J. Benefit Cost Anal. 2021; 12(1):170–198
doi:10.1017/bca.2020.26

Phoebe Koundouri, Nikos Chatzistamoulou*,
González Davila, Amerissa Giannouli, Nikolaos Kourogenis,
Anastasios Xepapadeas and Petros Xepapadeas

Open Access in Scientific Information: Sustainability Model and Business Plan for the Infrastructure and Organization of OpenAIRE

Abstract: In 2008 European Commission launches the open access infrastructure for research in Europe project (OpenAIRE), supporting open access (OA) in scientific information and research output. In this paper, we assess the economic sustainability of the OpenAIRE project. The empirical strategy is developed through a Cost–Benefit Analysis framework to evaluate and compare the costs and benefits of OpenAIRE services to provide recommendations on the project's economic

Phoebe Koundouri: School of Economics, Research Laboratory on Socio-Economic and Environmental Sustainability (ReSEES), Athens University of Economics and Business, Athens, Greece
International Center for Research on the Environment and the Economy (ICRE8), Maroussi, Greece
and
Grantham Research Institute, London School of Economics and Political Science, London, United Kingdom
***Corresponding author: Nikos Chatzistamoulou**, School of Economics, Research Laboratory on Socio-Economic and Environmental Sustainability (ReSEES), Athens University of Economics and Business, Athens, Greece, e-mail: chatzist@aueb.gr
González Davila: Department of Economics & IFCELS UK SOAS, University of London, London, United Kingdom
Amerissa Giannouli: School of Business, Athens University of Economics and Business, Athens, Greece
Nikolaos Kourogenis: Department of Banking and Financial Management, University of Piraeus, Piraeus, Greece
Anastasios Xepapadeas: Department of International and European Economic Studies, Athens University of Economics and Business, Athens, Greece
Petros Xepapadeas: Department of Management Science and Technology, Athens University of Economics and Business, Athens, Greece

efficiency and sustainability, a non-market valuation method based on the results of a "Choice Experiment" to calculate the Total Economic Value generated by OpenAIRE and a full preference ranking approach. Findings indicate that stakeholders prefer interoperability between research platforms and output, better access to scientific results and compliance to OA mandates. Furthermore, net social benefits for the basic services for 15 years are at least five times higher than costs' present value while the potential R&D effect from research suggests even larger benefits in the long run. Subscriptions based on the estimated willingness to pay and cost, institutional subsidies and public awareness are the main recommendations for the sustainable operation of OpenAIRE. This study contributes to the literature on monetary valuation of the benefits and costs of OA to scientific knowledge.

Keywords: choice experiment; cost benefit analysis; open access; OpenAIRE; rank-ordered logit; research and economic valuation

JEL classifications: C25; C35; C51; D61; D80; L17

1. Introduction

In the literature there is a general agreement that "publicly funded research data are a public good, produced in the public interest, and as such they should remain in the public realm" (OECD, 2006). This applies to all research results, data, and literature, as scientific and technical advances are made possible only by sharing research results. There is a consensus among many authors (e.g., Cavaleri *et al.*, 2009; Conley & Wooders, 2009; Willinsky, 2009; McCabe *et al.*, 2013; Odlyzko, 2013) that despite the easier and faster access to a wide range of research information because of technological innovation and digitization, there are many problems in scientific publishing. Commercial journal prices tend to go ever higher and their publishers earn huge profits by charging libraries a large amount of money.

Considering market imperfections, the market for scientific publishing is not an ideal, perfectly competitive private market. The European Commission (2006) stresses the three main features of the market that cause imperfections. First, the material published in scientific journals is mostly publicly funded. Second, considering that authors are consumers of scientific output as well as producers of it, the private and the social values of publications may differ according to how the individual researcher behaves (as an author or as a reader). Third, the market is intermediated. Market imperfections tend to strongly weaken the price elasticity of consumers (authors or readers). The current price evolution reflects the ability of publishers to take advantage of the relative price elasticity of demand, and especially for their most popular journals.

Many authors (e.g., Getz, 2005; Kircz, 2005; Houghton & Sheehan, 2006, 2009) support open access (OA) and believe in OA's benefits for the community and for research. There are benefits for a number of stakeholders as well: researchers and research-performing institutions, research funders and society in general. In addition, Houghton and Sheehan emphasize the OA's potential impact in Research and Development (R&D) as well as in economic and social development.

In August 2008 European Commission launched the open access infrastructure for research in Europe (OpenAIRE) project, supporting OA in scientific information and research output. OA corresponds to unrestricted online access to peer-reviewed scholarly research. OpenAIRE connects institutional and thematic repositories, OA journals and Current Research Information Systems, developing and promoting interoperability mechanisms for the efficient dissemination of scientific content.

Evidence suggests that publishers are concerned about their journals' financial viability, which could be challenged by the OA repositories. On the other hand, libraries using OA repositories can lower their expenses. Funders care about their investments in research and how OA affects the research results. The difference between the stakeholders' needs and expectations creates different attitudes toward OA, which brings to the forefront the discussion about who is affected more from OA and how much.

To answer these questions we need to evaluate the benefits and costs of OA relative to the stakeholders involved in scientific publishing and scientific dissemination. For this purpose, a sustainability study was conducted in order to provide OpenAIRE with an accurate estimation of the benefits and costs of the OpenAIRE infrastructure and to build a sustainable business model for the continuation of OpenAIRE beyond the life-time of the project funding. Koundouri *et al.* (2012), measured people's willingness to pay (WTP) in order to gather more scientific information before they decide on the management scheme on climate change mitigation effects on Rokua esker in Northern Finland. However, this study is the first attempt to value the benefits and costs of OA to scientific knowledge monetarily.

Given that OA in research output is a public good and the benefits do not have market values, a Choice Experiment Method (CEM) was used to value the benefits associated with the OpenAIRE infrastructure monetarily. Pre-requisite for the Choice Experiment was the stage of identifying, prioritizing, and mapping the relevant stakeholder groups and their needs. After identifying and monetarily valued the benefits of OpenAIRE, we compared them with the costs produced from the OpenAIRE infrastructure applying a Cost Benefit Analysis. It should be noted that a choice experiment is adopted for the first time to value benefits and costs of scientific research output and knowledge.

2. The OpenAIRE infrastructure

There is a need to coordinate the development and sustainability efforts of OA e-Infrastructure initiatives, within as well as across initiatives. The EC (2006) states that “the goal of OpenAIRE is to advocate and enable science via an interoperable data infrastructure capable of collecting publications and data and interlinking and contextualizing them.”

OpenAIRE has built up a participatory infrastructure of people, repositories, and technologies, which provides OA to publications and a support network for the implementation of the EC’s OA policies across 32 European countries and it is extending its scope with the OpenAIREplus project by connecting publications to contextual information, such as research data.

OpenAIRE is attached to a number of benefits as it maximizes the discoverability and accessibility of research outputs, enhances research dissemination reduces the project coordination costs, enables institutions to offer services to their researchers, and enables research funders and institutions to monitor the research output. Furthermore, it enables publishers to offer add-on services to authors, providing opportunities for reimbursement of article processing charges.

Willinsky (2011) says that “the EC’s launch of OpenAIRE provided an encouraging and enlightening moment of thinking about how greater access to knowledge will contribute to the educational and democratic quality of our lives.” The funding model of OpenAIRE currently relies on an EC FP7 grant matched by institutional in-kind contributions. Additional resources are now further explored based on the results of this study.

3. Methodology

The methodology is based on the following objectives; a graphical illustration of the process can be found in the online appendix.

Stakeholder definition and given benefits: it is important to identify and prioritize the stakeholder groups and their needs. This study seeks ways to assign a monetary value to the benefits these stakeholders gain from OpenAIRE.

Accounting: how much does the current system setup, operation, and maintenance cost? It is important that this study distinguishes between the operation of existing services, upgrades of the system, and the development of new services.

Cost benefit analysis: how do the system costs respond to the benefits of the identified stakeholders?

Table 1 Summary of stakeholder categories.

1	Scientists and researchers
2	Research funders
3	Research centers and laboratories
4	Publishers
5	Scholarly and learned societies
6	Research communities
7	Libraries and library organizations
8	Repository service providers and standards groups
9	National OA desks
10	University administration and university organizations
11	OA organizations
12	Preservation services
13	Other repositories
14	Primary and secondary education instructors and students
15	Patent, trademark, and technology transfer, commercialization offices

Abbreviation: OA, open access.

Revenue channels: identify the best and most viable model for OpenAIRE to spread the costs among beneficiaries (including service charges where appropriate) for its services. Who contributes, how much and when?

3.1 Stakeholder analysis

Using the latest research results from the relevant literature, in this study, we identify the initial stakeholders' categories as shown in Table 1.

Then, a specific questionnaire was implemented to those focus groups, providing information for the stakeholder analysis. The most important stakeholders were identified as well as their expertise, how important is their role in OA and OpenAIRE, as well as their willingness to engage (WTE) with/support them. Based on their characteristics, an initial two-dimensional (2D) stakeholder mapping was provided in order to identify and depict the most relevant stakeholders to implement the questionnaire. The results from the stakeholder questionnaire were used in a 3D stakeholder mapping for a more complete representation of those stakeholders and their relevance with OpenAIRE.

The mapping allows seeing where stakeholders stand when evaluated by the same key criteria and compared to each other, and helps to visualize the complex interplay of issues and relationships created according to their contribution, their legitimacy, their WTE, their influence and their necessity of involvement.

We map the stakeholders according to their expertise, WTE, and to their type and category. "Expertise" $\left(e_{ij}\right)$ is assigned to the Z-axis $\left(0 \leq e_{ij} \leq 1,\right.$ where 0 means no expertise and 1 full expert). "Willingness to engage WTE"$\left(w_{ij}\right)$ is assigned to the

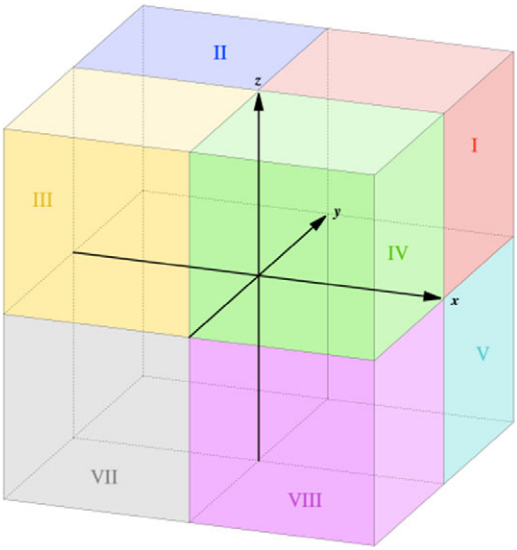


Figure 1 Location of octant (+, +, +).

X-axis $\left(0 \leq w_{ij} \leq 1\right.$, where 0 means no WTE and 1 full WTE). “Stakeholders” (SH_{ij}) are assigned to the Y-axis where i identifies the 15 stakeholder categories$(i = 1, 2, 3, \ldots, 15)$ and j identifies the set of “n” stakeholders in each category $(j = 1 \ldots n)$. The only relevant space since we define $eij \geq 0$, $wij \geq 0$ and $\exists\, eij \wedge \exists wij \leftrightarrow SHij \neq \{\emptyset\}$ is the one labeled as I in Figure 1 (Octant (+, +, +)). Then, using the stakeholder questionnaire the following questions were used to assess expertise:

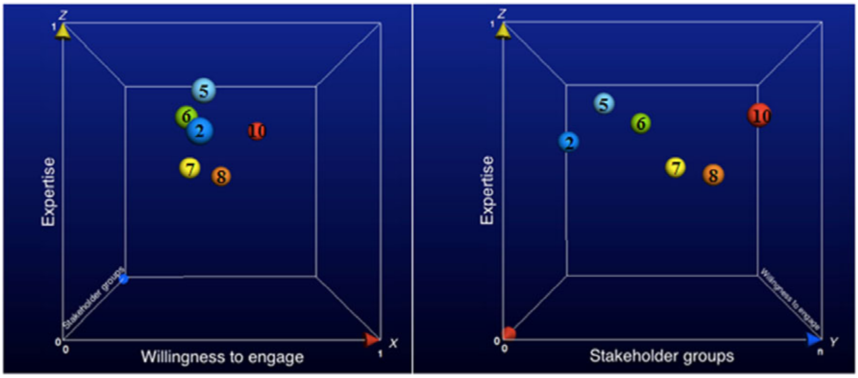
Question 1. *How would you describe your knowledge level of OA and related initiatives?*
The answers were coded between 0 and 1 (where 0 = Not knowledgeable, 0.25 = Have only a general sense, 0.5 = Somewhat knowledgeable and 1 = Very knowledgeable).

Question 2. *Are you familiar with the OpenAIRE initiative?*
The answers were coded between 0 and 1 (where 0 = Not familiar at all, 0.5 = Somewhat familiar and 1 = Yes, I know the initiative).

Then, an average for each stakeholder category for questions 1 and 2 was obtained and the maximum was identified and plotted in axis Z.

On the other hand, the answers to Question 3 were used to assess WTE:

Question 3. *How does/will OpenAIRE support your (or your organization’s) work?*
The answers were coded between 0 and 1 (where 0 = Not at all, 0.5 = Not much and 1 = Very much).



Stakeholder (Y)		Expertise (Z)	WTE (X)
Library or Archive	2	0.68	0.42
Repository Service Provider	5	0.86	0.43
Research Funder	6	0.78	0.35
Research Organization or Laboratory	7	0.56	0.35
Researchers and Scientists	8	0.53	0.50
University administration (dean, provost, chair, etc.)	10	0.77	0.69

Figure 2 Stakeholders mapping.

Then an average for each sub-question was obtained and the maximum was identified and plotted in axis *X*. The illustrations in Figures 1 and 2 present the parameters obtained from the Stakeholder Questionnaire and the 3D mapping.

It must be stressed that the responses in the following categories: one Researcher in a Corporation, three National Open Access Desk (NOAD), four Publisher and nine Teacher (elementary/middle/high school), were excluded from the mapping since the rate of response was very low and it was not possible to use their information for the mapping exercise.

The most relevant stakeholder categories identified were six: Libraries or archives, Repository service providers, Research funders, Research organizations/ laboratories, Researchers or scientists, and the University administration. The university administration had the highest level of WTE with OpenAIRE as well as a high level of knowledge about its services. Researchers and scientists showed a low level of WTE but they were marginally more optimistic considering the possible benefits that OpenAIRE could create. The rest of the categories were quite knowledgeable and interested in OpenAIRE but they were not very willing to engage. Additionally, research funders and research organizations/laboratories were not very optimistic concerning OpenAIRE's benefits. Most of the stakeholders contacted were not aware of the existence or services provided by OpenAIRE. Given the results of this analysis, it is suggested that a campaign of awareness should be implemented in order to disseminate services and tools provided by OpenAIRE among the scientific community.

3.2 Monetary valuation of the benefits: choice experiment

The CEM is a Stated Preference Method (Birol and Koundouri, 2008) that elicits the total economic value of non-market goods, which can in turn be used to produce effective policies for sustainable management and conservation. It has a theoretical grounding in Lancaster's characteristics theory of value (Lancaster, 1966), which states that any good can be described in terms of its characteristics, or attributes, and the levels that these attributes take. CEM provides information about the significant determinant attributes of the values that stakeholders place on a public good, the implied ranking of these attributes amongst the relevant stakeholders, the value of changing more than one of the attributes at once and the total economic value of the public good (Bateman *et al.*, 2003).

CEM has an econometric basis in models of random utility theory that derives from Luce (1959) and McFadden (1973). Suppose that we can represent an OpenAIRE stakeholder's preferences by the following utility function:

$$U = U(X_1, \ldots, X_m; Z_1, \ldots, Z_n). \tag{1}$$

The utility for this stakeholder depends on the levels of $X = 1, \ldots, m$ OpenAIRE goods and services consumed and on $Z = 1, \ldots, n$ available goods. Utility is assumed to be a function of observed and unobserved factors relating to choice alternatives and decision-makers. Because as already mentioned the researcher cannot observe all the factors that determine utility toward the alternative choices, we divide the conventional utility function $(U(.))$into a non-deterministic, observable part $(V(.))$, and an error, unobservable part $(e(.))$:

$$U = V(X) + e(X, Z), \tag{2}$$

where *X*, *Z* represent vectors.

3.3 Construction and implementation of the choice experiment questionnaire

3.3.1 Experimental design

The target population consists of those who receive benefits or costs of the OA in scientific information using OpenAIRE. The choice experiment questionnaires of the study are addressed to the European Population. The implementation of the OpenAIRE choice experiment questionnaire was based on the relevant stakeholders' categories provided by the stakeholder analysis. A quota sampling technique was used, while European stakeholders from all the stakeholder categories were asked via email to respond to the questionnaire. The ultimate selection of respondents was not made by a probability mechanism (non-probabilistic sampling design), although the sample frame population was divided into researchers and non-researchers (stratified probabilistic sampling). The separate stratum estimates were combined (weighted) to form an overall estimate for the entire population based on the stakeholder analysis results.

The Researchers' questionnaire consists of 21 and the Non-researchers' questionnaire of 20 questions. Both questionnaires were implemented using the SurveyMonkey®. Questionnaires are separated in sections (A, B, C, D, E). There is an introduction about the scope of the survey and then a few basic questions (Section A). In Section B, we investigate the general attitude of the respondent toward OpenAIRE and OA. After that, OpenAIRE is described to the respondent in case he/she is unfamiliar with the initiative (Section C). In order to minimize the respondent's confusion we used videos for the description as well as different icons to explain who and how can get involved with OpenAIRE. Section D, includes the main valuation method of the study and the choice cards based on the occupational characteristics of the respondent, accompanied by an irrelevant example of a choice card and follow-up questions. E, the final section of the questionnaires contains the socio-economic characteristics questions.

Face to face interviews and Skype calls/webinars were also provided to respondents because email surveys tend to elicit very low response rates, although email surveys of special populations generate significantly higher response rates (Bateman *et al.*, 2002). For the questionnaire that is addressed to researchers, budget reallocation was used because researchers do have different attitudes toward sources of funding for OpenAIRE and OA in general. The different approaches of the two questionnaires indicate that elements related to the payment conditions, influence preferences for OpenAIRE infrastructure.

The survey design included the selection of attributes, the definition of attribute levels, the choice of the experimental design to allocate alternative scenarios to choice tasks to present to respondents, and the elicitation of preferences by asking respondents to rank the alternative scenarios in each choice task. When selecting the choice experiment attributes, the development of OpenAIRE's services was considered in the construction of the scenario descriptions that were used in the survey design. We choose to have two levels for each attribute. Level 1 represents the current OpenAIRE's services status, assuming that the attributes could not be provided separately. This was assumed in order to avoid a more complicated choice experiment questionnaire, since the object of the study supports many details that could puzzle the respondents (see Bateman *et al.*, 2002).

Also, the basic services of OpenAIRE are complementary, meaning that you cannot provide the services on Level 1 separately. Level 2, on the other hand, represented possible future services, completely independent from each other in terms of infrastructure and cost. The non-researchers were presented with nine separate choice cards and the researchers with eight. This was done because, based on the cost analysis and stakeholder analysis, the non-researchers could represent categories with lower income than the researchers. The attributes and their levels associated with different OpenAIRE profile options are presented in Table 2. Data for the final survey were collected from October to November 2014. The average completion time for an interview across those who completed the ranking tasks was 30 min.

3.4 Choice experiment data collection and descriptive statistics

Descriptive statistics were based on 196 completed questionnaires. The sample collected consisted of 105 non-researchers and 91 researchers. Comparison of the complete with the incomplete responses showed that respondents more familiar with the current publishing condition and in favor of OA and/or OpenAIRE were more likely to complete the questionnaire.

Differences between complete and incomplete responses cease to exist when we examine the valuation part of the questionnaire and the choice cards, indicating that stakeholders who are interested in OpenAIRE continue answering. However, another indicator of selectivity bias is the fact that more than 50 % of the sample states that all attributes affected very much the way they made their choices (Figure 3).

12.76 % of the respondents chose not to contribute to OpenAIRE, 40 % of the respondents that chose not to contribute stated that the reason was because OpenAIRE would take them too much time to understand and use. Finally, the selectivity bias was more obvious when we estimated separately the percentage of responses that came from South Europe. South European research sector and university professors tend to earn less than the rest of Europe, which is depicted on the low income statistical results.

Full ranking approach

The experimental design approach used in construction of both choice experiment questionnaires for OpenAIRE is the full ranking approach. Scarpa *et al.* (2009) used a high quality rank-ordered data in which the ranking of alternatives is elicited by means of the best-worst approach to alternative selection. The same approach was

Table 2 OpenAIRE attributes and levels.

	Attribute levels		Level description
Access to scientific results and compliance to OA mandates	1	Access OA Science + Deposit + Deposit with Embargo	i. Easy access to literature and connection with similar research/ scientists ii. Option to self-deposit allowing compliance with funder and institutional mandates iii. Option to restrict access to the results for some period.
	2	Access OA Science + Deposit + Deposit with Embargo + Supplemental material	Access OA Science, Deposit, Deposit with Embargo (+iv) Possibility to deposit more than the article itself (including appendices, data sets, programming code, related funding information)
Interoperability	1	Retrieve scientific impact	(i) Retrieve article, project, institution citations and alternative types of metrics (e.g., number of downloads, tweets, social/ researcher networks).
	2	Retrieve Scientific impact + Create research profile + Publication-Dataset Resolver	Retrieve Scientific impact (+ii) Option to create a profile to brand the research and the impact of the author's institution (i.e., link to funding). (+iii) Provides links from publications to related data sets and vice versa.
Misc Services and OA Costs	1	Check Copyright + Research Analytics + H2020 reporting tool	(i) Verifies that copyright is respected, (ii) Provides research analytics for the funders (iii) Reporting tool for H2020, with possible extension to future frameworks
	2	Check Copyright + Research Analytics + H2020 reporting tool + Publishers compensation + EC post project publication APCs for Gold OA.	Check Copyright, Research analytics and Reporting tool for H2020 (+iv) Services to publishers (peer-review etc.) to compensate for the OA. (v) Processes EC APCs for Gold OA for the EC for after the end of the project in order to continue publishing.

Abbreviations: APCs, Article Processing Charges; OA, open access.

used in these questionnaires. Rank-ordered choices are well-known (Hausmann & Ruud, 1987) to provide researchers with richer preference information than simply asking a respondent to state their favorite alternative and/or provide partial rankings.

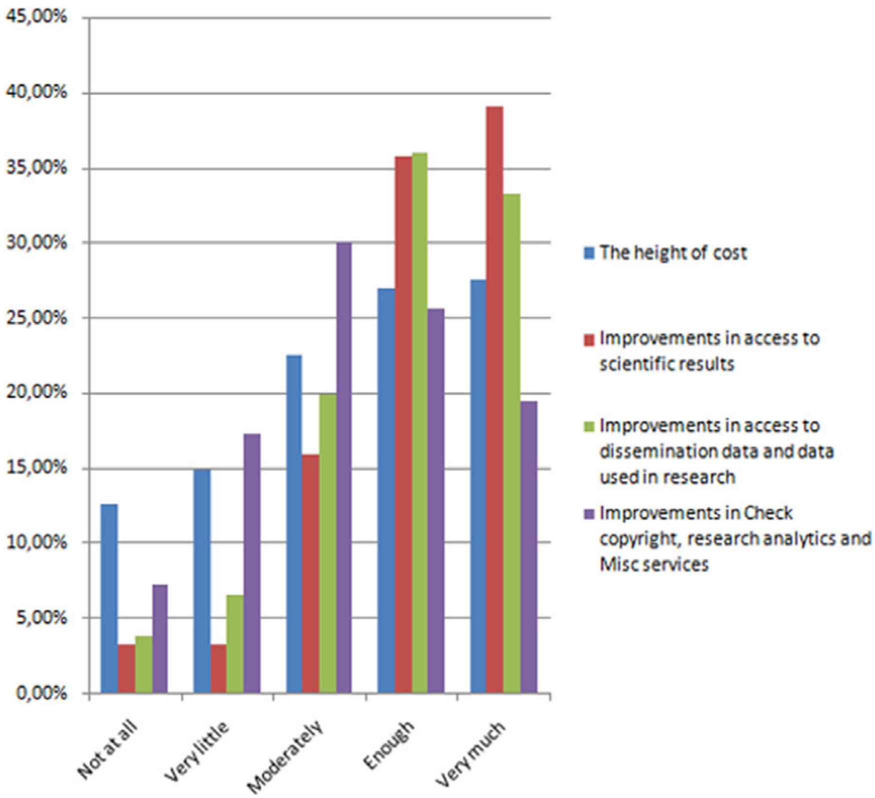


Figure 3 Question: To what degree do the following reasons affect the way you made your choices in the choice cards above?

Although, respondents might be reluctant to engage in such time consuming surveys, this approach provides rich statistical information.

Under this approach, the interviewer presents a set of five choice cards to the respondent who is instructed to follow the following sequential choice process: First, the respondent chooses the most preferred alternative out of the initial five alternatives in the choice set. This best alternative is then excluded from the choice set and the respondent is asked to select the least preferred out of the remaining four, which is also excluded. This process is repeated for the remaining three alternatives from which the respondent selects the second most preferred out of the remaining three, and finally the second least preferred out of the remaining two cars.

In the case of data obtained with this twice repeated best-worst approach on a choice set with five alternatives denoted (A_1, A_2, A_3, A_4, SQ), the analyst identifies four responses (x^{1b}, x^{1w}, x^{2b}, x^{2w}), where the subscripts denote first best, first worst, second best, and second worst. This leads to the following preference ordering ($x^{1b} > x^{2b} > x^{2w} > x^{1w} > x^{r}$), where the subscript r denotes the residual alternative. The rank-ordered logit model can be described with:

$$\begin{aligned}\Pr\left(x^{1b} > x^{2b} > x^{2w} > x^{1w} > x^{r}\right) = \Pr\left(x^{1b} | x^{1w}, x^{2b}, x^{2w}, x^{r}\right) \times \Pr\left(x^{1w} | x^{2b}, x^{2w}, x^{r}\right) \\ \times \Pr\left(x^{2b} > x^{2w}, x^{r}\right) \times \Pr\left(x^{1w} | x^{r}\right).\end{aligned} \tag{3}$$

Using the assumptions of a sequence of independent logit choice probabilities, each full ranking gives the following product of logits, where v denotes the indirect utilities of the relevant alternatives:

$$\begin{aligned}\Pr\left(x^{1b} > x^{2b} > x^{2w} > x^{1w} > x^{r}\right) = \left(\frac{\left(\exp\left(v^{1b}\right)\right)}{\sum_{j \in 1b,2b,2w,1w,r}\left(\exp\left(v^{j}\right)\right)}\right) \times \left(\frac{\left(\exp\left(v^{2b}\right)\right)}{\sum_{j \in 2b,2w,1w,r}\left(\exp\left(v^{j}\right)\right)}\right) \\ \times \left(\frac{\left(\exp\left(v^{2w}\right)\right)}{\sum_{j \in 2w,1w,r}\left(\exp\left(v^{j}\right)\right)}\right) \times \left(\frac{\left(\exp\left(v^{1w}\right)\right)}{\sum_{j \in 1w,r}\left(\exp\left(v^{j}\right)\right)}\right).\end{aligned} \tag{4}$$

The ranking model relies critically on the IIA assumption, which permits the multiplication of the successive probabilities. The parameters of the utility function can be estimated by maximizing the log-likelihood function, where j denotes the different alternative choices and i denotes the specific stakeholder:

$$\log \ L = \sum_{i=1,2,3,\ldots} \sum_{j \in 1b,2b,1w,r} \log \left[\frac{\exp\left(v_i^{1b}\right)}{\sum_{j \in 1b,2b,2w,1w,r} \exp\left(v_i^{j}\right)}\right]. \tag{5}$$

Using the specification model of Scarpa *et al.* (2009), the first preferred choice is a selection out of five alternatives and relates to the specification of the scale parameter $\lambda = \exp\left(\sum_k q_k\right)$ with $j = 1,2,3,4,5$ via coefficient $q_5 q_5$ and a dummy-coded indicator function for that choice made in the context of the five alternatives. The second preferred choice is a selection from the remaining four (q_4). The third preferred choice is from the remaining three (q_3). The fourth preferred choice is from the remaining two alternatives (q_2), representing the least favorite alternative and the baseline for the scale effects (Table 3).

3.5 Econometric analysis

For individual i the rank of the j alternative is given by:

$$prob\left(Rank_{i,j} = k\right) = f\left(AltSp_j, LnSp_{i,j}\right) + \varepsilon_{i,j}. \tag{6}$$

Table 3 Mapping best/worst choice.

Instance	Choice	Rank	Alternatives in exploded logit choice set	Scale coefficient
1	Best 1	1	5	$q5$
	Worst 1	5	2	$q2$
2	Best 2	2	4	$q4$
	Worst 2	4	2	$q2$
Residual alternative		3	3	$q3$

where $AltSp_j$ *and* $LnSp_{i,j}$ are vectors of alternative specific and individual-specific variables. In our case, all alternative specific variables are binary except of the one that corresponds to the cost. Specifically, we consider as baseline option the one where OpenAIRE stops after the end of the funding from the European Commission. The option corresponds to $AltSp_j = 0$. Vector $AltSp_j$ can be represented as follows:

$$AltSp_j = [ALL\_Low, AccessHigh, MetaDataHigh, ServicesHigh, Cost]', \tag{7}$$

where LL_Low equals 0 under the baseline option and 1 otherwise. Variables $AccessHigh, MetaDataHigh, and\ ServicesHigh$ equal 1 only when the option involves the highest level of the corresponding attribute. Otherwise they equal 0.

Concerning the individual specific variables are also binary:

$$InSp_{i,j} = [RegionW, RegionS, RegionE, RegionG, Inc0\_Inc5, Inc5\_Inc10, Inc10\_Inc20, Inc20\_Inc40, IncG40]'. \tag{8}$$

They correspond to the region of the respondent (West, East, South, and Greece, because around 30 % of the sample were Greeks) and to his/her monthly income level (less than 500 euros, 500–1000, 1000–2000, 2000–4000, and >4000). Here, if a respondent did not answer these questions the corresponding variables are all zeros.

In the current model, we assume that the individual utility is adequately approximated by a linear function of the alternative specific variables around a region that corresponds to the baseline option and the proposed changes. In this case, the ratio of the estimated coefficient of a variable of interest (alternative specific) over the estimated cost coefficient times -1 provides us with the marginal rate of substitution which is the marginal WTP:

$$MWTP_{AltSp_j} = -\frac{b_{ALtSp_j}}{b_{Cost_j}}. \tag{9}$$

3.5.1 Econometric results: alternative specifics

A total number of 1225 answered choice cards were collected from 192 respondents. Seven hundred and thirty four choice cards came from 104 non-researchers and 491 came from 88 researchers. Because of the full ranking model, each choice card corresponded to four data points, making the results more robust. Based on the econometric results, OpenAIRE has a positive effect on respondents' utility (see Tables 1–3 in the online appendix). Higher level of interoperability and more access to scientific results, as well as compliance to OA mandates have a positive effect on respondents' utility. Higher level of miscellaneous services and OA costs have a negative effect but the coefficient is insignificant.

In order to derive the benefits from the services provided by OpenAIRE, we calculated the WTP using weights, since each stakeholder has a different WTE (see stakeholder analysis). For the alternative specific variables, we multiplied the estimated WTP from the stakeholder analysis by the WTP that resulted from the econometric analysis of the choice experiment results, and by the number of stakeholders.

The estimated WTP from the estimated econometric results was approximated using for the standard error an alternative variance expression, due to existing non-significant estimates. The following expression was used for the variance of the ratio of two estimates (Bateman *et al.*, 2002):

$$var\left(\frac{b_{ALtSp_j}}{b_{Cost_j}}\right) = \left(\frac{b_{ALtSp_j}}{b_{Cost_j}}\right)^2 \left(\frac{var\left(b_{ALtSp_j}\right)}{{b_{ALtSp_j}}^2} + \frac{var\left(b_{Cost_j}\right)}{{b_{Cost_j}}^2} - \frac{2Cov\left(b_{ALtSp_j} b_{Cost_j}\right)}{b_{ALtSp_j} b_{Cost_j}}\right). \tag{10}$$

Tables 4–6 include the total WTP from Level 1 to Level 2, the estimated WTP based on the different assumptions and scenarios of the services provided from OpenAIRE, respectively.

In order to approximate the WTP for each alternative scenario, the levels of the attributes based on the different scenario needed to be combined. The equation proposed by Bateman *et al.* (2002) slightly changes, using the variance of the sum of the coefficients:

$$Var\left(\frac{\sum b_{AltSp\,j}}{b_{Cost\,j}}\right). \tag{11}$$

By distinguishing Researchers and non-researchers, we observe that researchers are willing to contribute for higher level in each of the attribute categories (Table 7), in contrast to the non-researchers for whom more Miscellaneous Services and OA Costs

Table 4 WTP for the additional features offered.

Attribute	Additional feature	WTP	WTP (s.e.)	WTP *t*-stat
Access to scientific results and compliance to OA mandates	Supplemental material	252.02	119.51	2.11
Interoperability	Create research profile and publication-dataset resolver	203.06	104.29	1.95
Misc services and OA costs	Publishers compensation and EC post project publication APCs for Gold OA	−38.87	57.66	−0.67 (non-significance)

Abbreviations: OA, open access; WTP, willingness to pay.

Table 5 WTP for each alternative scenario.

	Alt1	Alt2	Alt3	Alt4	Alt5	Alt6	Alt7	Alt8
WTP	1763.13	2015.15	1966.20	1724.26	2218.21	1976.28	1927.32	2179.34
(s.e)	779.34	873.66	853.72	768.66	952.06	863.27	844.12	942.66
t-stat	2.26	2.31	2.30	2.24	2.33	2.29	2.28	2.31

Abbreviation: WTP, willingness to pay.

Table 6 Alternative OpenAIRE service scenarios.

Alternative scenarios	
1	Attribute 1 (Level 1) + Attribute 2 (Level 1) + Attribute 3 (Level 1)
2	Attribute 1 (Level 2) + Attribute 2 (Level 1) + Attribute 3 (Level 1)
3	Attribute 1 (Level 1) + Attribute 2 (Level 2) + Attribute 3 (Level 1)
4	Attribute 1 (Level 1) + Attribute 2 (Level 1) + Attribute 3 (Level 2)
5	Attribute 1 (Level 2) + Attribute 2 (Level 2) + Attribute 3 (Level 1)
6	Attribute 1 (Level 2) + Attribute 2 (Level 1) + Attribute 3 (Level 2)
7	Attribute 1 (Level 1) + Attribute 2 (Level 2) + Attribute 3 (Level 2)
8	Attribute 1 (Level 2) + Attribute 2 (Level 2) + Attribute 3 (Level 2)

is statistically insignificant. In addition, for researchers alternative 8, which includes higher level for the third attribute, is more preferable than the alternative 5, which is the most preferable for the non-researchers (Table 8). The covariance matrix of coefficients b_{AltSp_j} is introduced to estimate them. For more details about the calculations see Tables 4–9 in the online appendix, which present the extra calculations needed to calculate the WTP and the covariance matrix of coefficients.

Table 7 Researchers and non-researchers: WTP for additional features offered.

	Attribute	Additional feature	WTP	WTP (s.e.)	WTP *t*-stat
Researchers	Access to scientific results and compliance to OA mandates	Supplemental material	47.07	34.09	1.38
	Interoperability	Create research profile and publication-dataset resolver	56.19	43.24	1.30
	Misc services and OA costs	Publishers compensation and EC post project publication APCs for Gold OA	14.41	28.32	0.50
Non-researchers	Access to scientific results and compliance to OA mandates	Supplemental material	474.44	299.90	1.58
	Interoperability	Create research profile and publication-dataset resolver	302.94	201.33	1.50
	Misc services and OA costs	Publishers compensation and EC post project publication APCs for Gold OA	−24.75	92.77	−0.27

Abbreviations: OA, open access; WTP, willingness to pay.

Table 8 Researchers and non-researchers: WTP for alternative scenarios.

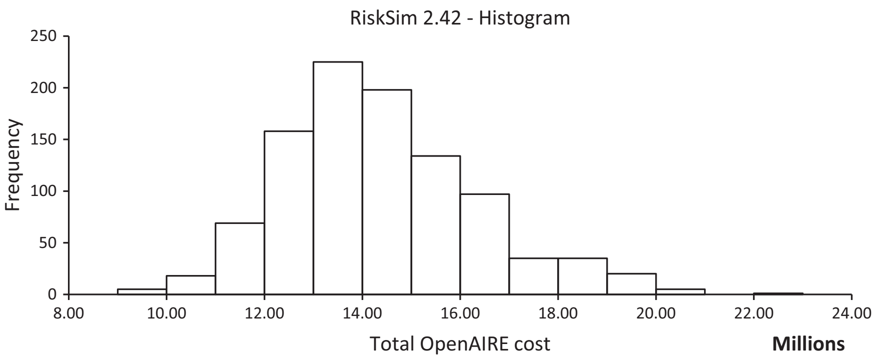
		Alt1	Alt2	Alt3	Alt4	Alt5	Alt6	Alt7	Alt8
Researchers	WTP	1222.35	1269.42	1278.55	1236.77	1325.62	1283.84	1292.96	*1340.03*
	(s.e)	539.36	559.21	562.66	547.35	583.60	567.91	571.72	593.29
	t-stat	2.27	2.27	2.27	2.26	2.27	2.26	2.26	2.26
Non-researchers	WTP	1232.11	1706.55	1535.05	1207.36	*2009.48*	1681.80	1510.30	1984.73
	(s.e)	840.27	1072.75	982.39	836.26	1233.73	1068.32	980.07	1230.74
	t-stat	1.47	1.59	1.56	1.44	1.63	1.57	1.54	1.61

Abbreviation: WTP, willingness to pay.

3.5.2 Econometric results: individual specific variables

The individual specific variables concern the region, income, and researcher or non-researcher status of the respondent. This stratification was used in order to correct for the sampling bias and the heterogeneity. According to the region, four variables were identified: South, Greece, North West, and East. The income was separated into five different variables ($Inc0\_Inc5, Inc5\_Inc10, Inc10\_Inc20, Inc20\_Inc40, IncG40$), corresponding to 0–500, 501–1000, 1001–200, 2001–4000, more than 4000 euros, respectively. For the researcher or non-researcher variable, a dummy variable was used, where one indicates researcher status and 0 indicates non-researcher status (see Tables 1–3 in the online appendix).

Table 9 Monte Carlo simulation for the stochastic model with the same growth rate.



3.6 Cost analysis and cost benefit analysis

3.6.1 Cost questionnaire

A cost questionnaire was implemented on the NOADs in order to extract the main costs of having the OpenAIRE infrastructure. The structure of the questionnaire was based on the technical and financial aspects of the OpenAIRE platform. Eighteen out of 29 NOADs responded to the Cost Questionnaire. These are the NOADs of Bulgaria, Sweden, Switzerland, Cyprus, Finland, Germany, Greece, Iceland, Italy, Lithuania, Luxembourg, Malta, Norway, Portugal, Slovenia, Spain, Turkey, and Belgium. Three regional coordinators responded to the questionnaire. Specifically, we received responses from: Universidade do Minho, Portugal (regional coordinator of the South), Ghent University, Belgium (regional coordinator of the West), and EIFL (regional coordinator of the East).

The questionnaire on the Technical cost was sent to the European Organization for Nuclear Research (CERN), Institute of Information Science and Technologies of the Italian National Research Council (CNR-ISTI), Uniwersytet Warszawski (ICM), University of Bielefeld, and University of Athens. The reported costs of CERN have been calculated per PByte of data, while the current unit of measurement is the TByte. Therefore, CERN costs could not be compared with the costs of the other partners.

It was observed that the NOADs' responses concerning the person-months (p-ms) for similar tasks exhibited large deviations. In order to rationalize the corresponding labor costs we first calculated for each task the median of the person-months provided from each respondent. In the calculation of the median we excluded any zero values. Then, we replaced the person-month values that exceed their corresponding median with the value of the median. This approach on the NOADs responses resulted to an aggregate labor cost of 250,914.31 euros representing a reduction of 116,678.76 euros on the total annual labor cost with respect to the 18 original NOADs responses. The total annual cost for the 18 NOADs under harmonized labor cost is 316,513.60 euros. In order to estimate the total annual cost for all NOADs, we have to take into account that 13 NOADs did not respond, namely the NOADs of Denmark, Czech Republic, Hungary, Romania, Slovakia, Estonia, Latvia, Poland, Austria, France, Ireland, Netherlands, and UK. Cost extrapolation yields a cost of roughly 500,000.00 euros.

OpenAIRE's annual operation cost was estimated to be 1,100,306.57 euros, including the cost of NOADs, the regional coordinators' cost, the total financial cost, the management cost, the cost for marketing and sales, as well as the technical cost, excluding CERN as the outlier. In addition, the new functionalities estimated to cost 151,203.32 based on R&D expenses. So, the total annual OpenAIRE's operation cost with the new functionalities is 1,251,509.89. Tables 10 and 11 in the online appendix present the estimation results.

3.6.2 Cost analysis

The cost analysis is based on the different costs derived from the cost questionnaire analysis (18 NOADs, regional coordinators, financial, management, marketing sales and dissemination, and technical cost) and two additional variables that define the outcome, labeled as annual rate of growth of cost item and discount rate. The first one indicates the annual expected rate of change of the cost items, in real terms, during the time horizon of the analysis. It could be positive or negative. The discount rate reflects, on the first level of analysis, the real cost of capital for the OpenAIRE initiative. On the level of social cost benefit analysis it will reflect the social discount rate. In the sensitivity analysis we examine a deterministic scenario for the change in the cost items, while in the Monte Carlo simulations we examine fully stochastic scenarios. We examine three different cost scenarios over a 15-year period. A sensitivity analysis is also performed on the deterministic model.

In the purely deterministic model the cost items evolve under the same baseline growth rate of -1 % and a discount rate of 2 % during the next 14-year period. The Net Present Value (NPV) is estimated for all 15 periods separately and once overall for the expected total cost of the OpenAIRE initiative. The expected total OpenAIRE cost, under these assumptions, is 13,308,699.85 euros.

In the sensitivity analysis, we define "extreme" values, a minimum value (-3 %, 1 %) and maximum value (3 %, 5 %), for the growth and discount rates, respectively. By performing sensitivity analysis we basically obtain the partial derivatives of the total cost of OpenAIRE for given growth and discount rates. That way, we can observe the

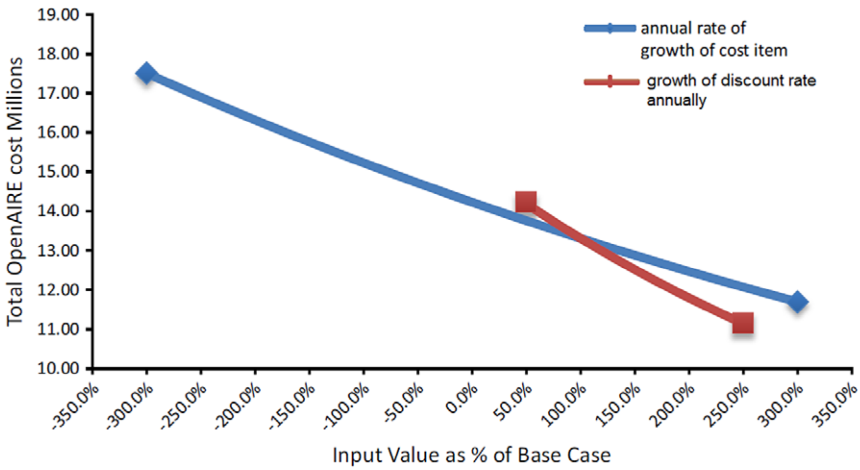


Figure 4 Sensitivity analysis for the purely deterministic model.

"extreme" values of the total cost for ranges of the growth and discount variables as shown in Figure 4.

The baseline case is defined 100 %, with the "extreme" values being the edge points in the graph. The graph can be interpreted in the following way. An annual increase of the growth rate to the "extreme" of 3 % relative to the baseline case will increase the present value of costs from 13,308,699.85 to 17,504,929.09 euros. The discount rate has a similar interpretation. Notice that the slope of the discount rate sensitivity line is steeper than the corresponding growth rate sensitivity line. This means that the present value of costs is more sensitive to discount rate changes relative to changes in the cost item growth rates.

In the second scenario, we deal with a stochastic model where all periods operate under the same growth rate and a discount rate of 2 %. However, this time the growth rate is randomly distributed with a mean 0 and standard deviation 2. The discount rate remains steady. We performed a Monte Carlo simulation on the NPV of the total OpenAIRE cost based on these parameters. The mean present value of costs is 14,373,565.13 with a standard deviation of 1,977,090.47. Assuming normality for the distribution of the present value of costs, the 95 % confidence interval for the present value of costs is (10,498,467.81,18,248, 662.45).

In the third scenario, we deal with a stochastic model where each of the 15 periods operates a different growth rate, but the same discount rate of 2 %. The growth rate is randomly distributed with a mean 0 and standard deviation 2. The discount rate remains steady. The Monte Carlo results a mean present value of costs 14,411,699.33 with a standard deviation of 1,045,867.97. Assuming normality for the distribution of the present value of costs, the 95 % confidence interval for the present value of costs is (12,361,798.11, 16,461,600.55).

4. Financial sustainability

Following the EU "Guide to Cost Benefit Analysis of Investment Projects," a project is sustainable when it does not incur the risk of running out of cash in the future. The crucial issue here is the timing of cash proceeds and payments. Sustainability occurs if the net flow of the cumulative cash flow generated is positive for all the years considered.

For the OpenAIRE initiative, we have an estimate of the cash flow of costs, but a scheme for revenue generation is not operating. Thus we approach the sustainability issue by determining a minimum amount of annual cash inflows that will secure the financial sustainability of the project. In particular, we consider a scheme where the costs that occur once every 5 years, which refer mainly to infrastructure, are covered by EU support. These costs are not substantial, they are approximately 84,000 euros and they occur every 5 years. The annual costs, which are mainly operation and maintenance, should be covered by annual subscription by the institutions participating in the initiative.

Alternatively the once-every – five-years costs can be covered by borrowing from the capital markets, and then the annual interest payments will be covered by annual subscriptions. In any case, the amount corresponding to the once-every – five-years costs is not substantial relative to the annual costs.

In the scenario for the deterministic case, the cost items are decreasing by 1 % per year and the real discount rate is 2 %. It is clear that with financing of 85,000, 80,000, and 76,000 euros in years 1, 6, and 11, respectively and annual subscriptions of 1,100,000 euros the project is financially sustainable, since the cumulative cash flow is positive for the whole 15-year period.

In the worst-case scenario for the project within the context of the sensitivity analysis assumptions, costs items could increase as much as 3 % per year and the discount rate would be 5 %. With financing of 85,000, 80,000 and 76,000 euros in years 1, 6, and 11, respectively and annual subscriptions of 1,350,000 euros the project is financially sustainable under this worst-case scenario since the cumulative cash flow is also positive for the whole 15-year period.

For the stochastic case where the rate of change of all the cost items is subject to stochastic shocks distributed normally with zero mean and standard deviation equal to 2 and running Monte Carlo simulation for 1000 times, the project can be regarded as financially sustainable with aggregate annual subscriptions with the range of 1,350,000 from all users. Specifically, financing of 85,000, 80,000, and

76,000 euros in years 1, 6, and 11, respectively and annual subscriptions of 1,350,000 euros the 95 % confidence interval for the cumulative cash flow in years 5, 10, and 15 are (712,576.03, 1,621,677.77), (344,700.43, 4,236,892.03), (−489,443.54, 8,477,2016.16), respectively.

On the other hand, when the financial sustainability for the stochastic case where the rate of change of each of the cost items is subject to stochastic shocks distributed normally and independently of the other items with zero mean and standard deviation equal to 2, project is also financially sustainable with aggregate annual subscriptions from all users with the range of 1,350,000. Again running the Monte Carlo simulation 1000 times, financing of 85,000, 80,000, and 76,000 euros in years 1, 6, and 11, respectively and annual subscriptions of 1,350,000 euros the 95 % confidence intervals for the cumulative cash flow in years 5, 10, and 15 results in (1,178,886.65, 1,640,940.89), (1,687,829.99, 3,803,255.53), (1,409,286.82, 6,478,080.42), respectively.

5. Cost benefit analysis

The general approach is to estimate the present value of costs and benefits for the 15-year period and then estimate the benefit costs (B/C) ratio. A B/C ratio greater than one indicates that for each euro of costs used to finance the OpenAIRE alternative, the corresponding benefit is more than one euro. Thus the project is beneficial and desirable if the B/C ratio is greater than one.

As a first stage of the analysis, we estimate B/C ratios for the deterministic model of the first scenario and we perform the corresponding sensitivity analysis. Second, by using the mean WTP and the corresponding standard deviation for each alternative, we perform Monte Carlo simulations by assuming that the WTP in each alternative is distributed normally. Monte Carlo simulations are performed for both the second and the third scenarios and the corresponding mean B/C ratios are estimated as the ratio of the mean present value of benefits to the mean present value of costs (Tables 9 and 10).

All B/C ratios are greater than 4 (Table 11), suggesting that the benefits of the project are sufficient to render the project highly acceptable on a cost–benefit basis. Sensitivity analysis is performed for the alternatives corresponding to the minimum and the maximum B/C ratios. In all the runs of the sensitivity analysis the B/C ratio was above 3.4, confirming the acceptability of the project at this level of analysis.

The cost benefit analysis indicates that the benefit–cost ratio in terms of present value of benefits and costs is substantial with benefits reflecting the WTP of the users of the OpenAIRE services. Furthermore, the ratio of benefits to the annual subscriptions required for the financial sustainability of the project is also around four and above.

Table 10 Monte Carlo simulation for the stochastic model with the same growth rate.

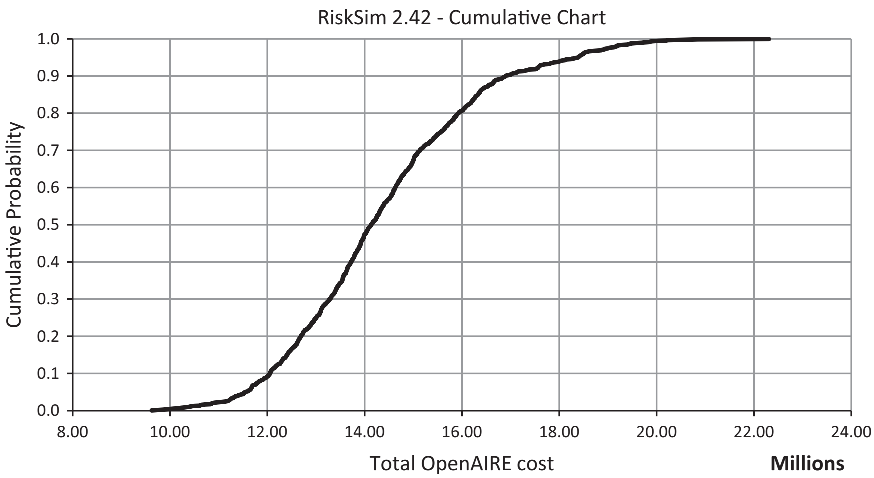


Table 11 Mean B/C ratios.

	B/C ratio		
	Deterministic model	**Second scenario**	**Third scenario**
Total Alt 1	4.59	4.26	4.24
Total Alt 2	5.24	4.87	4.85
Total Alt 3	5.12	4.73	4.71
Total Alt 4	4.49	4.17	4.12
Total Alt 5	5.77	5.33	5.35
Total Alt 6	5.14	4.75	4.74
Total Alt 7	5.02	4.62	4.62
Total Alt 8	5.67	5.25	5.27

6. Long run effects and knowledge spillovers

The close relationship between economic growth and knowledge goes back to Arrow's (1962) learning by doing models, where the production of a new good creates knowledge that could be used for the successful production of the next generation of goods. In this context, knowledge is a non-rival public good (e.g., Romer, 1986, 1990). R&D based growth models that developed in the 1990's consider economic growth as driven by R&D in the advanced developed world (e.g., Grossman & Helpman, 1991; Aghion & Howitt, 1992; Jones, 1995). More recently, Lucas (2009) and Lucas and Moll (2014) link growth with deliberate actions of individuals to allocate a certain part of their time in production-related knowledge creation. Knowledge created in this way is "rival" and has private good characteristics in the short run, but it is "non-rival" with public good characteristics in the in the long run.

We consider the OpenAIRE initiative as a route which, through the facilitation of diffusion of exiting knowledge to researchers, is a way to make the knowledge creating effort of individuals more productive and to generate a larger stock of knowledge with public good characteristics. This knowledge is non-rival knowledge which facilitates the creation of further knowledge.

Although the concepts are quite clear, the quantification of the impact of the OpenAIRE initiative on knowledge creation and eventually on growth, with the purpose of accounting for these benefits in the cost–benefit analysis is a very complicated task. This is because of modeling complexities and information requirements. Nevertheless, it is useful as a first approximation to provide an example of a possible approach in quantifying such benefits. These benefits will represent the OpenAire long run benefits, due to knowledge creation effects, on the benefit–cost ratios related to the project.

In developing our example, we follow Jone's (1995) R&D based growth model. The per GDP per capita steady state growth rate in the economy can be defined in terms of the growth rate in the stock of ideas or knowledge, g_A, as

$$g_A = \frac{\lambda n}{1-\phi}, \tag{12}$$

where λ is the elasticity of the growth in knowledge with respect to labor dedicated to R&D of new ideas, n is the rate of growth of population and $\phi > 0$ indicates that the productivity of research increases with the stock of ideas that have already been discovered. Along a balanced growth path the rate of growth of knowledge A is determined as

$$\frac{\dot{A}}{A} = \delta \frac{L^{\lambda}}{A^{1-\phi}}, \tag{13}$$

where δ is a positive parameter and L_A is the labor force in R&D. The OA initiative by facilitating the diffusion of knowledge and improving the efficiency in the exchange and the development of new ideas is expected to have a positive effect on the growth rate of knowledge accumulation. It is reasonable to assume that this effect will be realized through an increase in ϕ. The impact from an increase in ϕ on the growth rate of the economy is given by

$$\frac{dg_A}{d\phi} = \frac{\lambda n}{(1-\phi)^2} > 0. \tag{14}$$

As an example of the importance of this impact, we try to calibrate the effect from an increase in ϕ through the OpenAIRE initiative on a steady state balanced growth path, using the EU economy as reference.

Assume that the long run average annual growth rate in the EU on a balanced growth path will be 1 %, which is a rather conservative estimate. Use $n = 0.1$ % as the average population growth rate, and assume $\lambda = 1$.[1] Then the first equation implies that $\phi = 0.9$. Assume now that the OpenAIRE initiative increases ϕ by the very small amount of 0.005 %. This implies that the average annual growth rate of per capita GDP on the balanced growth path will increase from 1 % to 1.00045 %.

Using the value of 25,700 for per capita GDP (PPP) in the EU for 2013, the above result implies that the gain in per capita GDP along the balanced growth path will be given by

$$\Delta(GDP \text{ per capita}) = 25,700\left[e^{.0100045t} - e^{.01t}\right]. \tag{15}$$

This reflects the spillover effects of the OpenAIRE initiative in terms of facilitating the accumulation of knowledge and new ideas. This value, projected on a part of the EU population, which is currently at 505.7 million, represents a considerable flow of benefits. We continue the example by incorporating the knowledge spillover benefits in the cost benefit analysis of the OpenAIRE initiative considering very long time horizons of 50, 75, and 100 years.

The following assumptions are made when we extend the time horizon.

The WTP remains constant at the average of the eight alternatives, but the users increase with an average rate of 5 % per 10 years. In the Monte Carlo simulation, this rate is subject to an additive stochastic shock, which is distributed normally with mean zero and standard deviation 0.005.

Costs remain constant during the first 15-year period in line with the cost questionnaire and are reduced 1 % per 15 years.

The rate of growth of per capita income, which was estimated through the calibration at 1.00045, is subject to an additive stochastic shock, which is distributed normally with zero mean and standard deviation 0.000002.

The spillover benefits accrue to approximately 5 % of the EU population. This value is subject to an additive stochastic shock, which is distributed normally with zero mean and standard deviation 0.0025.

A declining discount rate was used with discount factors (Groom *et al.*, 2007; Gollier *et al.*, 2008; Hepbum *et al.*, 2009; Koundouri *et al.*, 2009).

The simulation results for the benefit–cost ratio for the 50, 75, and 100-year time horizon resulted a very high B/C mean and the 95 % confidence intervals do not

[1] This means that labor as an input in the production of new knowledge exhibits constant returns.

extend to negative values (see Tables 12, 13, and 14 in the online appendix). The inclusion of knowledge spillover-benefits in this example makes the project highly valuable from a social point of view.

7. Results

The results suggest that the average OpenAIRE stakeholder is WTP 1763.13 €/institution/year for the basic services provided by OpenAIRE.

We calculated the WTP for the additional potential services in each attribute category. The average OpenAIRE stakeholder is WTP only for higher interoperability (publication-dataset resolver) 252.01 €/institution/year and only for better access to scientific results (supplemental material) and compliance to OA mandates 203.06 €/institution/year. The average OpenAIRE stakeholder is not WTP for higher miscellaneous services.

This study details stakeholders' WTP for different combinations (scenarios) of OpenAIRE services provision. In general, OpenAIRE stakeholders prefer to have more interoperability, access to scientific results, and compliance to OA mandates. The upper bound of average WTP between considered scenarios for more interoperability, more access to scientific results, and compliance to OA mandates and basic level of miscellaneous services is 2218.21 €/institution/year. The lower bound of average WTP between considered scenarios for basic level of interoperability and access to scientific results, and higher level of miscellaneous services is 1724.26 €/institution/year.

The cost questionnaire allowed detail calculation of the cost of the OpenAIRE coordination platform that used in the cost benefit analysis. The total annual OpenAIRE's operation cost with the new functionalities is 1,251,509.89 euros (1,100,306.57 euros without the new functionalities). Including the estimated benefits of OpenAIRE, the main result of the cost benefit analysis shows that the discounted cost of OpenAIRE's 15-year operation is small compared to the corresponding benefits, considering. This result stays robust under the different scenarios.

The annual aggregate subscription for attaining financial sustainability for the OpenAIRE initiative is around 1,350,000 euros. This implies a per institution annual fee of 675 € with a participation of 2000 users. The annual fee per institution required for financial sustainability of OpenAIRE is lower than the lower bound of annual WTP per institutional stakeholder. The latter means that OpenAIRE is finanacially sustainable and social welfare increasing. Specifically, net social benefits are estimated at: 5,724,000 € for the provision of OpenAIRE basic services, and range between 7,222,500 € and 5,562,000 € for different combination of OpenAIRE services.

The simulated potential R&D effect from the existence of OpenAIRE suggests even larger net social welfare benefits in the long run. Using The EU economy as a basis, we provide an example which suggests that the knowledge spillover-benefits of the OpenAIRE initiative make the project highly valuable from a social point of view. We have estimated Benefit/Cost ratios of 71.82 for 50 years, 95.75 for 75 years, and 115.58 for 100 years. Finally, risk analysis supports the robustness of the study's results under different assumptions on future costs and benefits.

8. Discussion

We have estimated that the required cost for the OpenAIRE's 15-year operation is very small with respect to the corresponding benefits. In terms of long-term benefits, we consider the OpenAIRE initiative as a route which, through the facilitation of diffusion of existing knowledge to researchers, makes the knowledge creating effort of individuals more productive and helps to generate a larger stock of knowledge with public good characteristics. This has important positive effects on the long-run balanced growth rate of the economy.

We recommend the implementation of institutional fee via budget reallocations from institutional services that are substitutable from OpenAIRE services and a fee discrimination between stakeholders according to stakeholder specific WTP, which allows higher revenues for OpenAIRE. In case of implementing a subscription strategy, an average institutional fee of 675 €/institution/year is proposed for sustainable operation assuming 2000 institutional subscribers.

Given the estimated net social benefits in the short and medium run, as well as the huge long-term benefits for R&D, European or state subsidies could be provided for OpenAIRE subscription for countries or institutions, respectively, which cannot afford the minimum fee.

Finally, it is important to invest in making OpenAIRE less complicated and more user-friendly. We propose the increase of public engagement through awareness campaigns in order to attract more users and capitalize on spillover effects, protecting simultaneously the OA nature of OpenAIRE.

Supplementary Materials

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/bca.2020.26.

References

Arrow, K. J. 1962. "The Economic Implications of Learning by Doing." *Review of Economic Studies*, 29: 155–173.

Aghion, P. and P. Howitt. 1992. "A Model of Growth Through Creative Destruction." *Econometrica*, 60: 323–351.

Bateman, I. J., R. T. Carson, B. Day, W. M. Hanemann, N. Hanley, T. Hett, and S. Swanson. 2003. *Guidelines for the Use of Stated Preference Techniques for the Valuation of Preferences for Non-Market Goods*. Cheltenham, UK: Edward Elgar Publishing.

Bateman, I., R. T. Carson, B. Day, M. Hanemann, N. Hanley, T. Hett, M. Jones-Lee, G. Loomes, S. Mourato, E. Ozdemiroglu, D. W. Pearce, R. Sugden, and J. Swanson. 2002. *Economic Valuation with Stated Preference Techniques: A Manual*. Cheltenham, UK: Edward Elgar Publishing.

Birol, E. and P. Koundouri (Eds.). 2008. *Choice Experiments Informing Environmental Policy: A European Perspective*. Cheltenham, UK: Edward Elgar Publishing.

Cavaleri, P., M. Keren, G. B. Ramello, and V. Valli. 2009. "Publishing an E-Journal on a Shoe String: Is It a Sustainable Project?." *Economic Analysis and Policy*, 39(1): 89.

Conley, J. P. and M. Wooders. 2009. "But What Have You Done for Me Lately? Commercial Publishing, Scholarly Communication, and Open-Access." *Economic Analysis and Policy*, 39(1): 49.

Hausman, J., P. Ruud. 1987. "Specifying and Testing Econometric Models for Rank-Ordered Data." *Journal of Econometrics*, 34: 83–103.

Hepburn, C., P. Koundouri, K. Panopoulou, and T. Pantelides. 2009. "Social Discounting Under Uncertainty: A Cross Country Comparison." *Journal of Environmental Economics and Management*, 57(2): 140–150.

Houghton, J. and P. Sheehan. 2009. "Estimating the Potential Impacts of Open Access to Research Findings." *Economic Analysis and Policy*, 39(1): 127–142.

Houghton, J. W. and P. J. Sheehan. 2006. The Economic Impact of Enhanced Access to Research Findings, CSES Working Paper No.23, Victoria University, Melbourne.

Getz, M. 2005. Open Scholarship and Research Universities. Department of Economics Working Paper No. 0517, Vanderbilt University, Nashville.

Gollier, C., P. Koundouri, and T. Pantelides. 2008. "Decreasing Discount Rates: Economic Justifications and Implications for Long-Run Policy." *Economic Policy*, 56: 758–795.

Groom, B., P. Koundouri, K. Panopoulou, and T. Pantelides. 2007. "Discounting the Distant Future: How Much Does Model Selection Affect the Certainty Equivalent Rate?" *Journal of Applied Econometrics*, 22: 641–656.

Grossmal, G. and E. Helpman. 1991. *Innovation and Growth in the Global Economy*. Cambridge, MA: MIT Press.

Jones, C. E. 1995. "R & D-Based Models of Economic." *Journal of Political Economy*, 103(4): 759–784.

Kircz, J. 2005. Institutional Repositories, A New Platform in Higher Education and Research. Discussion Paper Commissioned by SURF, KRA Publishing, Amsterdam.

Koundouri, P. 2009. "Introduction to the Special Issue on 'Discounting the Long-Run Future and Sustainable Development'." *Economics (e-Journal)*, Vol.3, July 16. http://www.economics-ejournal.org/special-areas/special-issues/discounting-the-long-run-future-and-sustainable-development.

Koundouri, P., E. Kougea, M. Stithou, P. Ala-Aho, R. Eskelinen, T. P. Karjalainen, and P. M. Rossi. 2012. “The Value of Scientific Information on Climate Change: A Choice Experiment on Rokua Esker, Finland.” *Journal of Environmental Economics and Policy*, 1(1): 85–102.

Lancaster, K. J. 1966. “A New Approach to Consumer Theory.” *The Journal of Political Economy*, 74(2): 132–157.

Lucas Jr, R. E. 2009. “Ideas and Growth.” *Economica*, 76: 1–19.

Luccas Jr., R. E. and B. Moll. 2014. “Knowledge Growth and the Allocation of Time.” *Journal of Political Economy*, 122: 1–51.

Luce, R. D. 1959. Individual Choice Behavior. Chap. 1: 1–37.

McFadden, D. 1973. “Conditional Logit Analysis of Qualitative Choice Behavior.” In Zarembka, P. (Ed.) *Frontiers in Econometrics*. New York: Academic Press.

McCabe, M. J., C. M. Snyder, and A. Fagin. 2013. “Open Access versus Traditional Journal Pricing: Using a Simple ‘Platform Market’ Model to Understand Which Will Win (and Which Should).” *The Journal of Academic Librarianship*, 39(1): 11–19.

Odlyzko, A. 2013. Open Access, Library and Publisher Competition, and the Evolution of General Commerce.

OECD. 2006. Recommendation of the Council Concerning Access to Research Data from Public Funding, C(2006)184, OECD/LEGAL/0347. Available at: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0347.

Romer, P. M. 1986. “Increasing Returns and Long-Run Growth.” *Journal of Political Economy*, 94, 1002–1037.

Romer, P. M. 1990. “Endogenous Technological Change.” *Journal of Political Economy*, 98: S71–S102.

Scarpa, R., S. Notaro, R. Raffelli, D. Pihlens, and J. Louviere. 2009. Exploring Scale Effects of Best/Worst Rank Ordered Choice Data to Estimate Visitors’ Benefits from Alpine Transhumance. In Proceedings of the International Choice Modeling Conference, Leeds, UK, March (Vol. 30).

Willinsky, J. 2009. “The Stratified Economics of Open Access.” *Economic Analysis and Policy*, 39(1): 49.

Willinsky, J. 2011. The Enlightenment 2.0: An Open Letter on OpenAIRE. A Message Posted on SLAW.