



**FUTURE
CIRCULAR
COLLIDER**
Innovation Study



The value estimation of an open scientific data and documentation platform in a global project

Irene del Rosario Crespo Garrido, CERN



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.



This work has been carried out in the frame of CERN's doctoral student program. Supervisors: Dr. Johannes Gutleber (CERN) and Prof. María Loureiro García (USC)

Introduction

- ❖ **Open Science** was defined by Saez and Martinez-Fuentes (2018) as “transparent and accessible knowledge that is shared and developed through collaborative networks”.
- ❖ **CERN's most important intangible asset is its research and technological developments.** Since its inception, the organization has become a global reference for the "Open Science" movement.
- ❖ **The most common way** to disseminate knowledge is through the use of digital technology. An example of this technology are virtual repositories, such as **Zenodo**.



Introduction

- ❖ **The benefits produced by these repositories** range from increasing the **accessibility and visibility of scientific production** both in general and individually (e.g. the author), **collecting and preserving all scientific production**, to being able **to give back to society what has been invested in the form of research results**, since the funding of this research comes from each individual.
- ❖ These benefits are qualitative in their entirety, but **what if they could be quantified?**
- ❖ This study has **converted these qualitative benefits** into a quantifiable estimation of the **Socio-Economic Impact of a virtual repository**, taking **Zenodo as an example**.



Overview

❖ Case study of the Zenodo, an open and free virtual repository.

- Economic model
 - Assumptions
 - Results
- Common good value results



Case study of Zenodo



The Future Circular Collider Innovation Study (FCCIS) receives funding from the European Union's Horizon 2020 research and innovation programme under grant No 951754. The information herein only reflects the views of its authors and the European Commission is not responsible for any use that may be made of the information.

Economic Model

- ❖ Economic model based in expected net present value the Socio-Economic impact of a free virtual repository.
- ❖ **Baseline period of the study 2028 – 2057** will be defined by:

$$NPV = \sum_{t=0}^T \text{Discounted Benefit } i - \sum_{t=0}^T \text{Discounted Cost } i$$

- ❖ **Cost of the repository:** Costs related to the development and maintenance of the repository.
- ❖ **Benefits of the repository:** Three variables have been identified as having a socio-economic impact benefit measuring the benefits of open access.
 - **Data storage:** Data storage includes all files stored in the repository. The quantification of the impact is determined by establishing a “virtual” price based on market values from “Dryad”.
 - **Downloads for off-line use of information:** This variable makes it possible to quantify the value of information made available through the repository via downloads.
 - **Online usage:** This variable provides a quantification of the platform value for the online use of the platform via the time spent by unique visitors to the website. The web usage is based on Piwik data. Piwik is a free and open-source web analytics application used by Zenodo for its metrics. The value of time was calculated, based in the average salaries per minute for researchers in EU countries.



Future collider project impact model assumptions

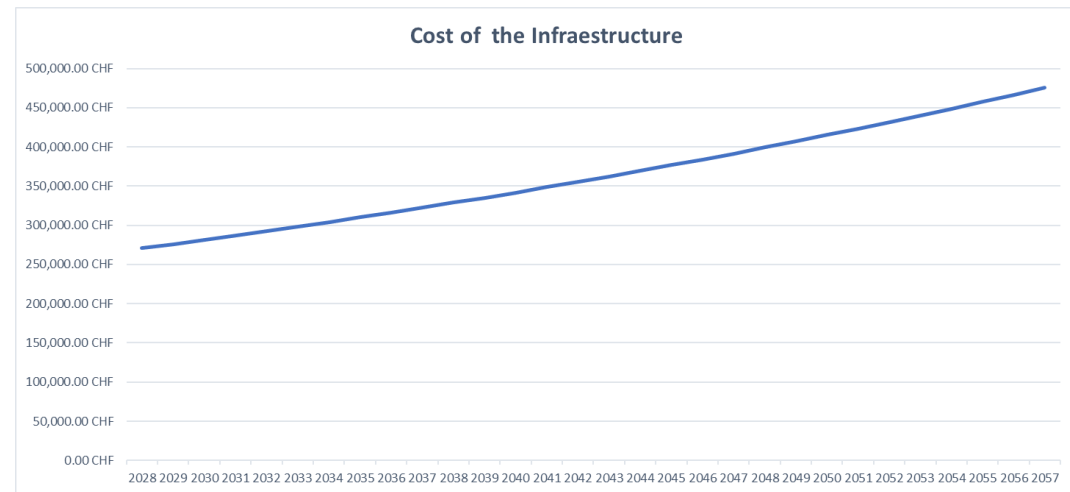
Timeline	2028 – 2057 with 2021 as a base year
Annual variable value change	The variables analyzed in the study will gradually increase until 2040 and then remain constant, except the costs of the infrastructure.
Social Discount Rate	2 % established by an economist expert panel in the EU project in which this work is carried out.
Currency Exchange rate	Fixed at 1 € = 1.07 CHF
Value of time per user	0.30 €/min extracted from the Eurostat database.
Virtual value of a single download	7 € established by analysing comparable platforms (e.g. IEEE Xplore) and a common good value survey



Infrastructure costs

Assumptions

- ❖ Constant annual increase of 4% throughout the analysis period.
- ❖ Costs are based on the assumption of salaries for three constantly needed full time equivalent (FTEs).
- ❖ Total cost of infrastructure (2028 – 2057) discounted: **12.9 MCHF**



Quantified of the repository's benefits

Data storage value

- ❖ **Market prices:**
 - < 50 GB or 50 GB per document upload = **111.72 €**
 - > 50 GB = **46.55 €** per +10 GB
- ❖ **10% annual increase** in documents **under 50 GB in size** until 2040, then constant until 2057.
- ❖ **3% annual increase** in documents **over 50 GB in size** until 2040, then constant until 2057.
- ❖ Total discounted benefit of **1.3 billion CHF** for the observation period of **29 years**.

Online usage value

- ❖ **Value of time: 0,30 €/min.**
- ❖ Total time spent by a single visitor on average per visit → **4 minutes**
- ❖ % annual increase of visitors until 2040, then constant until 2057 → **10 % annual increase.**
- ❖ Total discounted benefit of **1.4 billion CHF** for the observation period of **29 years**.

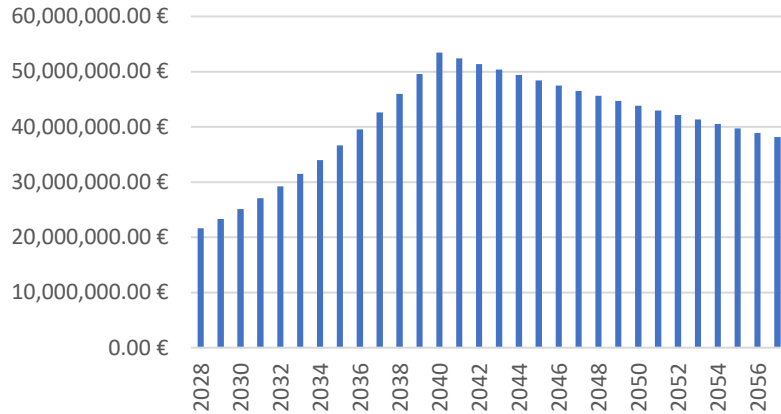
Offline information value

- ❖ **Average value of the download of one information record (paper, data, software): 7 €.**
- ❖ Given the impossibility of computing unique downloads per document, it was decided to consider only one single download per individual record stored in the repository throughout the entire lifetime of that information record.
- ❖ Total discounted benefit of **82 million CHF** for the observation period of **29 years**.

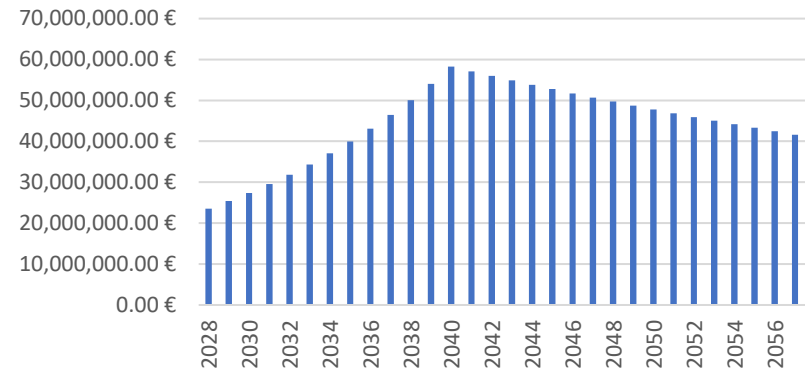


Impact potential overview

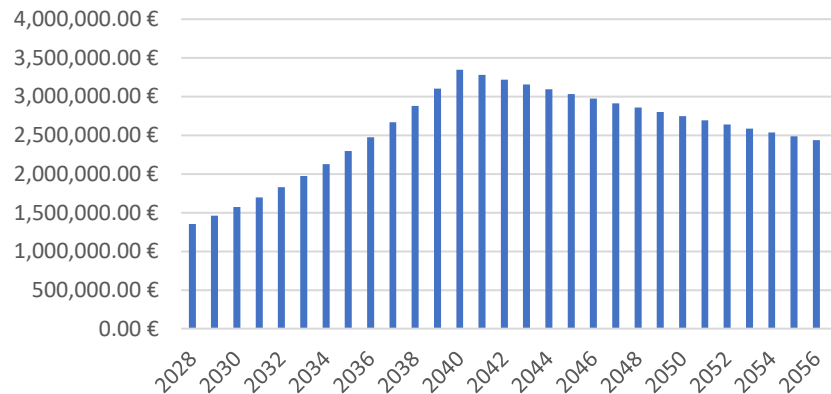
Benefit of data storage discounted



Benefit of online usage discounted



Benefit of downloads discounted



The combined discounted socio-economic impact potential of all model variables results in **2.8 billion CHF** for the 29 years observation period **between 2028 and 2057** or an **average annual impact potential value of about 95 million CHF**.

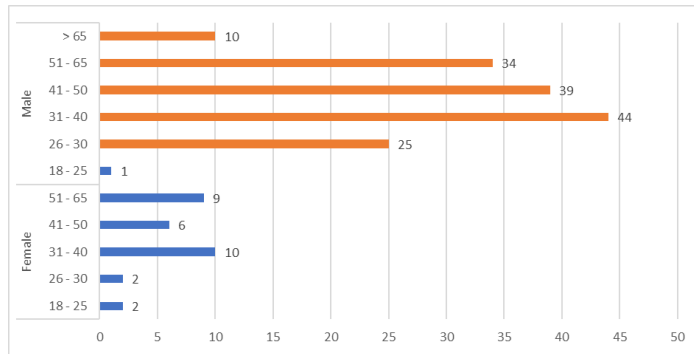


Zenodo as a common good value identification

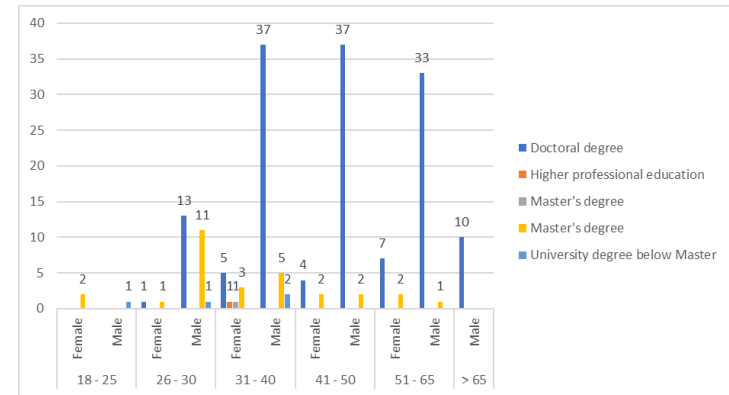


❖ Sample → 182 respondents that collaborate in the frame of the FCC study

Age range of the respondents by gender



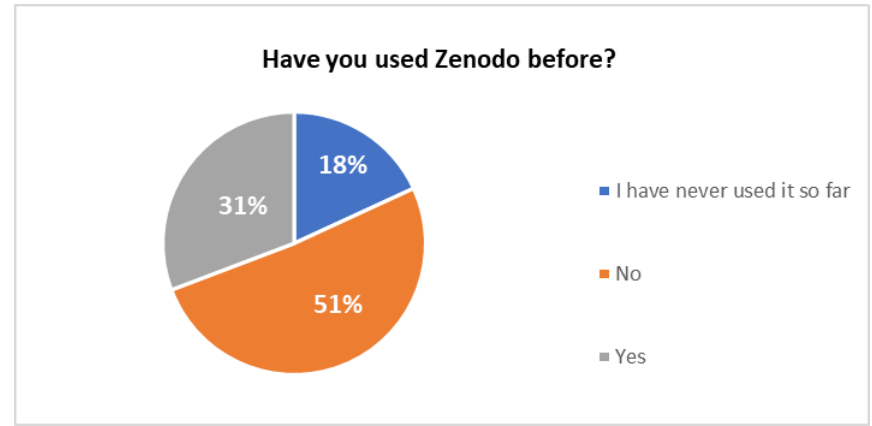
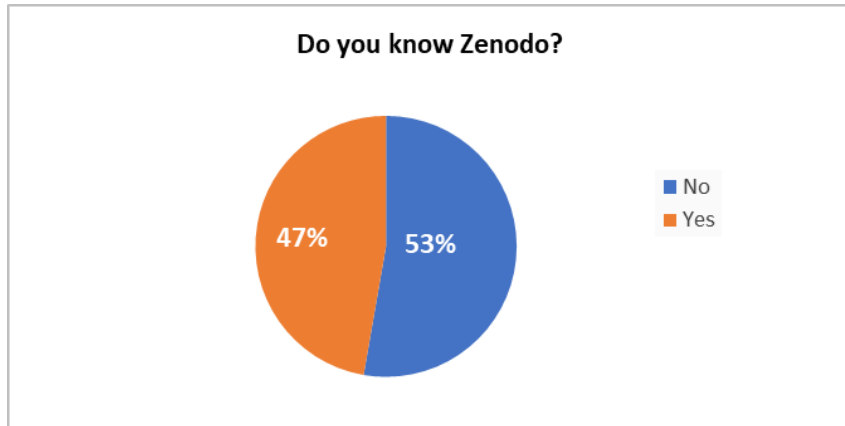
Highest level of education of the respondents in function of gender and age range



Current occupational status in the function of the type of firm/organization and gender

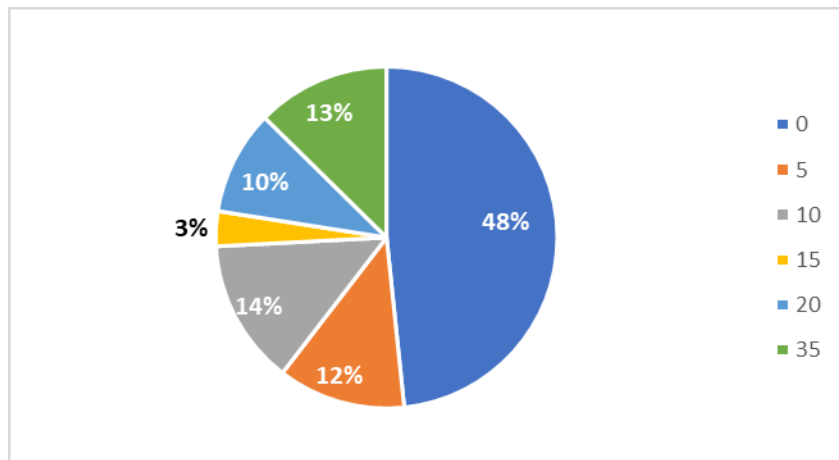
Row Labels	Employed	Retired	Self-employed	Student	Unemployed	Working at CERN
Female	18	1	1			9
Research (public or private university, research centre)	12	1				8
Company	2		1			1
Education (public or private school)	3					
Public administration	1					
Male	107	2	1	1	1	41
Research (public or private university, research centre)	93		1		1	41
Company	1					
Education (public or private school)	11	2		1		
Public administration	2					
Grand Total	125	3	2	1	1	50



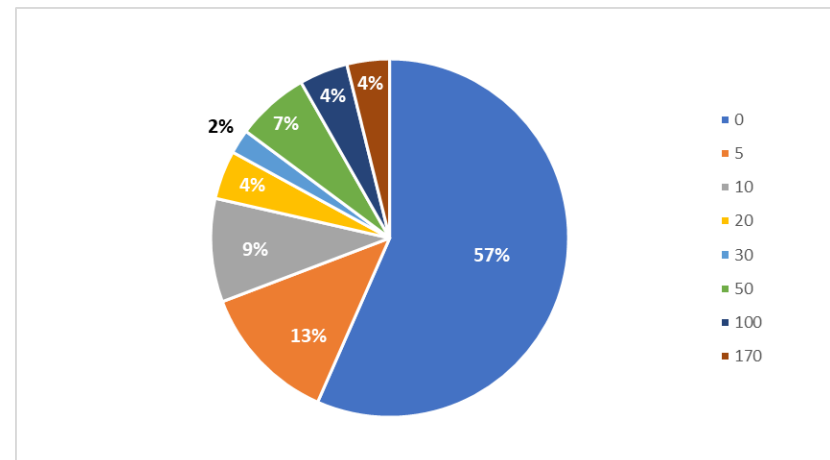


“Willingness To Pay” (WTP) is a means to reveal the **common good value of a public item**. The indication of a zero WTP is expected and is included using a mathematical formula.

WTP levels per year to have unlimited data storage for scientific information in Zenodo:



WTP levels per year in order to obtain unlimited access to scientific information in Zenodo:



Regression model

A regression analysis was carried out with the following variables:

- **Respondents' salary** → This value is taken from the Glassdoor database. The average salary considered is that of a researcher in the countries of residence provided in the survey.
- **Gender → Dummy variable:** 1 = male, 0 = female
- **Age**
- **Education background → Dummy variable:** 1 = natural sciences, mathematics and statistics, 0 = other
- **Level of education → Dummy variable:** 1 = doctorate level, 0 = other
- **Type of occupation → Dummy variable:** 1 = employed and retired, 0 = working at CERN
- **Respondent knows the repository → Dummy variable:** 1 = Yes, 0 = No
- **Respondents uses the repository → Dummy variable:** 1 = Yes, 0 = No and never used



Regression model – Results

WTP of downloads = 17,34 €

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	37.9119079	14.9497144	2.536	0.0121 *
Average.salary..	-0.0002237	0.0001384	-1.616	0.1080
Value.gender	1.0833594	8.3250936	0.130	0.8966
Average	0.1775843	0.2398328	0.740	0.4600
Value.educ.back	-12.6918199	8.3757281	-1.515	0.1315
Value.level.education	-10.5984002	7.9794152	-1.328	0.1859
Value.occupation	-3.8330448	6.4212479	-0.597	0.5513
Value.know	14.4725224	8.1126245	1.784	0.0762 .
Value.use	-7.7266875	8.9129292	-0.867	0.3872

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

WTP of storage = 8,87 €

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.88962012	4.55392829	1.293	0.19763
Average.salary..	-0.00002316	0.00004217	-0.549	0.58358
Value.gender	2.48285577	2.53596010	0.979	0.32892
Average	0.05689920	0.07305701	0.779	0.43714
Value.educ.back	-2.71057583	2.55138419	-1.062	0.28954
Value.level.education	0.10433976	2.43066080	0.043	0.96581
Value.occupation	-1.63634112	1.95601747	-0.837	0.40399
Value.know	7.56168677	2.47123852	3.060	0.00257 **
Value.use	-2.04009180	2.71502448	-0.751	0.45343

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	43.8015280	17.9429574	2.441	0.0156 *
Average.salary..	-0.0002468	0.0001661	-1.486	0.1392
Value.gender	3.5662152	9.9919500	0.357	0.7216
Average	0.2344835	0.2878523	0.815	0.4164
Value.educ.back	-15.4023957	10.0527226	-1.532	0.1273
Value.level.education	-10.4940604	9.5770597	-1.096	0.2747
Value.occupation	-5.4693859	7.7069150	-0.710	0.4789
Value.know	22.0342092	9.7369402	2.263	0.0249 *
Value.use	-9.7667794	10.6974826	-0.913	0.3625

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

WTP total = 26,21 €



Conclusions

- ❖ **The survey sample** could have been **more varied** and more information could have been obtained. Even so, we were able to obtain significant results.
- ❖ The **only variable** that influences the model is **knowledge of the platform**. This shows that the choice of WTP is not influenced by the type of education or the age of the respondent.
- ❖ And what is **most remarkable** is the response of the respondents to the knowledge and use of the platform. Based on the fact that all respondents have a relationship with CERN and many of them work in the organization, **it is surprising that most of them do not know about the existence of the virtual repository.**





THANK YOU