The Value of an Open Scientific Data and Documentation Platform in a Global Project: The Case of Zenodo

Irene del Rosario Crespo Garrido, María Loureiro García, and Johannes Gutleber

Abstract Open Science is a movement aimed at promoting public access to all scientific research products, without barriers or restrictions. Open Data refers to the practice of sharing research data in a way that assures that the research is accessible, reusable, and reproducible for everyone. Leveraging these two principles, scientists can validate results, and findings, conduct new research, and promote scientific progress. Open data also enables interdisciplinary collaborations and the exploration of research questions beyond the original scope of the data. The most appropriate means used for implementing Open science and open data are digital, collaborative technologies. One notable example of a platform facilitating information dissemination is Zenodo, a free virtual repository based on the CERN developed Invenio software suite. Zenodo serves as an open access and open data platform, offering researchers, scientists, and individuals a centralized, durable, reliable, scalable, free, and accessible space to share, publish, and preserve their research outputs. Zenodo provides various features and benefits that foster knowledge advancement and collaboration within the research community. By promoting open access, Zenodo enables the global dissemination of research findings, eliminating obstacles such as geographic and financial constraints. It is challenging to accurately capture the impact of scientific dissemination, both social and economic. This is particularly the case for a free, "catch-all" repository, which permits any user to supply and access non-reviewed information. This report provides a quantitative estimate of the monetary value that a virtual repository represents based on a multi-component model in which the different parts of the system are quantified using appropriate distinct methods. This study uses the virtual repository Zenodo as a reference case for the ex-ante societal impact analysis for the Future Circular Collider (FCC) at CERN, assuming that in the lifetime of such a new research infrastructure, at least one comparable development will be required due to the collaborative nature of scientific physics

I. R. Crespo Garrido (✉) · M. Loureiro García
Universidade de Santiago de Compostela (USC), Santiago de Compostela, Spain
e-mail: irene.crespo.garrido@cern.ch

I. R. Crespo Garrido · J. Gutleber
The European Organization for Nuclear Research (CERN), Geneva, Switzerland

J. Gutleber and P. Charitos (eds.), *The Economics of Big Science 2.0*, Science Policy Reports, https://doi.org/10.1007/978-3-031-60931-2_14

research with particle accelerators and colliders. Our results indicate a discounted socio-economic impact potential of about 2.8 billion CHF for an observation period of 29 years, from 2028 to 2057.

Keywords Zenodo · CERN · LHC · FCC · FCCIS · OpenAIRE · Open science · Open-access · Economic impact · Social impact · Virtual repository

1 Introduction

Open Science (OS) [1] is a movement that aims to transform the way scientific research is conducted, disseminated, and accessed. It advocates transparency, collaboration, and inclusiveness in the scientific process, to accelerate scientific advances and maximize their societal impact. By removing barriers and fostering accessibility, OS strives to liberate knowledge and cultivate a more inclusive and efficient scientific community.

Typically, access to reliable scientific research results has been limited to those who subscribe to paid journals. This hinders the flow of knowledge and impedes collaboration among researchers. OS aims to change this model by promoting open-access publishing, which makes research articles and the underlying data freely available to all. In this way, scientists, students, and the public can access and benefit from the latest scientific discoveries. Nevertheless, the information provided is free of charge and is funded. For Open Access articles in peer-reviewed journals, this happens typically through institutional funding of the publication process (e.g. SCOAP3 collaboration[1]). In the case of a free virtual repository, it happens through the institutional funding of the system development and operation (e.g. CERN and OpenAIRE[2]).

OS adopts the long-known principles of open-source and free software. Open-source software refers to any freely available, but not necessarily free-of-charge software, allowing researchers to view, modify, and distribute it. In the context of open science, open-source tools help the sharing of scientific work, methodologies, and computational models. Free software is openly available and free of charge. Opening data involves sharing it in structured, documented, and reusable formats, allowing other researchers to validate findings, perform additional analyses, and conduct further research. This practice drives transparency, reproducibility, and

[1] SCOAP3 is a partnership of over three thousand libraries, funding agencies and research centers in 44 countries, regions, or territories and three intergovernmental organizations. It supports OA publishing in a set of journals at no cost for authors. In addition, existing Open Access journals, books, and monographs are centrally supported, removing existing financial barriers for authors and allowing a free and easy scientific discourse in High-Energy Physics. Each country, region or territory contributes in a way commensurate to its scientific output in the field.

[2] OpenAIRE (www.openaire.eu) is a pan-European infrastructure for research.

collaboration while facilitating interdisciplinary research. Eventually, the concept helps accelerate scientific progress.

OS encourages open peer review, which makes the review process transparent and accountable. This makes reviewers' comments and identities openly accessible, encouraging constructive feedback and scientific discourse. Open peer review improves the quality of published research, in addition to developing a sense of community and collaboration among researchers.

Citizen Science [2] is another integral facet of OS, promoting active public participation in scientific research. This term defines the involvement of the public in scientific research projects, engaging so-called "citizen scientists" in various aspects and phases of the research process, such as data collection, analysis, and interpretation. Citizen science projects can cover a wide range of disciplines, such as ecology and the environment, as well as medicine, astronomy, or social sciences. This approach promotes public engagement, scientific literacy, and equal participation in the research process.

Despite OS's wide range of benefits, it also faces major challenges. The traditional publication model and current academic reward systems can stand in the way of the transformation to open science. The non-enforcement of a review process and the absence of community-based quality processes (for instance implemented in Wikipedia[3]) lead to an inflation of products and an issue of referencing credible, reliable, and accurate information. Concerns around intellectual property, privacy, and data security need to be raised to ensure a judicious sharing of research results. In addition, achieving cultural and systemic changes in the scientific community requires collective efforts and ongoing advocacy.

Virtual repositories are a suitable channel to capture and make information available openly and freely. These repositories, also known as digital repositories or online repositories, are digital platforms and systems to store, manage, and provide access means (organize information, search information, link information) to a wide range of digital content, such as documents, files, datasets, images, and multimedia. They serve as a centralized location for preserving digital resources in durable and reliable ways, making them easily findable, referenceable, and accessible to users. Virtual repositories are designed to facilitate the storage, retrieval, and sharing of digital content, providing an efficient solution for managing large volumes of information. They usually include search functions, metadata management, version control, and access control mechanisms to ensure the organization and security of stored content.

The purpose of the study presented in this report is to analyze and quantitatively estimate the socio-economic potential of a free virtual repository in monetary terms. The work is motivated by the fact that a Future Circular Collider (FCC) [3–5] is assumed to require during its lifetime at least one comparable information management infrastructure to satisfy the needs to a worldwide collaborative scientific research activity as was the case with CERN and the LHC program so far. The solution may be a new type of information management platform, but it may also be the continued development and use of the virtual repositories such as the "CERN

[3] Quality process of Wikipedia: https://en.wikipedia.org/wiki/Wikipedia:Quality_control.

Document Server" and "Zenodo" built on the Invenio[4] software that has been developed at CERN for this purpose. This study takes the Zenodo virtual repository as a case for the investigation that provides ex-post socio-economic impact data as input for the ex-ante socio-economic impact study for the FCC.

The report first gives an overview of the Zenodo open and free virtual repository. Then it presents the approach to elucidate and quantify the socio-economic impact potentials that can be associated with this platform with an outlook on the FCC program period. To conclude, the results of a survey conducted to estimate the common good value of the repository via a willingness to pay (WTP) approach are presented.

2 Zenodo: An Open Virtual Repository

Zenodo [6] is a free virtual repository and open-access data platform that allows, researchers, scientists, and individuals to share, publish, preserve, find, reference, and access information. The underlying software and the system were developed by CERN. The virtual repository is managed by the OpenAIRE (Open Access Infrastructure for Research in Europe) project. The name Zenodo derives from Zenodotus, who is said to have been the first librarian of the Library of Alexandria.

The Zenodo source code is openly accessible. It is based on the Invenio digital library, which is also an open-source project managed by CERN. Work in progress is openly shared on GitHub; anyone can contribute to any aspect, but the source is controlled by CERN. Metadata is openly available under a CC0 license, and all content is accessible through open APIs. This engaging process promotes that any individual or institution can have access to the platform to either use it as is or tailor it to its needs under the condition that the original name, Zenodo is not used. The Zenodo deployment at CERN is the EU-recommended repository for all EU co-funded research project results. Anyone in the world can freely deposit information on this platform that is hosted on CERN's computing infrastructure. Zenodo is integrated into the European Open Science Cloud (EOSC).

Zenodo offers researchers a user-friendly, reliable, and scalable platform for sharing and preserving virtually unlimited amounts of research outputs. Its commitment to open access, long-term preservation, version control, and integration with other platforms makes it a valuable tool in the research community. The following are some of the key features of Zenodo:

- **Open access**: Zenodo follows the principle of open access, meaning that the research outputs shared on the platform are freely accessible to anyone without paywalls or subscription requirements. This promotes the dissemination and accessibility of research worldwide.

[4] Invenio is a free and open-source software developed by CERN for building digital repositories and information management systems. https://invenio-software.org.

- **Wide range of research outputs**: It supports various types of research outputs, such as datasets, software, papers, posters, presentations, and multimedia files. This flexibility makes it suitable for different disciplines and research fields.
- **Open access and preservation**: The virtual free repository follows open access principles and makes research results publicly available. It assigns a Digital Object Identifier (DOI) to each uploaded item, ensuring persistent and citable links across versions. Zenodo also guarantees long-term preservation, ensuring that shared research results remain accessible and discoverable for the foreseeable future.
- **Integration with other platforms**: Zenodo integrates seamlessly with other research infrastructures and platforms, such as ORCID (Open Researcher and Contributor ID), GitHub, and EOSC. This allows researchers to link their Zenodo profiles to their ORCID IDs and connect their code repositories directly to Zenodo for versioning and archiving.
- **Version control and DOI**: The platform supports version control, allowing researchers to upload multiple versions of their research results. Each version receives a unique DOI, ensuring proper citation and referencing. This feature allows researchers to update and improve their work while maintaining previous versions.
- **Licensing**: The platform facilitates collaboration by allowing multiple collaborators to be associated with shared research output. In addition, it offers several licensing options, including open licenses such as Creative Commons, which allow researchers to define the conditions under which their work can be reused or shared.
- **Discovery and citation**: The virtual free repository ensures that research results are findable through its search interface and its integration with other indexing services. It promotes the use of standardized metadata and encourages proper citation of shared research results, contributing to the scholarly record and recognition of contributors.
- **Statistics and metrics**: Zenodo provides usage statistics and metrics for uploaded articles, allowing researchers to track the impact and visibility of their work.
- **Community support**: The platform has an active community of users and developers who provide support, guidance, and feedback. The platform regularly incorporates suggestions and updates from users to improve its functionality.

It is impossible to talk about Zenodo without discussing the FAIR principles. These principles emphasize data management and sharing, aimed at promoting the findability, accessibility, interoperability, and reuse of research data.

2.1 FAIR Principles

The FAIR principles were initially proposed by [5] to address the challenges associated with data sharing and reuse in the scientific community. They are a set of guiding principles that promote the Findability, Accessibility, Interoperability, and

Reusability of research data. They were developed to address the challenges associated with data management and sharing in the scientific community. The concept emerged as a response to the growing volume of scientific data and the need to unlock its potential for advancing research and innovation. These principles aim to ensure that data is effectively managed, shared, and utilized by researchers, institutions, and organizations. Each of the principles will be explored in more detail below:

- **Findable**: Research data and resources should be easy to find, both for humans and machines. To achieve findability, data should be assigned persistent identifiers (such as DOIs), and metadata should be provided to describe the data and its context. The metadata should be sufficiently rich and standardized, enabling effective data search, retrieval, and linking.
- **Accessible**: Research data and resources should be openly accessible to both humans and machines. This principle emphasizes the removal of barriers to access, enabling unrestricted access to the data without requiring unnecessary permissions or restrictions. Open access facilitates broader use, analysis, and validation of research outputs.
- **Interoperable**: Research data and resources should be structured and represented in a way that enables their integration and interoperability. Interoperability allows data from different sources to be combined and reused effectively. It involves the use of standardized data formats, vocabularies, and ontologies that promote compatibility and facilitate data integration and exchange.
- **Reusable**: Research data and resources should be well-described and provide sufficient context and documentation to facilitate their reuse. This involves providing clear and rich metadata, including information about the data's provenance, methods of collection, and conditions of use. Licensing and permissions should be clearly defined to enable others to understand and comply with the terms of reuse.

Adhering to the FAIR principles means that research data and resources become more valuable and usable for both researchers and the broader community. FAIR principles support open science practices, enhance collaboration, enable data-driven discovery, and promote reproducibility and transparency of research. They are fundamental to maximizing the impact and potential of research results and fostering a culture of openness, sharing, and innovation in the scientific community.

3 Estimated Value of Zenodo's Socio-economic Impact Potentials

This section sheds light on the quantitative estimation of the socio-economic impact potentials of the Zenodo repository as an example of an OS platform.

To our best knowledge, no studies exist so far that monetize the value of the impact potentials of this type of infrastructure. A comparable analysis has been

carried out in 2021 on the Benefit/Cost of the entire OpenAIRE infrastructure [5]. Several studies analyze the impact of open-source software such as [2, 6–8] among others, but none provides a ready-to-use prescription or guideline for the elucidation of the quantitative socio-economic value of an open and free data platform. Due to a lack of an existing method to capture the quantitative value, we had to devise a viable set of complementary methodologies to estimate the values of the individual segments of the platform.

This study aims to answer the question: What is the estimated value of the socio-economic impact potentials produced by open scientific data and documentation platforms, taking the Zenodo case study as a reference? We provide one possible lower limit by formulating an economic model based on the expected net present value model.

3.1 Methodology

Knowing what a virtual repository is and the benefits potentials it comes with at the scientific and social levels, we conceived an econometric model to provide an estimate of its socio-economic impact value, based on the sum of estimated socio-economic impact values of individual segments of the platform for which we use models. Eventually, this estimation is compared with the revealed common good value that has been approximated with a Willingness to Pay survey.

The monetary equivalent value has been estimated through the expected Net Present Value (NPV) adopting an observation period 2028–2057, i.e., ex-ante given continued use and further development in the frame of the FCC program, based on the past, known evolution of the platform. This financial metric is used to calculate the present value of the future investment. Both, the benefits, and costs of an investment are estimated, and a social discount rate (SDR). It is used as a measure of the avoided cost and therefore as an investment decision criterion to update the collections and payments and to know how much will be gained or lost with such an investment. Reference [9] sheds light on the use of this financial metric in socio-impact analysis, demonstrating its efficiency and optimal results. A project holds social value when its benefits consistently outweigh costs over time, indicated by a positive Net Present Value (NPV). When considering benefits (B_{ti}) and costs (C_{ti}) occurring at different instances represented by time *ti*. Below is the formula (1) used for the final calculation of the socio-economic impact study, where SDR represents the social discount rate.

$$NPV = \sum_{i}^{t} \frac{B_{ti} - C_{ti}}{(1 + SDR)^{t}} \tag{1}$$

One first issue that arises in setting up the econometric model is the chronology to which we are exposed. Setting up an economic model for such a long period

means that we are faced with the possibility of very large and unforeseeable changes in future years. For this reason, several assumptions have been implemented in the model to be able to address future changes in the economy and to be transparent about the presented results and how they have been devised. The assumptions considered are:

- The reference period considered for the study, 2028–2057, has been established by an expert panel of economists in the EU project Horizon 2020 research and innovation action "Future Circular Collider Innovation Study" (FCCIS), in which this work is carried out, to homogenize the criteria among the other studies developed in the same project.
- The base year is 2021 and the discount rate is 2%, established by [10].
- The exchange rate EUR/CHF used in the analysis is 1 € = 1.07 CHF, representing an average exchange of the year 2021.
- It is assumed that data archived in Zenodo will grow until 2040 and then remain constant. This development assumes that twenty years from now, virtual repositories will be fully integrated into research domains, which means that their user base will be mature and will therefore only grow marginally. Our estimates aim to be conservative lower-bound limits for socio-economic value and therefore we assume a constant use base rather than a marginally growing one.
- For the computation of the benefit value from the online use of the repository (explained in detail below), the value of the time visitors spend on the platform must be estimated. For this study, the time value is set to 0.30 €/min. This value has been extracted from the Eurostat database based on the average salaries per minute of researchers in EU countries in 2018 [11]. The average salaries per minute have been updated to 2021 using the GDP deflator following the methodology applied by [12].
- For estimating the economic benefit of downloads made in Zenodo it is necessary to establish a monetary value per download. This monetary value has been set at 7 € per download. The choice of this value will be explained in detail below.

3.2 Description of Relevant Variables

The resulting expected net present value between the years 2028 and 2057, taking 2021 as the base year, was obtained by subtracting the discounted estimated costs of the infrastructure from the sum of the discounted estimated benefits of the free virtual repository. The variables chosen to formulate the model, as well as the monetary values, were:

3.2.1 Development and Operation Costs

The virtual repository's annual development, maintenance, and operation costs have been known since 2012. The value is rather modest since Zenodo is a "by-product"

of developments carried out at CERN to manage the knowledge produced in a global high-energy and particle physics community of almost 20,000 people over several decades. Physical assets required for the development and operation such as computing and data management infrastructures, networks, offices, end-use devices, and even travel represent a marginal cost in a common effort of a community of nations. In other words, the cost base, platform, and deployment can de facto be considered a gift to society.

The cost level for 2022 has been maintained in the analysis. Consideration should be given to the fact that the creation of a virtual repository requires mainly an initial budget to bootstrap the development, but as the architecture is settled, a working code base is available, and the repository becomes operational the expenditure starts to decrease and only the salaries of the software engineers and technicians has been noticeable in the accounting. Given this assumption and observing the long timeline applied to the study, an annual incremental cost of 4% has been applied from 2023 to 2057 to account for cost increases additional features, and code consolidation. The future budget estimated with formula (2) corresponds to three full-time equivalent individuals working on the development of the virtual repository every year.

$$Total\ Cost = \sum_{i}^{t} ((Cost_{t-1} \times 4\%) + Cost_{t-1}) \tag{2}$$

3.2.2 Data Storage Benefit

One of the noteworthy impacts of the platform is the assurance of long-term preservation and accessibility of information. This aspect can be captured by estimating the value of the persistent data store segment of the virtual repository. As the repository is free, estimating the monetary term is challenging, but not impossible. Currently, several information platforms apply prices for the preservation of scientific documents and data. Repositories such as Arxiv [13] and IEEE Xplore [14] store text documents. Zenodo also stores datasets, software, video, images, and files in HTML or archives for instance in ZIP format, differentiating itself from other repositories. For that reason, we chose the international repository Dryad [15], which provides functionalities approaching Zenodo as a comparable reference to establish the monetary value of the impact produced by this variable. Dryad is an international open-access repository of research data, especially data underlying scientific and medical publications (mainly evolutionary biology, genetics, and ecology). Dryad is a general-purpose curated repository that makes data discoverable, freely reusable, and citable. Dryad's scientific, educational, and charitable mission is to provide the infrastructure necessary to promote the reuse of scholarly research data.

Dryad publishes a price list [16] for the storage of these files that were used as a reference to establish a monetary value for the case study. The prices established by Dryad for uploading files to its repository are as follows:

- For less or equal to 50 GB the price is 111.72 € per document.
- For more than 50 GB it charges 46.55 € for each additional 10 GB.

A conservative approach has been adopted, considering previous years of available data, to implement an annual growth rate of 10% for data storage under or equal to 50 GB and 3% for data storage over 50 GB. Using the formula (3) the total estimated benefit was obtained.

$$\begin{aligned} Data\,storage\,benefit = \sum_{i}^{t} (((Data \leq 50\,Gb_{t-1} \times 10\%) + Data \leq 50\,Gb_{t-1}) \\ * 111.72\,€) \\ + \sum_{i}^{t} (((Data > 50\,Gb_{t-1} \times 3\%) + Data > 50\,Gb_{t-1}) \\ * 46.55\,€) \end{aligned} \tag{3}$$

3.2.3 Online Use Benefit

The purpose of this variable is to measure, through the time users spend on the web platform, the impact potential of the use of an open-access repository by users via a web browser.

The actual past and current time of web usage are captured using Piwik's web interaction activity recording. This free and open-source web analytics tool used by Zenodo and other platforms for its metrics, measures the time users spend during each visit to the free virtual repository. Using this method, we were able to establish an average interaction time per unique visit of 4 min. The annual increment of individual users for the analysis period is 10% until 2040. The percentage increase is based on a moderate growth of the deposited information observed in recent years. We set the monetary value estimate of time for interacting with the web interface to 0.30 € per minute, based on the average salaries per minute of researchers in EU countries [11].

$$\begin{aligned} Online\,usage\,benefit = \sum_{i}^{t} \left(\left(unique\,visitors_{t-1} \times 10\%\right) + unique\,visitor_{t-1}\right) \\ \times 0.30\,€ \times 4\ \text{min}) \end{aligned} \tag{4}$$

3.2.4 Download Benefit

The monetary value of downloads is another variable that we considered as a measure of the socio-economic impact potential. It allows measuring the impact of using information stored in the virtual repository.

Table 1 Common average prizes to access information in research repositories and paid scientific journals

Repository or scientific journal	Average item price (€)
Science Direct	31
IEEE Xplore subscription fee	6.22
Nature subscription fee	7.7
Springer	29.95

We assigned a monetary value of 7 € for an individual download, irrespective of its type, contents, and size. We chose this value after comparing the prizes to access information in several repositories and paid scientific journals (Table 1).

Our original estimations pointed to an average value of 30 € per download, however, Zenodo does not feature content and format quality assurance and peer reviewing, which is the main cost driver of paid information resources. Two relevant paid platforms, the non-profit IEEE Xplore digital library, and the commercial Nature platform, see Table 1, both offer the possibility of an individual subscription, which facilitates the estimation of the value per user and download. Estimating the costs of quality management with experienced copy editors and reviewers of scientific publications and deducing it from the individual subscription costs brings the price estimate to 7 € on average, considering the IEEE Xplore and Nature prices as a reference.

The second assumption for the estimation of the value of this variable is that we do not have an unambiguous figure for individual downloads, but only for all downloads of an information set. A single document or data set may be downloaded by the same user multiple times. If it is downloaded only a single time it may also be consumed multiple times by the person having downloaded the information or further shared the information. Zenodo cannot distinguish individual downloads. Consequently, there exists no reliable method to estimate the use of the information by counting downloads only. Therefore, we decided to consider a simplified approach, counting only a single download as consumption per individual record stored in the repository over the entire lifetime of that information record, taking as a reference the number of documents evaluated in the data storage variable.

4 Results

The results of the study have shown that open-access virtual repositories, considering Zenodo as a case study, promote social benefits that when transformed into monetary values, bearing in mind the assumptions mentioned above, yield the results (Table 2).

The combined discounted socio-economic impact of all model variables results in an estimated monetary value of about 2.8 billion CHF from the period spanning 2028 to 2057, considering a SDR of 2% [10]. This translates into an annual average value of approximately 97 million CHF discounted.

Table 2 Discounted monetary values of model variables from the period 2028 to 2057

Model variables	Discounted values (CHF)
Development and operation costs	10,917,550
Data storage benefit	1,309,037,916
Online use benefit	1,426,391,881
Download benefit	81,998,069

5 Assessing the Use and Perception of Zenodo in the Scientific Community

In addition to the study described above, we designed and surveyed with the purpose of estimating the use and perception of the Zenodo virtual repository in the scientific community. Elucidating the common good value of a virtual repository by using the Willingness to Pay (WTP) approach helped us to verify and fine-tune our monetary value estimates and assumptions.

We administered the survey to a random sample of the international scientific Future Circular Collider (FCC) study collaboration, which federates more than 150 universities and research centers worldwide, active in many diverse disciplines of science and engineering. Respondents were asked about their use and knowledge of the Zenodo free virtual repository, as well as their willingness to financially contribute, with a view to establishing a monetary value for the data storage, the online use, and the download of information. WTP in economics is a metric that refers to a consumer's willingness and ability to pay a certain price for a product or service [17]. It can vary depending on factors such as individual preferences, perceived value, affordability, and market conditions. This metric was previously used by [18] to contingently assess citizens' attitudes towards CERN. The survey indicated clearly that the question on willingness to pay (WTP) would be used to assess the added value of an open data platform for the scientific community and that no intent exists to make the platform a paid service. Despite this fact, as will be shown later, a significant number of respondents misinterpreted the survey intent as an attempt to test for introducing a service fee. This required us to identify such respondents and exclude their responses from the evaluation. However, the strong reactions are also a sign for the high value of Open Science and free access to scientific research results, thus supporting at the same time the socio-economic value of such platforms.

Finally, we obtained a sample of 182 valid responses, which include the ones who reported a zero WTP. The larger part of the respondents were men, aged between 30 and 65 years with a doctoral degree and working in physics research.

It is noted that the survey was aimed primarily at the scientific community and therefore biased by this community. After all, the Invenio software emerged from this discipline. This means that administering such a survey to a more heterogeneous set of actual or potential users could lead to different results when including people from substantially different disciplines such as life sciences, social sciences, economics,

law, and business administration. It is therefore not surprising that the respondents have an overwhelming number of Ph.D.'s (81%) or other university degrees (Fig. 1). 157 of the 182 respondents work or have worked in research centers, as is the case of one of the retired respondents. Of those who report working in scientific research, 31% work at CERN. A suspicion of a CERN bias due to the amount of respondents working at CERN could, however, be ruled out after analysis of the data.

It was surprising that 53% of the participants in the study did not know about Zenodo when they were contacted, despite the large set of persons working at CERN and being involved in the FCC study. From our experience of using the WTP method in previous investigations, the fact of not knowing the subject for which a person is asked to financially participate explains an overall low willingness to financially participate. People need to know a service or system, understand its functions and services and need to have experienced at least once its potential benefits to be able to attribute a monetary value reliably to it (Fig. 2).

Out of the 51% of respondents who do not know Zenodo, only 18% consider using it in the future. On the other hand, of the 86 respondents who do know Zenodo, 59% use the repository routinely for professional reasons. All the respondents who do know Zenodo are in favor of sharing their data for free because they believe that knowledge should be accessible to everyone.

Figures 3 and 4 show the respondents' WTP levels. 48% of respondents state a zero WTP to have unlimited data storage on Zenodo for scientific information. 57% respond with a zero WTP for unlimited access to scientific information. When asked for their main reason, 97% of respondents say that "*access and use of the virtual repository should be free of charge and that it goes against the principles of the OpenAIRE movement*". This shows that the respondents misunderstood the survey, thinking that they were asked how much they would pay in case a service feed would be introduced for the use of Zenodo, rather than how much is it worth to them. This

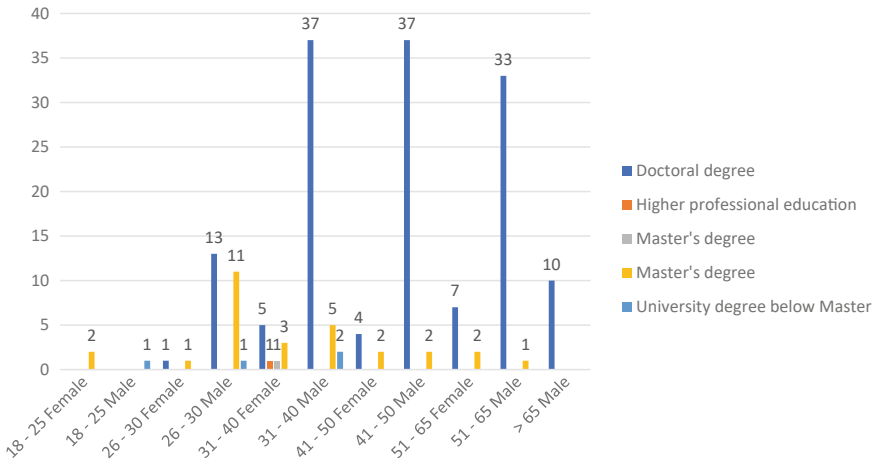

Fig. 1 The highest level of education of the respondents in the function of gender and age range

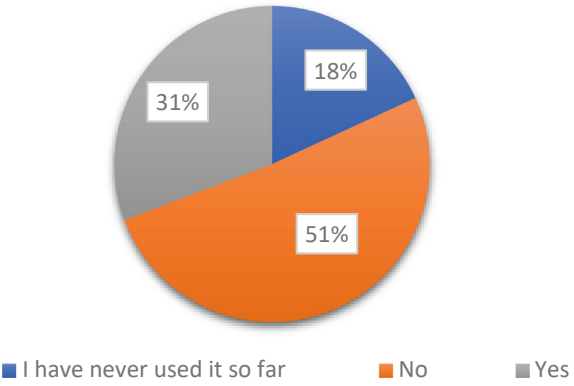

Fig. 2 Previous use of the virtual repository based on the question "Have you used Zenodo before?"

exhibits a limitation of the WTP survey approach. On the other hand, the results demonstrate that people generally favor scientific research being shared openly and freely, as mentioned above.

For the WTP analysis of the survey, we considered only the valid responses with WTP values above 0 € and respondents who indicated 0 € for reasons that make clear that they have understood the purpose of the survey (i.e., the goal to elucidate the value for the respondent). This approach is consistent with the recommendations found in the literature [19] not to exclude zero value responses, on condition that the arguments for answering with zero demonstrate the correct interpretation of the question. All other respondents have been removed from the analysis.

Next, we established an OLS regression model to obtain overall respondents' willingness to pay and to elucidate, which variables directly influence it, based on testing different dependent or endogenous variables. The independent or explanatory variables considered in the model are as shown in Table 3.

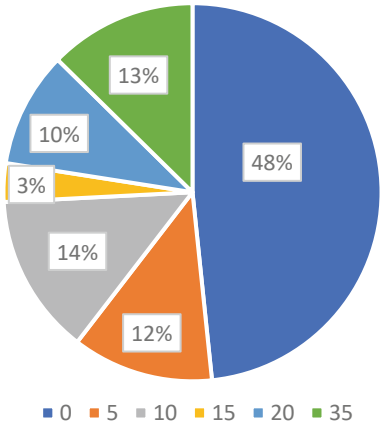

Fig. 3 Willingness to pay levels in euros per year for unlimited data storage in Zenodo for scientific information

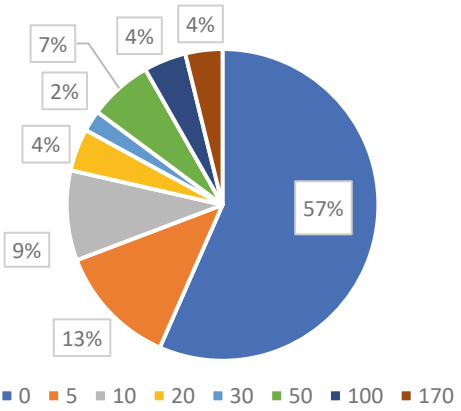

Fig. 4 Willingness to pay levels in euros per year for unlimited access to Zenodo

Table 3 Independent or explanatory variables of the regression model formulated

The independent or explanatory variables	Variable description	Mean	Standard deviation
Respondents' salary	This value is taken from the Glassdoor[5] database. The average salary, in euros, considered is that of a researcher in the country of residence the survey provides	53,313.1526	20,899.5773
Gender	Dummy variable: 1 = male, 0 = female	0.8517	0.3716
Age	Age of the respondents	44.12	12.97
Education background	Dummy variable: 1 = natural sciences, mathematics, and statistics, 0 = other	0.8297	0.3769
Level of education	Dummy variable: 1 = doctorate level, 0 = other	0.8077	0.3952
Type of occupation	Dummy variable: 1 = employed and retired, 0 = working at CERN	0.7253	0.4476
Knowledge = respondent knows the repository	Dummy variable: 1 = yes, 0 = no	0.4725	0.5006
Use = respondent uses the repository	Dummy variable: 1 = yes, 0 = no and never used	0.3077	0.4628

[5] For more information: https://www.glassdoor.com/.

Using the variables outlined in Table 3, the results shown in Tables 4, 5, and 6 were obtained.

The results of the three models shown in Tables 4, 5, and 6 indicate that only the variable "previous knowledge of the platform" has the effect of influencing the users' decision on their WTP. The Variance Inflation Factor (VIF) [20] of the models is in the range of 1.03 and 2.12, which shows that the variables are moderately correlated.

The outcomes related to this important variable reveal that with each additional person informed about the Zenodo virtual repository, there is an observed willingness to pay an average of 14.47 € for downloading a document, 7.56 € for accessing unlimited storage, and approximately 20 € for utilizing a combination of features within the Zenodo virtual repository.

Using the derived estimators for each variable, we conduct the calculation to determine the respondents' WTP for each formulated model. The ensuing outcomes

Table 4 Parameters used for the calculation of WTP for downloads

Independent or explanatory variables	Estimate	Std. error	t value	Pr(> |t|)
Constant	37.9119	14.9497	2.536	0.0121*
Respondents' salary	− 0.0002	0.0001	− 1.616	0.1080
Gender	1.08336	8.3251	0.130	0.8966
Age	0.1776	0.2398	0.740	0.4600
Education background	− 12.6918	8.3757	− 1.515	0.1315
Level of education	− 10.5984	7.9794	− 1.328	0.1859
Type of occupation	− 3.8330	6.4212	− 0.597	0.5513
Knowledge	14.4725	8.1126	1.784	0.0762
Use	− 7.7266	8.9129	− 0.867	0.3872

Signif. codes: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$
Num. of observations: 182

Table 5 Parameters used for the calculation of WTP for data storage

Independent or explanatory variables	Estimate	Std. error	t value	Pr(> |t|)
Constant	5.8896	4.5539	1.293	0.19763
Respondents' salary	− 0.0000	0.0000	− 0.549	0.58358
Gender	2.4828	2.5359	0.979	0.32892
Age	0.0569	0.0731	0.779	0.43714
Education background	− 2.7106	2.5514	− 1.062	0.28954
Level of education	0.1043	2.4306	0.043	0.96581
Type of occupation	− 1.6363	1.9560	− 0.837	0.40399
Knowledge	7.5617	2.4712	3.060	0.00257**
Use	− 2.0401	2.7150	− 0.751	0.45343

Signif. codes: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$
Num. of observations: 182

Table 6 Parameters used for the calculation of combined WTP

Independent or explanatory variables	Estimate	Std. error	t value	Pr(> |t|)
Constant	43.8015	17.9429	2.441	0.0156*
Respondents' salary	− 0.0002	0.0001	− 1.486	0.1392
Gender	3.5662	9.9919	0.357	0.7216
Age	0.2344	0.2879	0.815	0.4164
Education background	− 15.4024	10.0527	− 1.532	0.1273
Level of education	− 10.4940	9.5771	− 1.096	0.2747
Type of occupation	− 5.4694	7.7069	− 0.710	0.4789
Knowledge	22.0342	9.7369	2.263	0.0249*
Use	− 9.7668	10.6975	− 0.913	0.3625

Signif. codes: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$
Num. of observations: 182

Table 7 Results of the WTP models

Willingness to pay for a function	Model output monetary value (€)
Willingness to pay for a single download	17.34
Willingness to pay for a single data storage	8.87
Combined willingness to pay for a publication on the platform	26.21

are outlined, showing respondents' WTP for each segment defined in the survey (Table 7).

6 Conclusions

Open Science is a paradigm shift in the way scientific research is conducted and shared. It promotes transparency, collaboration, and accessibility by making scientific knowledge and data freely available to the public. The dissemination of information that promotes open science can be carried out through many means of dissemination, but virtual repositories are key enablers of open science.

In this study we presented an estimation of the monetary value that an open and free virtual repository represents for the society, taking the Zenodo platform as an example. We conceived a methodology that applies to any open platform for scientific data and documentation. It consists of decomposing the task of quantifying the socio-economic impact potential into individual segments of the repository and establishing models based on past and factual observations and reliable references. The assumptions are verified with a common good value analysis based on the WTP approach.

The total discounted estimated monetary value of the socio-economic impact potentials of a virtual repository for the period 2028–2057 amounts to about 2.8 billion CHF. This value must be compared to the total discounted estimated cost of developing, maintaining, and operating the platform, represented by a marginal value of about 11 million CHF, for the same period. It must be considered that this initial socio-economic value analysis is a first and lower-bound estimate and that the true value potential is most likely significantly larger as annual accounting data of the IEEE Xplore shows.

The results demonstrate therefore convincingly that large-scale, fundamental science projects contribute to society rather than take from the society. The value added generated by useful by products of fundamental scientific research and management of international science collaborations dwarfs the investment costs.

The insights garnered from this study furnish us with the basis to articulate specific recommendations aimed at amplifying the influence of advancements like Zenodo:

- **Enhance investment in development**: It is recommended to bolster investment in the advancement of free and open virtual repositories. This will serve to enhance functionalities and overall quality, thereby rendering the platform more appealing to a broader spectrum of users.
- **Incorporate quality management features**: Introducing quality management attributes is advised. These could include provisions for both anonymous and identified comments on deposited content, potentially incorporating a voting mechanism and facilitating moderated comments.
- **Boost awareness and adoption**: Investment in augmenting awareness and usage of the platform is crucial. Our study underscores that the perceived value among individuals is intrinsically tied to their familiarity and utilization of the platform.

We propose that the financial burden for these initiatives should be borne by stakeholders with an interest in fostering such infrastructures. This could involve international and national science funding agencies, as opposed to the organizations responsible for platform development, operation, and utilization. This objective can be accomplished by offering targeted funding to key communities for leveraging Zenodo's capabilities.

Acknowledgements This project has received funding from the European Union's Horizon 2020 research and innovation program under the European Union's Horizon 2020 research and innovation program under grant agreement No 951754.

References

1. Burgelman J, Pascu C, Szkuta K, Von Schomberg R, Karalopoulos A, Repanas K, Schouppe M (2019) Open science, open data, and open scholarship: European policies to make science fit for the twenty-first century. Front Big Data 2. https://doi.org/10.3389/fdata.2019.00043

2. Bonney R, Cooper CB, Dickinson JL, Kelling S, Phillips T, Rosenberg KV, Shirk J (2009) Citizen science: a developing tool for expanding science knowledge and scientific literacy. BioScience 59(11):977–984. https://doi.org/10.1525/bio.2009.59.11.9
3. Abada A, Abbrescia M, AbdusSalam SS et al (2019) FCC physics opportunities. Eur Phys J C 79:474. https://doi.org/10.1140/epjc/s10052-019-6904-3
4. Abada A, Abbrescia M, AbdusSalam SS et al (2019) FCC-hh: the hadron collider. Eur Phys J Spec Top 228:755–1107. https://doi.org/10.1140/epjst/e2019-900087-0
5. Abada A, Abbrescia M, AbdusSalam SS et al (2019) FCC-ee: the lepton collider. Eur Phys J Spec Top 228:261–623. https://doi.org/10.1140/epjst/e2019-900045-4
6. Zenodo—Research (n.d.) Shared. https://zenodo.org/
7. Koundouri P, Chatzistamoulou N, Dávila O, Giannouli A, Kourogenis N, Xepapadeas A, Xepapadeas P (2021) Open access in scientific information: sustainability model and business plan for the infrastructure and organization of OpenAIRE. J Benefit-Cost Anal 12(1):170–198. https://doi.org/10.1017/bca.2020.26
8. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J, Da Silva Santos LOB, Bourne PE, Bouwman J, Brookes AJ, Clark TW, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R et al (2016) The FAIR guiding principles for scientific data management and stewardship. Sci Data 3(1). https://doi.org/10.1038/sdata.2016.18
9. Carrazza S, Castelnovo P, Catalano J, Florio M (2019) Assessing the value of CERN's free and open-source software: the case of ROOT. Study of the socio-economic impact of CERN HL-LHC and FCC-HH. CERN technical report
10. Blind K, Böhn M, Gzegorzewska P, Katz A, Muto S, Pätsch D, Schubert T (2021) The impact of open-source software and hardware on technological independence, competitiveness, and innovation in the EU economy. Final study report. European Commission. https://digital-strategy.ec.europa.eu/en/library/study-about-impact-open-source-software-and-hardware-technological-independence-competitiveness-and
11. Robbins R, Korkmaz G, Santiago Calderón JB, Chen D, Kelling C, Shipp S, Keller S (2018) Open-source software as intangible capital: measuring the cost and impact of free digital tools. International Association for Research on Income and Wealth. https://www.semanticscholar.org/paper/Open-Source-Software-as-Intangible-Capital%3A-the-and-Robbins-Korkmaz/eab74edcd67ee5659db40de02cc52c5afa77c505
12. Gosh K, Ramakrishnan T, Chawla S (2012) An investigation of the impact of open source software support on its market share. In: AMCIS 2012 proceedings, vol 2. https://aisel.aisnet.org/amcis2012/proceedings/StrategicUseIT/2/
13. Florio M (2019) The (expected net present) value of investing in discovery. In: Investing in science: social cost-benefit analysis of research infrastructures. The MIT Press, pp 273–298. https://direct.mit.edu/books/book/4583/chapter/204569/The-Expected-Net-Present-Value-of-Investing-in
14. Catalano G, Pancotti C (2022) Estimations of SDR in selected countries. Zenodo. https://doi.org/10.5281/zenodo.6675063
15. Eurostat—Database (n.d.) Statistical classification of economic activities in the European Community (NACE Rev. 2). https://ec.europa.eu/eurostat/data/database
16. Morretta V, Vurchio D, Carrazza S (2022) The socio-economic value of scientific publications: the case of earth observation satellites. Technol Forecast Soc Change 180. https://doi.org/10.1016/j.techfore.2022.121730
17. arXiv (n.d.) Cornell University. https://arxiv.org/
18. IEEE Xplore (n.d.) Research. https://ieeexplore.ieee.org/Xplore/home.jsp
19. Dryad (n.d.) Research. https://datadryad.org/stash
20. Dryad (n.d.) Research. Frequently asked questions. Overage fees. https://datadryad.org/stash/faq
21. Catalano G, Florio M, Giffoni F (2016) Willingness to pay for basic research: a contingent valuation experiment on the large hadron collider. RePEc Res Pap Econ. https://EconPapers.repec.org/RePEc:mil:wpdepa:2016-03

22. Giffoni F, Florio M (2023) Public support of science: a contingent valuation study of citizens' attitudes about CERN with and without information about implicit taxes. Res Policy 52(1):104627. https://doi.org/10.1016/j.respol.2022.104627
23. Arrow K, Solow R, Portney PR, Leamer EE, Radner R, Schuman H (1993) Report of NOAA panel on contingent valuation. Federal Register. https://ci.nii.ac.jp/naid/10003241251
24. VIF (n.d.) Investopedia. https://www.investopedia.com/terms/v/variance-inflation-factor.asp#citation-1