

The Economic Impact of Open Science: A Scoping Review

Lena Tsipouri ^{1,2 *}, Sofia Liarti ², Silvia Vignetti ³, Izabella Martins Grapengiesser ⁴

1) National and Kapodistrian University of Athens, Greece

2) OPIX PC, Greece

3) CSIL, Italy

4) Technopolis Consulting Group, Belgium

Keywords: open science, open access, open data, open-source software, open methods, economic impact

Abstract

This paper summarised a comprehensive scoping review of the economic impact of Open Science (OS), examining empirical evidence from 2000 to 2023. It focuses on Open Access (OA), Open/FAIR Data (OFD), Open Source Software (OSS), and Open Methods, assessing their contributions to efficiency gains in research production, innovation enhancement, and economic growth. Evidence, although limited, indicates that OS accelerates research processes, reduces the related costs, fosters innovation by improving access to data and resources and this ultimately generates economic growth. Specific sectors, such as life sciences, are researched more and the literature exhibits substantial gains, mainly thanks to OFD and OA. OSS supports productivity, while the very limited studies on Open Methods indicate benefits in terms of productivity gains and innovation enhancement. However, gaps persist in the literature, particularly in fields like Citizen Science and Open Evaluation, for which no empirical findings on economic impact could be detected. Despite limitations, empirical evidence on specific cases highlight economic benefits. This review underscores the need for further metrics and studies across diverse sectors and regions to fully capture OS's economic potential.

1. Introduction

The aim of this paper is to review and summarise the literature that provides evidence for the economic impact of Open Science (OS) and identifying gaps where further research is needed. This review is a complement to others on academic (1) and societal (2) impacts of OS.

OS has a long history, and its definitions have evolved. In this paper we use the OECD definition “*Open Science is the practice of making scientific inputs, processes, and outputs available to all with minimal restrictions. This includes open access to publications, open research data, open software, open educational resources, and open collaboration among scientists.*”(3). This openness can be applied across the full research workflow, bringing greater transparency, accessibility and participation through practices like open protocols and methods, co-

*Author for correspondence (lena.tsipouri@opix.ai).

†Present address: OPIX PC, Aigialeias 52, Maroussi, 15125, Greece

creative and participatory approaches like Citizen Science, sharing of open data and code, transparent evaluation via Open Peer Review, and making scholarly publications Open Access (OA). The academic community pioneered OS and a large number of declarations¹ and funders' policies² have been adopted. These initiatives aimed to make research freely accessible to anyone with internet access, driving advances in all scientific fields.

Benefits of OS go beyond the scientific community (4). Since endogenous growth theory considers knowledge an essential driver of innovation, productivity enhancement, and long-run economic expansion (5) reducing the barriers to its access and stimulating wider dissemination was seen as essential to accelerate and facilitate this positive process. UNESCO describes OS as a "critical accelerator for the achievement of the United Nations Sustainable Development Goals and a true game changer in bridging the science, technology and innovation gaps and fulfilling the human right to science".(4)

The economic rationale for OS builds on the idea of knowledge as a public good³ (6,7). This is related to making its access and dissemination open. The economic justification for making publicly funded research results openly available stems largely from Paul David (8–10). Suber (11) further contended that taxpayers should not be charged twice for access to publicly funded research. This line of reasoning led to growing interest in assessing the economic impact of open versus closed science, prompting a wave of research on the benefits and challenges of mandating open availability of publicly funded scientific results.

Since then, a substantial body of theoretical studies and models has emerged, examining the benefits of OS through the lens of economics, aiming to demonstrate the benefits of opening up research results. Academic research papers on the economics of technological change make strong theoretical arguments in support of OS (12–14). Overall, the main economic rationale is an increase in efficiency in science production (3). By reducing the cost of reproducing, collecting and processing data, OS is then thought to foster research productivity (1).

Some authors theorise the economic value of OS using applications and examples from their respective industries, like McManamay and Utz (15) for fisheries, Harding (16) for medicine, and Chan (17) on the economics of the transition to the OS Model. However, such papers provide no precise quantitative evidence in support of the argument that free-of-cost access to new knowledge accelerates technological progress. Arguments for the economic benefits of OS include the willingness of individual firms and government services to benefit from research results or cooperate with academia or other companies (18). In addition, expressing views or forming alliances for FAIR data (19) are considered as evidence of the economic value companies - innovative enterprises in particular, expect from OS.

In addition to these positive arguments for the economic impacts of OS, others have flagged potential economic downsides. As an example, Mueller-Langer and Andreoli-Versbach (20) used a theoretical model to find that while mandatory disclosure of research data is an essential feature for credible empirical work, it may come at a cost. They claimed that authors might invest less in data generation if they are not the full residual claimants of their data after the first journal publication and they might "strategically delay" the time of submission of papers to fully exploit their data in subsequent research. This means that the welfare effects of universal mandatory data disclosure are ambiguous and risk being welfare-reducing, unless

¹ Budapest Open Access Initiative Declaration in 2002 and the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities in 2003.

² The U.S. National Institutes of Health (NIH) was one of the first major funding agencies to mandate Open Access to research. Many more followed including the European Commission.

³ The public nature of knowledge is given by its inherent properties of being i) non-rivalry: knowledge can be shared indefinitely without being depleted since a person's use of knowledge does not reduce its availability for others and ii) non-excludable – once knowledge is made available, it is difficult (or costly) to prevent others from accessing or using it.

accompanied by appropriate incentives which deter strategic delays. No empirical evidence is however provided to support this argument.

Given this wealth of theoretical claims, assessment of the evidence supporting or refuting them is essential for orienting and finetuning future OS policy. Previous reviews have indicated that this evidence base is thin. Other reviews have considered the impact of OS as a whole (i.e., academic, societal and economic impacts (21)), or economic impact of specific components, i.e., Open Access (22), Open/FAIR data (23) and Citizen Science (24). One previous study, Fell's (25) "rapid review" targeted the economic impact of OS specifically. While it found some evidence that OS enabled reductions in costs (access, labour and transaction) or could enable new research, products and service, it also admits that evidence is only indicative, being "patchy and diverse". In this literature a prevalent theme is that, despite abundant theoretical arguments, the potential economic impacts of OS are still poorly documented. More evidence on how OS partnerships drive innovation and economic growth is urgently needed (26) to sustain growing support for OS and to strengthen the case against critics.

Creating such evidence is the aim of the PathOS project⁴, an EC-funded initiative to better understand and measure OS impacts and their causal mechanisms. Initiation of this work included a series of scoping reviews of current available evidence. We hereby present our scoping review of the economic impact of OS, complementing already published reviews on academic (1) and societal (2) impacts. This review was conducted to both update the available evidence and align the synthesis with an impact pathway logic. By structuring the findings around main OS dimensions and causal relationships between different impact aspects, this review provides a systematic and structured approach to understanding the economic implications of OS. This perspective enhances clarity on how various dimensions of OS contribute to economic outcomes, advancing and complementing existing literature.

Our main research question: *What evidence exists in the literature regarding the effect of OS on the economic impact of research?* This can be decomposed into secondary research questions:

- SRQ1: What types of positive or negative, direct or indirect impacts are observed?
- SRQ2: What kinds of mechanisms produce them?
- SRQ3: What specific enabling and/or inhibiting factors (drivers and barriers) are associated with these impacts?
- SRQ4: What knowledge gaps emerge from this analysis?

2. Methods

The study followed a systematic approach, based on PRISMA-ScR, to answer its research questions through four key stages: identifying relevant studies, selecting those that met eligibility criteria, extracting data into a data chart, and then summarising and reporting the findings. The study's protocol (27) was pre-registered on October 31, 2022, and updated with an addendum (28) on June 29, 2023, which provided additional details regarding the search process for grey literature and snowballing techniques. Both documents are available on the Open Science Framework (OSF) platform (<https://osf.io/>). Any deviations from the original plan are outlined in Supplement 1.

⁴ <https://pathos-project.eu/>

2.1. Identifying relevant studies

The Scopus and Web of Science (WoS) databases were used to identify relevant peer-reviewed literature published in English from January 1, 2000, up until November 8, 2022, when the results were collected. The search included broad terms related to impact (impact, effect, and outcome), and key terms specifically focused on economic impact, including economic impact, financial/monetary impact, cost-benefit analysis (CBA), benefit-cost analysis (BCA), input-output, return on investment, patenting, innovation, new products, efficiency gains, and savings⁵.

Table 1. Primary search strings for Web of Science (WoS) and Scopus

Web of Science	(TI= ("open scien*" OR "science 2.0" OR "open data" OR "FAIR data" OR "open access" OR ("open code" OR "open software" OR "open tool*") OR "open method*" OR "citizen science" OR "open peer review" OR "open metric*") OR AB= ("open scien*" OR "science 2.0" OR "open data" OR "FAIR data" OR ("open code" OR "open software" OR "open tool*") OR "open method*" OR "citizen science" OR "open peer review" OR "open metric*" OR "open access publ*" OR "open access paper*" OR "open access journal*" OR "open access book*")) AND TS= ((impact* OR effect* OR outcome*) AND (econom* OR financ* OR cost* OR mone* OR cba OR bca OR "input-output" OR "return on investment" OR "patent*" OR "innovation*" OR "product*" OR "efficiency gain*" OR "saving*"))
Scopus	TITLE-ABS ("open scien*" OR "science 2.0" OR "open data" OR "FAIR data" OR ("open access" W/1 publ* OR paper* OR journal* OR book*) OR ("open code" OR "open software" OR "open tool*") OR "open method*" OR "citizen science" OR "open peer review" OR "open metric*") OR TITLE ("open access") AND TITLE-ABS-KEY ((impact* OR effect* OR outcome*) AND (econom* OR financ* OR cost* OR mone* OR cba OR bca OR "input-output" OR "return on investment" OR patent* OR innovation* OR product* OR "efficiency gain*" OR saving*)) AND (PUBYEAR>1999) AND (LIMIT-TO (LANGUAGE, "English"))

During the second phase of the study, additional literature was identified through two approaches: a "snowball search", which analysed citations from previously reviewed or included studies, and a systematic search of "grey literature" which included non-peer-reviewed materials from organisation websites (e.g. EC, UNESCO, OECD, UKRI etc.). This phase also involved searching preprint servers (e.g., OSF) and conducting general web search using platforms like Google. Full documentation, including the code and data from these searches, is available in the shared dataset accompanying the study.

2.2. Selection of eligible studies

The selection process followed the PRISMA-ScR checklist, with specific inclusion criteria applied during both the title/abstract and full-text screening phases, as listed below:

⁵ The corresponding search terms in the automatic process were used with * as presented in Table 1 to capture all derivatives of the respective words.

- Studies must provide evidence of the economic impact of OS, which encompasses six OS types, namely Open Access, Open/FAIR Data, Open Methods, Open Code/Software, Citizen Science, and Open Evaluation.
- Conducted at either national or international levels.
- Published between January 1, 2000, and the search date.
- Written in English.
- Full-text availability.
- Study is either a research article, review article, conference paper, or other peer-reviewed output, or a grey literature study from a recognised stakeholder.
- Studies using any methodologies (quantitative, qualitative, mixed or others).

In addition, we specifically excluded the following:

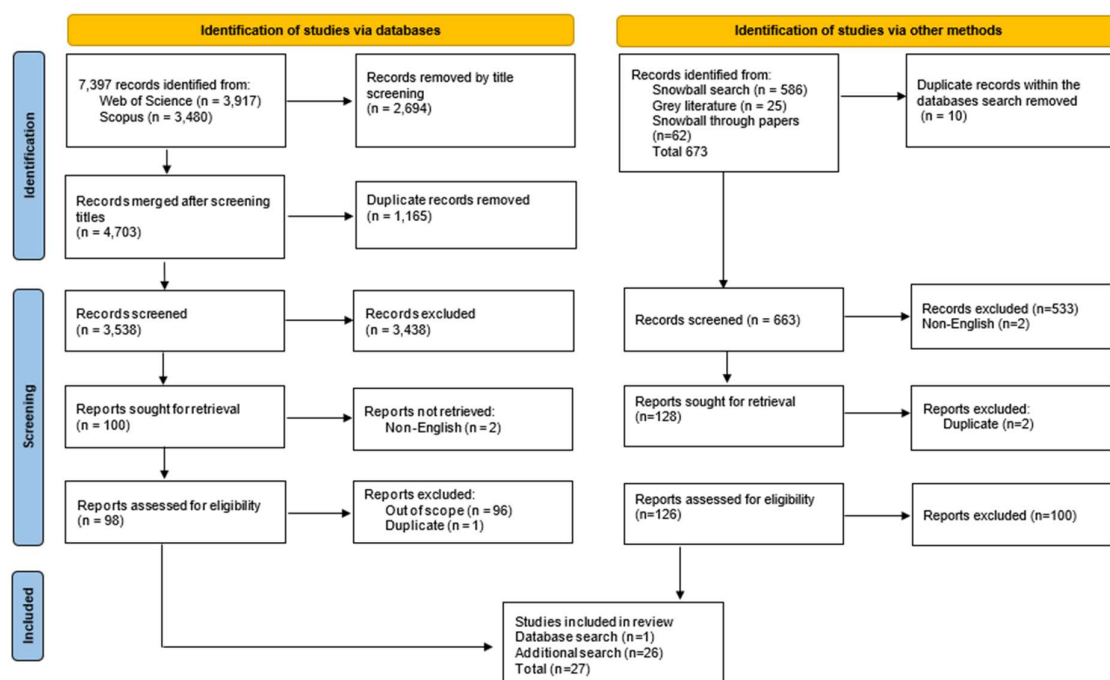
- *Concepts with Partial Similarities to Open Science*, which some authors relate to OS. Although some argue that corporate participation in crowdfunding (29), open innovation (30), and academic-industry collaboration (31,32) signal support for OS, these are broader concepts. Investigating how OS interacts with collaborations is an intriguing topic for future research but lies outside the scope of this review.
- *Open/FAIR Data from the Public Sector*, other than research results, is excluded from our analysis, as this paper focuses solely on academic research results. A particular challenge was to decide whether to include economic evidence from repositories mixing Open Data regularly generated by the public sector, which partly include research results (e.g. Copernicus). The decision was based on the observation of the limited share of academic research results, which was our main scope, within the repository.
- *Literature Reviews*: While not considered primary evidence, literature reviews played a critical role in guiding our snowballing technique, helping us identify relevant studies and prolific authors focused on the economic impacts of OS. When relevant, reviewed papers were included in our sample.
- *Indirect Economic Impacts stemming from Academic Benefits*: Topics such as the costs of publishing and Article Processing Charges are not explored.

The search in Web of Science (WoS) and Scopus yielded 7,397 results, which were then screened as per the PRISMA-ScR methodology. Initially, obvious false positives were eliminated by reviewing the titles. The remaining records were merged, and duplicates were deleted, yielding 3,538 entries. Following this, three researchers reviewed the titles and abstracts. Two researchers assessed each record, assigning a 'yes', 'no', or 'unsure' code to decide its inclusion. A third researcher assessed any data tagged as 'unsure' or with conflicting 'yes' and 'no' judgments and made the final decision on inclusion. During this process, the specific aspects of OS (e.g., Open Access, Open Methods etc.) of each study were also recorded.

The full texts of the initially selected studies were collected using various methods, such as library access, inter-library loans, and direct contact with authors. A total of 100 full texts were successfully retrieved and saved to a shared Zotero folder for reference management. Four researchers shared the studies and conducted a full-text screening, which led to the selection of one study for further analysis.

During the screening of snowball and grey literature, 128 studies were initially identified through title and abstract screening, which were narrowed down to 26, following full-text review. In total, across all search phases, 27 studies were included in the Scoping Review, as shown in the PRISMA-P chart (Figure 1).

Figure 1. PRISMA-P flow diagram



2.3. Extracting the data

The data extraction for the included studies was done collaboratively using an Excel file shared via Microsoft Teams. A pre-defined extraction form was used, covering study details, data sources, the study aims, types of impact, key findings, coverage, and confidence assessments (see Table 2). Each study was assigned to individual co-authors, responsible for extracting the relevant data.

Table 2. Categories extracted from included studies in the data charting process

Heading	Description
Author	Name of author/s
Date	Date article sourced
Title of study	Title of the article or study
Publication year	Year that the article was published
Publication type	Journal, website, conference, etc.
DOI/URL	Unique identifier
Exclusion	Out of scope, non-English, duplicate
Justification	If a study was deemed to be out of scope, justification had to be provided.
Study details and design (if applicable)	Type of study, empirical or review, etc. Notes on methods used in study (whether qualitative or quantitative, which population demographics studied, etc.)
Types of data sources included	Detail the data sources
Study aims	Overview of the main objectives of the study

Relevance to which aspect of Open Science	Open Access, Open/FAIR Data, Open Methods, Open Code/Software, Citizen Science, Open Evaluation
Relevance to which aspect(s) of economic impact	Efficiency gains, Enhancing Innovation, Economic growth
Key findings	Noteworthy results of the study that contribute to the scoping review question(s)
Coverage	Optional field to note any relevant information about the level of coverage of the study, e.g., only specific countries, disciplines, demographics covered
Confidence assessment	Optional field to note any concerns about reliability/generalisability of findings (e.g., conflict of interest, potential biases, small sample sizes, or other methodological issues) within the study

Economic impact categories were identified in an iterative way from the initial extraction through data screening and further analysis to better align with literature findings. This process led to the identification of three key categories: Efficiency gains, Enhancing Innovation, and Economic growth.

2.4. Limitations

Several limitations should be acknowledged. First, the review adhered to the PRISMA protocol, which comprises well-defined inclusion criteria. While this enhances the robustness of the approach, it may limit the inclusiveness of the exercise. For example, only studies published in English were considered, potentially excluding relevant research in other languages. In addition, grey literature was initially excluded and added only in a second phase based on a snowball strategy. Second, the scope of the review was limited to examining the economic impact of OS on users, such as researchers, companies, and policymakers. Consequently, the effects of OS on the publishing sector were not within the review's focus. While the publishing industry is an important stakeholder in the OS ecosystem, the decision to exclude its economic dynamics was made to ensure a clear and targeted analysis of user-centred impacts.

As a result, this scoping review did not aim to critically and systematically assess the robustness of the evidence or the methodological rigour of the included studies. As such, the findings presented should be interpreted as a comprehensive overview of the available evidence rather than a definitive assessment of its reliability or causal validity. Despite these limitations, this review contributes to the understanding of the economic implications of OS by providing a structured synthesis of existing research.

2.5. Summarising and reporting the results

Data extraction results from the initial database search, grey literature, and snowballed sources were compiled into a single Excel file, organised by OS aspects, and shared on Microsoft Teams. Co-authors were assigned to summarise and narratively report these results in a shared Microsoft Teams document, organised by OS and economic impact aspects, and then refine it into the final version.

3. Results

3.1 Overview

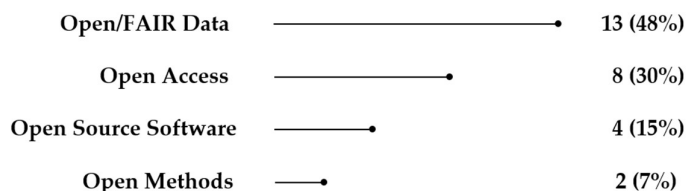
Our analysis draws on evidence from 27 studies, as illustrated in Figure 2. The economic impacts identified from the literature are classified according to two dimensions:

- The **Open Science dimension**: Empirical evidence was found for:
 - **Open/FAIR Data** (13 papers, 48% of OS type instances),
 - **Open Access** (eight papers, 30% of OS type instances),

- **Open Source Software** (four papers, 15% of OS type instances), and
- **Open Methods** (two papers, 7% of OS type instances),

while evidence is lacking for Open Evaluation and Citizen Science.

Figure 2: Number of studies reporting economic impact of Open Science by relevance to aspects of Open Science



- The **Economic Impact dimensions**, explained as follows:
 - **Efficiency gains** occur because time and costs are saved through accessing and reusing knowledge, data, software, or methods, rather than investing resources in recreating data or duplicating research.
 - **Enhancing innovation**, which as defined in the Oslo Manual, is “a new or improved product or process (or combination thereof) that differs significantly from the unit’s previous products or processes and that has been made available to potential users (product) or brought into use by the unit (process)” (33). In case OS reduces the cost of access and hence facilitates the use of knowledge for the creation of new product and processes creation, openness contributes to innovation.
 - **Economic growth** occurs when total factor productivity (or multifactor⁶) increases. It is measured by growth models based on the Cobb-Douglas production function (34–36).

3.2 Economic impact of Open/FAIR Data

Open data refers to research data that is freely accessible, usable, and shareable by anyone. The FAIR principles—Findable, Accessible, Interoperable, and Reusable—ensure that data is well-structured, properly documented, and can be effectively used by both humans and machines while respecting legal and ethical constraints. Our review identified 13 studies that assessed the economic impact of Open Data (13,37–48). Among these, eight studies reported impacts on efficiency gains, four focused on enhancing innovation, and three highlighted impacts on economic growth. Two of reported paper reveal impacts on both efficiency gains and economic growth.

⁶ “Multifactor productivity (MFP) reflects the overall efficiency with which labour and capital inputs are used together in the production process. Changes in MFP reflect the effects of changes in management practices, brand names, organisational change, general knowledge, network effects, spillovers from production factors, adjustment costs, economies of scale, the effects of imperfect competition and measurement errors. Growth in MFP is measured as a residual, i.e. that part of GDP growth that cannot be explained by changes in labour and capital inputs. In simple terms therefore, if labour and capital inputs remained unchanged between two periods, any changes in output would reflect changes in MFP. This indicator is measured as an index and in annual growth rates.” <https://www.oecd.org/en/data/indicators/multifactor-productivity.html?oecdcontrol-00b22b2429-var3=2021>

3.2.1 Efficiency gains

Although speed is often emphasised as a critical factor, there is a scarcity of studies providing robust estimations of actual time savings and their impact. Yozwiak et al. (13) argued that sharing data accelerated the development of therapies, diagnostic tools, and vaccines in both the public and private sectors during outbreaks, making the cases of Ebola epidemic in 2015 and H1N1 influenza in 2009. On contrary, a lack of or uncertainty in data-sharing practices during the Middle East respiratory syndrome (MERS) in 2012 slowed progress in addressing the challenges. However, the article did not provide quantitative evidence on the time saved and the related economic impact.

Available estimates of time savings rely on individual case studies. However, comparison is complicated by use of different metrics and a range of methods, even within individual cases. A number of studies were performed on UK open data services by Beagrie and Houghton (37–42) following the same methodology. For example, based on survey results and a mixed-method approach combining qualitative and quantitative analysis, Beagrie and Houghton (39) assessed the value and impact of the British Atmospheric Data Centre (BADC). The study found that the research efficiency reported by the users is high, but they estimated different values for it. A minimum value can be estimated by looking at the access costs defined as the time-related costs incurred by users to retrieve the data. Although these costs represent an expenditure, they can be interpreted as an implicit measure of the users' willingness to pay. The underlying rationale is that if users are willing to incur these costs, it suggests that the efficiency benefits they will derive from accessing the data are at least equivalent to this value, thus establishing a lower bound. Access costs of BADC are estimated at £ 2.3 million per year. However, when users are asked to provide an explicit value, their responses often provide a different perspective, also depending on the survey design and method. Here a distinction must be drawn between "willingness to pay" (WTP) to acquire something and "willingness to accept" (WTA) compensation for forgoing it. The primary difference lies in that WTP is limited by an individual's ability to pay, often determined by their disposable income, while WTA is not subject to such constraints. The choice between one approach or the other leads to different results. In the case of BADC, for example, the willingness to pay solicited through a contingent valuation (a survey-based economic technique for the valuation of non-market resources) shows a value of around £ 5.2 million. However, when asked about the willingness to accept, i.e. when capacity to pay is not considered, the value users place on BADC data and services rises to around £ 15 million per year. Finally, an upper bound of the efficiency gain was obtained by estimated time saving. Authors found that BADC data significantly improves research, teaching, and studying efficiency, with estimated potential impacts ranging between £ 10 and 58 million per annum. Adjusted estimates also suggest efficiency savings closer to £ 10 million annually, still a substantial benefit of - double the operational and access costs. All these estimates provide distinct perspectives on the value of efficiency gains by, on the one side, enriching the overall understanding of their magnitude but, on the other, posing challenges to comparability and validity of results.

In a similar study on the Economic and Social Data Service (ESDS) of the UK Economic and Social Research Council (ESRC), Beagrie and Houghton (37) adopted the same approach, assessing economic impact at multiple levels. The aim of the ESDS is to encourage broader and informed use of data for social science research and education while ensuring the long-term preservation and accessibility of such data. Based on the result of two surveys, the authors found evidence on the use value calculated at £ 24 million per year. This value is quite close to the willingness to pay, calculated with a contingent valuation and resulting in a value of around £ 111 million per year. An upper bound of value to the users was instead estimated at around £ 111 million per year if considering the willingness to accept. The analysis adopted a counterfactual approach: approximately 70% of users surveyed thought that if ESDS had not existed they would not have been able to obtain the data in any other way; in addition, 58% reported that it was beyond their capacity to

(re)create the data themselves. The extent to which these benefits are accurately reflected in the contingent valuation remains uncertain.

In 2013, Beagrie and Houghton (38) assessed and measured the value and impact of the Archaeology Data Service (ADS). As in previous studies, they used a mixed-method approach combining qualitative and quantitative analysis. The assessment of ADS data and services reveals direct use value to the ADS user community to £ 1.4 million per annum, while users' willingness to pay indicates a value of around £ 1.1 million per year and willingness to accept £ 7.4 million annually (38).

Finally, Beagrie and Houghton (41) showed that the services of the European Bioinformatics Institute of the European Molecular Biology Laboratory (EMBL-EBI) offering data openly are widely used and highly valued by their user community. The time users spent accessing data and services was valued at £ 270 million for 2015, while the willingness to pay was estimated at £ 322 million. The study was then updated in Beagrie and Houghton (42), employing multiple approaches in an attempt to look at more immediate/direct effects but also to wider impacts. It built on the results of a user survey which received almost 5,000 responses. The use value is estimated at £ 5.5 billion per annum, a willingness to pay of £ 1.25 billion per annum and a willingness to accept equal to £ 68,600 per respondent, per annum. The study estimated the total efficiency savings from time spent using data from EMBL-EBI, as reported by users, to range between £ 2.6 billion and £ 11 billion per year.

The analysis of the global utilisation of the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (43) echoed previous works done by Beagrie and Houghton and calculated the productive time researchers gain, associated with using RCSB-PDB data and services at \$ 2.5 billion annually, measured by the time allowing researchers more freedom for other activities. The analysis built on website analytics (user numbers and levels of use) and on estimates made by Beagrie and Houghton (41).

Available evidence shows that efficiency gains associated with OFD do not come without costs. Data creation, curation and deposition are costly activities. Moreover, accessing data (e.g. searching, screening, downloading and interpreting data) also requires time. The screened studies consider these aspects and also provide estimates of costs. The interesting result is that, compared with the benefits, the net result is always positive, although with variation in terms of magnitude. Beagrie and Houghton (41) showed that stated benefits are around six times the annual operational cost of EMBL-EBI. By adding access costs borne by the users, the result remains positive. Estimated efficiency gains of ADS is about five times the combined operational, depositor, and user access time costs (38). Returns calculated on the BADC data and services equates a rate ranging from 4 to 12 time the costs (39).

3.2.2 Enhancing innovation

Four studies investigated how OFD can enhance innovation. Using data on the Human Genome Project and the private firm Celera⁷, a paper by Williams (44) estimated the impact of Celera's gene-level IP on

⁷ The Human Genome Project was a landmark international scientific research initiative aimed at mapping and understanding all the genes in the human genome (the "genome"). The project began in 1990 and was completed in 2004, ahead of its original schedule. It stands as one of the most significant achievements in modern biology and genomics. The sequencing made simultaneous use of two different approaches with the competition and collaboration between HGP, a publicly-funded international collaboration, and Celera Genomics, a private company. Celera and the HGP

subsequent scientific research and product development. Genes initially sequenced by Celera were subject to intellectual property (IP) restrictions for up to two years but were made publicly available once re-sequenced by the public Human Genome Project. By mapping the timing of gene sequencing and Celera's IP and linking them to scientific research and product development, the author found that the IP constraints reduced scientific research and related product development by approximately 20 to 30 percent. Though the analysis admittedly does not assess the welfare consequences, these findings indicate that Celera's temporary IP protections had a lasting negative impact on innovation, compared to a scenario where Celera's gene sequences had been in the public domain from the outset.

Using patents as a proxy for innovation, Arshad et al. (45) found, in a case study on the use of a specific molecule, that the open dissemination of the molecule significantly increased its adoption among a diverse and multidisciplinary research community. A comparative examination of patent landscapes for various drug candidates revealed that the public release of molecule data spurred greater innovation, as evidenced by a higher number of downstream patents associated with it. These findings challenge the assumption that open data drug discovery compromises commercial IP. Instead, it exemplifies how open data can drive economic benefits by reducing the costs of both drug discovery and commercialisation.

Giovani (46) provided a detailed examination of how OFD influences innovation and IP strategies within the pharmaceutical industry. The study, based on a survey of 13 pharmaceutical firms, investigated the extent to which OFD (including both government and research results) affects companies' research activities or strategic monitoring, but also its impact of OFD on IP strategies. The findings revealed that pharmaceutical firms report leveraging OFD to complement proprietary data for research purposes, implement licensing-in and licensing-out strategies, map partnerships and networks among industry players, and identify key expertise for recruitment. Giovani argued that using and contributing to the OFD movement may change the approach of pharmaceutical firms to IP and innovation.

An alternative path to assess the innovation enhancing power of OFD is to apply a foresight methodology. Nielsen et al. (47), using workshops with industry and experts, found consensus among experts on the fact that open data accelerates innovation and enhances efficiency, potentially lowering product development costs and enabling better legislative and environmental decision-making. The panel of experts also agreed that open data can improve transport systems, enhance journey planning, and reduce societal costs, but it demands strategic frameworks to balance its benefits with risks like job displacement and cybersecurity concerns.

3.2.3 Economic growth

Three studies assessed the wider impacts generated by OFD on the economy as a whole, all of them relying on economic model simulations. A notable example is the study by Tripp and Grueber (48) which attempted to quantify the direct and indirect economic benefit of the abovementioned Human Genome Project providing open data on mapped sequences. The study concluded that between 1988 and 2010, the project and associated research and industry activity generated an economic impact of \$ 796 billion, personal income exceeding \$ 244 billion and 3.8 million job-years of employment. Considering that the federal government invested \$ 3.8 billion and that in 2010 alone the genomics-enabled industry generated over \$ 3.7 billion in

competed to finish the genome first. However, they eventually collaborated by sharing data, with Celera using HGP's publicly available mapped sequences to improve its assembly process.

federal taxes, the return on the investment was very high. The results related to the HGP are often cited as a benchmark for demonstrating the significant economic returns on investment in science. Some critics address the assumptions underlying the economic models which may lead to overestimations or lack sensitivity to alternative scenarios⁸. There are also attribution problems about the extent to which the outcomes quantified can be directly linked to the HGP and not to parallel private and public investments in genomics and related fields.

Wider economic growth estimates were also attempted for smaller initiatives. Beagrie and Houghton (42), with a macro-economic model, estimated the return on investment in R&D made possible by EMBL-EBI data. The study distinguished between return on R&D facilitated or depending on EMBL-EBI managed data, the first equalling £ 2.2 billion annually and the second being £ 1.3 billion annually or £ 9 billion over 30 years in net present value. These figures demonstrate substantial increases from the previous 2015-16 impact study (which estimated £ 920 million annually, or £ 6.9 billion over 30 years).

The analysis of the contribution to the local economy and the global utilisation of the Protein Data Bank (43) estimated an impact of R&D expenditure of \$ 1.1 billion annually, noting this as a conservative estimation. Over a period of 30 years, the estimated impact is valued at \$ 8.0 billion.

In all these simulations, attribution problems remain when assessing impact far beyond the direct use of the data, since other confounding or facilitating factors may intervene along the causal chain of effects.

3.2.4 Summary of the economic impact of Open/FAIR Data

The evidence collected is limited but convincing, since it is based on solid methodological design and empirical analysis. Evidence is particularly compelling in the case of life science research and applications to the pharma industry. Efficiency gains rise as researchers save both time and financial resources in conducting their research activity. Financial savings occur by avoiding expenses such as subscription fees to access the data necessary for their research. Time savings occur, not only because the search process is streamlined, but because researchers can also directly use existing data. Reproducing such data would require substantial time investment and the development and maintenance of adequate skills. Time savings can accelerate the uptake of research results by industries, which can in turn produce more innovation.

Using surveys and interviews, researchers can document and measure efficiency gains and innovation enhancement of individual initiatives. Most of the relevant included studies (see APPENDIX) come from large research centres or repositories and demonstrate a significant variance in the indicators calculated. In all cases efficiency gains are high, except for the Archaeology Data Service (ADS). However, no generalisations or comparisons can be made given the limited number of cases and their peculiarity. In addition, caution should be paid to assessing wider impacts because they are based on economic model simulations and may raise attribution problems.

3.3 Economic impact of Open Access

Open Access is the provision of free access to peer-reviewed, scholarly and research information to all. Our review identified eight studies examining the economic impact of OA (49–56). Of these, four studies

⁸ For example, Wadman has addressed these concerns. <https://www.nature.com/articles/nature.2013.13187>

highlighted efficiency gains, two emphasised innovation enhancement, and two focused on economic growth impacts.

3.3.1 Efficiency gains

Efficiency gains are achieved as researchers save both time and costs associated with accessing knowledge. OA enables them to explore a larger reservoir of information at no cost, allowing them to efficiently identify and utilise the resources most relevant to their research activities, saving time and money.

Houghton et al. (49) examined the levels of access to and utilisation of research results by small and medium-sized enterprises (SMEs) in Denmark. Through an online survey and in-depth interviews with knowledge-based SMEs, the authors assessed the cost and value of specific innovations, the role research publications played in it, and to what extent barriers to access was influencing the innovation capacity of SMEs. They concluded that without access to academic publications, replicating the same outcomes would have taken an extra 2.2 years, leading to significant opportunity costs. This efficiency gain was not broken down between *open access* in our sense and *availability*; the latter being also considered as open access in a broader sense by industry, since the knowledge is not proprietary but can be accessed at a relatively small cost. However, their respondents extensively used OA journals and repositories (more than half of respondents used OA journal monthly or more regularly), hence the efficiency gains can be partly attributed to OA.

Acceleration in scientific production is measured by Coccia (50), with specific reference to COVID-19 research. According to the author, scientific output increased by about 1.2% each day in the period April 2020-June 2021. Among the key characteristics influencing this peculiar evolution of research production, the author lists OA to publications, which facilitated the rapid dissemination of findings and contributed to the acceleration of research production. This effect was claimed to be crucial in supporting discoveries in the research arena over a short period of time, such as the development of innovative drugs and new antiviral COVID-19 treatments. Comparisons with other research fields during the same period, such as lung cancer, show a velocity of 0.4 of daily growth of scientific production (50). However, the author did not reach a monetary quantification of this gain and could not even identify to what extent the increased productivity attributed to the severity of the pandemic or to the degree of openness.

The economic benefits of OA are not confined to the business community. Look and Marsh (51) explored the impact of time savings associated with OA in the public sector, analysing both direct and indirect benefits. Based on case studies of selected organisations, they found that OA through both Green and Gold routes⁹ led to £ 28.6 million in direct cost savings for the UK public sector (£ 26 million in access fees and £ 2.6 million in time savings). Indirect benefits (such as improved decision-making and policy analysis) were suggested, but they were not quantifiable within the scope of this study.

Houghton et al. (52) conducted a comparative analysis of three models of scholarly publishing, concluding that the benefits of OA publishing are likely to outweigh its costs. Their study did not only investigate the

⁹ There are two main routes to making research outputs openly accessible. One involves publishing articles or books via the OA route on a publisher's platform (often referred to as gold open access). The other involves archiving a version of the manuscript in an OA repository (often described as green open access). Content published via the gold OA route is accessible immediately on publication, while manuscripts deposited via the green OA route may, in many cases, be made accessible only once a self-archiving embargo period has elapsed. The terms for onward sharing and re-use of OA content will depend on the license under which it has been made available <https://www.springernature.com/gp/open-science/about/green-or-gold-routes-to-oa>

academic impact for researchers, but assessed also the impact for funding agencies, research institutions, the publishing industry, research libraries, and government agencies. While the transition to OA publishing involves initial costs and a delay before benefits are fully captured, the long-term gains are significant.

3.3.2 Enhancing innovation

The impact of OA on innovation is challenging to quantify, because it is the accumulation of various sources of knowledge (open, closed, internal) that contribute to innovation over time. Hence the impact can only be assessed either by big data or by surveys, interviews, and case studies. The former refers to *Patent data analysis*, which offers the advantage of extensive data coverage but is primarily relevant to industrial sectors with a high propensity for patenting. The latter provides a broader perspective but requires significant effort and is hard to translate into standardised indicators. Unfortunately, very few studies have been identified with robust quantitative evidence.

Bryan and Özcan (53) took advantage of the National Institutes of Health (NIH) mandating free online availability of funded biomedical research in 2008, which provided a unique opportunity to conduct counterfactual analysis on OA. A difference-in-differences estimation was performed to assess the propensity of patent citations for articles published before and after April 2008, with and without NIH funding. The authors employed a novel tool that measured extracted in-text citations rather than the commonly used front-page patent citations, which they believed to be more closely linked to the inventor's knowledge. Using all US patent applications filed between 2005 and 2015, the model concluded that, after 2008, patents cited NIH-funded research 12 to 27% more frequently than before. The result was controlled for non-funded research, funded research in journals unaffected by the mandate, and academic citations. The findings suggest that the open-access mandate led to increased patentability, one of the most frequently used proxies for innovation in economic literature.

Using a more recent dataset of 22 million patent families from 2010 to 2020, Jahn et al. (54) matched publicly available data sources on patents and scholarly publications for a similar analysis. Unlike Bryan and Özcan (53), who focused on patents from a single country, they examined global patent data from Google Patents alongside scholarly data from Unpaywall, Europe PMC, and arXiv. Their analysis showed that the share of OA among scientific non-patent references in patents was above the general trend of all non-patent references. Their findings provided more detailed insights into the impact of OA across countries and disciplines. By analysing non-patent literature citations, they found that patent families in the fields of Human Necessities, Chemistry & Metallurgy, and Physics had an above-average open-access share compared to the overall average. In contrast, the fields of Textiles and Fixed Construction exhibited lower-than-average open-access shares. Geographical patterns also varied significantly. The US and the UK ranked highest in the percentage of open-access patent families, followed by technologically advanced OECD members, China, Russia, and Taiwan.

3.3.3 Economic growth

Houghton and Sheehan (55) developed a variation of the Solow-Swan growth model¹⁰ to calculate aggregate efficiency gains and social returns on research and development (R&D). In their model, income is a function of technology change, capital and labour, all contributing to growth. The authors applied three assumptions:

¹⁰ The Solow-Swan model is a highly respected economic model that explains long-term economic growth by examining capital accumulation, labour or population growth, and technological progress. Developed independently by Robert Solow and Trevor Swan in the 1950s, the model primarily focuses on how these factors contribute to a nation's output or GDP over time.

i) the use of an ‘accessibility’ parameter as the proportion of the R&D knowledge stock that is accessible to those who could use it productively, ii) the use of an ‘efficiency’ parameter as the proportion of R&D spending that generates useful knowledge, iii) that the move to OA leads to a one-off increase in both accessibility and efficiency. Based on the model specification and these assumptions, the authors calculated the potential returns on R&D investment for alternative scenarios (separately for GERD and GovERD¹¹), to a one-off increase in accessibility and efficiency in the OECD countries. They simulated the results using shares of returns to R&D for 25%, 40%, 50%, 60% and 75% and examined percentage changes in accessibility and efficiency for 1%, 2%, 5% and 10%. They considered that the choice of percentages depends on each country’s characteristics and hence should be left to national experts to assess which ones were most suitable for their country. The study demonstrated that, whether applied across the board or to sector specific research findings (e.g. OA to publicly funded research), there were substantial benefits to be expected from expanding OA. Using Germany with a GERD at \$ 58.7 billion as a showcase and choosing for social returns to R&D of 50%, and a 5% increase in access and efficiency. Their model suggested that OA would generate an additional \$ 3 billion. With the same shares of social returns to R&D and increase in access and efficiency Japan with GERD of \$ 112.7 billion would yield a gain of \$ 5.8 billion.

In a later study for Australia, using the same model, Houghton and Sheehan (56) introduced the difference of openness/closedness in the model’s variables. They introduced again the concepts of ‘accessibility’ and ‘efficiency’ into the traditional Solow-Swan model and estimated how OA could increase returns on investment in R&D. This time they broke down the impact of a one-off increase in accessibility and effectiveness between Government Expenditure on R&D, Higher Education Expenditure on R&D and the research funding by the Australian Research Council. Increased accessibility could lead to faster research processes, reduced duplicative research, and a higher rate of commercialisation of findings. Their estimates pointed to potential gains from openness: assuming 20% rate of return of R&D a 5% increase in accessibility and efficiency in Australia one could expect a yield of recurring annual gains of AUD 165 million (€ 101 million) from government R&D expenditures, AUD 111 million (€ 68 million) from higher education R&D, and AUD 12 million (€ 7,348,080 from Australian Research Council-funded research).

3.3.4 Summary of the economic impact of Open Access

Evidence on economic impact of OA remains limited. We identified eight papers assessing it with different methods and models indicating evidence on efficiency gains through accelerating research and cost saving, as well innovation enhancement. OA seems to contribute to patenting knowledge more than non OA. At an aggregate level, a growth model helps calculate OA contribution to growth under different assumptions of efficiency growth thanks to increased accessibility.

3.4 Economic impact of Open-Source Software

Our review identified four studies assessing the economic impact of OSS (57–60). OSS is defined as “software that anyone can study, inspect, modify, and distribute freely under very limited restrictions such as attribution”[(61) in (62)]. Platforms such as GitHub¹², GitLab, and Bitbucket have emerged as essential infrastructures for facilitating collaboration. While numerous science-based OSS projects exist—Linux and

¹¹ Gross Expenditure on Research and Development (all sectors); Government Expenditure on Research and Development (a sub-total of GERD).

¹² GitHub is a developer platform to build, scale, and deliver secure software. <https://github.com/about>

Python being prominent examples—there is no clear means of distinguishing between software derived from academic research from OSS in general.

Science-driven software development prioritises knowledge, research, and openness, often at the expense of usability and long-term support. In contrast, commercially developed software emphasises profitability, market demands, and scalability, resulting in more polished but often less open solutions. There are no systematic statistics quantifying the distribution between science-based and commercially developed OSS. One of the rare estimates [(63) in (60)] suggests that academic institutions in the United States contribute nearly one-third of the total OSS output. However, the absence of strict criteria or taxonomies to differentiate between research-based and commercially developed software necessitated a highly selective approach in our choice of papers.

3.4.1 Efficiency gains

Nagle (57) examined the firm-level productivity effects of non-pecuniary OSS, finding that its impact depends on whether firms possess complementary capabilities, such as skilled information technology (IT) personnel and supportive infrastructure. Firms with these capabilities experience a positive and statistically significant productivity impact. A 1% increase in the use of non-pecuniary OSS results in an increase in value-added productivity of between 0.002% and 0.008%. While these productivity increments may appear modest at the firm level, the aggregated economic impact across an industry or sector can be substantial, given the widespread adoption and reliance on OSS in IT-driven industries.

Conti et al. (58) analysed how startups, which utilise GitHub, can secure funding and reduce operational costs. Their study identified key mechanisms through which efficiency gains are realised, particularly in cost savings, as firms reduce expenditures on software development, licensing fees, and time-to-market. To rigorously assess these impacts, the authors employed a difference-in-differences approach, leveraging firm-level data to compare outcomes before and after engagement with Open-Source Communities.

Chesbrough (59) conducted a survey to examine the costs associated with OSS for large corporations, particularly Fortune 500 firms engaged in software development projects. The survey required participants to estimate costs and value comparisons for a specific "focal project," referring to a concrete software development initiative selected by the respondent. Participants were asked to calculate the cost of OSS access, which was defined as the sum of installation and maintenance expenses. For the focal projects identified, "almost two-thirds of [the respondents] reported that the benefits exceeded the costs, and only one-fifth reported that the perceived costs exceeded the benefits" (59). Additionally, 46% of respondents believed that developing the code internally would have cost at least twice as much as using OSS. Respondents also highlighted benefits related to faster development times.

3.4.2 Enhancing innovation

Conti et al. (58) combined data from GitHub and Crunchbase to conduct a counterfactual analysis assessing the likelihood of funding for early-stage startups engaging with open-source communities (OSCs) on GitHub compared to peers that were not similarly engaged. They analysed 160,065 U.S. startups. Their findings indicated that engagement was associated with an increase of at least 36% in the likelihood of receiving funding, with the relationship being strongest for firms working on novel technologies. They concluded that by integrating OSS into their operations, firms can minimise operational costs while maximising productivity and innovation opportunities.

The opportunity costs of engagement with OSS, as identified in their study, were primarily related to appropriability. Startups operating in less competitive markets benefit more from OSS, as they experience fewer pressures to protect intellectual property (IP), whereas firms in highly competitive industries perceive this as a significant challenge. OSS played a crucial role in technological development, particularly for startups working on novel technologies such as artificial intelligence (AI) and blockchain.

A report by the European Commission et al. (60) examined how OSS supports innovation processes in Europe. Their results, based on surveys and case studies, demonstrated that OSS fosters innovation through three key mechanisms: enabling collaboration, reducing barriers to entry, and establishing de facto standards. These mechanisms are particularly impactful in sectors such as healthcare and information and communication technology (ICT), where the collaborative nature of OSS accelerates technological

advancements and facilitates the development of socially relevant products. The authors estimated that a 10% increase in GitHub contributions by EU Member States could lead to the creation of over 650 additional IT startups within the EU. By leveraging the collaborative and accessible nature of OSS, these startups further strengthen innovation ecosystems and contribute to economic dynamism.

3.4.3 Economic growth

Only one report was identified (60) – which provided an estimate of the OSS-related impact on GDP. The authors found that within the EU, OSS contributed an estimated €65 to €95 billion to GDP in 2018. Using econometric modelling, they projected that a 10% increase in OSS contributions could boost GDP by an additional 0.4% to 0.6% annually. Unlike the respondents in Chesbrough’s survey, the dataset underlying these projections was predominantly composed of startups and micro-enterprises, which were strongly represented in the study.

The report concluded that OSS not only contributes to GDP growth but also enhances labour productivity within the EU. National contributions to OSS strengthen domestic competitiveness, as reflected in metrics such as exports and trade in value-added.

3.4.4 Summary of the economic impact of Open-Source Software

Based on the synthesis of the limited number of papers examined for our review, OSS appears to have economic impacts at both microeconomic and macroeconomic levels.

At the microeconomic level, OSS contributes to efficiency gains by reducing costs and saving time. It also plays a crucial role in fostering innovation, particularly in start-ups and in technology-intensive sectors. At the macroeconomic level, the available evidence, though limited with one report for the EU market only, suggests that OSS enhances labour productivity by lowering software development costs, providing free and reusable tools, and enabling a more efficient allocation of resources eventually leading to GDP growth.

However, unlike the cases of OFD, OA above, and Open Methods discussed below, conclusions regarding the economic impact of OSS are constrained by the challenge of distinguishing the extent to which OSS is science-based.

3.5 Economic impact of Open Methods

Open Methods in Open Science (OS) refer to sharing the entire research process transparently. This means making research steps, data analysis, and workflows publicly available so others can understand, reproduce, and build upon the work. Researchers do not only offer free access to their publications and data but also reveal in detail how to replicate the methods they used for their research. Our review identified only two studies that assessed the economic impact of Open Methods (64,65), both of which suggested that open methods contribute to efficiency gains and innovation enhancement.

3.5.1 Efficiency gains

Lee (64) addressed the problem of failure to rapid innovation in drug discovery, reporting on the establishment of the Structural Genomics Consortium (SGC), a UK-registered charity operating as a precompetitive public–private research partnership. SGC’s pharmaceutical partners provide funding and collaborate on scientific research, without retaining IP rights or exclusive access to findings. Lee’s findings in a case study of an SGC project, collaborating with Glaxo Smith Klein, showed that the adopted OA enabled significant timesaving by accelerating drug discovery. By using structure-guided methods, researchers rapidly developed highly potent and selective protein inhibitors and shared them with the scientific community without restrictions. He reported the design of a small molecule in 11 months, and half a year for compounds to be used by the community. The registration of the first clinical trial took place 16 months after the first seminal publication. This case suggests that OA accelerates scientific research and generates pioneering drug programmes and clinical studies faster.

3.5.2 Enhancing Innovation

Open methods do not only enhance innovation through efficiency gains as indicated by Lee's (64) but also through sharing information on research methods. In a publication addressing innovation enhancement directly, Murray et al. (65) developed a theoretical model suggesting that openness lowers the cost of accessing the ideas of others and nurtures new and speculative research lines facilitating exploration that leads to innovation. They had the opportunity to test their model with a natural experiment on genetically engineered mice because two IP holders (out of four methods developed for four categories of genetically modified mice) agreed in 1998 and 1999 with NIH to significantly lower the cost of access to the method or to acquiring mice directly. Prior to the agreement developing new mice would require significant resources and cause delays of the order of 18 to 24 months. This impacted more than 50 mice that had been developed and disclosed in the scientific literature allowing the authors to estimate pre- and post-NIH agreement citation rates to the treated mouse-articles and the control group. Robustness tests were conducted and confirmed the absence of selection bias in the control group. Using annual citations and new authors in the field as impact indicators the paper concluded that after the NIH agreement the two categories which benefitted from lowering costs produced comparatively more citations and increased citations by new authors. The paper also concluded that lowering costs did not affect all stages from research to innovation equally and that greater openness of early-stage research leads to an increase in the diversity and the exploratory nature of follow-on innovation. The increased openness in sharing information about genetically engineered mice fosters innovation at various stages by encouraging researchers to explore new directions.

3.5.3 Summary of the economic impact of Open Methods

Although only two papers were identified that studied open methods, both suggest benefits. One emphasised efficiency gains and innovation driven by time savings, the other stressed that lowering costs impact innovation because not only they enhance academic research but do so by speeding up access to innovation and support exploring new directions.

4. Discussion

The literature is abundant with theoretical considerations regarding the economic value of OS. Conceptually, this value is founded on the premise that the creation of new knowledge ultimately drives economic growth, while openness fosters a positive feedback loop: the more knowledge is openly accessible, the greater its potential to facilitate the generation of further knowledge. The scarce empirical evidence seems to substantiate the claim that OS positively impacts the economy. This is evidenced through case studies employing interviews and surveys and the comparatively higher prevalence of OS references in patent literature. In more detail:

- **The measured impacts found are all but one positive**, suggesting OS contributes to growth. The sole exception regarding OFD in the case of archaeology (38) where the authors calculated a negative net economic value when monetising results.
- **Economic benefits are generated by efficiency gains, in terms of time and cost saved proxied by salaries and by contribution to innovation.** The latter is measured by patents and new products developments. **The most precise impacts have been identified in detailed studies by large repositories**, which offer free access to data and publications and have an interest in proving the value of their business model.
- **Macroeconomic benefits can be assessed through standard growth models** enriched by "accessibility and efficiency parameters". The difficulty here obviously lies in determining the necessary parameters to run the simulations and the attribution claim.

Many of these studies rely on depositors' and users' surveys and on individual cases to empirically estimate impacts. This methodology however shows several limitations which the authors themselves recognise. First, respondents to these surveys are not a representative sample of the data services users, but a self-

selected sub-group. One might reasonably expect that those taking the time to respond are likely to use and value data and services more than those that did not respond, which may suggest a risk of overestimation. Moreover, stated preferences need to be elicited following a strict and complex protocol, in order to be considered reliable and avoid a risk of hypothetical bias, especially in addressing the counterfactual problem that may require measuring the time to generate again the same data made openly available. When datasets or repositories are as peculiar as the ones analysed in the reported papers, it may be difficult if not impossible to identify a proper counterfactual (or different counterfactual scenarios may exist depending on the type and frequency of use) and therefore respondents may not be in a position to provide a reliable answer.

This indicative evidence is not ubiquitous and there are still important knowledge gaps: i) the number of studies investigating economic impacts of OS is limited and concentrated in specific organisations, industries or countries; ii) collecting and/or generating raw data is costly and only few organisations have been willing to invest in understanding the value of their openness strategy; iii) the methods used are very dependent on surveys and interviews, hence subject to biases; iv) some studies lack explicit methodological transparency, making it difficult to assess the validity and rigour of the implied estimations.

Analysing the frequency of various components within the literature reveals several noteworthy patterns:

- **In terms of disciplines**, biomedical research represents the most frequently studied domain, theoretically (16,66) and empirically as reported above. This can be attributed to a longstanding culture of data sharing, strong regulatory frameworks and global health relevance. Other sectors, which are reported in the literature, but with less empirical evidence, are agriculture (23), fisheries (15), AI tools and machine learning (67).
- **In terms of timing**, the literature which was found relevant for our purposes shows a slightly increasing trend. Since 2021 there are eight relevant papers, which constitute 30% of the total between 2006 and 2023.
- **In terms of numbers and prolificness of authors**, there are only two authors (Neil Beagrie and John Houghton) who have invested substantial effort in developing and applying methodologies both for OA and OFD; all the others have published one or two papers each in the specific topic. This reflects the highly relevant role the prolific authors play and the need for them to train their successors. Regardless of the reasons for the limited engagement of researchers in assessing the economic impact of OS, it is evident that further investigation is required to achieve a comprehensive understanding of the subject.
- **In terms of geographic distribution**, the majority of the evidence has been gathered from the USA and the UK. Many authors agree that international collaboration increases the likelihood of OA publication from the Global South and better integrating researchers from these countries into global innovative activities (56,68,69).

Our literature review points at important gaps and promising topics for research in OS. Prolific theorising has helped set the stage but falls short of proving economic impact on its own. To truly demonstrate the economic value of OS, much more evidence is needed from models or case studies to capture the full extent of its effects. Expanding the tested methods (surveys, interviews, and growth models) to unexplored areas—different sectors, countries, and data repositories—could significantly enrich our understanding of the economic impact of OS. Similarly, other relevant economic factors, such as employment, are essential components for thoroughly understanding economic impacts and are currently missing completely.

An additional significant gap in research is how OS and private industry can mutually benefit. The question of how companies can contribute to OS, and how OS can support business needs, remains largely underexplored. Finding effective ways to balance the need for proprietary knowledge at the micro level with the widely recognised (albeit mostly theoretical) macroeconomic benefits of OS is an urgent research priority. Case studies by both large corporations and tech-oriented SMEs could provide initial insights before broader quantification efforts are pursued.

Finally, developing new methods and tools to assess economic impacts is essential—particularly for areas like Citizen Science and Open Evaluation, where there is barely any evidence at present. Developing systematic assessments could enable cost-benefit analyses for capturing its impacts in the future.

5. Conclusion

Based on a relatively low number of papers identified, the review confirms that the OS paradigm shift holds transformative potential for economic development through efficiency gains, innovation enhancement, and economic growth. By enabling unrestricted access to research outputs, OS accelerates the pace of discovery and application, reducing redundancies and fostering collaboration across disciplines and sectors.

Efficiency gains are a hallmark of OS economic benefits. OFD eliminate barriers to data access, allowing researchers and industries to save time and financial resources. Case studies such as the Human Genome Project and other life science experiments demonstrate that open data repositories catalyse research, with downstream effects on innovation and economic growth. Similarly, OSS contributes to productivity by lowering costs and enabling businesses to optimize internal resources.

Innovation is another critical outcome of OS. Enhanced access to research outputs drives the development of new products and services, particularly in fields like biomedicine and IT. OSS, for example, fosters collaborative environments where startups and established firms can innovate more rapidly. Open Methods, though less studied, reveal promising potential to accelerate scientific discoveries, as demonstrated in precompetitive collaborations in drug development.

Economic growth, while harder to quantify, is evident through macroeconomic analyses. Studies highlight significant returns on investment in OS initiatives, such as OA publications and data repositories.

Despite these encouraging findings, the empirical evidence base remains limited and unevenly distributed across sectors and geographies. Most studies focus on high-resource settings, leaving gaps in understanding OS's impacts in developing economies where its benefits might be even greater. Additionally, methodologies often rely on self-reported data, introducing biases that challenge the reliability of results. Areas such as Citizen Science and Open Evaluation remain underexplored, emphasising the need for new tools and metrics to assess their impacts.

Moving forward, researchers and policymakers should encourage the development of standardised metrics to measure the economic impacts of OS. Expanding research to include diverse sectors and regions will enhance our understanding of OS's global potential.

Acknowledgments

The authors gratefully thank Ioanna Grypari, Istvan Karasz and Nicki Lisa Cole for valuable comments on earlier drafts of this article, Eva Kormann for guidance on the process of uploading data for this preprint, Thomas Klebel for conceptualisation, data curation, methodology, software, editing and review and Tony Ross-Hellauer for conceptualisation, methodology, editing and review. Any errors remain entirely the authors' own.

Funding Statement

This work was supported by the project PathOS, funded by the European Commission under the Horizon Europe programme (grant no. 101058728).

Data Accessibility

Datasets and additional materials supporting this article are published on Zenodo, doi: 10.5281/zenodo.14892767

Competing Interests

We have no competing interests.

Authors' Contributions

LT: Conceptualisation, investigation, project administration, supervision, validation, writing (original draft), writing (editing and review)

SL: Conceptualisation, data curation, formal analysis, investigation, validation, visualisation, writing (original draft), writing (editing and review)

SV: Conceptualisation, investigation, validation, writing (original draft), writing (editing and review)

IMG: Formal analysis, investigation, writing (original draft), writing (editing and review)

References

1. Klebel T, Traag V, Grypari I, Stoy L, Ross-Hellauer T. 2024 The academic impact of Open Science: a scoping review. SocArXiv. (DOI: 10.31235/osf.io/ptjub)
2. Cole NL, Kormann E, Klebel T, Apartis S, Ross-Hellauer T. 2024 The societal impact of Open Science: a scoping review, Royal Society Open Science 11240286. (DOI: 10.1098/rsos.240286)
3. OECD 2015 Making Open Science a Reality. OECD Science, Technology and Industry Policy Papers 25, OECD.
4. UNESCO 2021 UNESCO Recommendation on Open Science. Available from: <https://unesdoc.unesco.org/ark:/48223/pf0000379949>
5. Romer P. 1990 Endogenous Technological Change. *Journal of Political Economy*. 98(5), 71–102.
6. Arrow KJ. 1962 'Economic Welfare and the Allocation of Resources for Invention', in: *The Rate and Direction of Inventive Activity: Economic and Social Factors*. Princeton University Press.
7. Stiglitz, JE. 1999 'Knowledge as a Global Public Good', in: Inge Kaul, Isabelle Grunberg, and Marc Stern (eds), *Global Public Goods: International Cooperation in the 21st Century*. (DOI: 10.1093/0195130529.003.0015)
8. David P. 1998 Common Agency Contracting and the Emergence of "Open Science" Institutions. *The American Economic Review*, 88 (2), 15–21. Available from: <http://www.jstor.org/stable/116885>
9. David P. 2003 The Economic Logic of "Open Science" and the Balance between Private Property Rights and the Public Domain in Scientific Data and Information: A Primer. SIEPR Discussion Paper No. 02-30.
10. David P. 2004 Understanding the emergence of 'open science' institutions: functionalist economics in historical context. *Industrial and Corporate Change*, 13 (4), 571–589. (DOI: 10.1093/icc/dth023)
11. Suber P. 2003 The taxpayer argument for open access, SPARC Open Access Newsletter.
12. Chataway J, Parks S, Smith E. 2017 How Will Open Science Impact on University-Industry Collaboration? *Foresight and STI Governance*, 11 (2), 44–53.
13. Yozwiak NL, Schaffner SF, Sabeti PC. 2015 Data sharing: Make outbreak research open access. *Nature* 518, 477–479.
14. Mazzucato M. 2011 The entrepreneurial state. *Soundings*. 49, 131-142.
15. McManamay RA, Utz RM. Open-Access Databases as Unprecedented Resources and Drivers of Cultural Change in Fisheries Science. *Fisheries*. 39(9), 417-425.
16. Harding RJ. 2017 Open Science and Accelerating Discovery in Rare and Neglected Diseases, in: *Proceedings of the 21st International Conference on Electronic Publishing. Presented at the Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices*, Limassol, Cyprus, 1–5. (DOI: 10.3233/978-1-61499-769-6-1)
17. Chan GR. 2015 Cost Impact in Managing the Transition to an Open Access Model, in: *The Importance of Being Earnest. Presented at the Charleston Conference, Against the Grain*, 358–362. (DOI: 10.5703/1288284315595)
18. Williams RJ, Walker I, Takle AK. 2012 Collaborative approaches to anticancer drug discovery and development: a Cancer Research UK perspective. *Drug discovery today*, 17(5-6), 185–187. (DOI: 10.1016/j.drudis.2012.01.020)

19. van Vlijmen H, Mons A, Waalkens A, Franke W, Baak A, Ruiter G, Kirkpatrick C, da Silva Santos LOB, Meerman B, Jellema R, Arts D, Kersloot M, Knijnenburg S, Lusher S, Verbeek R, Neefs JM. 2020 The Need of Industry to Go FAIR. *Data Intelligence* 2, 276–284. (DOI: 10.1162/dint_a_00050)
20. Mueller-Langer F, Andreoli-Versbach P. 2018 Open Access to Research Data: Strategic Delay and the Ambiguous Welfare Effects of Mandatory Data Disclosure. *Information Economics and Policy* 42, 20-34, Max Planck Institute for Innovation & Competition Research Paper No. 14-09. (DOI: 10.2139/ssrn.2458362)
21. Tennant JP, Waldner F, Jacques DC, Masuzzo P, Collister LB, Hartgerink CHJ. 2016 The academic, economic and societal impacts of Open Access: an evidence-based review. *F1000Res* 5, 632. (DOI: 10.12688/f1000research.8460.1)
22. Arzberger P, Schroeder P, Beaulieu A, Bowker G, Casey K, Laaksonen L, Moorman D, Uhlir P, Wouters P. 2004 Promoting Access to Public Research Data for Scientific, Economic, and Social Development. *Data Science Journal*, 3, 135–152. (DOI: 10.2481/dsj.3.135)
23. Ali B, Dahlhaus P. 2022 The Role of FAIR Data towards Sustainable Agricultural Performance: A Systematic Literature Review. *Agriculture*. Feb 21;12(2):309. (DOI: 10.3390/agriculture12020309)
24. Wehn U, Gharesifard M, Ceccaroni L, Joyce H, Ajates R, Woods S. 2021 Impact assessment of citizen science: state of the art and guiding principles for a consolidated approach. *Sustainability Science*, 16(5):1683-1699. (DOI: 10.1007/s11625-021-00959-2)
25. Fell MJ. 2019 The Economic Impacts of Open Science: A Rapid Evidence Assessment. *Publications*, 7(3):46. (DOI: 10.3390/publications7030046)
26. Ali-Khan SE, Jean A, MacDonald E, Gold ER. 2018 Defining Success in Open Science. *MNI Open Research*, 2(2). (DOI: 10.12688/mniopenres.12780)
27. Ross-Hellauer T, Klebel T, Pitelis A. 2022 Protocol for Scoping Reviews of (1) Academic, (2) Societal and (3) Economic impacts of Open Science. (DOI: 10.17605/OSF.IO/GM57C)
28. Cole NL, Klebel T, Ross-Hellauer T. 2023 Addendum to Protocol for Scoping Reviews of (1) Academic, (2) Societal and (3) Economic impacts of Open Science: PROTOCOL FOR GREY LITERATURE SEARCH AND INDEXING OF SOURCES, AND FOR SCREENING AND DATA CHARTING IN THE SYRF PLATFORM. *Open Science Framework*. Available from: <https://osf.io/3b6xj>
29. Ikkatai Y, Ono E. 2018 Exploring Characteristics of Academic Crowdfunding in Japan. In: 2018 7th International Congress on Advanced Applied Informatics (IIAI-AAI). Yonago, Japan: IEEE, p. 458-461. Available from: <https://ieeexplore.ieee.org/document/8693257/>
30. Chesbrough H. 2003 *Open Innovation: The New Imperative for Creating and Profiting From Technology*. Harvard Business School Press.
31. Bikard M, Vakili K, Teodoridis F. 2019 When Collaboration Bridges Institutions: The Impact of University–Industry Collaboration on Academic Productivity. *Organisation Science*, 30(2), 426-445. (DOI: 10.1287/orsc.2018.1235)
32. Arnone S, Buoni L, Manieri A, Mathieu PP, Mazzitelli G, Spagnoli F. 2016 University, Industry, and Research Cooperation: The Lazio Pulse Initiative. Presented at the International Technology, Education and Development Conference, Seville, Spain, 3808–3814. (DOI: 10.21125/iceri.2016.1902)
33. OECD, Eurostat. 2018 *Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation*, 4th Edition. Available from: https://www.oecd.org/en/publications/oslo-manual-2018_9789264304604-en.html
34. Solow R. 1957 Technical Change and the Aggregate Production Function, *Review of Economics and Statistics*.
35. Abramovitz M. 1956 Resource and Output Trends in the United States Since 1870. *American Economic Review*.
36. Jorgenson D, Griliches Z. 1967 The Explanation of Productivity Change. *Review of Economic Studies*.
37. Beagrie N, Houghton J. 2012 *Economic Impact Evaluation of the Economic and Social Data Service*.
38. Beagrie N, Houghton J. 2013 *The Value and Impact of the Archaeology Data Service: A Study and Methods for Enhancing Sustainability*. Available from: https://repository.jisc.ac.uk/5509/1/ADSReport_final.pdf
39. Beagrie N, Houghton J. 2013 *The Value and Impact of the British Atmospheric Data Centre*. Available from: https://vuir.vu.edu.au/39092/1/2013_Beagrie%26Houghton_Value%26Impact_Brit_Atmospheric_Data_Centre.pdf
40. Beagrie N, Houghton J. 2014 *The Value and Impact of Data Sharing and Curation. A synthesis of three recent studies of UK research data centres*.
41. Beagrie N, Houghton J. 2016 *The Value and Impact of the European Bioinformatics Institute*. Available from: https://www.embl.org/documents/wp-content/uploads/2021/09/EMBL-EBI_Impact_report-2016-summary.pdf

42. Beagrie N, Houghton J. 2021 Data-driven discovery: The value and impact of EMBL-EBI managed data resources.
43. Sullivan KP, Brennan-Tonetta, Marxen LJ. 2017 Economic Impacts of the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank. (DOI: 10.2210/rcsb_pdb/pdb-econ-imp-2017)
44. Williams HL. 2013 Intellectual Property Rights and Innovation: Evidence from the Human Genome. *Journal of Political Economy*. 121, 1–27. (DOI: 10.1086/669706)
45. Arshad Z, Smith J, Roberts M, Lee WH, Davies B, Bure K, Hollander GA, Dopson S, Bountra C, Brindley D. 2016 Open Access Could Transform Drug Discovery: A Case Study of JQ1. *Expert Opinion on Drug Discovery*. 11(3), 321-332. (DOI: 10.1517/17460441.2016.1144587)
46. Giovani B. 2017 Open Data for Research and Strategic Monitoring in the Pharmaceutical and Biotech Industry. *Data Science Journal*, 16. (DOI: 10.5334/dsj-2017-018)
47. Nielsen AF, Michelmann J, Akac A, Palts K, Zilles A, Anagnostopoulou A, Langeland O. 2023 Using the future wheel methodology to assess the impact of open science in the transport sector. *Scientific Reports* 13. (DOI: 10.1038/s41598-023-33102-5)
48. Tripp S, Grueber M. 2011 The Economic Impacts of Human Genome Project.
49. Houghton J, Swan A, Brown S. 2011 Access to research and technical information in Denmark. Available from: <https://eprints.soton.ac.uk/272603/>
50. Coccia M. 2021 Evolution and structure of research fields driven by crises and environmental threats: the COVID-19 research. *Scientometrics*. 126(12), 9405-9429. (DOI: 10.1007/s11192-021-04172-x)
51. Look H, Marsh K. 2012 Benefits of Open Access to Scholarly Research Outputs to the Public Sector. A report for the Open Access Implementation Group. Available from: <https://wiki.lib.sun.ac.za/images/e/e3/Report-to-oauk-benefits-of-open-access-public-sector.pdf>
52. Houghton J, Rasmussen B, Sheehan P, Oppenheim C, Morris A, Creaser C, Greenwood H, Summers M, Gourlay A. 2009 Economic implications of alternative scholarly publishing models: Exploring the costs and benefits. JISC EI-ASPM Project. Available from: https://vuir.vu.edu.au/15222/1/EI-ASPM_Report.pdf
53. Bryan KA, Ozcan Y. 2021 The Impact of Open Access Mandates on Invention. *The Review of Economics and Statistics*, 103(5), 954-967. (DOI: 10.1162/rest_a_00926)
54. Jahn N, Klebel T, Pride D, Knoth P, Ross-Hellauer T. 2022 Quantifying the influence of Open Access on innovation and patents. *Open Research Europe*, 2(64). (DOI: 10.12688/openreseurope.14680.1)
55. Houghton J, Sheehan P. 2006 The Economic Impact of Enhanced Access to Research Findings. CSES Working paper No 23.
56. Houghton J, Sheehan P. 2009 Estimating the Potential Impacts of Open Access to Research Findings. *Economic Analysis and Policy*, 39(1), 127-142. (DOI: 10.1016/S0313-5926(09)50048-3)
57. Nagle F. 2018 Open Source Software and Firm Productivity. *Management Science*, 65(3), 1191-1215. (DOI: 10.1287/mnsc.2017.2977)
58. Conti A, Peukert C, Roche M. 2021 Beefing IT up for your Investor? Open Sourcing and Startup Funding: Evidence from GitHub. *SSRN Electron Journal*. (DOI: 10.2139/ssrn.3883936)
59. Chesbrough, H 2023 Measuring the Economic Value of Open Source: A Survey and a Preliminary Analysis. The Linux Foundation.
60. European Commission, Blind K, Böhm M, Grzegorzewska P, Katz A, Muto S, Schubert T. 2021 The impact of Open Source Software and Hardware on technological independence, competitiveness and innovation in the EU economy. Brussels. Available from: <https://data.europa.eu/doi/10.2759/430161>
61. Laurent A. 2004 Understanding Open Source and Free Software Licensing. Sebastopol: O'Reilly Media Inc.
62. Calderón JBS, Robbins CA, Korkmaz G, Kramer BL. 2022 Measuring the Cost of Open Source Software Innovation on GitHub.
63. Korkmaz G. 2020 Measuring the Cost and Impact of Open Source Software Innovation on GitHub. Presentation at Harvard Business School.

64. Lee WH. 2015 Open Access Target Validation Is a More Efficient Way to Accelerate Drug Discovery. *PLOS Biology*, 13(6), e1002164. (DOI: 10.1371/journal.pbio.1002164)
65. Murray F, Aghion P, Dewatripont M, Kolev J, Stern S. 2016 Of Mice and Academics: Examining the Effect of Openness on Innovation. *American Economic Journal: Economic Policy*, 8(1), 212-252. (DOI: 10.1257/pol.20140062)
66. Arefolov A, Adam L, Brown S, Budovskaya Y, Chen C, Das D, Farhy C, Ferguson R, Huang H, Kanigel K, Lu C, Polesskaya O, Staton T, Tajhya R, Whitley M, Wong JY, Zeng X, McCreary M. 2021 Implementation of the FAIR Data Principles for Exploratory Biomarker Data from Clinical Trials. *Data intelligence*, 3(4), 631-662. (DOI: 10.1162/dint_a_00106)
67. Chen SC, Chen YC, Chen WL. 2017 Data Mining Techniques vs. Policy Development: Evaluating Advanced Applied Technological Policies and Emerging Communication Technology, in: 2017 IEEE International Symposium on Multimedia (ISM). Presented at the 2017 IEEE International Symposium on Multimedia (ISM), IEEE, Taichung, 469-474. (DOI: 10.1109/ISM.2017.93)
68. Iyandemye J, Thomas MP. 2019 Low income countries have the highest percentages of open access publication: A systematic computational analysis of the biomedical literature. *PLOS ONE*, 14(7), e0220229. (DOI: 10.1371/journal.pone.0220229)
69. Boufarss M, Laakso M. 2023 Open access and international coauthorship: A longitudinal study of the United Arab Emirates research output. *Quantitative Science Studies*, 4(2), 372-393. (DOI: 10.1162/qss_a_00256)

APPENDIX

Indicators	ADS (Archaeology Data Service)	BADC (British Atmospheric Data Centre)	ESDS (Economic and Social Data Service)	EMBL-EBI (European Bioinformatics Institute of the European Molecular Biology Laboratory)	Protein Data Bank	HGP
Annual Investment Value	£ 1.2 mn	£ 2.8 mn	£ 23 mn	£ 110 mn	\$ 6.9 mn	
Annual Use Value	£ 1.4 mn	£ 2.3 mn	£ 24 mn	£ 5.5 bn	\$ 5.5 bn	
Cost of obtaining	£ 1,439,091	£ 2,257,170	£ 4 mn			
Cost of Provision	£ 698,000	£ 2,049,600	£ 3 mn			
Annual Users' Willingness to Pay	£ 1.1 mn	£ 5.2 mn	£ 25 mn	£ 1.25 bn	\$ 760 mn	
Annual Willingness to Accept	£ 7.4 mn	£ 15 mn	£ 81 - £111 mn	£ 68,600 per respondent	-	
Consumer Surplus	-£ 312,000	£ 3 mn	£ 21 mn	-	-	1988-2010 \$ 796 bn
Net Economic Value	-£ 1 mn	£ 935,000	£ 18 mn	-	-	
User Efficiency Gain (minimum per year)	£ 13 mn	£ 10 mn	£ 68 mn	£ 2.6 bn	\$ 2 bn	
User Efficiency Gain (maximum per year)	£ 58 mn	£ 58 mn	£ 112 mn	£ 11 bn	\$ 3 bn	
Return on Investment (Additional Use, NPV over 30 years)	£ 2.4 mn to £ 9.7 mn	£ 11 mn to £ 34 mn	£ 58 mn to £ 233 mn	£ 9 bn	\$ 8 bn	
Return on Investment (Additional Use)	2.1 to 8.3-fold	4 to 12-fold	2.5 to 10-fold	-	-	
Return on Investment (Non-Recreate, cannot be easily recreated by others)	£ 1.5mn - £ 5.9mn (1.3 to 5.1-fold RoI)	-	-	£ 6 bn per annum from reduced duplication of effort.	-	