initial survey, a large number of databases – or data sources – were suggested. In this chapter, a closer look at a sub-group of these suggested data sources is taken into account. This allows for us to draw upon more focused conclusions and enables an in-depth analysis of, among other areas, the formats, technical implementation and accessibility.

4.1 Justification of the selection

Previously in this study, based on the responses from 13 researchers of the ECHOES project, we identified characteristics of data sources in SSH data and performed classifications. A not-so-trivial question is then, how to select the data sources from the long list of suggestions, based on the characteristics described above.

4.1.1 Classification of suggestions presented by the respondents

The answers regarding the data sources - generally and in foci areas - consisted altogether of 24 international suggestions and 84 national suggestions (Table 7 and Table 8). Firstly, all the suggestions were somewhat superficially reviewed. As a conclusion of this review, it was observed that data source suggestions showed a relatively wide spectrum of contents. Some answerers specified the data source exceedingly precisely by a link, whereas in other answers the data source was described in general terms by organisation name (e.g. “IEA”, “EU”, “OECD”). In line with this, the international data source suggestion were divided in “SSH data emphasis”, “Energy data” and “General” categories.

Table 7. International data source suggestion categorized by number. The highlighted green areas are of focus of this chapter.

	Number of suggestions	SSH data emphasis	Energy data emphasis	General	Of which national statistic offices
General (international)	24	5	4	15	

In accordance with the review of suggestions as collected from the respondents, the abundant general statistics infrastructures (Table 3) by, for instance, the United Nations or Eurostat are seen as non-optimal as a starting point to study the nature of SSH data related to low-carbon transformation. This is due to the fact that websites of data providers operating with wide focuses (such as the EU or the UN), without specifications to SSH or low-carbon related data, are viewed as inappropriately generic. Contrastingly, it is perceived as more efficient to start the analysis from databases that are specifically designed for SSH or low-carbon energy related questions. It must be emphasized that the afore-presented discussion does not signify that the general international data sources do not contain data relevant for the ECHOES project. Notwithstanding, in the initial phase dealing with definitions and high-level characterisations of the data sources, a review of more specific databases is preferred. This is due to the fact that, to a greater extent, the ECHOES project database development shares similar properties.

The national suggestions (appendix, Table 4, Table 5, Table 6) allowed different, more specific categorisation that is summarized in Table 8.[7]. Table 8 confirms that national statistical offices are clearly the most commonly

[7] One source (“Norsk elbilforening”) moved from “General” to “Mobility” category by the author

suggested source. Accordingly, the suggestions were categorised using the categorisation: (1) *Websites* including links to many forms and sources of data, (2) *Documents* group consisting of, among others, policy and strategy reports, scientific articles and market reports by consultancy organisations. (3) *Databases* presenting the most "genuine" characteristics in respect to database definition. In other words, they for instance allow for making queries, or downloading data in an organised form. Despite not being unambiguous, the categorisation chosen assists us with focusing on the data sources of significant relevance with regards to the target of this report

Table 8. National data source suggestion categorized by number. The highlighted green areas are focus in this chapter.

	Number of suggestions	„Website"	„Document"	„Database"	Of which national statistical offices
General (national)	32	16	5	11	6
Mobility	23	8	11	5	3
Buildings	20	5	6	9	6
Smart energy technology	9	4	3	2	0
Total (national)	84	33	25	27	15

As this report has a focal point of defining relevant characteristics for the build-up of databases of the ECHOES project, the *Databases* group was assessed as a natural primary target and focused group for the subsequent rigorous analysis. *Documents* as such may be interesting sources for the project but are not viewed as the primary class of sources to help in *designing* databases given their static nature. That is, documents originating from single studies or policy strategies may be very precise even for scientific standards. Despite this, recurrently, documents focus on narrow topics and they lack the option of sending queries and searches while providing no possibilities to organise the data according to user's choices. Therefore, despite the existence of documents that could be utilised as *content* of a database, they are barely instrumental in considering architecture or protocols needed in order to build up a database. In contrast, general *websites* with scarce specifications related to SSH or low-carbon energy related data are viewed as generally having too broad of a focus for the purposes of this report. In other words, given that any type of digital information could technically be included in generic websites, their appropriateness as sources for helping with the *design* of the database of the ECHOES project is viewed as secondary.

4.1.2 The data sources of primary interest

Resulting from the discussions in section 4.1.1, we have conclusively picked our selection of databases analysed in detail, from international and national data sources in the general segment of the survey. Specifically of international, more generic sources, the primary focus is targeted on those possessing SSH data of energy data focus. With regards to national sources, the sources with most genuine database characteristics are of primary interest.

There are *5+4= 9* international databases with focuses on SSH data and energy, respectively (presented in Table 10) that are seen to be of primary interest. Five international data sources are classified with an emphasis on SSH data: International Social Survey Programme, European Values Study, World values survey, Gesis Sowiport, World Input-Output Database, and, respectively, four data sources with an energy data emphasis, are suggested: ACEEE (American Council for an Energy-Efficient Economy), International Energy Agency (IEA), BP (formerly British Petroleum), EIA (U.S. Energy Information Administration).

Table 9. International databases

	Name	Content focus of database
1	International Social Sorvey Programme, ISSP	SSH
2	European Values Study	SSH
3	World values survey	SSH
4	Gesis Sowiport	SSH
5	World Input-Output Database	SSH
6	ACEEE	Energy
7	International Energy Agency (IEA)	Energy
8	BP	Energy
9	EIA	Energy

National general data sources suggestions (see appendix) included 32 alternatives in our sample of databases. 11 of the suggestions were classified as "Databases", 16 as "Websites", 5 as "Documents" by the authors of this report. Documents and websites were directly ruled out as irrelevant for further in-depth review due to their static and non-specified nature, as previously discussed above. Besides the national statistical offices, there are 5 databases in general, 2 in mobility, 3 in buildings and 2 in smart energy technology that are suggested. Therefore, the number of databases is deemed satisfactory in order to review the contents of the databases more closely. It is apparent that the number of databases, especially in foci areas, is too small to make any general conclusions of database characteristics in these areas. However, this is not the primary focus in the current explorative study. Thus, the results must be viewed as descriptive and as a basis for generating ideas for future phases of the project

Besides the national statistical offices, there are five general databases, that is, the Socio-Economic Panel (SOEP), ALLBUS, ISOC, Legal Encyclopedia, and the Council of Ministers' Legal Information System. Their characteristics are described in Table 10. Correspondingly, the six databases in foci areas are described in Table 11.

Table 10. General national databases

Data source	Qualitative/ Quantitative	Managed by/ Author	Nationality	Published
Socio-Economic Panel (SOEP) https://www.diw.de/soep	Quantitative	German Institute for Economic Research, DIW	National (Germany)	Constantly updating
Allgemeine Bevölkerungsumfrage ALLBUS The German General Social Survey www.gesis.org/allbus	Quantitative	GESIS - Leibniz-Institute for the Social Sciences	National (Germany)	Constantly updating
ISOC - Ciencias Sociales y Humanidades Research database http://bddoc.csic.es:8080/informe.html;jsessionid=19E2DDD4FEDAA408405F1	Reference database on studies on social sciences and humanities	Center of Humanities and social sciences	National (Spain)	Constantly updating

BF663FCBEDF?total=23&bd=ISOC&tabla=docu General/				
Legal encyclopedia http://www.juen.bg/	Including various legal information	Governmental	National (Bulgaria)	Constantly updating
Council of Ministers' Legal Information System http://pris.government.bg/prin/login.aspx?ReturnUrl=%2fprin%2fdefault.asp	Decrees, orders and decisions	Governmental	National (Bulgaria)	Constantly updating
Turkish Statistical Institute (TurkStat) http://www.turkstat.gov.tr/Start.do;jsessionid=lyT8YzTLRQrWwkNwGQ2r1JwW7TpH1Y3bnppZnGrTQv2JLj26C1Tt!1193456523	Quantitative	National statistical office	National (Turkey)	Constantly updating
Statistics Norway https://www.ssb.no/en/	Quantitative	National statistical office	National (Norway)	Constantly updating
Statistik Austria National statistical office www.statistik.at	Quantitative	National statistical office	National (Austria)	Constantly updating
Statistisches Bundesamt (Germany) https://www.destatis.de	Quantitative	National statistical office	National (Germany)	Constantly updating
El Instituto Nacional de Estadística (INE) http://www.ine.es/welcome.shtml	Quantitative	National statistical office	National (Spain)	Constantly updating
REPUBLIC OF BULGARIA National statistical office http://www.nsi.bg/en	Quantitative	National statistical office	National (Bulgaria)	Constantly updating

Table 11. National databases in foci areas.

Data source	Foci area	Qualitative Quantitative	Managed by/ Author	National/ International	Published
Austrian Institute of Construction Engineering (Building codes)	Buildings	Qualitative	Industry organisation	National (Austria)	Constantly updated
Immopreisatlas.at (Real estate prices)	Buildings	Quantitative	Banking institute	National (Austria)	Constantly updated
Real estate prices	Buildings	Quantitative	El idealista	National (Spain)	Constantly updated
e-control: prices for gasoline and diesel	Mobility	Quantitative	Energy regulation agency	National (Austria)	Constantly updated

Norsk elbilforening (Norwegian electric vehicle association) Purchase of electric vehicles	Mobility	Quantitative	NGO (Norwegian electric vehicle association)	National (Norway)	Constantly updated
TopTen.eu – Best Products in Europe	Smart energy technology	Quantitative	European Commission, WWF, European Climate Foundation, VDE Institute	International (Europe)	Constantly updated
e-control: (regulation)	Smart energy technology	Quantitative	Energy regulation agency	National (Austria)	Constantly updated

4.2 Analysis

4.2.1 Which indicators, variables or other data is stored in the databases?

A rigorous review of the SSH based studies indicated that their contents generally build on repeated questionnaires. Model-based or calculational tools are usually used to generate data in energy databases and in economic and industrial data, respectively. The main content of the databases is reviewed in the following:

4.2.1.1 International databases

The **International Social Sorvey Programme, ISSP** (http://www.issp.org) is a cross-national collaboration programme conducting annual surveys on diverse topics relevant to social sciences with founding members of Australia, Germany, Great Britain and the US. Since 1984, the ISSP has included 57 nations. The list of topic modules approached consists of Role of Government, Social Networks, Social Inequality, Family and Changing Gender Roles, Work Orientations, Religion, Environment, National Identity, Citizenship, Leisure Time and Sports, Health and Health Care with varying subtopics in the surveys conducted.

European Values Study (http://www.europeanvaluesstudy.eu/), according to its website, is the most comprehensive research project on human values in Europe. It is a large-scale, cross-national, and longitudinal survey research program based on how Europeans think about family, work, religion, politics and society. Repeated every nine years, in an increasing number of countries, the survey provides insights into the ideas, beliefs, preferences, attitudes, values, and opinions of citizens all over Europe. Respondents answered about 250 questions in the interviews.

World values survey (http://www.worldvaluessurvey.org/wvs.jsp) is a global survey on changing values and their impact on social and political life. Established in 1981, the survey consists of nationally representative surveys conducted in almost 100 countries and contain almost 90 percent of the world's population. The data is produced using a common questionnaire and a network of scientists.

GESIS Sowiport (http://sowiport.gesis.org/) is an online portal for social sciences on references to publications and research projects that are searchable by queries. According to its website, Sowiport currently contains approximately 8 million references on publications and research projects. Sowiport is based on projects funded by the German Research Society and the Federal Ministry for Education and Research.

World input-output database (http://www.wiod.org/home) is based on EU Commission funded project run in 2009-2012. The core contents consists of the data on economic, industrial sector inputs and outputs. More specifically, the core of the database is a set of harmonized supply and use tables, alongside with data on international trade in goods and services. The data covers 43 countries. As an addition, in accordance with the existing knowledge of the authors of this report, there also exists commercial **GTAP (Global Trade Analysis Project)** input-output database by Purdue University (https://www.gtap.agecon.purdue.edu/default.asp) with similar type of characteristics.

ACEEE (http://database.aceee.org/) State and Local Policy Database includes comprehensive information on energy efficiency policies currently implemented at the state and local level in the United Stated. The contents are text-based descriptions and rankings of the energy efficiency policies and it is updated annually.

BP data is provided by an energy company known formerly as British Petroleum. The contents include national annual energy balances, production, capacities, energy resources, prices and scenarios on energy futures. The data is updated by an annual energy outlook of the BP.

IEA data services (http://data.iea.org/) are very widely applied data sources in energy research. Variables available include typical variables in energy statistics, such as energy balances, price information, or CO_2 emission information, in annual, monthly or quarterly time resolution. The database enables country-or area specific queries on historical values.

EIA data (https://www.eia.gov/opendata/) are provided by U.S. Energy Information Administration. The contents consist of hundreds of thousands of data sets on energy system data, among others, such as hourly electricity operating, crude oil imports and natural gas series..

4.2.1.2 National databases

ALLBUS (The German General Social Survey) is a very large and versatile social science database. According to its description, it collects up-to-date data on attitudes, behavior, and social structure in Germany. The survey is conducted biannually: that is, every two years since 1980 a representative cross section of the population is surveyed using both constant and variable questions. The methods, questionnaires, as well as data are downloadable from the website. A large portion of the materials are translated to English. The database is large, for example, the database for 2014 ALLBUS alone, consist of 861 variables and 3471 units.

The German Socio-Economic Panel (SOEP) (https://www.diw.de/soep) study seems to having similar characteristics as a broad study based on household samples. It is a wide-ranging representative longitudinal study of private households, located at the German Institute for Economic Research, DIW Berlin. Every year, nearly 11,000 households, and more than 20,000 persons are sampled.

Of the foci-specific databases, the database maintained by **Austrian Institute of Construction Engineering** (https://www.oib.or.at/de/datenbanken) contributes to issues on qualitative databases to the sample. In particular, the database contains different documents on building codes and regulations. Almost 51,000 objects are stated to be registered, maintained and updated (Figure 5). The database allows information that can be found easily and quickly via searches. The documents and data search tools have been implemented in the German language and there are different parameters, for instance, on products and product numbers according to how the search can be fine-tuned. Additionally, **Immopreisatlas.at** is a database on real estate prices in Austria. **TopTen.eu** allows for making searches on most energy efficient products in different European countries. **E-control**, an Austrian energy regulator maintains statistics on energy and fuel prices. **Norsk elbilforening** ("Norwegian electric vehicle association") provides data on purchases of electric vehicles.

Figure 5. A screenshot of the databases maintained by the Austrian Institute of Construction Engineering. (https://www.oib.or.at/datenbanken/uea-expert)

"Legal encyclopedia" and "**Council of Ministers' Legal Information System"** (unofficial translations) include documents on various legal pieces of information; that is, they contain decrees, orders and decisions from Bulgaria and are available in the Bulgarian language.

National statistical offices carry out national mandates such as responsibility of national data delivery to European statistical officials. In the survey, the statistical officials of Norway, Austria, Germany, Turkey, Bulgaria, and Spain were explicitly mentioned. As a result, these data providers are subject to many standardized procedures and methods in statistical production. As national statistical offices are the main public provider of data in European countries, it is not surprising that many of the statistics produced by them fall under topics of the ECHOES project such as price data, housing, transportation, economy, population, energy, labour market. As an example, the main categories of statistics provided by Statistics Norway are depicted in Figure 6. A detailed review of all the statistics is neither possible nor necessary for this report. For example, the statistical authority of Finland, Statistics Finland maintains 154 official statistics and contain over 2000 open data tables.

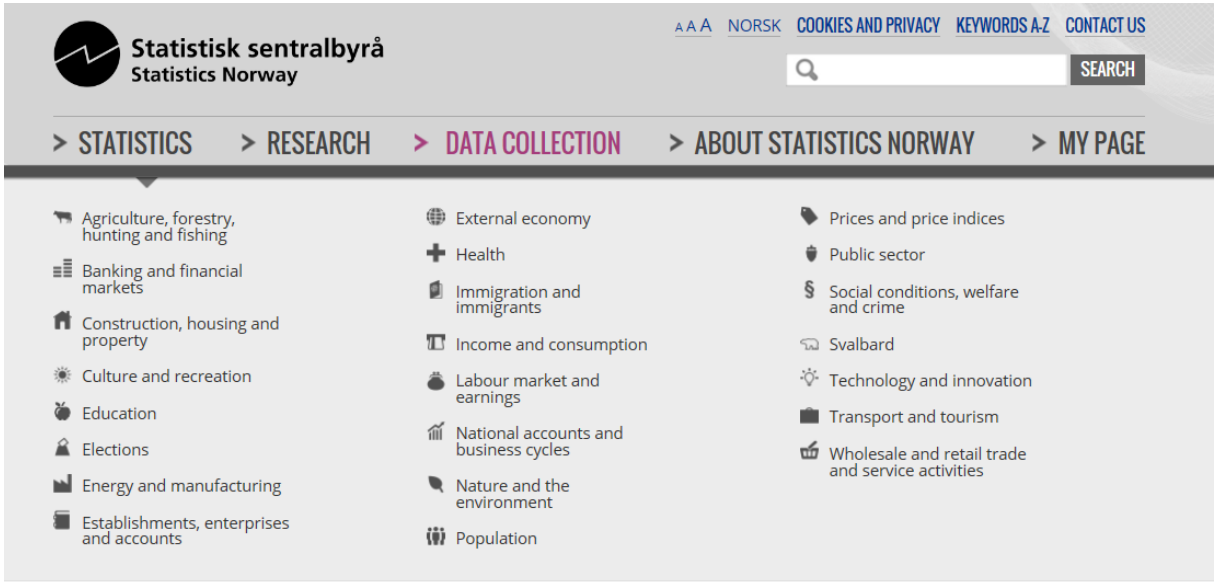

Figure 6. Main categories of statistics provided by Statistics Norway (https://www.ssb.no/en/innrapportering)

4.2.2 Are there any restrictions of use?

In order to build databases and develop indicators, collecting, synthesizing, and formulating relevant data and indicators related to individual and collective energy choices and behaviour, restrictions of use of databases may imply a barrier for data utilisation in the ECHOES project and therefore need to be taken into account.

As examples of data subject to terms and conditions, the IEA data is subject to terms and conditions and only for a small part available as publicly and free of charge accessible database. This is also true of the BP data. For example, in the 2017 edition of the annual "The statistical Review of World Energy" by BP, publishers are welcomed to quote from this review provided that they attribute the source to BP Statistical Review of World Energy 2017. However, restrictions are set by a note, saying"… where extensive reproduction of tables and/or charts is planned, permission must first be obtained …"

The ALLBUS, in contrast, is presented as being made available to interested parties for research and teaching immediately after processing and documentation. National statistical offices, as public authorities, generally, have an in-built property of many statistics all made available free of charge. This seems to be valid regarding other publicly funded sources as well (e.g. GESIS Sowiport). However, the terms must be meticulously checked case by case. In some cases, even with the licenced use, the terms and conditions may restrict the use of data and other information for non-commercial purposes only. As an example, publisher of a peer-reviewed article may restrict the use of it for commercial purposes, which further indicates that the terms and conditions of ECHOES database should be also be thoroughly checked.

4.2.3 Data formats and technical implementation

Another, more indirect barrier for data use is if the available data is not easily downloadable or query functionalities are substandard. In our studied small sample, for example, the quantitative data services such as IEA, EIA, ALLBUS databases, SOEP, as well as those provided by national statistical offices were obviously largely designed for data download and utilisation. Additionally, they generally contain versatile options for data utilisation formats. As such, this issue is not perceived as a major threat in these cases. The search functionalities were versatile also in the

case of database maintained by Austrian Institute of Construction Engineering. However, as this database contains mostly qualitative data (documents), it is apparent that the functionalities will differ to some extent (see Figure 5).

4.3 Conclusion

A closer look at a small selection of databases allowed us to make preliminary conclusions on database issues to be acknowledged in database development of the ECHOES project:

- Generally, suggested sources of data take on many forms. These vary from ”websites”, ”documents”, ”data providing actors/organisations”, to the most developed, “genuine databases”.
- Even the initial survey conducted suggests that there are extensive options for SSH data sources. The most commonly presented group are national statistical offices operating in different European countries. The most feasible choices for the ECHOES project database are to be determined finally in later phases of the project. A brief discussion for the next steps is provided in the following chapters.
- The potential space consists of both databases that are free-of-charge and that are subject to a charge within those considered to have potential. Consequently, terms of use need to be carefully reviewed. According to the working assumption of the ECHOES consortium, generally, to simply download other databases cannot be considered as an extensively available option. One potential solution proposed is to rather provide links/interfaces between the ECHOES database and selected others.
- A “link list” of “website” type of storing data appears the most convenient and flexible solution in order to include various types of information. However, this is highly maintainence intensive as links expire quickly and new links need to be included.
- Attention has to be paid on what is needed to qualify as a “sufficient” database. In all cases, choices such as metadata definitions, user interface design, categorising,and making the database instrumental, are essential. The analysis of sources collected for this initial study clearly demonstrated this need.

5 DEVELOPMENT NEEDS

5.1 General

As the ECHOES project aims at developing an open access SSH database focused on energy system transition, this initial report works as a tool to help its design based on state-of-the-art within the area. In particular, the national and EU level databases identified are used as a starting point to evaluate the existing gaps in quantitative data, which is subsequently required in policy and other decision-making and research communities. Therefore, the development requirements that facilitates SSH data being more instrumental is viewed as consisting of (i) data management and open data practices (ii) substance-related needs to improve the gaps in knowledge.

5.2 Data management needs

In response to the call for open data management, the European Commission has launched s.c. FAIR Data Management Principles (Findable, Accessible, Interoperable, and Re-usable). In light of these principles we are able to reflect on the suggested SSH data sources. Generally, the data located at national statistical offices' databases could be viewed as having a built-in incentive for a wide-range of usage. However, as there are several data sources related to commercial databases or those under other usage restrictions, caveats for open data management can immediately be recognized.

As the analysis revealed a wide variability of SSH data sources in terms of content and technical implementation, a need for interoperability should be carefully tackled in database development. Furthermore, as the sources included were collected with various methods and level of ambition, an inspection of the reliability should be taken into consideration prior to including them.

Generally, collecting, processing and maintaining – potentially large – SSH data sets requires substantial effort. Despite this, generally speaking, the data collected and produced within an individual project is not more widely shared and re-used within other projects or in public or private applications. Barriers for open data practices may arise from various sources –for example, commercial, competitive, conflicting interests or privacy reasons. Additionally, lack of incentives or suitable infrastructures may hinder data sharing. Based on the initial surveys, this issue is of relevance in the SSH area. More specifically, restrictions were identified on the use of the recognized data sources by the respondents.

Based on the afore-mentioned principles, there are several examples of EU H2020 projects that have made data sharing available in a sufficient manner. Based on internal scanning work at VTT, with a reliance on expert on knowledge services, some EU based, international and national examples are listed, providing a pool of options for the ECHOES project:

- Generic repositories:
  - **Zenodo**: funded by CERN & EU OpenAIRE project
    - This database is recommended; it is mentioned in the Guidelines for FAIR Data Management in Horizon 2020; furthermore, it is among the EU OpenAIRE recommendations for finding a research data repository
    - Additionally it is possible to deposit only metadata
  - EUDAT: a research data infrastructure initiative funded by European Commission
    - **B2SHARE**: simple solution for small-scale sharing of research data
    - B2SAFE: robust and safe replication service for local repositories with PIDs and metadata in place

    - **Figshare** (by UK based Digital Science & Research Ltd): offers easy-to-use data depositing and publishing solution
- Discipline based repositories: http://www.re3data.org/

As a next step in ECHOES database development, formats and specifications of data that will be finally delivered and filtered through from the other WPs, are to be determined. Moreover, taking the data requirements that are implied in the above-mentioned list into account, technical details/formulation/protocols will be consulted with database experts. Accordingly, more specific questions for the WP will be sent as soon as a concrete plan/suggestion can be developed. The work done so far in categorising data as well as scanning for the options can be utilised in this work. An initial proposal is presented in "Conclusions" chapter of this report.

5.3 Substance needs brought up by the surveys

The surveys suggested that several SSH data related substance areas appear to actually be covered. Typically, national statistical offices maintain wide population data registers. Not surprisingly, population was assessed as the best-covered topic, scoring high in almost every one of the eight European countries represented in the survey.

However, the linkage between low-carbon energy transition research and these rather generic databases is more challenging. Generally, areas of energy cultures, consumer behaviour, acceptance, and gender issues were assessed as areas where data coverage from existing sources is the weakest. The areas of acceptance and consumer behaviour were, in particular, additionally raised in the technological foci: buildings, smart energy technology, as well as (electrical) mobility.

Many answers highlighted the need/interest on data on energy behaviours (mobility, consumption, energy use in buildings) linked to:

- social norms, moral norms, values
- knowledge on environmental issues
- technology change

More specifically, independent of the attitude towards the technologies in the foci, actual follow-up in behaviour might not be reflected. The respondents identified a clear need for datasets in this area.

The need of synergetic dealings with the SSH data was also brought up in the survey. Firstly, this was reflected by an identified call for analysis of interdependencies between different consumption areas. Secondly, in addition to, among others, statistics on key technologies and cases, a need for databases that synthesize case studies in a good way, was raised.

One respondent highlighted the general lack of data for human and organisational factors: '*qualitative data regarding human factors (decision-making, user experience, etc.) as well as qualitative data related to organisational factors are needed. Most gaps seem to be related to new technology as the perspective is then easily focused on technical matters and the human or social perspectives are missing (with the possible exception of hype themes).'*

6 CONCLUSIONS

6.1 General

The aims and policy targets of the European Union set the scheme of the transition towards a low carbon and consumer-driven energy system. It is therefore evident that better understanding is needed on human behaviour and social cultures in formulating successful policies and fair governance of all of the citizens across the EU. However, the existing methodologies, analysis methods, and data that combine social science and humanities with techno-economic and socioeconomic analysis are clearly lacking especially when considering that the policy impact analysis barely takes these SSH dimensions into account. The aims of the WP2 of the ECHOES project is to take important steps to fill this gap by collecting and verifying SSH related data and indicators relevant for sustainable low carbon energy system transition. The final objective is to create an open-access database that includes both quantitative and qualitative SSH data. The ECHOES database will include relevant data (or www-links to other databases) collected from other open-access databases and all the open data that will be created within the ECHOES project. The results of the WP2 will give answers to the following questions:

- How to combine the best practises used in human-centred research approach with quantitative SSH data collection and by using quantitative *ex-ante* and *ex-post* analysis of energy system transition and governance?
- How to combine human-centered qualitative data with quantitative techno-economic and socio-economic data?
- What are the recommendations based on WP2 results to better include SSH dimension in the analysis of sustainable energy system transition?

A good example of inadequate methods and data is related to energy poverty. As part of the Energy Union Governance process presented in the 2016 Winter Package, Member States will have to monitor and report on energy poverty. However, it is recognized that monitoring is not possible with today's information and in order to fill the gap the Commission will need to facilitate the exchange of best practices and coordinate these monitoring efforts at the EU level via an Energy Poverty Observatory. Furthermore, this example is only related to ex-post analysis of made decisions and what presents itself as even more challenging is the need to link SSH related data and methods to our studies on future energy systems and, finally, to more efficiently link SSH foci to governance processes of energy system transition.

According to this report, the nature of SSH data can be described as fragmented, dispersed, consisting of varying technical implementations, varying languages, as well as of varying formats. Furthermore, providers of data portray a large variety. Particularly, national statistical agencies are mentioned frequently while ministries are another typical source. The spectrum of other data providers ranges from research institutes to private companies to NGOs and industry organisations. The reliability of classified data sources is not self-evidently homogenous between all different sources. This issue should be considered in the development of the database of ECHOES project.

Generally, potential sources of data take on several forms. These vary from ”websites”, ”documents”, ”data providing actors/organisations”, to the most developed “genuine databases” that allow queries, searches and organisation of the data. Quantitative databases, are typically run by national statistical agencies or institutes established to conduct a certain persistently important statistical task in society. Scientific bibliographical databases, on the other hand, are typical examples of qualitative databases that are identified.

Substance-wise, the results suggest that many SSH data related areas appear to be covered. Typically, national statistical offices maintain wide population data registers. In line with this, population was assessed as the best-

covered topic, scoring high in almost all of the eight European countries represented in the survey. However, a lot of gaps were identified, too. Generally, areas of energy cultures, consumer behaviour, acceptance, and gender issues were assessed as areas where data coverage is weakest. The areas of acceptance and consumer behaviour were mentioned repeatedly within the more specific analyses in the ECHOES technological foci: energy use in buildings, smart energy technology and electric mobility. Bearing in mind that the sample size of the questionnaire is small, the results are to be primarily used, among other things, in characterizing the nature of data and generating hypotheses, instead of deriving general results suitable for more generic conclusions. Therefore, this report is primarily serving the purpose of preparing the work on the ECHOES database.

Considering the development of the project's own database, "a link list" of "website" types of storing information appears to be the most convenient and flexible solution for the inclusion of the various types of information. However, a question of what is required to qualify as a "sufficient" database needs to be resolved. Moreover, there are both databases free-of-charge and subject to a charge among those considered having potential. Consequently, terms of usage need to be rigorously reviewed before secondary data is included in the ECHOES database.

All encompassingly, there is a need to pay attention to choices such as metadata definitions, user interface design, categorising, in order to make the database instrumental. The analysis of sources collected for this initial study clearly demonstrated this need. This report identified several EU, international and national alternatives as the building blocks of the data management infrastructure, and the strengths and weaknesses of which should be considered in the forthcoming phases of the project.

As a next step in the development of the ECHOES database, formats and specifications of data that will finally be delivered from the other WPs, will be determined. Paralleling this, and while taking the EU's requirements into account, the technical details/formulation/protocols will be consulted with database experts. Accordingly, more specific questions for the WPs will be forwarded as soon as a concrete plan/suggestion can be developed. The work done so far in categorising data as well as scanning for the options can be utilised in the forthcoming work. The following sections preliminary outline contemplations of the following phases. However, the suggestions are preliminary and open to discussions and suggestions for improvement.

6.2 Relevance of the results for database development

The responses of the initial questionnaire strengthened the view of an exceedingly wide spectrum of what is available as SSH data, and what can be included in this category. The suggestions included documents of varying quality, highly developed data infrastructures containing quantitative data such as national statistical offices, as well as websites of various organisations with no further specifications of data.

Accordingly, it is concluded that the mechanical storing of every suggestion in the database is not considered a feasible solution to contribute to making the ECHOES database user-friendly and utilisable. More importantly, there is a compelling need to highlight the added value of the ECHOES database in relation to existing data. While this task is non-trivial, as a necessary condition, the data needs to be categorised and organised. Two main approaches could fulfil this need: (i) "Top-down" approach: sharply specifying the definition of the ECHOES relevant SSH data and ask for data providers to follow the specifications (ii) "Bottom-up" approach: the data providers are requested to specify and define the types of their data, and the architecture to classify and organise data is designed upon this basis.

The initial survey suggested that there are varying definitions of qualitative and quantitative SSH data, additional to as the differing perceptions of borders between the foci areas in the ECHOES consortium. This can be viewed as an expected result for ECHOES as a multi-disciplinary project with researchers having varying backgrounds. Seemingly apparent is the unlikelihood that all the open questions related to data types, formats, and specifications,

as well as those related to user interfaces could be conveniently solved with a "command and control" type of operation. In this sense, an iterative approach is viewed as having superior advantages.

6.3 The road ahead and open questions

With the preliminary D2.1 results readily accessible, the more concrete database solutions were discussed with experienced database experts working at VTT. Six key points emerged from the discussions:

First, the functionalities of Zenodo data architecture were briefly reviewed in line with the ECHOES database goals. At a first glance, it was deemed sufficiently versatile and as a suitable platform for the ECHOES. Furthermore, the use of the Zenodo database was among the recommendations made by VTT knowledge service experts as well as fulfilling the open data principles of the H2020. However, should there be any alternatives or further suggestions from the consortium; VTT is very open to discussions on other alternatives.

Second, according to preliminary information, at the very least, the following type of data is to be produced in the different Work Packages of the ECHOES project.

Types of data in ECHOES

- Anonymized interview transcripts
- Database of relevant literature
- Policy list analyzed
- Data-input for meta-analyses (basically correlations between key variables extracted from different scientific sources)
- Experimental data (quantitative responses to questionnaires)
- Main survey data (Here, the consortium has decided to engage a professional company to host, recruit, collect and clean data, and translate text. The survey will cover at least 9,000 households in 30 countries.)

The list of potential ECHOES data, in accordance with current information, was introduced and discussed with experienced database experts working at VTT. Using a technical requirements point of view for the database, and on a rudimentary level, the data most currently, can be divided based on the following three types:

Categories of different data in the database

- "Strings", consisting of verbal information (e.g. interviews)
- "Numbers", meaning scalable data that can be integers or decimals. There can be many origins of data of this type (i.e. answers to questionnaires, model results, measured data etc.). However, from a technical data storing and processing point of view, the origins do not play major role.
- "Literature reference data". There are several examples of potential data fields for this type of data (e.g. Author, Publication year, Journal/Report series, Title, link, keywords)

Third, based on the aforementioned lists and discussions, the technical issues on data formats in any of the identified groups are not seen as the main sources of concern. Two options for technical implementation for data

storing were identified (i) manual storing by VTT or by an interface to be designed, or (ii) an automatized solution based on code implemented in the database software between the data stored in NTNU's server. Naturally, the larger the data to be stored is, the more emphasized are the assets of an automatized solution. In both cases, a structured form of data is viewed as necessary. This is of significant relevance at least when related to the main survey data.

Fourth, irrespective of technical details, structuring the data to be stored in Excel sheets assists considerably with the design of database (i.e. linking and categorizing the different fields)

Initial questions related to survey on ECHOES data structure are raised to discussion:

- Dataset description (e.g. "A survey of consumer attitude")
- Data fields ("quantitative responses to surveys")
- Country/geographic area (e.g. "Turkey")
- WP/task (for internal purposes) (e.g. "Task 4.2)
- Author (?) (e.g. "VTT")
- Level of "Energy collective" (e.g. "Individual decisions making")
- Foci area ("Buildings")

Fifth, user rights as a subject to be raised in discussion within the ECHOES consortium was identified. In particular the question of who will be granted with permissions for uploading/downloading of the data to the database, writing or reading access, needs to be resolved.

Initial recommendations

- As the advantages of iterative feedback are recognized, during the project, it seems relevant to keep the database open to project participants.
- After the project closure, the database is primarily open to everyone. However, regarding external users' uploads, there is no provision of guarantees of its feasibility. Therefore, only accessing the reading rights globally after the project's closure is proposed as a preliminary suggestion.

As a **sixth** point, the need to emphasize the ethical questions in the consortium (i.e. how can the privacy of the answerers guaranteed?) and IPR rights (i.e. are there any legal barriers of storing suggested data?), and carefully tackle them prior to a more concrete level of implementation of the database. Procedures for this are also defined in the data management plan (D8.1).

6.4 Suggestions for next steps for ECHOES database development

Based on the above discussions, and in order to summarize, the following preliminary "Action plan" was drafted to concretize the next steps in ECHOES database development (Table 12). The Action plan is subject to discussions and VTT is open to further development ideas.

Table 12.A preliminary suggestion for Action plan to develop the ECHOES database.

Step	Responsible	Timeline	Notes
1 Introduce the results of D2.1, raise up the open questions, introduce the action plan	VTT	ECHOES meeting In San Sebastian	
2 Developing an internal questionnaire for specifying the ECHOES database architecture.	VTT (incl. database expert)	9/2017	Input from the San Sebastian meeting taken into account. The relevancy of questions is to be verified with the VTT database expert
3 Answering the questionnaire	ECHOES partners.	10/2017	It is viewed as beneficial if one "data contact person"/organisation could be named instead of general e-mails.
4 Developing an architecture of the database.	VTT (incl. database expert)	? (to be confirmed)	Taking the questionnaire and characteristics of the Zenodo into account (if decided)
5 Publishing the database internally for further development	VTT (incl. database expert)	? (to be confirmed)	
6 Updating and storing data in the database once ready	VTT or ECHOES partners	? (to be confirmed)	
7 Publishing the database externally	VTT	11/2019	

7 REFERENCES

Bhattacherjee, A. 2012. Sccial Science Research: Pronciples, Methods, and Practises. Textbooks Collection. Book 3. http://scholarcommons.usf.edu/oa_textbooks/3Hertwich, E. G., Peters, G. P. 2009. Carbon footprint of nations: A global, trade-linked analysis. Environmental science & technology, 43(16), 6414-6420

EC (2011a) European Commission. Communication form the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. A Roadmap for moving to a competitive low carbon economy in 2050, Brussels, 8.3.2011, COM (2011) 112 final

EC (2011b) European Commission. Communication form the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Energy Roadmap 2050, Brussels, 15.12.2011 COM (2011) 885 final

EC2014. European Commission. A policy framework for climate and energy in the period from 2020 up to 2030. Impact assessment. Bussels: European Commission, SWD(2014) 15 final.

IEA 2008. Energy Technology Perspectives 2008. Scenarios & Strategies to 2050. In support of the G8 Plan of Action. International Energy Agency, OECD/IEA, Paris.

IPCC 2014. Edenhofer et al. Technical Summary. In: Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assess-ment Report of the Intergovernmental Panel of Climate Change. Cam-bridge, United Kindom and New York: Cambridge University Press.

Koljonen, T., Leinonen, A., Wessberg, N., Sokka, L., Eerola, A., Pursiheimo, EA. 2012. Analysing Transition Planning and Systemic Energy Planning Tools for the implementation of the Energy Technology Information System – ATEsT. Suggestions for tools and methodologies development to support SetPlan implementation. EU 7th Framework Programme. Project n:o 241382.

Koljonen, T., Hernandez, P.I., Kubeczko, K., Laes, E., Lechon, Y., Tomasgard, A., Velte, D., Virdis, M.R. 2016. How to tackle societal change in the transition to low-carbon societies? Proceedings of the 7th International Sustainability Transitions (IST) Conference 2016, September 6-9, 2016, Wuppertal.

Kriegler, E., O'Neill, B.C, Hallegaette, S., Kram, T., Lempert, R.J., Moss, R.H. & Wilbanks, T. 2012. The need for and use of socio-economic scenarios for climate change analysis: A new approach based on shared socio-economic pathways. Global Environmental Change 22, 807-822.Philibert, C. and J. Pershing (2002), Beyond Kyoto – Energy Dynamics and Climate Stabilisation, OECD/IEA, Paris.

Philibert, C. and J. Pershing (2002), Beyond Kyoto – Energy Dynamics and Climate Stabilisation, OECD/IEA, Paris.

Schweizer, V.J., O'Neill, B.C. 2014. Systematic construction of global socioecnomic pathways using internally consistent element combinations. Climatic Change 122, pp. 431-445.

Sovacool, Benjamin K. 2014. What are we doing here? Analyzing fifteen years of energy scholarship and proposing a social science research agenda. Energy Research & Social Science 1, 1-29.

8 APPENDIX

8.1 General data sources

Data source	Type	Qualitative Quantitative	Managed by/ Author	National/ International	Published
Socio-Economic Panel (SOEP) https://www.diw.de/soep	Database	Quantitative	German Institute for Economic Research, DIW	National (Germany)	Constantly updating
Allgemeine Bevölkerungsumfrage ALLBUS (The German General Social Survey) www.gesis.org/allbus	Database	Quantitative	GESIS - Leibniz-Institute for the Social Sciences	National (Germany)	Constantly updating
TurkStat (Turkish Statistical Institute) http://www.turkstat.gov.tr/Start.do;jsessionid=lyT8YzTLRQrWwkNwGQ2r1JwW7TpH1Y3bnppZnGrTQv2JLj26C1Tt!1193456523	Database	Quantitative	National statistical office	National (Turkey)	Constantly updating
Ministry of Energy and Natural Resources: electricity http://www.enerji.gov.tr/en-US/Pages/Electricity	Website	Including various information	Ministry	National (Turkey)	Constantly updating
Ministry of Energy and Natural Resources Strategic Plan http://www.enerji.gov.tr/en-US/Strategic-Plan	Document: Policy plan	Report with qualitative and quantitative information	Ministry	National (Turkey)	2014-2015
Statistics Norway https://www.ssb.no/en/	Database	Quantitative	National statistical office	National (Norway)	Constantly updating
Enova	Website	Including various information	Publicly owned energy solutions promoting office	National (Norway)	Constantly updating

https://www.enova.no/					
SIFO (National Institute for Consumer Research) http://www.sifo.no/page/Research//10435/10441	Website	Including various information	Research organisation	National (Norway)	Constantly updating
Statistik Austria (Statistics Austria) www.statistik.at	Database	Quantitative	National statistical office	National (Austria)	Constantly updating
Energy styles http://tinyurl.com/h6t59jz	Document: research report	Research report including quantitative and qualitative information	Research project	National (Austria)	2011
Statistisches Bundesamt (National statistical office of Germany) https://www.destatis.de	Database	Quantitative	National statistical office	National (Germany)	Constantly updating
SESAME https://www.sesame-project.eu/	Website	Including various information	Research project (Acceptance of electricity infrastructures)	EU-27	Project ended in 2014
E-control https://www.e-control.at/	Website	Including various information	Energy regulation agency Austria	National (Austria)	Constantly updating
"Klimabarometer" https://www.tns-gallup.no/tns-innsikt/klimabarometeret-2016-rapport-pdf?pid=TNS-Report-ReportFile	Document: report	Research report including quantitative and qualitative information	Research project (Acceptance etc.)	National (Norway)	2016

ISOC - Ciencias Sociales y Humanidades Research database http://bddoc.csic.es:8080/informe.html;jsessionid=19E2DDD4FEDAA408405F1BF663FCBEDF?total=23&bd=ISOC&tabla=docu General/	Database	Reference database on studies on social sciences and humanities	Center of Humanities and social sciences	National (Spain)	Constantly updating
El Instituto Nacional de Estadística (INE) http://www.ine.es/welcome.shtml	Database	Quantitative	National statistical office	National (Spain)	Constantly updating
REPUBLIC OF BULGARIA National statistical institute http://www.nsi.bg/en	Database	Quantitative	National statistical office	National (Bulgaria)	Constantly updating
Ministry of agriculture, food and forestry, Republic of Bulgaria Ministry http://www.mzh.government.bg/	Website	Including various information	Ministry	National (Bulgaria)	Constantly updating
Energy and Water Regulatory Commission http://www.dker.bg/docsbg.php?d=5	Website	Including various information	Energy regulation agency	National (Bulgaria)	Constantly updating
Ministry of Labour and Social Policy https://www.az.government.bg/stats/1/	Website	Including various information	Ministry	National (Bulgaria)	Constantly updating

Executive Environment Agency (ExEA http://eea.government.bg/en	Website	Including various information	Environmental agency	National (Bulgaria)	Constantly updating
The Center of Women's Studies and Policies (CWSP) http://www.cwsp.bg/htmls/page.php?category=445&id=1190 http://www.cwsp.bg/upload/docs/67.pdf	Website	Including various information	Research institute	National (Bulgaria)	Constantly updating
Open data portal of the Republic of Bulgaria http://opendata.government.bg/	Website	Including various information and databases	Governmental	National (Bulgaria)	Constantly updating
Open Society Institute Sofia http://osi.bg/downloads/File/2016/energy5.pdf	Document: report (on energy poverty)	Research report including quantitative and qualitative information	NGO	National (Bulgaria)	2016
Institute for Market Economics http://ime.bg/var/images/SocialAid_Final.pdf	Document: report	Research report including quantitative and qualitative information	Research org/think thank	National (Bulgaria)	2011
Ministry of Labor and Social Policy / Employment Agency https://www.az.government.bg/stats/1/	Website	Including various information and databases	Ministry	National (Bulgaria)	Constantly updating
Strategy.bg http://strategy.bg/StrategicDocuments/List	Website	Including various strategic documents, consultations.	Governmental	National (Bulgaria)	Constantly updating
Official Gazzette of the Republic of Bulgaria	Website	Including various information and databases	Governmental	National (Bulgaria)	Constantly updating

http://dv.parliament.bg/DVWeb/index.faces					
Legal encyclopedia http://www.juen.bg/	Database	Including various legal information	Governmental	National (Bulgaria)	Constantly updating
Council of Ministers' Legal Information System http://pris.government.bg/prin/login.aspx?ReturnUrl=%2fprin%2fdefault.aspx	Database	Decrees, orders and decisions	Governmental	National (Bulgaria)	Constantly updating
The National Center of Public Health and Analyses http://ncphp.government.bg/index	Website	Including various information	Research/Governmental	National (Bulgaria)	Constantly updating
Ministry of Health http://www.mh.government.bg/bg/ministerstvo/vtorostepenni-razporediteli/regionalni-zdravni-inspektsii/	Website	Including various information	Ministry	National (Bulgaria)	Constantly updating