Project title:	All Data 4 Green Deal - An Integrated, FAIR Approach for the Common European Data Space
Project number:	101061001
Project Acronym:	AD4GD
Type:	HORIZON-AG - HORIZON Action Grant Budget-Based
Work program topics addressed:	HORIZON-CL6-2021-GOVERNANCE-01

Deliverable No: D1.1

MULTI-ACTOR AND KNOWLEDGE CENTRES CO-DESIGN AND REQUIREMENTS DEFINITION REPORT

Due date of deliverable:	28/02/2023
Actual submission date:	28/02/2023
Version:	1.0
Main Authors:	Florian Jasche, Constanze Ritzmann, Lena Brüch, Jacqueline Ullmann

Co-funded by the European Union, Switzerland and the United Kingdom

Document Metadata

Project number	101061001
Project title	All Data 4 Green Deal - An Integrated, FAIR Approach for the Common European Data Space

Deliverable title	Multi-actor and knowledge centres co-design and requirements definition report
Deliverable number	D1.1
Deliverable version	1.0
Contractual date of delivery	28/02/2023
Actual date of delivery	28/02/2023
Document status	Final
Document version	1.0
Online access	
Dissemination	Public
Work package	WP1: Semantic Interoperability Space
Partner responsible	Fraunhofer Institute for Applied Information Technology FIT
Author(s)	Lucy Bastin, Lena Brüch, Florian Jasche, Vincent-Henri Peuch, Constanze Ritzmann, Hella Schwarzmüller, Jacqueline Ullmann
Editor(s)	
Reviewer(s)	Victoria Lush, Francesca Noardo
EC Project Officer	Lara Congiu

Abstract	The Green Deal Data Space (GDDS) addresses a wide audience of different stakeholders. We follow a Human-Centred Design (HCD) approach to understand the needs of these stakeholders. To identify a base set of relevant stakeholders, we conducted an in-person workshop with the consortium members. We then prioritised the stakeholders and classified their role in the GDDS. We separated global/EU-level stakeholders and stakeholders most relevant to pilots 1 and 2. As agreed within the consortium, pilots 1 and 2 (water quality and biodiversity) should be the starting point and therefore we conducted eight interviews with the partners (KWB and CREAF) to understand their working practices and interaction with their respective stakeholders. Based on the interviews, we identified an initial set of 30 user requirements. Our next steps will be to (a) do the stakeholder analysis for the third pilot and (b) conduct further interviews and workshops. We also plan to collect and refine the list of requirements for the GDDS by conducting a market analysis of other data spaces and open data portals as well as a design sprint activity that also could include our sister projects or other projects like BioAgora and GREAT.

Co-funded by the European Union, Switzerland and the United Kingdom

Keywords	stakeholder analysis, user requirement, Green Deal Data Space, Human-Centred Design
Disclaimer	Views and opinions expressed in this deliverable are those of the author(s) only and do not necessarily reflect those of the European Union, the United Kingdom or Switzerland. Neither the European Union nor United Kingdom nor Switzerland can be held responsible for them.

Co-funded by the European Union, Switzerland and the United Kingdom

Document version history

Version history			
Version	**Date**	**Modification reason**	**Modified by**
0.1	05/12/2022	Initial version of the document + ToC	Constanze Ritzmann
0.2	10/02/2023	Main content added	Florian Jasche, Constanze Ritzmann, Lena Brüch, Jaqueline Ullmann
0.3	14/02/2023	Added image of HCD process and minor editing	Florian Jasche, Constanze Ritzmann
0.4	24/02/2023	Internal review	Victoria Lush, Francesca Noardo
1.0	26/02/2023	Finalisation for submission	Constanze Ritzmann, Florian Jasche

Co-funded by the European Union, Switzerland and the United Kingdom

Abbreviations

Abbreviation	Definition
API	Application Programming Interface
AU	Aston University
CREAF	Centro De Investigacion Ecologica Y Aplicaciones Forestales
CSV	Comma-Separated Values
DG ENV	Directorate-General Environment
DG DEFIS	Directorate-General Defence Industry and Space
EC	European Commission
ECMWF	European Centre For Medium-Range Weather Forecasts
EEA	European Environment Agency
ESA	European Space Agency
FAIR	Findable, Accessible, Interoperable, and Reusable
FIT	Fraunhofer Institute for Applied Information Technology
FTP	File Transfer Protocol
GBIF	Global Biodiversity Information Facility
GD	Green Deal
GDDS	Green Deal Data Space
HCD	Human-Centred Design
IoT	Internet of Things
IUCN	International Union for Conservation of Nature
JRC	Joint Research Centre
KWB	Kompetenzzentrum Wasser Berlin
MEDEEC	Mediterranean Experts on Climate and environmental Change
NGO	Non-governmental organisation
UNFCCC	United Nations Framework Convention on Climate Change
WCMC	World Conservation Monitoring Centre
WP	Work Package

Co-funded by the European Union, Switzerland and the United Kingdom

Table of Contents

Co-funded by the European Union, Switzerland and the United Kingdom

Table of Tables

Table of Figures

Co-funded by the European Union, Switzerland and the United Kingdom

Executive summary

The Green Deal Data Space (GDDS) will address a wide audience of different stakeholders. As we follow a Human-Centred Design (HCD) approach, it is important to understand the needs and the requirements of the stakeholders. In the beginning of the project not all relevant stakeholders were already known and, therefore, identifying and collecting information about these stakeholders was the focus of the initial steps. As the first step, we conducted an in-person workshop at the kickoff meeting with the consortium members to brainstorm a base set of relevant stakeholders for the GDDS in general, including stakeholders that are more important to specific project pilots. Based on the collected stakeholder data, we conducted three additional online workshops to prioritise and classify each stakeholder's role in the GDDS (e.g. data provider and/or data consumer). These workshops were conducted for the global/EU-level stakeholders (organisations working on an EU-level such as EC Knowledge and Topic centres) and the stakeholders for pilots 1 and 2 (water pollution and biodiversity). Due to the lower maturity of the third pilot (air pollution) at the start of the project, we agreed within the consortium to focus on pilots 1 and 2. To date, we identified a total of 40 relevant stakeholders with almost equal split between data providers (25 stakeholders) and data consumers (19 stakeholders)[1].

The consortium members KWB and CREAF play leading roles in their respective pilots. Therefore, we conducted eight interviews with the partners (KWB and CREAF) to further understand their working practices and how they interact with other stakeholders. Based on the interviews, we identified a number of challenges, such as finding the right data source for a given problem or that data transformation and aggregation can be cumbersome due to non-standardized data formats or missing metadata. From these insights, we derived an initial set of 30 user requirements, however, further investigation is needed.

As Task 1.1 will continue for further 24 months, our next steps will be to (a) do the stakeholder analysis for pilot 3 and (b) conduct interviews with stakeholders relevant to this pilot. We also plan to collect further requirements for the GDDS by (c) conducting a market analysis of other data spaces and open data portals as well as (d) organising a design sprint that could possibly include contributions from our sister projects (FAIRiCUBE[2], USAGE[3] and B3[4]) as well as other projects like BioAgora[5] and GREAT[6].

[1] One stakeholder can be both, a data consumer and a data provider. You find a detailed list of stakeholders in Tables 1, 2 and 3 of this deliverable.
[2] https://fairicube.nilu.no/
[3] https://www.usage-project.eu/
[4] https://cordis.europa.eu/project/id/101059592
[5] https://bioagora.eu/
[6] https://www.greatproject.eu/

Co-funded by the European Union, Switzerland and the United Kingdom

1 Introduction - Purpose & Scope of the Deliverable

The deliverable presents the preliminary results of T1.1 Multi-actor and knowledge centres co-design and requirements definition. The purpose of this task is first, to identify relevant stakeholders and second, to elicit the initial set of requirements for the GDDS. The task contributes to Milestone 2: Project baseline established.

This document describes our planned approach and methodologies combining Human-Centred Design with the agile development process. Furthermore, it presents all our research activities and their current results. Those results illustrate the users' needs and interests, by elaborating initial descriptions of the context of use of a future GDDS and the initial set of requirements.

The section 2 briefly introduces the three AG4GD pilots. They will serve as starting points for developing an initial design of the GDDS. For the purpose of this deliverable, we focus on the first two pilots. As the project progresses, the third pilot will become more concrete, and additional requirements will be collected accordingly.

Section 3 depicts the approach and methodology used to retrieve the initial set of requirements from the given desk research results, stakeholder analysis and pilot partner interviews.

Then section 4 and 5 show the results of the stakeholder analysis and the prioritisation and classification.

Section 6, 7 and 8 depict the requirements elicitation, by presenting the context-of-use descriptions as a primary output from the interviews, then showing the related user stories and finally the initial set of requirements.

Finally Section 9 concludes the deliverable and outlines the next steps.

2 Understanding the Context - Introduction of the three pilots

The initial concept (design vision) of a GDDS-portal will be based on the needs expressed by the identified stakeholders and project pilot partners. A deep understanding of the three project pilots and their context therefore serves to break down the overall project idea of designing a suitable data space to support any Green Deal-related application into feasible and concrete initial building blocks. The following subsections are verbatim excerpts from the Grant Agreement for AD4GD and have to be attributed to the original authors; Lucy Bastin and Hella Schwarzmüller for Pilot 1, Lucy Bastin for Pilot 2, and Vincent-Henri Peuch and Lucy Bastin for Pilot 3.

2.1 Pilot 1 - Water Pollution

Berlin's ecosystems face challenges related to the city's rapid densification (300.000 new inhabitants in the last 10 years) and climate change. Amongst these are the extension of soil sealing, longer dry periods (e.g. 2018 and 2019), heat island issues, and low connectivity of blue/green areas, all of which may be aggravated by large-scale changes in the regional water balance due to climate change, the closure of open mining pits in the Lausitz region and related changes in natural flow conditions, groundwater recharge and quality of surface water and groundwater. 30% of the total area of Berlin is covered by forests and other green and open space and 6% is covered by water. The latter comprises not only the Spree-Havel river and lake system, but also more than 300 small urban lakes. Many of these are heavily dependent on groundwater or even solely on stormwater, and the increased occurrence of temperature and rain extremes promotes their qualitative and quantitative degradation. At the same time, small urban lakes have great potential for biodiversity and, together with their riparian areas, provide recreational spaces. In addition to the consequences of overuse, stormwater discharges, which feed rainwater from sealed surfaces

Co-funded by the European Union, Switzerland and the United Kingdom

(especially roads without cleaning), are often, according to the EEA[7], the cause of poor water quality. Typical consequences are heavily polluted sediments (e.g. with heavy metals or microplastics) and a pronounced lake electrophobia, which in turn leads to blue-green algae blooms, oxygen depletion and odour pollution. In addition, small lakes <50 ha are not part of the monitoring and action cycles of the EU Water Framework Directive and accordingly only receive higher attention in individual cases (e.g. where they function as bathing waters).

Today, no simple methods are known to identify particularly polluted lakes, to investigate the type of pollution and to prioritise and evaluate measures. Accordingly, solutions driven by combining available data and deploying AI technologies yield great potential to improve the management of small urban lakes, identify drivers and indicators for stressors of water quality and make better informed decisions.

This pilot aims to develop a condition index for polluted small urban lakes by combining IoT, CitSci and EO data with AI technologies to identify drivers and indicators for stressors of water quality. KWB will lead the strategic interactions with stakeholders and thus we started our requirements collection with KWB employees.

2.2 Pilot 2 - Biodiversity

Functional landscape connectivity for animal and plant dispersal is a fundamental consideration for national, regional and local governments in making strategic spatial decisions about protected area networks, industrial / residential zoning, agricultural practice regulations and incentives, and land remediation. These governmental stakeholders seek for standardised metrics on the state and protection of biodiversity for international reporting. At the same time, they need accessible information products, that are more sensitive to local contexts, to facilitate dialogue about policy with stakeholders in the countryside such as farmers.

Approaches to quantifying connectivity vary from graph-based models (e.g. Biodiversity Indicators Partnership's ProtConn) that generalise the exploration and exploitation of intervening matrix, to remote sensing approaches that represent habitats as a continuous field classified according to their accessibility and reciprocal affinity (such as the Ecological Connectivity Index currently used in Catalonia). ProtConn statistics for protected areas are available at country and ecoregion level through the JRC's Digital Observatory of Protected Areas[8], and are accepted internationally for reporting progress against Aichi Target 11 and Sustainable Development Goal 15. However it is difficult for regional governments to access the data and services which would allow consistent connectivity metrics to be re-generated and downscaled across a complex landscape mosaic which must support food security and a range of other human uses. There is significant potential for estimates of functional connectivity to be validated with IoT and CitSci observations that verify the usage of corridors by the target species. However, the integration of such data into workflows mainly based on satellite and cadastral data is complex and not well supported by existing tools.

This pilot aims to deliver FAIR and scalable computation services for assessing landscape functional connectivity, it intends to integrate CitSci, administrative, satellite, socio-economic and IoT data streams, and AI analysis services. While AU leads technical development and integration, CREAF leads strategic interactions with stakeholders and thus was our first point of contact for defining the initial context of use and requirements.

---

[7] EEA 2018 European Bathing water quality in 2017. Report 02/2018: https://www.eea.europa.eu/publications/european-bathing-water-quality-in-2017

[8] https://dopa-explorer.jrc.ec.europa.eu/, https://research.aston.ac.uk/en/publications/the-digital-observatory-for-protected-areas-dopa-explorer-10, https://dopa.jrc.ec.europa.eu/static/dopa/static/dopa/files/factsheets/en/DOPA%20Factsheet%20C1%20EN%20Connectivity.pdf

Co-funded by the European Union, Switzerland and the United Kingdom

2.3 Pilot 3 - Air Pollution

Curbing greenhouse gas emissions and monitoring progress and effectiveness of mitigation policies is a key need in the context of the UNFCCC Paris Agreement. Countries in Europe and in the world have committed to report their emissions and about progress towards implementation of their nationally determined contributions. As part of the European GD, the EC has tasked the Copernicus programme to support the monitoring and verification process of Member State countries by implementing an ambitious system based on Earth Observation -the future CO2M (CO2 Monitoring) satellite constellation- and very advanced processing systems in order to derive emissions estimates from measured atmospheric concentrations through so-called "inverse modelling" techniques. A number of European Research and Innovation projects (CHE, VERIFY and CoCO2 to mention the most directly relevant) have supported or are supporting the required underpinning research. Robustness, transparency and quality of the emissions estimates are pivotal and a growing variety of in situ observations could in principle be used for validation purposes or even as input for the inversion either for assimilation or as a priori information. Integration of many sources of observation is essential in order to ensure that all relevant available information is eventually integrated for providing the best possible input to reporting and to policy design and monitoring. In the context of the Copernicus Atmosphere Monitoring Service, ECMWF is already in contact with stakeholders at international, European, national and regional levels for co-designing solutions that meet their needs and requirements. ECMWF used the first 6 months of the project to define the exact scope of this third dedicated pilot study. The search for an area where a substantial amount of new low-cost sensors is available, is a crucial step to be taken, before the future operational emissions monitoring system can be tested.

This pilot focuses especially on evaluating the possible added value of socio-economic (e.g. traffic counts, energy production and transfer) and IoT (low-cost sensors) for estimating greenhouse gas concentrations and emissions in Europe. ECMWF will lead the task, building upon activities already in place to provide a CO2 Monitoring and Verification Support capacity, with workflows that do not currently include IoT sources and could include more socio-economic data, particularly in real-time.

3 Approaches and Methodology used in AD4GD

In AD4GD, we are following the Human-Centred Design approach according to ISO 9241-210 (International Organization for Standardization 2019). This approach is incremental and iterative, as is characteristic of agile development projects. This means that the project results are developed incrementally together with stakeholders, starting from requirements discovery to documentation of the solution concepts (Figure 1). The development of the operating and visualisation concept of the GDDS will take place in an iterative manner. It will start from a rough concept, developed based on the initial set of requirements defined in this deliverable. Based on user and stakeholder feedback on this rough concept an interactive prototype will be developed in order to avoid erroneous (software) development work.

As the name suggests, the human (user) is in the middle of the development process. The Human-Centred approach to system design adheres to the following principles (International Organization for Standardization 2019):

- the design is based upon an explicit understanding of users, tasks and environments;
- users are involved throughout design and development;
- the design is driven and refined by user-centred evaluation;
- the process is iterative;
- the design addresses the whole user experience;
- the design team includes multidisciplinary skills and perspectives.

Co-funded by the European Union, Switzerland and the United Kingdom

The process of the Human-Centred Design consists of four steps (Figure 1):

1. understanding and specifying the context of use;
2. specifying the user requirements;
3. producing design solutions;
4. evaluating the design.

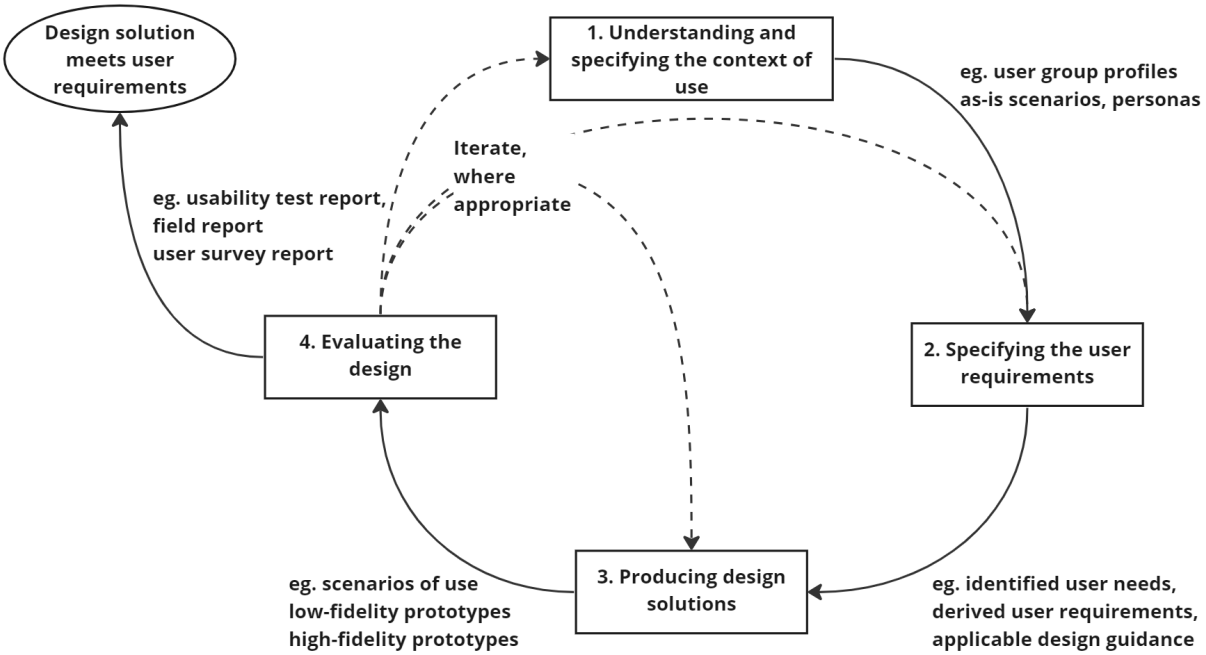

Figure 1: HCD activities in the development process of the project according to ISO 9241-210 (International Organization for Standardization 2019).

It is a framework that offers multiple methods that are all built on close interaction and discussion with users. This ensures the best information and feedback gathering possible, from a human perspective. Hereby, the application of specific methods, like interviews, focus groups or desk research depends on the level of already gathered domain knowledge as well as on available resources.

Within AD4GD the Human-Centred Design process is applied iteratively. This allows us to adapt to changing user needs and requirements as well as to limitations and problems which may occur during the project development at any stage. Iterations can happen between any phases in the process, but are usually triggered after evaluation.

This deliverable reports on the results of the first two steps. We present the context of use and specify a first set of user requirements that may be redefined and most probably will be extended in the course of the project and presented in the second version of this deliverable in M30.

With the help of all AD4GD consortium partners, FIT started working on step 1 **Understanding and specifying the context of use** and step 2 **Specifying the user requirements**. Three methods were applied within the first months of the project:

- **stakeholder analysis** that focused on identifying and prioritising relevant stakeholders together with the consortium members right in the beginning and thus establishing potential users of the GDDS that would be available for in-person interviews in November and December 2022. The prioritisation was done based on availability and proximity to the project use cases.
- context **interviews**
- a qualitative **content analysis** to define the initial set of user requirements (requirements definition process).

Co-funded by the European Union, Switzerland and the United Kingdom

3.1 Method 1: Stakeholder Analysis

The stakeholder mapping method is an activity to collaboratively collect names of stakeholders to a potential new product / system (Giordano et al., 2018). The method originates from the domain service design (Stickdorn and Schwarzenberger, 2016). Service design is a Human-Centred Design approach that focuses on the strategic and holistic planning, implementation, and management of the customer experience. Stakeholders are defined as legal bodies, organisations, institutions or possibly individuals that in the future may be interested to use a system or service product under development. The reason to use a stakeholder map is to visualise all stakeholders and their respective interrelations. Arrows between the stakeholders visualise their connection. Normally, the most important stakeholder is visualised in the centre of the map. This afterwards helps to identify the needs for the GDDS of the most important stakeholders. The mapping exercise is led by a neutral moderator who prompts a group of experts[9] with a related trigger question, like: "Who do you consider as a relevant stakeholder for pilot x?" The experts then, in a silent reflection, note down on post-its one stakeholder per post-it. Then each expert brings up their notes to a wall/whiteboard. The resulting list of stakeholders is then clustered by theme (e.g. rose, lavender and basil might build the cluster "plant"). Following this, prioritisation exercises help structuring a vast amount of information and deciding where to start.

3.2 Method 2: contextual interviews to create the description of context-of-use

The methods and procedures described in the following section as "contextual interview" are based on ISO 9241 on the ergonomic design of human-system interactions and best practice for the participatory development of innovative design solutions based on stakeholder requirements.

A design solution is never absolutely good or bad but must be assessed and developed relative to the requirements of a particular context of use. The context of use according to ISO 9241-11 comprises a combination of users, goals, tasks, resources, and the technical, physical and social, cultural and organisational environments in which a system, product or service is used (International Organization for Standardization 2018). The usability that results from use of a system, product or service depends on the particular characteristics of each of these components. Therefore, it is important to identify the relevant characteristics and their variability in the actual or intended context(s) of use. ISO/IEC 25063 identifies the information items to be included in a description of context of use.

Accordingly, it is of great importance for the success of a project to illuminate and precisely understand these aspects of the context of use first, in order to be able to set corresponding user requirements for the later design of an operating and visualisation concept. A dedicated AD4GD context of use interview guideline was created to lead and structure the context of use interviews. See Annex I for the full interview guide used. The interview guide was created based on FITs internal interview guide for context-of-use interviews (ISO 9241-210). The in-depth interviews are conducted and recorded by two interview moderators, one asking questions and one taking notes. At least three representatives of a stakeholder group are interviewed. A description of the context-of-use (context scenario) is the final result of this method step.

3.3 Method 3: Requirements definition process

The context scenarios are then evaluated in terms of hurdles or problems in the current work practice by usability and user experience experts. The identified problems imply user needs and requirements for the product under development. The identified needs and requirements for use then result in a user needs report (compliant to ISO/IEC 25064) and User requirements specification (compliant to ISO/ISE 25065). The

[9] Who will be considered as an expert is to be defined by the context.

Co-funded by the European Union, Switzerland and the United Kingdom

user requirements are clustered using so-called task models in order to make them manageable for the task-based development of prototypes.

4 Results of the Stakeholder Analysis

The initial workshop to identify relevant stakeholders took place on September 22, 2022. The workshop was integrated into the third day of the AD4GD kickoff meeting in Barcelona. The goal of the workshop was to structure various stakeholders in the context of three project pilots. The workshop was based on the list of known stakeholders created during the preparation of the DoA. The generated overview of stakeholders, that are involved in the pilots, serve as a basis for the later prioritisation step.

During the workshop, all project partners of AD4GD were present. The pilots and the related stakeholders were discussed one after the other and the respective pilot partners and other experts in the topic area participated in the recording and clustering of the stakeholders. The session was moderated by FIT staff and recorded on post-its on a whiteboard. The goal of the workshop was to create one stakeholder map for each pilot.

The stakeholder map has a circular structure. The various stakeholders are each described on a post-it note. Similar stakeholders are grouped close to each other on the map. Areas exist in which all stakeholders with a similar function are located, for example "Data provider", "NGOs", "Scientific Networks" or "Politics". Arrows are used to show initial connections, such as the information on how data flow between the stakeholders, in order to depict the basic data delivery processes as is at the moment. Furthermore, we differentiate the stakeholders by their scale.

Small scale stakeholders refer to stakeholders that are closely associated with the dedicated pilot, where needs can be ascertained in the further steps. They are interested representatives who are or might be (directly) involved in the pilots on a regional level. Examples include regional political representatives, research institutions that also work regionally in the context of the pilot, or local citizens.

Global/EU-level stakeholders, on the other hand, ensure the scalability and visibility of the project at EU and global level and are therefore also relevant to be included. Global/EU-level stakeholders focus on national, EU-wide or global impact. These are, for example, globally active political institutions, globally active research institutions or NGOs. Of course, there are stakeholders that are in between small and global/EU-level stakeholders. One example is the Deutscher Wetterdienst (German Meteorological Service) that operates on a national level. They collect data all over Germany but only in Germany.

As we used a slightly modified version of the stakeholder mapping method, we decided to place the small-scale stakeholders in the centre of the stakeholder map, as they are more tangible than the global/EU-level stakeholders. In this case, no statement is to be made about the importance of the stakeholders at this point in time, both the global/EU-level and the small-scale stakeholders are highly relevant for the success of the project.

Figure 2 shows the stakeholder map that was co-created with regard to the first pilot on the topic of "water pollution". For this stakeholder map, input was provided primarily by KWB.

Co-funded by the European Union, Switzerland and the United Kingdom

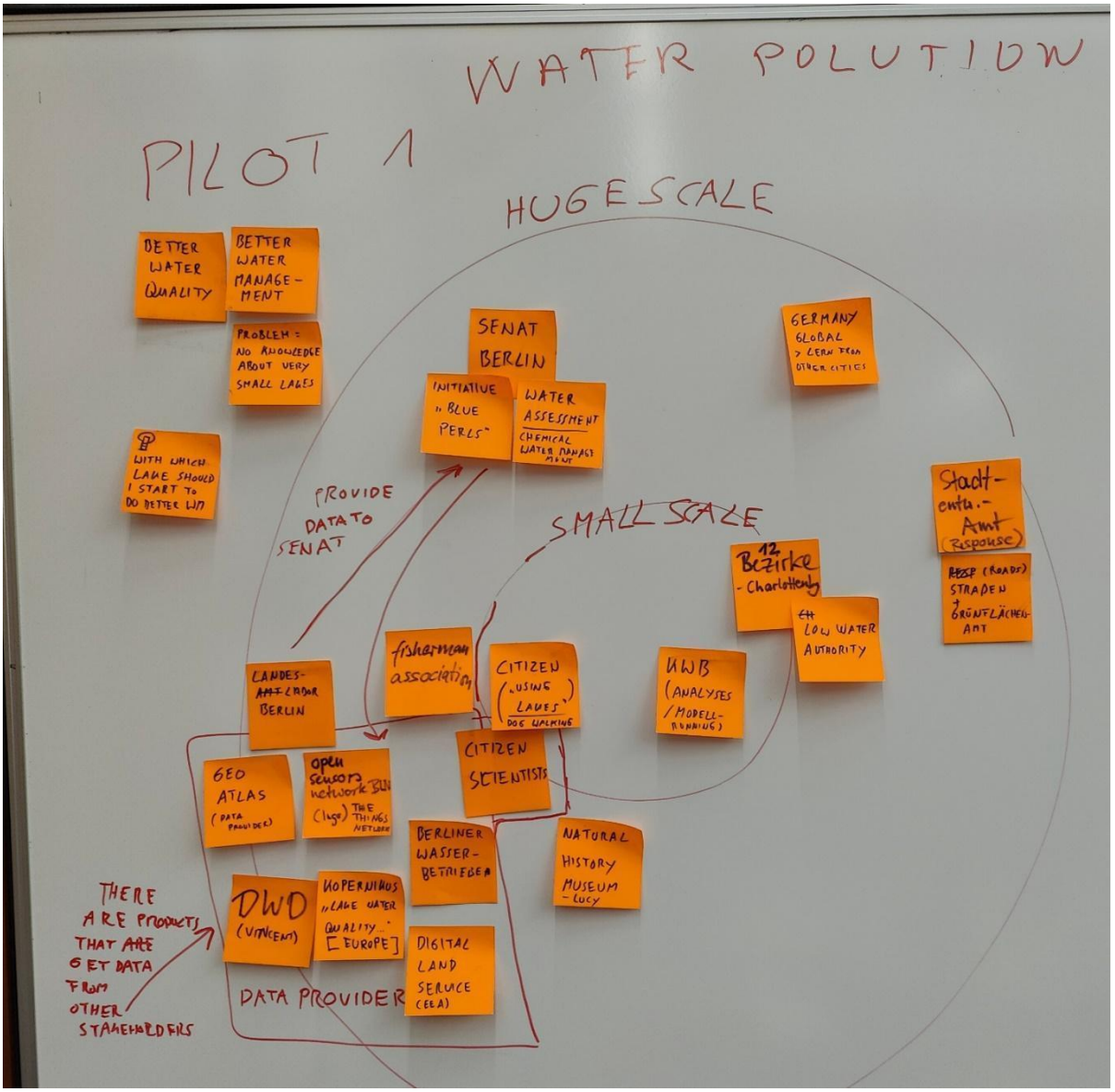

Figure 2: Stakeholder map for pilot 1 "water pollution".

The stakeholder map shows in the left upper corner the general goal and challenge of the pilot. The goals are to obtain a better water quality and to allow better water/ lake management and monitoring. Today, small lakes are not under the regulation of the EU Water Framework Directive and therefore they are not monitored which results in lacking knowledge about these lakes. For this reason, policy makers do not have the necessary information to make a proper action plan for effective water management.

On a small scale the main stakeholders are KWB, other water authorities and citizens. At global/EU-level level, stakeholders can be divided into several groups: on the one hand the data providers and on the other hand projects and political actors who consume data.

Figure 3 shows the stakeholder map that was co-created with regard to the second pilot on the topic of "biodiversity". For this stakeholder map, input was provided primarily by CREAF.

Co-funded by the European Union, Switzerland and the United Kingdom

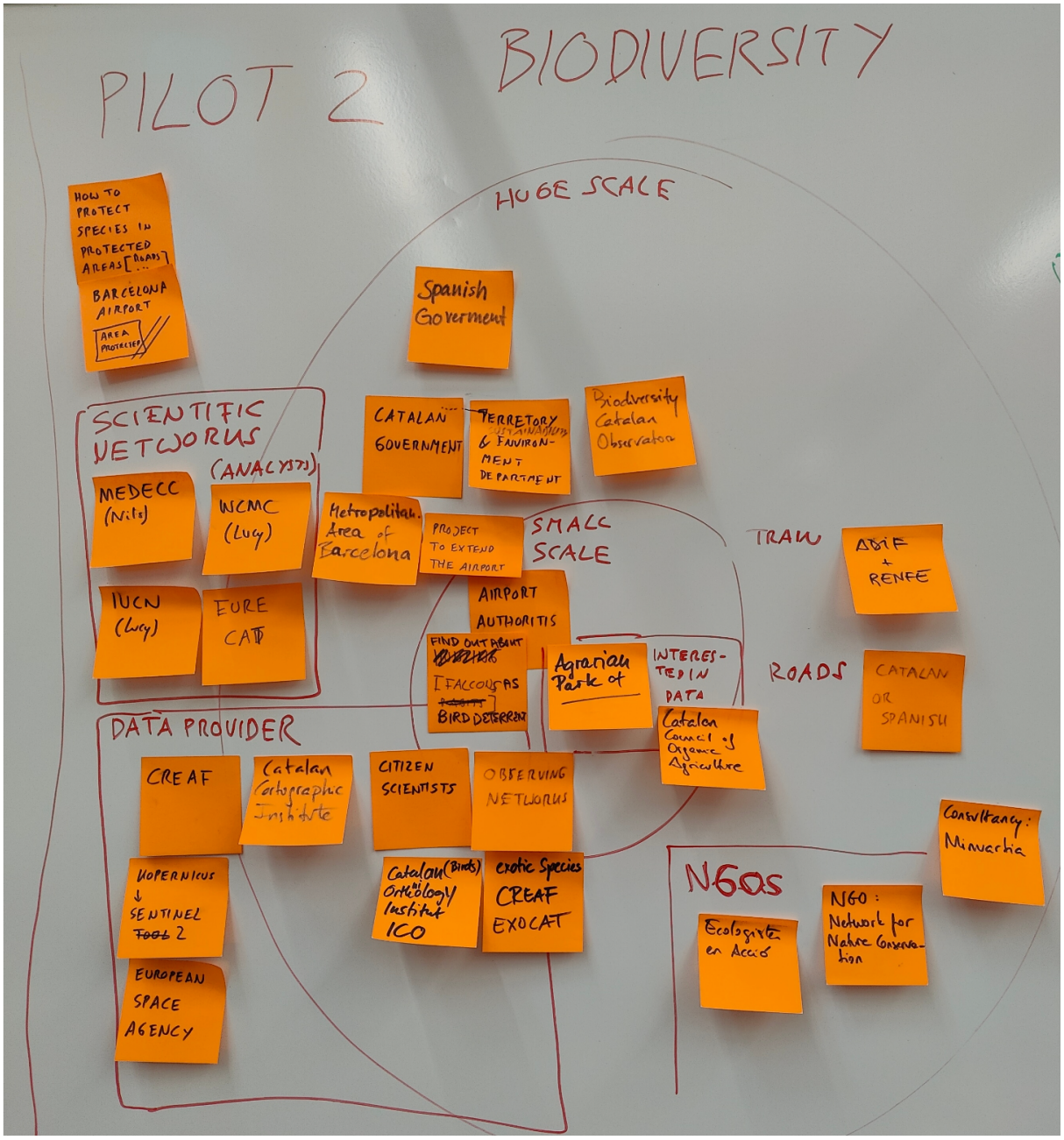

Figure 3: Stakeholder map for pilot 2" biodiversity".

The stakeholder map shows in the upper left corner the general goal and challenge of the pilot. At the date of the workshop, the overall goal of the biodiversity pilot is to protect species in protected areas. The pilot concentrates on the region of Catalonia and especially on the protected areas around the airport of Barcelona.

There were more global/EU-level scale stakeholder than small scale stakeholder identified for this pilot. The local (small scale) stakeholders are the airport authorities. Also the Catalan Council of Organic Agriculture and the Agrarian Park are described as small scale stakeholders that are interested in data.

At the global/EU-level different data providers were identified, such as Copernicus. Scientific networks MEDEEC, WCMC, IUCN and EURE CAT were detected. Also political actors, such a the spanish government, the Catalan government and the biodiversity Catalan observation were named as global/EU-level stakeholders. Relevant NGOs could be Ecologistes en acció or the network for nature conservation in

Co-funded by the European Union, Switzerland and the United Kingdom

Catalonia. The operators and owners of roads and train lines in Spain were identified as important stakeholders, as they contribute to the urban sprawl of protected areas.

Figure 4 shows the stakeholder map that was co-created with regard to the third pilot on the topic of "climate change and air quality". For this stakeholder map, input was provided primarily by ECMWF.

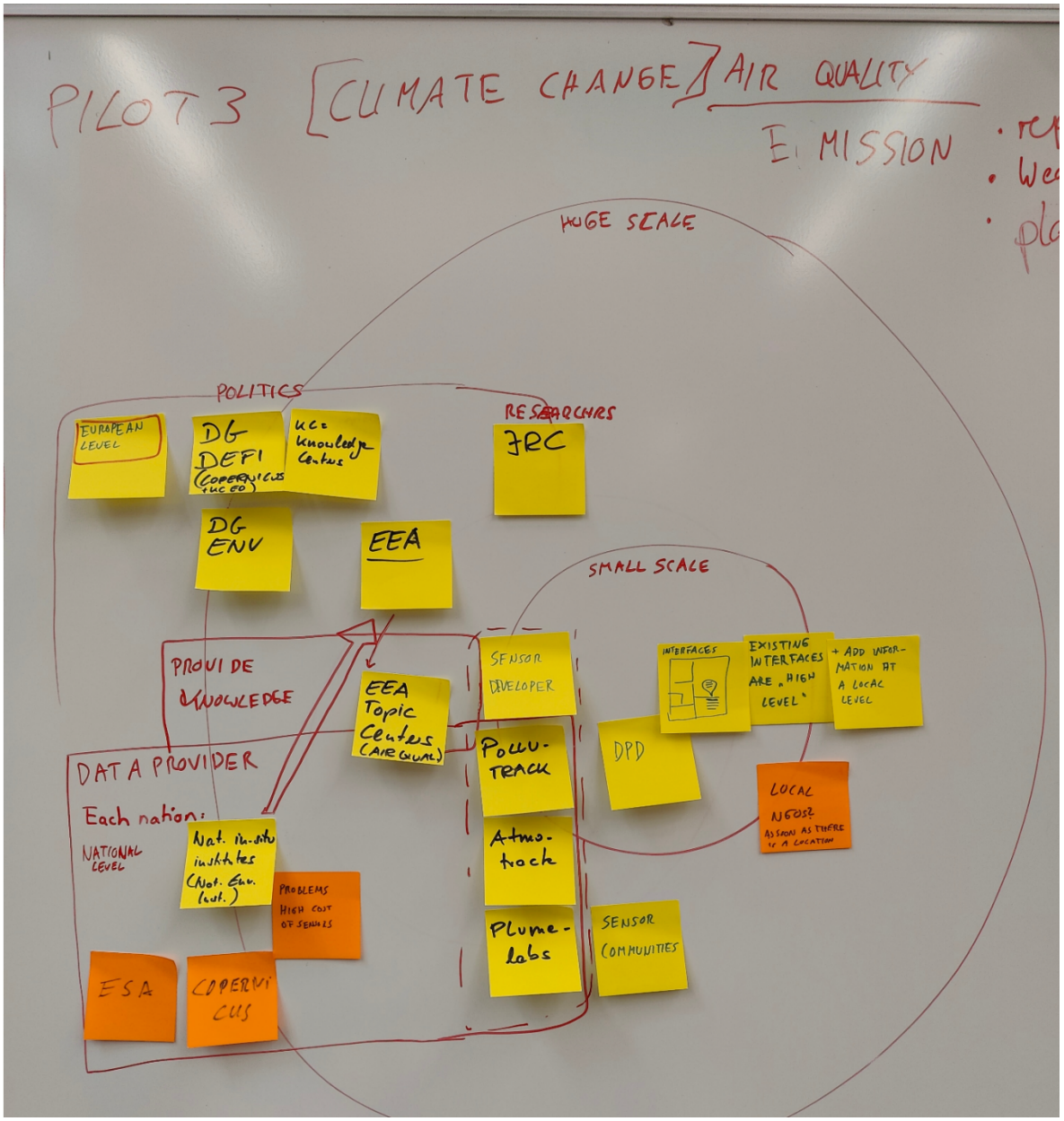

Figure 4: Stakeholder map for pilot 3 "air pollution".

At the time of the stakeholder discussion, it was not yet clear where the pilot would be located, but northern Italy was mentioned as an option.

The delivery service DPD was mentioned as a possible small scale stakeholder that could provide air quality data. The company could cooperate with the pilot by installing sensors on the transporters. The huge scale stakeholders were divided into three clusters: data providers, knowledge providers and politics. Various sensor manufacturers and large data collectors such as ESA or Copernicus were named as data providers. The EEA Topic Centres and EC Knowledge Centres were identified as data consumers that generate knowledge based on the data from the future GDDS. The DG ENV and DG DEFIS were identified as stakeholders closely related to politics. Overall, the stakeholder maps help to visualise identified

Co-funded by the European Union, Switzerland and the United Kingdom

stakeholders and their relation to the pilots. Because the workshop encouraged an open discussion, the number of identified stakeholders is high and the relevance of these stakeholders for the GDDS vary greatly. To further refine the list of stakeholders, we prioritised them in the follow-up workshops.

5 Results of prioritisation and classification of Stakeholders

Since a large number of stakeholders was identified in the first workshop, it was necessary to define their priority in order to limit the initial interviews number, allowing to respect the planned goal of interviewing at least five to six individuals from key stakeholder organisations by the end of month 4. Therefore, FIT conducted three more online workshops after the kickoff meeting. In these workshops, we discussed the identified stakeholders and their role in the future GDDS. Each workshop was dedicated to one of the pilots to assess the relevance of the stakeholders to the GDDS in general but more specifically to the pilot cases and how important these stakeholders are for the success of the planned activities in the pilots. The prioritisation for each stakeholder was collaboratively set from low to high. The responsible pilot partners as well as the pilot supervisor (AU) participated in the workshops.

Based on the stakeholder maps from the workshop at the kickoff meeting, AU identified multiple stakeholders that were present in several pilots. Lucy Bastin created a separate list for these stakeholders and checked which other institutes or organisations were closely related to them (Table 1). This list of stakeholders was discussed in the first internal online workshop with the project partners AU, CREAF and FIT on October 18, 2022. The goal of this workshop was to prioritise the stakeholders and to classify their role regarding the GDDS. Furthermore, additional stakeholders were added to the list during the discussion.

Table 1: High-level stakeholders relevant to the GDDS in general.

Stakeholder name	Relation to pilots	Relevance to project
EC Knowledge Centre on Biodiversity	2, possibly 1	very high
EC Knowledge Centre on Earth Observation	1,2, especially 3	very high
EEA European topic centre on inland, coastal and marine waters	1	very high
EEA European topic centre on biological diversity	2	very high
EEA European topic centre on climate change mitigation	3	very high
GBIF	2	high
Bio-Agora	2, possibly 1	high
Copernicus (lake quality and land service)	1,2,3	high
INSPIRE portal	1,2,3	high
Eurostat	1,2,3	high
Directorate-General Environment	1,2,3	medium
International Union for Conservation of Nature IUCN	2	medium
UN Environment Programme World Conservation Monitoring Centre	2	medium

The list presented in Table 1 does not claim to be complete and it will be extended during the project.

Co-funded by the European Union, Switzerland and the United Kingdom

In the second workshop, we discussed the stakeholders that are relevant for the first pilot case about small lakes in Berlin (Table 2). This workshop was an internal online workshop with partners from AU, KWB and FIT and took place on October 27, 2022.

Table 2: Stakeholders relevant to pilot 1.

Stakeholder Name	Type in relation to the GDDS	Relevance for pilot
Kompetenzzentrum Wasser Berlin (KWB)	Data Consumer + Data Provider	very high
Senat Berlin	Data Consumer + Data Provider	high
Deutscher Wetterdienst (German Meteorological Service)	Data Provider	high
Geoportal Berlin	Data Provider	high
Copernicus Land service	Data Provider	high
Natural History Museum Berlin	potential Data Providers + connection to citizen science activities	medium
Stiftung Naturschutz Berlin (Foundation for Nature Conservation Berlin)	Data Consumer	medium
Regenwasseragentur Berlin (Rainwater Agency)	Data Consumer + Data Provider	medium
Citizens (e.g. Fisherman association)	Data Provider	medium
Landeslabor Berlin	Data Provider	medium
Berliner Wasserbetriebe	Data Provider	medium
Lower water authority	Data Provider	low
Roads and green spaces Office	Data Provider	low
The Things Network	Data Provider	low
Urban Development Office	Data Provider	low

The list presented in Table 2 does not claim to be complete and it will be extended during the project.

In the third online workshop we discussed the stakeholders that are relevant for the second pilot case about biodiversity (Table 3). The workshop was conducted with partners from AU, CREAF and FIT and took place on October 20, 2022.

Table 3: Stakeholders relevant to pilot 2.

Stakeholder Name	Type in relation to the GDDS	Relevance for pilot
Ecological and Forestry Applications Research Centre (CREAF)	Data Provider + Data Consumer	very high
Catalan Ministry of Territory and Sustainability (DTES)	Data Provider + Data Consumer	high
Network for Nature Conservation in Catalonia	Data Consumer	high

Co-funded by the European Union, Switzerland and the United Kingdom

Stakeholder Name	Type in relation to the GDDS	Relevance for pilot
Exocat - Sistema d'informació de les espècies exòtiques de catalunya	Data Provider	medium
Metropolitan Butterfly Monitoring Schemes	Data Provider	medium
Catalan Council of Organic Agricultural Production	Data Provider	medium
Catalan Institute of Ornithology	Data Provider	medium
L'Observatori del Patrimoni Natural i la Biodiversitat - Mammals Atlas	Data Provider	medium
Observatorio de la Biodiversidad Agraria	Data Provider	medium
Minuartia	Data Consumer	medium
Institut Cartogràfic i Geològic de Catalunya	Data Provider	medium
Ecologistes en acció	Data Consumer	low

The list presented in Table 3 does not claim to be complete and it will be extended during the project.

6 Results of Contextual Interviews

6.1 Data collection (Pilot 1 + Pilot 2)

The initial data collection happened via eight semi-structured interviews that were scheduled in November and December 2022.

Five interviews with KWB employees were conducted in person during a field visit in Berlin on November 15th and 16th, 2022. While three interviews were held at the KWB office, the other two were "walking interviews". The two walking interviews and observations served for a profound understanding of the context of use. The first one was a visit to a small lake (Halensee in Berlin, see Figure 5) exploring and observing KWBs' work and their ambitions for small lakes. The second one was a field visit to a project site, that allowed in-depth understanding of a previous project called FLUSSHYGIENE[10] on the influence of heavy rains on the water quality of bathing lakes connected to the river Spree.

[10] https://www.kompetenz-wasser.de/en/forschung/projekte/flusshygiene

Co-funded by the European Union, Switzerland and the United Kingdom

Figure 5: Walking interview with KWB at the Halensee in Berlin, 15 December 2022.

The other three interviews were conducted with CREAF employees remotely using Microsoft Teams on November 18th, December 12th and 15th, 2022. The tools currently in use were demonstrated during the online interview.

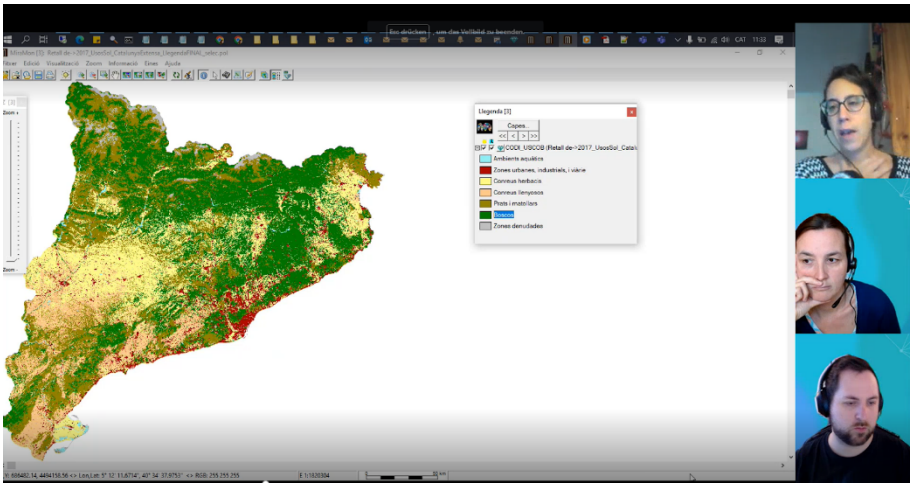

Figure 6: Screen shot of the interview with CREAF employee Ivette Serral, showing the Miramon Software, 18 November 2022.

The interviews lasted 75 minutes on average. All interviews were audio recorded (in the case of KWB) and video recorded (in the case of CREAF) and afterwards written transcripts for qualitative data analysis were created. All interviews were conducted following the interview guide as shown in Annex I.

6.2 Preliminary context-of-use Descriptions

"A context-of-use description includes information about the users and other stakeholder groups, the characteristics of each user group, the goals of the users, the tasks of the users, and the environment(s) in which the system is used"(ISO/IEC 25063:2014).

The interview participants identified themselves as research associates, data scientists, research technicians with a background in environmental science and remote sensing and one as a full stack

Co-funded by the European Union, Switzerland and the United Kingdom

developer, with a degree in biology. While the roles and activities of interviewees differed, a common schema within their work related to handling green deal data could be identified.

Their shared general task can be called "researching and analysing data". This task consists of the following three steps.

1. collecting data,
2. processing data, and
3. delivering an interpretation or recommendation based on the findings.

Based on these activities, we created a context of use description called: "a researcher as data consumer".

Based on our stakeholder prioritisation, we decided to start with the data consumer perspective but as the GDDS cannot live without data, we also wanted to understand the data provider perspective. We therefore decided to acquire one interviewee from CREAF who had a slightly different role in handling green deal data. This interviewee is responsible for maintaining the technical part of a citizen science data collection portal[11]. Since we have interviewed only one person to date, the context of use: "a data provider" is very preliminary, and needs more profound research, but allows to shed light on the requirements for this important part of a future GDDS as well. We therefore call this preliminary context of use description: "a data provider".

The following section contains both contexts of use descriptions.

6.2.1 "A researcher as data consumer"

A researcher receives a research job or a research question from an advice seeking body. In order to answer the research question or sometimes just deliver the requested processed data, the researcher **first identifies potential data sources and then downloads all available data**.

Depending on the project, the required data varies from data about rainfall, soil and geological information, to water quality data in general for KWB (pilot 1), to land use data like satellite pictures, land set images or data on the status of the vegetation often based on remote sensing images for CREAF (pilot 2). The sources for these data sets also vary. For example, in the case of KWB, data on rainfall is often retrieved from the Deutsche Wetterdienst (German Meteorological Service). In this case the data is publicly available[12] via different interfaces (APIs/FTP). Depending on the data set, the formats may vary. KWB interviewees reported that they normally use FTP access to download historical meteorological data as CSV files. However, downloading files manually is very time consuming and therefore KWB created R scripts[13] to automatically download reports for a given time frame. The GeoPortal Berlin, also called FIS-Broker[14], is also often used by KWB to retrieve data. It may also happen that data is provided by clients or project partners. This usually takes place by email or via file servers using CSV or Excel files. These data sets often are not available publicly. KWB employees also take water samples themselves and have them tested by the country laboratory. The results are then sent via email to the dedicated employee of KWB. In some cases, data sets may also contain sensitive information, because they relate to critical (water) infrastructures.

The second step for the researcher is to **check the quality of the available data sets and determine whether the available data can help answer the research question.** Metadata describing the actual data play a

[11] http://www.mosquitoalert.com/

[12] "As laid down in the Ordinance Setting the Terms of Use for the Provision of Federal Spatial Data (GeoNutzV), all spatial data and spatial data services available for free access may be used without any restrictions provided that the source is acknowledged. When speaking of spatial data, this also includes any location-related weather and climate information presented on our [DWD] open web pages."

[13] https://github.com/KWB-R

[14] The FIS-Broker is an open data portal operated by Senatsverwaltung für Stadtentwicklung, Bauen und Wohnen (Senate Department for Urban Development, Building and Housing). It is accessible via WMS or WFS and licensed under DL-DE BY 2.0.

Co-funded by the European Union, Switzerland and the United Kingdom

crucial role in this process but often metadata documents are not available, and the researcher needs to make sense of data by investigating its content, for example, what kind of unit a measurement has. Overall, this is a complex decision and often a collaborative process. Close discussions with the respective research team and the client or the advice seeking entity take place to determine whether the data set is suitable to solve the research question.

The researcher then spends a major part of their time **processing the data**. This may refer to data cleaning activities, such as deleting unneeded parts, consolidation activities, such as copying data from several time periods or several national data repositories into one single file or transforming the resolution of the data in order to create comparable clusters. This accounts as well for geospatial variables as well as, for example, grouping species. This process, also called harmonising, is the most crucial activity as it results in the final cleaned data set that is sometimes directly shared with clients and is used as a basis for interpretation and recommendations. Data interpretation and recommendations are mostly delivered in written report documents.

6.2.2 "A Data Provider"

The main task of a data provider is the upload of aggregated cleaned data sets to a data sharing portal. The example we base this context of use description on comes from CREAF and is based on regular GBIF data contributions with cleaned and standardised data sets coming from the european project Mosquito Alert[15].

Private citizen science enthusiasts take pictures from mosquitos and upload those using a dedicated app to a CREAF database. The observations are qualified either by the enthusiast first, meaning that the species name is added. After an expert approval process the observation data is published to the report and thus a unique detailed species observation report is accumulated from all European regions.

This accumulated data is then processed by the data provider, meaning that the data set is manually transformed into a predefined standard and then uploaded to the GBIF database and thus made available publicly. The manual transformation of the data may include: deleting unnecessary information, renaming table headings for standard conformity and/ or adding metadata.

The data upload happens as a bulk upload that does not follow a regular timetable, but is triggered by the number of data entries collected and the time available by the person in charge. Thus, at the moment, the uploads happen about twice per year.

7 User Stories

Based on the contextual interviews we synthesised user stories that describe user needs. Because the user needs in both pilots were very similar, we do not differentiate between the pilots here. Each user story was assigned a priority. The user stories will be revised and refined as the project evolves. Furthermore, we derived a first set of concrete requirements for the GDDS.

Table 4: Preliminary collection of user stories.

#	User Story	Priority	Label
1	As a researcher, I would like to find data by providing a location, because I lack the experience where to find it (e.g. when I work in an unfamiliar region).	High	Data consumer

[15] Mosquito Alert is a citizen science project to investigate and control disease-carrying mosquitoes in Europe. http://www.mosquitoalert.com/en/

Co-funded by the European Union, Switzerland and the United Kingdom

#	User Story	Priority	Label
2	As a researcher, I would like to find data by providing a topic/keyword, because I lack the experience where to find it (e.g. when I work on a new topic).	High	Data consumer
3	As a researcher, I need sufficient metadata because I need to know what the data is about.	High	Data consumer
4	As a researcher, I need sufficient metadata because I need to know if the data meets my quality requirements.	High	Data consumer
5	As a researcher, I would like to see when the data set was last modified to determine if I have the most recent data.	Medium	Data consumer
6	As a researcher I would like to validate the data, towards various variables (e.g. completeness of dataset), to make sure that the dataset meets my quality requirements.	Medium	Data consumer
7	As a researcher, I need sufficient metadata to know how the data was collected, because I need to understand if the data meets my trustworthiness requirements.	High	Data consumer
8	As a researcher, I would like to see who authored the data set to determine if I trust this data provider.	Medium-High	Data consumer
9	As a researcher, I would like to get all the necessary data from a single point of interaction, so I don't have to ask many different sources.	High	Data consumer
10	As a researcher, I need access to historical (meteorological, geological, ...) data to create my models and show changes over time.	High	Data consumer
11	As a researcher, I need easy access to private/confidential data without requesting them via email every time.	Low-Medium	Data consumer
12	As a researcher, I would like to get help in combining/harmonising data from different sources because the data sources often don't use standardised formats (e.g. timestamp format, decimal separator, satellite image resolution, ...).	High	Data consumer
13	As a researcher, I would like to be able to share my data (preliminary results) with clients and/or project members to get feedback on my work.	Low	Data provider
14	As a researcher, I would like to share my data/results to contribute to my research community/field.	Medium-High	Data provider

The list presented in Table 4 is based on the contextual interviews of pilots 1 and 2 and does not claim to be complete. The list will be extended during the project. There were some additional user stories identified, however, these were out of scope and are not listed here.

Co-funded by the European Union, Switzerland and the United Kingdom

8 First Set of Requirements

As user stories are formulated more problem-oriented to provide more context, we also provide a preliminary list of more concrete requirements for the GDDS that are based on the user stories and comments during the interviews. However, some of the requirements listed below are still generic and broad.

1. The GDDS needs search functions.
    - 1.1. The users should be able to provide a location by text to get all available datasets for this location.
        - 1.1.1. Country
        - 1.1.2. Region name
        - 1.1.3. City name
        - 1.1.4. Zip Code
        - 1.1.5. Coordinates
    - 1.2. The users should be able to select a region on a map to get all available datasets for this region.
        - 1.2.1. There could be different modes of dataset selection: Contains – the dataset is completely in the selected region, Overlaps – the dataset only partly overlaps with the selected region.
2. The GDDS needs a list of themes of all available datasets.
    - 2.1. Topics should be semantically grouped.
3. All datasets registered in the GDDS need metadata.
    - 3.1. The most important metadata (list below) of a selected data set should be shown in a condensed way.
        - 3.1.1. Author
        - 3.1.2. Date of last modification
        - 3.1.3. Temporal coverage
        - 3.1.4. Licence
    - 3.2. The complete metadata of a selected dataset should be collapsed first.
    - 3.3. The metadata should include information about the data quality (tbd based on data set) – provided by the data owner and/or assessed by the GDDS.
    - 3.4. The metadata should include information about how the data was collected.
4. The GDDS should include as many data providers as possible.
    - 4.1. The GDDS should integrate the Deutscher Wetterdienst (German meteorological service) as a data provider.
    - 4.2. The GDDS should integrate the FIS-Broker as a data provider.
    - 4.3. The GDDS should integrate the Copernicus Open Access Hub as a data provider.
    - 4.4. The GDDS should integrate the USGS EarthExplorer as data provider.
    - 4.5. The GDDS should integrate GBIF as a data provider.
    - 4.6. The GDDS could integrate iNaturalist.org as a data provider.
    - 4.7. The GDDS could integrate Green Deal relevant data from the EU open data portal data.europa.eu.
5. The GDDS should provide a functionality to do basic data transformation (e.g. converting units or timestamps).
6. The GDDS needs a mechanism to define the access to a dataset (access control system).

The list of requirements does not claim to be complete and will be extended and refined during the project.

To make the requirements available to all consortium members, we are using the issue tracker functionality of a shared GitLab repository. Each requirement is converted into an issue, stating the problem and, if already known, a possible solution. Furthermore, the issues are prioritised and tagged to show to which component the requirement belongs to (frontend, backend, data consumer, data provider, ...). Each issue can be discussed to determine if and how it should be implemented. In addition, the issues can be used to visualise and check the progress of the software implementation on a Kanban board.

Co-funded by the European Union, Switzerland and the United Kingdom

9 Conclusion & Next Steps

The topic of the Green Deal is very broad and building a data space for this topic involves a lot of different stakeholders. Therefore, we first focused on identifying important stakeholders for our GDDS. In a series of workshops, we were able to identify, prioritise, and classify (data consumer vs data provider) an initial set of 40 stakeholders. The stakeholders range from small local institutions over the EC Knowledge centres to other globally operating initiatives. Topic wise, the stakeholders are aligned to the topics of our three pilots: zero pollution (focusing on water quality), biodiversity and air quality.

To understand the context of use for the future GDDS, we conducted contextual interviews with the pilot partners of pilots 1 and 2. With the interviews we could uncover current problems and hurdles in the work with Green Deal related topics mainly from the data consumer perspective. These problems lead to the first set of requirements for the GDDS. Main problems are the findability of data (required data is often distributed on multiple platforms) but also the interoperability of the data including missing metadata.

To further deepen the understanding of the context of use of the GDDS we will conduct further interviews. This includes interviews with the local project "Blaue Perlen" (blue pearls) in Berlin because they are also interested in the water quality of small lakes in Berlin. With these interviews we search for further collaboration on this topic. In addition to the interviews regarding pilot 1, we also want to interview the Catalan Ministry of Territory and Sustainability because we identified them as a very important stakeholder for pilot 2. As in the meantime pilot 3 was further concretized, we aim to do the stakeholder prioritisation and context interviews with the pilot partner ECMWF. Based on these activities, we aim to create user journeys to show the current state but also how the work practices could look like with a functional GGDS.

Besides the perspective from the pilots, we plan further interaction with related projects like BioAgora and the EC Knowledge Centres to integrate a more high-level perspective on the GDDS. This also includes exchange with the GREAT project.

Other planned activities to collect requirements for the GDDS include a market analysis of other data spaces and open data portals as well as an intensive design sprint that also could include our sister projects or other projects like BioAgora and GREAT.

10 References

AD4GD Project 101061001. GRANT AGREEMENT. European Research Executive Agency (REA)

Giordano, Fanny, Nicola Morelli, Amalia De Götzen, and Judith Hunziker. 2018. "The stakeholder map: A conversation tool for designing people-led public services." In Service Design and Innovation Conference: Proof of Concept.

International Organization for Standardization. 2018. ISO 9241-11:2018 Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts. https://www.iso.org/standard/63500.html.

International Organization for Standardization. 2019. ISO 9241-210:2019 Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems. https://www.iso.org/standard/77520.html.

Stickdorn, Marc, and Klaus Schwarzenberger. 2016. "Service design in tourism." In Entrepreneurship und Tourismus: Unternehmerisches Denken und Erfolgskonzepte aus der Praxis.

Co-funded by the European Union, Switzerland and the United Kingdom

Annex I

In the following we are presenting the interview guideline. The questions are formulated openly and are intended to stimulate narrative. The interviewer should follow these narratives with "why" and "how" questions but without drifting too far from the original topic.

Introduction

1. What is your exact role and activities at your institute? Phrase it in one or two sentences.
2. Which tasks do your activities consist of? List typical core tasks, i.e. if large proportion of time or frequently recurring or very important
3. How are these activities organised? E.g. as mixed work, as a sequence of tasks, as a monotonous individual task?

Requirements

4. What qualifications are required to complete your tasks (Task management / software use)?
5. Who (or which incident) determines what to do? (Who makes the selection? Autonomy of processing, division of the work, external data sources)
6. What tools are required (for the task management / software use)? Which of these are missing, which are additionally desired?

Normal execution

7. Which work steps are to be carried out?
    a. About which production processes do you receive information or data and when? (What is the data? What data is important or necessary?)
8. Which work steps recur frequently? (Automation desired / required?)
9. Which work steps are carried out automatically? Are there / allowed / desired / necessary possibilities for the user to influence automated work steps?
10. Does it happen that several users have to work on the same object (e.g. process, file, document, data record) at the same time?
    a. What overview does the user have with regard to the overall activity?
11. Is there a fixed sequence of work steps and if so, what does it look like? (Is flexibility useful / necessary?)
12. Which results / partial results are created and how are they used / continued?
13. What feedback do you get regarding the work results and the impact of your work?

Special features of the execution

14. What are the interruptions and why? What troubles occur (organisational / social / technical)?
15. How are errors reported and fixed (organisational / social / technical)?
16. Which important special cases must be considered (do the user spontaneously think of them, e.g. by division of work/collaboration)?

Organisational framework / overall context

17. What organisational goals are there with regard to the activity?
18. Are there mechanisms for performance management / performance monitoring? (If yes, which ones? Are these required?)

Co-funded by the European Union, Switzerland and the United Kingdom

19. Which changes affecting the processing of tasks are to be expected or desired? What suggestions does the interviewed person have?
20. Which work results / work steps directly affect third parties (e.g. customers)? And what follows from this?
21. What stress factors are there and how are they dealt with?

Conclusion

22. If you could have one wish based on what we discussed today, what would you wish for?