Data Management Plan

Prepared by Pietro Zambelli, Simon Pezzutto, Giulia Garegnani, Alessandro Pignatelli, Liise Lehtsalu (EURAC), Lukas Kranzl, and Sara Fritz (TUW)

Reviewed by Marco Ciurcina (StudioLegale.it), Vincent Roch and Gabriel Ruiz (CREM)

31.01.2017 (First draft)
15.02.2017 (Version sent for peer-review)
31.03.2017 (Final version)
28.09.2018 (Fist update)

Funded by the Horizon 2020 programme of the European Union

PROJECT INFORMATION

- **Project name** Hotmaps
- **Grant agreement number** 723677

- **Project duration** 2016-2020
- **Project coordinator**

Lukas Kranzl
TU Wien
Energy Economics Group – EEG
Gusshausstrasse 25-29/370-3
A-1040 Wien / Vienna, Austria
Phone: +43 1 58801 370351
E-Mail: kranzl@eeg.tuwien.ac.at

Legal notice

The Hotmaps project

The EU-funded project Hotmaps aims at designing a toolbox to support public authorities, energy agencies and urban planners in strategic heating and cooling planning on local, regional and national levels, and in line with EU policies.

In addition to guidelines and handbooks on how to carry out strategic heating and cooling (H&C) planning, Hotmaps will provide the first H&C planning software that is

- **User-driven**: developed in close collaboration with 7 European pilot areas
- **Open source**: the developed tool and all related modules will run without requiring any other commercial tool or software. Granting free in use and access to the source code.
- **EU-28 compatible**: the tool will be applicable for cities in all 28 EU Member States

The consortium behind

Scientific partners

TU WIEN TECHNISCHE UNIVERSITÄT WIEN Vienna | Austria

AALBORG UNIVERSITET

e3m

energycities WHERE ACTION & VISION MEET

Hes·so Haute Ecole Spécialisée de Suisse occidentale Fachhochschule Westschweiz University of Applied Sciences and Arts Western Switzerland

EURAC research

Fraunhofer ISI

ethink ENERGY RESEARCH

PlanEnergi

energy center

Pilot areas for developing and testing the tool

Aalborg Kommune

VILLE DE GENÈVE

Bistrita Municipality

STADT FRANKFURT AM MAIN Energiereferat > Die kommunale Klimaschutzagentur

MILTON KEYNES COUNCIL

Kerry County Council

donostiasustapena fomentosansebastián DONOSTIAKO GARAPEN EKONOMIKOA

Table of Contents

According to the ”Guidelines on FAIR Data Management in Horizon 2020 (Version 3.0, 26 July 2016)” the Data Management Plan should address following issues - next to the listed points, please find the text section of the given Data Management Plan in which those are treated:

DATA SUMMARY	
State the purpose of the data collection/generation	1. Objectives, Page 10
Explain the relation to the objectives of the project	4 Best practice and guidelines to data management in relation to Hotmaps, Page 13
Specify the types and formats of data generated/collected	4.4 Data formats, Page 22
Specify if existing data is being re-used (if any)	3.1 External data source, Page 11
Specify the origin of the data	3.1 External data sources, Page 11
State the expected size of the data (if known)	4.5 Data preservation, Page 23
Outline the data utility: to whom will it be useful	2. Types of managed data in Hotmaps, Page 10
MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA [FAIR DATA]	
Outline the discoverability of data (metadata provision)	4.1 Data vocabularies and metadata, Page 17
Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?	4.11 Data identification, Page 30
Outline naming conventions used	4.2 Name conventions, Page 17
Outline the approach towards search keyword	4.1 Data vocabularies and metadata, Page 17
Outline the approach for clear versioning	4.10 Data versioning, Page 28

Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how	4.1 Data vocabularies and metadata, Page 17
MAKING DATA OPENLY ACCESSIBLE [FAIR DATA]	
Specify which data will be made openly available? If some data is kept closed provide rationale for doing so	4.8 Data license, Page 27
Specify how the data will be made available	4.4 Data formats, Page 22
Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)	4.4 Data formats, Page 22
Specify where the data and associated metadata, documentation and code are deposited	4.4 Data formats, Page 22
Specify how access will be provided in case there are any restrictions	4.4 Data formats, Page 22
MAKING DATA INTEROPERABLE [FAIR DATA]	
Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability	4.1 Data vocabularies and metadata, Page 17
Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?	4.1 Data vocabularies and metadata, Page 17
INCREASE DATA RE-USE (THROUGH CLARIFYING LICENSES) [FAIR DATA]	
Specify how the data will be licenced to permit the widest reuse possible	4.8 Data license, Page 27
Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed	4.8 Data license, Page 27
Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why	4.8 Data license, Page 27
Describe data quality assurance processes	3.2 Data quality process, Page 12

Specify the length of time for which the data will remain re-usable	4.5 Data preservation, Page 23
ALLOCATION OF RESOURCES	
Estimate the costs for making your data FAIR. Describe how you intend to cover these costs	5.1 Cost estimate to make the open-data FAIR, Page 31
Clearly identify responsibilities for data management in your project	5.2 Data management responsibilities, Page 33
Describe costs and potential value of long term preservation	5.1 Costs estimate to make the open-data FAIR, Page 32
DATA SECURITY	
Address data recovery as well as secure storage and transfer of sensitive data	4.12 Data access, Page 30 and 4.5 Data preservation Page 23
ETHICAL ASPECTS	
To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former	4.3 Sensitive data, Page 19
OTHER	
Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)	3.2 Data quality process, Page 12

Executive Summary

Hotmaps is part of the H2020 Pilot on Open Research Data - a plan established to maximise access and reuse of projects research data. Data Management Plans (DMPs) describe how to make research data findable, accessible, interoperable and reusable (FAIR). Towards this goal, we present the following DMP, which describes the life cycle management of data sets processed or generated during/beyond Hotmaps. The DMP also documents how data types, methods and standards are exploited for verification and re-use and how they are curated and preserved.

Hotmaps' DMP implements the "Guidelines on FAIR Data Management in Horizon 2020 (Version 3.0, 26 July 2016)" and follows the new W3C draft on "Data on the Web Best Practices (31 January 2017)" and refers to their content as baseline for its research and innovation actions.

Hotmaps discusses specifically the twelve challenges for web data publishing brought up by the W3C guideline (key player in semantic web technologies), in particular referring to their harmonization within the project. Providing information on our data and the manual data collection method is a focus of the Hotmaps community, even after project's completion.

Hotmaps collects data establishing their relation to other Research & Innovation Action (RIA) and Cooperation Support Action (CSA) activities – THERMOS, Planheat, Heat Roadmap Europe 4 (HRE 4) – reviewing all together 26 projects.

Hotmaps will open the data sets used as inputs for the Hotmaps toolbox at EU28 level and open the results of the model and scenarios for the entire EU.

All the data coming from the Pilot areas (WP6) or uploaded by the user of the Hotmaps toolbox platform will be treated as sensitive and confidential data.

Introduction - Hotmaps objectives

The overarching goal of Hotmaps is the development of an open source heating / cooling mapping and planning toolbox and to provide default data for EU28 at national and local level. These data and tool allow public authorities to identify, analyse, model and map resources and solutions to supply energy needs within their territory of responsibility in a resource and cost efficient way. Those results will help authorities to develop heating and cooling strategies on local, regional and national scale which are in line with RES and CO_2-Emission targets on national and EU level.

This leads to three main project objectives:

1. Develop an open source toolbox (Hotmaps toolbox) that will effectively and comprehensively support local, regional and national heating and cooling planning processes.
2. Provide a default open data set to lower the initial barrier in applying the tool for regions across EU-28 member states and include the ability that the users can adapt and provide more accurate, large and complex data for data for a specific area.
3. Provide a tested and user friendly open source software tool which is based on user needs. Guarantee wide usability, flexible adjustability and concrete application of the tool within and beyond the project duration.

1. Objectives

The DMP defines and describes data management policies applied by the Consortium to the datasets generated by the project. It represents a novel and important tool of Horizon 2020 projects to describe data generation, exploitation, accessibility, re-use and preservation. Hotmaps, a Research & Innovation Action (RIA) project, investigates Energy Efficiency (EE) and energy demand for heating, cooling and domestic hot water (DHW) in EU households, services, industry and transportation. Hotmaps collects and generates new data at a spatially disaggregated level (with vector data at NUTS0, NUTS1, NUTS2, NUTS3 level or with raster data with a resolution of 250 x 250 m or higher) to allow especially public authorities to identify, analyse, model and map resources and solutions covering responsibly and cost-efficiently their local supply needs. These results will help authorities to develop heating and cooling (H&C) strategies on local, regional and national scales in line with RES and CO2-Emission targets at national and EU levels.

Data of the default data set, should be publicly available, comparable, correct, up-to date, complete, compelling and, most importantly, findable, accessible, interoperable and reusable (FAIR). The data will be ideally preserved by Hotmaps' active community of users and developers. Hotmaps extracts and harmonizes data from different sources like, for example, project websites, deliverables, publications (journal papers, conference proceedings etc.), tools and feedbacks from partners. This strategy will foster reuse of results, increase collaboration and help identifying common data requirements.

This DMP shows the approach that Hotmaps has been chosen to deal with expected project results. It first clarifies the types of data managed and the methods used to collect and harmonize them and then explains how Hotmaps deals with the challenges of data management and publication (Guidelines on FAIR Data Management in H2020 [1] and W3C Proposed Recommendation "Data on the Web Best Practices" [2]). It should be noted that the data types Hotmaps will collect and generate do not necessarily endorse all the recommendations laid out by the European Commission (EC) and W3C.

2. Types of managed data in Hotmaps

Hotmaps, as a RIA, will collect data in the form of vector (GIS vector format) or raster. It will also generate new data applying statistical models to fill in missing information (WP2), energy models (WP3) to estimate the main energy parameters and generate data for scenarios (WP5). Collected and processed data, available in vector form, raster maps and downloadable databases, will especially benefit from:

- Public authorities: administrative staff in the public authorities at local, regional and national levels, responsible for energy and/or spatial planning, strategy and policy development in the heating and cooling sector;

- Planners and consultants: although public authorities represent the main target group of the toolbox, its main users in the case of small cities/municipalities will be professional planners and consultants;

- Energy agencies: staff of local, regional and national energy agencies assisting public authorities in the planning and strategy development processes for H&C;

- Open source modelling community: staff of scientific and academic institutions as well as professionals from the private sector focusing on the development of methods and tools for the analysis of energy systems, energy management and spatial representation tools, dedicated to the open source concept and to the extensive sharing of ideas and developments.

- The updated data will be made available to the Technical Board, composed by the project partners (TUW, EURAC, CREM, HES-SO, Fraunhofer, AAU, PlanEnergi and e-think).

Hotmaps aims to build - around the data collected and the tool developed during the project - an open community of users and developers interested in the project's topic. The creation of an open community of people working at different aspects of Hotmaps energy systems (data, tools, models, scenarios, etc.) will be paramount to update and improve the main outputs beyond the end of the project.

3. Data collection framework

3.1 External data sources

As shown in the previous section, a main outcome of Hotmaps in terms of managed and published data is H&C demand data at spatially disaggregated level (vector data at NUTS0, NUTS1, NUTS2, NUTS3 level in CSV format with NUTS code and year or raster data with a resolution of 250 x 250 m or higher in compressed format). Data will concern four sectors of the EU market: residential, service, industry and transportation. Hotmaps will review, categorize and compare different RIA projects to encourage knowledge sharing and to increase the impact of project results. In particular, due to a number of shared points, we describe how H2020 HRE4 shares methods, approaches, data and results with Hotmaps. We compared data requirements and expected results of both projects and identified their possible synergies. We also described how the two projects differ and where/in what measure data exchange is limited (see working document "Deliverable 2.0 Synergies between Hotmaps and HRE4").
Data will also be collected from national statistics to increase data coverage and availability and create a list of national contact points (NCPs), which can offer support to this and other projects during data collection.

Concerning the spatial data several sources like EU agencies (EUROSTAT, European Environment Agency - EEA) and data of previous EU-projects (e.g. EEA [3]) will be used to provide/integrate information to the Hotmaps toolbox concerning: land-cover, population, urban-atlas, urban morphological zones, solar potentials, shallow geothermal potentials, etc.

The list of input datasets to be used in the project will be included in deliverable D2.2 "Open datasets" due on month 18 (March 2018).

3.2 Data quality process

Data quality, completeness, accuracy, and reliability are very important aspects in the process of establishing the datasets within the project. Among others, the following main points will be taken into consideration: i. Data inventory, ii. Data processing, iii. Data definition and comparability:

i. Data inventory
The output datasets to be generated within the project will be published in the GitLab public repository within the deliverable D5.2 “Scenarios for EU-28” and D2.2 “Open dataset ”due on month 18 (March 2018).

One of the major challenges in developing an inventory of energy demand and resource data for heating, cooling, and DHW supply in different sectors, is to provide a complete list of all existing information. In general, the advantage of using data that comes from EU wide information providers and EU projects is that this data is available for a large territory (e.g. ESPON [4]).

Filling in the data gaps implies not only extrapolating and assembling data from large data tools available online (e.g. Building observatory [5]), but also researching data source-by-source from single scientific literature sources such as journal papers, conference proceedings and project deliverables. Only through such an in-depth approach can the already mentioned lacks of data per energy type and nation be filled.
One important aspect of the data inventory is to ensure that the information can be understood and interpreted correctly by users. This requires a compilation of clear metadata description, annotation, contextual information and documentation [1].

The data documentation will provide standardized structured information explaining the purpose, origin, time references, geographic location, creator, access conditions and terms of use of the data collection, following, as much as possible, the W3C Proposed Recommendations on the Data on the Web Best Practises.

The whole open dataset collected and produced within the Hotmaps project will comply with the INSPIRE Directive 2007/2/EC [6] protocols and standards. CREM in WP4 is responsible to develop and deploy the Hotmaps unique interface to access all data available in the public GitLab repositories, the open data sets will be published using the protocols defined by the Open Geospatial Consortium that are compliant to the INSPIRE directive.

ii. Data processing
Data processing are performed mainly in WP2 (Open Data Set for EU-28), WP3 (Development of energy system planning modules) and WP5 (Scenarios and regional scale consistency).

In WP2, much effort is going to be allocated to analyse sources, assess the reliability of the gathered data and complement the missing gaps by in-depth local surveys/inquiries. The Consortium will correctly distinguish between different kinds of information by analysing the methodology related to the data found. In case of missing data documentation, these will be excluded from the database. All collected information on heating, cooling and DHW (with a

unit of kWh/m² y, Mm², etc.) will be filtered and evaluated statistically. As far as the number of sources allows, the coefficient of variance will be used as statistical indicator of uncertainty to exclude values outside a range of plus or minus one standard deviation around the average. The filtered values will be used to compute a more robust average. Further types of information will be collected to validate the indications obtained per country and sector to assess the reliability of the data. (E.g. population size per county, percentage of heated, cooled and total floor area, gross domestic product - GDP, economic growth indicators and energy related funding/financing schemes).

In WP3, a set of tools and modules will be developed to process and analyse the default data set produced in WP2 in order to support the planning process of the energy systems.

WP5 will prepare and compute the scenarios (e.g. on the development of buildings renovation activities and related heat demand or the expansion of district heating) using the modules and tools developed in WP3.

iii. Data definition and comparability

Despite most data providers utilizing standardized data formats and units, this does not necessarily mean that data are fully comparable. Adjusting differences and inconsistencies among different measures, definitions, assumptions, methods, time references and specifications to improve data comparability is one of the most important aspects of the whole process of data elaboration.

Data will be collected for each country with reference to the most recent year and to a previous baseline year. Apart from the use within the Hotmaps toolbox and other existing tools the developed database including the documentation in this project is expected to improve data quality for user, add value to already existing data and provide data needed to monitor the progress towards the achievement of the goals defined in EU energy related Directives.

4. Best practices and guidelines to Data Management in relation to Hotmaps

The DMP describes what data will the project generate, whether and how will it be exploited or made accessible for verification and re-use, and how will it be curated and preserved. Moreover, a crucial point of the DMP is to describe how research data are made findable, accessible, interoperable and reusable (FAIR). The Consortium is expected to [1]:

- Deposit research open-data in a repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate;
- Use the open-data, including associated metadata, to validate the results presented in scientific publications;

- Treat other data, including associated metadata, as specified and within the deadlines laid down in the DMP.

Hotmaps manages the generated data using the following web platforms:

1. Hotmaps website (public) http://www.hotmaps-project.eu/: the website provides information about the project's objectives, partners and results, as well as about upcoming events related to project topics.

2. LiveLink - developed and maintained by OpenText: the platform is meant for private and project internal use only to easily share documents and contacts between the partners during the project (e.g. deliverable drafts, project management documents, presentation slides, project related literatures, etc.).

3. The Hotmaps toolbox - hosted by Energy Cities after the project's end and by HES-SO during the project - will publish tools and models developed by Hotmaps enabling users to execute and test different models for their region of interest, to refine their own data analysis and to export the results in a common format to support the strategic development of their area. Any other user data compatible with the Hotmaps toolbox, and complementing the default dataset, will remain private and treated as confidential.

4. All the datasets of the Hotmaps projects will be shared under the following GitLab web-page (https://gitlab.com/Hotmaps). Each dataset will have a dedicated Git repository. The repository will be freely accessible without authentication for all the datasets that the project will released as open-data, and will required user authentication for all the datasets that are classified as not publishable due to data sensitiveness and confidentiality. In case the terms and conditions underlying the use of a GitLab repository are not satisfying to the user's requirements, the HotMaps toolbox should foresee other possibilities to work with data stored locally. The open-data datasets will be available for all regions of EU-28 without any specific condition and restriction. The use of a Version Control System such as git to handle the public project datasets can help the Hotmaps project follows the Guidelines on FAIR Data Management in Horizon 2020 (Version 3.0, 26 July 2016) [1] - as recommended by European Commission - and the Proposed Recommendations "Data on the Web best practices" by the W3C communities (31 January 2017) [2], promoted by Hotmaps' partners through its dissemination activities in WP7.

The Horizon 2020 guidelines address the following topics:

- Data set reference and name: to enable identification, search, and retrieval of data. Each published data set is named and accessible through a unique and identifiable Uniform Resource Locator (URL) provided by git and published and reachable on the GitLab web-platform;

- Data set description: addressed by Hotmaps in section 4.4 to define data format and information included;

- Standards and metadata: Hotmaps defines a dedicated vocabulary and set of metadata schemes to describe energy data collected at EU28 level. Each published

dataset present in the git repository and published on the GitLab platform will use a unified communication format, for instance the JSON for Linking Data (json-ld) [7] format to provide metadata based on a common format easily readable by both humans and machines. EURAC in coordination with the other technical partners will prepare a document with the vocabulary and the main features that are needed to describe and classify consistently the energy data sets. The vocabulary will use and extend the wiki-page defined by the Open Energy Model Initiative (https://wiki.openmod-initiative.org/wiki/Category: Glossary). The list of required features and classification of the energy data sets will be shared and discussed in a public dedicated Hotmaps repository managed by EURAC. CREM and HESO will support the project in converting the previous documents into a specific schema. The schema will be described in section 4.1 and will be defined in Task 4.4 and delivered at the end of the project. The definition of a shared specific schema and vocabulary to describe energy data sets will provide the consumer with a better understanding of the collected and enriched data;

- Data sharing: this term refers to the way data are shared - including access procedures, license, and management of sensitive data - and will be further developed in section 4.3, 4.4, 4.7, 4.8, 4.9 and 4.12;

- Archiving and preservation deals with the procedures for long-term preservation of the data. It includes for how long data should be preserved, their approximate end volume, associated costs and recovery plan. The implementation in Hotmaps is described in section 4.5.

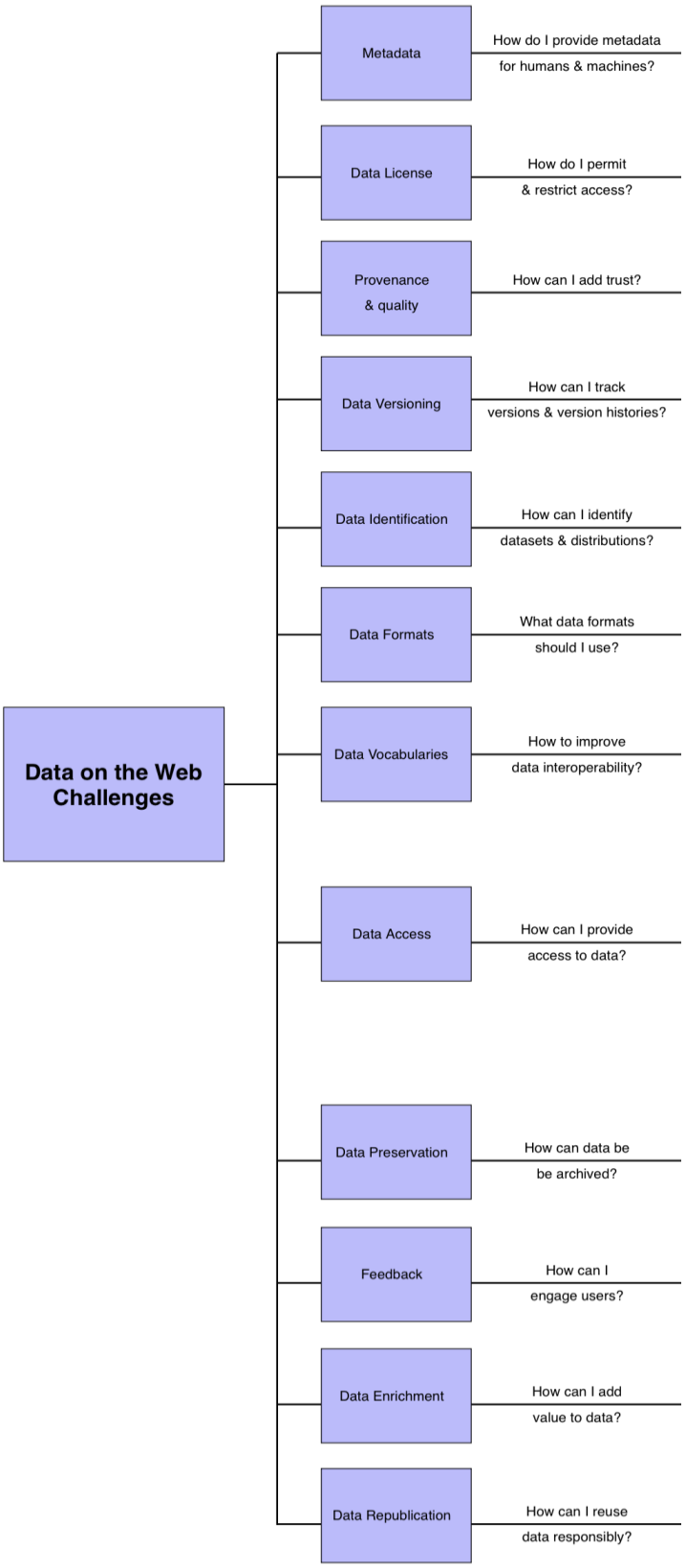

Figure 1 Best practices addressing the challenges faced when publishing data on the web [2]

4.1 Data vocabularies and metadata

This challenge is relevant to achieve semantic interoperability between data producers and consumers. The solution proposed in the semantic web community is to agree on a shared vocabulary and to make it available in an open format. Hotmaps will use Join UP Glossary [8] and enrich its existing vocabulary adding, when needed, new definitions to the wiki page of the OpEnMod Initiative [9] community. The keywords used by the project to describe the dataset will be chosen from the wiki page of the OpEnMod Glossary.

Sharing common data is important both among Hotmaps Consortium partners and among external data consumers, also to avoid ambiguity and clashes. This remains however a challenge, due to the wide range and diversity of covered topics, and requires finding a good compromise at the level of common comprehension.

Hotmaps will define its own set of schemas to semantically describe the data collected and produced within the project whenever there is a lack of vocabulary definitions that satisfy the project requirements. The term Energy is missing in both the Data Catalog Vocabulary (DCAT) [7] and the Semantic Interoperability Community (SEMIC) [10] vocabulary. schema.org provides a basic definition for Energy [11], without however providing any distinction on the form of energy (i.e. electric, thermal, kinetic, chemical etc.), the balance role (i.e. produced, consumed, imported, exported), the resources used (i.e. fossil, nuclear, renewable, etc.), the technology used to produce it (i.e. photovoltaic, gas turbine, hydro, etc.), the data type (i.e. measured, estimated, forecasted, etc.), the time range (i.e. yearly, monthly, daily, etc.) and the spatial resolution (i.e. single component, an entity as a plant or a building, an area as a municipality, a region or a country). The schema definition will be done together with other EU funded projects (i.e. HRE4, ExcEED, etc.) and open-source communities interested in the topic (i.e. OpEnMod [11], etc.).

4.2 Name Conventions

The consortium defines two distinct name conventions: one is dedicated to the project documents and one to the data collected and elaborated that are used as default inputs for the Hotmaps toolbox and that will be released as open-data or the inputs used internally assess some specific parameter for the models and scenarios.

Hotmaps documents

The Hotmaps documents are organized in a hierarchical file tree that reflects the main project structure. At the first level there is the name of the WP, inside there is a dedicated directory for each task and a dedicated directory with the minutes of the WP meetings and telephone conferences. Each task's directory can contain another directory with the minutes (if necessary) and the other documents and investigations performed during the task's activities. See the structure reported below as an example of the hierarchical structure of the Hotmaps documents.

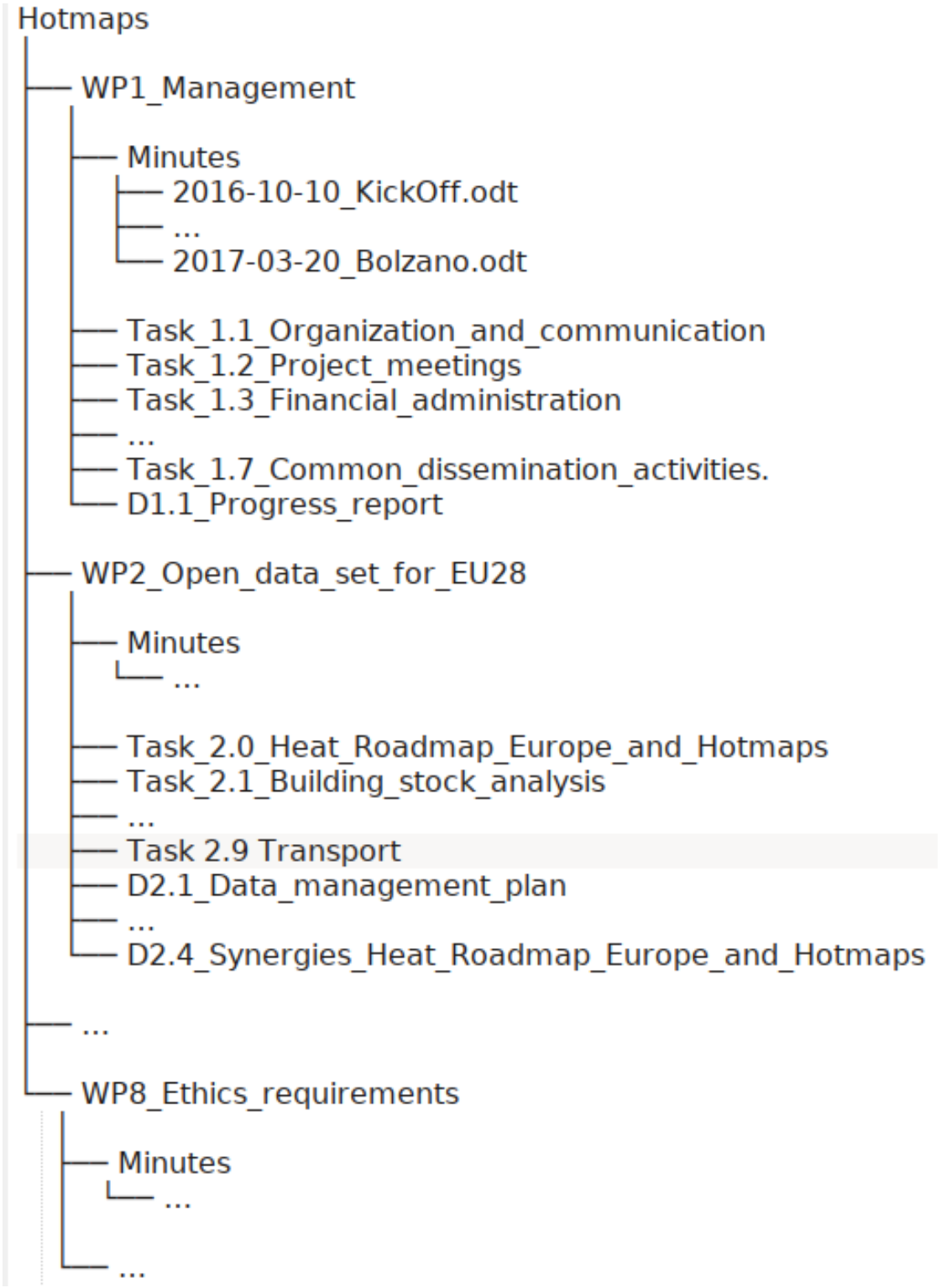

Figure 2 Hotmaps hierarchical file tree

Hotmaps data

Hotmaps data can be private or public and the potential users must not be aware of the project work packages and tasks. The title of all the data repositories will start with: "Data package: ...".

The name of the repository will use keywords defined in a dedicated wiki page of the repository containing the list of all the Hotmaps datasets. The wiki page will be created and managed by EURAC, all partners will contribute defining and refining the names.

The first element will define the spatial resolution of the data set (e.g.NUT{0|1|2|3} or LAU{1|2}). Then we will use the keywords defined in the wiki-page to describe the dataset.

So for example the H&C demand of the building sector at NUTS3 could be: NUTS3_building_HC_residential_demand.

4.3 Sensitive data

Hotmaps is committed to follow best practices in collecting, processing, and sharing personal data[1] and confidential data[2] that may be generated or processed further during the project. In this DMP, we designate such data as sensitive data. The Consortium is aware that mishandling of sensitive data may lead to damages to individuals or organizations and may breach national, EU and international law.

To support best practices for sensitive data, data publishers should identify all sensitive data, assess their risk exposure, determine their intended use, user audience and any related policies; obtain appropriate approval and determine appropriate security measures for data protection (which should also account for secure authentication and use of HTTPS).

Use-cases generated during the Hotmaps originate from publicly available deliverables. Internal project documents and personal data, in the form of contact details (addresses), are only shared through the internal LiveLink and in any case cannot be disclosed without a

[1] As defined in letter (a) of article 2, Directive 95/46/EC, that states 'personal data' shall mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.
This definition will be replaced from the 25th of May 2018 by the slightly different definition adopted in point (1) of article 2, Regulation (EU) 679/2016, that states 'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

[2] We refer to data that is excluded from access on the grounds of protection of national security (i.e. State security), defence, or public security, or statistical confidentiality, it is used under a confidentiality agreement or protected by rights of third parties, like the right on trade secrets as defined in point (1) of article 2, Directive (EU) 2016/943 that states 'trade secret' means information which meets all of the following requirements: (a) it is secret in the sense that it is not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question; (b) it has commercial value because it is secret; (c) it has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret.

permission by the data subject or in breach of the data license, if processing previously collected secondary data.

Data coming from the Pilot areas during the Hotmaps project will be used to test and tune the tools and algorithms that will be included in the Hotmaps toolbox and shared between the project partners using the data package format as described in section 4.5 through a private GitLab repository as described in section 4.12.

Sensitive data collected through interviews and questionnaires will be stored in a dedicated GitLab private repository using the frictionless data formats for (meta-)data. The data will be anonymized before sharing and publish the dataset. The Consortium will review all questionnaires or surveys before ensuring that they comply with the EC standards [13]. This sets down strict policies for managing and anonymizing personal data.

Hotmaps investigates energy efficiency and energy demand for heating, cooling and DHW in EU households, services, industry and transportation. Hotmaps collects and generates new data at a spatially disaggregated level (with vector data at NUTS0, NUTS1, NUTS2, NUTS3 level or with raster data with a resolution of 250 x 250 m or higher) to allow especially public authorities to identify, analyse, model and map resources and solutions covering responsibly and cost-efficiently their local supply needs. Therefore the main ethical issues arising from the project are concerning data protection and privacy as well as informed consent.

In course of the project human participants are involved through interviews, training and workshops. All other data necessary for the determination of the project results will be at higher aggregation level and compiled by external information sources such as statistical data or published results of other research projects.

Particular attention will be devoted to check that datasets to be published do not include personal data (for example, because clusters of data could refer to individual persons) or other sensitive data.

The data and information collected in interviews, surveys and workshops will only be shared anonymised or, where this cannot be the case, the relevant persons/organizations will be asked for permission to publish related information. No sensitive data will be published without explicit permission of the affected persons.

This data collection may include sensitive information, e.g. information about personal values. It will be made clear to the participants that participation is entirely voluntary, and that they will be able to withdraw from the study at any time and can ask researchers to delete their data from the study at any time.

In processing, analysing and publishing the data from the research, we will follow standard procedures such as assigning numbers and/or pseudonyms to participants. We will not use the responses of any one individual, household or organisation in publications in such a way that they could be identified. To avoid the possibility of connecting reported data to specific companies or persons we will just report aggregated data. Participants will be informed that the personal data will be used for research purposes only and kept confidential at all times. Personal data, which is necessary for the interviews, is collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes.

And they are kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data were collected or for which they are further processed. Personal data will be processed pursuant to Directive 95/46/EC and Regulation (EU) 679/2016 (that will apply and replace the Directive from the 25th of May 2018) and applicable national laws.

The ethics and privacy requirements that the project must comply with will be included in deliverable D8.1 POPD - Requirement No. 1 to be delivered in month 6 (March 2017):

- Copies of opinion or confirmation by the competent Institutional Data Protection Officer and/or authorization or notification by the National Data Protection Authority must be submitted (whichever applies according to the Data Protection Directive EC Directive 95/46, currently under revision, and the national law);
- If the position of a Data Protection Officer is established, their opinion/confirmation that all data collection and processing will be carried according to EU and national legislation, should be submitted;
- If the project involves sensitive data, justification must be given for the collection and/or processing of personal such data;
- Detailed information must be provided on the procedures that will be implemented for data collection, storage, protection, retention and destruction and confirmation that they comply with national and EU legislation;
- Detailed information on the informed consent procedures that will be implemented in regard to the collection, storage and protection of personal data must be submitted on request;
- Templates of the informed consent forms and information sheet must be submitted.

With regard to the Pilot areas activities (WP6), planning processes will be carried out applying the Hotmaps database and toolbox, and thus demonstrating its use in course of strategic heat planning. During the WP6 activities each Pilot Areas will try to characterize: What kind of H&C are in use? How are H&C distribution systems in buildings? Which fuels are used? Analysis of stakeholders within H&C sector (utilities, authorities) and decision makers. Organisation and regulation of the sector. Infrastructure for electricity, gas and heat. Description of barriers for district heating and cooling. Map the present H&C demand in buildings and industrial process. Identify and locate resource and fuel potentials (excess heat, geothermal heat, biomass for burning and gasification, biomass for biogas, heat sources for heat pumps, solar, wind). Calculate a reference situation for the city/region, including fuel consumption, environmental consequences and economic consequences. Define and compare district heating and cooling scenarios with individual supply. As last activities of WP6 each Pilot Area elaborates and formulate a H&C strategy and description of how it can be realized stepwise.

All the data and information collected and elaborated for the Hotmaps Pilot Areas and more generally collected by the Hotmaps toolbox web platform will be treated as sensitive not publishable data sets, unless the related partners agree on a deviating procedure, in line with privacy guidelines.

4.4 Data formats

Hotmaps collects data and organizes information in the form of vector data (geometry & attribute table at NUTS 0, 1, 2 and 3 or different geographical unit/geometry) or raster data (covering the whole EU28 with a spatial resolution of 250 x 250 m up to 100 x 100 m).

Any collected, enriched and opened data is published on https://gitlab.com/Hotmaps. Making the data sets openly accessible over the internet, although available for editing only to community members. The main audience of this information is humans and machines (for automatic evaluation); its main aim is to trigger further discussions and information exchange within the Hotmaps community and other research projects/institutions.

A machine readable format is needed for automatic evaluation. Each dataset is a self-contained repository following the standard Data Package format defined and promoted by Frictionless data [14].

The Hotmaps project decided to adopt the Data Package format for its clear and simple principles and because is based on the lessons learned by the CKAN and Open Knowledge Foundation (OKF) during the last years of data collection, harmonization and publications. As stated by Frictionless data the inspiring principles of the Data Package format are:

1. Focused: Focus on one part of the data chain, one specific feature (e.g. packaging), and a few specific types of data (e.g. tabular).

2. Web-oriented: Build for the web using formats that work naturally with HTTP such as JSON, a common data exchange format for web APIs, and CSV, which is easily streamable.

3. Distributed: Design for a distributed ecosystem with no centralized, single point of failure or dependence.

4. Open: Make things that anyone can freely and openly use and reuse with a community that is open to everyone.

5. Built Around Existing Tooling: Integrate with existing tools while also designing for direct use—for example, when a Tabular Data Package integration is unavailable, fall back to CSV.

6. Simple: Keep the formats and metadata simple and lightweight, and make things easy to learn and use by doing the least required.

Each data set will be published as a self contained git repository respecting the technical specifications defined by the Data Package format. The repository consists of a set of text files encoded in Unicode Transformation Format (UTF-8):

- README.md: using MarkDown format, provides basic information such as dataset title, content description, methodology/process for data generation; references to input data sources;

- The datapackage.json file: following the Frictionless Data specifications, provides information in a machine-friendly format. The json file will contain the following fields: name, title, description, license, homepage, author, contributors, sources and schemas. The collected data will be offered as a Data Package format - a standard defined and promoted by Frictionless data using the json-ld format to for the main metadata, based on an ontology derived from the vocabulary and schema discussed in section 4.1;

- A directory data that contains the raw data, preferably in text format, so for instance instead of using shape file that is a binary format and required a dedicated software, use a CSV file with a column with the geometry inserted as Well Know Text (WKT) format, or use a format that can be directly be rendered in a web-browser such as geo-json or topo-json. For the raster map instead due to their size we will use a compressed format such as GeoTiff. Not external software is required to access the raw data except to a text editor, if a binary format is required in the README.md file there will be a section indicating which open-source tools are available to open and work with this special file format;

- If some scripts are needed to load and processing the data these scripts will be included in the repository under a dedicated directory called scripts.

EURAC will be in charge to check the conformity to the Frictionless data specifications and to foster the use of common keywords and definitions, to achieve a data set that is harmonized and consistent.

4.5 Data preservation

This section describes best practices related to open-data preservation:

- The coverage of a dataset should be assessed prior to its preservation - check whether all the resources used are either already preserved somewhere or provided along with the new dataset considered for preservation.

- Data depositors willing to send a data dump for long term preservation must use a well-established serialization - Web data is an abstract data model that can be expressed in different ways (RDF, JSON-LD etc.). Using a well-established serialization of these data increases its chances of re-use without temporarily limit the dataset re-usability.

- Preserved datasets should be linked with their "live" counterparts - a link is maintained between the URL of a resource, its most up-to-date available description, and other preserved descriptions. The description should notify in case the resource does not exist anymore and refer to its last available description.

Energy Cities is responsible for Hotmaps Communication and Dissemination, starting with the third year of the project, Energy Cities will host the Hotmaps database, containing the open-

data set as well as private data sets uploaded by users and pilot areas. CREM, EURAC and TUW will support Energy cities with initiating the data management task. User data from Pilot areas will be hosted locally or in private git repositories (to be created by the user), and accessed by the Hotmaps toolbox and partners.

The following personnel will work on this task:

- Pietro Zambelli (EURAC): holds a PhD in Environmental Engineering. He is expert in developing new GIS models and tools for the spatial and temporal data analysis. Pietro Zambelli designed several spatial database structures to collect energy data coming from monitored systems and to collect data on produced/consumed energy per different sectors and from different resources. He is full skilled in several computer languages: Python, Rust, C, SQL, Fortran, Matlab, R, Bash, Latex, Logo, Javascript and golang mainly for GIS and data processing and analyses.

- Sara Fritz (TU Wien): holds a PhD in energy economics. Her research focus on energy system modelling with special emphasis on spatial highly resolved district heating expansion analyses. She participated in an interdisciplinary doctorate course at TUW, with the aim to research and develop an interactive environment for analyzing scenarios. Thus, she was involved in setting up a data exchange structure for the developed URBEM Smart City Application and the corresponding compliance and security criterias. She is experienced in Python, Matlab, R, QGis and GAMS.

- Olivier Lagarde (Energy Cities): holds a Master in Resources Management and a Master in Digital Communications. He has 15 years of experience in software development, IT coordination and digital communications. He has contributed to develop softwares in the field of energy efficiency and coordinated the IT management and related promotion and dissemination for initiatives and campaigns of the European Commission - DG ENER over the last 10 years. He is skilled in several computer languages (C, Java, Javascript, Postgresql, SQL), data management and programmes assessment.

- Lesly Houndole (CREM): holds a Master in computer science. He has 8 years of experience in software development. He is an expert in data modelling and Database conception. He has been developing a lot of software for companies including an application for the Technical Center of the European Space Agency (ESTEC-ESA) using GPS, GALILEO, GLONASS data for analysing Relative RAIM (Receiver Autonomous Integrity Monitoring) on a multi constellation (GPS + Galileo + SBAS). He is experienced in object oriented programming (OOP), designs pattern, test driven development , GIT, C, C++, Python, JAVA, JAVASCRIPT, Postgresql, PostGIS, SQL.

The database will be backed up at least weekly, depending on data traffic and volume.
The data volume is expected to be small for vector data, especially if we split the geometry and the attribute table the NUTS3 areas at EU28 are less than 2000 items, therefore in most of the case should be possible to store this data as plain uncompressed UTF-8 text documents in CSV format (with a total size that is less than 250 Kb per data set). However for data with high resolution in space (NUTS3) and time (hourly) the data size can became big and therefore a different file format can be considered. On GitLab the free limit of the single file size in is set to 100 MiB and the maximum size of the repository is set to 10 GiB. Raster data

can reach a problematic size, for example the Corine Land Cover (CLC) is around 6.6 GiB for a raster of 100 x 100 m of spatial resolution in the ASCII GRID uncompressed format. To avoid to meet the GitLab size limits the Hotmaps raster data we will use a compressed binary format (GeoTiff generated with the following options: ZLEVEL=9, COMPRESS=DEFLATE, PREDICTOR=1, TILED=YES, TFW=YES), to contains as much as possible the final size of the data set. These binary files will be managed with git Large File Storage [13] version control system tool, that split and track the binary file in smaller binary chunks.

TITLE	RASTER TYPE	BYTES	SIZE	RESOLUTION	FORMAT
CLC	i32	6504320	6.5 GiB	100 x 100 m	ascii
CLC	i32	13316	89.2 MiB	100 x 100 m	ascii.7z LZMA2
CLC	f32	8021200	7.6 GiB	100 x 100 m	ascii
CLC	f32	108636	106.1 MiB	100 x 100 m	ascii.7z LZMA2
CLC	u16	83916954	80.0 MiB	100 x 100 m	GeoTiff DEFLATE1
CLC	f32	97028412	92.5 MiB	100 x 100 m	GeoTiff DEFLATE1
Temperatu re	f32	56672	55.3 MiB	250 x 250 m	GeoTiff DEFLATE1
Random	u16	1264940	1.2 GiB	100 x 100 m	GeoTiff DEFLATE1
Random	i32	1414792	1.3 GiB	250 x 250 m	ascii
Random	i32	146308	83.0 MiB	250 x 250 m	ascii.7z LZMA2
Random	i32	1994848	1.9GiB	100 x 100 m	GeoTiff DEFLATE1
Random	f32	2521628	2.4 GiB	100 x 100 m	GeoTiff DEFLATE1
Random	u16	203008	198.2 MiB	250 x 250 m	GeoTiff DEFLATE1
Random	i32	320024	312.5 MiB	250 x 250 m	GeoTiff DEFLATE1
Random	f32	403852	394.4 MiB	250 x 250 m	GeoTiff DEFLATE1

Table 1 Raster file size and formats for a single temporal step.

Assess the total volume of the raster data sets repository size can be difficult because can change a lot by the raster type (Int32, UInt16, Float32, etc.) and by the raster value see for example raster with the same dimension and resolution can vary from 55 MiB up to 394.4 MiB. Therefore, if we estimate an average repository dimension for raster data of around 500 MiB and assess at the end of the project to open 100 raster dataset we will reach a total volume for repository of the generated dataset of around 60 - 50 GiB.

As it concerns free and open data, there will not be any specific control to monitor data copying on external personal devices (laptops).

Data will be available for import/export from and to existing energy models. The use of a public interface to communicate from and to the developed modules will make the platform easy to integrate and extend it to non-open-source solutions (e.g. EnergyPRO [15], Times [16], etc.).

4.6 Feedback

The Hotmaps website is open for the community to contribute and provide feedback. Feedback is also being elicited through the use of questionnaires and surveys. The latter data are stored on a dedicated private GitLab repository, where they are analysed by the Technical Board. All the published datasets can receive comments and open discuss on specific dataset issues using the GitLab Issue page of each dataset repository.

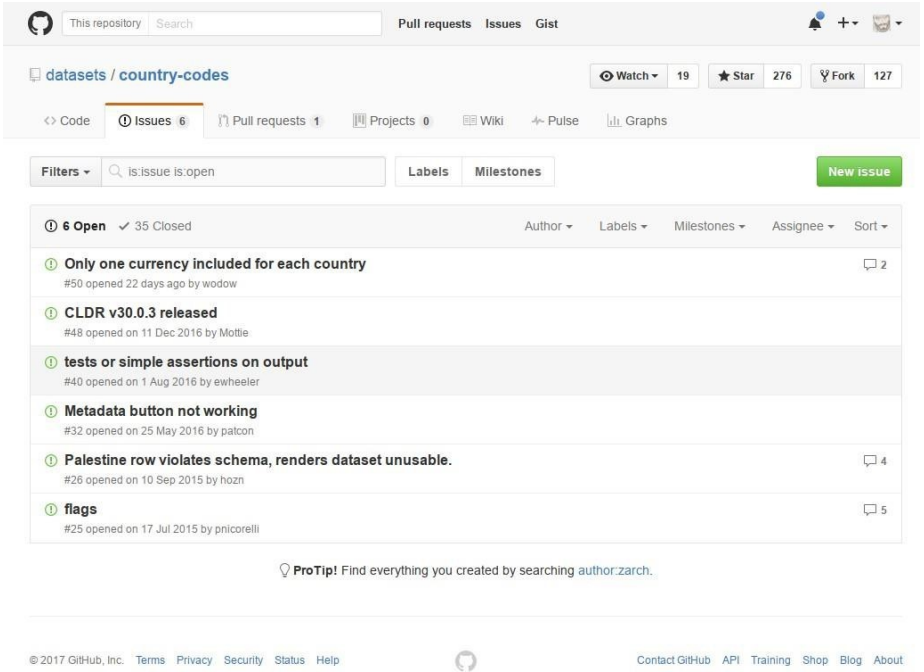

Figure 3 GitHub list of open issues concerning a specific dataset, similar interface is provided by GitLab.

4.7 Data enrichment

Data enrichment is defined as a set of processes that can be used to enhance, refine or otherwise improve raw or previously processed data [2]. Original Hotmaps documents (deliverables, websites, and specifications) provide the necessary input to extract, categorize and publish the required data, harmonizing information and, when necessary, enriching data by getting feedback from project partners. References to original information sources are always provided for verification. Moreover, the use of Git repository on GitLab, open the possibility to people from outside the project to discuss and purpose data refinement/enrichment through the use of the pull-request system.

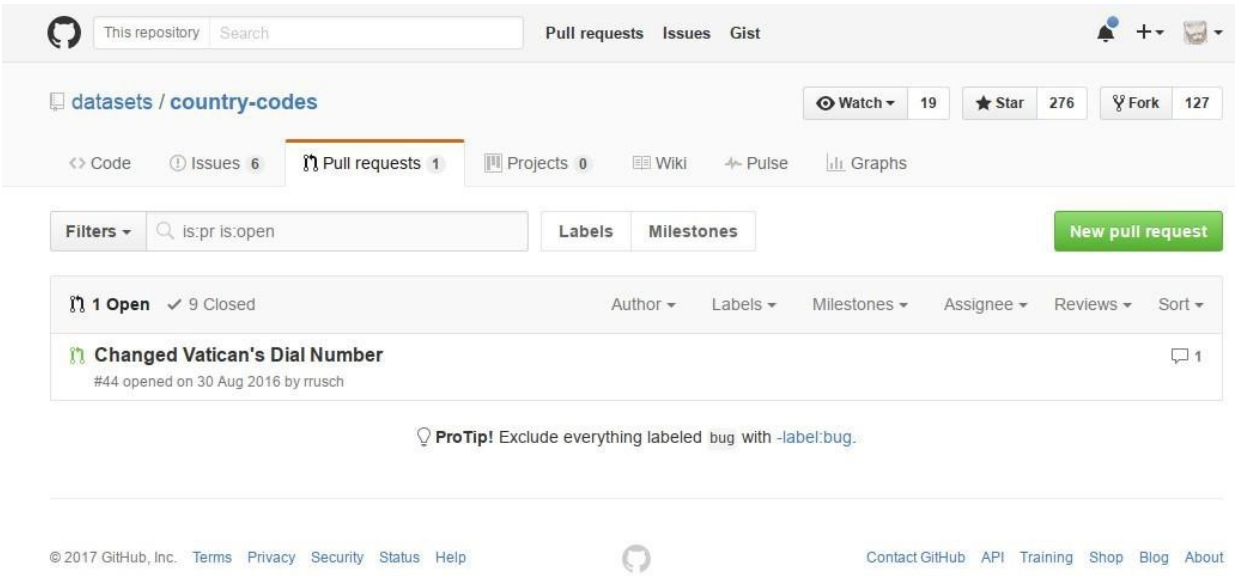

Figure 4 GitHub list of open pull-request to change and improve the dataset, similar interface is provided by GitLab.

4.8 Data license

A license is a legal document giving official permission to use data generated or used in a project. According to the type of license adopted by the publisher, there might be more or fewer restrictions on data sharing and reusing. In the context of Web available data, the license of a dataset can be specified within the data, or in a separate document to which it is linked. The HotMaps project will specify for each dataset repository the specific license of the data, both in the README and in the meta-data json file.

All the data generated within the project will be released under an open-license (e.g. Creative Commons Zero [17] or Open Data Commons Open Database License [18). Hence, they will be potentially freely used, modified and shared by anyone for any purpose as stated by the Open Knowledge Foundation in the Open Definition. The open data will be also shared as linked open data through existing dedicated portals in line with the European Union Open Data Portal [19].

Within the deliverable D2.2 “Open Dataset”, all the input data sets for the Hotmaps toolbox will be published on the GitLab dedicated page and the deliverable will include a detailed table with the license for each input dataset present in the repository. Moreover, the main Hotmaps data outputs, such as the scenarios will be published in the GitLab repository in concomitance with the deliverable D5.2 “Scenarios for EU-28” that will include a table with the licenses for each scenario. The Table in deliverable D2.2 “Open dataset” and D5.2 will also explicitly describe which dataset will be not accessible or accessible only to a restricted number of parties and explain why.

Output data sets generated within the project will be analysed in order to check if they are derivatives of reused datasets and, if this is the case, to comply with the license terms of the reused datasets and to check mutual compatibility of the licenses of reused datasets with the license adopted for the output dataset.

In particular, if reused datasets are licensed under the terms of copyleft licenses, the output dataset could be required to be licensed under the terms of the reused datasets. On the other hand, if the output dataset is to be licensed under a copyleft license, there could be incompatibilities with licenses of the reused datasets. Licenses conflicts (use of datasets licensed under conflicting licenses in violation of their licenses) should be avoided: in order to achieve this a check should be conducted before publication.

4.9 Provenance and quality

Data provenance allows providers to pass information on data origin and history to consumers and it is an important information especially when shared data collaborators do not have direct contact with each other [2]. In Hotmaps, the author’s contact and link to the project homepage are included in the datapackage.json file (as described in section 4.4): this allows data consumers to access the original information sources from project home pages and to contact the use case author if necessary.

Data quality affects the suitability of data for specific applications. Documenting data quality significantly eases the process of datasets selection, increasing the chances of re-use. Independently of domain-specific peculiarities, the quality of data should be documented and well known quality related issues should be explicitly stated in metadata [2]. Hotmaps ensures data quality requiring a feedback from authors/project owners (appearing in the author’s field of collected use cases).

4.10 Data versioning

As the data on the web collaboration platform are continuously changing over time, version information is an essential specification to uniquely identify a dataset, informing data users on the history and determining the exact version they are working with [2]. Good data versioning also enables consumers to learn about the existence of a newer available dataset version. Explicit versioning enables repeatability in research, comparisons and avoids confusion.

All the published dataset are published on GitLab and are under the control of the Version Control System called Git. Git, combined with the use of the GitLab web platform, takes care to authenticate the user and store all the data changes. Git allows to univocally identify the status of file in each time of his history, each status is identified by an unique hash string such as: de1eaf515ebea46dedea7b3ae0e5ebe3e1818971.

Git has been developed to coordinate and share a decentralized development of the Linux Kernel and provides a tool to easily view, compare and share changes between text documents like source code.

All the changes are published using GitLab, from the online web portal an anonymous user can easily retrieve information on the history of each document, who published the document, who modify which line and when, were there were any discussion connected to this change, who was in favour who was against, who at the end decided to accept or refuse a certain change.

The use of git for large binary file, like possible GeoTiff raster maps can be a problem, all the repository using data in a binary format will use a git dedicate extension called: git Large File Storage (git-lfs) [20] that has been specifically developed to efficiently handle this kind of files, more details on this point in section 4.5.

To promote the use of a high quality of the commit messages the data repository will use Git Journal template and rules [21] to auto-generate the changelog of each data release. The Git Journal tool required that each single commit message contains one single logical change described in standardized way.

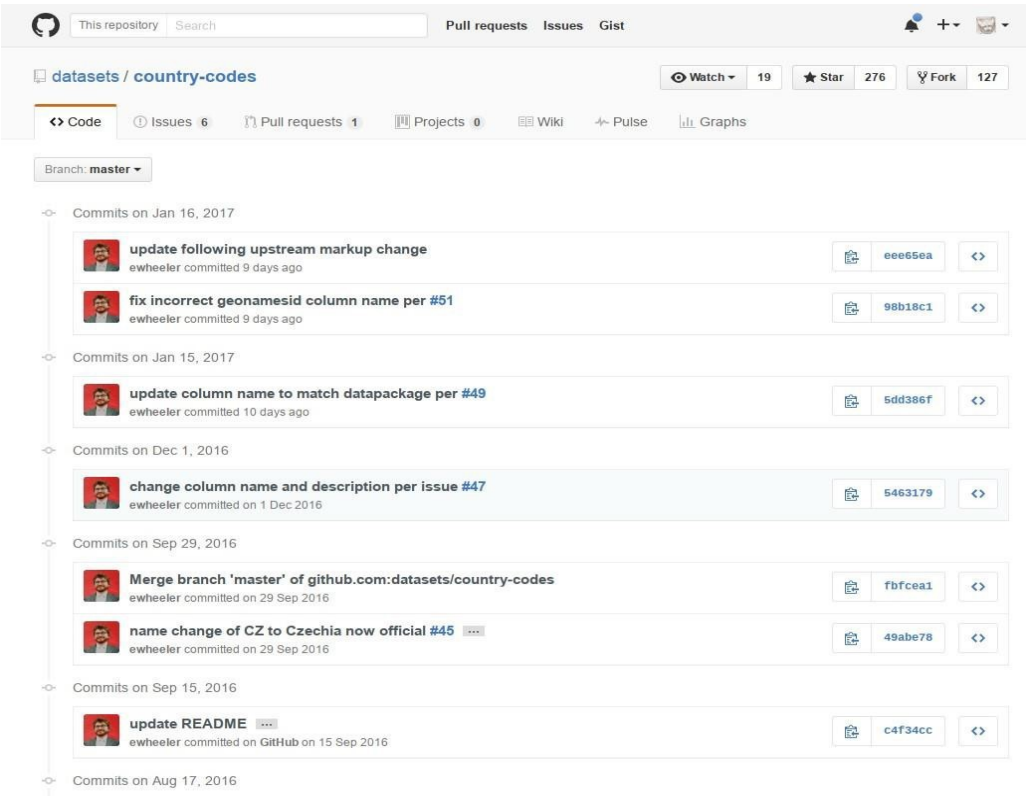

Figure 5 GitLab list of dataset changes, similar interface is provided by GitLab

4.11 Data identification

A common identification system helps data consumers to identify the data and compare them reliably. Data have to be discoverable and citable throughout time. Authors of Hotmaps published datasets use the services offered by Zenodo [22] to manually assign a unique Document Object Identifier (DOI) to the GitLab repository for each stable release of the data set.

4.12 Data access

Data consumers usually require a simple and near real time access to data on the web. The Hotmaps project website and the GitLab portal are accessible from anywhere via web browser without any read protection and provide the ability for bulk download through HTTP or using git without authentication.

The private GitLab repositories used to organize and share sensitive data within the project activities allow a granular access to the dataset that can be defined person by person. The GitLab authentication system support a HTTPS authentication with user/password but support also a strong authentication through a private/public key via a SSH connection.

Both public and private GitLab repositories will remain available beyond the project's end.

The Hotmaps LiveLink is also accessible - only to project partners - from web browser through an HTTPS connection after a one step user/password authentication process. No bulk download services are provided to the public. The project's shared LiveLink (see Figure 6 below) is modelled on a simple folder structure recalling the breakdown of the work packages (WPs). Each WP folder contains subfolders corresponding to documents and deliverables. Additional folders contain information on meetings minutes, logos, budget related documents, etc., see section 4.3 for further details. The Hotmaps LiveLink will not be accessible after the project's end.

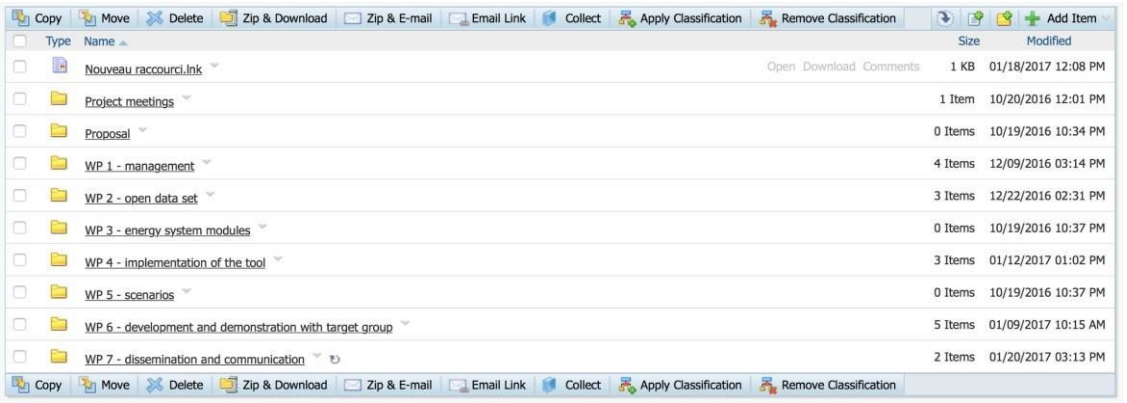

Figure 6 Hotmaps LiveLink Folder Structure for internal documents management

4.13 General Data Protection Regulation

The Horizon 2020 Project Hotmaps fully complies with the EU General Data Protection Regulation (GDPR) [23]: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC. The GDPR applies to the processing of personal data (information relating to an identified or identifiable natural person - so called data subject).

The only personal data stored by the Hotmaps service is the email address that it is necessary to manage the account (i.e. account confirmation and password reset) and to provide the platform service (i.e. platform notifications). A further personal data that it is not required at the moment could be a mobile phone number or with an ID card/key to implement a two factor authentication. Both information are private and accessible only to the user, through the user profile page. The registered user must consent to process the email and the phone number to join and use the platform. The withdraw of the consent imply the user account deletion.

4.14 Conclusion of best practices and guidelines

The previous section introduced the best practices guidelines on data management in Horizon 2020 as recommended by the EC [1] and the proposed recommendations on data on the web best practices by the W3C communities [2]. It addressed these aspects with respect to the type of data generated by Hotmaps, namely: heating, cooling and DHW demand values for residential, service, industry and transportation at EU level. These data will be stored on a public/private git repository hosted on the GitLab platform, while the internal documents and contacts will be shared through the internal LiveLink service. Data types related to projects which Hotmaps is clustering though will also refer to the same guidelines, which Hotmaps will be actively promoting during events held as part of WP7 (Communication and dissemination).

5. Allocation of resources

5.1 Costs estimate to make the open-data FAIR

The Hotmaps project will use free online services and standards:

- GitLab to publicly store, versioning, discuss and revise the data;
- Data Packages format as defined by Frictionless Data (it is free to use and does not require any royalties);
- Zenodo to assign an unique DOI for each data repository and tool developed within the project.

The solution described in this DMP is therefore:

Findable: because there are open-data communities which regularly scan the GitHub repository for datasets published using the Data Package format (https://github.com/datasets/registry). EURAC with the support of CREM/HESO will extend the script to scan also all the GitLab public repositories, before the end of the project.

Accessible: because Hotmaps' use of a format designed to reduce the effort in obtaining information and metadata from the dataset and in accessing raw data, includes when necessary the main scripts for reading and processing.

Interoperable: because the formats chosen for the data and the metadata are agnostic to the Operative System (OS) used or to the application used to read and process the data. The metadata use public vocabularies to describe datasets.

Reusable: because all dataset are published with the relevant attributes included in the metadata file, indicating usage license, dataset provenience and generation sources and because there is an open-data community already using and promoting this standard to publish and share data.

The use of free platforms such as GitLab reduces the cost of long term preservation to zero publicly discussing the changes and updates of the datasets. The ability to visualize the whole history of a dataset increases the chances that other research projects and/or data scientists will use and improve the dataset for their own purposes, creating an open community of users and developers of Hotmaps datasets. Moreover since git is a distributed version control system each user that clone for their own use the data repository has a complete copy of the data set (history, changes, meta-data, etc.). The repository can easily published under a different git-host (e.g. bitbucket.org, SourceForge.net, RocketGit, repo.or.cz, Savannah, etc.) making the proposed solution vendor-neutral.

However, for a complete cost estimation, it is important to include the following costs:

- Person month of researcher and data manager working on the data (harmonizing, storing, enriching, describe, full fill meta-data files, populate and manage the public repository, etc.);
- Legal cost to advice on compliance of legal obligations deriving from reused datasets licenses for output datasets;
- Legal cost to advice on privacy law compliance in output datasets publication.

5.2 Data management responsibilities

EURAC, leader of the WP2 on data collection, is in charge of:

- Monitoring all the data repository that are using and are compliant to the Data Package format as defined in the technical specification;
- Check the quality of the metadata used to describe the datasets. EURAC is in charge to check the (meta-)data quality of the open-data sets and foster the use of a common and harmonised vocabulary for description of the data set.;
- Include in deliverable D2.2 Open Dataset a table with the summary of the data published and under which conditions[3].

Each dataset author is in charge of opening a dedicated repository on https://gitlab.com/Hotmaps. She/he becomes the contact person coordinating the discussion on the data through the GitLab Issues page and accepts or rejects the requests to modify/update the dataset. The Author of the dataset is also responsible for data quality.

[3] The publication conditions will include the performance by the person responsible, before publication of each dataset, of predefined tasks (following procedures and checklists according to the best practises generally adopted) in order to verify that the dataset publication complies with the applicable license (see 4.7 "Data license"), it complies with the applicable privacy law, and does not include confidential data (see 4.2 Sensitive data").

CREM, leader of WP4 (development and deployment of the Hotmaps toolbox platform), is in charge of developing a set of tools in Hotmaps to automatically consider the latest data uploaded and updated from the public/private GitLab repositories by the Hotmaps toolbox platform. CREM will support the project partner with the definition of the meta-data schema for the energy data-sets. CREM will publish the open (meta-)data sets present in the data warehouse using the Open Geospatial protocols that are compliant to the INSPIRE directive.

TUW, leader of the WP3 (development of specific modules), is in charge of checking that each module generates the outputs using the Data Package format and verifying the quality of the metadata provided. Each partner involved in the development of the Hotmaps toolbox modules is in charge to develop a tool that provides the generated outputs using the Data Package format.

TUW, leader of the Task 5.4 on the generation of the scenarios, is in charge of:

- Publishing the GitLab repositories with the final scenarios generated within Hotmaps' activities;

- Including in deliverable D5.2 "Scenarios for EU-28" a summary of data published specifying and publication conditions[4] and justifying and explaining why a certain dataset may not be accessible or under what restrictions it is.

5.3 Data management plan updates and revisions

As stated in the "Guidelines on FAIR Data Management in Horizon 2020":

"...The DMP is intended to be a living document in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur. Therefore, DMPs should have a clear version number and include a timetable for updates. As a minimum, the DMP should be updated in the context of the periodic evaluation/assessment of the project. If there are no other periodic reviews envisaged within the grant agreement, an update needs to be made in time for the final review at the latest."

The current document will be revised with the first publication of the data sets that are used as inputs for the Hotmaps toolbox (M18 - March 2018), with the publication of the project scenarios (M30 - March 2019) and at the end of the project (M48 - September 2020).

[4] The publication conditions will include the performance by the person responsible, before publication, of the tasks indicated in note 3.

Conclusions

Pilot on Open Research Data, aims to maximize access to and re-use of the research data generated during the course of the project. A crucial point of the DMP is to describe how research data are made findable, accessible, interoperable and reusable (FAIR). The Hotmaps project is a Research and Innovation Action (RIA) focusing on energy demand for heating, cooling and domestic hot water (DHW) in the household, service, industry and transportation sectors at EU level.

Hotmaps' aim is to collect and generate new data on spatially disaggregated level (with vector data at NUTS0, NUTS1, NUTS2, NUTS3 level or with raster data with a resolution of 250 x 250 m or higher) to allow especially public authorities to identify, analyse, model and map the resources and solutions needed to supply energy within their territory of responsibility in a resource and cost efficient way.

These results will help authorities to develop heating and cooling (H&C) strategies at a local, regional and national scale, which are in line with renewable energy sources (RES) and CO2-emission targets at national and EU levels. The data generated by Hotmaps will be accessible on the project's website and within a LiveLink repository for internal use.

This deliverable documents a methodology that captures data and provides a description on how generated data are stored, made accessible for verification and re-use, and how they being curated and preserved via the Hotmaps community.

The document is an attempt to follow the best practices as set by the W3C on publishing data on the web, which addresses the same concerns as the DMP guidelines. As a Research & Innovation Action (RIA), Hotmaps will be actively promoting these best practices amongst the wider project communities and providing with new expertise and tools the projects that are unfamiliar with these practices to apply them for their own project-generated data. By so doing, these projects will support a greater exploitation of their results and increase the impact of their outcomes.

All the data and information collected concerning the Pilot areas or through the Hotmaps toolbox platform will be treated as confidential and therefore not be publicly available.

References

[1] EC. H2020 Programme, Guidelines on FAIR Data Management in Horizon 2020, Version 3.0, 26 July 2016. 2016 http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

[2] W3C. Data on the Web Best Practices W3C, Proposed Recommendation 15 December 2016 https://www.w3.org/TR/dwbp/

[3] EEA. European Environment Agency. 2016 http://www.eea.europa.eu/

[4] EC. ESPON. 2016 http://www.espon.eu/main/

[5] EC. ENERGY, EU Building Stock Observatory. 2016 https://ec.europa.eu/energy/en/eubuildings

[6] EC. INSPIRE. Infrastructure for spatial information in Europe. 2016 http://inspire.ec.europa.eu/

[7] W3C, Data Catalog Vocabulary (DCAT), 2014 https://www.w3.org/TR/vocab-dcat/

[8] JoinUp, Share and reuse interoperability solutions for public administrations, Glossary, 2017 https://joinup.ec.europa.eu/glossary

[9] Semantic Interopearability Communities (SEMIC), Core Vocabularies, 2017 http://joinup.ec.europa.eu/site/core_vocabularies/registry/corevoc/

[10] JSON-LD Community Group. JSON for Linking Data. 2016 http://json-ld.org/
[11] schema.org. Energy. 2016 http://schema.org/Energy
[12] OpEnMod Intiative, Glossary, 2017 http://wiki.openmod-initiative.org/wiki/Category:Glossary

[13] EC. HORIZON 2020, Ethics. 2016 https://ec.europa.eu/programmes/horizon2020/en/h2020-section/ethics

[14] Frictionless Data, Specifications for Packaging and Transporting Data, 2017 http://specs.frictionlessdata.io/

[15] EMD International A/S. energyPRO. 2016 http://www.emd.dk/energypro/

[16] IEA. TIMES. 2016 http://iea-etsap.org/index.php/etsap-tools/model-generators/times

[17] Creative Commons. Public Domain Dedication. 2016 https://creativecommons.org/publicdomain/zero/1.0/

[18] Open Data Commons. Open Data Commons Open Database License (ODbL). 2016 http://opendatacommons.org/licenses/odbl/

[19] EU. European Union Open Data Portal. 2016 https://data.europa.eu/euodp/en/data

[20] Git - fast version control, 2017 https://git-scm.com/

[21] Git Large File Storage, 2017 https://git-lfs.github.com/

[22] Git Journal, 2017 - https://github.com/saschagrunert/git-journal

[23] EU. The general data protection regulation applies in all Member States from 25 May 2018. https://eur-lex.europa.eu/content/news/general-data-protection-regulation-GDPR-applies-from-25-May-2018.html

www.hotmaps-project.eu

Funded by the Horizon 2020 programme
of the European Union