# Towards a sustainable Open Data ECOsystem

## D3.2

## Closing the cycle: Promoting open data users' contribution from a technical perspective

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.

| 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** | D3.2 |
| **Deliverable Title** | Closing the cycle: Promoting open data users' contribution from a technical perspective |
| **Dissemination Level** | Public |
| **Deliverable Leader** | University of Camerino (UNICAM) |
| **Submission Date** | 11-04-2024 |
| **Author** | Andrea Polini (UNICAM) (Task Lead)<br>Umair Ahmed (UNICAM)<br>Abdul Aziz (UNIZAR)<br>Dagoberto Herrera (UNIZAR)<br>Mohsan Ali (UAEGEAN) |

**Document history**

| Version # | Date | Description (Section, page number) | Author & Organisation |
|---|---|---|---|
| V0.1 | 05-02-2024 | First draft | Andrea Polini (UNICAM) (Task Lead)<br>Umair Ahmed (UNICAM)<br>Abdul Aziz (UNIZAR)<br>Dagoberto Herrera (UNIZAR)<br>Mohsan Ali (UAEGEAN) |
| V0.1a | 13-02-2024 | Review | Bastiaan van Loenen (TU DELFT)<br>Manolis Ktistakis (HUFFPOST) |
| V0.2 | 12-03-2024 | Second draft. Processed feedback obtained from the peer review process. Second draft. Integration of feedback received from beneficiaries and collected during Steering Supervisory Board meeting. | Andrea Polini (UNICAM) (Task Lead)<br>Umair Ahmed (UNICAM)<br>Abdul Aziz (UNIZAR)<br>Dagoberto Herrera (UNIZAR)<br>Mohsan Ali (UAEGEAN) |
| V0.2 | 26-03-2024 | Review | Bastiaan van Loenen (TU DELFT)<br>Manolis Ktistakis (HUFFPOST) |
| V0.3 | 03-04-2024 | Submitted version. Processed feedback obtained from the second review round. | Andrea Polini (UNICAM) (Task Lead)<br>Umair Ahmed (UNICAM)<br>Abdul Aziz (UNIZAR)<br>Dagoberto Herrera (UNIZAR)<br>Mohsan Ali (UAEGEAN) |
| V0.4 | DD-MM-YYYY | Approval | Bastiaan van Loenen, TUD |
| V1.0 | DD-MM-YYYY | Final editing | Danitsja van Heusden, TUD |

# Table of Contents



**List of figures**

**List of tables**

## Abbreviations

| | |
|---|---|
| D | Deliverable |
| ESR | Early Stage Researcher |
| M | Milestone |
| ODECO | Open Data ECOsystem |
| WP | Work Package |
| ODP | Open Data Portal |
| EU | European Union |
| AI | Artificial Intelligence |
| CI | Collective Intelligence |
| EDP | European Data Portal |
| FDP | French Data Portal |
| SDP | Swedish Data Portal |

| 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 |
| **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 |

# 1 Introduction

The significant social, economic, and environmental benefits associated with open data are increasingly recognised across institutions and sectors [1]. The fundamental principle of open data revolves around facilitating the free use, sharing, and accessibility of data in a variety of formats [2]. Since the start of the open data movement, numerous public datasets have been made available across Europe, fostering the development of innovative applications and generating valuable knowledge. Open data has become a crucial component of the European digital agenda, leading Member States to align their national strategies with this paradigm [3].

Open data ecosystems (ODEs) focus on the interplay between social and technological elements that influence the effectiveness of open data initiatives. The ecosystem perspective plays a key role in reusing public sector information. Open Data Portals (ODPs) are part of the infrastructure of ODEs that facilitates open data access. Governments at all levels are increasingly launching open data initiatives and establishing portals dedicated to distributing open data in reusable formats, allowing citizens to use it for a variety of purposes [1]. As a result, numerous open data repositories, catalogues, and websites have emerged to serve this purpose. ODPs function as online repositories that provide detailed descriptions of datasets, including key attributes such as authorship, provenance, and licence [4]. These catalogues simplify the exploration and management of metadata records, providing valuable information about datasets available for download in various distribution formats. Open data initiatives anticipate that the publication of open data through ODPs will stimulate demand for high-quality data, thereby improving the overall quality of ODPs. In addition, the publication of public sector data serves as an important driver behind the ongoing movement to open government data through open data portals [5].

While the number of open government data initiatives has notably increased in the past decade, the actual impact of these open data initiatives is difficult to measure/track [1]. The initially anticipated high and tangible economic impact has proven to be more gradual and not immediately apparent, as it often manifests subtly or is hidden [3]. The advantages of open government data manifest through insights that enhance research and guide decision-making, services in the form of applications and websites, and improved products and processes that boost productivity, efficiency, well-being, health, safety, and sustainability [3]. Quantifying these benefits to articulate the economic impact of open government data is challenging because the most significant advantages are often indirect [3]. Measuring the impact of open data poses difficulties, considering the diverse areas in which open data can be applied for beneficial purposes and the challenge of selecting universal indicators to gauge its impact, among other factors [6]. Nonetheless, support to circularity permits to enlarge the spectrum of possible users with clear beneficial impacts on the economic side.

The provisioning of data sets generally follows a linear approach where one stakeholder, typically a public administration, publishes a data set that is then consumed by a different stakeholder. Data use and transformation generally produces new data. For instance, open geographic data can be used by companies delivering services to support mobility. The mentioned services generate highly valuable data, as the one related to people moving within a city. However, such data are generally not fed back in the open data ecosystem by the service providing company. This may partly be due to the lack of mechanism in the open data ecosystem to support users of OGD to deliver back to the ecosystem facilitating a circular scenario.

ODECO's WP3, "From a linear to a circular open data ecosystem", investigates the topic of circularity in open data ecosystems from different technical perspectives. Task 3.1 - Closing the cycle: understanding potential contributions of open government data users to the open data ecosystem focused on potential contribution of eight types of OGD users and made an in-depth analysis on how OGD circularity could derive value for the different stakeholders and the ecosystem as a whole. The conclusion of task 3.1 invites the development of technical aspects that facilitate ODE participants to bring value back to the ecosystem, including interfaces for circular open data portals, feedback tools for open data, and artificial and collective intelligence systems to interact directly with the ecosystem.

It is important to keep in mind that the provision of OGD on its own cannot fulfil the desired results or objectives that are intended to be achieved from OGD initiatives. The first US Chief Technology Officer presented the idea of open data portal not just for a data repository instead of that it should be a way "to foster a thriving ecosystem that creates opportunities in research and development" [7]. To create opportunities from the open datasets, OGD value-chain adoption is necessary, in which open data analysed, processed, and re-used by the ecosystem of producers, infomediaries, and users [8]. The relationship between the involved stakeholders in the value chain is also important, in this way one can understand what values and contributions they can create in the open data ecosystems.

For closing the lifecycle of the open data from a technical perspective, it is required to understand what the requirements/needs of circular open data ecosystems are: data sharing, data reuse, values and benefits (and fair distribution among contributors/stakeholders), community engagement, innovation, feedbacks loops, open standards and formats (helps in interoperability), and data tools and methods and analysis frameworks (to ease communication among stakeholders to share information and knowledge and also data some time for re-useability purposes) [18], [9], [10].

In this context, ODECO Task 3.2 introduces technical means that should make it easy to deliver value back to the open data ecosystem by applying the principles of circularity. This includes four components: (1) designing user interfaces for open data portals where stakeholders can readily materially add value to the ecosystem, (2) researching appropriate feedback tools, (3) assessing the tools and technologies used for the analysis of different kinds of open government data and non-government dataset analysis, and (4) assessing the technical requirements of artificial and collective intelligence systems to directly interact with the open data ecosystem.

Following the comprehensive introduction, this document aims to navigate the intricacies of circular open data ecosystems, starting with foundational concepts like the circular economy and the DIKW chain in Chapter 2 to lay the groundwork. Chapter 3 advances into a more technical analysis, identifying the requirements for circularity through the lens of user interface design, feedback mechanisms, and analytical tools, alongside integrating AI and (CI) through APIs. Chapter 4 critically evaluates the alignment of current Open Data Portals (ODPs) with these identified requirements, assessing their performance across various dimensions. Building on this assessment, Chapter 5 proposes targeted strategies to enhance the circularity of open data ecosystems, focusing on design modifications, feedback model conceptualization, and the incorporation of cutting-edge technologies through APIs. This exploration provides a comprehensive perspective, equipping readers with the insights and tools necessary to support a more sustainable and circular open data landscape.

# 2 Background

## 2.1 Circular open data ecosystems

Van Loenen et al. [11] argued that circularity is one of the key ingredients of a sustainable open data ecosystem. Current open data ecosystems tend to follow a linear structure, where users extract value without actively contributing to the system. This results in a skewed distribution of benefits, as only a few stakeholders benefit from open data, and often not those who have invested the most effort, time and money [12]. This inequality can make stakeholders reluctant to contribute.

In contrast to a linear open data ecosystem, a circular open data ecosystem works when stakeholders can use the data for as long as possible [13]. This circular approach allows users to extract maximum value from the data and derived services. Most importantly, stakeholders bring additional value to the ecosystem, for example by creating value-added products based on the open data or by evaluating data use processes with data providers [14]. This raises the question of how to realise circularity in the technical implementation of open government data initiatives.
When data is opened by data providers and subsequently processed to enable intermediaries, value-adding resellers, and enrichers to retain processed data for an extended period of time, an OD ecosystem is considered circular [13]. The utmost value is extracted by users from derived data and services. In the final step of the cycle, stakeholders contribute further value to the ecosystem [39]. The open data ecosystem is circular when it builds upon a complex network of values produced by diverse stakeholders those who are creating valuable products, information, and services. To generate these values, analysis of the data is required to generate the required information and knowledge.

## 2.2 Circular economy

Circularity is inspired from the concept of circular economy. Pressure on natural resources has motivated the society to rethink the way to create the economic value. The notion of circular economy offers an alternative to the conventional linear economy approach. It advocates the extended use of resources, extracting maximum value during their use, and then recovering and regenerating products and materials at the end of their useful life [15]. Although the concept initially originated in the context of natural resource management, the underlying principle of maximising value through cyclical processes can be extended to the digital economy, where data emerges as the fundamental resource.

The circular economy system diagram is one of the most popular graphical representations of the circular economy [16]. This diagram visually represents the continuous flow of materials within the circular economy, comprising two main cycles: the technical cycle and the biological cycle. Within the biological cycle, biodegradable materials are returned to the Earth. In contrast, the technical cycle ensures that products remain in circulation within the economy through practices such as reuse, repair, remanufacturing and recycling. This approach aims to keep materials in active use, preventing them from becoming waste. The diagram in Figure 1 highlights the technical cycle of the Butterfly Diagram. Below, the fundamental concepts of each process loop are outlined [17].

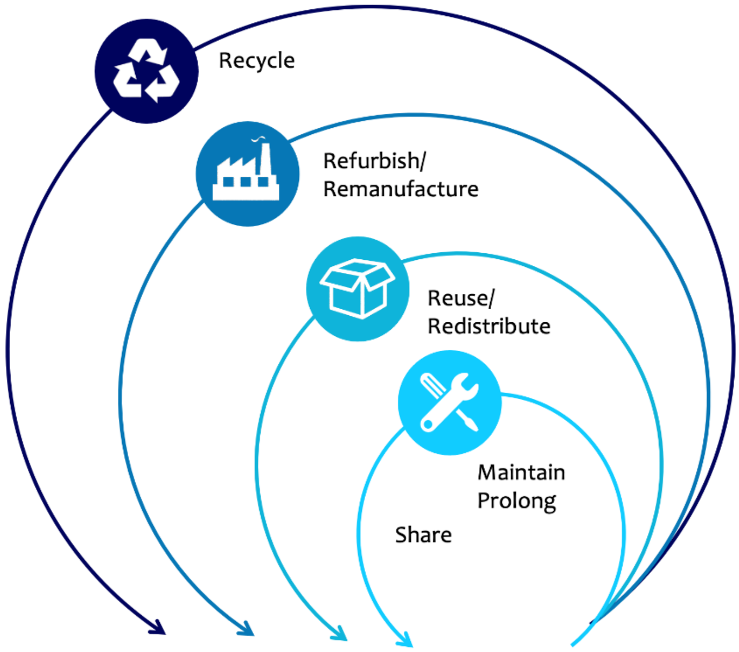


*Figure 1: Adaptation of the Circular Economy Technical Cycle Butterfly diagram by the Ellen MacArthur Foundation [11]*

- **Sharing**: this process encourages collaboration between users to improve the utilisation and value derived from their products.
- **Maintenance**: process aimed at prolonging the useful life of a product, preserving its maximum economic value and preventing the deterioration of its value due to possible damage.
- **Reuse**: this process emphasises the repetitive use of a product for the same or different purposes.
- **Redistribution**: this process keeps products in use and prevent them from becoming waste. Products that are unnecessary in their primary market are redirected to other markets where they can remain valuable.
- **Refurbishment**: this process involves returning products to a renewed working condition, which includes repairing or replacing components, upgrading specifications and improving aesthetic appearance, thus preserving their value.
- **Remanufacturing**: this process occurs when products can no longer be circulated in their current state and require more intensive work to be reused.
- **Recycling**: this process is the collection and reprocessing of materials to create new products or components, closing the cycle of material use.

The circular economy pays attention to the value chain of materials and keeps them active in one of the available value cycles. The following subsections explore concepts related to the value of data that can be useful for the design of circular technical solutions in ODEs: (1) the DIKW chain, (2) the Data Life Cycle (3) and Data Science as a profession. All these aspects should be considered in shaping a circular framework for open data. General concepts are reported here and successively referred in the document.

### 2.3 The DIKW chain

The DIKW chain (See Figure 2), also represented as a pyramid, is commonly used to represent the functional relationships between data, information, knowledge and wisdom. Usually, the movement through the stages of the chain involves changes in key variables. As the chain progresses, variables such as value, meaning, applicability, and human input increase while the volume of computer input decreases.

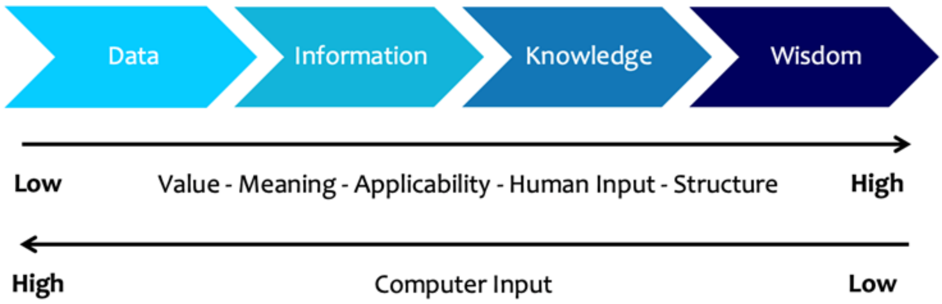


*Figure 2: The DIKW chain*

A) **Data:** the content of the first stage is composed of raw symbols that represent properties of objects, events, and their environments that result from the observation.
Example: a sensor collects raw data on the concentration of certain pollutant gases in an avenue.

B) **Information:** this stage adds meaning about the "who", "what", "where", "when", or "how many" around the data.
Example: an air monitoring centre receives readings from all sensors in a city, aggregates them and displays relevant air quality indicators in an organised dashboard.

C) **Knowledge:** this stage results from analysing and interpreting information to provide understanding of "how" and "why" certain events occur in form of patterns, trends, or relationships
Example: officials at the centre interpret the dashboard and identify a pattern of high concentrations of pollutant gases in the city centre on weekday mornings.

D) **Wisdom:** this last stage translates into the ability to decide the most appropriate course of action from the underlying knowledge.
Example: the findings on high concentrations are transferred to the city council to decide on actions to reduce vehicle traffic in the areas affected by air pollution.

### 2.4 The Data Life Cycle

The data life cycle identifies the sequence of activities that data goes through from its initial generation to its deletion or archiving at the end of its useful life. There are numerous efforts to characterise this life cycle with slight variations in aspects such as the terminology used, the emphasis on certain phases and the domain of application or context [10], [18], [19], [20].

To better understand the tools and technologies in the open data ecosystem, the open data life-cycle stages can help us. Charalabidis et al. proposed a multi-stage open data life cycle (ODLC) in which different tools and methods were suggested [10]. These tools and methods are helpful in the creation of values from the open data. Figure 3 attempts to identify and Table 1 attempts to describe phases common to various efforts made to understand the open data life cycle (ODLC):

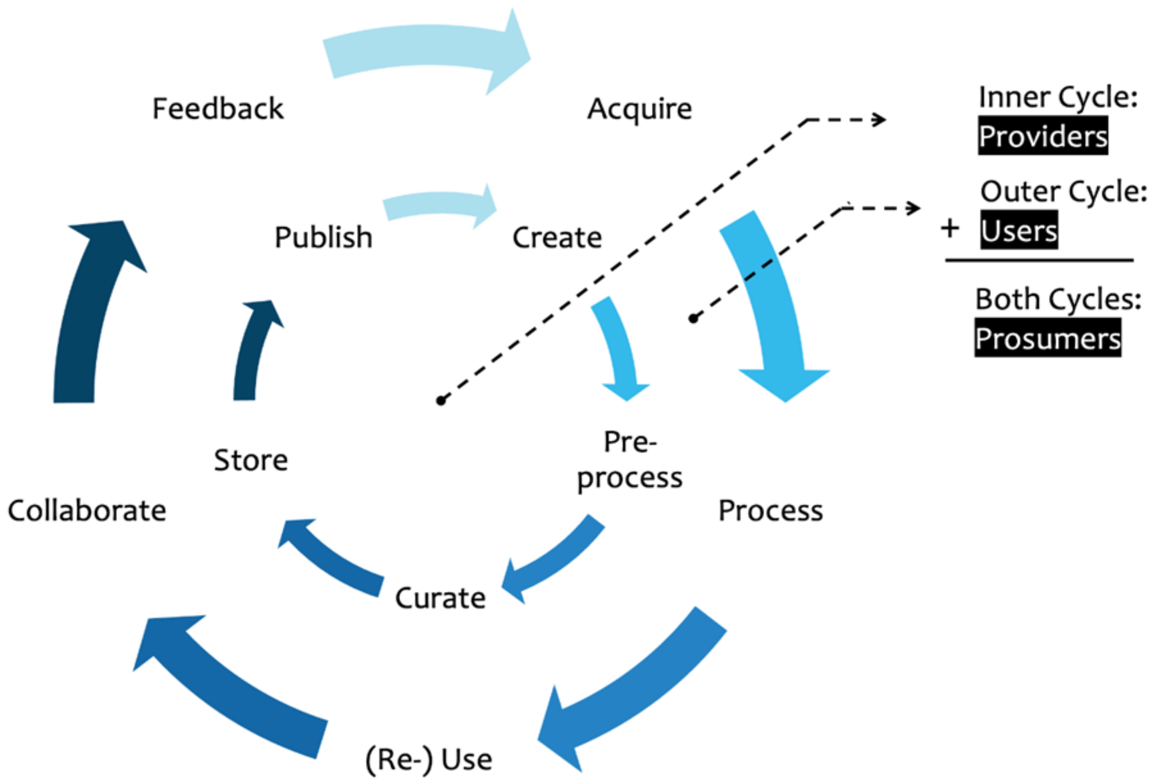


*Figure 3: The Open Data Life Cycle*

*Table 1: Steps of the open data life cycle (ODLC) (adapted from Charalabidis et al. 2018)*

| INNER CYCLE: PROVIDERS | OUTER CYCLE: USERS |
|---|---|
| **Create:** The process of creating data | **Acquire:** The process of data acquisition through OD portals |
| **Pre-process:** The managerial process of defining data quality | **Process:** The process of data analysis |
| **Curate:** The process of meeting the required data quality and legal requirements | **Use:** The process of presenting the analysis outcomes |
| **Store:** The decision-making process of storing | **Collaborate:** The process of communicating with other data users |
| **Publish:** The process covering legal issues | **Feedback:** The process of evaluating and providing feedback to data providers |

## 2.5 Data Science

The explosive availability of data in recent times has given rise to a multidisciplinary academic and professional field known as data science. This concept was popularised in the early 2010s with publications such as "Data Scientist: The Sexiest Job of the 21st Century" by Thomas Davenport and D.J. Patil [21] or Drew Conway's Venn diagram of data science (Figure 4) [22]. According to its proponents, the concept of Data Science integrates three dimensions: (1) Math & Statistics knowledge to model the data; (2) Hacking/computational skills to manage data at scale; (3) Substantive expertise to give real and useful meaning to the data and its processing. OD initiatives are somehow contributing to such an explosion of availability of data, and circularity adds a further dimension to the reflection and the manipulation of datasets.

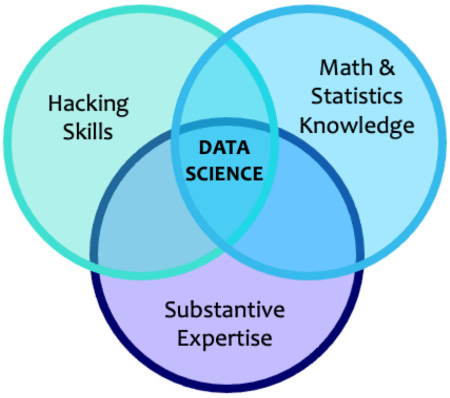


*Figure 4: Drew Conway's Venn diagram of data science*

# 3 Circularity in Open data ecosystems: A technical perspective

This chapter develops requirements for the technical design of circular open data ecosystems with a focus on ODP by analysing circularity from four perspectives: (1) user interfaces, (2) feedback mechanisms and tools, (3) tools and technologies used for open data analysis, and (4) artificial and collective intelligence systems to interact directly with the open data ecosystem.

## 3.1 User interface design for open data portals to promote circularity

A user interface is the part of a computer application that the user sees and interacts with. A graphical user interface refers to "a type of user interface that allows people to interact with a computer through a metaphor of direct manipulation of graphical images and widgets in addition to text" [23].

Open data portals are web-based interfaces designed to make it easier to find reusable information. Like library catalogues, they contain metadata records of datasets published for reuse, mostly relating to information in the form of raw, numerical data rather than textual documents. In combination with specific search functionalities, they facilitate finding datasets of interest. Application programming interfaces (APIs) are often available as well, offering direct and automated access to data for software applications.

### 3.1.1 Requirements for promoting the reuse of open data in open data portals

The report Rethinking the impact of open data [1] compiles the work of multiple authors of dimensions of a user interface design (See Table 2).

*Table 2: Dimensions of a user interface design that promotes the reuse of open data (adapted from van Ooijen et al. [1])*

| Metric | Description |
|---|---|
| Social media links | The portal integrates with social media platforms, establishing a dynamic social distribution channel for open data dissemination. |
| Feedback and support | Users of the portal benefit from real-time online support for feedback, enabling them to request or suggest the publication of new datasets easily. Additionally, assistance is readily available to address any issues encountered during use. |
| Newsfeed | The portal provides a way for users to stay informed and engaged with the latest information about data. |
| Guidance | Each dataset on the portal is thoughtfully accompanied by links and resources that offer user guidance and support. |
| Examples | To illustrate the practical application of the datasets real or fictitious examples of successful re-use are provided, demonstrating the versatility and impact of the available data. |

## 3.2 Enhancing Interface design through advanced feedback mechanisms

In the realm of open data, circularity embodies the philosophy of data as a continuous and dynamic resource, constantly flowing through a cycle of generation, dissemination, and repurposing. This notion underpins a constructive feedback loop (as represented in Figure 3 above), fostering an ongoing process of data collection, processing, and utilization. The strategic significance of establishing a Circular Open Data Ecosystem (CDO) becomes apparent in the context of Open Government Data (OGD) portals, which were envisioned to revolutionize the accessibility and utilization of government-owned data [3], [24], [25].

The open government data portals were launched with the noble goal of allowing the public to have the easy access and utilize government data, but these portals have faced challenges in fulfilling their complete capacity for impact change [26]. Despite the articulated goals of fostering innovation, promoting public engagement, and elevating government transparency, the empirical evidence for substantial impact remains elusive. One significant obstacle identified is the limited re-utilization of data [27]. It is crucial to recognize the need for circular strategies that empower both users with technical skills and those without, who may possess valuable perspectives and diverse needs. Hence, technically and also from T3.1, it is clear that the open data portals should have the feedback mechanism through which the portals can be updated by receiving feedback on the dataset level and at the portal level. Moreover, the conceptualized Circular Open Data Ecosystem promotes a proactive stance, encouraging active engagement and the improvement of users' technical proficiencies. By adopting this circular paradigm, data is rendered more readily available, scalable, and searchable across various formats, which promotes innovation and initiates an ongoing cycle of enhancement. Eventually, by adopting this comprehensive viewpoint on the open data ecosystem, the identified constraints are effectively resolved, and a more dynamic and influential OGD landscape is created. This means, in such a landscape, open government data is not only available but also utilized effectively by various stakeholders, including government agencies, researchers, businesses, civil society organizations, and the general public. An influential OGD landscape fosters transparency, accountability, and innovation, driving positive change and promoting citizen engagement in governance processes.

Furthermore, the evolution of open data ecosystems through iterative feedback loops, discourse, and dynamic interactions between data users and producers are integral to their maturation. Leveraging these user-centric characteristics enables effective communication of user requirements by both the open data community and the public sector entities responsible for data publication. Addressing these needs through appropriate public sector entities or harnessing community resources has the potential to significantly facilitate and expedite innovation. Regrettably, these factors appear to play a marginal role in current open data policies implemented within the public sector [9].

To develop user-friendly feedback mechanisms in portals, it is essential to consider several technical requirements. Feedback systems should be simple, not time-consuming, and use plain language to enhance user engagement [28]. Ensuring ease of navigation and user-friendliness of the portals is crucial for both data consumers and data producers [29]. Designing platforms with visual and quantitative outputs in an intuitive and user-friendly manner can enhance user interaction [30]. Additionally, incorporating user feedback to address issues like confusing test result displays can significantly improve the user-friendliness of portals [31]. Employing user-centred design approaches can foster the adoption and sustained use of patient portals [32].

"Technical aspects for inclusiveness across user domains in data portals" is an ODECO project tackling the challenge of making open data accessible and valuable for a diverse range of users through open data portals. Its core objective is to research new methods and tools that, from a technological point of view, close the open data cycle by delivering actionable feedback to open data publishers across user domains. So, open data portals that have resilient architectures and defined formats that make it simple for users to access the data need to be implemented. To equip users with the skills necessary to successfully contribute to the circular ecosystem, it is necessary to establish and implement (open) data literacy programs and training efforts. Because data literacy can help users to understand the data generated in various stages of the circular economy (e.g., resource extraction, product life cycles, waste management) and can help in analysing and interpreting data to identify opportunities for improvement and efficiency within the circular ecosystem.

### 3.2.1 Literature Review

This section provides a summary of a literature review on the need of a feedback mechanism in open data portals. Open data portals can significantly enhance circularity within the Circular Open Data Ecosystem by strategically incorporating and optimizing feedback mechanisms. However, the success of these mechanisms hinges on a user-friendly and intuitive UI/UX design.

Understanding user needs is crucial: Data consumers may want to suggest improvements to data quality or request specific datasets. Data providers might need feedback on dataset accessibility. By catering to these diverse needs, the UI/UX design should provide a variety of easily accessible feedback channels, such as clear contact forms, intuitive chatbots, or well-designed surveys [33].

Metadata Quality: It is important to improve the metadata quality because high-quality metadata promotes clear and actionable user feedback, leading to better data utilization and portal improvements [34]. It is essential to acknowledge that some Open Data Portals (ODPs) may indeed have excellent metadata documentation. Therefore, it is suggested that metadata should adhere to standards like ISO19115 and be regularly updated, presented in user-friendly language, and optimized for user comprehension. This approach ensures that users can easily navigate the portal and derive value from the available data, regardless of the portal's existing metadata quality.

Transparency and engagement are key: Displaying user feedback publicly (with permission) can build trust and encourage further participation. Providing regular updates on how feedback is used demonstrates the portal's commitment to improvement [35] [36].

Here are ways and levels at which open data portals can improve circularity through feedback mechanisms:

**Quality Enhancement of Data [36], [37]:**

- How: Feedback mechanisms can collect insights from users regarding the quality of the data available on the portal. Users can report inaccuracies, suggest improvements, or point out inconsistencies.
- Level: Implementing automated data quality checks based on user feedback, ensuring that the data provided is accurate, reliable, and up to date.

**Usability and Accessibility of the Portal [38]:**

- How: Users can provide feedback on the portal's interface, navigation, and accessibility features. This input helps in identifying areas for improvement and optimizing the user experience.
- Level: Regularly updating the portal's user interface based on feedback, ensuring it remains intuitive, responsive, and accessible to a diverse user base.

**Communication and Engagement [39]:**

- How: Feedback mechanisms enable users to communicate with portal administrators, fostering a two-way communication channel. Users can ask questions, seek clarification, or provide suggestions for additional features.
- Level: Establishing responsive communication channels, acknowledging and addressing user inquiries promptly, and incorporating user suggestions to enhance engagement.

**Iterative Development [40]:**

- How: Continuous feedback loops facilitate an iterative development process. Regularly obtaining user input allows for incremental improvements and ensures that the portal remains aligned with user expectations.

- Level: Embracing agile development methodologies, where feedback is incorporated into frequent release cycles, allowing for rapid and continuous improvement.

### 3.2.2 Feedback Mechanism for open data portals

The concept of circularity in open data ecosystems can be seen as an extension of the idea of feedback mechanisms. In a circular open data ecosystem, data is not only shared but also continuously refined through feedback from users and other stakeholders. This feedback loop helps to ensure that the data is of high quality and is being used in a way that benefits society. Also, feedback mechanisms allow stakeholders to provide input on the portal's functionality, usability, and content. Figure 5 shows the importance of feedback mechanism in the context of circular open data ecosystem. The figure illustrates the pivotal concept of a feedback mechanism within the context of an open data portal, elucidating its significance and various implementation strategies. At its core, a feedback mechanism constitutes a set of actions and tools designed to facilitate administrators in enhancing the quality, usability, accessibility, and communication of information on the open data portal [39], [41].

The figure emphasizes the essential role of a Feedback Mechanism in the improvement process, similarly like it is shown in Figure 5 about open data life cycle, highlighting its impact on crucial aspects such as quality enhancement and user experience. By providing a structured framework for communication between administrators and users, the feedback mechanism becomes instrumental in fortifying the open data portal's effectiveness.

The methods for implementing feedback mechanisms are detailed within the figure, showcasing a diverse array of tools. These include traditional approaches such as Feedback Forms and Contacting Portal Administrators, as well as more interactive methods like User Surveys and Interviews. Additionally, the Figure 5 underscores the importance of digital channels with the category "More Digital," which encompasses contemporary tools such as social media, chatbots, download statistics, multilingual support, data visualization, and data versioning.

- **Social media** is used to collect feedback for specific open data initiatives where users interact and share best practices, and foster community engagement.
- **Chatbots:** Open data portals can leverage chatbots to answer user questions about circularity principles, recommend datasets based on user preferences, and collect feedback on specific aspects of the datasets.
- **Download statistics:** Tracking downloads of datasets can identify user interests and it can help inn promoting the circular ecosystem.
- **Multilingual support:** It can make datasets more accessible to a wider audience and facilitate knowledge sharing across diverse communities.
- **Data visualization:** Interactive visualizations can help users understand complex data, identify trends, and make informed decisions.
- **Data versioning:** Tracking dataset versions ensures transparency and facilitates analysis of circularity progress over time.

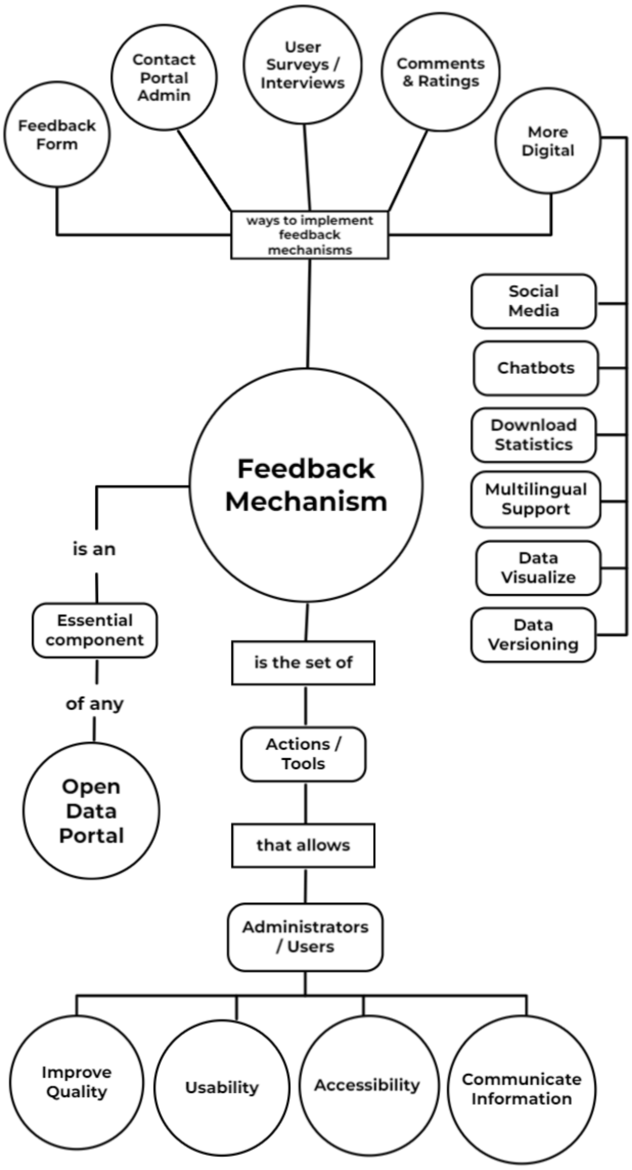


*Figure 5: Generalized Feedback Mechanism*

Open government data initiatives have made government information more accessible to the public, it is necessary to establish a consistent method for evaluating their influence on various user demographics. Although various feedback mechanism provides useful insights, a fully integrated solution would facilitate improved comparison and comprehension of user requirements. Keeping in mind the user requirements in the literature review section, Figure 6 illustrates iterative feedback loops that can help in building a circular open data ecosystem, which in details means a circular open data ecosystem with dynamic interactions between data consumers and producers contains the following properties:

1. Showing the interconnections and dependencies between different datasets, highlighting how they feed into one another to create a closed loop system.

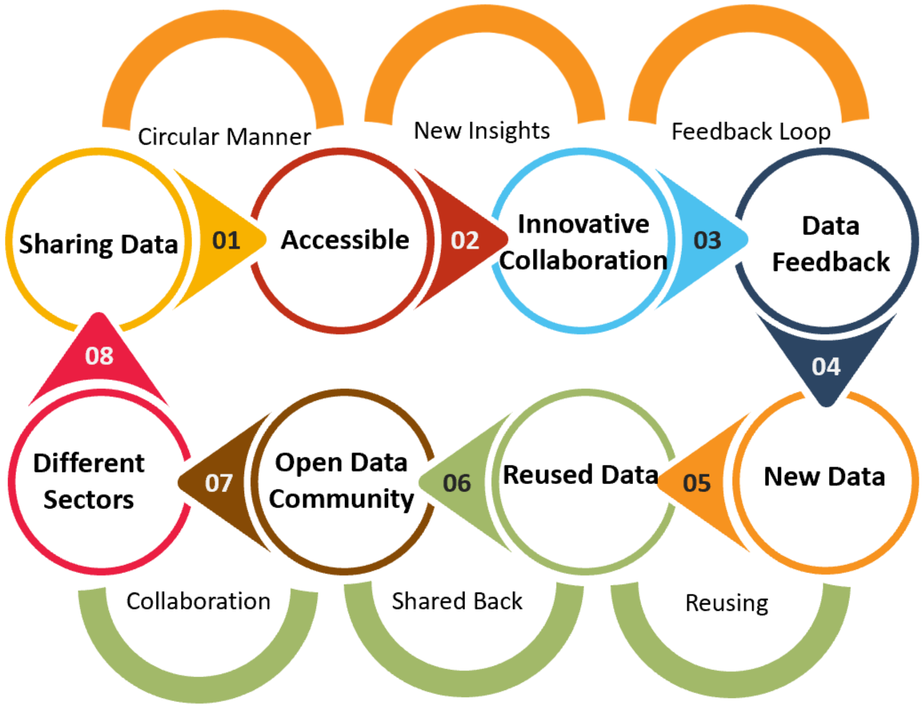


*Figure 6: Iterative Feedback Loops for a Circular Open Data Ecosystem*

2. Facilitating data sharing and collaboration between levels of government (i.e., central and regional) and the public and the private sector, and other organizations to promote circular open data ecosystem models.
3. Encouraging and enabling data reuse by making data openly available and easily accessible through the open data portal. This promotes circularity by reducing the need to collect new data and promoting the efficient use of resources.
4. Ensuring that data is properly described and catalogued using metadata standards, making it easier for users to find, understand, and use the data. Furthermore, by adopting quality control procedures, one can assure the accuracy, reliability, and currency of the data.
5. Encouraging the linking of datasets to promote a more interconnected and interrelated data landscape, promoting circularity by enabling data to be reused in new and innovative ways.
6. Establishing a user and stakeholder community for the open data portal, fostering circularity via the promotion of collaboration, feedback, and contribution, and facilitating a more agile and adaptable open data ecosystem.

The exploration of the concept of circularity in open data ecosystems reveals a paradigm shift in the way Open Government Data (OGD) portal's function. The strategic imperative arises from the conceptualization of a Circular Open Data Ecosystem, which entails the continuous refinement of data via dynamic interactions and iterative feedback loops.

## 3.3 Assessment of analytical tools and technologies to analyse OGD and non-government datasets to close the cycle:

In this subtask, a focus was given to the open data analytical tools and technologies and their importance to promoting open data user's contributions which further leads to the closing the cycle.

Deliverable 3.1 of the ODECO project elaborated on how different stakeholders contribute to the ecosystem and what values they could potentially generate. Figure 7 explains the relation between the values and contributions of open data user groups. ODEA stands for the open data ecosystems actors in the values and contributions matrix. The first column represents the contributions of the open data ecosystems. The first row presents the expected values generated by the stakeholders. It is self-explanatory from the matrix that a chain of stakeholders is involved in the value-contribution matrix based on their interest. Open data analysis tools and technologies can be located in the row "Data" for the creation of values such as "knowledge enrichment", "formed decision making", and "service enhancement" [42]. These three values ("knowledge enrichment", "formed decision making", and "service enhancement") are related to open data analysis tools and technologies. Hence, its worthy to note that open data analysis tools can contribute towards the value-creation in the open data ecosystem. It must be noted though that different open datasets require different analytics tools and technologies. Moreover, row "technological infrastructure" is also related to open datasets analytics as well. In this contribution ("technological infrastructure") open data users can create values in the form of "knowledge enrichment", "informed decision-making", and "service enhancement". These values and contributions are not exhaustive, and the list will be increasing with the passage of time, but to achieve these values and to contribute to them, technical, technological, and analytical support is required in the open data ecosystem. For instance, Big Data refers to the "emerging technologies that are designed to extract value from data having four Vs characteristics: volume, variety, velocity, and veracity" [43]. From the data we need some emerging technologies to extract values.

The circularity of the created values and data itself is the topic of this report, in which fair distribution of values among the stakeholders based on their contributions in open data ecosystems. The proper use of open data analytical tools and technologies can help the open data value-creation (and its circularity) and enhance/facilitate? contributions of different stakeholders. Van Loenen et al. [11] describes the current open data system as a linear system, which means that values are not created and distributed among the stakeholders fairly. For instance, Open data providers supply data to stakeholders, but they often do not receive corresponding benefits in return. The analytical technologies and tools are beneficial in this way to increase the participation of stakeholders to the open data portals (ODPs). Nowadays, technological tools are incorporated into portals to help stakeholders find the information they need about open data. Linked data technologies play a crucial role in the circularity of open data, encompassing Semantic Web standards, RDFs, query languages such as SPARQL, ontologies for mapping and alignment, navigation of linked data, data integration to explore and link datasets, and inference and knowledge discovery, which are made more accessible through linked data. By adopting these technologies and tools, circularity in the open data ecosystem and also the distribution of the values among stakeholders will be more fairly.

### 3.3.1 Literature Review

It is critical and essential to use open data analytics tools to improve circularity in open data ecosystems. According to Smith et al. (2016) [44], open data marketplaces can improve the value of open data by making data and related support services more accessible. This, in turn, promotes knowledge transfer within the ecosystem. According to Janssen et al. (2016) [45], policy guidelines aim to support commercial value generation in open data ecosystems by raising stakeholder understanding, ensuring resource availability, boosting cooperation between citizens and enterprises, and limiting negative consequences. When it comes to processing massive amounts of data, open-source tools such as Hadoop, Spark, and Presto play a crucial role. These tools enhance the efficacy and efficiency of data processing within open data ecosystems [46]. The growing prevalence of analytics practices and frameworks underpins the criticality of leveraging analytical tools and technologies for boosting openness and decision-making in open data ecosystems [47]. Ultimately, utilizing open data analytics techniques in open data ecosystems can promote circularity by enhancing data accessibility, encouraging cooperation, boosting innovation, and generating sustainable value. Open data ecosystems can fully fulfil their potential in achieving great societal and economic outcomes by following policy standards, encouraging interoperability, and successfully engaging stakeholders.

### 3.3.2 Tools and technologies for the open data analysis

This sub-task involves researching open data analytics tools and technologies through desk-based research. It includes examining their usage and describing the tasks for which they can be employed within the open data lifecycle (ODLC) as described in outer cycle of Figure 3 (section 2.4), to generate value from the open data. Furthermore, attention will be given to open data stakeholders to generate insights from the existing datasets using these tools and technologies. Before diving into the details of open data analytics tools and technologies, define these two terms. Open data analytics tools are computerized-programmed software and applications that help data analysts all along the open data life cycle (e.g., data collection, cleaning, analysis, and visualization). Some examples of open data analytics tools are MS Excel, Tableau, and Power BI. On the other hand, data analytics technologies are a broader term that encompasses the frameworks, concepts, and methods that promote, support, and underpin the data analytics process (e.g., database management systems, cloud computing, machine learning, and big data technologies). The open data tools are listed in Table 3, while the open data analytics technologies are detailed in Table 4. The information contained in these tables is gathered from desk research regarding the technologies and tools utilized by open data ecosystems. Additionally, the integration of the tools and technologies with open data portals or other frameworks and tools was described, along with their application. The utilization of these tools and technologies throughout the life cycles of open data is explicitly stated, as is the requisite expertise to operate these tools and technologies for open data analytics. It is important to note that the listings of tools and technologies are not exhaustive. They represent a blend of tools and technologies that resonate with stakeholders' expertise, current trends in tool and technology adoption, and the nature of the underlying datasets.

*Table 3: Some of the open data analytics tools*

| Tools | Usage and functionalities | Integration with ODP or other tools or frameworks | Use in ODLC (Outer Cycle, Figure 3) |
|---|---|---|---|
| Microsoft Excel | Data analysis, visualization, reporting | Microsoft Office, Power BI | Process: Used for data analysis, visualization, and reporting, making it versatile for preprocessing and basic analysis tasks. |
| Jupyter Notebook | Data exploration, prototyping, collaboration | Apache Spark, Google Cloud Platform | Process: Facilitates data exploration, prototyping, and collaboration, ideal for interactive data analysis and algorithm development. |
| Apache Spark | Big data processing, analytics, machine learning | Hadoop, Kafka, TensorFlow | Process: Enables big data processing, analytics, and machine learning, suitable for scalable data processing and complex analytical tasks. |
| Google Cloud AutoML | Automated machine learning, less expertise required in the | Google Cloud Platform | Process: Automates machine learning tasks, simplifying model development and analysis. |

| Tools | Usage and functionalities | Integration with ODP or other tools or frameworks | Use in ODLC (Outer Cycle, Figure 3) |
|---|---|---|---|
| | predictive analytics and machine learning. | | |
| SAS | Statistical analysis, data management | SAS Viya, SAS Studio | Process: Supports statistical analysis, data management, and modelling, useful for advanced analytics tasks. |
| Python | Data manipulation, machine learning, scripting | TensorFlow, PyTorch, Keras, Scikit-learn | Process, Use: Versatile for data manipulation, machine learning, and scripting, making it suitable for various preprocessing and analysis tasks. |
| R | Statistical analysis, data visualization | RStudio, Shiny | Process: Widely used for statistical analysis, data visualization, and modelling, ideal for exploring data patterns and relationships. |
| JavaScript | Web development, data visualization | D3.js, Chart.js | Use: Applied for web development and data visualization, primarily utilized in the "Use" stage for presenting analysis outcomes through interactive web applications or dashboards. |
| KNIME | Data blending, analytics, reporting, no coding required | KNIME Server, KNIME Hub | Process: Used in the "Process" stage for data blending, analytics, and reporting, providing a no-coding-required environment for data processing tasks. |
| https://streamlit.io/ | Building interactive web applications | Streamlit Sharing | Use: Employed for building interactive web applications, mainly utilized in the "Use" stage for sharing analysis outcomes through user-friendly web interfaces. |
| Microsoft Power BI | Business intelligence, data visualization, data integration from multiple sources. | Microsoft Office, Azure | Process: Utilized in the "Process" stage for business intelligence, data visualization, and data integration, facilitating data analysis and reporting tasks. |
| Tableau | Data visualization, analytics | Tableau Server, Tableau Public | Use: Applied in the "Use" stage for data visualization and analytics, enabling users to explore and interact with data through visualizations. |

| Tools | Usage and functionalities | Integration with ODP or other tools or frameworks | Use in ODLC (Outer Cycle, Figure 3) |
|---|---|---|---|
| Qlik Sense | Data visualization, analytics | Qlik Cloud, Qlik Analytics Platform | Use: Utilized in the “Use” stage for data visualization and analytics, enabling users to create interactive data visualizations and dashboards. |
| Looker | Data exploration, analytics | Looker Blocks, Looker API | Process: for data exploration and analytics, providing tools for analysing and visualizing data insights. |
| Zoho Analytics | BI, reporting, analytics | Zoho CRM, Zoho Projects | Process: for business intelligence, reporting, and analytics, facilitating data analysis and reporting tasks. |
| Domo | Business intelligence, data visualization | Domo Appstore, Domo APIs | Process: for business intelligence and data visualization, aiding in analysing and visualizing data insights. |
| Sisense | Embedded analytics, BI | Sisense APIs, Sisense BloX | Process: Employed in the “Process” stage for embedded analytics and BI, enabling organizations to integrate analytics into their applications or workflows. |
| Google Analytics Solutions | Web analytics | Google Tag Manager, Google Data Studio | Process: providing insights into website traffic and user behaviour. |
| SAP Analytics Cloud | BI, planning, predictive analytics | SAP ERP, SAP HANA | Process: For business intelligence, planning, and predictive analytics, supporting data analysis and forecasting tasks. |
| Microsoft Azure Analytics | Data analytics, AI, machine learning | Azure Machine Learning, Azure Data Lake | Process: for data analytics, AI, and machine learning, facilitating advanced analytics and predictive modelling. |
| RapidMiner | Data science, machine learning | RapidMiner Server, RapidMiner Marketplace | Process: for data science and machine learning, supporting various data analysis and modelling tasks. |
| Redash | Data visualization, querying | Redash API, Redash Integrations | Use: for data visualization and querying, enabling users to visualize and analyse data through interactive dashboards. |
| Grafana | Metrics dashboarding, monitoring | Prometheus, InfluxDB | Use: for metrics dashboarding and monitoring, providing |

| Tools | Usage and functionalities | Integration with ODP or other tools or frameworks | Use in ODLC (Outer Cycle, Figure 3) |
|---|---|---|---|
| | | | visualization tools for monitoring system metrics and performance. |
| DataMelt | Scientific computing, data analysis | Java, Jython | Process: for scientific computing and data analysis, providing tools for mathematical and statistical analysis. |
| Orange | Data mining, machine learning | Python-based, Integration with Python libraries | Process: for data mining and machine learning, facilitating various data analysis and modelling tasks. |
| Weka | Machine learning, data mining – required less expertise to use it. Usually drag and drop. | Java-based, Integration with Java environment | Process: Applied in the "Process" stage for machine learning and data mining, providing tools for building and evaluating predictive models. |
| IBM SPSS Statistics | Statistical analysis, predictive modelling | Integration with Python, R, and SQL | Process: for statistical analysis and predictive modelling, assisting in analysing data and building predictive models. |
| MATLAB | Numerical computing, data analysis | MATLAB Production Server | Process: for numerical computing and data analysis, providing tools for analysing and visualizing data. |
| Databricks | Unified data analytics platform | Integration with Apache Spark | Process: for unified data analytics, providing a platform for processing and analysing large-scale datasets. |
| Alteryx | Data blending, analytics, predictive modelling | Alteryx Server, Alteryx Designer | Process: for data blending, analytics, and predictive modelling, assisting in data preparation and analysis tasks. |
| H2O.ai | Machine learning, predictive analytics | Integration with Python, R, and Spark | Process: for machine learning and predictive analytics, providing tools for building and deploying machine learning models. |

*Table 4: Open data analytics technologies*

| Technologies | Use | Integration with ODP or other tools or frameworks | Use in ODLC (Outer Cycle, Figure 3) |
|---|---|---|---|
| CKAN | Open data management platform used for publishing, sharing, and accessing data. It provides metadata management, data organization, and access control features. | APIs, Plugins | Use: CKAN is commonly used by data publishers and organizations to publish and share open datasets. It facilitates data discovery and access for users. |
| DKAN | Open data platform and data management system that offers features for publishing, sharing, and managing datasets. It includes metadata management, visualization tools, and access control mechanisms. | APIs, Plugins | Use: DKAN is widely used in government and non-profit sectors for managing and sharing open datasets. It helps organizations in data publication and transparency efforts. |
| SOCRATA | Open data management and visualization platform used by government agencies and organizations for data sharing and visualization. It offers data publishing, visualization, and data analysis capabilities. | APIs, Plugins | Use: SOCRATA is commonly used by government agencies to publish and share open data with the public. It enables data visualization and analysis for better decision-making. |
| opendatasoft | Open data sharing and visualization platform designed for publishing and exploring datasets. It provides tools for data visualization, API access, and collaboration features. | APIs, Plugins | Use: OPENDATASOFT is used by organizations to create data portals for sharing open data. It enables data exploration, visualization, and collaboration among users. |
| Kaggle | Data science competitions platform and dataset repository. Kaggle hosts machine learning competitions, datasets, and notebooks, allowing data scientists to collaborate and showcase their work. | APIs, Kaggle Kernels | Process Kaggle is utilized by data scientists and researchers for analysing datasets, participating in competitions, and sharing insights through notebooks. |
| Elasticsearch | Distributed search and analytics engine used for real-time data analysis and full-text search. Elasticsearch is commonly used for indexing, searching, and analysing large volumes of | REST APIs, Client libraries | Process: Elasticsearch is utilized for real-time data analysis, log monitoring, and search functionalities in open data platforms, facilitating rapid access to |

| Technologies | Use | Integration with ODP or other tools or frameworks | Use in ODLC (Outer Cycle, Figure 3) |
| --- | --- | --- | --- |
| | structured and unstructured data. | | insights and information retrieval. |
| Apache Kafka | Distributed streaming platform used for building real-time data pipelines and streaming applications. Kafka enables high-throughput, fault-tolerant, and scalable messaging systems for data processing. | Kafka Connect, Kafka Streams | Process: Apache Kafka is integrated into open data platforms for real-time data ingestion, event processing, and building data streaming pipelines for analytics and monitoring. |
| Database management systems | Software systems for storing, managing, and retrieving data. They include relational databases (e.g., MySQL, PostgreSQL) and NoSQL databases (e.g., MongoDB, Cassandra). | APIs, Database connectors | Process: Database management systems are fundamental for storing and managing data in various applications, including open data platforms. They facilitate data storage and retrieval. |
| Cloud computing | Delivery of computing services over the internet on an on-demand basis. Cloud platforms (e.g., AWS, Azure) offer resources such as storage, processing, and analytics tools. | APIs, SDKs, Integration platforms | Process: Cloud computing provides scalable infrastructure for data storage, processing, and analysis, making it integral to open data platforms and analytics workflows. |
| Distributed computing | Processing large datasets across multiple nodes or computers. Distributed computing frameworks (e.g., Hadoop, Spark) enable parallel processing and analysis of big data. | Distributed computing frameworks | Process: Distributed computing frameworks are used for processing and analysing large volumes of data in open data platforms, enabling scalable and efficient data processing. |
| Machine Learning | Algorithms that learn patterns and make predictions from data. Machine learning libraries (e.g., TensorFlow, scikit-learn) provide tools for building and deploying ML models. | Machine learning libraries, SDKs | Process: Machine learning is employed in open data platforms for predictive analytics, anomaly detection, and pattern recognition tasks, enabling data-driven insights and decisions. |
| NLP | Natural language processing techniques for analysing and interpreting human language data. NLP libraries (e.g., NLTK, spaCy) enable tasks such as text classification and sentiment analysis. | NLP libraries, SDKs | Process: NLP is utilized in open data platforms for processing and analysing textual data, extracting insights, and facilitating information retrieval and understanding. |

| Technologies | Use | Integration with ODP or other tools or frameworks | Use in ODLC (Outer Cycle, Figure 3) |
|---|---|---|---|
| Neo4j | Graph database management system used for storing, querying, and analysing graph data structures. Neo4j enables efficient graph algorithms and queries for exploring relationships and networks in data. | Neo4j drivers, Cypher query language | Process: Neo4j is employed in open data platforms for graph data analysis, network visualization, and relationship exploration, facilitating insights into complex interconnected data. |
| BigQuery | Fully managed data warehouse and analytics platform used for storing, querying, and analysing large datasets. BigQuery offers SQL-like querying and real-time analytics capabilities for big data processing. | BigQuery API, Google Cloud SDK | Process: BigQuery is integrated into open data platforms for data warehousing, ad hoc querying, and analytics, enabling rapid insights and data-driven decision-making. |
| Databricks | Unified analytics platform for big data processing and machine learning. Databricks provides an Apache Spark-based environment for data engineering, data science, and collaborative analytics workflows. | Databricks Connect, Databricks API | Process: Databricks is used in open data platforms for scalable data processing, machine learning model development, and collaborative analytics, facilitating insights and innovation. |
| Redis | In-memory data store and caching system used for real-time data processing and caching. Redis enables fast data access and caching of frequently accessed data for performance optimization. | Redis clients, Integration frameworks | Process: Redis is integrated into open data platforms for caching, session management, and real-time data processing, enhancing performance and scalability of data-driven applications. |
| Neo4j | Graph database management system used for storing, querying, and analysing graph data structures. Neo4j enables efficient graph algorithms and queries for exploring relationships and networks in data. | Neo4j drivers, Cypher query language | Process: Neo4j is employed in open data platforms for graph data analysis, network visualization, and relationship exploration, facilitating insights into complex interconnected data. |
| Prometheus | Monitoring and alerting toolkit used for recording and querying time-series metrics. Prometheus enables monitoring of distributed | Prometheus APIs, Client libraries | Process: Prometheus is integrated into open data platforms for monitoring data ingestion, processing, and |

| Technologies | Use | Integration with ODP or other tools or frameworks | Use in ODLC (Outer Cycle, Figure 3) |
|---|---|---|---|
| | systems and applications, providing insights into performance and reliability. | | |
| R | Programming language and software environment for statistical computing and graphics. R offers extensive libraries for data manipulation, statistical analysis, and visualization. | R packages, RStudio | Process: R is commonly used in open data platforms for statistical analysis, data visualization, and predictive modelling, enabling advanced analytics workflows. |
| Python | General-purpose programming language widely used in data analytics and scientific computing. Python offers libraries such as NumPy, pandas, and scikit-learn for data manipulation, analysis, and machine learning. | Scikit-learn, NumPy, SciPy library, Matplotlib, Pandas | Process: Python is extensively used in open data platforms for data manipulation, statistical analysis, machine learning, and visualization, enabling diverse analytics workflows. |
| Apache Spark | Unified analytics engine for large-scale data processing. Apache Spark offers distributed computing capabilities and high-level APIs for processing and analysing big data. | Spark MLlib, Spark SQL | Process: Apache Spark is utilized in open data platforms for distributed data processing, machine learning, and real-time analytics, enabling scalable and efficient data analysis workflows. |
| Jupyter Notebook | Interactive computing environment for creating and sharing documents containing live code, equations, visualizations, and narrative text. Jupyter Notebooks support various programming languages including Python, R, and Julia. | Jupyter extensions, kernels | Process: Jupyter Notebook is commonly used in open data platforms for exploratory data analysis, data visualization, and collaborative research, facilitating interactive and reproducible analytics workflows. |
| Apache Hadoop | Distributed storage and processing framework for big data. Hadoop enables distributed processing of large datasets across clusters of computers using simple programming models. | Hadoop ecosystem components, Hadoop connectors | Process: Apache Hadoop is integrated into open data platforms for distributed data storage, batch processing, and analysis, enabling scalable and fault-tolerant data processing workflows. |
| Apache Beam | Unified programming model for batch and streaming data | Beam SDKs, Beam runners | Process: Apache Beam is used in open data |

| Technologies | Use | Integration with ODP or other tools or frameworks | Use in ODLC (Outer Cycle, Figure 3) |
|---|---|---|---|
| | processing. Apache Beam provides a portable API for building data processing pipelines that can run on various execution engines such as Apache Spark, Apache Flink, and Google Cloud Dataflow. | | platforms for building data processing pipelines, enabling unified batch and streaming data processing workflows across different execution engines. |
| Apache Cassandra | Distributed NoSQL database designed for handling large amounts of data across multiple nodes. Cassandra offers high availability and scalability for distributed data storage and querying. | Cassandra drivers, CQL | Process: Apache Cassandra is integrated into open data platforms for scalable and fault-tolerant data storage, enabling efficient data management and querying in distributed environments. |
| PyTorch | Open-source deep learning framework for building neural networks and machine learning models. PyTorch offers flexibility and ease of use for research and production deployments in various domains. | PyTorch APIs, TorchServe | Process: PyTorch is utilized in open data platforms for deep learning, neural network training, and model deployment, enabling advanced analytics and predictive capabilities. |
| Knowledge Graph | Google Knowledge graphs, Bing's Satori (from Microsoft), Freebase, DBPedia, and wiki data | Google Knowledge Graph API, Wikidata (REST API), Freebase API, | Use, Process: It can be used to rank the entities that match certain criteria. Content annotation and organization using the knowledge graph entities. |

Keeping in mind the decryption of open data tools and technologies and their usage, let us assess an open data portal and the usage of the above tools and technologies. The Croatian open data portal (data.gov.hr) is developed using the CKAN technology (which is a framework to develop content management systems, also listed in Table 4 as a technology). The portal uses PHP, Java, and Python as programming languages and storage functionalities (PostgreSQL), and they also provide data review functionalities using JavaScript frameworks and libraries. All this information regarding data.gov.hr has been collected from the wappalyzer.

In a nutshell, a wide variety of tools and technologies are used to cover the whole of the open data lifecycle, but if we stick to open data analytics, then Table 3 and Table 4 assess the open data analysis tools and technologies with respect to the open data life cycle, their use, and their integration functionalities within the open data portals and across the tools and technologies. Integration with open data portals and with other tools and frameworks is equally necessary in this multi-discovery world because sometimes standalone tools and technologies cannot generate enough value; they need to be integrated with other tools and technologies to achieve that.

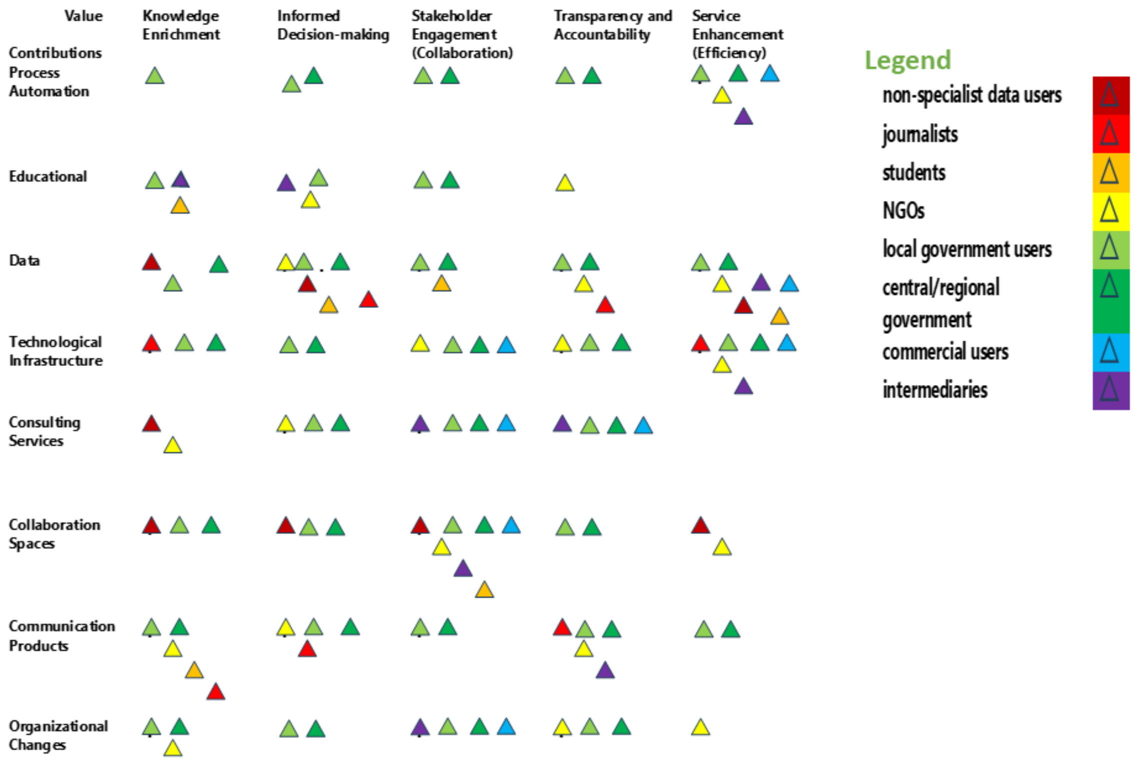


*Figure 7: Matrix of values and contributions of open government user groups in the ODEA [23]*

### 3.3.3 Technical requirements for circular open data ecosystem from data analytics perspectives

By extracting technical requirements from an examination of extant literature and a synthesis of previous ODECO deliverables, it is possible to determine which tools and technologies are necessary to transform the current open data ecosystem into a circular one. These requirements are interconnected with user interfaces, the implementation of CI and AI, and feedback mechanisms for open data ecosystem circulation. The subsequent points outline the technical requirements that must be incorporated in ODP pertaining to data analysis tools and technologies to transform an open data system into a circular open data ecosystem:

- **Req1: Open Data accessibility and interoperability**: The Open Data Platform (ODP) requires backing for data integration and interoperability from a range of stakeholders, including governmental and non-governmental entities. Employing open data standards and formats can streamline data accessibility and sharing within open data ecosystems, enhancing their circularity.
- **Req2: Comprehensive dataset coverage**: availability of a wide variety of thematic datasets such as from different domains such as from science and transport and so on.
- **Req3: Scalability and performance of the portal**: the open data volume, variety and velocity increases by the time, so open data portals should be scalable, and performance should be high to accommodate these variations over the time.
- **Req4: User-friendly interfaces**: Establish an easy-to-navigate interface for the datasets and analysis tools that anyone can understand and use. Facilitate better data exploration and comprehension by including interactive visualisation tools.
- **Req5: Advanced searches and filtering capabilities**: Users should be able to swiftly locate pertinent datasets by utilizing strong search capabilities that allow them to enter keywords,

categories, or other criteria. Make it possible to narrow your search and zero in on certain data sets by including filtering options.

- **Req6: Data governance (data quality assurance):** Find ways to check if datasets from the government and private organizations are up to par. To help consumers evaluate the data's dependability, include metadata standards and quality indicators. For instance, for linked open data, verification of links could be beneficial for the open data interoperability and further analysis of the data.
- **Req7: Analytical tool integration**: Assist in the incorporation of numerous data analysis tools that address certain dataset types.
- **Req8: Collaboration and knowledge sharing**: With the help of these capabilities, users will be able to work together and exchange ideas, analysis methods, and results. Put in place commenting systems, collaborative workspaces, or discussion forums to encourage the sharing of knowledge.
- **Req9: Feedback mechanisms**: Establish feedback mechanisms to elicit users' opinions on the portal's features, datasets, and analytical tools. Maintain the portal's currency and enhance it based on user input and advancements in technology.
- **Req10: Documentation and Training Resources**: Offer extensive documentation and training materials to aid users in proficiently utilizing the portal's features and functionalities.
- **Req11: Technological infrastructure**: technological infrastructures such as cloud computing and distributed computing for the analysis of the datasets. this can help in the scalability and performance of the open data portals. For instance, in cloud computing users have a wide variety of analytics tools and technologies available.
- **Req12: Data categorization with respect to possible analytics**: (Regression, Classification, Time series, Unsupervised, Natural language processing etc.).

### 3.3.4 Conclusion

Open data ecosystems require the circularity of values being generated and distributed fairly among stakeholders. These values are not directly created by the data itself; they were somehow processed to generate the information, knowledge, and products based on the OD. The identified 13 technical requirements and their adoption in the open data ecosystem facilitate the open data ecosystem circularity. Specifically, integrated tools within open data portals are useful to generate or somehow increase these values for technical and non-technical users of open data. The utilization of linked open data tools and technologies, such as RDFs, query languages, graph analytics, ontologies (mapping and alignments), data science methods, and NLP techniques, aids in fostering the circularity of open data. Open data interoperability is also an important factor in circularity, as well as an important technical requirement for circularity. In the future, it would be beneficial to explore different tools and technologies that could be beneficial for diverse types of open datasets, like, for instance, linking specific data analysis tools to different types of datasets). Some portals also specify the datasets which are readily available for the analysis (for example, dataset X is available for the regression analysis, dataset Y is available for supervised machine learning and so on). This categorization of datasets can add value to the open data ecosystem by providing the involved stakeholders with the potential usage of open data.

## 3.4 Technical requirements of AI and CI open data ecosystems

### 3.4.1 Introduction

In evaluating the technical requirements for the integration of Artificial Intelligence and Collective Intelligence (AI and CI) to directly interact (through APIs) with the open data, it is essential to understand the pivotal role of these open data ecosystems in the current digital era. These networks, consisting of data providers, users, and intermediaries, are crucial for promoting transparency, innovation, and collaboration. With the ever-increasing volume and significance of data, ensuring the sustainability and circularity of these ecosystems is paramount. The circularity in this perspective, mirroring the principles circular economy, underscores the necessity of transforming data's lifecycle to maximize efficiency and value retention.

Integrating methodologies from Artificial Intelligence use cases, such as automated metadata extraction, intuitive search engines, and recommendation systems, along with Collective Intelligence use cases like Wikipedia's collaborative editing and Open Street Map's collaborative mapping, into these ecosystems can significantly enhance their circularity. Specifically, AI enhances circularity by improving data accessibility by facilitating faster, more relevant data retrieval, which in turn, boosts reuse across various contexts [48], [49]. On the other hand, CI contributes by continuously updating and verifying data through communal effort, ensuring its relevance and extending its utility. The employment of APIs underscores this synergy, enabling robust data exchange and integration that enhance data's interoperability and lifecycle management. Together, these integrations foster a regenerative data use cycle, where information is constantly refined, shared, and optimized, embodying circularity principles by preserving data's value and utility over time within the ecosystem.

However, to leverage the full potential of this integration, a structured framework ensuring data quality, data interoperability, data accessibility, and compliance to legal standards is essential. The vast processing capabilities of AI, combined with the contextual enrichment provided by CI, can lead to transformative insights and applications. Yet, this integration is not without challenges. In our below literature review we have included some literature pertaining to the challenges for AI and CI in promoting circularity. Our literature review strategy involved using pertinent keywords such as AI, CI, Open data, Government Data, API, Crowdsourcing, ML, LLM and examining titles for initial identification, followed by a detailed review of abstracts and conclusions to verify their relevance and value to our topic of study.

### 3.4.2 Literature Review

The following literature focuses on the technical requirements of AI and CI to improve the circularity within open data ecosystems, addressing a range of aspects related to technological integration and innovation. These studies underline the essential role that AI and CI play in promoting the reuse and recycling of data, reflecting the principles of the aforementioned circular economy model. These studies highlight the pivotal contribution of AI and CI to circularity, while shedding light on the necessary requirements for their effective integration and the challenges they encounter in this process.

The study by Jin Li et al. explores the symbiotic relationship between big data and AI, underscoring the AI requirements to harness and analyse vast datasets effectively, which is crucial when interacting with APIs for meaningful data insights and decision-making [50]. Richard Absalom et al.'s research emphasizes the role of CI in gathering and synthesizing intelligence from various sources, which enhances API interactions by providing a richer, more nuanced understanding of data [51]. Joutsenlahti et al. pinpoint the governance structures necessary for AI and CI applications to ensure reliable, ethical, and compliant data exchange via APIs, highlighting the need for transparency and accountability in data handling [52]. They also discuss the importance of adaptable governance and service frameworks to facilitate user-centric and efficient API interactions within AI/CI-driven data services [52].

Zhengbao Jiang et al. showcase how AI and CI can advance data processing techniques, enabling sophisticated relation extraction and integration, which is vital for extracting actionable insights from complex data sets accessed through APIs [53]. Tan explores the requirements for AI compatibility in open government data ecosystems. He sheds light on the specific needs for AI systems to effectively integrate into governmental data ecosystems, which often involves complex data [54].

Building upon these findings, further studies emphasize the need for robust data platforms to support the integration of AI and CI. Vassilev et al. discuss the development of AI-driven platforms that can enhance data integration and management [55]. This advancement is crucial for the efficient utilization of AI and CI in handling extensive data sets. Abedjan underlines the importance of data quality and literacy in implementing effective AI solutions. He establishes that for AI and CI to be effectively integrated into open data ecosystems, there is a fundamental need for ensuring high-quality data and enhancing data literacy among stakeholders [56]. Traub et al. emphasize the importance of open-source tools and collaborative platforms [57]. These platforms are vital for sharing and analysing large datasets, enabling stakeholders to engage in circularity initiatives more effectively. Trevisan et al. focus on leveraging digital platforms for efficient network management and collaboration [58]. It highlights how AI technologies can facilitate efficient information and resource exchange among businesses, enhancing circularity in business ecosystems.

In summary, these studies collectively paint a comprehensive picture of the challenges and requirements for integrating AI and CI into open data ecosystems. They emphasize the need for managing data complexities, addressing biases in AI, ensuring data quality, and establishing robust governance frameworks. These elements are crucial for the successful and ethical integration of AI and CI technologies in open data environments, thereby paving the way for more sustainable and efficient data ecosystems.

### 3.4.3 AI, CI and their technical requirements

**Artificial Intelligence:** Artificial Intelligence (AI) has potential to significantly enhance the functionality of open data portals, driving the findability, accessibility, usability, and circularity of data [59]. By improving search capabilities, AI enables users to efficiently locate and utilize diverse datasets through advanced search mechanisms. It also plays a crucial role in generating and enriching metadata, making data more accessible and understandable. AI-driven recommendation systems further aid in discovering relevant datasets, encouraging reuse and recycling of data. Additionally, AI's ability to extract actionable insights from data transforms raw information into valuable knowledge. This multifaceted contribution of AI not only maximizes the utility of data but also fosters a self-sustaining cycle of data creation, sharing, reusing and innovation within these portals.

**Collective Intelligence:** Collective Intelligence (CI) has potential to improve the circularity of open data portals by harnessing the power of crowdsourcing, as demonstrated by platforms like OpenStreetMap [60] and Wikidata [61]. In this model, users globally contribute, refine, and verify data, ensuring its accuracy and richness. This collaborative approach can be applied to open data portals, allowing users to identify and correct inaccuracies or gaps in datasets, similar to the collaborative contributions in GitHub.

Furthermore, CI enables users to add value to these datasets. Individuals with diverse expertise can analyse the data, derive unique insights, and share their findings back with the community. This could include creating new data visualizations or developing advanced predictive models. To encourage and track these contributions, we propose features like a history of changes for each dataset and a recognition system for contributors, fostering a transparent, community-driven, and continuously improving data ecosystem. Figure 8 reflects the contribution chart for GitHub commits, which is what inspires our proposal for a similar mechanism for open datasets [62].

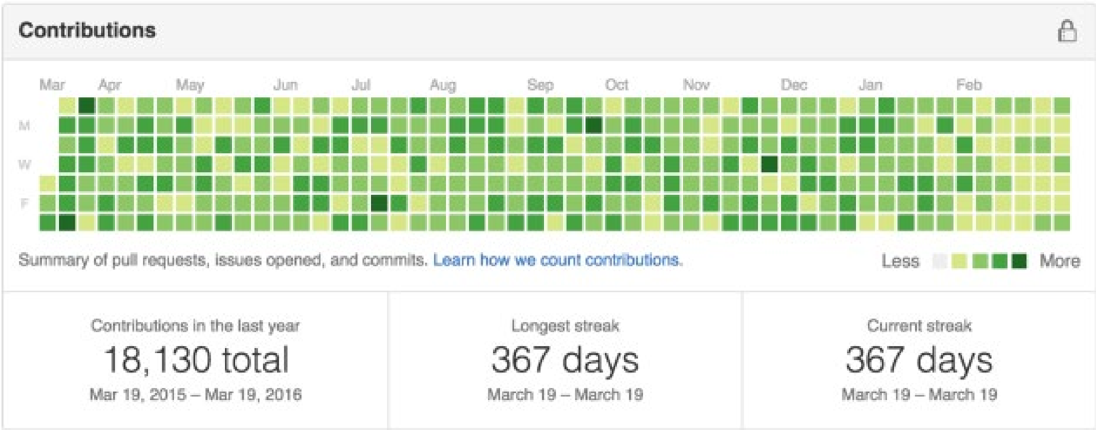


*Figure 8: Contribution chart reflecting GitHub commits [38]*

**Technical Requirements:**

To effectively harness the potential of AI and CI within OD ecosystems concerning their interaction with APIs, it is essential to identify and address specific requirements tailored to each domain. AI and CI serve distinct yet complementary roles in leveraging open data: AI focuses on automated data processing and insight generation, while CI emphasizes human collaboration and collective knowledge building. By delineating clear, dedicated requirements for each, we can ensure that both AI and CI not only coexist but synergize within the OD ecosystem, enhancing data utility, accessibility, and innovation. Below are the requirements specifically curated for AI and CI in the context of their interactions with OD APIs:

**I. AI Requirements:**

- **High-Quality and Structured Data:** AI systems require clean, structured, and high-quality data from OD APIs for reliable analyses and outputs. "High-quality" here refers to data that is accurate, complete, and consistent, essential for the effective functioning of AI applications like recommendation systems and conversational search engines, ensuring they deliver precise and actionable insights.
- **Comprehensive Metadata:** AI applications rely heavily on detailed metadata and their adherence to standards to navigate and interpret open data accurately, necessitating that APIs furnish extensive, contextual information about the datasets they present. Here, "comprehensive" metadata implies that the information provided should encompass all necessary details that define and describe the dataset's content, context, and structure. Such depth and breadth in metadata are crucial for tasks like identifying or associating similar datasets within an open data portal, significantly enhancing data discoverability and usability. This level of detail ensures that AI systems can understand the relevance, provenance, and applicability of the data they access, facilitating more informed and accurate analyses.
- **Data Interoperability:** AI's ability to integrate and analyse diverse data sets hinges on semantic and technical interoperability, necessitating that OD APIs facilitate seamless data exchange and synthesis across varied formats and standards.
- **Real-time Data Processing Capabilities:** For AI to respond dynamically to evolving situations, OD APIs must offer capabilities for real-time data processing, enabling immediate analysis and decision-making. A prime example is traffic management, where real-time data processing is essential for optimizing traffic flow, minimizing congestion, and improving road safety.
- **Robust and Scalable Infrastructure:** AI interactions with OD APIs demand a strong, scalable infrastructure to support intensive computational tasks and large-scale data analyses, ensuring efficient and timely data processing. It entails powerful machines with high

processing power and significant storage to be able to cater to the demands of ever-growing open data.

**II. CI Requirements:**

- **User-friendly Interfaces and Tools:** CI thrives on community engagement, requiring OD APIs to be accessible through interfaces and tools that are intuitive and easy for all users, facilitating active participation and collaboration.
- **Feedback Mechanisms for Continuous Improvement:** To cultivate a progressive and interactive CI landscape, it is crucial for OD ecosystems to embed feedback functionalities within their APIs. These features should empower users to offer insights and recommendations seamlessly through the API interface, thereby advancing data quality and augmenting community engagement. Such mechanisms ensure that user feedback is directly integrated into the data cycle, promoting an iterative enhancement process that is vital for the sustained evolution of the OD ecosystem.
- **Compliance with Legal and Ethical Standards:** Ensuring that CI interactions via OD APIs adhere to legal and ethical guidelines is paramount, promoting a secure, respectful, and transparent environment for all participants.
- **Data Interoperability:** Central to fostering a robust CI environment is ensuring that users can easily share, access, and collaborate on diverse datasets. Interoperability across different data formats and standards is critical, enabling seamless integration and analysis of various data sets, thus enhancing collective knowledge creation and decision-making processes.

Meeting these requirements ensures that OD ecosystems are conducive for AI and CI, facilitating advanced data analysis, better decision-making, and enhanced community engagement. Figure 9 reflects the requirements of AI and CI to interact with open data through APIs.

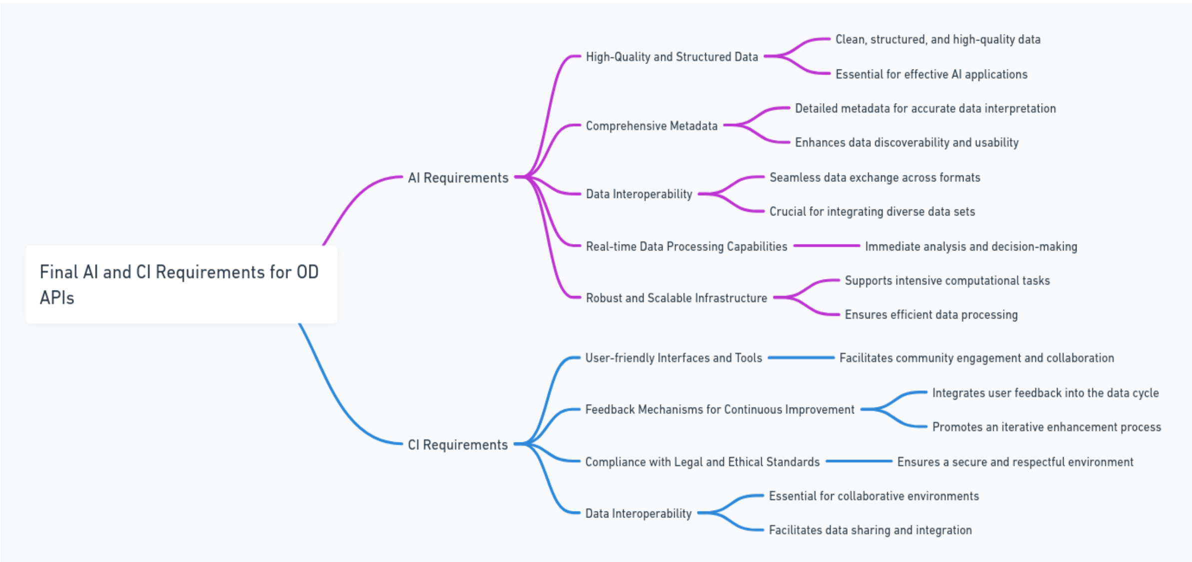


*Figure 9: Technical requirements of AI and CI to seamlessly interact with open data through APIs and enhance the ecosystem*

### 3.4.4 Conclusion

The technical intricacies of APIs are fundamental in effectively integrating Artificial Intelligence (AI) and Collective Intelligence (CI) with open data ecosystems. This is an essential step to unleash their collective transformative capabilities. To achieve this, APIs must provide robust data accessibility, ensuring that AI algorithms can access the high-quality, structured data they require for precise analysis. Equally important is the integration of comprehensive metadata through APIs, which enhances data contextualization for AI while bolstering understanding and participation in CI endeavours. Furthermore, APIs need to uphold interoperability standards, facilitating the seamless integration of varied data formats and systems, a necessity for the collaborative and cross-functional nature of AI and CI interactions. Support for real-time data processing is another crucial feature, particularly vital for AI applications where timely data analysis is imperative. Additionally, the scalability and reliability of APIs are paramount, ensuring they can handle escalating data volumes and maintain performance consistency. Embedded security measures within APIs are also essential, safeguarding data integrity and user trust. Lastly, APIs should be designed with a user-centric approach, especially to enhance CI participation, making them intuitive and accessible for diverse user groups. By prioritizing these technical attributes, APIs can significantly bolster the integration and functionality of AI and CI within open data frameworks, fostering innovation and a more interconnected data landscape.

# 4 Current state of Open Data Portals

In this section, the Early Stage Researchers (ESRs) conducted an analysis of three notable European data portals: the European Data Portal (EDP) [63], the French Data Portal (FDP) [64], and the Swedish Data Portal (SDP) [65], from their topics' perspective. The current status of these data portals reveals significant room for improvement in aspects of circularity and closing the loop.

## 4.1 Circularity compliance of ODPs from the perspective of user interfaces

For our analysis, we navigated these data portals as case studies to assess their alignment with the dimensions of promoting reuse in open data portals defined in Table 5.

*Table 5: Compliance assessment of EDP, FDP and SIDP with the dimensions of promoting reuse in open data portals*

| S.no | Requirement | European Data Portal | French Data Portal | Swedish Data Portal |
|---|---|---|---|---|
| 1 | Technology | CKAN plus other technologies | Udata plus other technologies | Ad hoc development |
| 2 | Social media links | The portal has links to its own Facebook, X, LinkedIn and YouTube profiles. It does not have the functionality to share individual datasets through any social media platform. | The portal has links to its own Facebook, LinkedIn, Mastodon and RSS profiles. It does not have the functionality to share individual datasets through any social media platform. | The portal has links to its own Facebook, and X profiles. It does not have the functionality to share individual datasets through any social media platform. |
| 3 | Feedback and support | Using the contact form, users can create different kinds of tickets: (1) request to update a dataset, (2) request for a specific dataset, (2) suggestion for a new dataset, (3) request for another format, (4) error report, (5) question on SPARQL API, (6) question on copyright, (7) general information, (8) feedback/other questions, and (9) request to be harvested by data.europa.eu | Using the contact form, users can create different kinds of tickets: (1) I have a question, (2) Request to open data, (3) Submit a bug, and (4) Give feedback on data.gouv.fr | The portal offers a general contact form. |
| 4 | Newsfeed | A newsfeed section with a wealth of current publications. | A newsfeed section with a wealth of current publications. | No newsfeed |
| 5 | Guidance | It has a documentation section with general site data, metadata models, | The documentation includes technical guides, data | The portal has a documentation section with |

| S.no | Requirement | European Data Portal | French Data Portal | Swedish Data Portal |
|---|---|---|---|---|
| | | search procedures, publication procedures, API documentation, metadata quality, data quality, data citation, data visualization, accessibility and glossary of terms.<br><br>The portal's academy offers educational material for different audiences such as academics, civil servants, data providers, developers, journalists, private sector, NGOs, and others.<br><br>It does not provide documentation at the level of individual datasets. | schema and an initiative called Open Data University that helps teachers and students design interesting school projects using open data.<br><br>It also has a resource section for individual datasets. | specifications of information and data models.<br><br>It does not provide documentation at the level of individual datasets. |
| 6 | Examples | The portal has general sections on apps and use cases.<br><br>No examples are available at the level of individual datasets in the catalogue. | Reuses are available both in a general section and linked to individual datasets. | There is no general examples section or at the level of individual datasets. |

As can be seen, the French open data portal is the one that offers the best conditions for the reuse of data. Its greatest distinction is that usage functions are directly associated with individual datasets, which facilitates the transfer of knowledge and resources between users to add value to the data.

### 4.2 Circularity compliance of ODPs from the perspective of feedback mechanisms

We performed desk research using the chapter 2 technical requirements for feedback mechanisms in ODP as the assessment measure for the feedback mechanisms of the three open data portals. The analysis of requirements' compliance is reflected in Table 6.

*Table 6: Technical requirements for the circular open data ecosystem with respect to feedback mechanism*

| S.no | Requirement | European Data Portal | French Data Portal | Swedish Data Portal |
|---|---|---|---|---|
| 1 | Quality Enhancement | Yes: The EDP strives for quality by publishing data according to specific standards and | Yes: Users can report data inaccuracies and suggest | Yes: Public and data supplier feedback tools exist for reporting |

| S.no | Requirement | European Data Portal | French Data Portal | Swedish Data Portal |
|---|---|---|---|---|
| | | metadata guidelines. However, user feedback on data accuracy and completeness is limited, potentially hindering further improvement. | improvements through forms. | data issues and suggesting improvements. |
| 2 | Usability and Accessibility | Yes: The EDP offers basic accessibility features like keyboard navigation and facet search compatibility. However, the interface could be further optimized for different user needs and abilities. | Yes: The portal offers commenting and feedback forms, potentially aiding user interaction. | Yes: Hierarchical feedback structure with clear categories aids user navigation. |
| 3 | Communication and Engagement | No: Users currently have limited options for providing feedback, primarily through email or forms, which may not encourage active engagement. | Yes: Multiple feedback channels (feedback forms, discussion forum, email, etc.) exist for user communication. | Yes: The portal counts user posts, suggesting user engagement is tracked and potentially used. |
| 4 | Iterative Development | No: While the ODP releases regular updates, there's limited evidence of user feedback directly influencing these updates, suggesting a less iterative development process. | No: Information on roadmaps or timelines for incorporating feedback is unavailable. | No: While feedback is collected, the text does not clarify its integration into updates or roadmaps. |

### 4.3 Circularity compliance of ODPs from the perspective of analysis tools

Desk-based research has been performed to check the compliance of the technical requirements with respect to analytical tools and technologies. The results are depicted in Table 7. The first column presents the technical requirements that are necessary to be integrated as functionalities in the current open data systems to make them more circular and enhance the integration of analytical tools and technologies. This integration, in return, can make the open data ecosystem circular with respect to value distribution and sustainability.

Symbols represent the following:

- "☑" indicates technical requirements are fulfilled.
- "☒" indicates technical requirements are not fulfilled.
- "±" indicates limited fulfilment of technical requirements.

*Table 7: Technical requirements for the circular open data ecosystem with respect to data analytics*

| S.no | Requirement | European Data Portal | French Data Portal | Swedish Data Portal |
|---|---|---|---|---|
| 1 | Req1:<br><br>Open Data accessibility and interoperability | ☑ Yes, these requirements are available in the EDP but with limited fulfilment. | ☑ Data is accessible and interoperable with some of the features of the interoperability. | ☑ Data is accessible and interoperable with some of the features of the interoperability. |
| 2 | Req2:<br><br>Comprehensive dataset coverage | ±Different thematic datasets are given around 13 categories are covered within the EDP, but these are not exhaustive categories. | ± The have themes which are available on the EDP but other than that they cover some hot themes as well. Such as machine learning useable datasets, covid themes datasets and so on. | ± The portal has fixed themes, which are expandable in the future. |
| 3 | Req3:<br><br>Scalability and performance of the portal | ± These are subjective requirements. And EDP follows cloud infrastructure which is scalable, and performance is also good. | ± These are subjective requirements. And French open data portals follow cloud infrastructure which is scalable, and performance is also good. | ± These are subjective requirements. And Swedish follows cloud infrastructure which is scalable, and performance is also good. |
| 4 | Req4:<br><br>User-friendly interfaces | ☑ Reasonable user-friendly interfaces; requires an improvement over the time. | ⬇ Reasonable user-friendly interfaces; requires an improvement over the time. | ⬇ Reasonable user-friendly interfaces; requires an improvement over the time. |
| 5 | Req5:<br><br>Advanced searches and filtering capabilities | ☑ EDP provides with the functionality of advanced searches and filtering capabilities. | ☑ French OD portal provides with the medium functionality of advanced searches and filtering capabilities. | ☑ Swedish OD portal provides with the medium functionality of advanced searches and filtering capabilities. |
| 6 | Req6:<br><br>Data governance (data quality assurance) | ☑ Data governance procedures implied within the EDP. | ☑ National and European data governance procedures are being followed in the French Open data portal. | ☑ National and European data governance procedures are being followed in the Swedish Open data portal. |
| 7 | Req7:<br><br>Analytical tool integration | ± Limited availability of analytical tools integrated within the EDP. For instance, | ± Limited availability of analytical tools integrated within the FDP. For instance, | ± Limited availability of analytical tools integrated within the SDP. For instance, |

| S.no | Requirement | European Data Portal | French Data Portal | Swedish Data Portal |
|---|---|---|---|---|
| | | various kind of datasets which requires different types of analysis tools. So, it is very difficult to build a generic tool. | various kind of datasets which requires different types of analysis tools. So, it is very difficult to build a generic tool. | various kind of datasets which requires different types of analysis tools. So, it is very difficult to build a generic tool. |
| 8 | Req8: Collaboration and knowledge sharing | ±Limited collaboration and knowledge sharing. | ±Limited collaboration and knowledge sharing. | ± Limited collaboration and knowledge sharing. |
| 9 | Req9: Feedback mechanisms | ± Limited feedback mechanisms implemented. | ±Limited feedback mechanisms implemented. | ± Limited feedback mechanisms implemented. |
| 10 | Req10: Documentation and Training Resources | ☑ documentation and training resources are available on the website. | ☑ documentation and training resources are available on the website. | ☑ documentation and training resources are available on the website. |
| 11 | Req11: Technological infrastructure (cloud computing) | ☑ Technological infrastructure is implemented based cloud computing tools. | ± Not sure about technological infrastructure. | ±Not sure about technological infrastructure. |
| 12 | Req12: Data categorization with respect to possible analytics | ☒ Not available data categorization. | ☑ Yes, available data categorization. | ☒ Not available data categorization. |

Based on Table 6 (Technical requirements for the circular open data ecosystem with respect to data analytics) and compliance with current open data portals (EDP, FDP, and SDP), a smaller number of technical requirements are being fulfilled by these portals for better circularity. These technical requirements revolve around data analytics overall. The adoption of these technical requirements and the development of functionalities can improve the circularity of the open data by means of data analytics at the end.

### 4.4 Circularity compliance of ODPs with technical requirements of AI and CI

In this section, we investigated these portals as case studies to evaluate their alignment with the established technical requirements for integrating AI and CI. This assessment is aimed to determine the extent to which Europe's most prominent data portals meet these critical criteria.

The findings of our evaluation are reflected in Table 8, which presents a detailed analysis of the compliance with each requirement.

*Table 8: Compliance assessment of EDP, FDP and SDP with established AI and CI technical requirements*

| S.no | Requirement | European Data Portal | French Data Portal | Swedish Data Portal |
|---|---|---|---|---|
| 1 | High-Quality and Structured Data | • Datasets with varying languages depending on the source.<br>• Loosely structured Data.<br>• A lot of data sets are not well defined. | • Unilingual data Well defined and well-structured data. | • Unilingual data.<br>• Mostly well-structured data. |
| 2 | Comprehensive Metadata | • Comprehensive metadata attributes.<br>• Incomplete and minimal metadata. | • Well-defined and mostly complete quality metadata. | • Basic but well defined and complete metadata. |
| 3 | Data Interoperability | • Many datasets were in limited formats.<br>• Abundantly lacked machine readability.<br>• A lot of them did not adhere to DCAT-AP principles. | • Better adherence to DCAT-AP.<br>• Format variability poses challenges. | • Moderate adherence to interoperability standards with some exceptions. |
| 4 | Real-time Data Processing Capabilities | • The concept of real-time data processing for AI and CI was minimal and virtually non-existent. | • Some initiatives for real-time data, but not widely implemented. | • Non-existent real-time data processing capabilities, mainly static datasets. |
| 5 | Robust and Scalable Infrastructure | • The infrastructure resembled a conventional data portal.<br>• Not optimized for AI operations.<br>• CI integration also appeared limited. | • Primarily conventional infrastructure with limited AI optimization.<br>• It is conducive for CI as it engages community for discussion and | • Infrastructure more tailored to conventional data storage than AI processing.<br>• It also aids CI with community engagement with giving |

| S.no | Requirement | European Data Portal | French Data Portal | Swedish Data Portal |
|---|---|---|---|---|
| | | • Limited SPARQL API query time | integrates reuses of datasets.<br>• General search API not apparent | hierarchical commenting tool to engage users.<br>• General search API not apparent |
| 6 | User-friendly Interfaces and Tools | • The interface did not seem conducive to integrating AI or CI effectively. | • Interface is user-friendly but lacks advanced tools for AI and CI integration. | • User-friendly interface; however, advanced tools for AI integration are limited. |
| 7 | Compliance with Legal and Ethical Standards | • Data from various sources showed inconsistencies in licensing and compliance. | • High compliance with legal standards, but ethical guidelines are less clear. | • Generally compliant with legal standards, but some data sources show inconsistencies. |
| 8 | Feedback Mechanisms for Continuous Improvement | • Although data visualization was present, there were no clear feedback mechanisms reflecting crowdsourcing principles. | • Integrates CI well from the perspective of community engagement and value integration.<br>• Focuses most on user needs. | • Integrates community engagement with commenting and discussion options but it is hierarchical. |

Our analysis of these European portals highlights distinct levels of compliance with AI and CI technical requirements, as summarized in Table 4.

The EDP, despite offering multilingual datasets, falls short in high-quality, structured data and real-time data processing capabilities. Its traditional infrastructure limits its effectiveness in AI and CI integration. On the other hand, the FDP stands out for its strong compliance with CI principles. It features a user-friendly interface, effective feedback mechanisms, and integrates data reuses well, showcasing a robust model for CI integration. However, it faces some challenges in data interoperability.

The SDP, while providing well-structured data and comprehensive metadata, struggles with interoperability and real-time data processing. Its conventional infrastructure, though somewhat conducive to CI, lacks in AI optimization.

In conclusion, while the FDP demonstrates strong CI integration, all portals need further development in certain areas, particularly in data quality, interoperability, and structure to support AI capabilities, to fully leverage the potential of AI and CI in European open data ecosystems.

# 5 Strategies to enhance circularity in open data portals

## 5.1 Designing user interfaces for open data portals where stakeholders can readily materially add value to the ecosystem

In this section, we present a proposal to guide the design of open government data interfaces to foster circularity. Our goal is not to prescribe every detail of a platform, but rather to invite leaders of open government data initiatives to reimagine the paradigm needed to cultivate a true ecosystem that creates and captures value from different types of users. At the centre of our proposal is the notion of the open data portal, which is a prominent technological facet of open government data initiatives. However, it is important to note that, while significant, the open data portal should not be perceived as the only channel for achieving this goal.

One strategy often employed by designers to manage complexity is to anchor user interface actions, tasks and goals in a recognisable set of concepts [66]. This framework, known as a user interface metaphor, has significantly influenced interface design practices. Metaphors facilitate the transfer or alignment of knowledge from a known domain to an unknown domain, allowing people to apply existing knowledge and experience to understand and navigate new or unfamiliar situations. It is precisely this metaphor-driven design model that we will use in the following lines. The initial phase of this process is to map the conceptual metaphor prevalent in existing data portals (e.g. EDP, FDP, SDP). This can be achieved through a close examination of the platform's user interface [67]. We can critically evaluate the strengths and weaknesses of the metaphors identified from the perspective of circularity. We then look for an alternative metaphor that can effectively address the limitations of the initial model, ideally complementing its strengths. Finally, we propose a course of action to implement this alternative metaphor in real portals.

For this exercise, we have chosen to use metaphors associated with physical meeting places. Physical metaphors can serve as valuable tools, as tangible objects tend to resonate more intuitively with humans compared to the abstract nature of most modern digital data sources [68]. A preliminary examination of the open data portals analysed in this research reveals the presence of at least three distinct metaphors that effectively describe the information architecture of these platforms. The main metaphor observed is that of **the library** - collection of resources in a variety of formats conveniently organised for physical, digital, bibliographic or intellectual access-[69]. It is joined by two other metaphors of lesser relevance but notable influence: **the school** -space dedicated to learning and the cultivation of knowledge and skills-, and the **community centre** - public space where community members convene for group activities, social support, access to public information, and various communal endeavours-[70]. Figure 10 shows the correspondence between the various sections of the open data portals in question and the identified design metaphors.

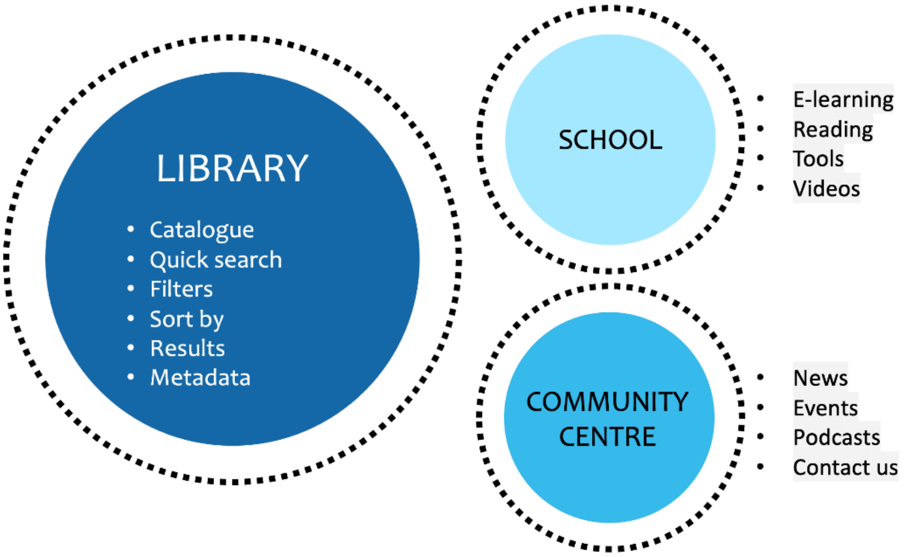


*Figure 10: Elements of open data portals that map to design metaphors*

Note that these metaphors aim at satisfying the value creation objectives of different user groups. However, bottlenecks in value creation and circulation can arise when these metaphors are implemented as separate sections that exist within the same website but lack organic interaction. This restriction in the circulation of value is precisely evident in the EDP, where the **library**, **school** and **community centre** manifest themselves as separate tabs with no direct interconnection. In contrast, the FDP presents a different scenario. Each dataset is directly linked to reuses, discussions and community resources, which facilitates the capture of value from users for the benefit of other users and the provider.

The next step is to identify an alternative metaphor. The challenge is to reconceptualise the portal beyond a mere catalogue, which typically presents data sets for individual consumption through browsing and consultation. Instead, while still emphasising convenience factors, user interfaces should encourage visitors to engage in a more communal experience of interacting with data. This means fostering an environment where users can not only access data, but also learn and share knowledge in collaboration with others. This interaction helps to bridge the technical and contextual knowledge gap needed to extract value from data at different stages of its lifecycle.

As a competing metaphor, we propose **the makerspace** (also known as hackerspace, hack lab or FabLab). One of such workshops at hackerspace is reflected in Figure 11. This physical collaborative workspace serves as a hub for creativity, learning, exploration, and knowledge sharing [71]. It allows people to come in with an idea and leave with products or prototypes with varying degrees of fidelity. The makerspace has a wide range of tools and resources, from high-tech equipment to basic materials, for users of all ages and backgrounds. There is an inherent connection between makerspaces and circularity [72]. This connection is not only present in the reuse of physical materials and tools but also in core practices such as exchange, collaboration, education, and guidance.

*Figure 11: Workshop for beginners at the Kyiv hackerspace held in January 2023 (Source: Wikimedia Commons)*

A makerspace has five key components: (1) space, (2) members and community, (3) resources, (4) activities, (5) values/philosophy [73]. Table 9 maps these makerspace components to a standard government open data portal.

*Table 9: Mapping of makerspace components to a standard government open data portal [67]*

| | Makerspace | Open Government Data Portal |
|---|---|---|
| **Members & Community** | Maker, hacker, artists, engineer | Government, NGOs, companies, students, journalists, non-specialised users, intermediaries |
| **Resources** | Supplies (e.g. plastic, metal, wood) and tools (e.g. welders, laser cutters, microcontrollers, 3D printers, digital design tools) | Data as raw material and tools for data cleaning, analysis, modelling and visualisation |
| **Activities** | Making, hacking, programming, entrepreneurship | Open data lifecycle activities [See section 2.4] |
| **Values** | Open, sharing, learning, collaboration. | |

Makerspaces are defined by the proximity of materials such as plastic, metal and wood, together with a range of tools such as welding machines, laser cutters, microcontrollers, 3D printers and digital design tools, which facilitate the creation of products. Translating this concept to the government's open data portal means ensuring that datasets are easily accessible along with tools for cleaning, analysing, visualising and modelling them. This can be achieved by providing on-site tools or by streamlining the process of exporting data to compatible tools.

The iterative process adopted in makerspaces recognises that products evolve gradually rather than reaching their final form abruptly [74]. Hence, makerspaces offer prototyping facilities with levels of fidelity varying accordingly. Among these facilities, rapid prototyping occupies a central place. Similarly, in the context of an open government data portal, the selection of tools should prioritise options that offer users a complete view of the data, allowing them to quickly assess its potential for a given purpose and further exploration.

Makerspaces prioritise learning and adopting methodologies aimed at fostering product creation, including Design Thinking [75]. Illustrated in Figure 12. Design Thinking is a user-centred problem-solving approach characterised by a systematic flow comprising three key stages: 1) understanding, 2) exploration and 3) creation. Within these general phases, the process unfolds through six distinct steps: empathise, define, ideate, prototype, test and implement [76]. Regarding this, the information, complementary materials, and tools around a dataset could be organised to facilitate users to seamlessly navigate through the phases of understanding, exploration, and creation.

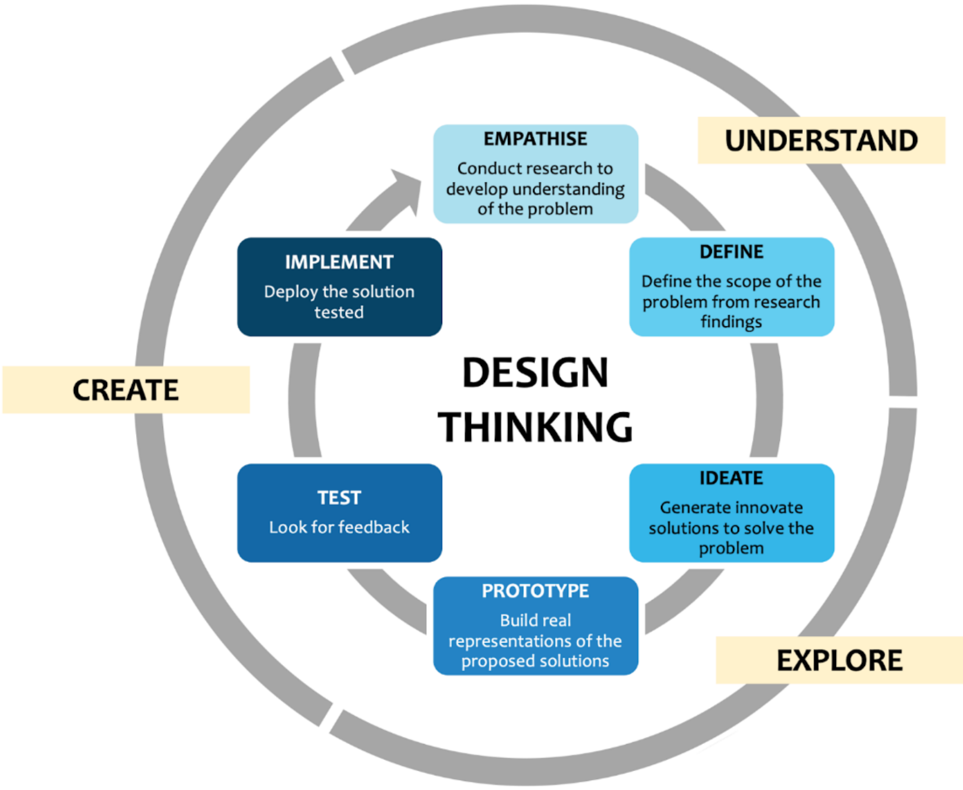


*Figure 12: Design Thinking [76]*

Below is a sequence of sketches to illustrate some elements of the **makerspace** metaphor described previously (proximity between inputs and, resources and tools; fast prototyping; design process facilitation) put into practice in a governmental open data portal. It is by no means intended as a detailed design specification. The sketch in Figure 13 shows the record of a dataset. At the top are the characteristic elements of current dataset sheets present in portals, containing titles, descriptions, and other important metadata, as well as buttons for basic operations such as downloading the dataset. The lower part shows a sequence of tabs that from left to right evoke the stages of understanding, exploration and creation typical of design thinking. Note that the specific labels and number of compartments into which the steps of a design process are distributed is debatable, the important idea is that the resulting experience is a flow that provides users with the context and resources to create value from the data and return it to the ecosystem.

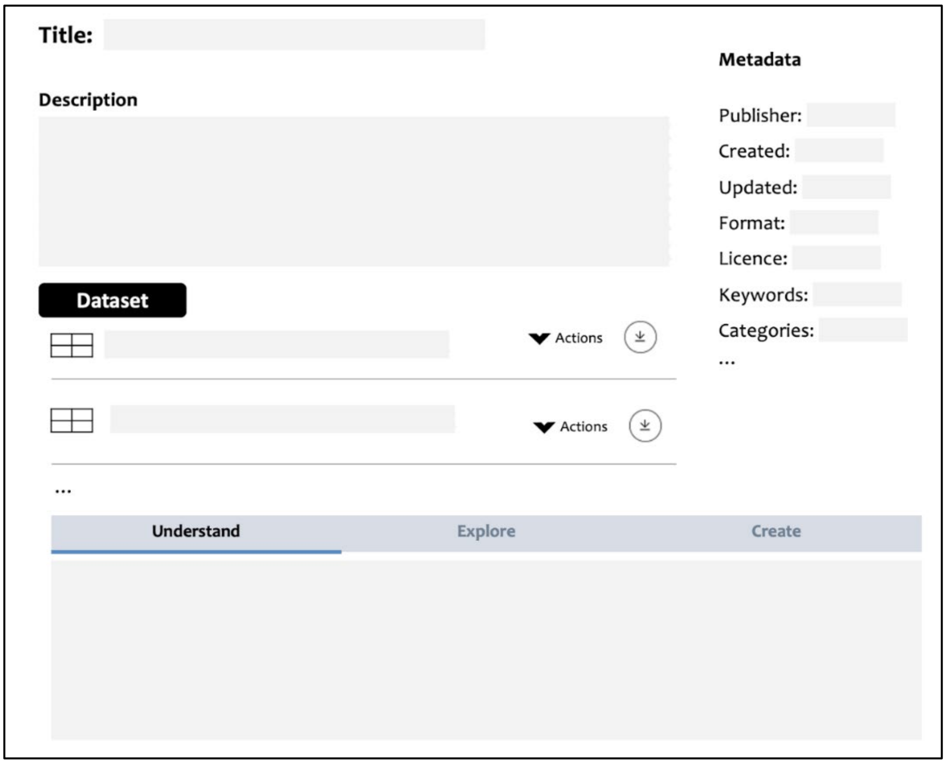


*Figure 13: Dataset record*

The sketch in Figure 14 elaborates on the concept of the **Understand tab**. The purpose of this section would be for the user to quickly discover what the data consists of from a technical and domain expert point of view. The technical description of the data is given by data profiling (variable dictionary, size, data preview, quality indicators). Domain knowledge is supplied d by a) Documents relevant to the interpretation of the data (e.g. a dataset on public contracts could show the public procurement laws and regulations of the corresponding jurisdiction), b) Use cases that help the user to draw inspiration from previous exploitations of the dataset. In this section, users should be able to connect or upload their re-uses in as many formats as possible. c) Communities of users related to the dataset either by thematic proximity or by having developed previous projects with this resource, d) Related data sets that could be combined with the current one to generate products with greater added value, and e) Projects where users or Communities could advertise initiatives that require volunteers or collaborators.

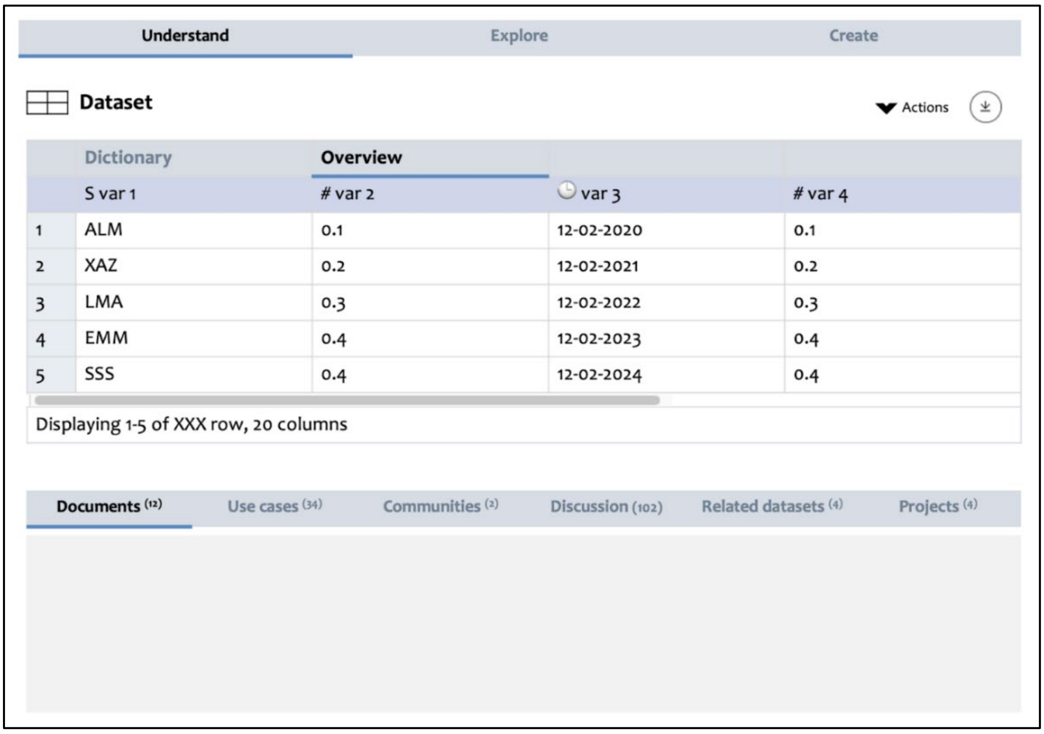


*Figure 14: Understand tab*

In the sketch of the **Explore tab** shown in Figure 15, the user, once understanding the context of the data, has a toolkit to carry out typical operations for working with data (cleaning, analysis, visualization, modelling). This section could also have an artificial intelligence assistant that, knowing what the user intends to do with the dataset, can advise them on its feasibility, strategy for its execution, useful resources (documents, communities, projects, related datasets, etc.). In this section the user can create data artifacts of varying complexity (tables, visualizations, models.). These artifacts must have a modular nature and be highly portable to be able to be integrated into larger products, shared or transferred to other platforms in various formats.

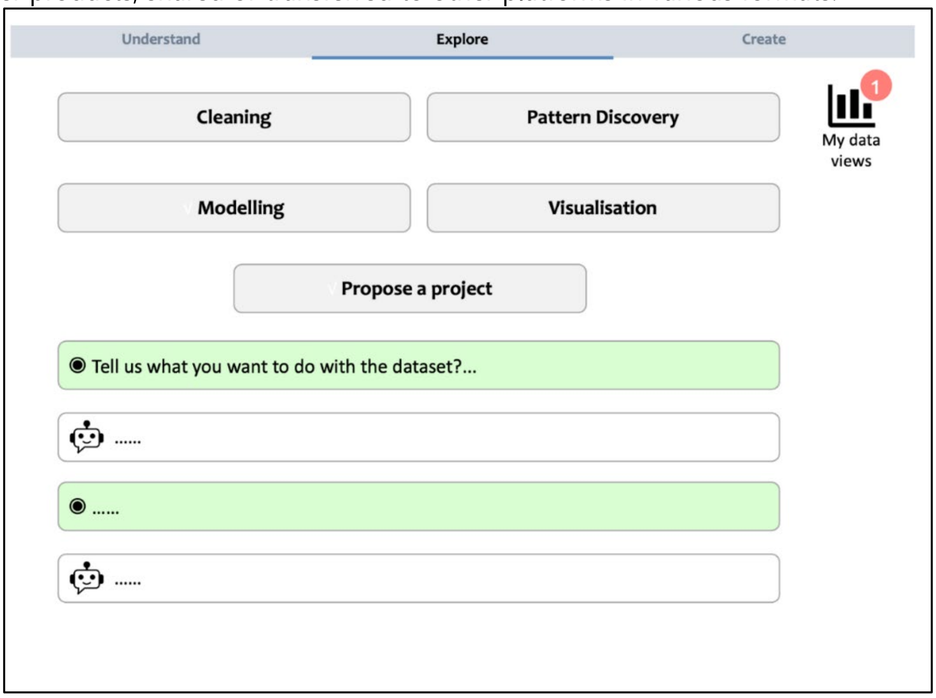


*Figure 15: Explore tab*

Finally, the **Create tab** view of Figure 16 offers a space for the user to assemble higher value-added data products from smaller pieces like those that were generated in the previous stage. A possible implementation of this idea is data notebooks, which in their ultimate expression are a narrative product that integrates language, data and multimedia whose purpose is to facilitate the communication of findings. The creation of these products can be done collaboratively and invite other users to provide feedback. The result could be shared in a wide number of formats and of course incorporated into the list of reuses present in the **Understand tab**.

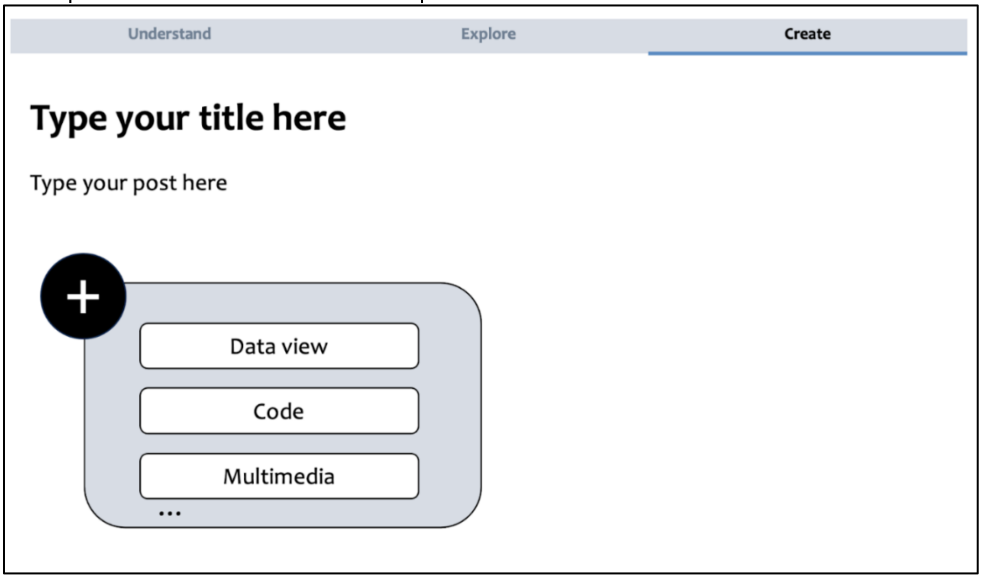


*Figure 16: Create tab*

## 5.2 Proposed conceptual model of feedback mechanism to make circular open data ecosystem

Figure 17 illustrates the system architecture designed to efficiently process data from multiple sources (S1, S2, S3) and transform it into actionable insights accessible through a dashboard interface. To better understand the conceptual model, let us dive into the data source integration. For example, data S1 can crawl the tweets related with open data by the users and by the portal administrators. While data source S2, can harvest the themes of the datasets within open data portal and the last data source S3 can provide a user opinion about the dataset as well as regarding the portal itself. Once data is collected in a database from all these sources, this database will act as a hub, facilitating seamless communication between the data collection components and downstream processing modules. Furthermore, upon ingestion into the database, the stored data undergoes algorithmic analysis to extract meaningful insights and patterns. The results of this analysis form the basis for decision-making and further action. Finally, the insights derived from the algorithmic analysis are presented to users through a dashboard interface. The dashboard provides a user-friendly way to interact with the data, visualizing key metrics, trends, and insights in a comprehensible manner.

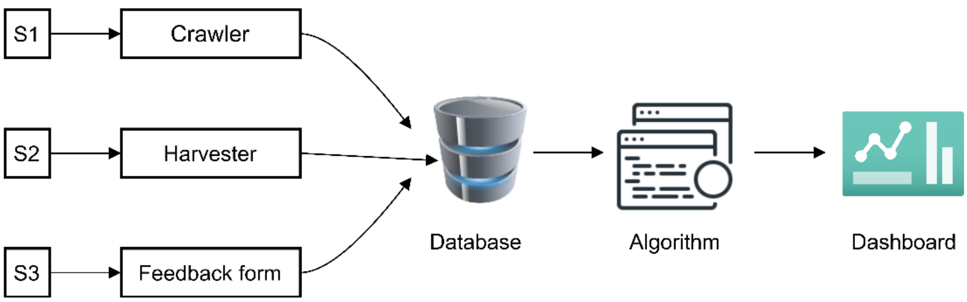


*Figure 17: Proposed Conceptual Model*

### 5.3 Enhancing Open Data Ecosystems Circularity by incorporating open data tools and technologies within open data portals

Merely offering data to users is not sufficient without considering how they can leverage the data for valuable insights and creations. In today's rapidly advancing landscape, where new tools and technologies continually emerge for handling datasets of all sizes and types, it is crucial to strategize open data provision. Enhancing circularity in current open data spaces involves ensuring availability of open-source analytical tools and technologies. Few important recommendations to enhance the ODECO circularity:

- Providing open data based on use-cases.
- Recommendations for tools and technologies that end-users can utilize to analyse particular datasets.
- Providing applications (data products) developed using open datasets.
- Incentivise the open data analytics (e.g. Kaggle competitions).
- Integration of ease-to-use data analytical tools for non-expert stakeholders.
- Offering courses and training modules on open data analytics within open data environments to facilitate capacity building.
- Highlighting the expertise level required for the specific open datasets.

Figure 18 illustrates the aspects of open dataset provisions, highlighting strategies aimed at enhancing circularity. These strategies emphasize the importance of making datasets accessible or categorized to various stakeholders based on their expertise, incentives, and specific use cases (the list is expanding denoted by dotted lines).

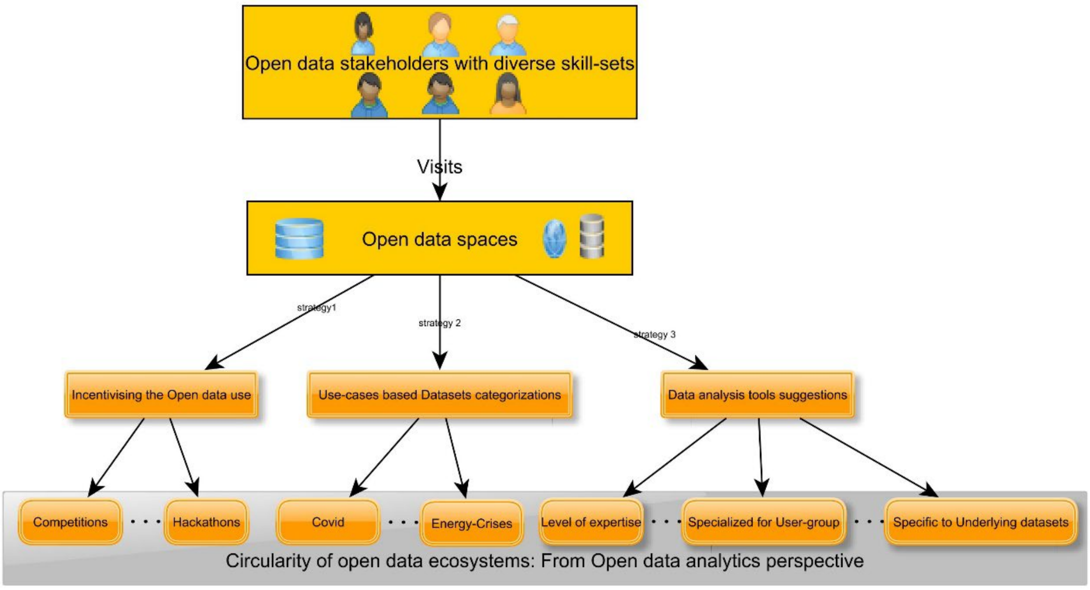


*Figure 18: Enhancing circularity in open data ecosystem with respect to data analytics (Source: Authors)*

### 5.4 Enhancing Open Data Ecosystem by employing AI and CI through APIs

The strategic integration of Artificial Intelligence (AI) and Collective Intelligence (CI) through APIs presents a transformative approach to enriching the Open Data Ecosystem, as reflected in Figure 19. By harnessing AI, we aim to elevate the ecosystem's capabilities in real-time data searching, data discovery, and metadata generation, ensuring fast, secure, and intelligent accessibility. This integration facilitates a more nuanced and dynamic interaction with data, enabling AI to provide predictive insights, advanced data analytics, and tailored data recommendations. On the Collective Intelligence front, inspired by platforms like GitHub, we envision an inclusive environment where users can actively engage in data editing, refinement, and contribution. Through these enhanced APIs, users can not only manipulate and analyse open data, but also feed their analytical outcomes and enhancements back into the system, fostering a continuous loop of improvement and innovation. This approach not only democratizes open data engagement, allowing for a broader spectrum of user interaction and contribution but also ensures that the open data ecosystem evolves to become more circular, inclusive, insightful, and impactful.

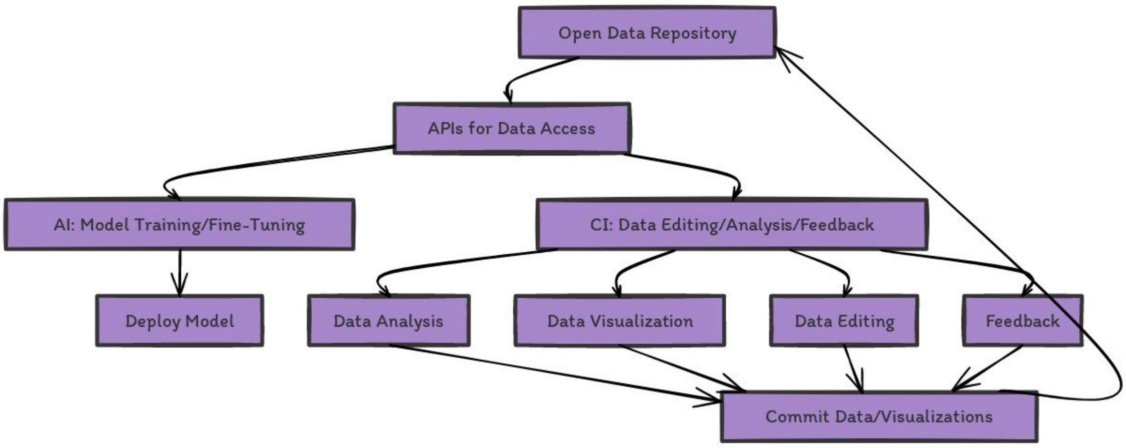


*Figure 19: Proposed methodology to enhance ODECO by using AI and CI through API*

# 6 Conclusions

The comprehensive exploration of open data portals (ODPs) and their role in fostering circular open data ecosystems reveals a multifaceted landscape of challenges and opportunities. This document has delved into various aspects of ODPs, including user interface design, feedback mechanisms, analytical tools, and the integration of Artificial Intelligence (AI) and Collective Intelligence (CI). The findings show the importance of circularity, and how it can be facilitated technically in open data ecosystems, highlighting the need for continuous refinement and dynamic interactions between data consumers and producers.

The analysis of the European Data Portal (EDP), the French, and Swedish open data portals from different technical perspectives has shown varying degrees of compliance with the principles of circularity. While some portals excel in certain areas, such as community engagement and user-friendly interfaces, others face challenges in data quality, interoperability, and real-time data processing capabilities. The integration of AI and CI into these ecosystems is identified as a key factor in enhancing their functionality and sustainability. However, this integration requires addressing several technical requirements, including high-quality and structured data, comprehensive metadata, data interoperability, and robust infrastructure.

The concept of circularity in open data ecosystems, inspired by the principles of the circular economy, emphasizes the need for a paradigm shift from linear data consumption to a more sustainable, collaborative, and value-creating approach. This shift is crucial for maximizing the benefits of open data, fostering innovation, and ensuring fair distribution of value among stakeholders. The document highlights the role of feedback mechanisms in enhancing the quality, usability, and accessibility of data, thereby transforming open data from a static resource into a dynamic catalyst for innovation and societal progress.

In conclusion, the journey towards achieving circularity in open data ecosystems is ongoing, with significant progress made in some areas and potential for improvement in others. The insights gained from this analysis provide a roadmap for future developments, emphasizing the need for collaborative efforts, continuous improvement, and the integration of advanced technologies to realize the full potential of open data. By embracing these principles, open data ecosystems can evolve into more efficient, transparent, and user-centric platforms, contributing to a more knowledgeable, interconnected, and sustainable global community.

# 7 Bibliography


[1] Publications Office of the European Union *et al.*, *Rethinking the impact of open data: a first step towards a European impact assessment for open data*. LU: Publications Office of the European Union, 2023. Accessed: Jan. 29, 2024. [Online]. Available: https://data.europa.eu/doi/10.2830/911822

[2] European Commission, "Digital Agenda: Commission's Open Data Strategy, Questions & answers," European Commission. Accessed: May 04, 2023. [Online]. Available: https://ec.europa.eu/commission/presscorner/detail/en/MEMO_11_891

[3] Publications Office of the European Union and E. Huyer, *The economic impact of open data: opportunities for value creation in Europe*. LU: Publications Office of the European Union, 2020. Accessed: Jan. 29, 2024. [Online]. Available: https://data.europa.eu/doi/10.2830/63132

[4] S. Neumaier, J. Umbrich, and A. Polleres, "Automated Quality Assessment of Metadata across Open Data Portals," *J. Data and Information Quality*, vol. 8, no. 1, p. 2:1-2:29, Oct. 2016, doi: 10.1145/2964909.

[5] European Commission. Directorate General for the Information Society and Media. *et al.*, *Creating value through open data: study on the impact of re use of public data resources*. LU: Publications Office, 2015. Accessed: Jan. 31, 2024. [Online]. Available: https://data.europa.eu/doi/10.2759/328101

[6] Publications Office of the European Union., *Open data maturity report 2023.* LU: Publications Office, 2023. Accessed: Jan. 29, 2024. [Online]. Available: https://data.europa.eu/doi/10.2830/384422

[7] C. Hendler, "Report card: Obama's marks at Transparency U," *Columbia Journalism Review*, vol. 5, 2010.

[8] B. Ubaldi, "Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives," OECD, Paris, May 2013. doi: 10.1787/5k46bj4f03s7-en.

[9] C. Alexopoulos, E. Loukis, and Y. Charalabidis, "A Platform for Closing the Open Data Feedback Loop Based on Web2.0 Functionality," *JeDEM - eJournal of eDemocracy and Open Government*, vol. 6, no. 1, Art. no. 1, Nov. 2014, doi: 10.29379/jedem.v6i1.327.

[10] Y. Charalabidis, A. Zuiderwijk, C. Alexopoulos, M. Janssen, T. Lampoltshammer, and E. Ferro, "The Multiple Life Cycles of Open Data Creation and Use," in *The World of Open Data: Concepts, Methods, Tools and Experiences*, Y. Charalabidis, A. Zuiderwijk, C. Alexopoulos, M. Janssen, T. Lampoltshammer, and E. Ferro, Eds., in Public Administration and Information Technology. , Cham: Springer International Publishing, 2018, pp. 11–31. doi: 10.1007/978-3-319-90850-2_2.

[11] B. Van Loenen *et al.*, "Towards value-creating and sustainable open data ecosystems: A comparative case study and a research agenda," *JeDEM*, vol. 13, no. 2, pp. 1–27, Dec. 2021, doi: 10.29379/jedem.v13i2.644.

[12] R. Pollock, "Building the (Open) Data Ecosystem – Open Knowledge Foundation blog." Accessed: May 16, 2023. [Online]. Available: https://blog.okfn.org/2011/03/31/building-the-open-data-ecosystem/

[13] B. S. I. Group, "Executive Briefing: BS 8001–a Guide: The world's first standard for implementing the principles of the circular economy in organizations."

[14] Y. Charalabidis, C. Alexopoulos, V. Diamantopoulou, and A. Androutsopoulou, "An Open Data and Open Services Repository for Supporting Citizen-Driven Application Development for Governance," in *2016 49th Hawaii International Conference on System Sciences (HICSS)*, Jan. 2016, pp. 2596–2604. doi: 10.1109/HICSS.2016.325.

[15] J. Kirchherr, N.-H. N. Yang, F. Schulze-Spüntrup, M. J. Heerink, and K. Hartley, "Conceptualizing the Circular Economy (Revisited): An Analysis of 221 Definitions," *Resources, Conservation and Recycling*, vol. 194, p. 107001, Jul. 2023, doi: 10.1016/j.resconrec.2023.107001.

[16] "How to Build a Circular Economy | Ellen MacArthur Foundation." Accessed: Jan. 31, 2024. [Online]. Available: https://www.ellenmacarthurfoundation.org/

[17] "The technical cycle of the butterfly diagram." Accessed: Jan. 31, 2024. [Online]. Available: https://www.ellenmacarthurfoundation.org/articles/the-technical-cycle-of-the-butterfly-diagram

[18] M. Cioffi, J. Goldman, and S. Marchese, *Harvard Biomedical Research Data Lifecycle*. Zenodo, 2023. doi: 10.5281/zenodo.8076168.

[19] "8 Steps in the Data Life Cycle | HBS Online," Business Insights Blog. Accessed: Jan. 31, 2024. [Online]. Available: https://online.hbs.edu/blog/post/data-life-cycle

[20] K. Rahul and R. K. Banyal, "Data Life Cycle Management in Big Data Analytics," *Procedia Computer Science*, vol. 173, pp. 364–371, Jan. 2020, doi: 10.1016/j.procs.2020.06.042.

[21] T. Patil and T. Davenport, "Data scientist: The sexiest job of the 21st century," *Harvard business review*, vol. 90, no. 10, pp. 70–76, 2012.

[22] "The Data Science Venn Diagram," Drew Conway. Accessed: Jan. 31, 2024. [Online]. Available: http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

[23] "Bachelor thesis: Conception and prototyping of a modern browser's user interface." Accessed: Mar. 08, 2024. [Online]. Available: https://www.greinr.com/blog/2012/conception-and-prototyping-of-a-modern-browsers-user-interface/

[24] A. Zuiderwijk, M. Janssen, and C. Davis, "Innovation with open data: Essential elements of open data ecosystems," *Information Polity*, vol. 19, no. 1–2, pp. 17–33, 2014, doi: 10.3233/IP-140329.

[25] J. Crusoe, "Open Data Ecosystem – The Data Market between Municipalities and Businesses," Linköping, Sweden, 2016.

[26] Capgemini Invent, European Data Portal, Publications Office of the European Union, E. Simperl, and J. Walker, *The future of open data portals*. Publications Office of the European Union, 2020. Accessed: Apr. 03, 2024. [Online]. Available: https://data.europa.eu/doi/10.2830/879461

[27] "Sharing and re-using open data: A case study of motivations in astrophysics - ScienceDirect." Accessed: Apr. 03, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0268401218311836

[28] "Do patient engagement interventions work for all patients? A systematic review and realist synthesis of interventions to enhance patient safety - Newman - 2021 - Health Expectations - Wiley Online Library." Accessed: Mar. 08, 2024. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/hex.13343

[29] C. S. Kruse, D. A. Argueta, L. Lopez, and A. Nair, "Patient and Provider Attitudes Toward the Use of Patient Portals for the Management of Chronic Disease: A Systematic Review," *Journal of Medical Internet Research*, vol. 17, no. 2, p. e3703, Feb. 2015, doi: 10.2196/jmir.3703.

[30] H. Li, Y.-Y. Chang, J. Y. Lee, I. Bahar, and L.-W. Yang, "DynOmics: dynamics of structural proteome and beyond," *Nucleic Acids Research*, vol. 45, no. W1, pp. W374–W380, Jul. 2017, doi: 10.1093/nar/gkx385.

[31] J. L. Baldwin, H. Singh, D. F. Sittig, and T. D. Giardina, "Patient portals and health apps: Pitfalls, promises, and what one might learn from the other," *Healthcare*, vol. 5, no. 3, pp. 81–85, Sep. 2017, doi: 10.1016/j.hjdsi.2016.08.004.

[32] K. M. Nazi, C. L. Turvey, D. M. Klein, and T. P. Hogan, "A Decade of Veteran Voices: Examining Patient Portal Enhancements Through the Lens of User-Centered Design," *Journal of Medical Internet Research*, vol. 20, no. 7, p. e10413, Jul. 2018, doi: 10.2196/10413.

[33] A. Simonofski, A. Zuiderwijk, A. Clarinval, and W. Hammedi, "Tailoring open government data portals for lay citizens: A gamification theory approach," *International Journal of Information Management*, vol. 65, p. 102511, Aug. 2022, doi: 10.1016/j.ijinfomgt.2022.102511.

[34] "Quality of Metadata in Open Data Portals | IEEE Journals & Magazine | IEEE Xplore." Accessed: Mar. 08, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9405650

[35] "Trust in open data applications through transparency - Christian Wiencierz, Marco Lünich, 2022." Accessed: Mar. 08, 2024. [Online]. Available: https://journals.sagepub.com/doi/full/10.1177/1461444820979708

[36] M. Lněnička, R. Machova, J. Volejníková, V. Linhartová, R. Knezackova, and M. Hub, “Enhancing transparency through open government data: the case of data portals and their features and capabilities,” *Online Information Review*, vol. 45, no. 6, pp. 1021–1038, Jan. 2021, doi: 10.1108/OIR-05-2020-0204.
[37] H. Zhang and J. Xiao, “Quality assessment framework for open government data: Meta-synthesis of qualitative research, 2009-2019,” *The Electronic Library*, vol. 38, no. 2, pp. 209–222, Jan. 2020, doi: 10.1108/EL-06-2019-0145.
[38] A. Nikiforova and K. McBride, “Open government data portal usability: A user-centred usability analysis of 41 open government data portals,” *Telematics and Informatics*, vol. 58, p. 101539, May 2021, doi: 10.1016/j.tele.2020.101539.
[39] A. Purwanto, A. Zuiderwijk, and M. Janssen, “Citizen engagement with open government data: Lessons learned from Indonesia’s presidential election,” *Transforming Government: People, Process and Policy*, vol. 14, no. 1, pp. 1–30, Jan. 2020, doi: 10.1108/TG-06-2019-0051.
[40] “Full article: Improving the speed and ease of open data use through metadata, interaction mechanisms, and quality indicators.” Accessed: Mar. 08, 2024. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/10919392.2015.1125180
[41] M. Alnajjar and S. S. A. Naser, “Improving Quality Of Feedback Mechanism In Un By Using Data Mining Techniques,” *International Journal of Soft Computing, Mathematics and Control*, vol. 4, no. 2, 2015, Accessed: Mar. 08, 2024. [Online]. Available: https://www.researchgate.net/profile/Samy-Abu-Naser/publication/301324472_Improving_Quality_of_Feedback_Mechanism_in_UN_by_Using_Data_Mining_Techniques/links/5717612f08aefb153f9edf91/Improving-Quality-of-Feedback-Mechanism-in-UN-by-Using-Data-Mining-Techniques.pdf?origin=journalDetail&_tp=eyJwYWdlIjoiam91cm5hbERldGFpbCJ9
[42] D. Di Staso *et al.*, “Closing the cycle: Understanding potential contributions of open government data users to the open data ecosystem”, Accessed: Jan. 29, 2024. [Online]. Available: https://lirias.kuleuven.be/4132428
[43] R. Mehmood, M. A. Faisal, and S. Altowaijri, “Future Networked Healthcare Systems: A Review and Case Study,” in *Handbook of Research on Redesigning the Future of Internet Architectures*, IGI Global, 2015, pp. 531–558. doi: 10.4018/978-1-4666-8371-6.ch022.
[44] G. Smith, H. A. Ofe, and J. Sandberg, “Digital Service Innovation from Open Data: Exploring the Value Proposition of an Open Data Marketplace,” in *2016 49th Hawaii International Conference on System Sciences (HICSS)*, Jan. 2016, pp. 1277–1286. doi: 10.1109/HICSS.2016.162.
[45] A. Zuiderwijk, M. Janssen, G. van de Kaa, and K. Poulis, “The wicked problem of commercial value creation in open data ecosystems: Policy guidelines for governments,” *Information Polity*, vol. 21, no. 3, pp. 223–236, Jan. 2016, doi: 10.3233/IP-160391.
[46] “Evaluation of Open-Source Tools for Big Data Processing | Dutse Journal of Pure and Applied Sciences”, Accessed: Mar. 08, 2024. [Online]. Available: https://www.ajol.info/index.php/dujopas/article/view/235496
[47] M. Lnenicka and J. Komarkova, “Big and open linked data analytics ecosystem: Theoretical background and essential elements,” *Government Information Quarterly*, vol. 36, no. 1, pp. 129–144, Jan. 2019, doi: 10.1016/j.giq.2018.11.004.
[48] M. Ghoreishi and A. Happonen, “New promises AI brings into circular economy accelerated product design: a review on supporting literature,” *E3S Web Conf.*, vol. 158, p. 06002, 2020, doi: 10.1051/e3sconf/202015806002.
[49] S. S. Frug and T. R. Bruce, “A Virtuous Circle: Artificial Intelligence and Accessibility for Administrative Applications.,” in *AIAS@ ICAIL*, 2019, pp. 14–17.
[50] “Study on the interaction between big data and artificial intelligence - Li - 2022 - Systems Research and Behavioral Science - Wiley Online Library.” Accessed: Mar. 07, 2024. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/sres.2878

[51] "Collective Intelligence or Collecting Intelligence? | SpringerLink." Accessed: Mar. 07, 2024. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-45982-0_10
[52] J.-P. Joutsenlahti, T. Lehtonen, M. Raatikainen, E. Kettunen, and T. Mikkonen, "Challenges and Governance Solutions for Data Science Services based on Open Data and APIs," in *2021 IEEE/ACM 1st Workshop on AI Engineering - Software Engineering for AI (WAIN)*, May 2021, pp. 1–4. doi: 10.1109/WAIN52551.2021.00012.
[53] Z. Jiang, J. Han, B. Sisman, and X. L. Dong, "CoRI: Collective Relation Integration with Data Augmentation for Open Information Extraction," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, C. Zong, F. Xia, W. Li, and R. Navigli, Eds., Online: Association for Computational Linguistics, Aug. 2021, pp. 4706–4716. doi: 10.18653/v1/2021.acl-long.363.
[54] E. Tan, "Designing an AI compatible open government data ecosystem for public governance," *Information Polity*, no. Preprint, pp. 1–17, 2022.
[55] V. Vassilev, S. Ilieva, I. Krasteva, I. Pavlova, D. Petrova-Antonova, and W. Sowinski-Mydlarz, "AI-Based Hybrid Data Platforms," in *Data Spaces: Design, Deployment and Future Directions*, Springer International Publishing Cham, 2022, pp. 147–170. Accessed: Jan. 29, 2024. [Online]. Available: https://library.oapen.org/bitstream/handle/20.500.12657/58395/1/978-3-030-98636-0.pdf#page=162
[56] Z. Abedjan, "Enabling data-centric AI through data quality management and data literacy," *it - Information Technology*, vol. 64, no. 1–2, pp. 67–70, Apr. 2022, doi: 10.1515/itit-2021-0048.
[57] J. Traub, J.-A. Quiané-Ruiz, Z. Kaoudi, and V. Markl, "Agora: Towards an open ecosystem for democratizing data science & artificial intelligence," *ArXiv, abs*, 1909, Accessed: Jan. 29, 2024. [Online]. Available: https://www.researchgate.net/profile/Jonas-Traub/publication/335689751_Agora_Towards_An_Open_Ecosystem_for_Democratizing_Data_Science_Artificial_Intelligence/links/5e05fcd04585159aa49e7692/Agora-Towards-An-Open-Ecosystem-for-Democratizing-Data-Science-Artificial-Intelligence.pdf
[58] A. H. Trevisan, I. S. Zacharias, C. G. Castro, and J. Mascarenhas, "Circular economy actions in business ecosystems driven by digital technologies," *Procedia CIRP*, vol. 100, pp. 325–330, 2021.
[59] J. Dam and H. Rickon, "Innovation in Artificial Intelligence and the Catalyst of Open Data Sharing: Literature Review and Policy implications," 2023, Accessed: Mar. 01, 2024. [Online]. Available: https://thesiscommons.org/a3zwu/download?format=pdf
[60] "OpenStreetMap," OpenStreetMap. Accessed: Jan. 31, 2024. [Online]. Available: https://www.openstreetmap.org/
[61] "Wikidata." Accessed: Jan. 31, 2024. [Online]. Available: https://www.wikidata.org/wiki/Wikidata:Main_Page
[62] "GitHub: Let's build from here · GitHub." Accessed: Jan. 31, 2024. [Online]. Available: https://github.com/
[63] European Commission, "The official portal for European data | data.europa.eu." Accessed: May 16, 2023. [Online]. Available: https://data.europa.eu/en
[64] "Home - data.gouv.fr." Accessed: Jan. 31, 2024. [Online]. Available: https://www.data.gouv.fr/en/
[65] "Swedish Data Portal," Sveriges Dataportal. Accessed: Jan. 31, 2024. [Online]. Available: https://www.dataportal.se/about-us
[66] D. C. Neale and J. M. Carroll, "Chapter 20 - The Role of Metaphors in User Interface Design," in *Handbook of Human-Computer Interaction (Second Edition)*, M. G. Helander, T. K. Landauer, and P. V. Prabhu, Eds., Amsterdam: North-Holland, 1997, pp. 441–462. doi: 10.1016/B978-044481862-1.50086-8.
[67] D. Herrera-Murillo, J. Nogueras-Iso, P. Abad-Power, and F. Lopez-Pellicer, "User Interaction Mining: Discovering the Gap Between the Conceptual Model of a Geospatial Search Engine

and Its Corresponding User Mental Model," 2023, pp. 3–15. doi: 10.1007/978-3-031-43126-5_1.

[68] L. Oehlberg *et al.*, "Making with Data (and Beyond)," in *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, in CHI EA '23. New York, NY, USA: Association for Computing Machinery, Apr. 2023, pp. 1–5. doi: 10.1145/3544549.3583748.

[69] A. L. A. Library, "LibGuides: Definition of a Library: General Definition." Accessed: Mar. 08, 2024. [Online]. Available: https://libguides.ala.org/library-definition/general

[70] T. D. GLOVER, "The 'Community' Center and the Social Construction of Citizenship," *Leisure Sciences*, vol. 26, no. 1, pp. 63–83, Jan. 2004, doi: 10.1080/01490400490272486.

[71] V. Holm and E. Joseph, "What are Makerspaces, Hackerspaces, and Fab Labs?" Rochester, NY, Nov. 07, 2014. doi: 10.2139/ssrn.2548211.

[72] S. Prendeville, G. Hartung, C. Brass, E. Purvis, and A. Hall, "Circular Makerspaces: the founder's view," *International Journal of Sustainable Engineering*, vol. 10, no. 4–5, pp. 272–288, Sep. 2017, doi: 10.1080/19397038.2017.1317876.

[73] "Reconstructing makerspaces in China: mass innovation space and the transformative creative industries | Humanities and Social Sciences Communications." Accessed: Mar. 08, 2024. [Online]. Available: https://www.nature.com/articles/s41599-022-01383-2

[74] "Journey of Product in Maker Spaces—A Case Study | SpringerLink." Accessed: Mar. 09, 2024. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-16-0119-4_58

[75] "Makerspaces and Design Thinking: Perfect - ProQuest." Accessed: Mar. 09, 2024. [Online]. Available: https://www.proquest.com/docview/1840668236?sourcetype=Magazines

[76] W. L. in R.-B. U. Experience, "Design Thinking 101," Nielsen Norman Group. Accessed: Mar. 09, 2024. [Online]. Available: https://www.nngroup.com/articles/design-thinking/