Towards a sustainable Open Data ECOsystem

D4.2
An approach to steer the behaviour of non-government data holders towards open data through a technical strategy

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 955569.

ODECO

Project Acronym	ODECO
Project Title	Towards a sustainable Open Data ECOsystem
Grant Agreement No.	955569
Start date of Project	01-10-2021
Duration of the Project	48 months
Deliverable Number	D4.2
Deliverable Title	An approach to steer the behaviour of non-government data holders towards open data through a technical strategy
Dissemination Level	Public
Deliverable Leader	University of the Aegean (UAEGEAN)
Submission Date	22-10-2024
Author (s)	Charalampos Alexopoulos - UAEGEAN, Maria Ioanna Maratsi - UAEGEAN, Mohsan Ali – UAEGEAN, Georgios Papageorgiou - Farosnet S.A., Abdul Aziz – UNIZAR, Dagoberto Herrera Murillo - UNIZAR

Document history

Version #	Date	Description (Section, page number)	Author & Organisation
V0.0	06-06-2024	Kick-off Meeting	Charalampos Alexopoulos - UAEGEAN
V0.1	05-06-2024	First draft table of contents	Charalampos Alexopoulos - UAEGEAN, Maria Ioanna Maratsi - UAEGEAN
V0.1a	06-06-2024	Review of table of content	Andrea Polini - UNICAM, Charalampos Alexopoulos - UAEGEAN, Maria Ioanna Maratsi - UAEGEAN, Mohsan Ali – UAEGEAN, Georgios Papageorgiou - Farosnet S.A., Abdul Aziz – UNIZAR, Dagoberto Herrera Murillo - UNIZAR
V0.2	20-08-2024	First draft	Charalampos Alexopoulos - UAEGEAN, Maria Ioanna Maratsi - UAEGEAN, Mohsan Ali – UAEGEAN, Georgios Papageorgiou - Farosnet S.A., Abdul Aziz – UNIZAR, Dagoberto Herrera Murillo - UNIZAR
V0.2a	13-09-2024	Second draft. Restructuration after the feedback from the first peer-review.	Charalampos Alexopoulos - UAEGEAN, Maria Ioanna Maratsi - UAEGEAN, Mohsan Ali – UAEGEAN, Georgios Papageorgiou - Farosnet S.A., Abdul Aziz – UNIZAR, Dagoberto Herrera Murillo – UNIZAR

Version #	Date	Description (Section, page number)	Author & Organisation
			Reviewers: Javier Nogueras - UNIZAR, Francisco Javier López Pellicer - UNIZAR
V0.3	18-10-2024	Final report. Editorial work after the feedback from the second peer-review	Reviewers: Bastiaan van Loenen - TUDELFT, Javier Nogueras, UNIZAR, Francisco Javier López Pellicer - UNIZAR, Birger Larsen - AAU
V1.0	22-10-2024	Final editing.	Danitsja van Heusden-van Winden – TUDELFT

Table of Contents

List of figures

List of tables

Abbreviations

AI	Artificial Intelligence
API	Application Programming Interface
AR	Augmented Reality
D	Deliverable
ESR	Early-Stage Researcher
GDPR	General Data Protection Regulation
GenAI	Generative Artificial Intelligence
HOT	Humanitarian OpenStreetMap Team
HOT-TM	Humanitarian OpenStreetMap Team Tasking Manager
IDIP	Intelligent Data Interaction Platform
IOT	Internet of Things
LLM	Large Language Model
LODC	Linked Open Data Cloud
MS	Milestone
MVP	Minimum Viable Product
NGD	Non-Government Data
NGO	Non-Governmental Organisation
NLP	Natural Language Processing
NPO	Non-Profit Organisation
OD	Open Data
ODP	Open Data Portal
ODK	Open Data Kit
ODECO	Open Data ECOsystem
OSM	OpenStreetMap
PII	Personally Identifiable Information
RAG	Retrieval-augmented generation
TAM	Technology Acceptance Model
TRL	Technology Readiness Levels
TSLO	Technical, Semantic, Legal, and Organizational
VR	Virtual Reality
TW	Training Week
UCD	User-centred design
VGI	Voluntary Geographic Information
WAI	Web Accessibility Initiative
WP	Work Package

Nr	Partner	Partner short name	Country
Beneficiary			
1	Technische Universiteit Delft	TU Delft	Netherlands
2	Katholieke Universiteit Leuven	KUL	Belgium
3	Centre National de la Recherche Scientifique	CNRS	France
4	Universidad de Zaragoza	UNIZAR	Spain
5	Panepistimio Aigaiou	UAEGEAN	Greece
6	Aalborg Universitet	AAU	Denmark
7	Università degli Studi di Camerino	UNICAM	Italy
8	Farosnet S.A.	FAROSNET S.A.	Greece

Nr	Partner	Partner short name	Country
Partner organisations			
1	7eData	7EDATA	Spain
2	Digitaal Vlaanderen	DV	Belgium
3	City of Copenhagen	COP	Denmark
4	City of Rotterdam	RDAM	Netherlands
5	CoC Playful Minds	CoC	Denmark
6	Derilinx	DERI	Ireland
7	ESRI	ESRI	Netherlands
8	Maggioli S.p.A	MAG	Italy
9	National Centre of Geographic Information	CNIG	Spain
10	Open Knowledge Belgium	OKB	Belgium
11	SWECO	SWECO	Netherlands
12	The government lab	GLAB	United States of America
13	Agency for Data Supply and Infrastructure	ADSI	Denmark
14	GFOSS Open Technologies Alliance	GFOSS	Greece
15	Inno3 Consulting	IC	France
16	Regione Marche	RM	Italy
17	Open Data Institute	OCI	United Kingdom
18	Swedish National Archives (Riksarkivet)	SwNA	Sweden

1. Introduction

1.1 Problem Definition

Open Data initiatives often focus on a narrow selection of providers and user groups while neglecting other stakeholders in the ecosystem. This exclusion is evident in the dominance of governmental actors in providing open data. Sustainable open data ecosystems advocate for inclusivity, suggesting that non-government actors should also share their data in open formats (Charalabidis et al., 2018, Runeson et al., 2021). However, encouraging non-government actors to share their data as open data remains a largely unexplored challenge. The ODECO project aims to address this issue through a multidisciplinary approach that includes: (1) understanding the barriers and motivations for non-government actors to contribute open data to the ecosystem, (2) developing the technical mechanisms to facilitate this process, and (3) establishing governance structures to support and sustain it.

Specifically, Task 4.2 examines the technical mechanisms that can be used to encourage non-government data holders to share open data. Task 4.2 considers the barriers and motivations of task 4.1, and technical requirements for non-government actors to become active contributors to the open data ecosystem (by releasing their data into the open data ecosystem) and proposes the corresponding technical steering mechanisms to suitably prepare the ground towards this direction. The technical steering mechanisms presented in this Deliverable are of varying Technology Readiness Levels (TRL), those developed within the ODECO project being of higher TRL.

Governmental actors have historically been the dominant providers of open data. In many cases, such responsibility is engraved in policy documents, from non-binding strategic plans to binding laws and regulations such as the EU Open Data Directive. On the other hand, while there have been some initiatives by non-governmental actors to publish their data as open data, it is still a limited practice and remains largely desired (van Loenen et al., 2018). Non-governmental actors are actors that are outside the public/governmental sector, such as companies, civil society organisations, and the media. Exploring the motivations and challenges for non-governmental actors to contribute open data is thus essential to develop sustainable open data ecosystems that incorporate both government and non-government open data.

Aligned with the ODECO's Description of Action, this deliverable aims to answer the following **main objective:**

- **Identify and further develop technical mechanisms (solutions) to facilitate OD sharing by non-government data holder groups.**
  These solutions will entail quality assurance methods, the development (or embedment) of data integrator services, as well as new user (data holder) interfaces from at least the working prototypes of the engaged ESRs. These solutions will help the non-governmental data holders to more easily create and share their data as open data using these novel and automated solutions. First and foremost, the developed solutions could be embedded to the already existing open data spaces (e.g. Zenodo, OpenAIRE, wikidata etc.) that enable the upload of datasets from any non-governmental stakeholder. Secondly, they could be used towards the (further) development of their own open data portals (or websites sharing open data) within their organisations.

In order to address the main objective, three sub-objectives are necessary:

- **Sub-objective 1**: Analyse the identified motivations and technical requirements for non-governmental data holders to contribute open data. The identified motivations are briefly presented in section 3 of this deliverable and in more detail in D4.1.
- **Sub-objective 2:** Analyse the technical requirements and barriers which hinder non-governmental data holders to contribute open data. The developed and proposed technical mechanisms should not disregard the requirements of the data users to build "user-driven"

solutions. Many of these requirements derive from ODECO D2.2. Furthermore, it should be noted that open data ecosystemic "circularity" will be achieved through the non-governmental data holders too serving their dual role in the ecosystem as users and providers of open data. This means that specific open data user requirements are also important for the development of technical mechanisms.

- **Sub-objective 3:** Define the main challenges under which the solutions will be developed. This sub-objective identifies and describes the major challenges for non-governmental data holders. The proposed and developed solutions are meant to address one challenge.

To address the main objective and sub-objectives, this deliverable follows the methodological approach presented in section 1.2.

1.2 Role of this deliverable in the ODECO project

The ODECO deliverable 4.2 is part of Work Package 4, "From an Exclusive to an Inclusive Open Data Ecosystem". D4.1 explores the motivations of non-government actors to become active contributors to the open data ecosystem. The relation to the other deliverables in WP4 is as follows. In D4.2, we explore technical mechanisms to steer the behaviour of non-government data holders towards open data. It will report on technological ways to promote the inclusion of non-government data holders in the open data ecosystem. D4.3 explores a governance strategy to steer the behaviour of non-government data holders towards open data. It will report on steering mechanisms and approaches for activating NGD holders in the open data ecosystem. Connecting to the work conducted in ODECO in the other deliverables of the same Work Package (WP4), for the development of this Deliverable (D4.2), the background knowledge derived from D4.1 is used for the identification of non-governmental data holder groups, as well as their barriers and motivations to share data. Deliverable D4.3 ran in parallel with D4.2.

1.3 Structure

This report is structured as follows:

- Chapter 1 defines the problem and research questions, as well as it explains the role of this deliverable within the ODECO project.
- Chapter 2 explains the overall methodological strategy and processes followed to produce the results of this deliverable.
- Chapter 3 presents the motivations and barriers of each given stakeholder type to contribute to the Open Data Ecosystem, as those were identified during Task 4.1. Each sub-chapter starts with a brief definition of the given data holder, and their respective barriers and motivations for open data sharing.
- Chapter 4 describes the technical requirements concerning the Data Holder Groups of Chapter 3 and these act as the underlying technical basis and gap which needs to be bridged by the technical mechanisms which are presented in Chapter 5.
- Chapter 5 presents the Challenges derived from the analysis performed in Chapters 3 and 4.
- Chapter 6 defines the proposed technical steering mechanisms, corresponding to the Challenges of Chapter 5.
- Chapter 7 serves as the discussion and insights for the whole document. Shared perspectives and conclusions are also presented, as well as limitations of the scope of this report, and a future research agenda based on the findings.

2. Methodology

The overall methodological approach is following the systematic analysis of previous research results as it is presented in Figure 1. The sources of information (highlighted in green) are coming from previous deliverables and ESRs related research using different methodologies (like SLR-systematic literature reviews, interviews, workshops). After the gathering of these lists, we proceeded to the evaluation of the preliminary results under the aspect of "Sharing process (from creation to publication and re-use) for non-governmental stakeholders". Next, we synthesised and validated the most prominent challenges for our scope. Finally, ODECO ESRs participated into 2 tasks in order to identify and propose the associated with the challenges technical steering mechanisms (green highlighted stars): (a) review of already existing solutions and (b) proposition of solutions on which they are currently working on. In addition to that, an Ideathon was organised (described below as part of the general methodology) to include more ideas from experts outside the consortium.

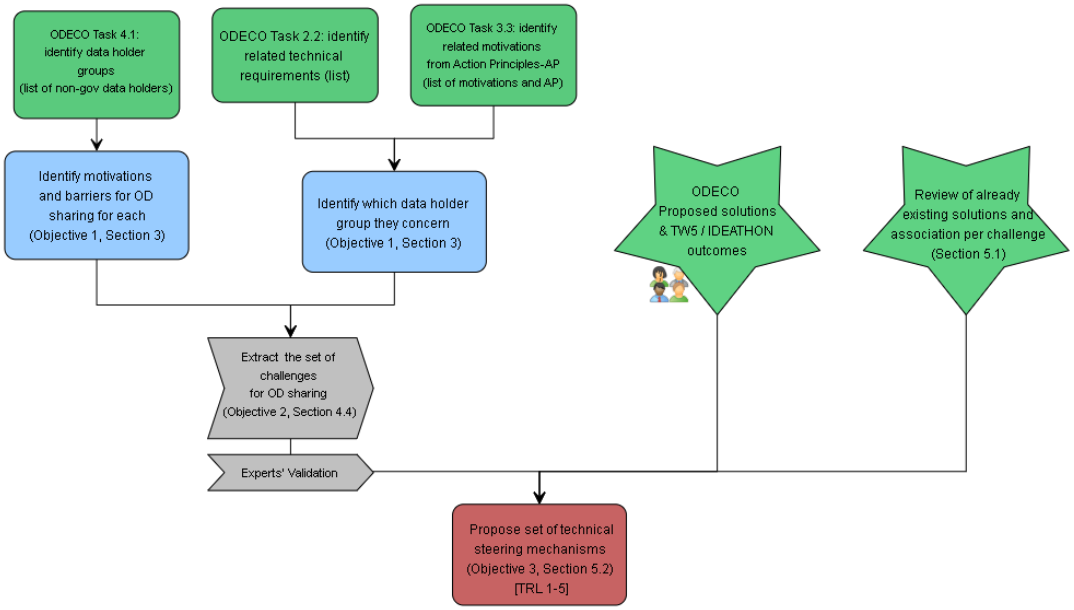

Figure 1: The methodological process of Task 4.2

More specifically, the work was carried as follows: First, the identified non-governmental data holder groups from ODECO Task 4.1 are integrated, along with the respective motivations and barriers for data sharing for each of them. In addition, Task 3.3 results on the Action Principles of the identified target groups of stakeholders have been further analysed, while Task 2.2 set the ground for the extraction of related technical requirements organised according to the FAIR principles (Findable-Accessible-Interoperable-Reusable) (go-fair.org, 2024) and matched to corresponding concerned data holder groups. In order to make this deliverable clearer and more readable as well as self-contained, we present the necessary results in sections of 3 and 4. Based on these three sources of information, a set of challenges with respect to OD data sharing concerning the non-governmental data holder groups was extracted out of the results synthesis, accompanied afterwards by proposed corresponding solutions (technical steering mechanisms) for each challenge group.

Regarding the synthesis to extract the challenges for OD sharing, the methodological process for this step included the high-level grouping of the related technical requirements, barriers, and motivations. This task entails (i) the development of the first version of the categorisation (groups)

created by the authors and (ii) the validation of the created groups by the ODECO experts resulted into the same 5 groups.
The validation process included 5 external open data experts (2 from the European Commission, 1 from a Greek technology facilitator company- including open data, 1 from a major research centre in Greece, 1 from the United Nations). The experts took part in the validation process by proposing changes and commenting on the initially proposed challenges by the ODECO group. The internal experts' group from the ODECO project included 5 ESRs and 5 supervisors (professors). The validation process took place both online (internal validation during the initial phases of the deliverable development and online meetings) and in person (external and internal experts feedback during the ODECO TW5). The final synthesis resulted into the proposed challenges described in Chapter 5. This synthesis was conducted by the core author team of this Deliverable (D4.2).

It needs to be noted that every proposed solution has its own methodological approach and research plan in order to be developed. This deliverable serves as a report of the main challenges identified and of the main solutions addressing them (existing, under development, and proposed).

The proposed corresponding technical mechanisms, apart from the individual research within ODECO (studied from different research field perspectives, e.g., technical and semantic interoperability, user interface design, data portal feedback mechanisms, etc.), were reinforced by the results of an Open Data Ideathon which took place during ODECO Training Week 5 (TW5 was organised and hosted by the University of the Aegean), where the set of challenges was provided to all ESRs and other participants and students attending the event (as ODECO TW5 was organised alongside the 11th International Summer School on Digital Government, organised by UAEGEAN). More information on the Ideathon Structure can be found in Appendix A. The technical steering mechanisms presented in this Deliverable are of varying Technology Readiness (TRL) levels, those developed within the ODECO project being of higher TRL.

3. Barriers and Motivations for Data Sharing for Non-Governmental Data Holder Groups

Governmental actors have historically been the dominant providers of open data. In many cases, such responsibility is engraved in policy documents, from non-binding strategic plans to binding laws and regulations such as the EU Open Data Directive. On the other hand, while there have been some initiatives by non-governmental actors to publish their data as open data, it is still a limited practice and remains largely desired (van Loenen et al., 2018). Non-governmental actors are actors that are outside the public/governmental sector such as companies, civil society organisations, and the media. Exploring the motivations for non-governmental actors to contribute open data is thus essential to develop sustainable open data ecosystems that incorporate both government and non-government open data.

Motivations and barriers were identified for the following non-government stakeholders: **non specialist users, data journalists, students, NGOs, commercial organisations, and open data intermediaries**. The data presented in this section has been gathered from previous research of the ODECO project, namely Task 4.1 "Motivations of non-government actors to become active contributors to the open data ecosystem". The rest of this Chapter refers to the results of ODECO Task 4.1, identifying the barriers and motivations for each data holder group to share their data as open data.

3.1 Non-Specialist Citizens

In the context of open data ecosystems, non-specialist citizens are individuals who are interested in accessing open data, or can benefit from open data, while lacking the specialized skills needed to analyse datasets. Barriers and motivations for non-specialist citizens to share data are shown in Table 1 and Table 2 respectively.

Table 1: Barriers to non-expert citizens' contribution to OD ecosystems

Technicality of the prototypes (high vs low fidelity)	The production of high-fidelity prototypes makes it harder for non-specialist citizens (problem owners) to share their knowledge of the issue.
Collaboration tools	Hard to manage and integrate multiple contributions into the same digital prototype.
Availability of participants	Non-expert citizens may not have enough time and resources to attend open data events.

Table 2: Motivations for non-expert citizens to contribute to OD ecosystems

Learning digital skills	Learning new digital skills and the basics of coding through a beginner friendly approach and tools.
Enjoyment	Enjoyment of the creative process.

3.2 Data Journalists

Journalism plays a pivotal role in keeping citizens informed, updated, and interested in the events and developments of their community. As our societies become increasingly reliant on data, journalism must evolve and accommodate this momentous shift. The next big step for journalism goes hand in hand with the open data movement. Almost at the same time when the open data movement were popularised by the US with the launch of the US open data portal (data.gov), the Guardian quoted the term data journalism (Rogers, 2008). The most important difference with previous forms of data journalism was that now the journalists were not focused on the collection of data but on their analysis. Since then, data journalism has been adopted by other media

organisations and although there is a lot of interest for the domain in academia and the industry, the use of open data and the motivations of the journalist to contribute to the open data ecosystem are not explored. However, as journalists start to recognise the increasing value of open data and realise their potential to enhance transparency, accountability, and public engagement, they can gradually take a more active role in the ecosystem. The identified barriers and motivations for data journalists to become active contributors to the open data ecosystem are the following:

Barriers

- **Open Data are Not the Only Source of Information** Journalists are not using only open data. Although they use data from official sources, these are not only the data published on open data portals but often include data acquired by requesting them from other European or governmental agencies.
- **Lack of Skills to Use and Analyse Open Data**
  Journalists do not possess the skills required to find, analyse, and use open data themselves. In all cases, the introduction of experts from other fields is required (data analysts, visual artists). Although this is an easily bypassed barrier, it creates other problems; it increases the complexity of using open data as more people must collaborate and coordinate and increases the cost of the published articles as management has to add more people to the payroll.
- **Lack of Interest**
  A lack of interest in open data has been observed. Journalists in the newsroom are not interested in getting involved in activities related to open data.
- **Limited Time**
  Journalists' main activity is to present the news, but as analysing data and compiling comprehensive infographics is a time-consuming process, it is a frequent phenomenon that when an article with results extracted from data analysis is prepared, other news is more relevant to the public, and therefore the impact of the article is reduced.
- **Not Willing to Share Their Data**
  Journalists are not that keen to share the datasets they compile as they considered them to be an asset for the media organization. The main reasoning behind this stance is their concern that the datasets, which they have compiled through extensive research and effort, could be utilized by their competitors, and sharing them without compensation would eliminate the strategic advantage that their organization holds.

Motivations

- **Transparency and Accountability**
  The main reason that journalists must get involved and contribute to the open data ecosystem is to promote a culture of transparency and accountability in society. Using open data in journalistic activities displays social issues and exposes the deep roots of the problems with the use of infographics and data. By communicating complex data to a wider audience, journalists cultivate a more informed and engaged citizenry. Their contribution to the open data ecosystem empowers citizens to advocate for transparency and better governance.
- **Enhanced Credibility**
  Another reason that journalists want to use open data in their work is the boost to their credibility. By supporting their opinions with verifiable data, journalists can transform their articles from mere opinion pieces into well-substantiated analyses, thereby enhancing their trustworthiness and authority in the eyes of the public. This requires journalists to include references to their data sets and highlight their methodology of analysis so that their work is reproducible by the audience.

3.3 Elementary School Students

Students can be defined as individuals actively engaged in a learning process in formal or hybrid (formal/informal) educational environments, ranging from basic to higher education. In the Open Data context, they have been seen as part of the large percentage of citizens without technical backgrounds, often referred to as non-specialists, non-data experts or lay audiences (Boyles, 2020; Concilio & Mulder, 2018). Especially young students in basic school education have been revealed as significant actors in Open Data (OD) and Data Literacy initiatives (Celis Vargas et al., 2023). Building a larger OD-literate community is essential for fostering citizens able to participate and benefit of OD. Although the OD field has recognized students as a strategic user group to promote the skills and competencies necessary for increasing citizen's participation and ensuring the long-term sustainability of OD ecosystems, they have been participating as users of OD rather than active contributors in OD ecosystems.

Research of Pellegrino & Antelmi (2023) has shown that OD initiatives in school level primarily are focused on the use of open datasets or data exploitation rather than on their production. Although in few learning activities elementary school students not just use open government data but also create their own data, their data is not currently opened or shared outside the classroom. The identified students' motivations and barriers behind potentially sharing their data in OD ecosystems are the following.

Barriers

- **Lack of technical skills from teachers and significant training.** Teachers have an essential role in educational designs. Considering different pedagogical approaches, teachers lead or facilitate learning activities and propose the main tools and resources. Several studies have pointed out their lack of technical skills for managing data and digital skills, as a primary barrier for achieving the potential of OD as an educational resource.
- **Updating classroom technology.** Depending on the specific context, the change of tools, platforms, and methods for adapting classrooms to the fast-changing technology, could be at the same pace. Nevertheless, investment, skills and administration are factors to consider. The most traditional educational systems are characterized by slow adaptation and low insertion of technology.
- **The concept of open data being highly abstract.** Several authors have stressed the challenge that understanding and using OD presents for students due to its high level of abstraction (Atenas et al., 2015; Coughlan, 2020; Saddiqa et al., 2021a). For example, Saddiqa et al. (2021a) and Wolff et al. (2016) have suggested contextualizing the data for better understanding, using OD from student's own municipalities. Furthermore, to overcome this barrier, the need of customized hands-on open data collection, interpretation and exploitation tools and methods has been made explicit. However, the development of tools and methods at the same entangles new challenges for the usually steady educational systems.
- **Low awareness about what OD is.** Students and teachers refer to Open Data as any kind of information found on the internet. For example, teachers claimed to use OD for their teaching, but when asking more in depth their sources and process for managing the OD, it was explicit that they understand OD as any available information on the internet. On the other hand, it is a completely new term for the pupils.
- **Risk of disclosing personal data from pupils.** In the user context of elementary school students' ethical management of data is essential, since children are usually a vulnerable group. In connection to low awareness on data management, school, teachers, and parents are at risk of violating GDPR regulations when sharing open data.

Motivations

- **Being active citizens.** Celis Vargas et al. (2023) have identified that specially OD learning activities seeking the development of competencies for active citizenship address activities for collection of own data. In those cases, students have been involved in both creating simple spreadsheets and collecting more complex data by using tools such as sensors, games, or mobile applications (Badioze Zaman et al., 2021; Chicaiza et al., 2017; Saddiqa, Larsen, et al., 2019; Vallejo-Figueroa et al., 2018). The motivation is actively participating as citizens to create a better world. For example, students during the focus interviews wondered about their school projects "How is this going to create a better world?".
- **Raising awareness of local issues around students' context and daily life.** Considering students as experts of their own local experience, they can potentially create and share local datasets addressing aspects of their environment and daily life experiences. From their perspective, they want to raise awareness and provide contextual understanding of local issues. Their motivation behind is raising their voice and being heard. Making "Children's voice as important as others".
- **Helping the community around the school, considering students as important actors in local ecosystems.** Students are motivated by helping the community around them, implicitly for a feeling of belonging, building their identity and their place. Creating and sharing data has been identified as an opportunity to also building networks in their local communities by addressing problems from other actors and contributing to solve them with data.
- **Seeing what students learn in schools as useful in the real world.** For young pupils in elementary school, it was relevant to see what they do in school being used in the real world. It increases the authenticity of their learning experience.
- **Making school activities more relevant, interesting, and fun.** To sum up, creating and sharing their data might increases their motivation for learning by fulfilling their intrinsic motivation for making something relevant, being heard and connecting to their communities. Overall, students are motivated by active learning experiences, were they have freedom to experiment and learn by themselves.

3.4 Non-Governmental Organisations

Non-Governmental Organisations (NGOs), also interchangeably called Non-Profit Organisations (NPOs) in this section, take up an intermediary role in the open data ecosystem, where they bridge the gap between open data providers and users (González-Zapata and Heeks, 2015). NPOs are unique as intermediaries because there are specific user communities they are focusing on to address a social issue (Enaholo, 2017) while also not seeking to gain any profits from it (Salamon and Anheier, 1992). As such, NGOs can be seen as groups of individual users that are with one more level of explainability and technical assistance in the use of available tools. The difference with non-specialist citizens is that the services and guidelines could be drafted in a more technical format. If the aim of NPO is to improve overall openness and transparency, it can push them to aim for a variety of projects and have open data and open source on that principle (Baack, 2015). However, NGOs may face barriers that prevent them from contributing (Chattapadhyay, 2014). The identified barriers and motivations are presented in Table 3 and Table 4 respectively.

The barriers found are summarised in Table 3 ; there are two barriers to the motivations that stop NGOs from sharing the data as open: (1) the lack of financial resources, limits NGOs ability to pursue open data projects, (2) the lack of knowledge and technical skills of NGOs' employees, affecting the NGOs capacity to correctly license and publish dataset, and (3) the lack of an existing common portal for NGOs to share their data.

Table 3: Identified barriers

Barriers	Description
The lack of financial resources	The nature of non-profit organisational model means that NGOs need to secure funding in a form of grants or individual donations. Some NGOs may not be able to obtain enough funding for their open data projects.
The lack of knowledge and technical skills of NGOs' employees	The lack of knowledge and technical skills of NGOs' employees prevents them from publishing the data they have as open, or they do not know how to provide it properly licensed as open. NGOs that are short on their financial resources are unable to hire additional employees with the required skills.
The lack of an existing common portal for NGOs to share their data	NGOs cannot add to the open government data portals and there is no existing common portal for NGOs and civil societies. However, creating such a platform and/or maintaining it would require a lot of financial and human resources that the average NGO cannot afford.

Table 4: Motivations for NPOs to contribute open data

Motivation	Description
Show the power of open data	Creating a project with enhanced or reused open data can engage other stakeholders or make them more aware of the open data available.
Help with the local and global societal issues	There is open government data that might be available on societal issues, but not utilised. By finding ways to enhance the data and publish it as open, the NPO can highlight it and help those affected by the issue.
Receive positive feedback from the community	The community can give feedback on the issues with the data and give suggestions when they are interested and involved which motivates NGOs to continue the project.
Follow organisational goals and the personal beliefs of the employees in openness and transparency	Some NPOs have transparency and openness of data and knowledge as their main goals and employees join NPOs because they share the vision.
Create opportunities for other stakeholders	NGOs want their data to be reused widely by and benefit a variety of communities, NGOs, private and governmental organisations.

3.5 Commercial Organisations

Commercial organisations are defined as those whose goal is to make an economic profit. This is illustrated in contrast to users who intend non-commercial use, which Creative Commons (Creative Commons, 2023) defines as "means not primarily intended for or directed towards commercial advantage or monetary compensation." Commercial organisations may have different motivations to contribute to OD ecosystems, depending on their field and own necessities. The identified barriers and motivations for commercial organisations to share data are presented in Table 5 and Table 6 respectively.

Table 5: Barriers of commercial organisations to contribute open data

Barrier	Description
Technical/ Tools	Lack of data integration tools, and tools for importing large datasets.
License compatibility	Datasets generated by commercial organisations are sometimes a combined product of different sources, for which the license may not be compatible.
Insufficient resources [SMEs]	Limited money, labour and time, compared to large organisations.
Resistance by other community members [Big corporations]	Resistance by community members as they fear big corporations may dominate the project landscape.

Table 6: Motivations for commercial organisations to contribute open data

Motivation	Description
Improving the data quality	By improving the data quality organisations can deliver better services to their customers.
Community building	Some organisations are interested in each topic or domain. They kickstart communities in that topic or domain to maintain data continuity and freshness.
Standardization and alignment with an open project schema	For some organisations, this is a motivation to contribute data to open projects, to lower the subjectivity of the data.
Being grateful and wanting to give back	Some organisations feel that open projects have provided value to them and their products, and feel it is only fair to give back and provide value to the project, to keep it going.

3.6 Open Data Intermediaries

Open data intermediaries are defined as "third-party actors who provide specialised resources and capabilities to (i) enhance the supply, flow, and/or use of open data and/or (ii) strengthen the relationships among various open data stakeholders" (Shaharudin et al., 2023). Examples are developers who process and include open data in apps/software, crowdsourcing platforms that gather and publish data as open data, and organisations that transform open data into easily digestible information such as visual forms. Open Data Intermediaries carry out a wide range of tasks depending on their specialised resources and capabilities such as compiling data, validating data, and improving technical openness of data (Shaharudin et al., 2023). Thus, through some of their tasks, some open data intermediaries (re-)produce data that could be contributed back to the open data ecosystem.

As all actors in the open data ecosystem, including open data intermediaries, have their own agency, and thus self-interested (Davies, 2011; Poikola et al., 2010), the motivations for them to contribute open data naturally have to be aligned with their interests. In other words, open data intermediaries either must be convinced that they can directly or indirectly capture value for themselves by publishing open data (intrinsic motivation) or they must be forced by external conditions such as through law and regulations to do so (extrinsic motivation). The identified barriers and motivations for open data intermediaries are presented in Table 7 and Table 8 respectively.

Table 7: Barriers for open data intermediaries to contribute open data

Barriers	Description
Protecting business interests	Some open data intermediaries are hesitant to release (some of) their data as open data because they want to protect their business interests.
No compelling legislation	There are limited or no legal requirements that compel open data intermediaries to provide open data.
Additional costs to develop and maintain open data platforms	Developing and maintaining data platforms to release open data would incur additional costs to open data intermediaries.

Table 8: Motivations for open data intermediaries to contribute open data

Motivation	Description
Support the visibility of their organizations	By providing open data, open data intermediaries can increase the visibility of their organization; hence, it is a form of marketing for their products and services.
Support other partners within their networks	Some open data intermediaries provide open data to support the business or operation of their partners, which they would also get the benefit from. This may further strengthen their position within their network.
The desire to contribute to society	Some open data intermediaries are driven by philanthropic or altruistic motivations to contribute open data that could benefit society.
The availability of open data platforms to share open data	Open data intermediaries could be more motivated to release open data if there are open data platforms that could facilitate them to do so. This is because not all open data intermediaries have the capability and resources to develop and maintain their own data platforms.

4. Technical Requirements for Non-Government Data Holder Groups

For many non-government actors studied, motivations related to own benefit were identified in D4.1. Such motivations include supporting other partners within an organization network for their own benefit, the private value of contributing, the feeling of belonging, and enjoyment.
Aligning these own benefit motivations of the different stakeholders is key to achieving the goal of open data sharing. These own benefits match with the intrinsic motivations from the Self Determination Theory defined in Task 3.3 (D3.3). Own benefit motivations are supported by the desire to create an impact, mentioned in the clusters of supporting other partners, supporting the community, and creating social impact. These can be classified as extrinsic motivations in the Self Determination Theory.

Finally, the availability of and desire to improve the technical skills and solutions are mentioned as motivations, and potential motivations. This means that creating the correct technical environment can be an enabler for open data sharing, as well as including the potential stakeholders in the process, with the desire to improve the technical environment.

4.1 Open Data Action Principles

Based on the findings of ODECO Task 3.3, where the commons-based governance model for the open data ecosystem has been refined, six action principles have been unveiled related to boundary-making, community support, participation and decision-making, legal mechanisms, interoperability and interoperable projects, and sustainability. In our analysis, we have focused on the following principles which are the most relevant to the non-governmental data holders:

1. Encouraging participation and shared decision-making,
2. Considering appropriate legal mechanisms,
3. Designing an ecology of interoperable projects, and
4. Ensuring sustainability of open data ecosystems.

The aim is to extract the main key points and propose the development and develop technical solutions that can motivate each group to be an active participant in the open data ecosystem.

Encouraging participation and shared decision-making

The key point of this action principle is the opportunity for less powerful users of the open data ecosystem (particularly the ones that in a hierarchical system have no control over the open data) to raise concerns, and needs, and provide feedback. It is important to mention that for public and non-profit organizations, technical solutions may motivate them to give voice to citizens and underrepresented groups, but for commercial users (intermediaries, companies, etc.), they must have profit as an incentive or other extrinsic motivations (laws, regulations, etc.).

Journalists, NGOs, local governments, and open data intermediaries can all benefit from collaborative platforms that enable public contribution and feedback on open data. For journalists, such platforms offer valuable crowdsourced insights, though robust moderation is crucial to maintain data integrity. NGOs require simplified, low-maintenance feedback mechanisms to efficiently gather input without straining resources. Local governments can leverage citizen feedback to improve services and decision-making. Open data intermediaries play a key role by providing technical solutions for these feedback systems, bridging the gap between data providers and users. While students may not directly implement technical solutions, they can draw inspiration from these key points to understand the importance of public engagement in

open data ecosystems. Ultimately, a well-designed feedback mechanism can enhance data quality, increase transparency, and foster greater civic participation across various sectors.

Considering appropriate legal mechanisms

Two integral issues were uncovered in this legal mechanism action principle. The predominant one revolves around the lack of data literacy and the impact that legal mechanisms can have, especially in government structures. The second impediment is the discovery and adoption of sustainable business models for intermediaries and commercial users. This obstacle involves finding economically viable ways for these entities to operate while adhering to open data principles and legal requirements.

The primary challenge faced by all users is the requirement for compatible licenses. A potential solution to this issue could be the development of an ontological model for open data licensing. This model could be applied to open data sets, thereby automating the process of checking compatibility between different licenses.

Designing an ecology of interoperable projects

The integral focus of this action principle is the adoption of interoperable standards between datasets so they can be exchanged, combined, and analysed even when they are sourced by different organizations. This initiative can boost the use of open data across various user types as it will significantly reduce the technical skills and expertise currently required for users to effectively utilize the open data they can acquire and use. The most crucial problems mentioned in this action point in D3.3 are, as already noted, the lack of personnel with the appropriate skill set and knowledge to utilize open data. Secondly, there are concerns among users regarding potential security and privacy issues that may arise from combining different datasets, which could indirectly expose data that was obscured through anonymization techniques in the original datasets.

Ensuring sustainability of open data ecosystems

The primary emphasis of this action point is to ensure the sustainability of the open data ecosystem and the continued participation of users within it. This mainly involves the struggle for funds and resources, particularly for non-profit organizations. The focus is on advocacy for public funds, while for commercial users, the emphasis is on the viability of business models that centre on active participation in the open data ecosystem.

4.2 Technical Requirements for OD Holders and Users

In this Section, the technical requirements for Open Data holders were mainly derived from ODECO Task 2.2 (D.2.2) and were mapped afterwards with the data holder group they concern. Analysing the requirements regarding their importance (number of NGD user group reporting this requirement) and connection to the NGD holders (authors' interpretation), the most relevant requirements derived from D2.2 are those related to the Reusability and Interoperability aspects. They are mapped regarding their respective stakeholder group shown in Table 9. Since, NGD holders cannot share data through the open government data portals, we specifically mention the already existing open data spaces (e.g. Zenodo, OpenAIRE, wikidata etc.) that enable the upload of datasets from any non-governmental stakeholder. In this case, the available tools and registries for NGD holders should comply to the FAIR principles.

As already mentioned, the developed and proposed technical mechanisms should not disregard the requirements of the data users in order to build "user-driven" solutions. Furthermore, to achieve circularity in the open data ecosystem, we need to consider the dual role of the non-governmental stakeholders of the ecosystem serving both as users and providers of open data. This means that specific open data user requirements are also important for the development of technical mechanism to stimulate NGD Holders to provide their data as open data. For instance, one requirement from the user perspective is something that needs to be resolved by the provider

or data holder. In this case, we use these requirements to identify the challenges that occurred for the NGD data holders.

Table 9 - Technical Requirements for NGO Open Data users (taken from ODECO deliverable D2.2)

Requirements	Non-specialist actors	Data journalists	Elementary school students	Non-governmental organisations	Commercial organisations	Open data intermediaries
R1 - Facilities for data processing and for checking the quality of data		X	X	X	X	X
R2 - Datasets should have schema.org tags to provide a semantic meaning				X	X	X
R3 - Visualize a dataset on the data portal		X	X	X	X	X
R4 - Data provenance		X		X	X	
R5 - Technical standards for data formats, metadata, APIs		X		X	X	X
R6 - Easily publish data in the required formats	X	X	X	X	X	X
R7 - Engage with data holders to understand their technical needs and challenges	X	X	X	X	X	X
R8 - Feedback mechanisms for data holders to suggest improvements	X	X	X	X	X	X
R9 - Reducing the gap between NGO user groups and their technological needs can help identify the common requirements of all stakeholders	X	X	X	X	X	X
R10 - High quality data provision	X	X	X	X	X	X
R11 - The development of data integrators to enable the technical interoperability of open data stemming from different domains		X	X	X	X	X
R12 - Widely adoptable data formats	X	X	X	X	X	X
R13 - Metadata suitability checker for data publishers	X	X	X	X	X	X
R14 - Quality enhancement tools (data augmentation tools) should be available						X
R15 - Users should be allowed to upload updated versions of existing datasets	X	X	X	X	X	X
R16 - Consistent and reliable delivery of datasets should be provided to the artificial users					X	

These requirements related mainly to reusability and interoperability are described in more detail as follows:

User needs for Reusability for an Inclusive Open Data Ecosystem from a technical perspective

- R1: **Facilities for data processing and for checking the quality of data:** Currently, portal administrators and data users must have the basic skills for data processing and for checking the quality of data. For open data reusability, data quality, validity and completeness of the data may be evaluated with the use of technologies available via the data portal. This requirement addresses the services that need to be developed to reduce the data processing capabilities of the portal administrators.
- R2: **Datasets should have schema.org tags to provide a semantic meaning:** All the datasets available on the open data portal, even third-party datasets, should have schema.org tags to provide a semantic meaning because with this requirement we can prioritize in searches the comments that can be more relevant to their profile. It can also help in reusing the dataset.
- R3: **Visualize a dataset on the data portal:** If a user wants to visualize a dataset on the data portal, then the data portal must have visualisation tools and resources. For better insights of the dataset, it will be better to visualize it in the portal before downloading and processing it for further tasks.
- R4: **Data provenance:** Data provenance with respect to the reusability of open data within open data portals ensures data authenticity and quality. This involves clearly documenting the dataset's origin, ownership, and updates. Metadata and reuse standards like open data licenses promote data sharing and accessibility.
- R5: **Technical standards for data formats, metadata, APIs:** Implementing CSV, JSON, and XML data formats provides platform portability. Searching for data and context comprehension improve significantly by using Dublin Core or schema.org metadata standards. Strong APIs like REST or GraphQL provide automated data access, integration, and scalability while retaining security and control. Data is readily accessible, interpreted, and used by varied people and systems because of these standards.
- R6: **Easily publish data in the required formats:** An Open Data Ecosystem requires data to be publishable in the right formats. Data vendors may alter and disseminate their data in CSV, JSON, and XML using straightforward tools and services that simplify data transformation and publication. This streamlines the process, lowers technical barriers, and ensures data format compliance, improving end-user accessibility and usability.
- R7: **Engage with data holders to understand their technical needs and challenges:** Engaging data holders to understand their technological requirements and difficulties is essential for an inclusive Open Data Ecosystem. Conducting surveys, seminars, and feedback sessions helps data providers determine their needs and challenges. This knowledge enables the creation of customized assistance, training, and technological solutions to help data holders publish and distribute high-quality data.
- R8: **Feedback mechanisms for data holders to suggest improvements:** Continuously improving the Open Data Ecosystem requires data holders to provide input. Discussion forums, feedback forms, and writing to data publishers enable data suppliers to share their knowledge. Feedback is crucial for improving data standards and tools, ensuring the open data ecosystem meets users demands, and identifying areas for development.

User needs for Technical Interoperability for an Inclusive Open Data Ecosystem from a Technical Perspective

Technical requirements outlined in Table 10 from ODECO Deliverable 2.2, relevant to the open data users from a technical standpoint, are presented. Some parts of these requirements are listed below with their explanation, related to the non-government stakeholders:

- R9: Reducing the gap between NGO **user groups and their technological needs** can help identify the common requirements of all stakeholders. A one-size-fits-all technological approach would be more beneficial for stakeholders from government or non-government sectors. The solutions should be ready to enhance both types of platforms (national or governmental data portals but also other open data portals like Zenodo and OpenAire).
- R10: **High quality data provision** – From the government to the non-government stakeholders, and vice versa. The creation of open data has three different perspectives: (a) the actual creation of data (e.g. from IoT sensors), (b) the creation of DCAT-complied datasets (flat and contextual metadata levels) and (c) the linked open datasets (detailed metadata level) creation. Solutions should target all of the above.
- R11: The development of **data integrators to enable the technical interoperability** of open data stemming from different domains. In this stream, non-government stakeholders can integrate their data in a better way. The proposed solutions should target easier (automated if possible) creation of data and metadata from their systems.
- R12: **Widely adoptable data formats** – deciding upon uniform adoptable data formats can resolve several problems in the current open data ecosystems and can make open data interoperability much easier.
- R13: **Metadata suitability checker for data publishers –** metadata selection is a very important phase of the open data publication. For non-governmental organizations, there is also a need for a metadata suitability checker before adding the data to main streams of open data portals (National or governmental open data portals).
- R14: **Quality enhancement tools (data augmentation tools) should be available –** With the rise of AI and LLMs, data augmentation tools can enhance the quality of open data. Leveraging AI to achieve semantic and syntactic interoperability is highly sought after in the market. An example of this scenario is AI for data and data for AI.
- R15: **Users should be allowed to upload updated versions of existing datasets.** It has been noted that users currently cannot upload their improved data back to the portal, whether the improvements pertain to data quality, metadata, or applications.
- R16: **Consistent and reliable delivery of datasets should be provided to the artificial users –** Is there a way to bridge the gap between LLMs and open data portals? By establishing a technical interoperability channel, LLMs can effectively utilize data from open data portals.

4.3 Synthesis

This subsection includes the summary of the technical requirements (D2.2), action principles (D3.3), barriers, and motivations (D4.1) for non-governmental data holder groups to share their data as open data, grouping the various elements according to an experts' grouping as shown in Table 10. As explained in the methodology Section, the synthesis of the Challenges was conducted by the core authors team of this Deliverable. The Challenges and their detailed description are presented in the next Chapter (Chapter 5).

Table 10: Technical requirements, action principles, barriers, and motivations per Open Data Holder Group

Requirements + Action Principles /Non-Governmental Data Holder Group	Challenge Group	Non-Specialist Citizens	Data Journalists	Non-Governmental Organisa-tions	Schools	Commercial Organisa-tions	Open Data Intermediaries	Importance count
Easily publish data in the required formats	1	X	X	X	X	X	X	6
The interoperability needs of a government body are different from the interoperability needs of a company (fragmented open data ecosystems)	1	X	X	X	X	X	X	6
Useful data discovery tools (e.g., Contextualized discovery)	1	X	X	X	X	X	X	6
High quality data provision	1	X	X	X	X	X	X	6
The ODP should provide technical assistance for the contributor to prepare the data resource to be added to the portal (data preparation, cleaning, labelling)	1	X	X	X	X	X	X	6
Widely adoptable data formats	1	X	X	X	X	X	X	6
The ODP should help the data contributor discover useful connections from the data resources it contributes to other resources already present in the portal	1	X	X	X	X	X	X	6
Basic skills for data processing and for checking the quality of data.	1		X	X	X	X	X	5
The development of data integrators to enable the technical interoperability of open data stemming from different domains	1		X	X	X	X	X	5
Technical standards for data formats, metadata, APIs.	1		X	X		X	X	4
Datasets should have schema.org tags to provide a semantic meaning.	1			X		X	X	3
Addressing technical data challenges	1	X	X					2
Face challenges with data quality, timeliness, and resources	1		X					1
Benefit from seamless data exchange	1					X		1
Face challenges adopting new interoperable formats	1						X	1

Requirements + Action Principles /Non-Governmental Data Holder Group	Challenge Group	Non-Specialist Citizens	Data Journalists	Non-Governmental Organisations	Schools	Commercial Organisations	Open Data Intermediaries	Importance count
Need to provide data in diverse formats	1						X	1
Engage with data holders to understand their technical needs and challenges.	2		X	X	X	X	X	6
Feedback mechanisms for data holders to suggest improvements	2		X	X	X	X	X	6
Reducing gap between the user groups and their technological user needs	2		X	X	X	X	X	6
The ODP must notify the user community about new data resources added by NGD contributors	2		X	X	X	X	X	6
The ODP should allow users to provide feedback (ratings, reparation…) on the resources provided by contributors	2		X	X	X	X	X	6
The ODP should track usage statistics of data resources provided by contributors	2		X	X	X	X	X	6
Demands for data Hubs (Centralized hub and to consolidate more reliable and necessary data sources in one place	2		X	X	X			4
Require input from experts to understand and analyse open data	2			X				1
Require simple tools and guidance	2				X			1
Unclear licensing can be a barrier to use/contribution.	3					X		1
Diverse stakeholder opinions on license types.	3					X		1
Open licensing may negatively impact some business models.	3						X	1
Consider providing open licenses where it does not affect business model.	3						X	1
Open data shared at hackathons should use CC licenses	3	X						1
Need to balance open licensing with protecting commercial interests of some participants	3	X						1

Requirements + Action Principles /Non-Governmental Data Holder Group	Challenge Group	Non-Specialist Citizens	Data Journalists	Non-Governmental Organisa-tions	Schools	Commercial Organisa-tions	Open Data Intermediaries	Importance count
Enabling non-governmental data incorporation into ODPs	2		X	X	X	X	X	X
The ODP must display the data input of its contributors in a public personal profile	2		X	X	X	X	X	X
Visualize a dataset on the data portal.	2			X	X	X	X	X
Require user-friendly tools and interfaces	2		X					
May face barriers due to complex implementations	2		X					

5. Challenges for OD Creation and Sharing

As already explained in the Methodology section 2, the synthesis of the main challenges for the non-governmental data holders towards supporting the creation and sharing of open data was a challenging task, analysing all the data presented in previous chapters. It was a qualitative synthesis. The first version was created by the main authors of this deliverable. The next version until the final one is a result of 3 validation workshops including internal and external open data experts. The process included 5 external open data experts (2 from the European Commission, 1 from a Greek technology facilitator company- including open data, 1 from a major research centre in Greece, 1 from the United Nations). The experts took part in the validation process by proposing changes and commenting on the initially proposed challenges by the ODECO group. The internal experts' group from the ODECO project included 5 supervisors and 5 ESRs. The workshops have been facilitated in both virtual and in-person modes.

Considering the requirements, action principles, barriers, and motivations for non-governmental Open Data Holder Groups presented in the previous sub-sections, the synthesis of information resulted in three challenges. The challenges are targeting the creation of a more inclusive and user-driven open data ecosystem. The data sharing process (from data creation to data publication) for non-governmental Open Data Holder Groups is put in the centre of this concept.

It needs to be highlighted that there is already a variety of existing tools and data repositories that could be used as platforms for data sharing from NGD holders. So, in this deliverable we focus on enhancing the capabilities of these platforms and tools without proposing yet another open data platform for data sharing. The proposed challenges are focussed on the enhancement of the data creation and data sharing processes for the NGD holders. The proposed solutions are meant to be integrated to the existing tools and available repositories. Furthermore, the proposed solutions are targeting the automisation of the creation and sharing processes, as well as the integration of them to new OD platforms interfaces in order to minimise the skills gaps of NGD holders.

5.1 Enhanced data creation process through better quality, integration and validation tools

Open data platforms face significant technical and interoperability challenges, requiring data from different sources to be harmonized and integrated. Datasets need to be standardized and adopt widely recognized formats to ensure seamless data exchange. Technical standards for data formats, metadata, and APIs, as well as the creation of validators are crucial to achieve this, but varying requirements across organizations, such as government bodies versus private companies, create additional fragmentation. Developing data integrators that enable interoperability across domains is also essential.

There are three aspects to help the creation and sharing of open data: (a) actual data creation (from individuals or IoT devices), (b) portals' metadata compliance checking and (c) detailed metadata annotation at the level of the dataset columns description. The abovementioned developments will enhance the technological capabilities of the NGD holders in order to provide more usable data to the data users. They will bridge the gap of inclusiveness allowing NGDs to create data more easily and provide them with better quality); it will bridge the gap of circularity since there will places and tools to easily create and use open data.

The challenge lies in reducing this gap, ensuring the creation of data is of better quality and validated through automated tools, as well as clean, labelled, and discoverable answering the unique technical needs of the different user groups. Overall, this challenge requires a concerted effort to align technical standards, provide necessary tools, and support the integration process, enabling users to effectively utilize diverse data resources across multiple platforms and sectors.

This challenge is asking for technical mechanisms to improve and simplify the data creation process for NGD holders incorporating solutions and tools for the creation of better-quality data (including metadata), as well as integration and validation tools and techniques complying to the requested standards. Addressing this challenge will offer robust technical assistance and resource optimization, enabling contributors to produce high-quality, interoperable data that drives greater value across the open data ecosystem.

5.2 Enhanced data sharing through better licensing and anonymisation

Licensing and legal challenges are significant barriers to the effective use and contribution of open data. Ambiguities in licensing terms can deter potential contributors and users, as unclear guidelines create uncertainty about how data can be used or shared. Diverse stakeholder opinions on licensing types further complicate the situation, making it difficult to establish a consensus on best practices.

The Open Data Platforms must address these issues by providing clear, standardized licensing options and ensuring that contributors' preferences are respected, such as the choice to remain anonymous. Additionally, when non-governmental data is incorporated into ODPs, legal considerations around data ownership, privacy, and intellectual property must be managed carefully. Displaying contributor information publicly in a transparent manner while protecting privacy rights is a delicate balance that needs to be struck.

This challenge is about developing new technical mechanisms and solutions towards the creation of anonymised datasets in an automated way, as well as, the creation of clear legal frameworks and licensing policies that are acceptable to a diverse range of stakeholders, fostering a trustworthy environment for open data exchange.

5.3 Towards inclusive, user-driven and circular open data platforms through better interfaces and feedback mechanisms

Engagement is critical for the success of open data ecosystems. Addressing the dual role of non-governmental stakeholders in the open data ecosystem will create a circular and more user-driven ecosystem maintaining active participation and fostering a collaborative environment towards open data ecosystem development and function. Supporting the creation of inclusive OD ecosystems, the Open Data Platforms must provide the necessary mechanisms that will derive from challenges 1 and 2 addressing the skills gaps of NGD holders. Furthermore, to keep data users and holders engaged closing also the loop of interaction, open data platforms should also provide feedback mechanisms, such as allowing users to rate and suggest improvements on data resources. Tracking usage statistics can help understand how data is being used and by whom, which can inform platform improvements. Additionally, regular notifications about new data resources keep the community informed and engaged. However, ensuring that these interactions are meaningful and lead to actionable insights requires a well-structured feedback system. There is also a need to create a sense of community and belonging, encouraging stakeholders to support each other and contribute to the platform's growth. Achieving this requires balancing engagement and enjoyment towards the creation of social impact.

The challenge is to design and develop the necessary technical mechanisms including new interfaces and user interaction methods that enhance the way NGD holders are creating and sharing open data that is available, usable and valuable to a wide range of stakeholders, fostering a more inclusive and effective open data ecosystem. This way the system will sustain a vibrant, collaborative community of data holders and users that feels invested in the open data success. Data holders will be more engaged with continuous communication with their users and they as users will be more engaged with other data holders in the ecosystem.

6. Technical Steering Mechanisms per Challenge

In this Section, the proposed solutions (technical mechanisms), corresponding to the three identified challenges, motivations, and requirements presented in the previous Chapters, are presented. The first sub-section 6.1 is concentrated on already existing solutions and technical mechanisms. The second sub-section is providing the new developed or under development tools and Technical Steering Mechanisms organised per Challenge, initially derived from individual research of ODECO projects, while at a second phase, more solutions are added, as a result of the outcomes of the OD Ideathon during ODECO TW5.

6.1 Existing relevant solutions

In this Section, existing solutions, relevant to the identified challenges, are reviewed and presented. The process started by exploring several data tools and reading their documentation to get an understanding for the tool's basic features and its intended use. After having a good understanding of the tools' purpose and features, such as how easily the tool collects different kinds of data, sorts it, and displays it for further study, the tool examination phase took place. First, we examine the key characteristics of tools used for the creation and sharing of open data. Next, we explore how non-governmental data is integrated into both governmental and independent repositories. Finally, we analyse the insights gained from HOT-TM, an open data creation tool that employs a crowdsourcing approach.

6.1.1 General toolkits for open data sharing

The capacity to efficiently gather, handle, use and reuse data is now fundamental in a data-driven decision-making era across several industries. Building a welcoming open data environment relies heavily on tools that make data collecting, sharing and administration easier. The data in such an ecosystem is not only easily available, but it can also be reused, which increases its value in many applications. Datasette, Epicollect5, ODK (Open Data Kit), and QField are the four main tools surveyed in this section. They all have their own set of advantages and disadvantages when it comes to data administration and gathering. A set of criteria crucial to encouraging reusability in an open data ecosystem is used to assess these technologies. Skills in data processing, schema.org tagging for semantic meaning, visualization, data provenance, data publication ease, engagement with data holders, and feedback mechanisms are all part of the criteria. Technical standards for data formats and APIs are also covered. This section seeks to offer a thorough knowledge of how each tool may contribute to a successful and inclusive open data ecosystem by mapping these needs to the functionality of Datasette, Epicollect5, ODK, and QField.

- **Datasette**

The open-source application known as Datasette is largely geared toward SQLite databases and was developed with the purpose of exploring and publishing data. For interactive data exploration, it offers a web interface that is both sophisticated and user-friendly. This interface makes it easy for users to search, filter, and sort the data in which they are interested. With the help of Datasette faceted navigation feature, users are able to look deeper into particular datasets by utilizing a wide variety of filters and characteristics. When it comes to transforming static datasets into dynamic, shareable web apps that can be simply deployed on local servers or cloud platforms, this tool is an excellent choice. Following are the key attributes of Datasette:

- **Interactive Data Exploration:** Users can perform SQL queries directly through the web interface to validate and analyse data.
- **Faceted Search:** Enables efficient data filtering and navigation.
- **Plugins and Extensibility:** Supports customization through plugins for additional functionality.
- **Export Options:** Data can be exported in JSON and CSV formats.

- **API Access:** Provides a RESTful API for programmatic data access.
  Beside the key attributes here are a few use cases:
- Data journalism for publishing interactive data stories.
- Research and academic projects for sharing data.
- Creating open data portals for public data accessibility.

- **Epicollect5**
  Epicollect5 is a versatile mobile and online application designed specifically for the efficient gathering and management of data. It is particularly suitable for tasks that include data collection in the field. It offers a straightforward approach for creating, distributing, and managing forms and surveys. This allows for collecting data offline and synchronizing it later when an internet connection is accessible. Figure 2 shows the key attributes of the epicollect5 tool.

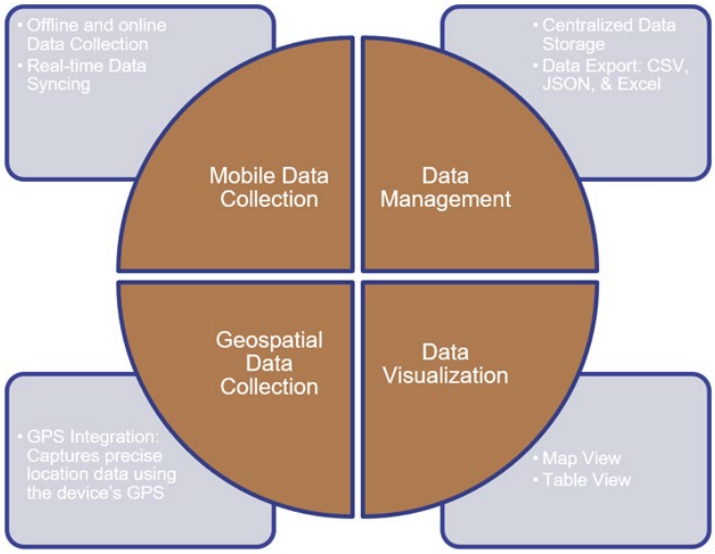

Figure 2: Main components of Epicollect5

Beside the key attributes here are a few use cases:
- Research projects requiring field data collection.
- Environmental monitoring and biodiversity surveys.

- **ODK (Open Data Kit)**
  The Open Data Kit (ODK) is a set of tools designed specifically for the collection, management, and analysis of data. The main focus of the ODK is to gather data under challenging circumstances. It is a robust option for complete data management requirements since it covers the whole data life cycle, beginning with the design of the form and continuing through the collecting and analysis of data.
  Here are the key attributes of ODK:
  - **Comprehensive Suite:** Includes ODK Collect (mobile app), ODK Build (form designer), ODK Aggregate (server component), and ODK Central (modern server).
  - **Offline Data Collection:** Supports offline data collection and synchronising.
  - **XLSForm Support:** Allows creating forms in Excel with complex logic.
  - **Multimedia Integration:** Captures photos, videos, audio, and GPS coordinates.
  - **Data Validation:** Supports validation rules and constraints in forms.
  - **API Access:** ODK Central provides a RESTful API for data access and integration.
  - **Data Export:** Supports CSV, JSON, and XML formats.
  - **Data Provenance:** Tracks data origin, ownership, and updates.
    Beside the key attributes here are a few use cases:
    - Humanitarian aid assessments and monitoring.
    - Environmental and agricultural studies.

- **QField**

  QField is a mobile application that is meant to integrate perfectly with QGIS, which is a popular desktop GIS software program. Its purpose is to collect geographical data. In addition to providing a variety of geographic data gathering and editing operations directly on mobile devices, it enables users to take their QGIS projects out into the field that they have created. The following Figure 3 highlights the key attributes of QField:

Figure 3: Key attributes of QField.

Beside the key attributes here are a few use cases:

- Urban planning and infrastructure mapping.
- Agricultural data collection and farm management.

Below, Table 11 compares the four data collecting and publishing tools (Datasette, Epicollect5, ODK, and QField) to the specific user requirements mentioned earlier. Table 11 also presents a technical assessment of each tool in correspondence to the Challenge it addresses:

Table 11: Evaluation of Open Data Tools for the Key Challenges

Challenge	Datasette	Epicollect5	ODK	QField
Challenge 1: Enhanced data creation process through better quality, integration and validation tools	Supports basic SQL queries for data validation and analysis	Basic data validation through form design and constraints	Supports data validation rules and constraints in forms. Supports exports to CSV, JSON, XML	Relies on QGIS for data validation, some in-app validation tools
Challenge 2: Enhanced data sharing through better licensing and anonymisation	Metadata fields for licensing, requires custom anonymization	No built-in support, export for external anonymization	Metadata support, potential for anonymization through form design	Managed through QGIS metadata and external anonymization
Challenge 3: Towards inclusive, user-driven and circular open data platforms through better interfaces and feedback mechanisms	GitHub issues, community forums, Built-in visualization tools	Community feedback through app, Basic visualization for data collection	Feedback via community channels, surveys, Limited in-app visualization, better with external tools	GitHub, QGIS forums for feedback, Limited in-app, relies on QGIS for advanced visualization

6.1.2 Integration of non-government data on open data repositories

Open Data portals are online platforms that enable the discovery, access, and management of metadata records for datasets (Kubler et al., 2018). These datasets may be available for download in various formats. These portals play a crucial role in making data accessible to the public. In the European Union, data.europa.eu is the leading open data initiative that brings together datasets from community providers at all levels of government (Carrara et al., 2015). In turn, all member states have a national open data portal (Herrera-Murillo et al., 2022).

Current user interfaces in open data ecosystems do not facilitate the addition of data by external actors outside government data providers (van Loenen et al., 2021). The latest edition of the Open Data Maturity Report (Publications Office of the European Union, 2023) offers several recommendations for leveraging collective intelligence by engaging the broader open data community in national open data programs. Key suggestions include allowing re-users to upload their own data and showcase their ideas and creations on national portals. Additionally, the report advocates for enabling users to comment on and rate datasets, with these inputs being integrated into search algorithms.

We conducted a review of the European Open Data Portal and 27 national portals, as compiled by Herrera-Murillo et al. (Herrera-Murillo et al., 2022), totalling 28 portals. During this inspection, we specifically assessed whether these portals explicitly and visibly offer users the following features: a) the ability to request the addition of specific datasets, b) the option to submit user-generated use cases or applications, and c) the opportunity for non-governmental users to contribute their data to the portal.

Out of the 28 portals reviewed, 17 (61%) provide users with the option to request the addition of datasets. However, the implementation of this feature varies significantly across portals. In some instances, a generic help request form is used, where users select the category of their request, with dataset incorporation requests being one of the options. Other portals offer more sophisticated mechanisms; for example, the Estonian portal displays all received dataset proposals and allows users to vote on their favourites, while the Lithuanian portal presents a request history that includes the outcomes—whether favourable or unfavourable—along with explanations for each decision.

Eighteen of the 28 portals (64%) allow users to submit applications or use cases to be showcased on the site. However, the methods for submission vary widely. Some portals simply invite users to share their use cases via email, offering little guidance on the required structure. In contrast, other portals provide a dedicated form for submissions. A notable example is the Austrian portal, which features a comprehensive form with multiple fields, including title, description, nature of the application, open datasets used, links to the application, links to the source code, and administrator information.

Only 6 of the 28 portals (21%) allow providers outside the public administration to ingest their own data into the portal. These countries are France, Spain, Austria, Poland, Luxembourg and Portugal. The portals of France, Austria, Luxembourg, and Portugal permit uploads from all types of users. The Polish portal extends this capability to companies. In contrast, Spain restricts the uploading of non-governmental data to select private institutions, primarily foundations, associations, and professional bodies that generate data from publicly funded projects.

To illustrate how a non-governmental provider can add its data to an open data portal, we will use the French portal as a reference. The portal offers a range of resources detailing best practices for open data publishing, as well as specific guidelines on how the platform operates. Data providers begin by registering an account, which allows them to add datasets. The process starts

with including the metadata for the resource, including the title, description, license, update frequency, keywords, time coverage, and spatial granularity. After this, the file itself is uploaded (See Figure 4). Finally, the publisher is encouraged to share the newly created resource. Once the dataset is published and accessible to the public, the data provider gains access to a dashboard where they can edit the entry and monitor various metrics related to the dataset's usage and interaction. These metrics include views, subscribers, use cases, comments, community contributions, quality indicators and recommendations to improve data and metadata quality (See Figure 5). Non-government providers can also use the portal API to automate data ingestion into the portal.

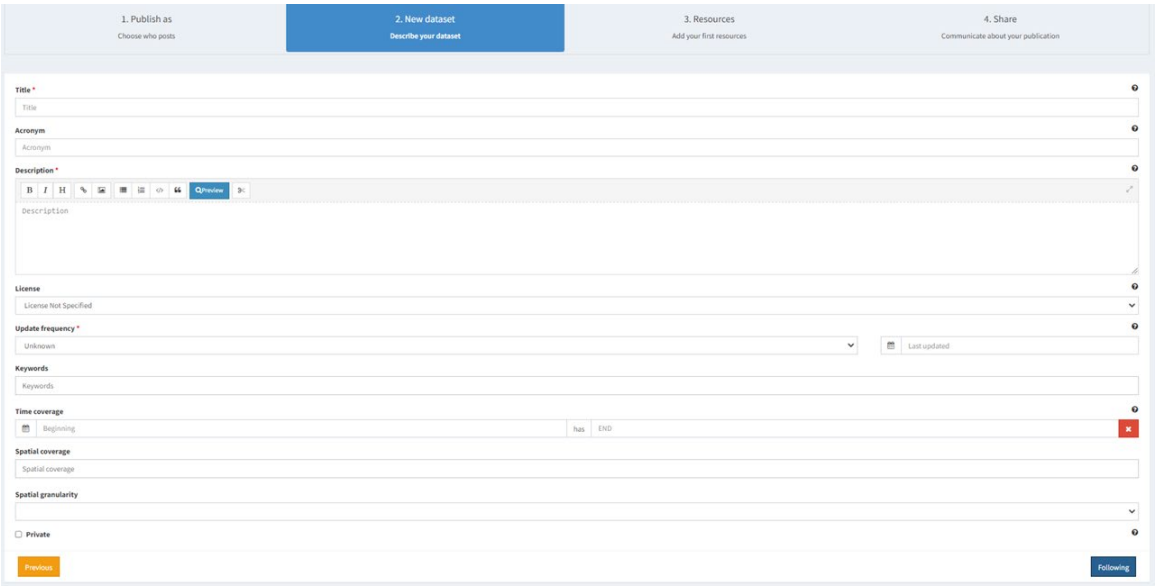

Figure 4: Screenshot of the process of ingesting individual datasets into the French open data portal

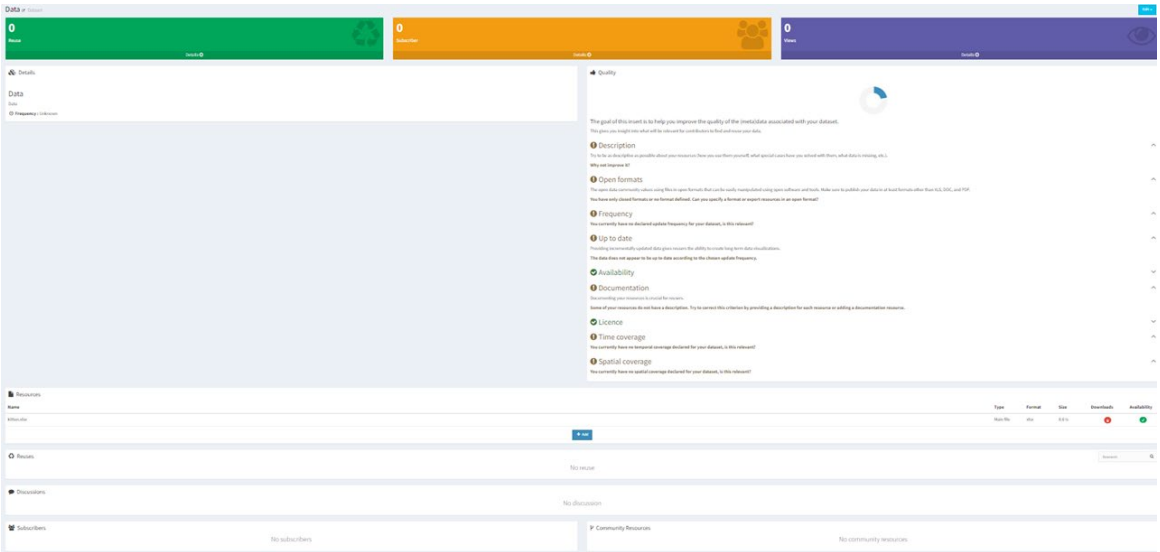

Figure 5: Screenshot of the dashboard of the datasets ingested by a user on the French open data portal

The addition of third-party datasets is not limited to government open data portals; there are also numerous open-access repositories that enable a wide range of users, particularly researchers, to share datasets generated from their work and research processes. A notable example is the Zenodo repository, managed by CERN, which currently allows the creation of entries containing

up to 100 files and 50 gigabytes of data. These entries can include both final and intermediate research outputs, such as datasets. Other general repositories include Figshare, Dryad, and the Open Science Framework. In addition, there are domain-specific repositories, such as ICPSR for social sciences, as well as institutional repositories hosted by various academic and research institutions.

6.1.3 Open data generated by volunteers through micro tasking: the case of the HOT Tasking Manager

This section reviews the main lessons learned from the ODECO project research on user interfaces for open data generation, specifically in the context of humanitarian mapping Herrera-Murillo et al. (2024).

Voluntary Geographic Information (VGI) is a philosophy that harnesses the collaborative contributions of volunteers to collect, analyse and distribute geographic data. This is one of the most successful examples of open data contribution by non-governmental actors. VGI plays with the idea of "citizens as sensor" which appeals to the ability of human beings to synthesize and interpret local information in an intelligent way (Goodchild, 2007). In turn, the Humanitarian OpenStreetMap Team (HOT) is one of the most recognized VGI initiatives. It leverages the OpenStreetMap (OSM) platform to enable the development of freely accessible global maps, with a special emphasis on regions in critical need of humanitarian assistance. A relevant tool in this process is the HOT Tasking Manager (HOT-TM), which coordinates a worldwide network of volunteers to support focused and efficient mapping efforts. Figure 6 and Figure 7 shows the project tasks screen where volunteers select the mapping task to be performed and the screen corresponding to an individual task.

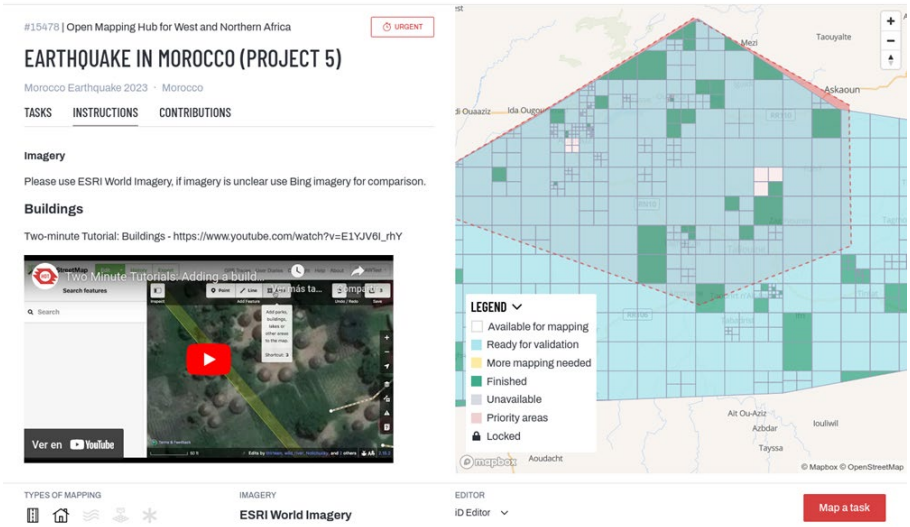

Figure 6: The project tasks screen in HOT-TM displays key information in two sections. On the left sidebar, users can find project details, task statuses in list form, step-by-step task instructions, and contributor statistics

Figure 7: Screenshot of the mapping phase in HOT-TM. Source: https://tasks.hotosm.org/projects/15476/tasks

As part of the ODECO project, Herrera-Murillo et al. (2024) conducted a systematic analysis of the humanitarian mapping process within HOT-TM, aiming to understand the interactions between volunteers and the user interface of the platform. Based on this research, the authors developed a set of practical recommendations for designing platforms and user interfaces that facilitate data contributions from non-governmental users, especially those that seek to involve the public, regardless of whether geographic information is involved:

Technology is a powerful enabler for building a community of data contributors: HOT-TM serves as a model that can be replicated by other non-governmental open data initiatives. Its technological platform acts as a bridge between a robust volunteer community, like OSM, and humanitarian organizations responding to global disasters. The platform equips both groups with the necessary user interface, tools and information to create and validate open data.

Volunteer experience is a key asset: when volunteers sign up for HOT-TM, their previous mapping experience within OpenStreetMap is automatically shared through indicators. This enables project managers to set experience requirements for participation in projects based on their complexity. The task validation phase is reserved for more experienced users, who possess the necessary judgment to evaluate and correct mapping products effectively. Research shows that this distinction is well-founded, as advanced users consistently demonstrate higher productivity, better task success rates, and sustained participation over time. Open data initiatives by non-governmental actors can greatly benefit from identifying key variables to assess data contributor experience. Doing so allows for more efficient task assignment, ensuring that projects are matched with the right level of expertise. Managing the expertise of data contributors also involves addressing the challenge of retaining and motivating novice users, many of whom stop contributing after their first project. This suggests that initiatives should focus on flattening the learning curve for newcomers, providing them with constructive feedback and friendly user interfaces, and sharing indicators of the real-world impact of their contributions.

Micro tasking can simplify participation for data contributors: HOT-TM employs a micro tasking approach that, unlike regular OSM dynamics, directs users toward specific, simple mapping objectives. Most tasks can be completed in just a few minutes, allowing even inexperienced users to contribute meaningfully to the project goals. This micro tasking model can be applied to other non-governmental open data aggregation initiatives, where participants are asked to make small and manageable contributions. When combined, these individual efforts can have a significant impact, making large-scale projects more accessible and efficient through the power of crowdsourcing.

Data quality matters: a common concern with open data generated by non-governmental stakeholders is whether it meets the necessary quality standards for future use. While many of these stakeholders possess valuable data, they often lack the skills and tools for effective data and metadata quality management. Overburdening them with additional quality control tasks could discourage their participation or divert their focus. To address this, HOT-TM has developed a two-stage quality control mechanism. In the first stage, during the mapping phase, contributors focus on achieving the mapping objectives as quickly as possible, ensuring that geographic information for emergency response is available immediately. In the second stage, during the validation phase, project organizers and experienced contributors validate and correct the mapping tasks, ensuring that the final product meets quality standards. This approach demonstrates how any open data initiative involving non-governmental actors can benefit from similar quality control mechanisms, whether through automated rule-based checks or reviews by human experts.

Process data is a useful tool for continuous improvement: HOT-TM collects detailed data on contributors, the work process, and the resulting outputs. This information is shared with relevant communities via a well-documented API (see Figure 8). This approach enables data-driven management of non-governmental open data initiatives, allowing external contributors to engage in process improvement within an open innovation framework. Key improvement decisions, such as those related to user interface design, can be effectively informed by analysing interaction data collected from users.

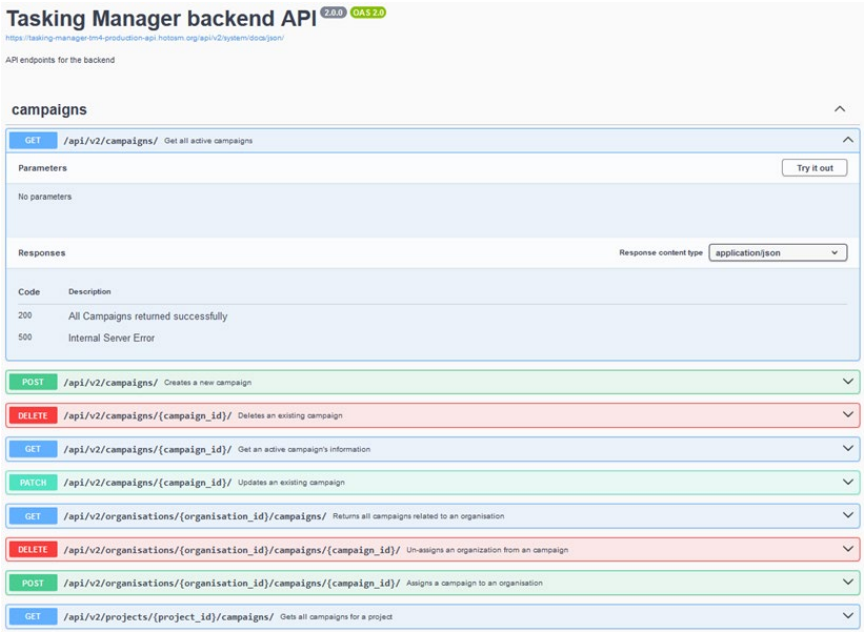

Figure 8: HOT-TM API that allows access to data on campaigns, projects, organizations, tasks, comments, among others. Source: https://tasks.hotosm.org/api-docs

Harnessing the collective intelligence of non-governmental open data contributors: Research indicates that HOT-TM is highly efficient in achieving broad mapping objectives within a short timeframe. However, its success relies heavily on experienced mappers, while novice mappers—who make up most contributors—tend to have a lesser impact, partly due to suboptimal interaction with the platform. As platforms like HOT-TM develop more sophisticated mechanisms for aggregating user input, they can tap into the "wisdom of the crowds," where the collective contributions of a diverse group result in an outcome of higher quality than the sum of individual inputs.

After identifying the tools for data sharing described above, it is noticeable that these tools allow for the data creation and sharing types of platforms, however, the need to make the created datasets more accessible in general open data portals remains. Many of the technical steering

mechanisms (or solutions) proposed in the following Section, are not limited to (even though they could improve those as well) government data sources (e.g., government open data portals), but they are applicable for use on other non-governmental data sources and providers, which is mainly the focus of this Deliverable.

6.2 Proposed technical solutions within the ODECO Project

In this Section, proposed technical solutions corresponding to each of the Challenges described in Chapter 5 are presented. The proposed solutions are either a result of ongoing research within the ODECO project (thus not finalised but under progress), or a result of a proposed solution from the ODECO TW5 Ideathon.

An important aspect to consider is that there is a distinction between the technical solutions within the ODECO project and the ones proposed during the OD Ideathon in Samos. The former ones are ongoing research of the ODECO team, and they will continue to progress while the project is still ongoing, but the latter are solution ideas to be considered and might be (usually the case) at a much earlier stage of conception, whose progress into deeper levels of progress is out of the scope of this deliverable.

Another important note for the following Section of this document is the distinction between the term "user" in its conventional use and meaning (so anyone, including citizens, non-specialist user groups etc.) and the use for this deliverable. In this Section, whenever the term "user" is encountered, it is not limited to the user as simply a data user, but rather, it refers to a system user, so the target system user who will make use of the proposed technical solutions presented. In our case, and as the main purpose of this deliverable, this of course refers to the non-government data holder groups which we aim to motivate to share their data, using the proposed and envisioned technical solutions.

6.2.1 Proposed solutions for Challenge 1 - Enhanced data creation process through better quality, integration and validation tools

6.2.1.1 Augmenting semantic interoperability for datasets

This proposed semantic interoperability solution aims to target one of the common problems that arises when non-government data must be combined and integrated with other the data sources: the semantic heterogeneity between the schemas (conceptual models) of two data sources that must be combined. The main challenge related to semantic interoperability that organisations and data portals face is related to the inconsistent use of data vocabularies and metadata schemas (Maratsi et al., 2024 a); Organizations use different (custom) data and metadata schemas that do not match existing standardised vocabularies and ontologies to express domain-specific knowledge.

For instance, taking an example from the Cultural Heritage domain, and assuming we have two organisations (e.g., art gallery, museum, etc.), the artifacts exhibited in Art Gallery 1 and Art Gallery 2 of Figure 9, are not automatically interoperable when exchanging data from these two institutions, due to the lack of linking the custom schemas to known, standardised ones of this specific domain (e.g., CIDOC-CRM and more, in this case). The quest for standardisation and interlinking existing ontologies and schemas mainly in the cultural heritage domain, but also in other domains (e.g., legal domain, psychology), has been actively researched within the ODECO project, the results of this study to be found in (Maratsi et al., 2024 b, Loutsaris et al., 2023, Maratsi et al., 2024 c, Maratsi et al., 2023).

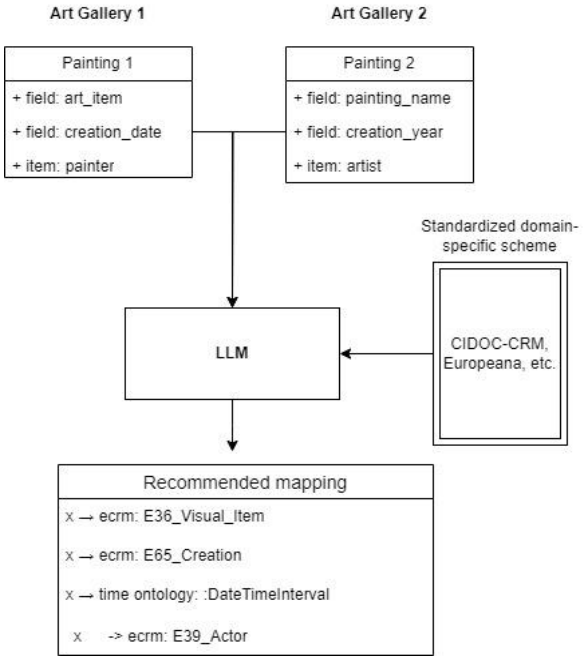

Figure 9: A simple example of different data schema use in a Cultural Heritage case

If the custom schema fields of the two different institutions could be mapped in an automated or semi-automated way to existing, standard schemas upon dataset upload for instance, then the new data would instantly become semantically interoperable with more data sources, and more easily discoverable by semantic search engines.

The proposed solution aims to use an LLM (e.g., LLAMA) to align and map these various schemas to standardized, vocabularies. This model would be trained using expert-curated data to ensure accuracy and would follow the process of Figure 10.

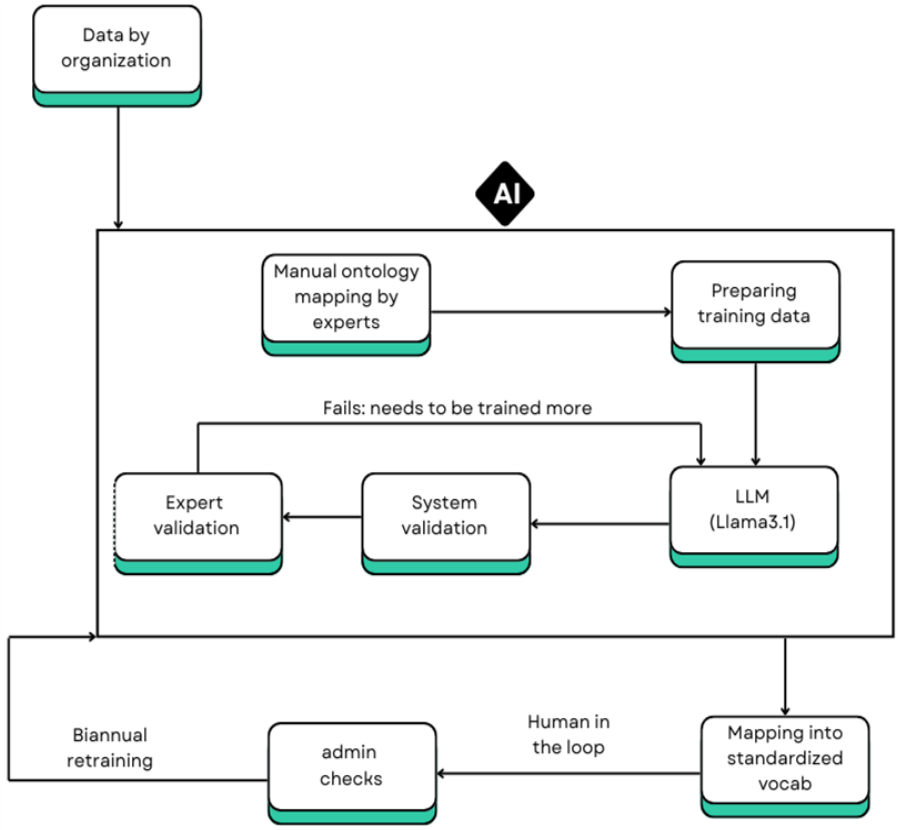

Figure 10: The proposed implementation process

Initially, the LLM model will be trained on curated expert-validated data (examples of experts' mappings), and then will go through a preliminary system validation process which will be supervised (approved or rejected) by a human. This is going to be an iterative process until consensus is reached regarding the produced mappings to standardised vocabularies and schemas. On a similar note, the use of Linked Open Vocabularies, for the facilitation of this process of ontology and standard domain-specific vocabulary re-use for improved semantic interoperability is described by (Maratsi et al., 2024 d). Some technologies to be used for the implementation of this solution are the following:

- Data Ingestion & Storage: AWS S3
- Model Training and Deployment:
    - Model Training: PyTorch, TensorFlow
    - Training Environment: GPUs and TPUs on AWS SageMaker
- Model Deployment: Docker
- Data Integration: Apache Kafka
- Monitoring & Logging: Prometheus

The outcome of this process is a coherent data integration which ensures all data is aligned with the proposed system's standard vocabularies. The form of outcome is a produced recommendation of mappings to standardised representation schemas of a specific domain (or more generic if this is not necessary). The results of this process will then be contributed to the Linked Open Data Cloud (LODC).

In practice, when an organisation aims to share their data (e.g., upload to an OD portal) or elsewhere, the produced system solution will provide them with a set of recommendations on how their custom schema can be improved in terms of semantic interoperability. Figure 11 shows an indicative screen prototype of the mapping recommendation process to the end user (e.g., organisation or data holder group which wants to share their data).

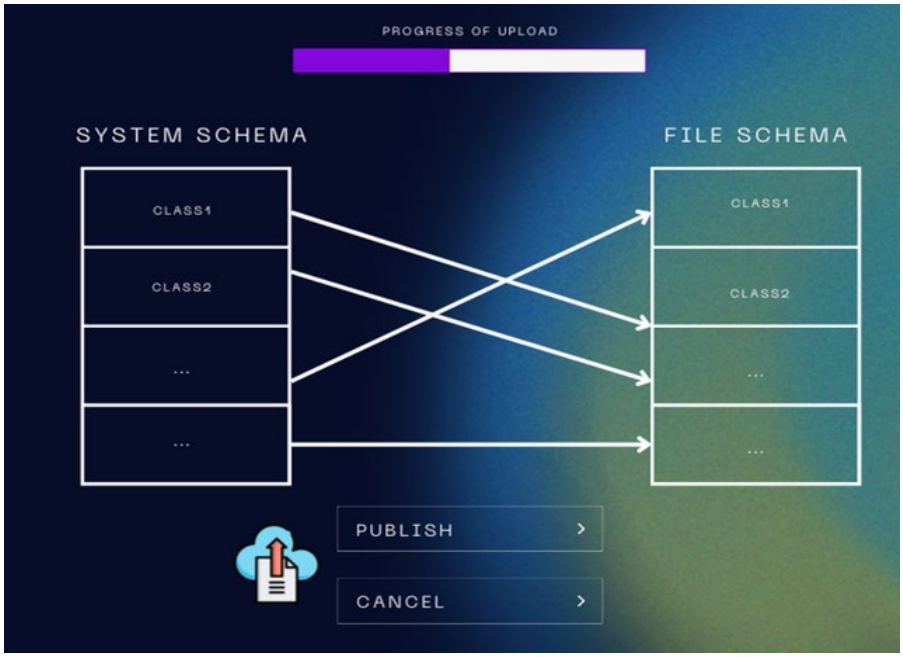

Figure 11: Indicative screen prototype of the mapping recommendation upon dataset upload

This proposed technological solution can be utilised, for instance, as a SaaS data standardisation platform, opening various opportunities to be integrated in existing open data portals, or as a separate service. This solution aims to improve the status on semantic interoperability of datasets and representation schemas, but also consequently improves dataset findability, while at the same time makes active contributions to the Linked Open Data paradigm.

6.2.1.2 Automatic Thematic Annotation of Open Data

Whenever a non-government data asset is integrated in an open data portal, it is essential to classify and annotate it appropriately to facilitate its findability and access by potential users. This problem is similar to the problem faced by Open Data portals such as the European Data portal (data.europa.eu) that must aggregate and classify the contents harvested from heterogeneous national and local portals to predefined list of themes. Although ideally the original datasets should be already annotated with a theme vocabulary and later matched with the European data themes, based on our desktop review of the datasets, it is evident that assigned themes are often either missing or incorrect. Incorrect theme assignment can hinder the findability of the datasets by other stakeholders.

To address these issues, we developed a solution that assigns proper themes to datasets, even when a dataset belongs to more than one category. This solution is a thematic classification system designed to resolve the problem of incorrect thematic assignment. We created a multi-class classification model using a training dataset of 29,793 records in RDF format. The system employs both classification techniques and cosine similarity, analysing the metadata (title and description) of datasets against the themes defined by the European Open Data Portal.

The developed system can assign themes to datasets based on their title and description with 92% accuracy. Additionally, the model can suggest a second theme that could be assigned to a dataset, improving its findability on the portal. All the code and data are available on the github repository.

For example, if a user owns a dataset and wants to determine the appropriate category based on the title, description, keywords, and publisher, our model will suggest one of the 13 themes required by the European Open Data Portal. In the second step, the model will also indicate if the dataset belongs to more than one theme.

Figure 12 illustrates the proposed methodology for the automatic thematic classification model. The process consists of five steps: (1) data collection for model training, (2) noise removal from the collected data, (3) inputting the data into two models—a supervised multi-class classification model and an unsupervised model—and (4) evaluating its performance against metrics such as accuracy and precision of the model (5) deploying the models to ensure correct thematic assignment for open datasets.

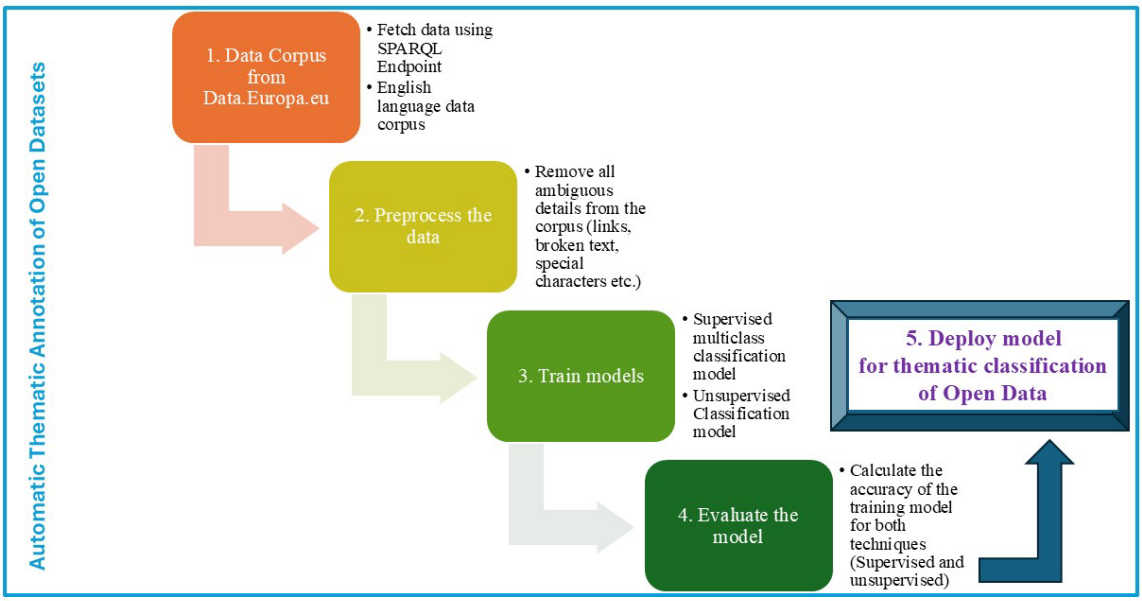

Figure 12: Proposed methodology for the automated thematic annotation of open datasets

This proposed and developed solution will enhance the findability of data and contribute to the accurate assignment of metadata (Theme) during the data publishing phase.

Another study (Ahmed et al. 2024) focuses on enhancing data findability by proposing a strategy for automatic keyword generation. It evaluated five existing keyword extraction methods (BERT, RAKE, YAKE, TEXTRANK, and ChatGPT (version 3.5)) and introduced a novel hybrid model called BRYT. The evaluation was conducted using datasets from the EU data portal, which featured high-quality metadata. The code for this experiment can be found on GitHub. Results showed that BRYT outperformed other algorithms in generating representative keywords, with 69.1% of keywords showing major matches, significantly improving data findability. ChatGPT performed well as a close second in major matches, and YAKE led in minor matches, with ChatGPT again coming close. The study concluded that BRYT consistently delivered better keyword extraction, enhancing the metadata's overall quality and making data more accessible. This research emphasizes the importance of high-quality metadata for improving the discoverability of open data and presents BRYT as an effective solution for keyword extraction. It sets the stage for future advancements in representative metadata extraction and population across open data platforms. Figure 13 shows the complete architecture of this research study.

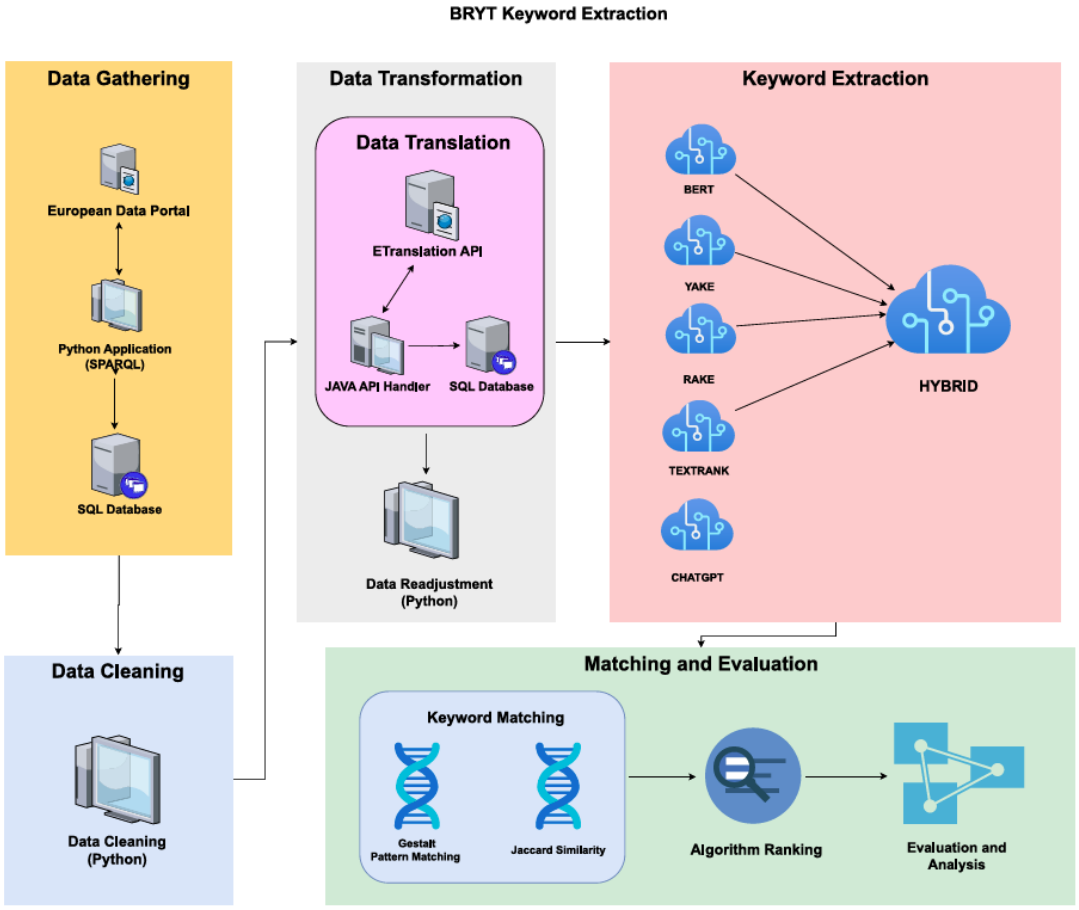

Figure 13: BRYT Keyword Extraction Architecture

6.2.1.3 Intelligent Validation of Data on the Web

The current process of checking data compliance for publication in open data portals is tedious and time-consuming, especially when it comes to adhering to standard best practices such as DCAT-AP and W3C guidelines. This challenge becomes even more pronounced if non-governmental data holders increase their contributions to these portals. Similar research within the ODECO project proposes a framework to automate the evaluation and standardisation of open data coming from local open government data sources and improve their infrastructure (Ali et al., 2024). To address this, we propose developing a tool that can assist both open data portals and data holders in validating metadata efficiently. The tool will automatically detect metadata on the web, assess its compliance with the required standards, and provide actionable recommendations to improve metadata quality. This solution aims to simplify the compliance process, making it easier and faster for data holders to meet the necessary standards.

Example Scenario: Figure 14 illustrates how the application can assist NGOs or open data publishers in ensuring their data meets required standards before publication. For example, when an NGO aims to publish data on data.europa.eu, they must first ensure that their data complies with established standards and protocols, such as DCAT-AP or the "Data on the Web" best practices by W3C organizations. The validator simplifies this process by checking the data's compliance with these standards. Additionally, the system supports publishers by identifying issues with metadata alignment, offering insights on how to improve compliance and ensuring that the data is properly structured for open data portals.

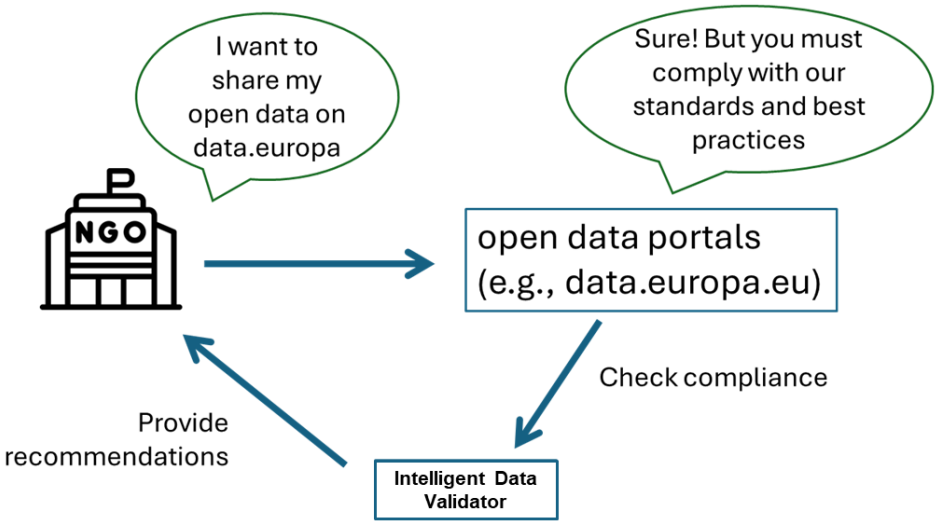

Figure 14: Example Scenario for the Application

Figure 15 illustrates the functionalities and workings of the Minimum Viable Product (MVP) of the application. The system is designed to incorporate various data compliance standards, such as W3C Data Practices and FAIR or DCAT-AP compliance checks. These functionalities are pre-integrated into the application, allowing data publishers to easily verify if their data complies with W3C's "Data on the Web" best practices and other standards like FAIR and DCAT-AP. The application generates reports that assess the data's alignment with these standards and provides recommendations and suggestions to improve compliance. Additionally, the application is capable of mapping dataset metadata to the DCAT-AP standard, further enhancing metadata accuracy and improving overall data practices for publishers. This ensures that datasets are more compliant, discoverable, and ready for publication on open data platforms.

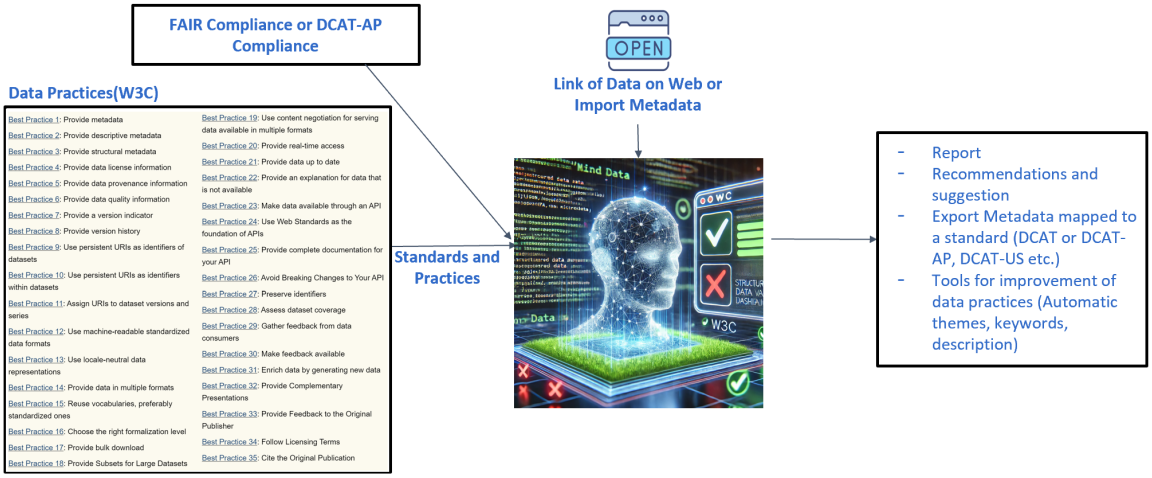

Figure 15: Minimum viable product for Application functionalities and working

Figure 16 outlines the technology stack required to develop the Minimum Viable Product (MVP) for the application. At the core, we need to store data related to compliance standards as well as the datasets themselves for future analysis. For this, SQL and NoSQL databases or services will be used to efficiently manage both structured and unstructured data.

The second layer involves the development of microservices, which are essential for checking compliance. For instance, the W3C "Data on the Web" best practices involve over 60 checks, and each of these checks will require a dedicated microservice to ensure thorough validation. The MVP will run on a Linux-based operating system, providing a reliable and scalable foundation for the application. For the application development itself, the MEAN stack (MongoDB, Express.js,

Angular, and Node.js) will be a suitable choice, offering a robust environment to build the graphical user interface and backend services needed for the application functionalities.

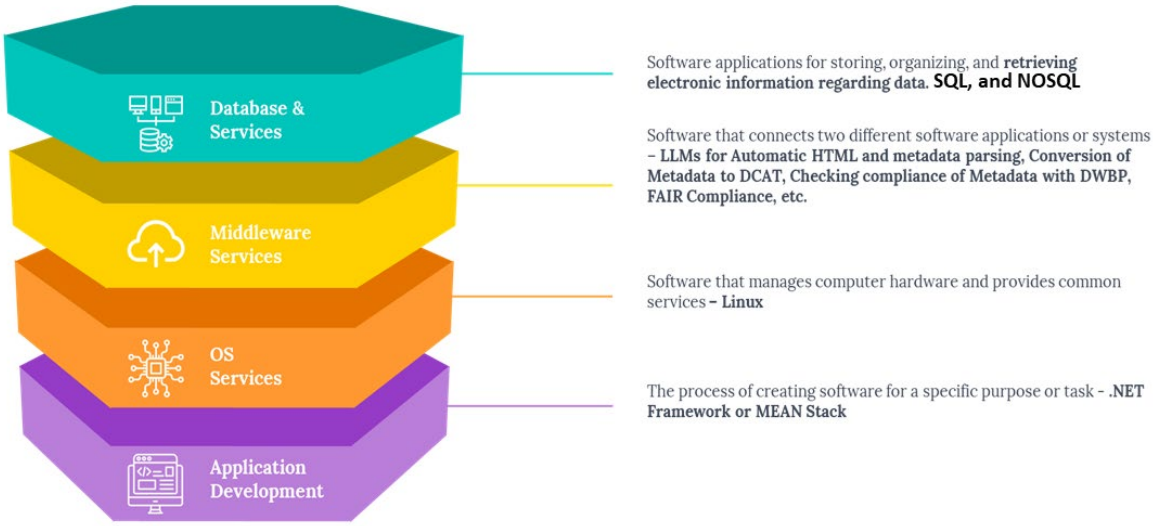

Figure 16: The technology stack required to develop the Application

By automating the compliance process, the application ensures that datasets meet established standards such as DCAT-AP, FAIR, and W3C "Data on the Web" practices. This reduces errors in metadata alignment and improves data quality.

6.2.1.4 Open Data Driven Citizen Science Applications as a mechanism for inclusion of non-governmental open data in open data portals

In their report, "Data.europa.eu and Citizen-generated Data," Corcho et al. (2022) highlight the minimal presence of citizen-generated data within European government open data portals, spanning all administrative levels from local municipalities to data.europa.eu. The report recommends developing tools and applications that leverage this type of data, enabling citizens to contribute directly—whether through data collection or curation—to the original data sources.

We address the key question of how can conventional open data portals benefit from citizen science initiatives and crowdsourcing to incorporate more non-government data? We believe that one way forward is to leverage the existing datasets on these platforms to inspire the development of citizen science applications, following the notion of "data invoking data." The building blocks of the proposal are described in the Figure 17 below:

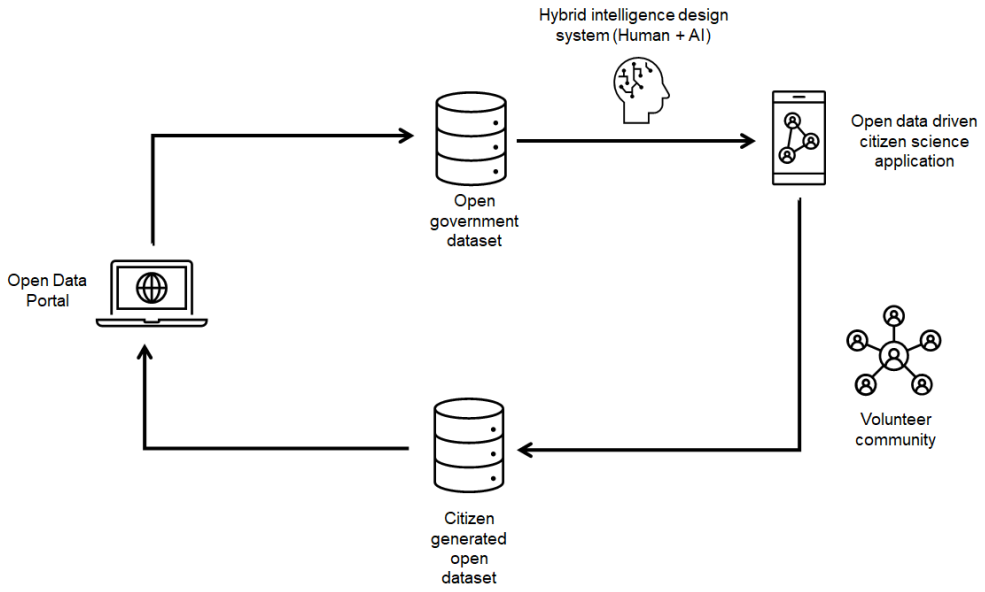

Figure 17: Building blocks of the Open Data Driven Citizen Science Applications generation model

- **Open Data Portal:** In its traditional form, an open government data portal facilitates the discovery, access, and management of metadata records for datasets. An expanded version—already adopted by many government data portals—goes further by linking available datasets to practical use cases and applications. Among these, applications with a focus on citizen science have the potential to be significantly strengthened and made visible, this idea is supported by Corcho et al. (2022).
- **Governmental Open Datasets:** When determining which open datasets are most suitable for driving the development of citizen science applications, those addressing topics of broad public interest—such as urban infrastructure, the environment, or public health—are ideal candidates. Datasets that include geographic information, such as point coordinates or defined geographic areas, are particularly valuable, given that geospatial datasets are also considered High Value Datasets (lex-europa.eu, 2023). It would be beneficial to include and highlight metadata that help identify datasets most suitable for developing citizen science applications. This metadata could include aspects such as geographic coverage.
- **Hybrid Intelligence Design System (Human + AI):** Designing data-driven applications is a complex process. To tackle this challenge, we propose an approach that combines user-centred design with hybrid intelligence (See Appendix B), harnessing the strengths of both human and computational inputs. This integration enhances the usability and effectiveness of the resulting applications, leading to more impactful solutions.
- **Open Data-Driven Citizen Science Application:** As a product of the design process, the application fosters a circular dynamic. It provides a benefit to users, typically offering information. In return, users contribute data —either actively or passively— while using the application. Incorporating micro tasking schemes can help alleviate the complexity burden on users, making participation more manageable and facilitating quality control.
- **Volunteer Community:** Citizen science applications should be entrusted to a community of users who not only benefit from the application but are also motivated to contribute to topics of public interest. These communities take full advantage of the concept of "citizens as sensor" proposed by Goodchild (2007). Users of citizen science applications can generate various types of data, including: a) numerical data from physical measurements or counts, b) geospatial data from recording exact observation locations or mapping areas and routes, c) audiovisual data such as photographs, videos, or audio recordings, d) descriptive data in the

form of comments and annotations, e) temporal data marking the date and time of observations or observation series, and f) data from surveys and questionnaires.

- **Citizen generated open dataset:** user-generated data, whether collected directly or through microtasks, is subsequently anonymized, cleansed, integrated, licensed, and properly tagged to form citizen-generated open datasets. These datasets are then made publicly available through the open data portal, contributing to the broader open data ecosystem.

6.2.2 Proposed solutions for Challenge 2 - Enhanced data sharing through better licensing and anonymisation

6.2.2.1. A deep learning application plugin to help users easily select and customize data licenses

This proposed solution aims to help users easily select and customize the licenses for their data, ensuring that they fully understand how it can be used. When referring to users here, it is important to note that users are not limited to users/citizens, but users as in system users, so in our case, the data holder groups which want to share their data. The proposed platform aims to offer two main functionalities:

- Automatic license selection. Users will know exactly how their data is used and will be able to customize the license to fit their needs. The platform will monitor compliance with the terms of the license and ensure that the data is used correctly.
- Anonymisation and privacy protection. The platform will provide a data anonymization tool to protect the privacy of contributors, and the users may choose the extent of anonymity for their own data, according to their preferences.

These two functionalities can be backed up on the following technologies:

- **Data Encryption Technologies**

Encryption is essential for protecting sensitive information and ensuring compliance with regulations, such as GDPR (Charalabidis et al., 2018). This technology converts data into an unreadable format, which can only be decrypted by authorized users.
Utilizing protocols such as SSL and TLS to secure data during transmission and storage.
Strong Encryption Algorithms: Implementing robust encryption standards to protect data integrity and confidentiality.

- **Generative AI - Supervised Machine Learning**

Generative AI and supervised machine learning can work together to create powerful models capable of producing new data that resemble a given dataset. Combining generative AI with supervised machine learning allows for the creation of models that not only understand and classify data but also can generate new data that mimics the original dataset. a) Training Phase b) Model Learning c) Generation Phase.

- **Interoperability Standards and APIs**

Implementing interoperability standards and APIs can enhance data sharing capabilities across different systems (Maratsi et al., 2024, Ali et al., 2024). This includes:
Standardized Data Formats: Ensuring that data can be easily shared and understood across various platforms.
API Integration: Allowing seamless data access and integration with existing systems, facilitating compliance with licensing agreements.

- **Data Masking and Anonymization Tools**

These tools contribute to protect personal data by obscuring identifiable information, making it suitable for analysis without compromising privacy. Key aspects include:
Data Masking: Creating a structurally similar but non-authentic version of the data for testing or training purposes.
Data Anonymization: Removing or altering personal identifiers in datasets to ensure that individuals cannot be identified from the data.

Figure 18 shows the proposed solution development process to create a first Minimum Viable Product (MVP) of the provided service:

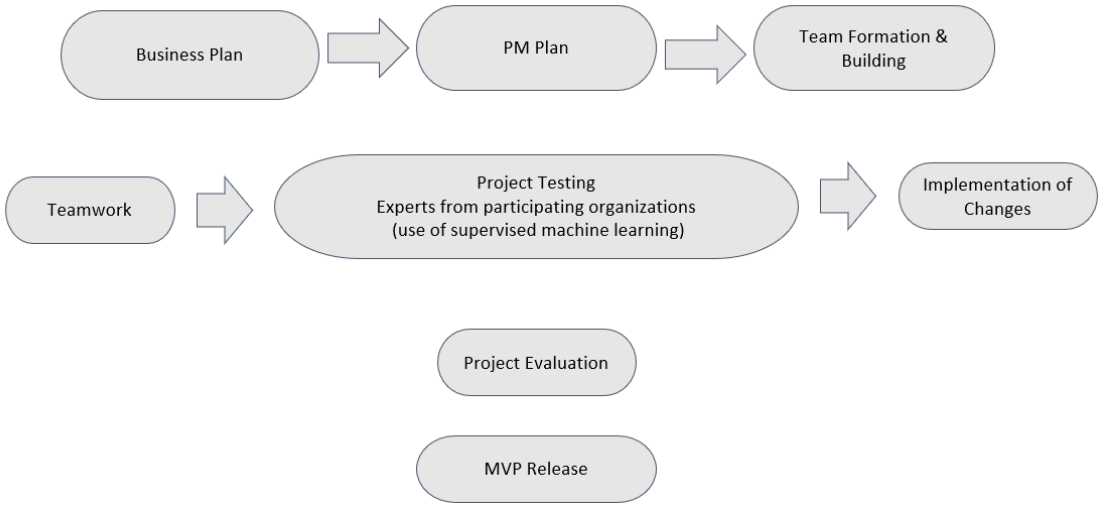

Figure 18: The development process of the proposed platform

This proposed solution reduces the risk of data exposure during development and testing phases, protecting sensitive information while it also improves data quality by ensuring high data quality and integrity, leading to better decision-making and operational efficiency. Moreover, it simplifies compliance with data-related regulations by establishing clear policies and procedures for data management, integrating data licensing sand compliance under the same platform.

6.2.2.2. An Open License Decision Support Software

One of the problems that system users (in this case, including interested non-government data holder groups to share data) are encountering is the complexity of licensing currently used in open data sets. Now, there are more than 300 open data licenses in use. Choosing the right license requires specialized knowledge to understand the various licenses accompanying open data sets and to properly reuse them. It becomes even more complicated in cases where users must combine data sets, especially from different providers and from different countries with different legal frameworks. This can be a severe impediment and discourages open data users not only from using open data but, more importantly, from becoming active contributors to the open data ecosystem. Therefore, the aim of this proposed solution is the creation of an open license decision support system that can help open data creators and users to handle the difficult decision of picking a data license.

Although there are some tools that provide this functionality, some drawbacks have been detected. The design of these tools is static – they either provide far too much information in a tabular form, or they exist as multi-step questionnaires where the respondent has no idea about their progress in the questionnaire or about how the questionnaire itself is structured. In this proposed solution, the goal is to develop a tool that can (Figure 19):

- Provide a decision-support system to dataset creators to choose the right license according to their specific needs.
- Build a system that will serve as a learning tool for dataset creators on how to navigate the different considerations when they must select a data license.

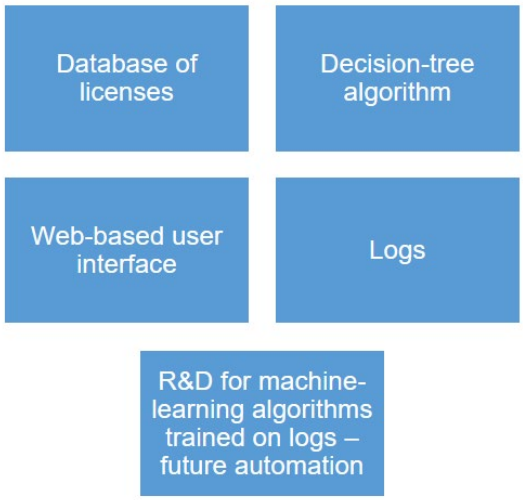

Figure 19: Key components of the system

First, the team aims to design a simple questionnaire to be answered by dataset creators or re-users. This questionnaire will prompt respondents to reflect on the source and provenance of each individual data element in their dataset, their objectives for creating and sharing the dataset, and their proprietary or other interests in the dataset. Each response will be mapped to nodes within the backend decision tree, generating a list of licenses that meet the respondent's parameters.

Second, the team aims for respondents to be able to "see" the internal workings of the decision-support software. This feature enables the software to serve as a learning tool for respondents as well. It also allows for more transparency of the software and moves away from a techno-solutionist "black box" design. The goal is to create a user interface where the respondent can always see each section of the questionnaire as well as their progress. The team also wants to enable respondents to toggle between questions and alter answers. Finally, they intend for the license decision tree to be dynamic for the respondent. In other words, as the respondent progresses through the questionnaire, the license decision tree will be generated with each response and modified with each alteration of a response. Instead of being visible to the respondent only at the end of the questionnaire, this dynamic license tree will always be visible as they go through the questionnaire, allowing them to understand the underlying logic of the recommendation system. Figure 20 and Figure 21 explain the key components of the system and three frame of the proposed user interface for the system respectively.

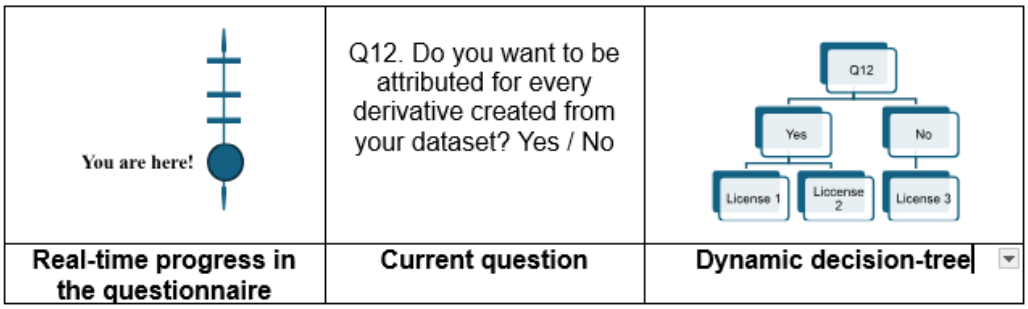

Figure 20: Three-part frame of the proposed user interface

In Figure 21, the development process of the system is depicted. The aim of the team was to adopt a process for the creation of the decision system that would "design solutions 'with' stakeholders, and not 'for' them." Therefore, a circular design for the development was adopted so that the system could be adjusted based on the requirements and needs of the users.

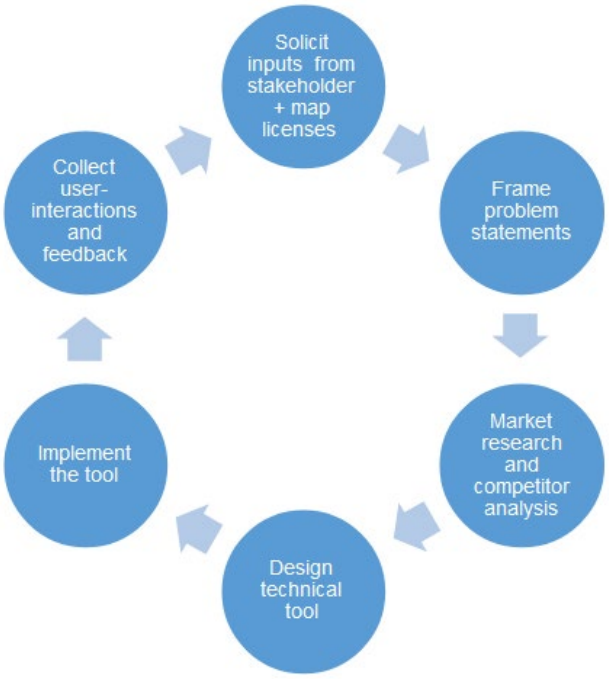

Figure 21: The development process of the proposed platform

The proposed open license decision support system addresses the complex challenge of navigating the multitude of open data licenses, offering a user-friendly solution for both creators and users of open data. By combining a simple questionnaire with a dynamic, transparent interface, the system not only recommends appropriate licenses but also serves as an educational tool, empowering users to understand the decision-making process.

6.2.3 Proposed solutions for Challenge 3 - Towards inclusive, user-driven and circular open data platforms through better interfaces and feedback mechanisms

Imagine an open data portal in which a variety of user groups, such as government agencies, researchers, developers, and organizations that represent civil society, are actively engaging with datasets. The feedback requirements of these user groups are diverse; for instance, developers are looking for more technical data formats, while researchers may seek improvements to metadata to facilitate data interpretation. Despite the availability of feedback channels such as discussion forums, emails, and surveys, there is frequently a lack of transparency regarding the processing and implementation of this input. Consequently, users are left with a lack of clarity regarding the enhancements or modifications that are the outcome of their contributions.

A possible feedback scenario might involve a user accessing a climate dataset through the national OGD portal. After working with the data, they notice inconsistencies in the date formats and provide detailed feedback via the discussion forum or fill in the feedback form. However, due to the fragmented nature of feedback processing across different channels (emails, feedback form, social media, discussion forums), their input is not addressed promptly, or the portal lacks transparency regarding what actions are taken based on such feedback. As a result, the feedback loop is incomplete, hindering data reusability and trust among users.

6.2.3.1. A ChatBot Guide for Open Data Portals

The motivation behind this proposed solution is the need of citizens without technical skills on open data to access information, tools and data available, knowing where and how to find them. Citizens without technical skills might need open data for understanding an issue, for example, climate change, compare Open Data and creating a solution to solve a problem.

The idea pertains to a Chatbot assistant which profiles the user towards the resources they need on the open data portal and externally. Figure 22 shows the screenshot of chatbot.

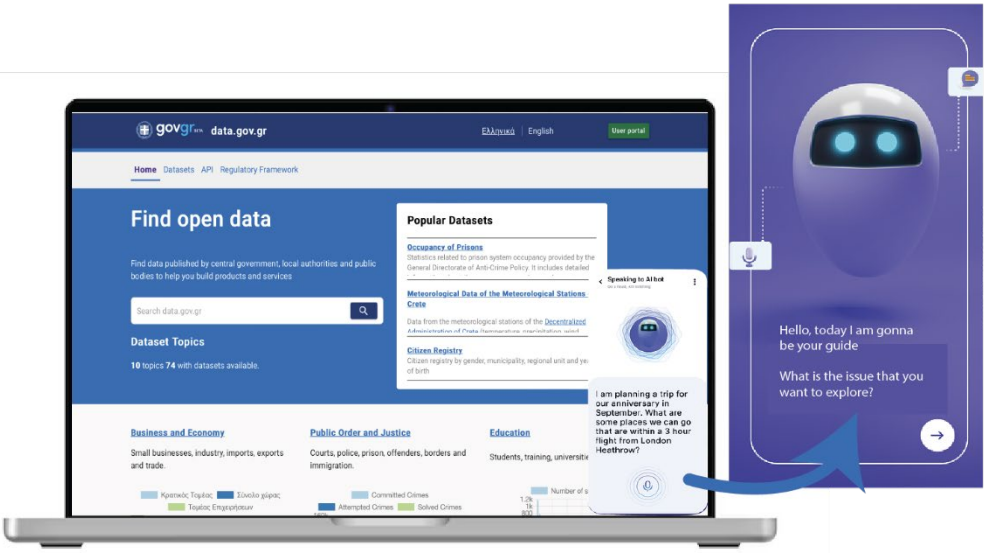

Figure 22: An example of a Chatbot assistant on the Greek open data portal

For instance, as also shown in Figure 22 the Chatbot could ask questions to understand the user needs and help navigate them accordingly through the available data. Below are some indicative questions the Chatbot could ask:

- What is your problem?
- Do you want me to suggest topics of interest?
- What data are you looking for?
- Do you need me to direct you to the tools and guides to help you analyse the data?
- There exists research on this topic using this data. Do you want me to direct you to it?

There can also be suggestions of questions that might be of interest to the user.

The technologies to be used for the implementation of this proposal are mainly Generative AI and Natural Language Processing (NLP). Creating a database of existing tools/visualisations and data research on the portal or from external stakeholders could enhance the trustworthiness of organisations, while the information for the Chatbot can be verified and/or provided by a list of organisations that the data portal owners deem as trustful.

The value of the proposed solution for citizens (users) is that it helps them gain understanding of a topic, while having data and research in the same place. The process of accessing and using data is facilitated. From the open data portals' side: new data users are being attracted. In addition, missing data or components can be discovered, as well as tools and other resources currently not available on the portal. The value of this solution might also be addressed towards private and third parties, such as through the creation of a database of stakeholders (NGOs, NPOs, volunteering organisations and more).

The development process of the Chatbot would include the affixing of an appropriate license, the preparation of the database, tools and research available to use, the verification of sources from external stakeholders, the creation of the algorithm and models, the design of the interface and guide character, the automation of solutions, conducting the appropriate usability tests, and forming a human support group of experts that would be available upon request.

6.2.3.2. Enhancing better interfaces and user interaction under feedback supported by AI

This proposed solution addresses the challenge of better interfaces and user interaction with open data, solutions focused on creating more intuitive, engaging, and accessible ways for users to interact with data. The proposed interface will be hosted by a local data portal and will be addressed to all citizens (including vulnerable groups). The rationale behind this is finding the appropriate mechanisms to support and develop more intuitive, AI-powered interfaces that allow users to query and interact with data in natural language. These interfaces could provide personalized recommendations, visualizations, and insights based on user queries, also allowing for AR (Augmented Reality) integration to create immersive data interaction experiences, where users can explore complex datasets in a 3D environment, making it easier to understand patterns and relationships. Furthermore, voice-activated assistants that guide users through datasets could be introduced, offering explanations, answering questions, and helping with data analysis tasks.

More specifically, the key features of the IDIP (Interactive Data Intelligence Platform) include:

1. **Interactive Data Visualizations**
    - **Dynamic Dashboards:** Create customizable, interactive dashboards that allow users to visualize data in real-time. Users can drag and drop data elements to create their own visualizations (charts, graphs, maps) and explore the data dynamically.
2. **Augmented Reality (AR) and Virtual Reality (VR) Integration**
    - **Interactive Data Models:** In AR, users could interact with holographic data models using gestures or voice commands. This is particularly useful for education, allowing students and professionals to engage with data in a hands-on, intuitive manner. The VR employs 3D near-eye displays and pose tracking to give the user an immersive feel of a virtual world and therefore the users can have a futuristic idea engaging also with data.
3. **Voice-Activated Assistants providing feedback**
    - **Personalized Data Assistant:** Introduce a voice-activated assistant that helps users navigate the open data platform, answers their queries, provides summaries, and even performs data analysis tasks. The assistant could also learn from user interactions to provide more personalized experiences over time.
    - **Hands-Free Interaction:** This feature is especially valuable for users who need to multitask or for those with accessibility needs. The voice assistant could execute commands, retrieve data, and explain complex datasets without the need for manual input.

The technologies the features will be based on include:

- **An LLM Fine tuning** - providing an LLM with additional and specialized data to refine the training.
- **Retrieval-augmented generation (RAG)** a specific architecture in which queries are "augmented" with additional and domain specific/contextual information. Implementing RAG in an LLM-based question answering system has two main benefits: It ensures that the model has access to the most current, reliable facts, and that users have access to the model's sources, ensuring that its claims can be checked for accuracy and ultimately trusted.
- **Technology Acceptance Model (TAM)** (perceived usefulness, perceived ease of use)

- **Impact Tracking**: Track the impact of feedback over time, showing how user suggestions have led to tangible improvements in data quality, usability, or features. This transparency encourages continued user participation.

Other technology characteristics which need to be taken under consideration are:
- Accessibility: Easily obtainable by everyone.
- Licensing: No restrictions on usage.
- Interoperability: Usable across different platforms and systems.

The perceived benefits of such an interface for the users include:
- **Increased User Engagement:** By providing more intuitive and interactive ways to engage with data, users are more likely to explore and utilize open data resources, leading to more informed decision-making and innovation.
- **Broadened Accessibility:** The inclusion of accessibility features and multi-language support ensures that a wider range of users, regardless of their technical expertise or physical abilities, can effectively use the platform.
- **Enhanced Data Utilization:** With tools that simplify data discovery, analysis, and visualization, users can more easily derive insights and value from open data, leading to greater use and impact of these resources.
- **Facilitated Collaboration:** The collaborative features foster a community around data exploration and analysis, encouraging knowledge sharing and joint problem-solving. By implementing an **Intelligent Data Interaction Platform (IDIP),** open data becomes more accessible, engaging, and useful for a broader audience, ultimately enhancing the impact and value of open data initiatives.

Development process parameters
Regarding the development process of this proposed solution, it will need to include the following:
- **User Training and Onboarding:** Provide tutorials or onboarding sessions to help users understand how to effectively use the feedback system. This could include demonstrations of how feedback leads to tangible improvements.
- **Privacy and Anonymity:** Ensure that users can provide feedback anonymously if they choose, especially when dealing with sensitive data or issues. Privacy considerations should be paramount, particularly in handling personally identifiable information (PII).
- **Open Source and Community Collaboration:** Consider making the platform open source to encourage community-driven development and innovation. This also allows for broader collaboration and integration with other open data initiatives.

For Natural Language Processing (NLP) Interfaces:
- **Conversational Querying:** Develop an NLP-based interface that allows users to interact with data using natural language (i.e. using BERT).
- **Contextual Suggestions:** The interface could offer suggestions as users type, predicting their needs based on the context of their queries. This feature helps guide users, especially those who may not be familiar with complex data structures or querying languages.

For Augmented Reality (AR) and Virtual Reality (VR) Integration:
- **3D Data Exploration:** Develop AR/VR environments where users can explore complex datasets in 3D space. For example, in a VR environment, users could "walk through" a city map that displays population density, pollution levels, and transportation networks in a visually immersive way.

Implementation and Deployment:

- **User-Centric Design Process:** Develop the platform with a user-centric approach, involving potential users in the design and testing phases to ensure the interface meets their needs and expectations.
- **Modular Architecture:** Build the platform using a modular architecture that allows for easy updates and the integration of new features. This ensures the platform can evolve with user needs and technological advancements: (*open data -> LLM model (BERT) –> RAG technique -> user interface*)
- **Cloud-Based Infrastructure:** Deploy the platform on a cloud-based infrastructure to ensure scalability, accessibility, and efficient handling of large datasets.

Sustainability:

- Ensuring that the rights will belong to the local government.
- A company will take care of the maintenance of the product, making necessary monitoring and adaptations using future science innovations and developments (after a public tender), ensuring the continuation of the funding).
- User interface and technical specs for to be kept to a certified repository (i.e. ZENODO) – Open Science EU.

6.2.3.3. A multi-channel inclusive feedback mechanism for diverse communities

The ability to provide feedback and user participation is critical for creating a fair digital future in a data-driven society. So, we propose the idea about a multi-channel feedback mechanism to foster inclusivity and engagement in open data sharing. It aims to empower both non-technical citizens and people with technical skills, to provide valuable feedback through open data portals. There are two main channels for this idea:

- First, we want to empower those who already have data literacy skills to provide their feedback through open data portals.
- Second, we want to find new ways to connect and include citizens with the open data portals through intermediation by public libraries. Public libraries offer a wealth of services to citizens and are, by design, inclusive places.

This idea proposes a transformed feedback loop that not only addresses the technical needs of open data portals but also strengthens community bonds, reduces digital inequities, and empowers citizens across all levels to participate in shaping the data landscape. It also has the potential to enhance transparency, elevate public trust, and ensure that data-driven governance is truly reflective of all voices.

Figure 23 and Figure 24 show the workflow of the described two channels.

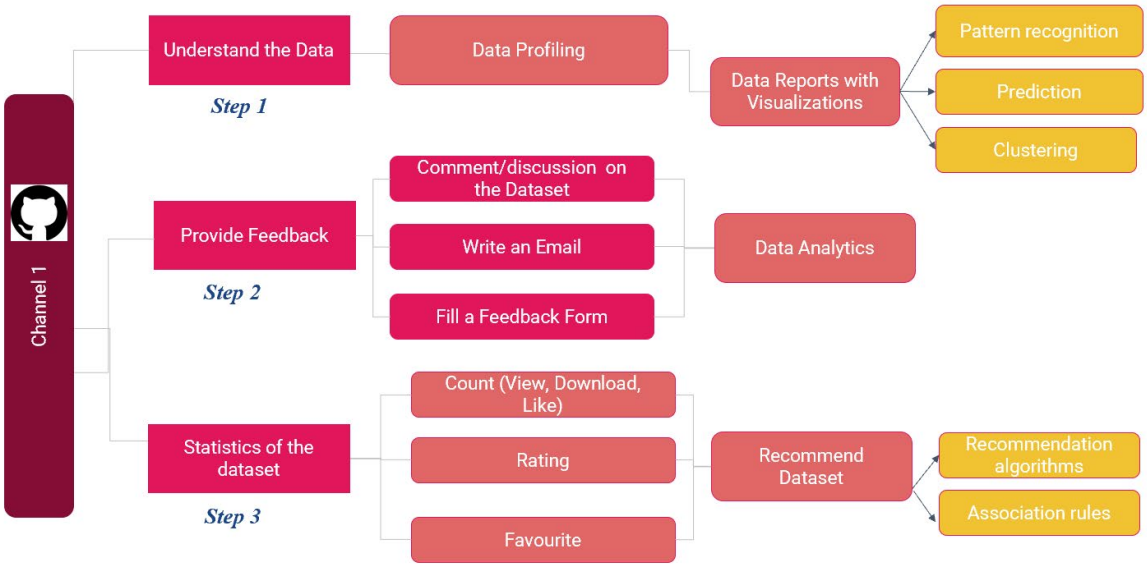

Figure 23: Channel 1 for a multi-channel inclusive feedback mechanism for diverse communities

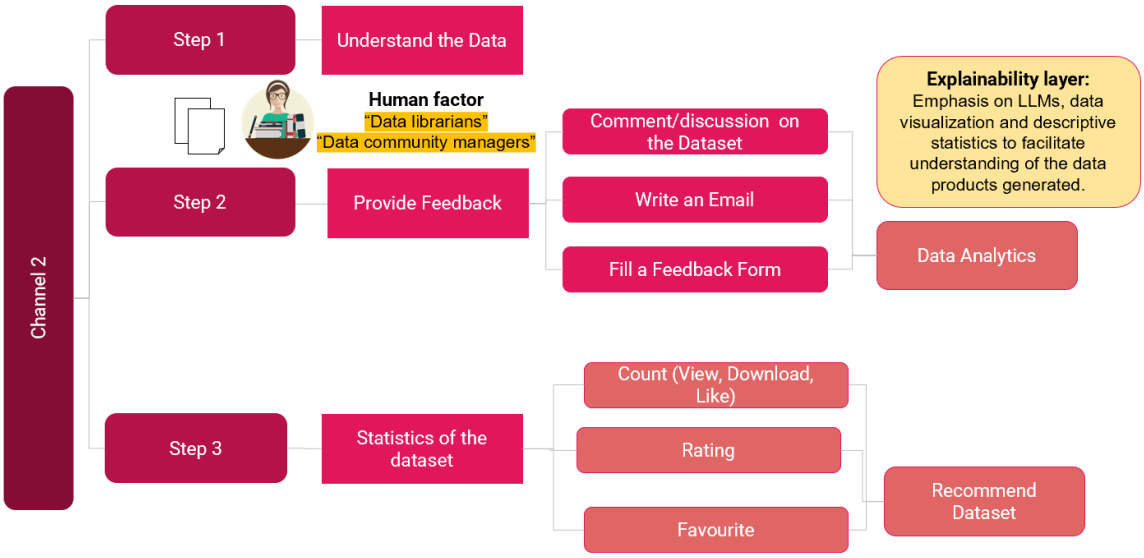

Figure 24: Channel 2 for a multi-channel inclusive feedback mechanism for diverse communities

6.2.3.4. A Plug-in for Feedback Mechanisms for Existing Open Data Portals

The motivation behind this idea is trifold. First, from the data users' side, the difficulty of providing feedback on the portal (lack of commenting features, lack of reporting tools), as well as the lack of awareness of how or why to interact with the portal, as well as what the effect of this interaction might be. In addition, users might include unofficial data sources, or wrong data. From the data holders' side, there is also little to no interaction with the data, and the risk of unofficial data sources being used is also high. The data also runs the risk of becoming easily outdated. From the data portal owners' side, the mentioned issues result in low traffic, less interaction with the website, low revenue from possible advertisement and cooperations, as well as problems with impact reporting (if relevant).

Figure 25 shows an example of a dataset interaction case at the European Data Portal. The current available interaction with the data is: download, metadata, and validation. This figure shows an example of the limited options for interactive processes regarding the in-word dataset each time.

There is no easy way to add feedback for this dataset, give recommendations on what is needed, or see other, existing commentaries from other users for this dataset.

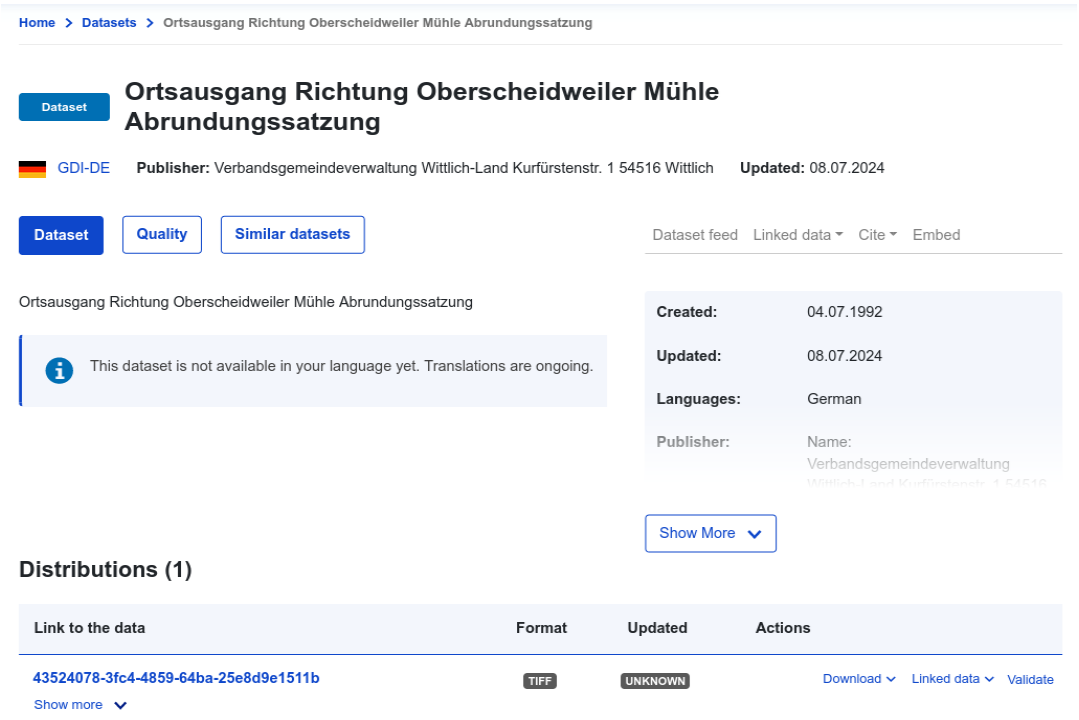

Figure 25: Example case of interaction with a dataset

The proposed solution requires that issues are integrated in the data portals, that users can add issues in a user-friendly way, and they can propose changes (in the form of pull requests), and that users are generally aware of these possible interactions. For instance, the French data portal, already implements a commenting system into the datasets, as shown in Figure 26.

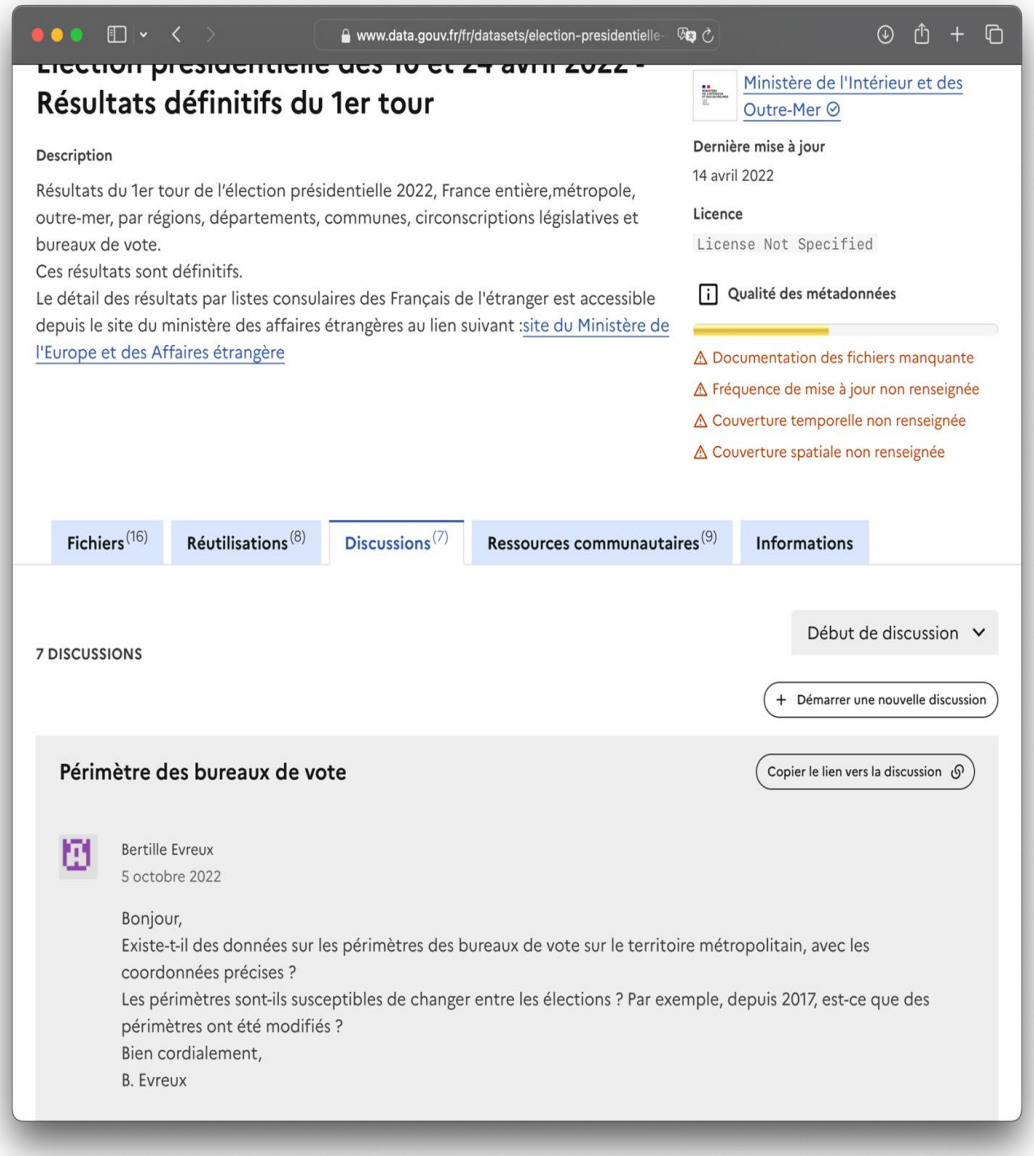

Figure 26: An example of a user interaction with a dataset on data.gouv.fr

Here we can see that under the Tab "Discussions" users can comment or start a feedback thread for this dataset. Technologies which can be used for this purpose include the W3C Web Accessibility Initiative (WAI), which provides tools for disabled people (e.g., people with hearing problems, impaired vision and other), Udata (pypi.org, 2024), a customisable social platform for open data, VCS software, for dataset version control, git, and more.

The solution proposed is a plugin for open data portals, improving user experience in an easy and efficient way of communication and interaction with data and the portal owners, as well as data quality for specific use cases. In addition, data holders (owners) can provide higher quality data for improved efficiency and/or internal operations and gain new ideas for data provision. Data portals will increase their user-friendliness, and the higher quality of datasets will result in more impact of the platform, something vital for improving the open data ecosystem.

6.3 Technology Readinessof the Proposed Solutions

The current technology readiness levels (TRLs) corresponding to each of the proposed solutions are shown in Table 12. In addition, this Table shows the TRL level these solutions aim to reach during the ODECO project timeframe. The proposed solutions which are currently in progress within ODECO are annotated in the last column of the Table.

Table 12: Technology Readiness of the Proposed Solution

Proposed Solution	Current TRL	Target TRL	ODECO Solution
Augmenting semantic interoperability for datasets	2	3	Yes
Automatic thematic annotation of open data	3	5	Yes
Intelligent validation of data on the web	3	4	Yes
A multi-channel inclusive feedback mechanism for diverse communities	2	3	No
A plug-in for feedback mechanisms for existing open data portals	1	3	Yes
A deep learning application plugin to help users easily select and customize data licenses	2	4	No
An open license decision support software	1	2	No
Open data driven citizen science applications as a mechanism for inclusion of non-governmental open data in open data portals	2	4	Yes
A chatbot guide for open data portals	2	5	No
Enhancing better interfaces and user interaction under feedback supported by AI	1	3	Yes

7 Discussion and Conclusion

7.1 Shared perspectives

The analysis in this deliverable is focused on the non-governmental data holders (NGDs) and the mechanisms they need towards the automation of data management tasks (or data creation and data sharing processes), minimising as much as possible the necessary skills an individual may possess to undertake these tasks. All the proposed solutions are aiming to enhance the interfaces through better services that will be incorporated in the current open data platforms. These solutions will improve the capacity of NGDs to create data of better quality (so more re-usable and easily validated), with enhanced feedback mechanisms (so user-driven), with easier tools for integration and description, and with better and automated tools for handling IPR and licenses.

Since this is a creative task, we deployed an innovation thinking methodology (Ideathon) to better develop the existing solutions and propose new ones. The methodological approach incorporates the dual role of the NGD in the open data ecosystem by covering also the user side from previous deliverables. The original descriptions of the current task and deliverable D4.2 stands including also the developments that surpass the understanding we had at that time. The challenges are based on the new requirements from the whole project and specifically the deliverables mentioned.

The proposed solutions address the identified challenges, and they are connected with the initial requirements of the task furthering our understanding on the services that could be provided through the open data platforms for NGD holders and users:

- New user interfaces were proposed and designed through the provision of specific services that will automate, as possible, the way NGD data holders interact with the system. Furthermore, section 6.1 identifies the existing solutions that could be used as a vehicle to incorporate the proposed services and mechanisms.
- New validation and integration services along with other ones on licencing and anonymisation as well as feedback mechanisms have been proposed and designed. The proposed solutions are a mix of different TRL levels that are also presented.

The synthesis of the motivations, requirements and action principles of the previous ODECO research was a very challenging task that resulted into four main challenges supporting the whole data sharing process (from data creation to data publication). These challenges were validated by a dedicated ODECO experts' group and a group of external experts. These challenges are a major outcome of this deliverable that drives the current and future developments towards providing support to non-governmental data holders. The whole deliverable is focused on making the open data ecosystem inclusive incorporating and proposing the necessary mechanisms to lower the barriers for non-governmental stakeholders (stakeholders that are not obliged to provide open data) to participate in the data sharing process.

The outcomes of this deliverable are presented mainly in Chapter 6 and they are divided into 2 categories: (a) the existing and (b) the proposed (developed or under development or new) technical steering mechanisms serving the above purpose. Four existing data sharing platforms were identified and analysed. Ten technical mechanisms are described that are either stand-alone or they can be incorporated into existing solutions. None of them is currently offered. The proposed technical mechanisms (both the ones in progress within the ODECO project and the ones beyond, proposed during the OD Ideathon) can be applied both as enhancements on the existing tools presented in Section 6.1, but also on the already existing open data provision platforms. The technological readiness levels (TRLs) of the proposed solutions vary from 1 to 5, with at least one solution per challenge to be at TRL4-5. This result highlights the great

developments within the ODECO project in terms of technical support and open data interoperability.

7.2 Limitations

This deliverable is focused on the stimulation of the non-governmental data holders to participate into the open data ecosystem (aiming towards the ecosystem inclusiveness) by lowering the skills and effort barriers. Even though several of the proposed steering mechanisms might also improve data usability and they could be considered contributory towards the open data ecosystem's user-drivenness, it is not the main focus of this deliverable. Although the research in the current deliverable analyses the incentives of the non-governmental data holders it does not include governance or management mechanisms and processes towards the inclusion of these stakeholder groups. The results (technical steering mechanisms) of the deliverable cover only the current understanding of the data holders' needs which may evolve in the future. Since this is a creative task, more innovative ideas and tools may be developed in due course. Not all the proposed solutions are meant to be implemented in during the lifetime of the ODECO project. Finally, this deliverable is not presenting any data stories and examples of data use and re-use that could be used as examples of value creation (or incentives) within the open data ecosystem that is deemed out of scope.

7.3 Future research directions and actions

As further research, ODECO anticipates:

- the development of the governance mechanisms to further support non-governmental data holders in terms of organisation setup.
- the generation of value creation examples to further stimulate the engagement of these stakeholder groups.
- the synthesis of all the WP2, WP3 and WP4 outcomes creating an inclusive, circular and user-driven open data ecosystem providing the necessary technical tools, and supporting the necessary governance mechanisms.

Acknowledgements

Special thank you to all the Open Data Ideathon participants for their contributions during ODECO TW5 in Samos. The participating teams, consisting of the 15 ODECO ESRs and the OpenGov2024 summer school participants (in mixed groups), exchanged valuable ideas and knowledge, several of which have been part of this deliverable.

In the Open Data Ideathon participated (in random order):

- Christos Vardatsikos
- Abdul Aziz
- Alejandra Celis Vargas
- Michail Skopelitis
- Caterina Santoro
- Maria Ioanna Maratsi
- Mohsan Ali
- Silvia Cazacu
- Liubov Pilshchikova
- Umair Ahmed
- Davide Di Staso
- Ahmad Ashraf Ahmad Shaharudin
- Dagoberto Jose Herrera Murillo
- Louise van der Peet
- Héctor Ochoa Ortiz
- Ioannis Deligiannis
- Christos Raftogiannis
- Thanos Panagopoulos
- Anna Thrasyvoulou
- Maria Karypidou
- Nikolaos Plessas
- Emmanouil Valis
- Orestis-Panagiotis Zois
- Sascha Kraus
- Georgios Papageorgiou
- María Elena López Reyes
- Jos Zuiderwijk
- Geraldina Mevoit
- Ramya Chandrasekhar
- Nikolaos Bermparis
- Dimitra-Styliani Kroustali
- Paola Azul
- Evangelia Alexopoulou
- Gennaro Angiello
- Mariia Rizun
- Kyriakos Vasileiadis
- Evangelos Georgatselos

Mentors:

- Yannis Kotsis-Yannarakis
- Charalampos Alexopoulos
- Nina Rizun
- Noella Edelmann
- Nataliya Rozbroj Jasinskaja

- Els Breedstraet
- Bastiaan van Loenen
- Antonis Fourlis
- Eleni Petra

Appendix A – The ODECO TW5 Ideathon Structure

After gathering (a) the already existing tools or mechanisms and (b) the already developed (or under development) tools within ODECO, we proceeded with the Ideathon organisation in order to gather more and innovative ideas for tools as well as to further develop the existing ones. The structure of the Ideathon held during the ODECO Training Week 5, which is part of the methodological framework of the current deliverable, is the following:

Aim: Ideathons aim to solve specific problems or challenges through brainstorming sessions, encouraging out-of-the-box thinking, and pushing boundaries. By creating an environment conducive to idea generation and collaboration, Ideathons have become popular in various fields, including technology, business, and social innovation [ref]. The aim of the ODECO TW5 Ideathon was to be a place where the teams, given the challenges identified for Open Data Sharing from a non-governmental data holder's perspective in advance, propose technical solutions (with a focus on AI-enabled solutions) to alleviate each of the identified challenges.

Participants: Organisations can benefit from involving both students and external experts in Ideathons. Students can have valuable and pure insights into operations and challenges, while external experts bring fresh perspectives from different fields. The Ideathon teams of students (mostly PhD candidates and MSc students), mixed with the ODECO ESRs, under the guidance of experienced mentors in the OD domain, and proposed solutions to the given challenges regarding data sharing for non-governmental groups.

Team size: The size of teams participating in an Ideathon can vary depending on the nature of the challenge. Smaller teams may foster more focused discussions and quick decision-making, while larger teams can bring together diverse expertise and skill sets. In our case we aimed at teams of 4 people, in a multidisciplinary setting, making sure that at least one team member has sufficient technical and technological background to facilitate the elicitation of ideas for potential technical solutions.

Nailing the Problem Statement for an Ideathon

A well-defined problem statement is crucial for a successful Ideathon, as it guides participants' creativity and leads to relevant solutions. The problem statement should be concise, specific, and effectively communicate the challenge or issue at hand. In order to formulate problem statements for Ideathons, involving subject matter experts or industry professionals can be beneficial. Their expertise can provide valuable insights and help ensure that the problem statement accurately reflects real-world challenges. By collaborating with these individuals, organizers can tap into their knowledge and experience to create problem statements that resonate with participants. Additionally, providing background information or context related to the problem is important. This additional information helps participants better understand the scope of the challenge and allows them to approach it from an informed perspective. By providing context, organizers can set the stage for innovative thinking and encourage participants to produce unique solutions. Overall, a clear and well-constructed problem statement sets the foundation for successful Ideathons by ensuring that participants fully understand the challenge they are trying to solve.

Provision of Challenge list in advance to participants

A Google Form with the challenges produced in the current Task (4.2) was shared with the participants of the Ideathon 10 days before the event, in order to give them some time to reflect and brainstorm ideas for solutions corresponding to one or more challenges. According to the information gathered by the Google Form, the participants were grouped in teams of 4 people each, each team working on one of the given Challenges. Each team was then assigned and mentored/guided by one Open Data expert on site during the event.

Ideathon Programme

Day 1

- Introduction and challenges presentation
- Teams' formation and facilitation
- Teamwork

Day 2

- Mentoring
- Teamwork
- Final presentation

Appendix B – User-centred Design

User-centred design (UCD) is an iterative design process in which designers prioritize users and their needs at every stage. In UCD, design teams engage users throughout the design process using various research and design techniques to create highly usable and accessible products. According to ISO 9241-210:2010 (See Figure 27), once a plan for the design process is established, four interconnected human-centred design activities are undertaken:

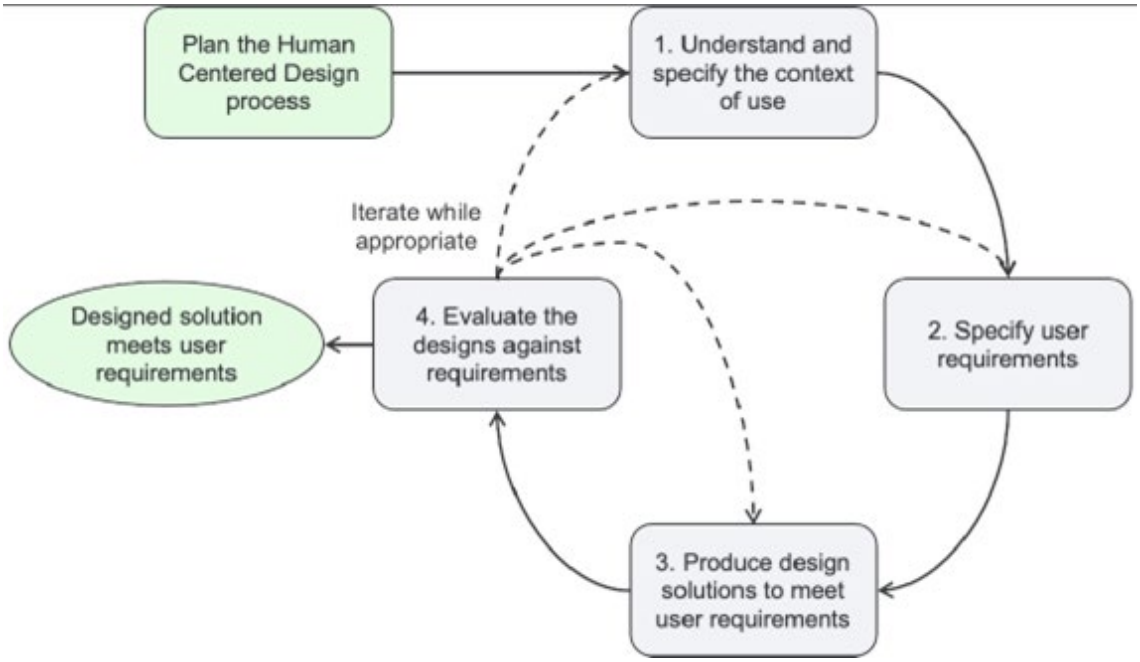

Figure 27: The user-centred design process based on ISO 9241-210:2010

Understand and specify the context of use: This activity involves identifying and describing the characteristics of the target users and other relevant stakeholders. It includes outlining the users' goals and tasks, as well as understanding the organizational, technical, and physical environments that define the system's context of use.

Specify user requirements: In this phase, user needs are identified, and functional and other requirements for the product are specified. This process considers the context of use and aligns with the system's business objectives.

Produce design solutions to meet user requirements: Design solutions are developed based on the context of use description. This begins with designing user tasks, user-system interaction, and the user interface to meet user requirements. Design solutions become more tangible through scenarios, simulations, prototypes, or mock-ups.

Evaluate the designs against requirements: User-centred evaluation is essential throughout the project's various stages, starting from an early stage. While the participation of end-users in the evaluation is highly desirable, alternative methods that do not require direct user involvement can be used when practical constraints prevent their participation.

These activities iterate until the designed solution meets user requirements.

Hybrid intelligence

Hybrid intelligence is the result of the integration between human and machine intelligence. This approach enhances human capabilities in the user interface design process. In particular,

generative artificial intelligence (GenAI) is gaining importance. This type of artificial intelligence can create products in forms such as text, images and videos by using generative models, often in response to humanly interpretable commands that describe the task to be performed or the desired output.

In 2023, a Spanish developer named Javi López created "Angry Pumpkins" (Marr, 2024), a parody of the popular game Angry Birds, entirely using artificial intelligence tools like Midjourney, DALL-E 3, and GPT-4. The development process involved creating graphical elements such as backgrounds, characters, and objects, as well as generating the source code of the game. López employed an iterative process, crafting a series of prompts to achieve the desired features without writing a single line of code. This example demonstrates the potential of generative AI to reduce time and effort in developing complex digital products. This is just one example of how GenAI is disrupting practitioners. Similarly, the academic community has responded with an emerging but expanding body of research on the implications of this technology for the design process.

The available research has explored both the positive and negative implications of introducing this new technology (Li et al., 2024). On the positive side, GenAI offers the potential to enhance the capabilities of design teams by reducing solution development time and enabling humans to focus on activities that add greater value and usability. However, if automation is not implemented correctly, it may lead to the loss of the human touch in final design products. In response, proposals such as that of Weisz et al. (2024) have emerged, emphasizing the need to establish design principles for GenAI applications. These authors propose six principles:

Responsible Design: Ensure that the AI system addresses real user problems and minimizes potential harm to users.
Design for Mental Models: Communicate effectively with users on how to interact with the AI system, considering their background and goals.
Design for Appropriate Confidence and Reliability: Help users determine when to trust or question the AI system's results by encouraging scepticism regarding quality issues, inaccuracies, biases, underrepresentation, and other potential problems.
Design for Generative Variability: Assist users in managing the capability of generative models to produce multiple distinct and varied results.
Design for Co-Creation: Enable users to influence the generative process and collaborate effectively with the AI system.
Design for Imperfection: Help users understand and work with results that may not always meet their expectations.

Stige et al. (2023) have explored the impact of artificial intelligence on the human-centred design activity cycle. In the context of the use understanding phase, AI is primarily used for the automatic summarization of research findings into artifacts commonly used in subsequent phases, such as user personas and user scenarios. During the requirements specification phase, interacting with a GenAI can assist the development team in refining the final set of system requirements. However, the application of GenAI is most concentrated in the production of design solutions to meet user requirements. In this stage, GenAI can generate prototypes that are iteratively modified to develop detailed solutions. GenAI can assist designers in progressing from low-fidelity solutions (e.g., sketches) to high-fidelity solutions (e.g., GUI code) or optimize their designs. Finally, in the evaluation phase, where designs are assessed against requirements, GenAI can automate the detection of common pitfalls, allowing the test team to focus on identifying major flaws.

Bleichner and Hermansson (2023) and Ho et al. (2024) have examined the application of GenAI in user interface design and identified the main commercial generative assistance tools: Galileo-AI, Uizard, and Midjourney. Table 13 highlights their main features. However, it is important to note that this list is not exhaustive, given the rapid emergence of new tools and the constant evolution of existing ones.

Table 13 - GenAI tools for user interface design

Tool	Description
Uizard (uizard.io)	Uizard is an AI-powered service designed to assist in creating user interfaces and interactive prototypes. In addition to generating interfaces from text prompts, Uizard can scan screenshots and wireframes, converting them into editable mock-ups with customizable components.
Galileo AI (usegalileo.ai)	Galileo AI, launched in 2023, is a tool for designing both web and mobile user interfaces based on text or image prompts. It generates high-fidelity mock-ups that can be exported as HTML code and edited in other interface design tools.
Midjourney (midjourney.com)	Midjourney is a general-purpose text-to- image model accessible via Discord. While it is primarily used for generating images from text prompts, its output can also serve as high-quality prototypes for user interface design when provided with well-crafted prompts.

References

Ahmed, U., Alexopoulos, C., Piangerelli, M., & Polini, A. (2024). BRYT: Automated keyword extraction for open datasets. Intelligent Systems with Applications, 23, 200421. https://doi.org/10.1016/j.iswa.2024.200421

Ali, M., Papageorgiou, G., Aziz, A., Loukis, E., Charalabidis, Y., & Lopez Pellicer, F., J., (2024). Towards the Development of Interoperable Open Data Ecosystems: Harnessing the Technical, Semantic, Legal, and Organizational (TSLO) Interoperability Framework. In Proceedings of the 25th Annual International Conference on Digital Government Research (dg.o '24). Association for Computing Machinery, New York, NY, USA, 909–919. https://doi.org/10.1145/3657054.3657160

Anderson, J., Sarkar, D., & Palen, L. (2019). Corporate Editors in the Evolving Landscape of OpenStreetMap. *ISPRS International Journal of Geo-Information*, *8*(5), Article 5. https://doi.org/10.3390/ijgi8050232

Atenas, J., Havemann, L., & Priego, E. (2015). Open Data as Open Educational Resources: Towards Transversal Skills and Global Citizenship. *Open Praxis*, *7*(4). https://doi.org/10.5944/openpraxis.7.4.233

Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society*, *2*(2), 2053951715594634. https://doi.org/10.1177/2053951715594634

Bleichner, A., & Hermansson, N. (2023). Investigating the usefulness of a generative AI when designing user interfaces (Dissertation). https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-504251

Boyles, J. L. (2020). Laboratories for news? Experimenting with journalism hackathons. *Journalism*, *21*(9), 1338–1354. https://doi.org/10.1177/1464884917737213

BRYT Keyword Extraction GitHub Repository: GitHub - umairahmedq/BRYT-AKE

Cambridge Dictionary. (2024). *Motivation*. https://dictionary.cambridge.org/dictionary/english/motivation

Carrara, W., Chan, Wae-San., Fischer, Sander., & Van-Steenbergen, E. Creating value through open data: Study on the impact of re-use of public data resources. European Commission, 2015.https://data.europa.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf

Celis Vargas, A., Magnussen, R., Mulder, I., & Larsen, B. (2023). Towards a framework for Open Data literacy in education: A systematic mapping review of Open Data skills and learning approaches. *Interaction Design and Architecture(s)*, *57*, 133–151. https://doi.org/10.55612/s-5002-057-008

Charalabidis, Yannis & Zuiderwijk, Anneke & Alexopoulos, Charalampos & Janssen, Marijn & Lampoltshammer, Thomas & Ferro, Enrico. (2018). The World of Open Data - Concepts, Methods, Tools and Experiences.

Chattapadhyay, S. (2014). Access and use of government data by research and advocacy organisations in India: A survey of (potential) open data ecosystem. *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance*, 361–364. https://doi.org/10.1145/2691195.2691262

Chicaiza, J., Piedra, N., Lopez-Vargas, J., & Tovar-Caro, E. (2017). Recommendation of open educational resources. An approach based on linked open data. *2017 IEEE Global Engineering Education Conference (EDUCON)*, 1316–1321. https://doi.org/10.1109/EDUCON.2017.7943018

Concilio, G., & Mulder, I. (2018). *Open4Citizens Policy brief.* OPEN4CITIZENS - Empowering citizens to make meaningful use of open data. http://open4citizens.eu/wp-content/uploads/2016/01/O4C_D4.8_30.06.2018_Final.pdf

Cook, K., Cakirlar, C., Goddard, T., Demuth, R. C., & Wells, J. (2018). Teaching Open Science: Published Data and Digital Literacy in Archaeology Classrooms. *Advances in Archaeological Practice*, *6*(2), 144–156. https://doi.org/10.1017/aap.2018.5

Corbett, J., Templier, M., & Takeda, H. (2018, January 3). The Making of a "Top" Open Data City: A Case Study of Edmonton's Open Data Initiative. *Proceedings of the 51st Hawaii International Conference on System Sciences*. https://doi.org/10.24251/HICSS.2018.308

Corcho, O., Jiménez Blanco, J., Morote, C., and Simperl, E. (2022). Data.europa.eu and Citizen-generated DataOpportunities and challenges associated to the inclusion of citizen-generated data in data.europa.eu. Luxembourg: Publications Office of the European Union. https://data.europa.eu/sites/default/files/report/data.europa.eu_Report_Citizen-generateddataondata_europa_eu.pdf

Coughlan, T. (2020). The use of open data as a material for learning. *Educational Technology Research and Development*, *68*(1), 383–411. https://doi.org/10.1007/s11423-019-09706-y

Creative Commons. (2023). *CC BY-NC-SA 4.0 Legal Code | Attribution-NonCommercial-ShareAlike 4.0 International | Creative Commons*. https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode

Datasette. (n.d.). Datasette: An open source multi-tool for exploring and publishing data. https://datasette.io

Davies, T. (2011). Open Data: Infrastructures and ecosystems. *ACM Web Science Conference*.

Davis, K., & Shneyer, E. (2020). Computer science as a tool for developing future civic change-makers. *SIGCSE 2020 - Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, 1344. https://doi.org/10.1145/3328778.3372636

Enaholo, P. (2017). Beyond mere advocacy: CSOs and the role of intermediaries in Nigeria's open data ecosystem. *The Social Dynamics of Open Data*. https://zenodo.org/records/1117797

Enaholo, P., & Dina, D. (2020). Journalists and the intermediation of open data: A Nigerian perspective. In *Situating Open Data: Global Trends in Local Contexts* (pp. 31–49). African Minds. https://www.scienceopen.com/hosted-document?doi=10.47622/978-1-928502-12-8_2

What is Epicollect5 | Epicollect5 User Guide. (n.d.). What is Epicollect5 | Epicollect5 User Guide. https://docs.epicollect.net

Falk, J. (2022). *How Game Jams and Hackathons Accelerate Design Processes* (arXiv:2205.04966). arXiv. https://doi.org/10.48550/arXiv.2205.04966

GO FAIR initiative. (n.d.). FAIR principles - GO FAIR. GO FAIR. https://www.go-fair.org/fair-principles/

Goodchild, M.F. Citizens as sensors: the world of volunteered geography. (2007). GeoJournal 69, 211–221. https://doi.org/10.1007/s10708-007-9111-y

González-Zapata, F., & Heeks, R. (2015, May). Understanding multiple roles of intermediaries in Open Government Data. *Proceedings of the 13th International Conference on Social Implications of Computers in Developing Countries*.

Gray, J., Bounegru, L., & Chambers, L. (2012). *The Data Journalism Handbook* (1st ed.). O'Reilly Media, Inc.

Herrera-Murillo, D. J., Aziz, A., Nogueras-Iso, J., & Lopez-Pellicer, F. J. Analysing user involvement in open government data initiatives. (2022). in *Proc. Int. Conf. Theory Pract. Digit.* Libraries, pp. 175–186. https://doi.org/10.1007/978-3-031-16802-4_14

Herrera-Murillo, D. J., Ochoa-Ortiz, H., Ahmed, U., López-Pellicer, F. J., Re, B., Polini, A., and Nogueras-Iso, J. (2024). Process Analysis in Humanitarian Voluntary Geographic Information: the case of the HOT Tasking Manager, *AGILE GIScience Ser.*, 5, 5. https://doi.org/10.5194/agile-giss-5-5-2024

Ho, C., Liu, X., Qiu, Y., and Yang, S. (2024). Research on Innovative Applications and Impacts of Using Generative AI for User Interface Design in Programming Courses. In Proceedings of the 2024 International Conference on Information Technology, Data Science, and Optimization (I-DO '24). Association for Computing Machinery, New York, NY, USA, 68–72. https://doi.org/10.1145/3658549.3658566

Implementing regulation - 2023/138 - EN - EUR-Lex. (n.d.). EUR-Lex — Access to European Union law — choose your language. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2023.019.01.0043.01.ENG

ISO 9241-210:2010, (2010). Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems. Standard. International Organization for Standardization. Geneva, CH. https://www.iso.org/standard/77520.html
Hou, Y., & Wang, D. (2017). Hacking with NPOs: Collaborative Analytics and Broker Roles in Civic Data Hackathons. *Proceedings of the ACM on Human-Computer Interaction*, *1*(CSCW), 53:1-53:16. https://doi.org/10.1145/3134688
Jaskiewicz, T., Mulder, I., Morelli, N., & Pedersen, J. (2019). Hacking the hackathon format to empower citizens in outsmarting "smart" cities. *Interaction Design and Architecture(s)*, *43*, 8–29. https://doi.org/10.55612/s-5002-043-001
Kubler, S., Robert, J., Neumaier, S., Umbrich, J., & Le Traon, Y. (2018). Comparison of metadata quality in open data portals using the Analytic Hierarchy Process. *Government Information Quarterly*, 35(1), 13 - 29. https://doi.org/10.1016/j.giq.2017.11.003
Kitsios, F., & Kamariotou, M. (2018). Open data hackathons: An innovative strategy to enhance entrepreneurial intention. *International Journal of Innovation Science*, *10*(4), 519–538. https://doi.org/10.1108/IJIS-06-2017-0055
Li, J., Cao, H., Lin, L., Hou, Y., Zhu, R., & El Ali, A. (2023). User Experience Design Professionals' Perceptions of Generative Artificial Intelligence. Proceedings of the CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3613904.3642114
Lodato, T. J., & DiSalvo, C. (2016). Issue-oriented hackathons as material participation. *New Media & Society*, *18*(4), 539–557. https://doi.org/10.1177/1461444816629467
Meijer, A., & Potjer, S. (2018). Citizen-generated open data: An explorative analysis of 25 cases. *Government Information Quarterly*, *35*(4), 613–621. https://doi.org/10.1016/j.giq.2018.10.004
Linked Open Vocabularies (LOV). (n.d.). Linked Open Vocabularies. https://lov.linkeddata.es/dataset/lov/
Loukis, E. & Saxena, Stuti & Rizun, Nina & Maratsi, Maria & Ali, Mohsan & Alexopoulos, Charalampos. (2023). ChatGPT Application vis-a-vis Open Government Data (OGD): Capabilities, Public Values, Issues and a Research Agenda. 10.1007/978-3-031-41138-0_7.
Loutsaris, Michalis Avgerinos & Alexopoulos, Charalampos & Maratsi, Maria & Charalabidis, Yannis. (2023). Semantic Interoperability for Legal Information: Mapping the European Legislation Identifier (ELI) and Akoma Ntoso (AKN) Ontologies. 41-53. 10.1145/3614321.3614327.
Maratsi, Maria & Ahmed, Umair & Alexopoulos, Charalampos & Charalabidis, Yannis & Polini, Andrea. (2024 b). Towards Cross-Domain Linking of Data: A Semantic Mapping of Cultural Heritage Ontologies. 165-176. 10.1145/3657054.3657077
Maratsi, Maria & Alexopoulos, Charalampos & Charalabidis, Yannis. (2024 d). A Structured Analysis of Domain-Specific Linked Open Vocabularies (LOV): Indicators for Interoperability and Reusability. 10.1007/978-3-031-56478-9_10.
Maratsi, Maria, Alexopoulos, Charalampos & Charalabidis, Yannis. (2024 a). On the Semantic Analysis of Open (Government) Data Portals' Metadata Provision and Schema. (to be published in a few weeks in the ICEGOV'24 ACM proceedings)
Maratsi, Maria, Gialoussi, Nina, Alexopoulos, Charalampos & Charalabidis, Yannis. (2024 c). A Proposed Methodology for Sub-Ontology Development in Comprehensive Scientific Investigation Methods and Tooling (to be published in a few weeks in the MTSR'24 Springer proceedings)
Maratsi, Maria & Lachana, Zoi & Alexopoulos, Charalampos & Charalabidis, Yannis. (2023). A functional and semantic analysis of artifact representation schemata in folklore museum websites. International Journal of Metadata, Semantics and Ontologies. 16. 105-117. 10.1504/IJMSO.2023.135339.
Marr, B. (2024). Generative AI in Practice: 100+ Amazing Ways Generative Artificial Intelligence is Changing Business and Society. John Wiley & Sons.
Meyer, P. (2002). *Precision Journalism: A Reporter's Introduction to Social Science Methods* (4th ed.). Rowman & Littlefield Publishers.

Navalkha, C. (2021). *Data for Housing Justice: Examining Activists' Use of Open Government Data for Housing Justice in Boston, MA and New York, NY* [Thesis, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/138973
ODK - Collect data anywhere. (n.d.). ODK - Collect data anywhere. https://getodk.org
OpenstreetMap Foundation. (2024). *Corporate Members—OpenStreetMap Foundation*. https://osmfoundation.org/wiki/Corporate_Members
OpenStreetMap Wiki. (2024). *Organised Editing/Activities—OpenStreetMap Wiki*. https://wiki.openstreetmap.org/w/index.php?title=Organised_Editing/Activities&oldid=2656488
Palova, D., & Vejacka, M. (2022). Experience with Open Data in Project Based Learning. *2022 45th Jubilee International Convention on Information, Communication and Electronic Technology, MIPRO 2022 - Proceedings*, 689–694. https://doi.org/10.23919/MIPRO55190.2022.9803595
Papageorgiou, G., Euripides, L., Magnussen, R., & Charalabidis, Y. (2023). Open data journalism: A domain mapping review. *Proceedings of the 16th International Conference on Theory and Practice of Electronic Governance*, 159–166. https://doi.org/10.1145/3614321.3614340
Park, C. H., Longo, J., & Johnston, E. W. (2020). Exploring Non-State Stakeholder and Community-Led Open Governance: Beyond the Three Pillars of Open Government. *Public Performance & Management Review*, *43*(3), 587–612. https://doi.org/10.1080/15309576.2019.1677253
Pellegrino, M., & Antelmi, A. (2023). At School of Open Data: A Literature Review: *Proceedings of the 15th International Conference on Computer Supported Education*, 172–183. https://doi.org/10.5220/0011747500003470
Pence, H. E., Williams, A. J., & Belford, R. E. (2015). New Tools and Challenges for Chemical Education: Mobile Learning, Augmented Reality, and Distributed Cognition in the Dawn of the Social and Semantic Web. In *Chemistry Education: Best Practices, Opportunities and Trends*. https://doi.org/10.1002/9783527679300.ch28
Poikola, A., Kola, P., & Hintikka, K. A. (2010). *Public data: An introduction to opening information resources*. Ministry of Transport and Communications.
Publications Office of the European Union. Open data maturity report 2023. Publications Office, LU, 2023. https://data.europa.eu/sites/default/files/odm2023_report.pdf
QField - Efficient field work built for QGIS. (n.d.). QField - Efficient field work built for QGIS. https://qfield.org
Ricker, B., Cinnamon, J., & Dierwechter, Y. (2020). When open data and data activism meet: An analysis of civic participation in Cape Town, South Africa. *Canadian Geographies / Géographies Canadiennes*, *64*(3), 359–373. https://doi.org/10.1111/cag.12608
Rogers, S. (2008, December 18). Turning official figures into understandable graphics, at the press of a button. *The Guardian*. https://www.theguardian.com/help/insideguardian/2008/dec/18/unemploymentdata
Rogers, S. (2010, August 13). Florence Nightingale, datajournalist: Information has always been beautiful. *The Guardian*. https://www.theguardian.com/news/datablog/2010/aug/13/florence-nightingale-graphics
Rogers, S. (2011, September 26). *The first Guardian data journalism: May 5, 1821*. The Guardian. http://www.theguardian.com/news/datablog/2011/sep/26/data-journalism-guardian
Romero, M., Usart, M., & Ott, M. (2015). Can Serious Games Contribute to Developing and Sustaining 21st Century Skills? *Games and Culture*, *10*(2), 148–177. https://doi.org/10.1177/1555412014548919
Runeson, P., Olsson, T., Linåker, J., (2021), Open Data Ecosystems — An empirical investigation into an emerging industry collaboration concept, Journal of Systems and Software, Volume 182, 111088, ISSN 0164-1212, https://doi.org/10.1016/j.jss.2021.111088.
Saddiqa, M., Larsen, B., Magnussen, R., Rasmussen, L. L., & Pedersen, J. M. (2019). Open data visualization in danish schools: A case study. *Journal of WSCG*, *2019*(WSCG2019CS), 17–26. https://doi.org/10.24132/CSRN.2019.2902.2.3

Saddiqa, M., Magnussen, R., Larsen, B., & Pedersen, J. M. (2021). Digital innovation in education: Perspectives, opportunities and challenges of educational open data and sensor data. *CEUR Workshop Proceedings*, *2991*, 74–83.
Saddiqa, M., Rasmussen, L., Magnussen, R., Larsen, B., & Pedersen, J. M. (2019). Bringing open data into Danish schools and its potential impact on school pupils. *Proceedings of the 15th International Symposium on Open Collaboration, OpenSym 2019*. https://doi.org/10.1145/3306446.3340821
Salamon, L. M., & Anheier, H. K. (1992). In search of the non-profit sector. I: The question of definitions. *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, *3*(2), 125–151. https://doi.org/10.1007/BF01397770
Shaharudin, A., van Loenen, B., & Janssen, M. (2023). Towards a Common Definition of Open Data Intermediaries. *Digital Government: Research and Practice*, *4*(2), 6:1-6:21. https://doi.org/10.1145/3585537
Stige, Å., Zamani, E.D., Mikalef, P., & Zhu, Y. (2023). Artificial intelligence (AI) for user experience (UX) design: a systematic literature review and future research agenda. Information Technology & People. https://doi.org/10.1108/ITP-07-2022-0519
Thematic Annotation of open Government data, GitHub repository created by IAAA lab, University of Zaragoza, Spain, https://github.com/IAAA-Lab/Thematic_Annotation_of_Government_Data, accessed on 26/09/2024
udata. (n.d.). PyPI. https://pypi.org/project/udata/
Vallejo-Figueroa, S., Rodriguez-Artacho, M., Castro-Gil, M., & Cristobal, E. S. (2018). Using text mining and linked open data to assist the mashup of educational resources. *2018 IEEE Global Engineering Education Conference (EDUCON)*, 1606–1611. https://doi.org/10.1109/EDUCON.2018.8363427
van Loenen, B., Vancauwenberghe, G., & Crompvoets, J. (Eds.). (2018). *Open Data Exposed*. T.M.C. Asser Press. https://link.springer.com/book/10.1007/978-94-6265-261-3
van Loenen, B., Zuiderwijk, A. ., Vancauwenberghe, G., Lopez-Pellicer, F. J. ., Mulder, I., Alexopoulos, C. ., Magnussen, R., Saddiqa, M., Dulong de Rosnay, M., Crompvoets, J., Polini, A., Re, B., & Casiano Flores, C. (2021). Towards value-creating and sustainable open data ecosystems: A comparative case study and a research agenda. JeDEM - EJournal of EDemocracy and Open Government, 13(2), 1–27. https://doi.org/10.29379/jedem.v13i2.644 Weisz, J.D., He, J., Muller, M., Hoefer, G., Miles, R., & Geyer, W. (2024). Design Principles for Generative AI Applications. Proceedings of the CHI Conference on Human Factors in Computing Systems. https://doi.org/10.48550/arXiv.2301.05578
Wolff, A., Cavero Montaner, J. J., & Kortuem, G. (2016). Urban Data in the primary classroom: Bringing data literacy to the UK curriculum. *The Journal of Community Informatics*, *12*(3). https://doi.org/10.15353/joci.v12i3.3278
Zaman, H. B., Baharin, H., & Ahmad, A. (2021). Fusion Technology and Visualisation to Share STEM Data Using PETS Robots (i-COMEL) for Open Data Readiness Amongst Primary School Children. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 13051 LNCS*. https://doi.org/10.1007/978-3-030-90235-3_16