technopolis group

January 2026

The Economic Benefits of Open Science

Final report (v3)

Version 3

January 2026

The Economic Benefits of Open Science

Final report (v3)

Cristina Rosemberg, Aphra Murray, Alexander Holmes, Shrishti Kajaria

Table of Contents

Tables

Figures

Executive Summary

Context: PLOS Redefining Publishing

This report is an output of PLOS's Redefining Publishing program, supported by the Gordon and Betty Moore Foundation and the Robert Wood Johnson Foundation. The program is exploring how scholarly communication can evolve to better support Open Science at scale, including moving beyond the article by recognizing a wider range of research contributions and identifying more sustainable financial models to support them. As part of this work, PLOS commissioned Technopolis to provide an independent study on the economic benefits of Open Science to clarify where economic value is created in an Open Science ecosystem, what conditions enable it, and where structural barriers and cost pressures remain.

Executive Summary

This report examines the economic implications of Open Science (OS) and what it would mean to move towards a research ecosystem in which research elements, including data, code, software, workflows, methods, and publications, are openly shared and valued. Because the OS ecosystem is vast and the ambition of this work is necessarily system-wide, the study focuses on the components most likely to yield actionable insight: (1) a rapid evidence assessment to bring together dispersed economic evidence into an actionable evidence base, and (2) a set of illustrative case studies to demonstrate how value is created in practice across different types of open outputs and infrastructures.

The study focused on three key areas:

- Understanding the impacts of migrating to an OS ecosystem in which all types of research outputs are valued.
- Exploring the costs of such a transition for the scholarly communication ecosystem.
- Evaluating the variety of economic impacts arising from an Open Science transition.

To ensure the findings reflect the breadth of Open Science beyond publications alone, the study includes five illustrative case studies spanning digital collections, computational workflows, open-source software, an open large language model, and open training resources. These were selected to represent different pathways to value, including efficiency, innovation, and human-capital benefits, and different patterns of cost and responsibility across the ecosystem (see Section 4 and Appendix C).

Headline findings

Open Science delivers significant economic benefits when it enables reuse at scale

Across the rapid evidence assessment and case studies, the analysis identifies consistent evidence that Open Science can deliver meaningful economic benefits when research outputs are designed and supported for reuse at scale, particularly for data, code, workflows, software, and training resources (see Sections 4.2 and 4.3; Appendix C). Quantified impacts are most robust where Open Science reduces duplication of effort, accelerates research processes, or enables downstream reuse across communities and sectors (see Section 4.2; Tables 3–4; Table 12).

At a macro level, the literature reviewed links Open Science practices to stronger long-run economic performance. The rapid evidence assessment reports macroeconomic modelling

evidence indicating that long-run GDP would be approximately 2% lower under more closed research systems, compared with more open conditions that embed Open Science practices (see Section 3; Appendix A). While this estimate is derived from system-level modelling rather than individual case studies, it reinforces the importance of Open Science beyond project-specific impacts.

These findings are consistent with a recent report from PathOS, while adding specificity on how value is realized in practice across different output types and where costs and responsibilities sit across the ecosystem.

Key takeaway: Open Science is not solely a transparency or access measure. When system conditions enable reuse, it functions as a mechanism for improving research productivity and strengthening long-term economic performance (see Sections 3, 4.2–4.3, and 5).

The most consistent and measurable benefits arise through efficiency gains

The evidence shows that efficiency gains are the most immediate and consistently quantified economic benefits of Open Science (see Section 4.2). Across multiple case studies, reuse of open outputs reduces duplication, shortens research timelines, and lowers transaction and travel costs, particularly in contexts involving digitized collections and shared computational workflows (see Section 4.2; Tables 3–4; Appendix C).

These efficiency gains are closely linked to the availability of shared infrastructure, common standards, and incentives that support reuse across institutional and disciplinary boundaries (see Sections 4.2 and 5.3).

Key takeaway: Efficiency-driven benefits depend less on individual openness decisions and more on coordinated investment in infrastructure and reuse-enabling conditions (see Section 5.3).

Open infrastructures enable innovation and downstream spillover value

The study finds that Open Science infrastructures can support innovation by enabling downstream reuse, adaptation, and extension of open resources, particularly open-source software, shared workflows, and open models (see Section 4.3; Appendix C). In several cases, estimated downstream value substantially exceeds initial development or training costs, indicating the potential for spillover benefits that extend beyond the original research context (see Section 4.3; Appendix C).

These innovation effects are strongest where adoption is widespread and where open resources are actively maintained and integrated into broader research ecosystems (see **Appendix B).**

Key takeaway: Open Science infrastructures can act as enabling platforms for innovation, but their economic value depends on sustained adoption and maintenance rather than one-off investment (see Section 4.3; Appendix B).

Network effects mean Open Science benefits can compound over time

A cross-cutting insight from the case studies is that many Open Science resources increase in value as more users adopt, contribute to, and reuse them. This dynamic is particularly evident in shared workflows, open-source software ecosystems, and open models, where broader participation enhances reuse and accelerates cumulative innovation (see Section 4.3; Appendix C; methodological discussion in Appendix B).

Key takeaway: The presence of network effects strengthens the case for sustained and coordinated investment in Open Science, rather than fragmented or short-term approaches (see Section 5.3).

Costs are unevenly distributed across stakeholders, creating barriers to scale

The report shows that the costs associated with Open Science are distributed unevenly across the ecosystem. Researchers bear time and skills costs, responsibility for long-term preservation is carried by libraries as well as other infrastructure providers and repository operators, and publishers invest in platforms and workflows to support a wider range of outputs (see Sections 5.1–5.3). In contrast, many of the benefits accrue broadly across society and across national boundaries, contributing to a mismatch between who pays and who benefits (see Section 5.3).

Key takeaway: Without coordination, this cost–benefit mismatch limits the scalability of Open Science. Funders are well positioned to address this through aligned incentives and shared infrastructure support (see Section 5).

The economic value of Open Science is systematically underestimated

The analysis shows clear evidence of significant economic benefit from Open Science where impacts can be quantified, particularly through reuse and downstream application of open outputs (see Sections 4.2–4.3; Appendix C). At the same time, it highlights gaps in how reuse, network effects, and longer-term spillovers are measured across the ecosystem. Practice-level monitoring efforts already exist, including PLOS's Open Science Indicators and international initiatives such as UNESCO's Open Science monitoring work, but the report finds that consistent mechanisms for capturing broader economic and societal value remain limited (see Section 4.1; Appendices A and B). This means that current assessments are likely to capture only part of the value Open Science creates, which in turn contributes to underinvestment in activities that generate the greatest long-term returns (see Sections 4.1 and 5.3).

Key takeaway: Strengthening measurement is critical to expanding and deepening assessment of Open Science value and to aligning policy, funding, and incentives with where economic value is created (see Sections 4.1 and 5.3).

Closing summary

Taken together, the evidence assessment and case studies show that the economic case for Open Science lies in its ability to enable reuse of research outputs at scale. When Open Science is implemented in ways that support reuse through appropriate infrastructure, incentives, and coordination, the sharing of data, code, software, workflows, methods, and publications can deliver measurable efficiency gains, support innovation, and strengthen long-term economic performance. The findings in this report provide a robust evidence base that supports the value of a transition toward Open Science and helps clarify where economic value is created and where barriers remain. The full report sets out the underlying methods and case study results in detail and provides the evidence underpinning each point in this Executive Summary, supporting deeper exploration of the analysis and sources.

As an output of PLOS's Redefining Publishing program, this analysis underscores the value of system-level change beyond the article and provides a strong foundation for the next phase of the program.

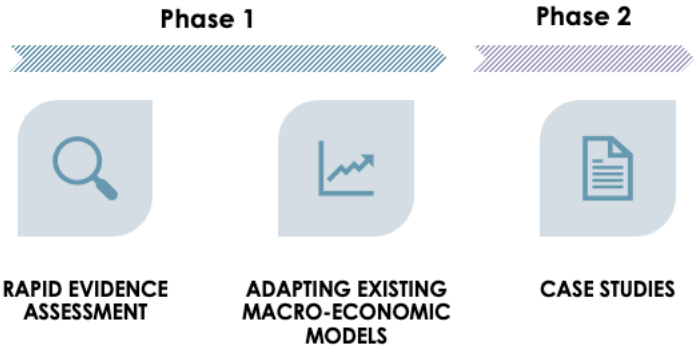
1 Introduction

1.1 This study

PLOS commissioned Technopolis to undertake a study to explore the **economic benefits of Open Science.** The study had the following objectives:

- Understanding the impacts of migrating to an OS ecosystem in which all types of research outputs are valued.
- Exploring the costs of such a transition for the scholarly communication ecosystem.
- Evaluating the variety of economic impacts arising from an Open Science transition.

We undertook the study between April and October 2025 across two mains phases. An overview of the study is presented in the figure below and then presented in more detail.



1.2 Phase 1 – scoping and initial analysis

The focus of the first phase of the study was to review pre-existing evidence on the economic benefits of Open Science. This involved a **Rapid Evidence Assessment (REA)**, to identify, compile and analyse recently published reports on approaches to assessing the economic benefits of Open Science outputs.[1]

The focus of the REA was to:

- Identify the different types of direct and indirect economic impact that Open Science has been demonstrated to have, informing the selection of case studies in the second phase of the study
- To identify methods and approaches to estimating the economic benefits of Open Science, with the aim of refining the approach that this study will take in assessing the economic benefits that might emerge from the migration to the new funding model.

Some of the results of the REA are presented in Section 4 of the report. These findings helped to refine the key research directions and overall approach for the remainder of the study, and also supported the identification of case studies.

[1] This study focused on reports that had been published between 8th November 2022 and April 2025. Additional detail on the approach to the REA is provided in Annex A.

1.3 Phase 2 – case studies

Phase 2 of the study extended the analysis of the economic benefits of Open Science. This phase involved the development of a series of impact case studies, informed by the outputs of the REA as well as comments and feedback from PLOS. The purpose of these case studies was to examine the benefits of Open Science in greater depth, providing concrete examples of some of the benefits identified through the REA. The case studies were selected to capture a variety of both Open Science outputs (e.g. research methods, software, data) and types of benefits.

The case studies focused on the following examples:

- **Distributed System of Scientific Collections (DiSSCo)** – a research infrastructure for natural science collections that digitally unifies all European natural science assets.
- **WorkflowHub** – a registry for describing, sharing, and publishing scientific computational workflows.
- **R** – a free and open-source software environment for statistical computing and graphics.
- **BLOOM** – a large language model trained in complete transparency, used to investigate the performance and behaviour of recently developed LLMs.
- **National Centre for Research Methods** – a UK centre providing training and resources across qualitative, quantitative, creative, visual, mixed, and multimodal methods.

1.4 This report

The remainder of this report is organised as follows:

- **Section 3, The Open Science Ecosystem** – this section presents an overview of the existing Open Science ecosystem, including the economic pressures in academic publishing and the growth of the Open Access movement
- **Section 4, Economic Impacts of Open Science** – this section presents the results of the REA and the case studies to provide a framework for understanding the types of economic benefits that emerge from Open Science
- **Section 5, Costs of Open Science** – this section provides an overview of the costs to key parts of the research ecosystem as the transition to Open Science increases, including libraries, scholarly publishers and researchers themselves
- **Section 6, Conclusions** – this section provides an overview of the evidence collected throughout this study

Separate **Appendices** contain the following supporting information:

- A – Rapid Evidence Assessment
- B - Approaches to measuring Open Science
- C – Case Studies

2 The Open Science Ecosystem

Open Science lacks a single, universally accepted definition, reflecting its nature as a broad and evolving movement. UNESCO provides a useful framework, **defining Open Science** as "an inclusive construct that combines various movements and practices aiming to make multilingual scientific knowledge openly available, accessible, and reusable for everyone, to increase scientific collaborations and sharing of information for the benefits of science and society, and to open the processes of scientific knowledge creation, evaluation, and communication to societal actors beyond the traditional scientific community."[2]

Contemporary understanding of Open Science has evolved beyond simply ensuring transparency of research outputs, such as sharing code and data. The movement is increasingly defined by its **capacity to achieve meaningful downstream objectives** including increased access to knowledge, enhanced collaboration across research communities, and improved scientific rigour through greater scrutiny and reproducibility.[3]

Open Science functions as an umbrella term encompassing multiple initiatives, each addressing different aspects of the research ecosystem. **Open access publishing** focuses on making research publications and articles freely available to the public, removing paywalls and subscription barriers that traditionally limit access to scientific knowledge. **Open source software initiatives** aim to dissolve costly barriers to computational tools and foster collaborative development communities that can advance research capabilities collectively.

Together, these complementary movements have coalesced into a broader transformation of how scientific research is conducted, evaluated, and disseminated. This shift represents a fundamental reimagining of the research process, moving away from closed, proprietary systems toward transparent, collaborative, and publicly accessible approaches to knowledge creation and sharing.

Open Science Publishing

The landscape of scholarly communication has undergone profound transformation over the past several decades, driven by escalating costs of subscriptions (for example to journals) and site licenses, technological advancement, commercial interest (as a blocker to OS advancement) and evolving expectations about research accessibility. Understanding this trajectory provides essential context for developing more comprehensive and sustainable approaches to research dissemination that maximise the societal value of investment (particularly public investment) in science.

2.1 Economic Pressures in Academic Publishing

For decades, the academic publishing landscape was dominated by a subscription-based model that created what scholars and librarians term the 'serials crisis' – a sustained period of dramatic price increases for scholarly journals that began in the 1980s and intensified through the 1990s and 2000s.[4] Under this traditional model, academic publishers consistently raised

[2] UNESCO, An introduction to the UNESCO Recommendation on Open Science (2022)

[3] Thibault R.T. et. al, Open Science 2.0: Towards a truly collaborative research ecosystem (2023)

[4] McGuigan, G. S. Publishing Perils in Academe: The Serials Crisis and the Economics of the Academic Journal Publishing Industry. (2004)

subscription prices at rates substantially exceeding inflation, creating unsustainable financial pressures for publicly funded institutions and their libraries.

The fundamental economics of scholarly journals create conditions that are ripe for exploitation. Unlike other markets, scholarly journals represent non-substitutable products with highly inelastic demand.[5] Academic libraries cannot purchase alternative journals to meet specific research needs, as each journal contains unique content essential for their collections. This monopolistic control over individual titles, combined with academic incentive systems that prioritise publication in high-impact journals, has enabled some publishers to implement sustained above-inflation price increases with limited competitive pressure.

Market concentration among a small number of large commercial publishers has further exacerbated these dynamics. While recent analysis suggests some diversification in overall journal ownership major publishers retain control over the most prestigious titles and have leveraged this position to implement bundled subscription packages (although this is not true of all fields or disciplines).[6] These multi-product, multi-year deals effectively prevent libraries from cancelling unwanted titles while systematically squeezing out smaller publishers and niche subjects. The result has been profit margins that significantly exceed industry norms – figures that are "unusually high" compared to firms in other industries. At the same time, the global expansion of research activity (with the researcher pool growing three times faster than the global population between 2014 and 2018) has required publishers to adapt their pricing models to reflect expanded usage and content volume.[7]

These economic pressures extend beyond immediate budgetary concerns to affect the fundamental infrastructure of scholarly communication. Rising costs create barriers to research access, potentially limiting scientific progress and widening gaps between well-funded and resource-constrained institutions. The crisis ultimately raises questions about the appropriate balance between commercial interests and the public benefit derived from scholarly research, setting the stage for alternative publishing models that would emerge in response.

2.2 The Open Access Alternative

In response to the challenges posed by traditional subscription-based publishing models, the Open Access (OA) movement has emerged as a significant alternative approach to scholarly communication. OA publishing eliminates reader paywalls, operating on the principle that research (and particularly publicly funded research) should be accessible to the public without financial barriers. This approach recognises that taxpayers support a large amount of the research enterprise and should therefore have unrestricted access to its outputs.

The OA movement has emerged as a response to three fundamental challenges in scholarly communication: accessibility, affordability, and equity.[8] By removing financial barriers to knowledge access, OA contributes to a more inclusive research environment that extends beyond well-resourced institutions to encompass researchers, practitioners, and citizens worldwide.

---

[5] Zarif, A., The Economics of Scientific Publishing (2023)

[6] van Bellen S, Alperin JP, Larivière V Scholarly publishing's hidden diversity: How exclusive databases sustain the oligopoly of academic publishers. (2025)

[7] UNESCO (2021) UNESCO Science Report: the Race Against Time for Smarter Development. S. Schneegans, T. Straza and J. Lewis (eds). UNESCO Publishing: Paris.

[8] The State of Open Access Today, KGL (2024)

OA publishing experienced substantial growth following its formal establishment in the early 2000s, with the movement explicitly committed to making scholarly research freely available to anyone with internet connectivity. The foundational principles of OA were articulated through several landmark declarations, including the Budapest Open Access Initiative (2002)[9], the Bethesda Statement on Open Access Publishing (2003)[10], and the Berlin Declaration (2003)[11]. These documents established the conceptual framework for unrestricted access to research outputs and provided guidance for implementation across different disciplinary contexts.

The financial sustainability of OA publishing has undergone considerable evolution since the movement's inception. Early OA journals typically operated through grant funding or direct institutional support to cover operational expenses. The next phase of the OA publishing development was in the early 2000s, as PLOS and BioMedCentral launched gold OA journals and established article process charge (APC) as a way of financing the publication of OA content. Under this approach, researchers or their affiliated institutions assume responsibility for publication costs, which is designed to shift the costs from readers to the author (or their funding source) and allow for content to be read free of charge.

The APC model presents a fundamental tension: while it enables free access to published research, it creates financial barriers for authors (and in particular) under-resourced authors and, more critically, incentivises journals to maximise publication volume.[12,13] Although OA articles still generate less revenue per paper than subscription articles, but that gap is likely to keep narrowing as publication fees increase.[14]

Despite these concerns, the research ecosystem continues to operate within a restrictive framework characterised by two reinforcing problems. First, the majority of research outputs remain inaccessible; either locked behind paywalls in the case of published articles, or simply not shared at all for non-article research outputs. Second, research assessment systems remain heavily dependent on quantitative metrics, particularly publication counts and Journal Impact Factors, within a hypercompetitive environment that rewards quantity and perceived prestige over actual research quality.[15]

These factors create a self-perpetuating cycle in which the research community feels constrained by existing systems yet remains resistant to fundamental change. The current model's focus on the publication of a final Version of Record as both the primary outcome of research and the key marker of scientific achievement exacerbates these systemic problems.

While OA publishers have made substantial investments in maintaining editorial standards and journal rigour, the underlying incentive structures that drive researcher behaviour and institutional assessment remain largely unchanged. This disconnect between publishing

[9] https://www.budapestopenaccessinitiative.org

[10] Statement on Open Access Publishing Archived 2012-03-11 at the Wayback Machine, by Patrick O. Brown, Diane Cabell, Aravinda Chakravarti, Barbara Cohen, Tony Delamothe, Michael Eisen, Les Grivell, Jean-Claude Guédon, R. Scott Hawley, Richard K. Johnson, Marc W. Kirschner, David Lipman, Arnold P. Lutzker, Elizabeth Marincola, Richard J. Roberts, Gerald M. Rubin, Robert Schloegl, Vivian Siegel, Anthony D. So, Peter Suber, Harold E. Varmus, Jan Velterop, Mark Walport, and Linda Watson. 20 June 2003

[11] https://openaccess.mpg.de/Berlin-Declaration

[12] Conflict between Open Access and Open Science: APCs are a key part of the problem, preprints are a key part of the solution (2020)

[13] Moving away from APCs: a multi-stakeholder working group convened by cOAlition S, Jisc, and PLOS, (2023)

[14] News & Views: Market Sizing Update 2025 – Has OA recovered its mojo? DeltaThink (2025)

[15] Sivertsen, G., et al, The Ongoing Reform of Research Assessment (2024)

innovation and systemic reform continues to limit the transformative potential of the Open Science movement.

2.3 Value of Open Science beyond the article

Contemporary research generates a wide array of valuable outputs beyond traditional articles, including research software, datasets, computational code, protocols, methodologies, and analytical tools. Additional valuable outputs include peer review reports, laboratory notebooks, negative results, replication studies and educational materials generated during the research process. These materials often represent substantial intellectual and technical contributions that enable other researchers to build upon existing work, verify findings, and accelerate scientific discovery. Research software, for instance, may embody years of development effort and represent novel algorithmic approaches that have applications across multiple disciplines. Similarly, carefully curated datasets can serve as foundational resources for entire research communities, enabling meta-analyses, comparative studies, and new theoretical insights.

The collaborative nature of modern research further amplifies the value of these diverse outputs. Open protocols enable standardisation across laboratories, reducing methodological inconsistencies and improving reproducibility. Shared analytical pipelines allow researchers to apply proven methods to new contexts, while open educational resources democratise access to specialised knowledge and techniques.

The temporal dimension of impact presents another challenge for traditional assessment approaches. While article citations may peak within a few years of publication, research software and datasets often demonstrate sustained or even increasing utility over longer timeframes, particularly as new analytical methods emerge or as research questions evolve.[16,17] This extended lifecycle of impact is poorly captured by conventional metrics.

The current emphasis on article-centric evaluation systems fails to recognise or reward these critical contributions, creating perverse incentives that discourage researchers from investing time and resources in producing high-quality, reusable research materials.[18] This system also fails to recognise the substantial effort required to properly document, maintain, and support these resources for community use; activities that are essential for maximising their scientific impact but receive limited institutional recognition. This misalignment between evaluation criteria and actual research practices undermines both scientific rigour and the efficient use of public research investments. Reform initiatives, such as the Coalition for Advancing Research Assessment (CoARA), are seeking to address this issue by advocating for evaluation frameworks that move beyond a narrow focus on articles and instead recognise a wider range of contributions, including datasets, software, and other reusable research outputs.[19]

Assessing the broader benefits of Open Science outputs is essential for understanding the full return on public investment in research. Traditional impact metrics focused solely on publication citations fail to capture the economic value generated when research software is adopted by industry, when open datasets inform policy decisions, or when computational tools

[16] Evidence for the importance of research software (2020)

[17] Aksnes, D. W., Characteristics of highly cited papers (2003)

[18] National Academies of Sciences, Engineering, and Medicine; Policy and Global Affairs; Board on Research Data and Information; Committee on Toward an Open Science Enterprise. Open Science by Design: Realizing a Vision for 21st Century Research. Washington (DC): National Academies Press (US); 2018 Jul 17. 2, Broadening Access to the Results of Scientific Research. Available from: https://www.ncbi.nlm.nih.gov/books/NBK525412/

[19] CoARA (The Coalition for Advancing Research Assessment)

enhance educational outcomes. These applications often generate network effects, where the value of a resource increases as more users adopt and contribute to it, creating compounding returns that extend far beyond the initial investment.

The COVID-19 pandemic provided compelling evidence of this broader value proposition, as openly shared viral genome sequences, epidemiological data, and analytical tools enabled rapid global research coordination and accelerated vaccine development in ways that would have been impossible under traditional closed-access models. [20]

These wider applications of research outputs often generate substantial social and economic benefits that justify continued public support for research activities. As such, this study seeks to contribute and complement the existing body of work on economic analyses of Open Science outputs.

[20] Philippe Buchy et. al., COVID-19 pandemic: lessons learned from more than a century of pandemics and current vaccine development for pandemic control (2021)

3 Economic Impacts of Open Science

3.1 Introduction

One of the main objectives of this study was to provide **evidence of the economic impacts of Open Science**. We have attempted to achieve this through two complementary approaches: conducting a **Rapid Evidence Assessment** (REA) of existing literature and developing a **series of case studies where we have attempted to provide routes to quantification of these benefits** using methodologies informed, in part, by the REA findings. This section presents a synthesis of both the types of economic impacts identified in the literature and the results from our case studies.

The REA revealed a diverse range of approaches and methodologies for assessing the economic impacts of Open Science. This methodological diversity likely reflects the challenges inherent in evaluating the growing variety of Open Science outputs (which extend far beyond traditional articles) and tracking their impacts and usage outside academic contexts, where bibliometric indicators can serve as proxies for impact.

A fundamental challenge in this field is the lack of formal indicators when users access Open Science outputs, unless there are explicit requirements to declare usage. Even where the use of Open Science outputs is clearly identifiable, linking the openness of these outputs to specific economic benefits remains methodologically challenging. Consequently, many attempts to quantify the economic benefits of Open Science rely on specialised methodologies tailored to specific output types, such as research data or software, rather than comprehensive cross-sectoral approaches. Moreover, many of these studies depend on large-scale primary data collection activities (e.g. surveys of users), which were not within the scope of this study.

These methodological constraints have important implications for how we understand and measure the economic value of Open Science. While individual studies may provide compelling evidence within their specific domains, the fragmented nature of current assessment approaches makes it difficult to develop a comprehensive picture of Open Science's overall economic impact and one which we have avoided in our own study.

For this study, we have grouped the types of economic benefits emerging from Open Science into three main categories. We recognise that these categories are not mutually exclusive and that overlaps are likely. The categorisation draws on insights from both our case studies and the literature review, and includes:

- **Efficiency:** These benefits arise from achieving the same research or innovation outputs with reduced inputs (primarily in the form of public research funding) or generating higher outputs for equivalent investment levels. Efficiency gains typically manifest through reduced duplication of effort, accelerated research timelines, and improved resource allocation across the research ecosystem.
- **Enhancing Innovation:** This category encompasses innovation outcomes that have led to economically impactful activities which would have been less likely to occur, or would have occurred more slowly, in a closed research environment. These benefits often emerge through increased collaboration, cross-disciplinary knowledge transfer, and the enabling of new research directions through improved access to foundational resources.
- **Societal and Indirect Economic Impacts:** These refer to long-term, often diffuse benefits that extend beyond the immediate research context to generate broader social and economic value. Such impacts may include improved policy decisions based on accessible

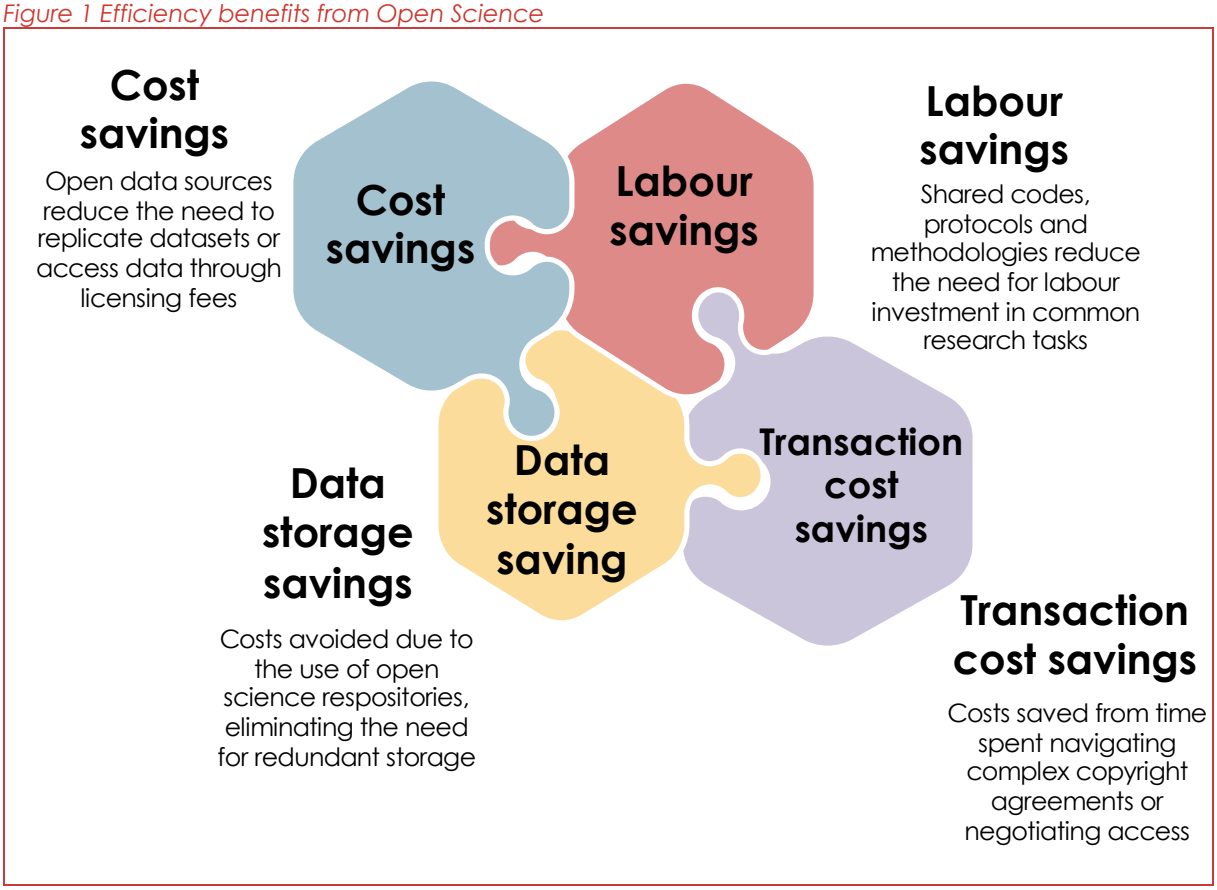
evidence, enhanced educational outcomes, and the creation of public goods that support wider economic activity.

This framework provides the structure for analysing both the literature findings and our case study results, enabling a systematic examination of how Open Science practices translate into measurable economic benefits across different contexts and timeframes.

3.2 Efficiency

In the broadest terms, efficiency refers to getting the same amount of output from research or innovation for less input (e.g. public research funding). Open Science outputs generate substantial efficiency gains by eliminating the need for duplicated efforts and reducing access barriers to research resources. The REA identified four distinct types of efficiency and cost-saving categories (although not all have quantified economic benefits) that demonstrate how Open Science practices create measurable economic benefits. A summary of the various ways in which Open Science can contribute to efficiency benefits is presented in the figure below.

Figure 1 Efficiency benefits from Open Science



Source: Technopolis (2025)

Empirical evidence identified through the course of this REA demonstrates the scale of the benefits derived from Open Science, which can be understood in terms of efficiency gains. In many cases, these approaches to estimating efficiency benefits are captured via large-scale surveys that seek to understand the users willingness to pay (the costs they would pay to retain

access). [21,22] For instance, in the case of open access to research data, studies have highlighted **cost savings** associated with the need for independent data generation. One way that these cost savings have been captured is through measurements of willingness to pay. In behavioural economics, willingness to pay is the maximum price a customer would pay for a product or a service. Research on the National Basic Science Data Centre in China estimated that the total economic benefits of open data provision at 165 million CNY (~£17m or $23m), with users collectively indicating a willingness to pay of 66 million CNY (~ £6.85m or $9.20m). This reflects not only direct cost savings, but also **labour savings,** as researchers avoid duplicating time-intensive data collection activities, and **data storage savings**, as centralised provision reduces the need for fragmented individual storage solutions.

Efficiency gains are even more pronounced in the domain of open source software (OSS). Here, the collaborative model reduces both **development costs** and **transaction costs,** by distributing shared investment across a global community and eliminating redundant parallel development efforts. The most widely used types of open source software required an estimated $4.15 billion (£3.09 billion) in collective development costs, yet the cumulative value to organisations using this software reached $8.8 trillion (£6.55 trillion).[23] This represents the scale of avoided costs that organisations would otherwise incur if required to independently develop equivalent capabilities. Such examples demonstrate how Open Science enables substantial cost, labour, storage, and transaction savings by pooling resources and maximising the value of shared outputs across diverse user communities.

Evidence from case studies: digitising scientific collections

Digital access to natural history collections generates significant economic value, with benefits ranging from millions to billions of pounds, depending on the size of the collection, how frequently it is used (by researchers, educators, etc.) and the proportion of research genuinely enabled by digitisation. Traditional research required costly visits to physical collections, involving travel expenses, accommodation, and time costs. Digital access eliminates these barriers, enabling broader research participation and increased research productivity.

The Natural History Museum (NHM) holds one of the largest and most diverse natural science collections in the world, with more than 80 million specimens. Over the past decade, it has pioneered digitisation methods, setting international standards for large-scale specimen imaging and data sharing. Its Digital Collections Programme (underway since 2014) has already made millions of specimens accessible online, which are widely used by researchers, policymakers, and businesses.[24]

The Royal Botanic Gardens, Kew, is home to approximately 8.5 million plant and fungal specimens, making it one of the largest and most comprehensive collections of its kind. Kew has long been a leader in biodiversity informatics, combining digitisation with genomic, ecological, and conservation research. Digitising a collection does not simply involve a digital image of a specimen, but includes core descriptive data (taxonomic identification, collection event), digital representations (high-resolution images, microscopy imagery), analytical and derived data (genomic sequences, morphological measurements) and links to literature where

---

[21] Tu, Z; Shen, J. Economic valuation of open research data: A conceptual framework and methodological approach. *Research Evaluation,* Volume 33, 2024, rvae033

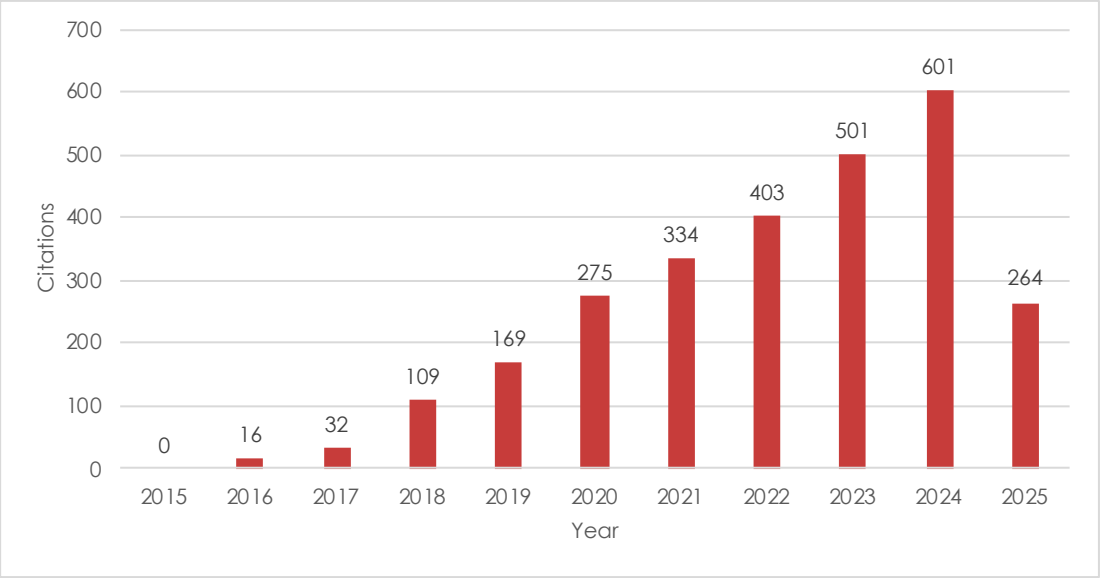
[22] Garrido, C; García, L; Gutleber, J. The Value of a Collaborative Platform in a Global Project. The Indico Case Study. In: Gutleber, J; Charitos, P. (eds) The Economics of Big Science 2.0. Science Policy Reports, 2025. Springer, Cham.

[23] Hoffman, M., Nagle, F., Zhou, Y., The Value of Open Source Software, Working Paper 24-038 (2024)

[24] Natural History Museum, Digital collections programme

the specimen was cited. Because digital specimens are machine-readable and linkable across collections, they open up new scientific possibilities as researchers can then query large datasets that were previously siloed. Since 2015, digitised collections from Kew Gardens has increasingly been cited in academic publications (presented in Figure 2 below).

Figure 2 Citations of Kew Gardens digital data



Source: GBIF Kew Gardens (2025)

The Natural History Museum's digital collections have demonstrated measurable economic impact, with 1,370 publications produced directly from downloaded digital data as of 2021. This research output represents approximately £4.8 million in saved travel-related costs that would otherwise have been incurred by researchers accessing physical collections.

Similarly, Kew Gardens' digital repositories have supported extensive research activity, with 2,706 publications citing their digital data in the Herbarium Specimens collection.[25] This scholarly output corresponds to at least £9.5 million in avoided travel expenses. The global reach of these digital resources extends well beyond the UK, with researchers from 19 different countries publishing papers that cite or rely on Kew Gardens digital data.

These figures likely represent conservative estimates of the true economic benefits. The £3,500 travel cost calculation used in these assessments may understate actual expenses, particularly for international researchers who face higher transportation, accommodation, and subsistence costs when accessing UK-based collections in-person. The international scope of usage suggests that the actual economic value generated by these digital collections significantly exceeds current estimates.

Evidence from case studies: sharing and re-using computational workflows

Making research workflows openly saves time and money, but it's hard to measure exactly how much. When researchers share their computational methods (like data analysis steps or software code) following FAIR principles (Findable, Accessible, Interoperable, Reproducible), it

[25] Royal Botanic Gardens, Kew – Herbarium Specimens, accessed September 2025

creates significant economic benefits. This happens because other researchers don't have to reinvent the same methods from scratch.

The WorkflowHub is a FAIR digital registry for describing, sharing and publishing scientific computational workflows and is part of the European Open Science Cloud (EOSC)-Life investment.[26] A workflow is a structured sequence of tasks or processes that automate the execution of data analysis or complex computational steps.[27] They form a key part in modern, data-intensive scientific research and are used across scientific domains such as the Life Sciences, astronomy, and physical sciences.[28] When considering the economic benefits that flow from the WorkflowHub, it is useful to consider the overall budget of the wider EOSC-Life project (EUR 26.1m), of which the development of the WorkflowHub represents a single work package out of a total 14 (i.e. the Workflow Hub, and therefore the economic benefits, represents a fraction of the full benefits of the investment).

To estimate the value of the platform content, we use the labour replacement cost. This is equal to the cost of reproducing all platform contents from scratch, by hiring a "workflow developer" at a competitive market wage. The replacement cost can be thought of the cost to restore the platform contents to its current state in a scenario where all workflows are permanently lost. This can be achieved through combining wage data with estimates for the time investment required for workflow development. To estimate the demand-side (usage) value of the WorkflowHub, we use workflow access data provided by the WorkflowHub project team. Specifically, we use the number of downloads as a measure of workflow usage or demand.[29]

Table 1 Estimates for the supply and demand-side values of Workflow Hub contents

Supply-side (replacement) cost	**Demand-side (usage) value**
£4.29m	£3,412m

Source: Technopolis (2025)

The economic benefit of the registry can be considered as the aggregate user surplus generated through open access to published workflows. Since usage value far exceeds replacement cost, the economic benefit is substantial. Under the assumptions discussed above, the benefits outweigh the costs by £3.4bn.

As the Workflow Hub user base continues to grow, the added value generated through scientific collaboration is also expected to increase. When viewed in the context of the wider EOSC-Life budget (EUR 26.1m), the disproportionate benefits of open workflow sharing become clear. This reflects the non-rivalrous nature of Open Science outputs, where initial investment yields social values that scales with repeated access and reuse.

[26] https://about.workflowhub.eu/

[27] Goble (2020) – FAIR Computational Workflows - https://direct.mit.edu/dint/article/2/1-2/108/10003/FAIR-Computational-Workflows

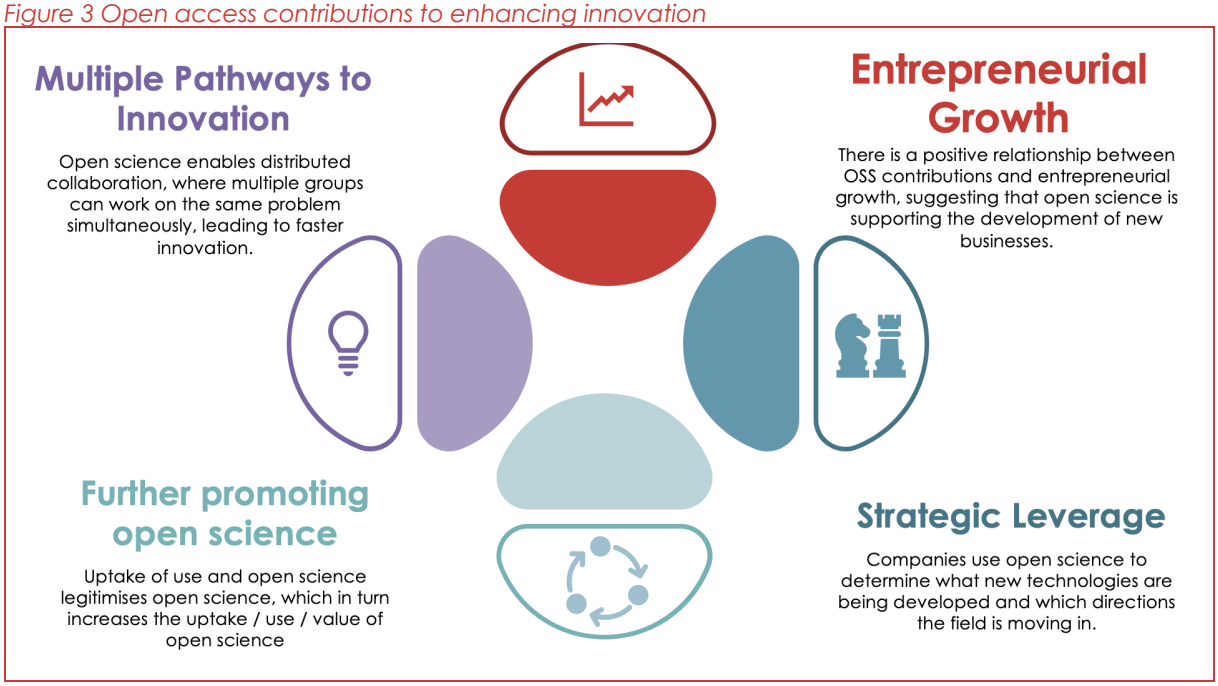
[28] Gustafsson, O.J.R., Wilkinson, S.R., Bacall, F. *et al.* WorkflowHub: a registry for computational workflows. *Sci Data* **12**, 837 (2025). https://doi.org/10.1038/s41597-025-04786-3

[29] *Note: there are a number of assumptions made in the modelling, including approximations of the assumed time in investment and categorisation of different types (simple or complex) workflows. These assumptions are detailed in full in the case study in Appendix C, but are simplified in the table and findings below.*

3.3 Enhancing innovation

A key enabling contribution of Open Science is its capacity to support further research and innovation that would not otherwise have been possible and, in the long-term, supporting wider economic growth. Our REA findings reveal that Open Science practices create multiple pathways for innovation enhancement, though quantitative modelling approaches used to explore these impacts often require substantial assumptions due to the complexity of measuring innovation outcomes. We have identified four ways in which Open Science outputs contribute to the wider innovation ecosystem: (1) by supporting multiple pathways to innovation, (2) catalysing entrepreneurial growth, (3) further promoting Open Science and (4) as strategic leverage for innovators. A summary of these benefits is presented in Figure 2 and explored in more detail below.

Figure 3 Open access contributions to enhancing innovation



Source: Technopolis, 2025

Research examining open source software contributions demonstrates a robust positive relationship between Open Science outputs and entrepreneurial growth, evident across both different contexts and over time.[30] This relationship highlights the existence of multiple pathways through which Open Science **promotes innovation,** ranging from the direct use of open resources in product and service development, to more indirect benefits such as credibility signalling, ecosystem formation and knowledge spillovers. The evidence suggests that entrepreneurs strategically leverage Open Science resources not merely as substitutes for internal knowledge or capabilities, but as tools for shaping technological frontiers, signalling credibility to investors, and attracting clients and users.[31] Crucially, companies use Open Science as a **strategic intelligence mechanism**, scanning emerging open datasets, code, and methodologies to identify what new technologies are being developed and how these can

[30] Wright, N.L., Nagle, F., Greenstein, S., Contributing to Growth? The Role of Open Source Software for Global Startups, Working Paper 24-040 (2024)

[31] Teigland, R., Networked Entrepreneurs: How Entrepreneurs Leverage Open Source Software Communities (2014)

be adapted or extended for their own aims.[32] This strategic usage indicates that Open Science serves as a catalyst for innovation rather than simply a cost-reduction mechanism.

Furthermore, Open Science outputs serve as a form of 'pre-competitive' collaboration space, where organisations can collectively advance foundational technologies and standards before competing on applications and implementations. This pathway has been particularly powerful in industries such as pharmaceuticals and automotive engineering, where pre-completive pooling accelerates collective progress, reduces duplicative costs, and sharpens competition at later stages of innovation.[33]

Economic analysis identifies several channels through which Open Science enhances innovation capacity. The transparency inherent in Open Science practices enables more efficient resource allocation, while reduced duplication of effort and data collection frees up resources that can be redirected toward innovation activities. Another pathway arises through the scaling of AI-driven innovation, as open datasets provide their raw material for machine learning models that can generate new insights and applications far beyond their original purpose.[34] Open Science also accelerates innovation cycles by enabling rapid prototyping and iterative development. When researchers and innovators can access and build upon existing open datasets, code, and methodologies, they can bypass lengthy data collection phases and move more quickly from concept to proof-of-concept. This compression of innovation timelines is particularly evidence in fields like genomics and climate science, where open repositories underpin entire ecosystems of academic and commercial innovation.

Open Science also enhances innovation efficiency by making research failures and negative results more visible. When researchers share what approaches did not work, they enable others to avoid duplicating unsuccessful paths, effectively creating a collective learning mechanism that reduces wasted innovation effort. This 'learning from failure' aspect of Open Science represents significant hidden value that traditional metrics often overlook.

The innovation benefits of Open Science compound through network effects: as more researchers and innovators contribute to open repositories, the value of these resources increases for all users, creating a virtuous cycle. This is particularly evident in open-source software ecosystems, where each contribution enhances the platform's utility and attracts additional users and contributors, thereby accelerating the pace of innovation across the entire ecosystem. Such network effects both lower barriers for new entrants and create fertile environments for entrepreneurial growth, as small firms can build upon large, shared infrastructures without prohibitive upfront costs.

Beyond digital resources, open scientific infrastructure generates innovation benefits through direct human capital development and knowledge transfer. A travel cost analysis of CERN visitors, for example, between June 2018 and May 2019 revealed that total visitor expenditures of EUR 46 million (EUR 552 per visitor on average) and estimated a consumer surplus of EUR 59,902, indicating visitors value their CERN experience at least EUR 0.72 more than their actual travel costs.[35] The positive consumer surplus offers quantitative evidence for the economic returns to open scientific infrastructure, supporting continued public investment in accessible research facilities that generate knowledge spillovers through visitor experiences. Here, the

[32] Sanabria-Z, J., et al, Research foresight in bridging Open Science and open innovation: Overview based on the complex thinking paradigm (2024)

[33] Warren, M., Big pharma is embracing open-access publishing like never before (2019)

[34] Vanschoren, J., et al, OpenML: Insights from 10 years and more than a thousand papers (2025)

[35] Creso Garrido, I.d.R, et al, The Value of Open Science at CERN: An Analysis Based on A Travel Cost Model (2025)

pathway is not only through data access but through human interaction, training, and mobility, which strengthen innovation systems by equipping individuals with skills and networks that travel with them into other sectors. Researchers trained at facilities like CERN often move into industry, bringing sophisticated analytical techniques, collaborative practices, and problem-solving approaches that enhance private sector innovation capabilities.

A cross-country analysis covering major economies from 2000 to 2018 reveals the dual nature of Open Science's innovation impact.[36] While the open-access character creates substantial learning opportunities by providing commonly accessible productive resources, it also generates knowledge spillovers that benefit the global innovation ecosystem. The analysis reinforces the idea of Open Science as a multi-channel driver of growth, directly by creating shared knowledge assets and indirectly by amplifying the absorptive capacity of firms and nations. The research indicates that if no countries contributed to Open Science development, GDP would be (on average) 2.2% lower in the long run, demonstrating the substantial innovation dividends generated by collaborative Open Science practices.[35]

Evidence from case studies: digitised scientific collections

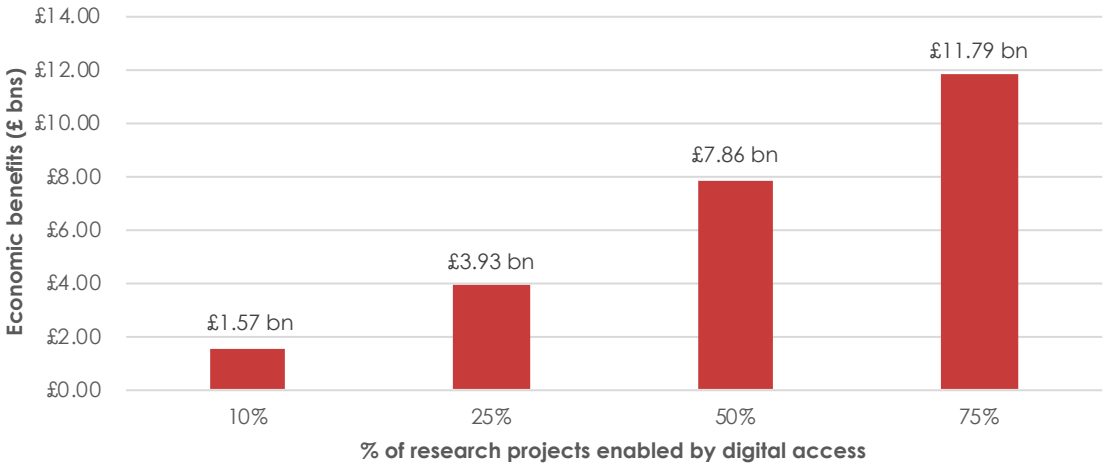
Digitised scientific collections not only reduce costs for researchers, but also enable research that would otherwise be too costly (or even impossible) to undertake. As an example of these benefits, we adopted an approach similar to that used in estimating the value of publications, applying a **travel cost method**. In this case, the **number of digital specimens downloaded** (or a fraction of these) was used as a proxy for research enabled by digital access. This reflects the fact that digital collections transform what is possible by enabling:

- **Data integration at scale,** as researchers can query specimens across thousands of collections simultaneously
- **Cross-disciplinary science,** by linking specimens with omics data, imaging, and other digital resources
- **Widened access,** enabling contributions from non-specialists and researchers who might otherwise lack access to physical specimens without loans or travel

Using download data, we can make considered assumptions about the scale of research projects enabled (for example, taking into consideration the number of specimens that contribute to a single study) and, in turn, estimate the economic benefits of improved access and innovation. Depending on project size, we can reasonable estimate that these range between £1.57 and £11.79 billion (below).

[36] Blind, K., Schubert, T. Estimating the GDP effect of Open Source Software and its complementarities with R&D and patents: evidence and policy implications. *J Technol Transf* **49**, 466–491 (2024). https://doi.org/10.1007/s10961-023-09993-x

Figure 4 Economic benefits for small scale projects (100 downloads / specimens per research project) using Natural History Museum data



Source: Technopolis (2025)

Evidence from case studies: open-access and pre-trained Large Language Models

The BigScience Large Open-Science Open-access Multi-lingual Language Model (BLOOM)[37] is the world's largest open multi-lingual language model. Developed by more than 1000 researchers representing over 250 institutions, the project aimed to democratise access to LLMs, allowing academia, nonprofits, and smaller research labs to study and use these models. LLM research and development is largely dominated by large, mostly American, tech giants, such as OpenAI, Google, and Meta. Prior to the project, the sector was characterised by a closed, culture, in which users have little to no understanding of how models operate or the data they are trained on. BLOOM's development was coordinated by BigScience, an open-science project which aims to promote collaborative practice in the AI/NLP community.

Whilst other models allow access to the underlying code, most (including models from Google and Meta) offer partial transparency, describing composition of training data, but do not disclose exactly what datasets were used. Training data is usually collected from public internet sources (books, websites, code repositories) but requires costly pre-processing. This, among other issues including data misuse or copyright challenges, disincentivises the public sharing of training material. The BLOOM developers offer a higher degree of transparency, documenting the primary training dataset and data preparation code.

Using cloud deployment costs, we estimate the user's Willingness to Pay (WTP) for access to the model, and find that the aggregate usage value ranges from $2.3 - $7.9 billion, depending on the assumed deployment time per user. We show that, even with extremely conservative assumptions on deployment time, the value generated for users outweighs the cost of training the model by a factor of ten.

BLOOM demonstrates how open access to foundational AI infrastructure can significantly enhance innovation. By making both the model and its training data documentation publicly available, BLOOM lowers the barriers to entry for researchers, developers, and organisations that would otherwise be excluded from frontier AI development. This transparency fosters

[37] https://bigscience.huggingface.co/

reproducibility, accelerates experimentation, and enables a broader range of actors to build domain-specific applications.

3.4 Indirect economic impact (wider societal benefits)

The economic benefits of Open Science extend far beyond immediate research applications through several key pathways that generate substantial (but difficult to quantify) returns on public investment. Studies have identified that open access publications show significantly higher citation rates in policy documents, indicating their enhanced role in evidence-based decision-making.[38] This policy influence translates directly into long-term economic growth as research findings inform regulatory frameworks and strategic planning across sectors.

In the context of this REA, we also identified a number of papers that discussed the socio-economic benefits of citizen science as a form of Open Science. Citizen science initiatives demonstrate another pathway to economic impact through community engagement and behavioural change. These participatory approaches build scientific literacy, foster evidence-based community decision-making, and create social capital that supports regional economic development. Enhanced community capacity for problem-solving, improved public health behaviours, and increased civic participation all contribute to economic resilience and growth over extended timeframes.

Knowledge spillovers and innovation ecosystems represent another critical dimension of Open Science's indirect economic impact. When research outputs are freely accessible, they catalyse innovation across unexpected sectors and disciplines, creating value chains that extend far beyond the original research domain. Small and medium enterprises, in particular, are impacted by journal subscription costs, as they typically lack the resources to maintain expensive journal subscriptions. As such, some argue that this provides evidence on the positive impacts that OA mandates then have on national innovation activity.[39] This democratisation of knowledge accelerates technology transfer, reduces duplication of research efforts, and enables faster problem-solving across industries. The cumulative effect manifests in enhanced regional innovation capacity, new business formation, and improved competitive positioning in global markets, though these benefits often materialise over decades rather than years.

The educational and workforce development benefits of Open Science generate long-term economic returns through enhanced human capital. Open educational resources derived from research outputs improve the quality and accessibility of higher education, while open datasets provide invaluable training opportunities for emerging data scientists and researchers. This expanded access to cutting-edge knowledge and methodologies helps address skills gaps in the workforce, particularly in data-intensive and technology-driven sectors. Furthermore, the collaborative practices inherent in Open Science cultivate transferable skills such as digital literacy, critical thinking, and interdisciplinary communication that are increasingly valued across the modern economy. These workforce capabilities contribute to productivity gains and economic adaptability, enabling societies to respond more effectively to technological disruption and economic transformation.

---

[38] Pablo Dorta-González et. al, Societal and scientific impact of policy research: A large-scale empirical study of some explanatory factors using Altmetric and Overton (2024)

[39] Does open access to academic research help small, science-based companies? ElSabry et al., Journal of Industry-University Collaboration (2020), DOI: 10.1108/JIUC-04-2020-0004

Evidence from case studies: open methodologies and training resources

The National Centre for Research Methods (NCRM) is a UK-based organisation that provides training, resources, and support to enhance research methods capacity across the social sciences. Established in 2004 and funded by the Economic and Social Research Council (ESRC), NCRM serves as a central hub for disseminating best resources and practices in research methodology. Now in its fourth phase of funding (which began in 2020), it focuses on meeting the diverse and changing needs of researchers through informed pedagogy and high-quality training and capacity building provision.

This study estimates the economic value generated by NCRM's Open Science initiatives by focusing on a defined set of freely available resources and activities. While the list below is exhaustive for the purposes of the analysis, it does not represent the full breadth of NCRM's work. Training programmes, although central to NCRM's mission, are low-cost rather than free and therefore excluded from this analysis. Additionally, some freely available resources, such as podcasts, have been excluded due to limitations in available data. The activities included in the analysis are:

- **Events** (both online and in-person focused on capacity building, networking, lectures, workshops)
- **Online Tutorials** (structured video series with supporting materials explaining various research methods.)
- **YouTube videos** (tutorials delivered through one or more videos, recordings of lectures and workshops)
- **E-publications** (freely accessible digital publications on research methods and related topics)
- **Training guides** (resources for trainers)

This analysis employs the **opportunity-cost method** to estimate the economic value of NCRM's open-access resources. The opportunity cost method is a well-established economic valuation technique that measures the value of a resource by examining the value of what individuals forgo to access it, which is typically their time that could otherwise be spent on alternative activities. By applying this approach, the analysis captures the implicit value that users place on NCRM's open-access offerings, reflecting their willingness to invest time in engaging with these resources.

This case study shows that NCRM generates high economic value through its Open Science activities. The economic benefits range from **£29.83 M to £80.22 M** between 2020 and 2025 (amounting to between £6m and £16m annually), depending on the level of engagement. The figures do not capture the full breadth of NCRM's activities and consequently underestimate the economic benefits. The findings affirm the significance of investment in Open Science.

Evidence from case studies: open software (R)

R is an open-source programming language and environment primarily used for statistical computing, data analysis, and visualisation. Its adoption is steadily increasing across academia and industry, with notable applications in sectors such as biotech, finance, and high technology.

R offers significant educational and societal benefits by providing free public access to tools for producing and consuming knowledge. It fosters innovation and collaboration through its

growing community of developers and users, who continuously enhance R's core functionality and offer efficient solutions to existing challenges. Additionally, R is a cost-effective tool for both commercial and non-commercial use.

As an open-source platform, traditional evaluation methods often fail to capture its full value. This study applies non-market valuation techniques to estimate the economic benefits derived from R. These benefits are measured as consumer surplus, calculated based on demand (number of users) and willingness to pay (cost of alternative proprietary software).

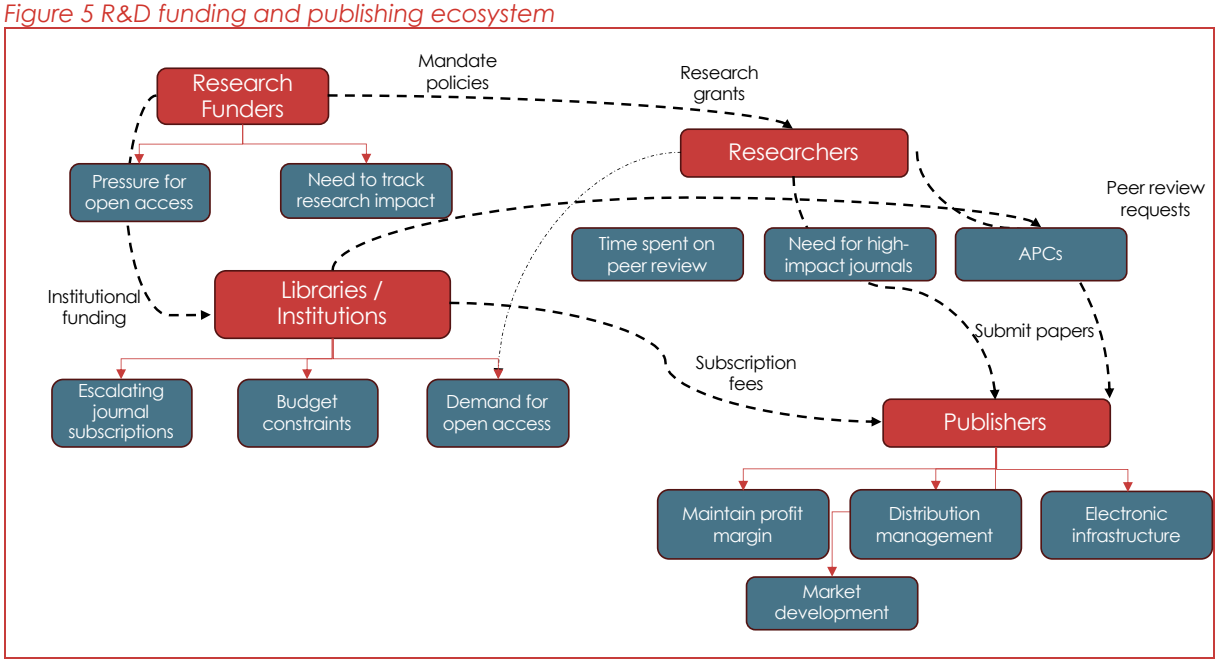
There are an estimated 2 million R users. Willingness to pay is approximated at £750, based on the pricing of proprietary alternatives such as STATA, SAS, and SPSS. To account for varying user preferences, three substitution scenarios are considered (50%, 30%, and 10%).

Overall, the economic benefits are estimated to range from £150 million to £748 million, even under conservative assumptions. This case study highlights that even modest estimates of R's replacement value emphasise the substantial economic impact of Open Science.

4 Costs of Open Science

In considering the transition to an Open Science ecosystem, it is clear that a focus solely on the benefits risks overlooking important questions about who bears the costs, how these costs are distributed across the R&D ecosystem, and what the long-term implications might be for system sustainability. The Open Science ecosystem is large and complicated, and includes research funders, institutions, researchers, publishers, industry and infrastructure providers (among others). For the purposes of this report, we have focused on three key stakeholders who directly manage the cost of a transition to Open Science, including researchers, libraries and publishers. This section of the report examines how the financial and organisational burdens of Open Science are distributed among those three key stakeholder groups.

Figure 5 R&D funding and publishing ecosystem



Source: Technopolis, 2025

4.1 Costs for Academic Publishers

Academic publishers play a multi-faceted role in the dissemination of scholarly outputs. Typical publisher activities include:[40]:

- **Journal and community development efforts**, including commissioning content, researching editorial board members and scope development
- **Submission to desk reject or acceptance**
- **Peer review management** by staff, including management of submissions that are ultimately rejected
- **Acceptance to publication,** including typesetting, conversation, and production tasks

[40] Breakdown of service developed by Information Power for cOAlition S's price transparency framework

- **Services after publication**, including ethics checks and queries, usage statistics, long-term preservation and access monitoring
- **Sales and marketing** to customers or of articles, including marketing campaigns
- **Author and customer support**, including queries about licensing, citations, and author system troubleshooting
- **Marketing and outreach:** Attracting authors, increasing readership, and developing subscriber bases

The transition to Open Science models represents a fundamental shift in the revenue structures and service offerings that have sustained academic publishing for decades. Traditional subscription-based models are being replaced or supplemented by APCs, fundamentally altering the financial relationship between publishers and the research community.

Between 2019 and 2023, an estimated $8.39 billion was spent on APCs paid to just six major publishers for open access publication.[41] This figure, derived from open datasets of annual APC list prices, represents only a portion of the global market and suggests the substantial scale of financial flows in the evolving Open Access landscape. These estimates are valuable for supporting more informed, data-based decision making by libraries, consortia, funders, and researchers as they navigate the changing publication ecosystem.

Beyond Open Access publishing, publishers may be increasingly expected to support the broader Open Science agenda by providing services that curate / host / link to other types of researcher outputs like data and code .[42] This includes developing platforms that can host or link to datasets, creating persistent identifiers for diverse research outputs, and implementing systems that verify the availability and accessibility of associated research materials. These technological investments represent significant operational costs, as do the development of new peer review processes that assess not just manuscripts but also accompanying data, code, and methodological documentation.

For publishers, the transition involves significant operational restructuring costs, including the management of hybrid models that maintain both subscription and open access content, and the need to demonstrate value in an increasingly competitive market where preprint servers, institutional repositories, and specialised data repositories offer alternative dissemination routes for multiple types of research outputs.[43] Fully open access publishers face a different but related challenge: sustaining operations and investing in new infrastructure and services without the buffer of subscription income, particularly as funder mandates and market competition place downward pressure on article processing charges (APCs).

4.2 Costs for Libraries

The role of the research library is likely to change substantially as Open Access (as part of the broader Open Science ecosystem) becomes more ubiquitous.[44] Instead of primarily acquiring materials produced elsewhere for the consumption of local patrons, libraries are assuming a larger role in ensuring that locally produced scholarly products are disseminated effectively

[41] Hausetein et. al, Estimating global article processing charges paid to six publishers for open access between 2019 and 2023 (2024)

[42] Late, E., Pölönen, J., How society publishers practice Open Science beyond open access publishing?

[43] Trial, H., Einseidler, J., The future of research dissemination: Innovation in publishing formats (2024)

[44] Hoving, S., Librarians in the age of open access: An evolving role (2024)

worldwide via Open Access publication channels. This shift represents both an opportunity and a challenge, requiring new skills, infrastructure, and strategic approaches.

Supporting comprehensive Open Science may involve libraries investing in and maintaining diverse infrastructure beyond traditional publication repositories. This could include institutional data repositories, code repositories, platforms for sharing methodological protocols, and systems for managing research software. Each type of repository needs specialised technical expertise, storage capacity, metadata standards, and curation practices. Some libraries may also provide support services to help researchers prepare and deposit various types of research outputs, necessitating staff with expertise in data management, software preservation, and metadata creation across multiple disciplinary contexts.[45]A separate, but important consideration for libraries is the financial landscape of Open Access (OA). Currently, institutions often face fragmented payment streams: libraries pay site licenses for subscription content, while researchers independently pay APCs. [46] This can create a risk of "double dipping", where publishers are effectively paid twice for the same content. Much of this expenditure comes directly from research funding, and libraries are often unaware of the total flow of money across APCs and subscriptions.

One potential solutions is the creation of centralised university-level OA funds. Such funds could collect and administer all OA-related payments, including APCs and membership fees. Centralisation would improve visibility of actual costs, enable more effective negotiation with publishers, and help prevent double payment scenarios, aligning financial management with Open Science goals.

Beyond the immediate financial challenges, the final success of Open Science depends on finding solutions for both the long-term preservation and knowledge organisation of diverse research outputs, including publications, datasets, software code, and methodological documentation.[47] These are major challenges that need to be solved by digital academic libraries and archives. Unlike traditional subscription models where publishers maintained archives as part of their business model, the distributed nature of Open Science publication and data sharing creates questions about who bears responsibility for ensuring perpetual access to multiple types of research outputs, each with different preservation requirements and technical challenges.

Software and code preservation presents particular difficulties, as these outputs may become unusable as computing environments evolve. Data preservation requires not only storage but also the maintenance of documentation and metadata that enables future use. Libraries may need to take on these preservation costs and technical responsibilities across all forms of research outputs, representing a significant long-term financial commitment that may not be immediately apparent in current budgets. In addition, specialist repositories (often operated at disciplinary or national levels) have emerged to provide domain-specific resources and expertise. Examples of these include Zenodo (operated by CERN), that accepts data, code and software) and complements institutional infrastructure and expanding the Open Science ecosystem.[48]

---

[45] Wong et. al., Research data management services in academic libraries to support the research data life cycle: A systematic review (2025)

[46] Fecher et. al., Open Access, Innovation, and Research Infrastructure (2016)

[47] Shanmugam, R., Data stewardship for Open Science: implementing FAIR principles / Data-driven storytelling (2018)

[48] Zenodo

4.3 Costs for Researchers and Authors

The most visible cost to researchers is the APC for open access publication, which can range from a few hundred to several thousand dollars per article depending on the journal and publisher. However, Open Science involves additional potential costs: fees for depositing data in specialised repositories, costs for obtaining persistent identifiers (such as DOIs) for datasets or software, and expenses associated with preparing research materials to professional standards for sharing. For researchers in well-funded institutions or those with substantial grant support, these costs may be manageable. However, for early-career researchers, those in less-funded disciplines, or scholars from institutions without robust library support or dedicated Open Science funds, these charges can represent a significant barrier to full participation in Open Science practices.[49]

Beyond direct fees, researchers face substantial time costs in preparing materials to align with various platform requirements across multiple types of research outputs. Open Science involves considerable additional work: preparing data for deposit in repositories with appropriate documentation and metadata; writing code comments and documentation that make software comprehensible and reusable by others; creating detailed methodological protocols; navigating different platforms with different technical standards and requirements; ensuring compliance with funder mandates, institutional policies, and ethical requirements for data sharing; and responding to requests for clarification or access to materials from other researchers.

While these activities may enhance the value, transparency, and reproducibility of research, they represent real opportunity costs, time that could otherwise be devoted to conducting research, writing additional papers, or pursuing other career-advancing activities. The effort required to make research truly open and reusable is often substantially underestimated in policy discussions, and researchers may not receive career recognition commensurate with the time invested in these activities.

Participating meaningfully in Open Science often requires researchers to develop new technical competencies. This includes learning version control systems for managing code and collaborative documents, understanding data management best practices, mastering metadata standards relevant to their discipline, and becoming familiar with various repository platforms and their requirements. Acquiring these skills requires time investment in training, which may not be readily available or may come at institutional cost. Researchers may also need to invest in learning new tools for creating reproducible computational analyses or for documenting their methodological workflows in standardised formats.

Beyond time and skill requirements, there are strategic costs on the researcher's side, including the opportunity cost of patenting and potential career implications of opening up developed knowledge instead of protecting it. These concerns are especially acute in academic fields where Open Science adoption is nascent or in sectors where scientific knowledge appropriation can yield significant economic benefits. For researchers working at the interface of academia and industry, or in fields with strong commercial applications or competitive research environments, the decision to openly share findings, detailed methods, or proprietary

[49] David Nicholas et. al., Purchase and publish: Early career researchers and open access publishing costs (2024)

software involves weighing the public benefit and scientific advancement against potential personal or institutional financial returns from intellectual property protection.[50]

In highly competitive fields, there may also be perceived risks in sharing code, data, or detailed methods before publication, including concerns about being "scooped" by competitors or having errors in data or code exposed to public scrutiny. While Open Science advocates argue that transparency ultimately benefits both individual researchers and the research enterprise, these concerns represent real psychological and professional costs that researchers must navigate.

The distribution of Open Science costs is not uniform across the global research community. Researchers in lower-income countries may face particular challenges in paying APCs and repository fees, even when fee waivers are nominally available.[51] They may also have less access to technical training and infrastructure support for managing and sharing diverse research outputs. Similarly, disciplines with different funding structures and research practices experience Open Science costs differently: a well-funded biomedical researcher may find APCs and data management costs easily absorbed within grant budgets and may work in a field with established data sharing norms and infrastructure, while a humanities scholar working without significant external funding may find the same charges prohibitive and may lack disciplinary repositories suited to their research materials.

Computational and data-intensive disciplines may face higher costs for data storage and management, while also potentially having better access to technical expertise and infrastructure. Experimental sciences may struggle with making physical materials or detailed protocols sharable in standardised formats. These disparities raise important equity questions about whose research becomes openly available across all dimensions of Open Science, and who is excluded from participating fully in the Open Science ecosystem, and points to areas that require further investment and support in order to enable Open Science expansion.

[50] Anneke Zuiderwijk et. al., What drives and inhibits researchers to share and use open research data? A systematic literature review to analyse factors influencing open research data adoption (2020)

[51] Juliet Nabyonga-Orem et al, Article processing charges are stalling the progress of African researchers: a call for urgent reforms (2020_

Appendix A Rapid Evidence Assessment

A.1 Methodological Approach

Our approach for this REA was modelled off of that set out in "The Economic Impact of Open Science: A Scoping Review" Tsipouri et al. (2025)[52]. The authors follow the PRIMA-ScR[53] framework, following four key steps: (1) identification of relevant studies, (2) assessment against pre-defined eligibility criteria, (3) data extraction, (4) reporting findings. This REA sought to provide an update on the state of play in a rapidly evolving field, and hence we modified their approach as follows:

- **Identification of relevant studies** – The scoping review mentioned above searched for papers published between January 1st 2000 and November 8th 2022. Our search was restricted to literature published after November 8th 2022, in order to deliver a new and up-to-date assessment of developments in the field, whilst avoiding repetition of findings. To complement this, we used a snowball approach, identifying key papers that were cited in those already found and had been published after November 2022.
- **Assessment against pre-defined eligibility criteria** – We made two key changes to the criteria set out in Tsipouri et al. (2025). Firstly, as mentioned above, the restriction on publication date was modified. Secondly, the criteria set out in the scoping review states "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." We adopted a similar approach in defining different types of OS, but relaxed the restriction on provision of evidence. Specifically, we included articles discussing methodological approaches to estimating the economic value of OS, given the primary objectives of the REA. Specific exclusions defined in the scoping review were maintained.

  Following the identification of relevant studies and subsequent eligibility assessment, a total of 22 reports and publications were selected.
- **Data extraction** – Studies were manually reviewed by a member of the research team, with the following information being extracted from selected studies:
  - General Scope & Aims
  - Empirical Study (Yes/No)
  - Description of Methodology
  - Data Sources
  - Methodological & Data Challenges
  - Relevance to OS (coded by the 6 OS types mentioned above)
  - Economic Impacts Covered
  - Key Findings

Data categories were defined prior to the review of selected studies and were chosen to reflect the primary objectives of the assessment.

[52] https://osf.io/preprints/metaarxiv/kqse5_v1

[53] https://www.prisma-statement.org/scoping

A summary of the reports is presented in Table 2 below, followed by a summary analysis of key findings from the review.

Table 2 Summary of literature identified in the REA

Study	Empirical Study	Methodology	Economic Impacts Covered
Open Access Publications			
The impact of open access mandates on scientific research and technological development in the US (Probst et al. 2023)	✓	Difference-in-difference	Enhanced Innovation
Linking Science and Industry: Influence of Scientific research on Technological Innovation through Patent Citations (Dorta-González et al. 2025)	✓	Poisson regression model, maximum likelihood estimation & delta method using Altmetrics	Enhanced Innovation
Societal and Scientific impact of policy research: A large scale empirical study of some explanatory factors using Altmetrics and overton (Dorta-González et al. 2024)	✓	Linear regression framework	Indirect economic effects (e.g. policy and societal impacts)
Using the future wheel methodology to assess the impact of open sceince in the transport sector (Nielsen et al. 2023)		Future wheel methodology	Enhancing innovation, efficiency gains
An Analysis of the Effects of Sharing Open Research Data, Code and Preprints on Citations (Colavizza et al. 2024)	✓	Linear regression framework	Indirect economic effects
A Multiple Linear Regression Analysis to Measure the Journal Contribution to the Social Attention of research (Dorta-González, 2023)	✓	Linear regression framework	Enhanced Innovation
Open-Source Software			
Estimating the GDP effect of Open-Source Software and its complementarities with R&D and patents: evidence and policy implications (Blind, 2024)	✓	Macro-econometric modelling (panel cointegration model)	Economic growth
The Value of Open-Source Software (Hoffmann, Nagle and Zhou, 2024)	✓	Labour replacement cost method	Efficiency gains
Measuring the Economic Value of Open Source - A survey and preliminary Analysis (Chesbrough, The Linux Foundation, 2023)	✓	Perceived costs/benefits to alternatives, counterfactual analysis	Efficiency gains
Contributing to Growth? The Role of Open-Source Software for Global Startups (Wright, Nagle and Greenstein, 2024)	✓	Panel correlation analysis, matching, firm-specific fixed effects regression framework	Enhancing innovation, economic growth
Measuring Software Innovation with Open-Source Software Development Data (Brown et al. 2024)	✓	Semantic Versioning analysis	Enhanced Innovation
Open/FAIR Data and Data Platforms			
Economic valuation of open research data: A conceptual framework and methodological approach (Tu and Shen, 2024)	✓	Contingent Valuation Method	Efficiency Gains, economic growth
FAIR data management practices to introduce circular economy in aquaculture: Benefits, barriers and a		Mixed methods approach	Efficiency gains, sustainability

preliminary roadmap (Giuffrida et al. 2025)			
The effect of Open Data on cost savings (Apartis et al. 2024)		Direct Acrylic Graph (DAG)	Efficiency gains, enhanced innovation
What is the value of data? A review of empirical methods (Coyle and Manley, 2023)		Methodological/Literature Review	
The Value of a Collaborative Platform in a Global Project. The Indico Case Study (Garrido, García and Gutleber, 2024)	✓	Contingent Valuation Method, Choice Experiment Method	Efficiency gains, enhanced innovation, local economic growth
The Value of an Open Scientific Data and Documentation Platform in a Global Project: The Case of Zenodo (Garrido, García and Gutleber, 2024)	✓	Estimation of Net Present Value (NPV), Contingent Valuation Method	Efficiency gains, enhanced innovation
Infrastructure			
The Value of Open Science at CERN: An Analysis Based on a Travel Cost Model (Garrido, Loureiro García and Gutleber, 2025)	✓	Travel Cost Method	Enhanced innovation
Costs and benefits of Open Science: contributing to the development of a rigorous assessment framework (Catalano, Delugas and Vignetti, 2025)		Cost-Benefit Analysis (CBA)	Efficiency gains
Citizen Science			
Beyond Science: Exploring the value of co-created citizen science for diverse community groups (Robinson, Delany and Sugden, 2024)		Value Creation Framework	Indirect economic impacts
Report on the costs and benefits of citizen social science: Analysis based on the YouCount project (Franco, Lorenz and Norvoll 2024)		Cost-Benefit Identification	Efficiency gains, enhanced innovation
Monitoring Citizen Science Performance: Methodological Guidelines (Skaržauskienė et al. 2024)		CS Performance Index - CSPI	Enhanced innovation

Source: Technopolis (2025)

Appendix B Approaches to evaluating the economic impact of Open Science

Open Science Output	Methodological Approaches	Strengths	Weaknesses	Examples
 **Open Access Publications**	• Linear regression frameworks • Difference-in-difference	• Concrete disparity between access types allows for the identification of average treatment effects • Flexibility in Control variables	• Focus on citation impact doesn't necessarily translate into economic impact • Causal identification requires strong assumptions in absence of natural experiments	• Impact of Open Access on citations of scientific papers in patents (Dorta-González et al. 2025) • Impact of adopting different OS principles on citations (Colavizza et al. 2024) • Factors affecting social impact of scientific research (Dorta-González, 2023)
 **Open Source Software**	• Macroeconomic modelling	• Provides estimates for Long Run macroeconomic impacts	• Identification reliant on strong theoretical assumptions	• Estimating the GDP effect of OSS and complementarities with R&D (Blind 2024)
	• Labour replacement cost method	• Direct, understandable approach • Obtains both supply and demand side estimates	• Demand side estimates assume that absence of OSS would require duplication by all users • Technological advances mean labour is no longer the sole input in software development	• Estimating the economic value of OSS (Hoffmann, Nagle and Zhou, 2024)
	• Survey-based approaches	• Incorporates user perspectives	• Respondents may have imperfect information • Reliance on subjective respondent views	• Survey and analysis investigating the private value of OSS (Chesbrough, The Linux Foundation, 2023)
	• Semantic versioning analysis	• Enables differentiation between major and incremental OSS innovations	• Self-proclaimed innovation labels risks endogeneity	• Impact of innovation complexity on downstream use of OSS (Brown et al. 2024)
	• Panel correlation, matching & firm-specific fixed effects	• Supports causal inference • Allows for longitudinal tracking of firm-specific outcomes	• Linking OSS contributions to firms presents a challenge	• Investigation into the relationship between Oss contributions and tech startup performance

			• Danger of selection bias – decision to contribute OSS is unlikely to be random	(Wright, Nagle and Greenstein, 2024)
 **Data**	• Contingent Valuation Method (CVM)	• Estimates direct value to users • Useful in estimating economic value of a specific repository/platform	• Respondents may have imperfect information • Reliance on subjective respondent views • Hypothetical bias – behaviour in a hypothetical situation differs from real economic decision making • Difficulty in extrapolating results	• Economic valuation of China's National Basic Science Data Centre (Tu and Shen, 2024) • Case study investigating socio-economic value of Indico, an open event management platform, operated by CERN (Garrido, García and Gutleber, 2024)
	• Estimation of Net Present Value (NPV)	• Allows for the socio-economic valuation of an investment over a defined time horizon • Segmented benefit estimation enables comparison between impact channels	• Focussed on a single platform • Difficulty in extrapolating results	• Case study investigating value of Zenodo, an open scientific data and documentation platform operated by CERN (Garrido, García and Gutleber, 2024)
	• Direct Acrylic Graph (DAG)	• Easily interpreted visual representation	• Theory based – does not allow for quantifiable results	• Exploration of the theoretical pathways from open data access to cost savings (Apartis et al. 2024)
 **Infrastructure**	• Travel Cost Method	• Provides revealed/realistic estimates of willingness to pay • Supports estimation of a demand function and consumer surplus derived from an infrastructure	• Focussed on a single infrastructure • Only applicable to physical sites attracting tourists	• Assessment of the economic value of CERN to visitors (Garrido, Loureiro García and Gutleber, 2025)
	• Cost-Benefit Analysis (CBA)	• Direct comparison of costs and benefits • Easily interpretable results	• Costs are typically borne by a single organisation, whereas benefits are available to all • Difficulty in providing a comprehensive view of all costs and benefits associated with an ecosystem such as Open Science	• Theoretical description of how CBA methodology can be applied to assess economic impact of OS (Catalano, Delugas and Vignetti, 2025)

 **Citizen science**	• Case study approaches • Cost-Benefit Identification	• Surveys and interviews incorporate participants • Qualitative analysis provides learnings rather than focussing on results	• Theory based – does not allow for quantifiable results • Monetisation of costs and benefits is challenging e.g. attaching cost to time devotion of citizens • Lack of control group for counterfactual analysis • Focussed on a single project	• Study exploring the social and environmental value of co-created CS (Robinson, Delany and Sugden, 2024) • Case study identifying the costs and benefits of the YouCount project (Franco, Lorenz and Norvoll 2024)

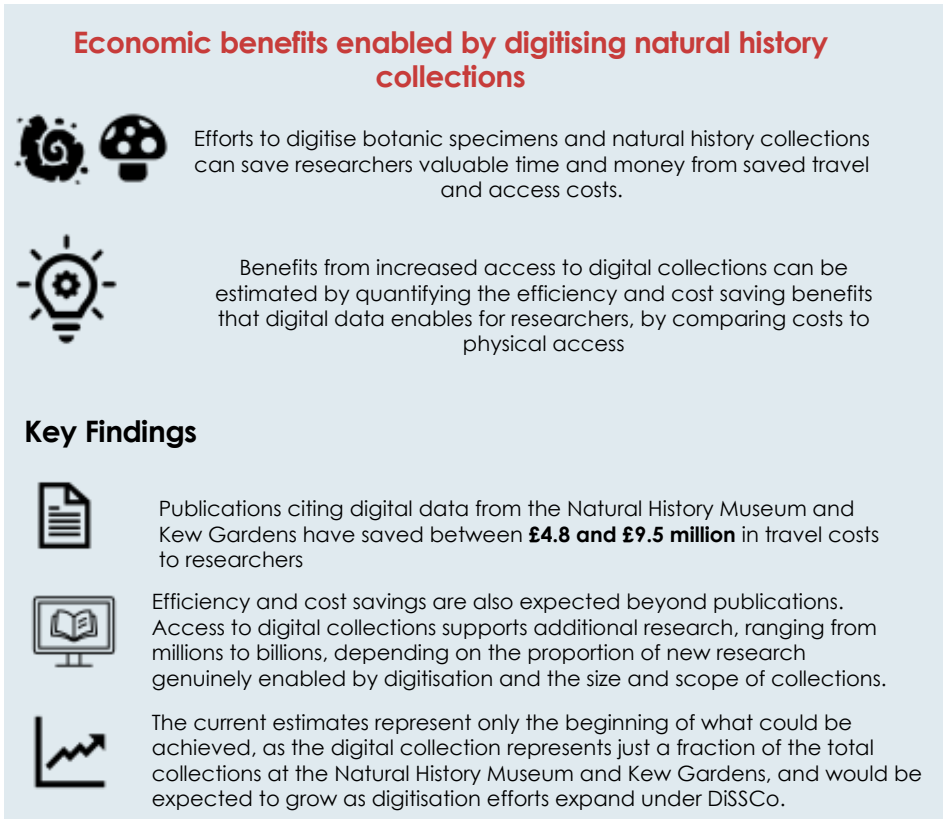
Appendix C Case Studies

C.1 Digitising scientific collections through DiSSCo

Headline Findings

Digital access to natural history collections generates significant economic value, with benefits ranging from millions to billions of pounds, depending on the size of the collection, how frequently it is used and the proportion of research genuinely enabled by digitisation. This case study examines the economic impact of natural history collection digitisation, drawing on evidence from institutions participating in the European Distributed System of Scientific Collections (DiSSCo) initiative. The findings demonstrate a compelling investment case for DiSSCo UK's £155 million digitisation programme.

As of 2021, at least 1,370 publications cite the Natural History Museum's digital data and 2,706 publications cite the Kew Gardens fungi collection, representing between £4.8 and £9.5 million in travel cost savings from avoiding physical collection visits that typically cost researchers £3,500 per project. These outcomes from European DiSSCo participation provide evidence supporting the business case for similar digitisation infrastructure investment in the UK.



Perhaps most significantly, digital collections are (and could continue to) fundamentally democratising research access by eliminating traditional barriers that have historically limited scientific investigation to those with substantial travel budgets. The data shows that small-scale research projects that make use of and access digital data could deliver the greatest proportional economic returns. This democratisation effect is evidenced by the scale of digital engagement: between 2015-2021, 28.6 billion record downloads occurred from just 4.9 million

specimens at the Natural History Museum, with projections suggesting that this will exceed 45 billion downloads in 2025. [54]

The current economic benefits represent only the beginning of what could be achieved, as they emerge from the digitisation of just 7% of the Natural History Museum's total collection. In the long-term, continued investment in digitisation efforts could yield economic returns that would be expected to scale as digitisation efforts expand through initiatives like DiSSCo UK.

DISSCo: Distributed System of Scientific Collections

The **Distributed System of Scientific Collections** (DiSSCo) is a pan-European research infrastructure designed to transform access to natural science collections by bringing them together into a single, open, and integrated digital platform.[55] In the UK, DiSSCo UK is now being developed as a national initiative led by the Natural History Museum, London, with a consortium of regional and local partners. Between September 2021 and March 2022, with funding from the Arts and Humanities Research Council (AHRC), the consortium carried out a comprehensive study to understand the scale and diversity of UK natural science collections and assess their readiness for digitisation, with plans to launch the huge £155m initiative in 2026. This case study examines the economic benefits that have emerged from the European DiSSCo programme and considers how these insights apply to the forthcoming UK initiative.

The UK's natural science collections represent a national asset of global significance. Together, they number in the hundreds of millions of specimens (spanning zoology, botany, palaeontology, geology, and mycology) and provide a physical record of life and environments on Earth across millennia. These specimens are critical for understanding biodiversity loss, climate change, food security, public health, and the sustainable use of natural resources. Yet, the majority remain inaccessible, stored in cabinets and vaults, available only to those able to travel and consult them in person.

DiSSCo UK aims to change this by creating a dynamic, one-stop e-science infrastructure, openly available to researchers, businesses, policymakers, and the public. Through a "hub and spoke" model of digitisation at national, regional, and local levels, the initiative will deliver:

- Improved data quality through standardisation and interoperability.
- Relevant, usable data for science, innovation, and decision-making.
- Enhanced national infrastructure for biodiversity and heritage information.

The economic case for investment in digitising such collections is compelling and many researchers have already attempted to capture this value.[56] DiSSCo UK estimates that a £155 million investment to digitise a critical mass of the UK's natural science collections could generate up to a ten-fold return, delivering substantial economic, social, and scientific benefits, including (i) time savings for researchers who can access specimens remotely, (ii) enabling new research that would not occur due to travel constraints. (iii) research efficiency gains and (iv) reduced research costs compared to physical collection visits. Considering the scale and breadth of the physical collections, digitised collections available at scale could support in areas including crop improvement, ecological restoration, and conservation

[54] Projections made by Technopolis, assuming a linear relationship between the number of new digital specimens and the download frequency. This is likely a conservative estimate – one study analysing downloads has shown an exponential growth in the number of downloads of specimen-derived digital data from the Naturalis Biodiversity Centre.

[55] The Distributed System of Scientific Collections UK

[56] Data integration enables global biodiversity synthesis, Heberling *et al.* 2021

planning. In short, DiSSCo UK represents a transformative opportunity to unlock the potential of the UK's collections for the digital age, ensuring that these important assets are accessible, usable, and impactful for generations of researchers. This case study therefore provides a timely assessment of the early economic signals emerging from existing digitisation initiatives.

The Natural History Museum

The Natural History Museum (NHM) holds one of the largest and most diverse natural science collections in the world, with more than 80 million specimens. Over the past decade, it has pioneered digitisation methods, setting international standards for large-scale specimen imaging and data sharing. Its Digital Collections Programme has already made millions of specimens accessible online, which are widely used by researchers, policymakers, and businesses.

NHM's track record demonstrates the scientific and societal value of digitisation. For instance, digitised entomology collections have been used to track long-term changes in pollinator populations, while plant and fungal collections contribute to research on food security and invasive species. By leading DiSSCo UK, NHM is extending this expertise across the national collections landscape.

Royal Botanic Gardens, Kew: A Digital Gateway to Plant and Fungal Diversity

The Royal Botanic Gardens, Kew, is home to approximately 8.5 million plant and fungal specimens, making it one of the largest and most comprehensive collections of its kind. Kew has long been a leader in biodiversity informatics, combining digitisation with genomic, ecological, and conservation research. Its digital collections provide global access to herbarium sheets, type specimens, and associated metadata, supporting research into taxonomy, agriculture, and climate resilience.

Kew's digitisation efforts underpin projects such as the *State of the World's Plants and Fungi* report and contribute to international initiatives on crop wild relatives, medicinal plants, and ecosystem restoration.[57]

Approach to estimating economic benefits

This analysis employs an adapted travel cost method to estimate the economic value of digitising natural history specimen collections. The travel cost method is an established economic valuation technique that measures the value of a resource be examining the costs people are willing to incur to access it.

The travel cost method is particularly relevant for evaluating digitised natural history collections because it captures the fundamental shift from physical to digital access. Traditional research required costly visits to physical collections, involving travel expenses, accommodation, and time costs. Digital access eliminates these barriers, enabling broader research participation and increased research productivity.

To provide clearer transparency, the approach used in this case study explicitly separates two categories of economic value: (1) direct cost savings from avoided physical visits, and (2) additional research activity that is only made possible because digital access removes logistical constraints.

Using data from the Natural History Museum dashboard (which tracks the number of specimens available online, among other statistics), we have developed a model based on key assumptions to capture economic benefits specific to digital collection usage. The assumptions

[57] State of the World's Plants and Fungi, Royal Botanic Gardens, Kew

below are evidence-based where possible; where empirical evidence is limited, we have used conservative values to avoid overstating benefits. These assumptions include:

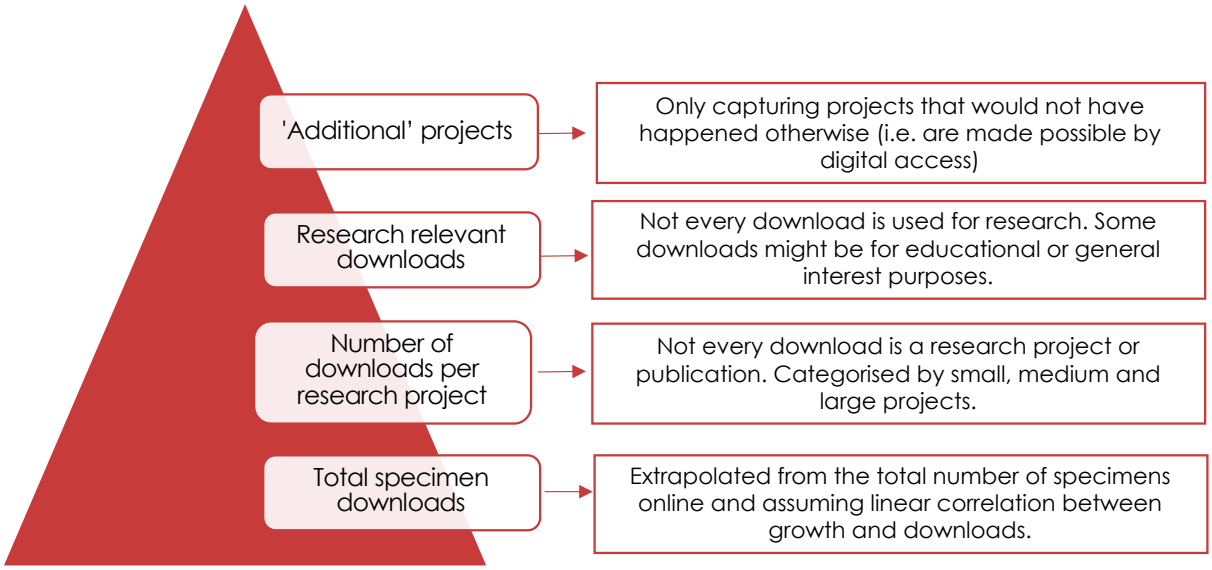
- **Download volumes and projections:** Between 2015-2021, there were 28.6 billion record downloads from 4.9 million online records. Using linear growth models, we project that there have now been upwards of 45 billion downloads by 2025 based on ~6 million online specimens.[58] This projection is conservative, as digital uptake often follows an exponential rather than linear trend.
- **Research project categorisation:** Since one download does not equal one research project, we have categorised projects that vary in their download volumes and specimen requirements per research output (between 100 and 500 downloads (specimens) per research project).[59] These categories reflect observed variations in project intensity across fields (e.g. taxonomy vs ecological modelling)
- **Research conversation rates:** Only a small proportion of downloads represent substantive research projects leading to publications. Our assumptions range from 1-20% for varying levels of research intensity.
- **Additionality assessment:** Not all digitally-enabled projects are genuinely additional – some would have proceeded using physical collections alone. We tested additionality rates between 25-50% to ensure robust estimates, supported by surveys indicating that nearly half of researchers could not undertake their work without digital access.
- **Cost savings calculations:** We applied the established value of £3,500 per research project in avoided travel and access costs, representing the total cost savings from digital rather than physical collection access.[60]

This approach to using the secondary data available is summarised in the figure below. This framework allows us to quantify the economic value generated by digital collection infrastructure while acknowledging the uncertainties inherent in translating download data into research impact.

[58] Digital Collections Programme, Natural History Museum

[59] We have made these assumptions by looking at a sample of publications that use museum specimens to determine appropriate 'buckets' or categories of projects.

[60] The Value of Digitising Natural History Collections, Frontier Economics, 2021



Economic Impact of Digitised Collections

Publications

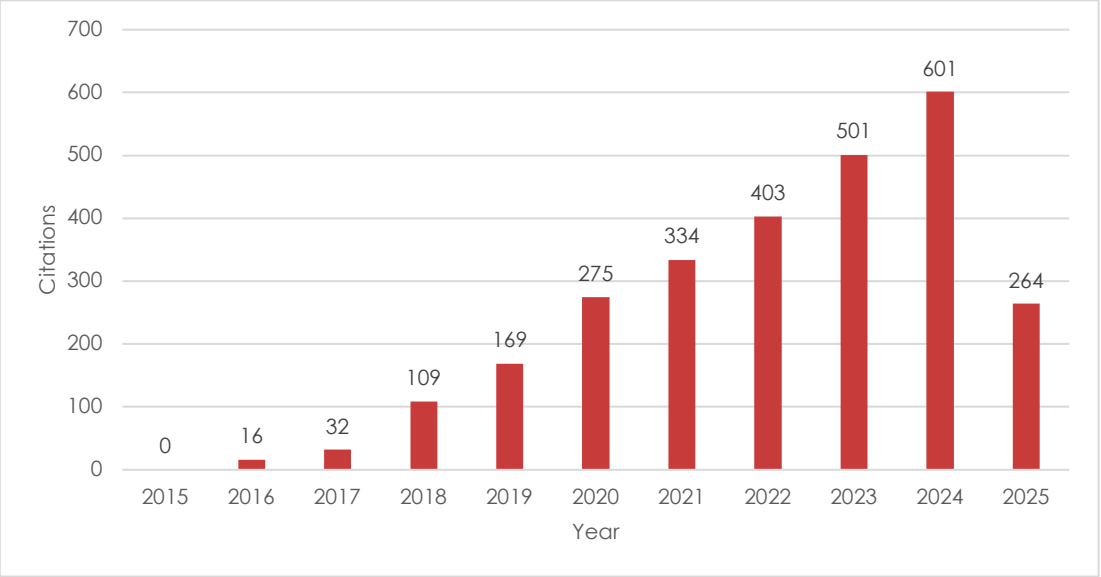
The economic impact of providing digital access to museum collections extends far beyond basic operational efficiencies. Data from the Natural History Museum demonstrates the substantial cost savings achieved through their digitisation efforts. As of 2021, the museum recorded 1,370 publications that resulted directly from downloaded digital data.

Research into the costs of accessing physical collections reveals that the average research project requires approximately 10 days of researcher time and incurs expenses exceeding £3,500 per project, primarily due to travel and accommodation costs. When applied to the documented publications alone, this suggests that digital access to the Natural History Museum's collections has saved researchers collectively around £4.8 million in travel-related expenses.

These figures represent a conservative estimate of the economic benefits, as they only account for tracked publications through 2021 and do not capture the full scope of research activities, preliminary investigations, or educational uses that digital access enables. Given the continued growth in digital collection usage over the past five years, the actual cost savings are likely substantially higher. In particular, the rapid expansion of global digital research workflows and increasing reliance on open biodiversity data suggest that current estimates understate both present and future value.

The herbarium collections (home to approximately seven million plant specimens collected from around the world) at the Royal Botanic Gardens Kew provides a clear example of the growing value and use of digital collections. Since 2015, the digitised collection has increasingly been cited in academic publications

Figure 6 Citations of Kew gardens digital data



Source: GBIF Kew Gardens (2025)

Using the same approach as earlier to determine the economic benefits, the 2,706 publications which cite digital Kew Gardens data suggest that at least £9.5 million in travel-related expenses have been saved through digital and open access. In addition to UK contributions, online repositories indicate that 19 different countries have published papers which either cite or rely on Kew Gardens digital data, suggesting that the value of £3,500 used to calculate travel costs is conservative and economic benefits are likely larger.

Broader Economic Impacts

Additional Research Enabled by Digital Access to Collections

The modelling results demonstrate substantial economic benefits from digitised collections, with total value ranging from £0.79 to £7.9 bn for collections at the Natural History Museum depending on assumptions about additionality (the proportion of research projects genuinely enabled by digital access rather than physical collections). Previous research reveals that almost half of researchers wouldn't be able to conduct investigations without digitised data, which we have used as a baseline for our economic modelling.[61]

Table 3 Economic benefits emerging from access to NHM collections, varying % of projects that additional due to digital access

% of projects that are additional due to digital access	Small (100 specimens per research project)	Medium (500 specimens per research project)
25%	£3.93bn	£0.79bn
50%	£7.86bn	£1.57bn

Note: These calculations are made by holding the % of research-relevant downloads (i.e. downloads that are specifically used for research) constant at 1%

[61] Digitising UK museum collections could help to boost the global economy, Natural History Museum (2023)

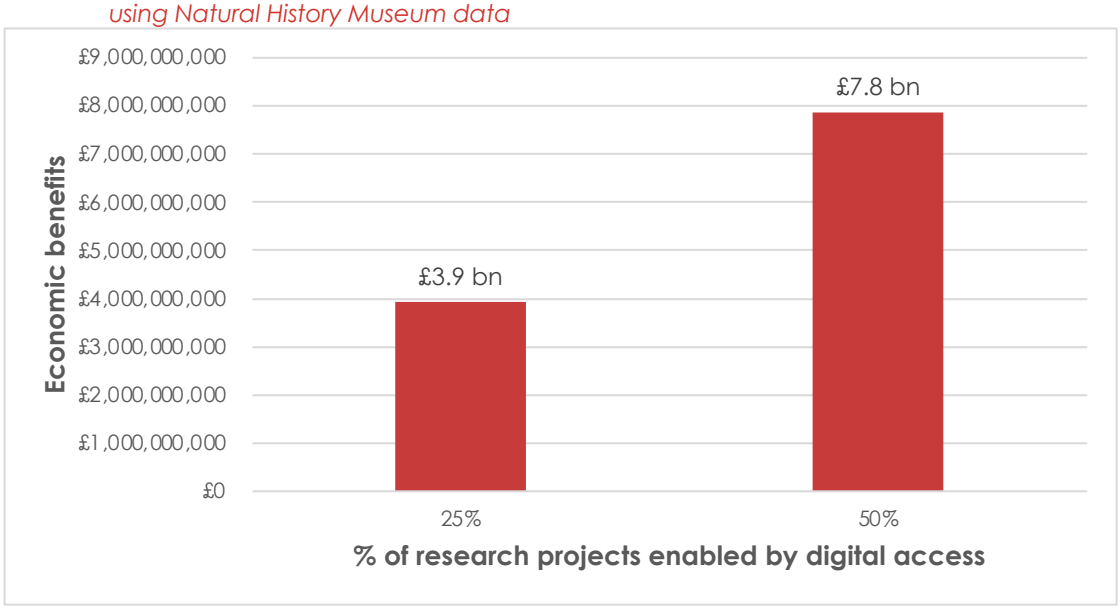
Table 4 Economic benefits emerging from access to Kew Gardens fungi collections, varying % of projects that additional due to digital access

% of projects that are additional due to digital access	Small (100 specimens per research project)	Medium (500 specimens per research project)
25%	£67m	£13.4m
50%	£134m	£27m

Source: Technopolis (2025)

The economic benefits scale directly with the percentage of projects that are additional due to digital access. At conservative estimates (25% additionality), the total economic values reaches £3.9 billion, rising to £7.8 billion when 50% of projects are considered genuinely enabled by digitisation. This demonstrates the importance of digital access as an enabler of research activity – even modest increases in the proportion of digitally dependent-research translates into billions in additional economic value.

Figure 7 Economic benefits for small scale projects (100 downloads / specimens per research project) using Natural History Museum data



Source: Technopolis (2025)

The greatest economic returns are also delivered by enabling numerous, smaller, research activities, rather than facilitating fewer large projects. This disproportionate impact on small projects indicates that digital collections have the potential to democratise research access, allowing researchers with limited travel budgets (including early-career academics, those at smaller institutions, and international researchers) to conduct meaningful studies.

Research Relevant Downloads

Determining what proportion of downloads constitutes "research relevant" activity presents significant methodological challenges. Digital collections serve multiple purposes beyond direct research, including education, public engagement, and general awareness. Evidence suggests considerable variation across disciplines and the types of data made available:

- Social sciences data shows approximately 40% of downloads were explicitly for research
- A survey on teaching vs research suggested that up to 85% of downloads (of qualitative data) serve teaching or learning purposes

Because of this variability, our modelling adopts a deliberately low central assumption of 1% research relevance to avoid overstating impacts. This means our estimates are conservative: as the proportion of downloads used for research increases, the associated economic benefits rise proportionally. Even small increases in research-relevant downloads can therefore generate substantial additional economic value.

Conclusions

The economic benefits demonstrated in this analysis likely only represent a fraction of the total potential value from the Natural History Museum's digital collections programme. At the time of the original study, the museum had digitised 6% of their total collection. Since then, the collection has continued to expand, with upwards of 6 million specimens now available online (an increase of 1.5 p.p.). Moreover, as of September 2025, the DiSSCo platform reports a total of over 17 million specimen records available across 9,134 UK natural science collections (which are not included in this analysis).

This growth trajectory suggests that the economic benefits calculated from current usage patterns reflects just a portion of the resources available digitally, and therefore only a fraction of the total achievable economic impact. As digitisation efforts such as DiSSCo continue to expand the proportion of collections available online, the economic returns are likely to scale proportionally, potentially reaching significantly higher values as more specimens become accessible to researchers worldwide.

C.2 Open Source LLMs: BLOOM

Headline Findings

Providing AI researchers and developers with open access to a pre-trained Large Language Model (LLM) generates significant value for users. Using cloud deployment costs, we estimate the user's Willingness to Pay (WTP) for access to the model, and find that the aggregate usage value ranges from $2.3 - $7.9 billion, depending on the assumed deployment time per user. We show that, even with extremely conservative assumptions on deployment time, the value generated for users significantly outweighs the cost of training the model.

BLOOM – Democratising AI

The BigScience Large Open Science Open-access Multi-lingual Language Model (BLOOM)[62] is the world's largest open multi-lingual language model. Developed by more than 1000 researchers representing over 250 institutions, the project aimed to democratise access to LLMs, allowing academia, nonprofits, and smaller research labs to study and use these models. LLM research and development is largely dominated by large, mostly American, hyperscalers, such as OpenAI, Google, and Meta. Prior to the project, the sector was characterised by a closed culture, in which users have little to no understanding of how models operate or the data they are trained on. BLOOM's development was coordinated by BigScience, an Open Science project which aims to promote collaborative practice in the AI/NLP community.

Open Source LLMs offer several advantages over proprietary models:

- **Transparency** – Users have full visibility into the model's architecture, training data, and algorithms. This enables peer review to ensure the model's integrity and performance.

[62] Introducing the World's Largest Open Multilingual Language Model - BLOOM

- **Customisation** - Users have the ability to tweak the model to better suit their requirements. This can be achieved through retraining models on sector-specific data to improve performance in a reduced context window.
- **Enhanced data security and privacy** - Deployment of open-source LLMs on private infrastructure eliminates concerns of third-party data access[63].
- **Cost savings** – Open access allows organisations to save on pay-per-use and licensing fees.
- **Sustainability benefits** – Open access reduces the need for duplicated training of large models in private settings, for which vast amounts of computing power and energy are required.

The BLOOM LLM had a significant impact on the attitudes towards collaboration in the AI sector. Since the launch of BLOOM in July 2022, major players, namely Google[64] and Meta[65], have released their own open-source LLMs. Furthermore, the main release paper - BLOOM: A 176B-Parameter Open-Access Multilingual Language Model[66] - has received significant academic attention, with a total of 2,517 citations as of September 2025. BLOOM differs from other open-source LLMs through its transparency and multi-lingual capabilities. Whilst other models allow access to the underlying code, most (including models from Google and Meta) offer partial transparency, describing composition of training data, but do not disclose exactly what datasets were used. Training data is usually collected from public internet sources (books, websites, code repositories) but requires costly pre-processing. This, among other issues including data misuse or copyright challenges, disincentivises the public sharing of training material. The BLOOM developers offer a higher degree of transparency, documenting the primary training dataset and data preparation code. A full description of the construction of the dataset is given in BLOOM: A 176B-Parameter Open-Access Multilingual Language Model[67]. Secondly, BLOOM is competent in 46 natural and 13 programming languages. This scale of language diversity is unusual, with English typically dominating Natural Language Processing (NLP) pipelines. This was an objective of the project from the beginning, increasing AI accessibility for non-English speakers.

LLM training makes up a significant proportion of the development cost, and have been found to be increasing at a rate of 2.4x a year since 2016[68]. The process requires access to state-of-the-art hardware facilities and is incredibly time and energy intensive. Training costs are increasing in both model size (measured in parameters) and the amount of training data used (measured in tokens[69]). Model performance is also increasing in these variables, creating a trade-off between cost and performance. The full BLOOM model has a size of 176 billion parameters and was trained using 350 billion tokens of data. For context, this architecture is similar to the leading model developed by OpenAI at the time, GPT-3, which has 175 billion

[63] This advantage can only be realised through local deployment on private infrastructure. This issue is discussed further in the Methodology section.

[64] Google DeepMind released the Gemma series of LLMs in February 2024, now on its third iteration - https://deepmind.google/models/gemma/gemma-3/

[65] Meta AI developed the LlaMa family, first released in February 2023, currently on version 4 - https://www.llama.com/

[66] BLOOM: A 176B-Parameter Open-Access Multilingual Language Model, BigScience Workshop, 2022

[67] BLOOM: A 176B-Parameter Open-Access Multilingual Language Model, BigScience Workshop, 2022

[68] The Rising costs of training Frontier AI Models, Cottier et al. 2025

[69] A token is a unit of text that the model processes. Inputs (words, part of a word, or punctuation) are "tokenised" by the LLM to make natural language machine readable.

parameters. The BLOOM LLM was trained using the Jean Zay supercomputer at IDRIS in France[70]. **The process took three and a half months at an estimated cost of over $9 million[71].**

Any individual or institution can access the pre-trained model through the Hugging Face repository[72]. Hugging Face is an online repository for the sharing of Machine Learning (ML) and NLP models, offering users several access options which are discussed in more detail in the methodology section below. The model can be accessed and used in six different specifications, ranging from the full 176 billion parameter model, down to the smallest version with 560 million parameters. The sharing of smaller versions has become standard with open-access LLMs, due to the resource constraints associated with running larger models. The Hugging Face repository plays a key role in our approach to quantifying the economic benefits of the BLOOM LLM.

Methodology

Willingness to Pay (WTP)

Our approach to quantifying the economic benefits of Open Source LLMs is based on estimation of the user's **Willingness to Pay (WTP)**. Specifically, we use the revealed preference method. This involves inferring the user's true value of access to the pre-trained model through observing an existing market price of a comparable good or service. In the case of LLM's, this could be achieved through the use of a comparable (in terms of size and capability) proprietary model. Here, the thinking is that the price an individual is willing to pay for use of a closed LLM reflects the value to that individual, and we can assume the value of the service is similar across individuals. However, open and proprietary models are inherently different in their value to the user. Proprietary models are closed and often referred to as a "black box" in which users have little to no understanding of how the output they see is generated. Customers are not granted access to the underlying code or the data used to train the model, suggesting the value is solely in the output. This is where open-source LLMs (and BLOOM in particular) differ – user value is generated through not only output but the ability to understand, modify and fine-tune the model.

In the context of BLOOM, we assume that a downloading user intends to deploy the model, and that the cost of deploying the model reveals the user's willingness to pay for using the LLM.

Our estimates include:

- Value: based on deployment costs (per hour)
- Usage: number of individuals accessing the models (based on downloads) and hours spent using the model

Deployment costs

Deploying or running inference on LLMs requires significant computing power, typically through Graphics Processing Units (GPUs)[73], which entail additional costs. Users can either deploy locally on private infrastructure, which is optimal for data sovereignty and may be preferrable for large

[70] Jean Zay: The supercomputer, IDRIS

[71] BLOOM Overview, Suvorov 2023

[72] https://huggingface.co/

[73] Not all models require GPU access – smaller or quantised versions of larger models can be run on a Central Processing Unit (CPU), but inference will be significantly slower.

scale commercial use, but this requires underlying infrastructure (GPU, CPU, memory) and consumes energy.

In most cases, users will "spin up" a GPU instance through the Hugging Face repository (or another cloud GPU provider). This is essentially users renting a virtual machine with all hardware requirements built in. **Virtual machines are charged at an hourly rate, providing a market price for an hour of "running" the model**.

The full model is too large to run efficiently on a single GPU, it is optimal to distribute the compute across GPU's using model parallelism techniques (essentially using multiple GPUs). The Incredibly Fast BLOOM Inference with DeepSpeed and Accelerate[74] article states that the **most efficient set-up is to use 8x80GB A100 GPUs;** therefore we use the Hugging Face hourly rental cost for this GPU[75]. **The hourly rate for a single 80GB Nvidia A100 GPU is $2.50[76], meaning the hourly rental cost of 8 is $20.** Again, we assume that downloading user intends to deploy the model, and this cost reveals the user's willingness to pay for one hour of LLM use.

This deployment cost only reveals the WTP of those deploying virtually. There will be some who download and run locally on private infrastructure. Local infrastructure is only economically viable with large-scale deployment, such as commercial or institutional use, due to the required infrastructure investment (a single A100 GPU can cost anywhere from $20,000 - $35,000). Due to difficulty identifying the deployment method of users, we assume WTP is consistent across users.

User access

To estimate the aggregate value of access to the pre-trained model, we combine WTP ($20per hour) with user access data, based on number of downloads. It is important to note that "downloads" from the Hugging Face repository does not only refer to downloads in a traditional sense (i.e. the transfer of files to a user's device), but it also counts HTTP requests to model files[77]. This means that the download volume allows the tracking of cloud deployment, in addition to a full download for local deployment, giving a complete view of true model access. Hugging Face provide information on download volume on the "model card", showing the number of occasions on which the model has been "downloaded" over the last month[78]. Our figure of interest is the cumulative download volume since launch (July 2022). We use the Internet Archive Wayback Machine[79] to observe monthly downloads for every month since the official launch date, allowing the computation of total downloads over BLOOM's lifetime. Total downloads amount to **4,866,603** since the model was launched.[80]

---

[74] Incredibly Fast BLOOM Inference with DeepSpeed and Accelerate, Bekman and Gugger, 2022

[75] Users are not limited to GPU rental through hugging face; it is possible to use other cloud providers. GPU rental through Hugging Face presents the easiest option for users, as all requirements are built into a single ecosystem.

[76] Hugging Face hardware cloud rental costs

[77] For a complete description of how downloads are counted, see Hugging Face Models download Stats

[78] It is unclear whether this is calculated monthly (e.g. downloads in October) or is measured on a monthly rolling basis (e.g. downloads in the past 30 days).

[79] The Internet Archive Wayback Machine lets you view historical versions of web page. For consistency, we collect downloads per month using the last available date in any given month. Snapshot availability depends on page traffic on any given day, meaning it is not possible to view previous versions of a web page on every day of the month.

[80] The download volume shown above only considers download volumes for the full 176 billion parameter model. Smaller versions see higher download volumes due to their relatively low deployment cost. For example, the smallest 560 million parameter model saw 95,463 downloads in September 2025, in comparison to just 3,785 for the full model.

The last element of the calculation requires estimating average usage time per user.

The time spent deploying the model will vary significantly between users – an AI researcher is likely to spend significantly more time deploying the model than a curious developer. In the scope of our study, we have no way of knowing the purpose of each download, and therefore we cannot make accurate estimates on the average usage time. To account for this, we apply an average deployment time for each user type (casual, intensive) and show estimates under varying assumed shares of casual and intensive users.

We assume that, on average, a casual user spends 5 hours actively deploying the model. Casual users are individuals who engage with the model on an occasional or exploratory basis. They may include curious developers experimenting with the model's capabilities or students learning about AI systems. Intensive users are individuals or organisations that rely on the model for sustained, high-volume tasks. This group may contain companies or institutional users retraining the model on sector-specific data, integrating the model into products or workflows, or running large-scale experiments. We assume intensive users deploy for an average of 100 hours.

Main findings

Combining total downloads (since July 2022), WTP estimates ($20 per hour), and assumed deployment time gives estimates for the usage value of the BLOOM LLM:

$$Aggregate\ Usage\ Value = Total\ downloads \times (WTP_{hourly} \times hours\ of\ use)$$

We consider three different scenarios:

- Casual and intensive users are equally distributed throughout the population
- 80% of users are casual, 20% are intensive
- 80% of users are intensive, 20% are casual

Estimates for the aggregate usage value under the different scenarios are presented in the table below.

Table 5 Estimates for aggregate usage value under different scenarios

Average deployment hours per user	Total downloads	Deployment cost	Aggregate usage value ($)
5	0.5 x 4,866,603	$100	$243.3m
100	0.5 x 4,866,603	$2,000	$4,866.6m
		Total	**$5,109.9m**

Scenario 1 – equal distribution

Average deployment hours per user	Total downloads	Deployment cost	Aggregate usage value ($)

Analysis of download volumes of all versions would lead to larger value estimates, however we choose to focus on the full BLOOM LLM rather than all project outputs.

5	0.8 x 4,866,603	$100	$389.3m
100	0.2 x 4,866,603	$2,000	$1,946.6m
		Total	**$2,336.0m**

Scenario 2 – Casual user weighting

Average deployment hours per user	**Total downloads**	**Deployment cost**	**Aggregate usage value ($)**
5	0.2 x 4,866,603	$100	$97.3m
100	0.8 x 4,866,603	$2,000	$7,786.6m
		Total	**$7,883.9m**

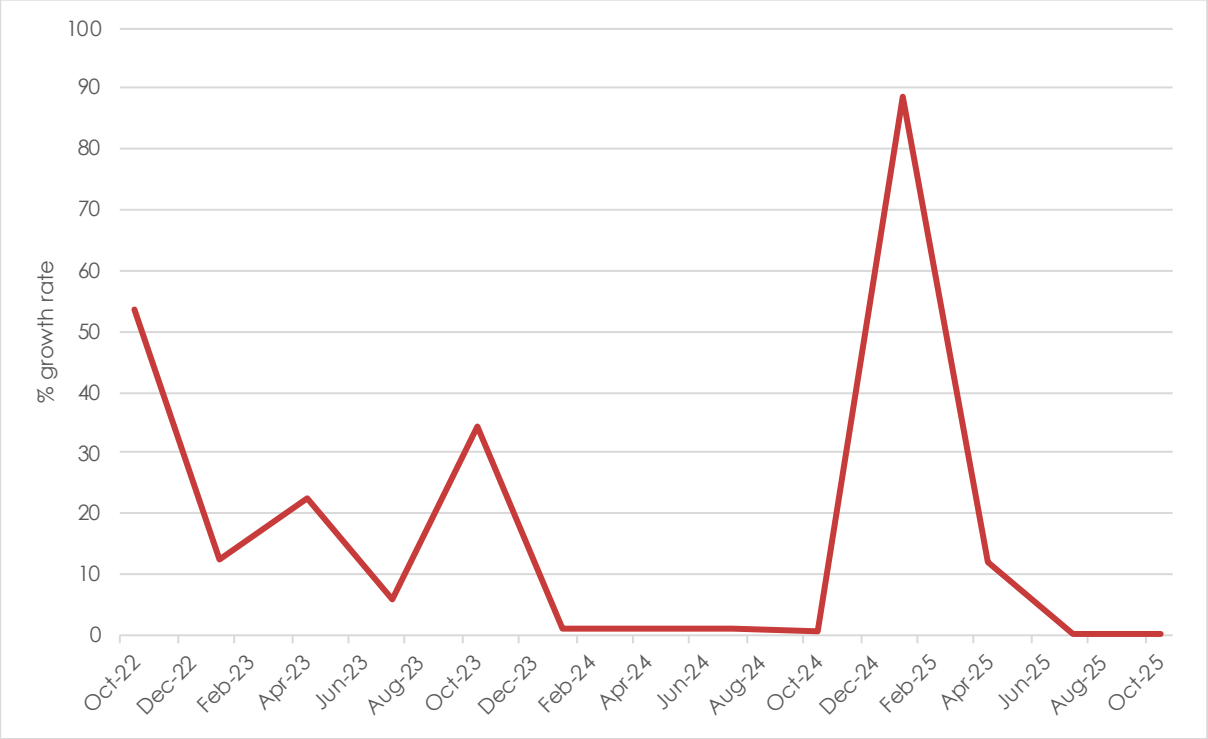
Scenario 3 – Intensive user weighting

The results presented above show the considerable value generated through open access to the BLOOM LLM, with estimates for the aggregate usage value ranging from $2.3 billion - $7.9 billion, depending on the assumed user distribution. When interpreting these results, it is useful to compare to the cost of training the model, estimated at $9 million. Table 5 shows that, even under our most conservative assumptions, the aggregate usage value significantly outweighs the training cost.

The results presented above are highly dependent on the assumed hours of deployment and could be significantly larger. We argue that our assumed average deployment time for intensive users (100 hours) is modest. The presence of a handful of highly intensive users who deploy the model for extended periods of time (1000+ hours for example) will result in a significant shift in the average deployment time per intensive user. That is, the distribution of deployment hours is concentrated at the lower end, but extreme values at the upper tail will significantly raise the average value, driving up estimates for the aggregate usage value.

Next, we look at the growth in usage value over time. Figure 8 below shows the three-month moving average of the growth rate in usage value, based on an assumption of 10 deployment hours per user. The rate at which value grows effectively tracks the growth rate of downloads, as WTP and average deployment time are constant.

Figure 8 Growth rate in usage value, presented as a three-month moving average



Source: Technopolis (2025)

Ignoring outliers driven by unusually high download volumes in January and February 2025, Figure 8 above shows a gradual decline in the growth of usage value generated by the BLOOM LLM. This trend would be observed at any chosen (constant) level of average deployment hours per user, as it is purely driven by download volumes. The trend can be seen to drop off about one a half years after launch (July 2022), reflecting the rapid pace at which the AI sector is evolving.

Conclusions

In this case study, we have attempted to place a monetary value on the aggregate usage value generated through granting users open access to a pre-trained LLM. Through the estimation of user's willingness to pay, we show that this value far exceeds the cost of training the model, even with stringent assumptions on the time for which users deploy the model. This analysis only considers the usage value, representing only a proportion of the total economic benefit to society.

C.3 Sharing of computational workflows - WorkflowHub

Headline Findings

The FAIR (Findable, Accessible, Interoperable, Reproducible) sharing of computational workflows generates value that significantly outweighs the required investment. The extent of this disproportionality is inherently difficult to assess, and estimates are highly sensitive to underlying assumptions. The estimated benefits range from £722m to £3.45bn depending on assumptions on the required time investment for workflow production. In all cases, the findings highlight the significant added value arising from collaborative workflow sharing, primarily through reducing duplicated effort and enabling substantial gains in research efficiency.

WorkflowHub

The WorkflowHub is a FAIR digital registry for describing, sharing and publishing scientific computational workflows[81]. A workflow is a structured sequence of tasks or processes that automate the execution of data analysis or complex computational steps[82]. They form a key part in modern, data-intensive scientific research and are used across scientific domains such as the Life Sciences, astronomy, and physical sciences[83].

The platform is domain and type-agnostic, allowing workflows to remain in their native format, with flexible search options to support findability. It allows workflows to be FAIR, citeable, and openly available, whilst requiring the upload of rich metadata to support user accessibility[84]. This machine-readable metadata enables systematic categorisation and efficient search, maximising the value of hosting large numbers of workflows in a single registry.

The WorkflowHub delivers significant benefits to computational researchers across both academic and industrial settings. Rather than each researcher developing their own analysis methods from scratch, scientists can now find and build upon existing workflows, dramatically reducing duplication of effort and accelerating research progress. This approach mirrors the advantages of open-source software, where collaborative development produces better tools more efficiently than isolated efforts. Since its inception as part of the European Open Science-Life (EOSC-Life) project in September 2020, the WorkflowHub has become a leading registry for the publication of computational workflows. The platform now hosts a total 1,390 workflows, with 1,179 registered users at the time of writing (September 2025).

Methodology

Our approach to quantifying the economic benefit of the WorkflowHub follows the Labour replacement cost method. Labour market approaches offer a direct quantification method and have previously been used to estimate the value of Open-Source Software[85]. Our approach does not attempt to provide a complete monetary valuation of the WorkflowHub, but rather use the registry as an example to highlight the economic benefits arising from the sharing of analytical methods. We combine platform statistics, labour market data, and time investment estimates to calculate (1) the supply-side (replacement) cost of the platform, (2) the demand-side (usage) value, and (3) the economic benefit derived from open access.

Supply-Side (Replacement) Costs

The total supply-side cost of the WorkflowHub registry is a combination of:

- The development and maintenance costs of the platform itself (inception, governance, platform/website design)
- The development and maintenance costs of the registry contents (the workflows themselves)

In this study, we focus on the registry contents. Development and maintenance of the underlying infrastructure represents considerable long-term investment, which is difficult to quantify in absence of compiled financial data. The WorkflowHub was developed as part of a wider Horizon 2020 project (EOSC-Life) with an overall budget of EUR 26m. EOSC-Life was a

[81] WorkflowHub project

[82] FAIR Computational Workflow, Goble et al. 2020

[83] WorkflowHub: a registry for computational workflows, Gustafsson, Wilkinson, Bacall et al. 2025

[84] WorkflowHub project

[85] The Value of Open-Source Software, Hoffmann, Nagle, and Zhou, 2024

large project with many elements; due to limited data, it is not possible to determine what proportion of this expenditure specifically funded WorkflowHub's development.

To estimate the value of the platform content, we use the labour replacement cost. This is equal to the cost of reproducing all platform contents from scratch, by hiring a "workflow developer" at a competitive market wage. The replacement cost can be thought of the cost to restore the platform contents to its current state in a scenario where all workflows are permanently lost. This can be achieved through combining wage data with estimates for the time investment required for workflow development.

Wage data is taken from the ONS 2024 Annual Survey of Hours and Earnings (ASHE)[86]. Annual wages have been converted to hourly rates under the assumption of 252 working days a year and a 40-hour working week. We acknowledge that using UK wage data simplifies a more complex international reality, but it provides a practical and transparent basis for estimation. Table 6 below shows a summary of the wage data collected.

Table 6 Gross (median) yearly salary across life science professions

Profession	Gross Yearly Pay	Gross Hourly Pay
Natural and Social Science Professionals	£41,495	£20.58
Biological Scientists	£40,347	£20.01
Chemical Scientists	£39,904	£19.79
Biochemists and Biomedical Scientists	£45,925	£22.78
Average	**£41,918**	**£20.79**

Source: ONS – Employee earnings in the UK: 2024

Estimating time investment

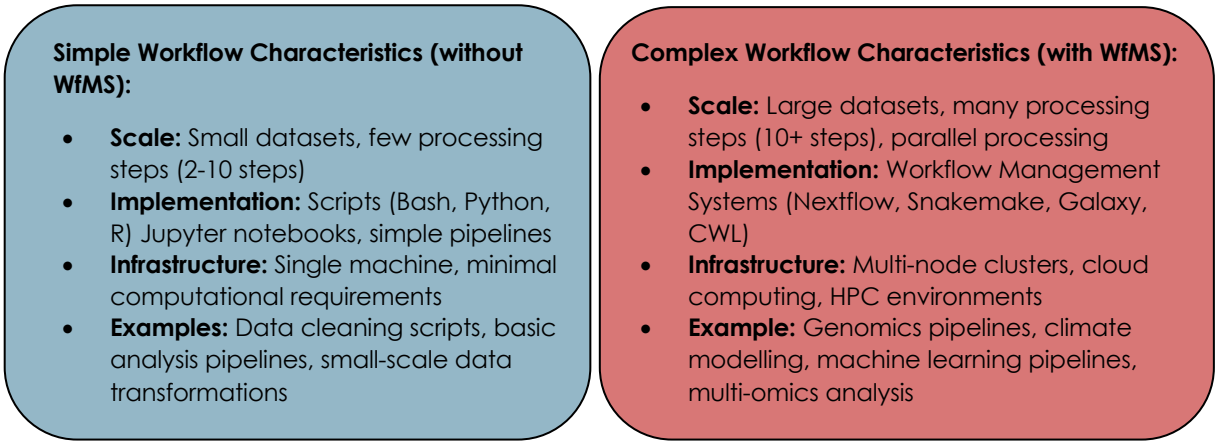
Next, we estimate the time investment required to develop a workflow. This is not straightforward and requires simplification. Computational workflows exist on a spectrum of complexity. Production workflows for large scale data collection and management are quite different to quick scripts written for simple data cleaning purposes. For example, the MGNify workflows, used by EMBL-EBI, for high-throughput microbiome data processing have taken 12 years to develop and are continuously being updated and maintained. Similarly, the Biodiversity Genomics Europe (BGE) project[87], which has spent 3 years developing its 17 workflows registered in WorkflowHub. These examples highlight that complex pipelines are constant developments that require incredible time investment. This heterogeneity in workflow complexity makes accurate estimation of the time investment extremely difficult.

One possible approach is to estimate an average development time; however, due to the vast heterogeneity discussed, such an approach risks misrepresenting true variation. Instead, we introduce differentiation by classifying workflows as "simple" or "complex", based on the Workflow Complexity Spectrum laid out in FAIR Principles for Computational Workflows. The guide, which references WorkflowHub as a key FAIR registry, differentiates between workflows based on their typical characteristics, with the defining factor being the implementation environment. Specifically, "complex" workflows require a Workflow Management System (WfMS) for execution, whereas "simple" workflows do not. Figure 9 below shows the typical characteristics of workflows in each category given in the guide.

[86] ONS, Annual Survey for Hours and Earnings (ASHE), 2024

[87] Biodiversity Genomics Europe

Figure 9 Classification of workflow complexity



Source: FAIR Principles for Computational Workflows

In WorkflowHub, published workflows are distinguished by type/language (Galaxy, Nextflow, Python, Jupyter etc.). This allows us to classify published workflows by complexity and calculate a proportion of workflows that belong in each classification. Out of a total 1,309 workflows in WorkflowHub, 1,286 were publicly visible. Table 7 below shows the number and proportion of workflows in each classification.

Table 7 Summary of workflow classification

Total public workflows	Simple	%	Complex	%
1,286	146	11.4%	1,140	88.6%

Source: Technopolis (2025) based on characteristics described in FAIR Principles for Computational Workflows

The FAIR Principles provide indicative ranges for development time: "hours to days" for simple workflows and "weeks to months" for complex ones. We assume a simple workflow requires a development time of 8 hours (1 working day) with a complex workflow requiring 180 hours (22.5 working days or 1 working month[88]). We use the upper tail of the indicative range given in the FAIR Principles to account for the existence of heavily complex workflows, such as examples previously discussed.

The cost of developing a "simple" workflow is calculated as the product of assumed time investment and average hourly wage. In this framework, we implicitly assume no knowledge spillovers in workflow production. This is a strong and unrealistic assumption, as workflow development is cumulative and often builds on existing tools. However, it is necessary for quantification and is consistent with methods used in previous OSS valuation studies.

Using the workflow development cost estimates, we calculate the replacement cost of registry contents through a weighted summation across simple and complex workflows. This yields the supply-side production value, excluding the platform's infrastructure costs.

Demand-side (usage) value

[88] On average, UK employees work between 20-23 days a month

To estimate the demand-side (usage) value of the WorkflowHub, we employ workflow access data provided by the WorkflowHub project team. Specifically, we use the number of downloads as a measure of workflow usage or demand. Table 8 below provides a summary of download data used in estimates.

Table 8 Downloads from WorkflowHub

Total Downloads	Total (Classifiable) Workflows	Downloads per Workflow
1,022,732	1286	795.28

Source: WorkflowHub project team

Workflow download volume is an imperfect measure of access because users may also view workflows, access external repositories, or run workflows on the platform itself.[89] Therefore, our usage estimates are likely conservative.

To translate the supply-side cost into the demand-side value, we multiply the content development cost by downloads per workflow. This assumes that the user's next best alternative is self-developing the downloaded workflow from scratch – a strong assumption but consistent with OSS valuation methods and appropriate for estimating the value of Open Science, rather than evaluating the platform itself.[90].

Main Results

Results for the estimation of both supply and demand-side value are presented in **Error! Reference source not found.** below.

Table 9 Estimates for supply and demand-side values of WorkflowHub contents

Supply-side (replacement) cost	Demand-side (usage) value
£4.29m	£3,412m

Source: Technopolis (2025)

The first column shows the estimated replacement cost of the registry contents, equal to the estimated cost to recreate the entirety of the registry contents once. The replacement cost can be thought of as a scenario in which the concept of Open Science (or the FAIR sharing of computational workflows) exists, but all registry contents have been lost. The second column in Table 12 presents the corresponding demand-side estimate. Demand-side value is the approximated cost of all users of a workflow self-developing from scratch, and can be thought of as the cost savings generated in comparison to a counterfactual scenario in which the concept or practice of the FAIR sharing of computational workflows ceases to exist.

The estimates presented in Table 12 are highly sensitive to the assumed time investment which, due to the heterogeneity in workflow complexity, cannot be estimated with certainty.

Economic Benefit

The economic benefit of the registry can be considered as the aggregate user surplus generated through open access to published workflows. Since usage value far exceeds

[89] For example, a Galaxy workflow can be executed directly on a Galaxy instance such as Galaxy Europe. The Galaxy Europe Instance hosts 163,521 workflows, has a user base of 137,256, and a total of 600,255 workflow runs.

[90] The Value of Open Source Software, Hoffmann, Nagle, and Zhou, 2024

replacement cost, the economic benefit is substantial. Under the assumptions discussed above, the benefits outweigh the costs by £3.4bn.

As the Workflow Hub user base continues to grow, the added value generated through scientific collaboration is also expected to increase. When viewed in the context of the wider EOSC-Life budget (EUR 26.1m), the disproportionate benefits of open workflow sharing become clear. This reflects the non-rivalrous nature of Open Science outputs, where initial investment yields social values that scales with repeated access and reuse.

Conclusions

This analysis identifies significant economic benefits of a single registry for computational workflows. Open access workflow registries share characteristics with OSS, and therefore similar valuation methods apply. Although attaching a monetary value to a diverse set of workflow types requires simplification (and likely underestimates true value), the results consistently show that usage value vastly exceeds development cost.

The study focusses purely on the registry contents and makes no attempt to analyse the impact of the underlying infrastructure required to execute computational workflows. Investment in such infrastructure is complex and collaborative, but the WorkflowHub is both a product of and contributor to these collaborative systems, and its benefits demonstrate the considerable economic gains enabled by shared computational resources.

C.4 Open-Source Software – R

Headline Findings

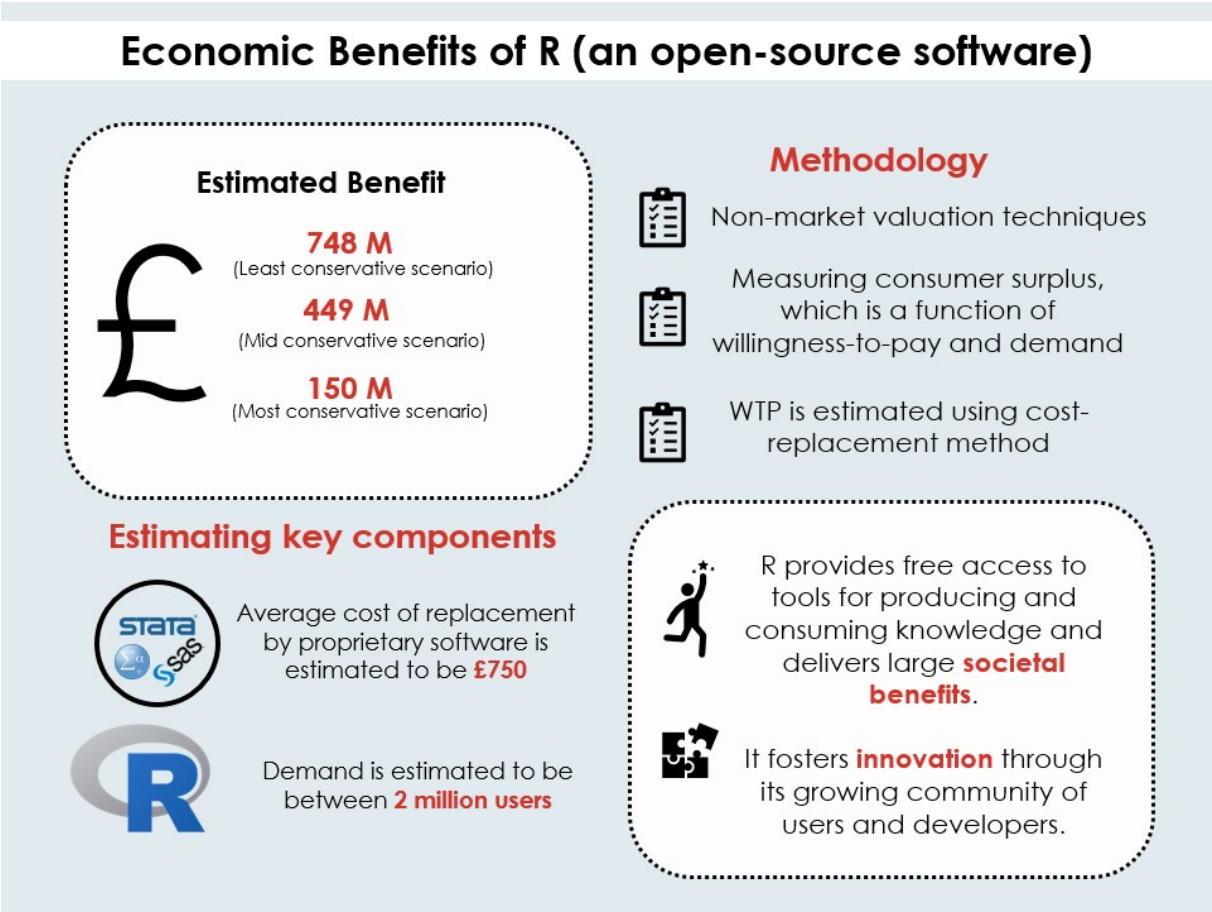
R is an open-source programming language and environment primarily used for statistical computing, data analysis, and visualisation. Its adoption is steadily increasing across academia and industry, with notable applications in sectors such as biotech, finance, and high technology.

R offers significant educational and societal benefits by providing free public access to tools for producing and consuming knowledge. It fosters innovation and collaboration through its growing community of developers and users, who continuously enhance R's core functionality and offer efficient solutions to existing challenges. Additionally, R is a cost-effective tool for both commercial and non-commercial use.

As an open-source platform, traditional evaluation methods often fail to capture its full value. This study applies non-market valuation techniques to estimate the economic benefits derived from R. These benefits are measured as consumer surplus, calculated based on demand (number of users) and willingness to pay (cost of alternative proprietary software).

There are an estimated 2 million R users. Willingness to pay is approximated at £750, based on the pricing of proprietary alternatives such as STATA, SAS, and SPSS. To account for varying user preferences, three substitution scenarios are considered (50%, 30%, and 10%).

Overall, the economic benefits are estimated to range from £150 million to £748 million, even under conservative assumptions. This case study highlights that even modest estimates of R's replacement value emphasise the substantial economic impact of Open Science.



R (Software)

What is R?

R is an open-source programming language and environment primarily used for statistical computing, data analysis, visualisation and more. It is widely adopted in academia, research, and industry[91]. A broad range of industries have adopted the R language, including biotech, finance, research and high technology industries, and its usage is continuously growing[92].

As an open-source software (OSS), R is freely available for anyone to use, modify, and distribute. Its core functionality is extended by a large number of software packages, which contain reusable code, documentation, and sample data. These packages are typically installed through the Comprehensive R Archive Network (CRAN), R's centralised software repository. CRAN is an archive of the latest and previous versions of the R distribution, documentation, and contributed R packages. While CRAN is the most well-known and widely recognised repository, other platforms such as GitHub also host a large volume of code and contributors (although not exclusive to R).

A large and growing community surrounds R. The R Consortium acts as a central hub for the R community bringing together industry leaders, academic institutions, and individual contributors to support the development of R. Industry giants such as Google and Microsoft are

[91] An Empirical Analysis of the R Package Ecosystem, Bommarito and Bommarito, 2021

[92] R Consortium, 2024

members of the consortium, underscoring R's widespread adoption. Beyond corporate involvement, individual users form a significant part of the ecosystem: there are at least 92 groups across 37 countries with over 79,000 members[92]. Numerous conferences, webinars, and forums further enrich this ecosystem and support development and learning.

Since there is no dedicated support for R, as is often the case with OSS, the R community also extends support through forums, free packages ad tutorials. This support is invaluable for troubleshooting and learning for users, and further development and adoption of the language.

How is R delivering impact?

R, being free and open-source, benefits everyone with access to tools to produce and consume knowledge, regardless of economic means. This democratisation of access benefits students, researchers, and professionals across the globe to engage with analytical software without the barrier of expensive licensing fees. By removing cost constraints, R contributes to more inclusive knowledge production and consumption, and delivers large educational and societal benefits.

R is also an environment that encourages innovation. R developers are continuously improving the software and expanding its scope. They develop 'packages' which extend R's utility and as well maintain active ones to provide more efficient solutions. A comprehensive empirical study of the R ecosystem found robust growth in all kinds of packages - active packages grew at a compound annual rate of 29%, new releases at 28%, and active maintainers at 26% - in the last two decades[91]. This shows continuous innovation in the language.

Moreover, it highlights collaboration between members and interconnectivity within the ecosystem. The study also showed that development of 60% of packages rely on at least one other package[91]. The platform is evolving through collaboration and contributions from a global community and is an active space for innovation.

For organisations, especially startups and SMEs, R offers substantial cost advantages over proprietary alternatives. By eliminating licensing fees, businesses can redirect resources toward product development or scaling operations. Similarly, pharmaceutical companies are also adopting R to reduce research and development costs[93]. R's flexibility and customizability allow tailored solutions without incurring additional costs, making it a financially efficient choice for both commercial and non-commercial use.

What are the alternatives to R?

While R is a powerful and widely used language, this case study does not aim to provide a comprehensive comparison with all available alternatives. Instead, it highlights a few notable options, recognising that each has distinct strengths and limitations. Python, also an open-source programming language, has become a strong alternative for data science projects in recent years. Proprietary alternatives include SAS, SPSS and STATA. SAS is a professional statistics software that is commonly used in biometrics, clinical research, and in the banking sector[94]. SPSS is considered to be particularly easy to use, and is one of the most widely-used statistics programs[95]. It was the most popular software but has seen a decrease in usage over the years.

---

[93] SAS to R Migration in the Pharma Industry, Bajuk, 2025

[94] The Software Behind the Stats: A Student Exploration of Software Trends Across Disciplines, Upton et al. 2025

[95] Trends in the Usage of Statistical Software and Their Associated Study Designs in Health Sciences Research: a Bibliometric Analysis, Masaudi et al. 2021

STATA is a commercial statistical software that is particularly favoured by econometricians[96]. These three software comprise a bulk of popular alternatives[97].

Approach to estimating economic benefits

The conventional method to evaluate software is in terms of the discount value of the future income stream earned through subscriptions. This methodology poses a challenge because OSS is free to access and use. Consequently, non-market valuation methods are employed. For example, using lines of code and a standard model to estimate package developer time, a study estimated that the resource cost for developing R exceed $3 billion dollars, based on 2017 costs[98]. Measuring the benefits of OSS is inherently challenging due to lack of reliable data, user demographics, and engagement. This study explores an alternative approach to capture the non-market use value of R.

What is the methodological approach?

This case study estimates the economic value of R by assessing the demand for it and the associated consumer surplus. In the case of OSS, the consumer surplus can be interpreted as the difference between user's willingness-to-pay and the price paid (which is zero)[99]. Willingness-to-pay for R is estimated using replacement cost method, i.e., the cost of obtaining same services from an alternative software. This method requires information about the price of replacement goods which are proprietary software for data analysis. Thus, there are two key components in the study:

- Cost of replacement (proxy for willingness-to-pay)
- Number of users (proxy for demand)

Estimating Economic Benefits

Estimating cost

Stata, SAS and SPSS are the three proprietary alternatives considered in this study. To estimate their cost, the price of a basic, single-user annual licence was collected from each software's official website. For Stata and SPSS, both professional and student pricing were obtained. Prices were either in pound sterling or converted from USD using Bank of England's exchange rate.

To calculate at an estimate for replacement cost, two steps were undertaken. First, a **50:50** professional-to-student ratio was assumed. While the actual demographic split is unknown, this equal split is a very conservative estimate of student usage. Companies are likely to sell more professional licences than student licences because professionals drive revenue, students often have access through their universities and do not have to buy individual licences, and professionals are more likely to buy new licences regularly.

Second, a **0.56: 0.22: 0.22** proportion for SPSS: Stata: SAS was calculated using data from three different sources. Although each source had limitations, ranging from overly narrow academic contexts to broad search-based data, the similar trends across all three lent credibility to the estimates. The three sources were:

---

[96] Trends in the Usage of Statistical Software and Their Associated Study Designs in Health Sciences Research: a Bibliometric Analysis, Masaudi et al. 2021

[97] Trends in the Usage of Statistical Software and Their Associated Study Designs in Health Sciences Research: a Bibliometric Analysis, Masaudi et al. 2021

[98] Open-Source software as intangible capital: Measuring the cost and impact of free digital tools, Robbins et al. 2018

[99] Valuing the User Benefits of Companies House Data, BEIS, 2019

- A study of software use in medical literature100
- Wikipedia search trends (Sep 2024 – Sep 2025)
- Google trends data (Sep 2024 – Sep 2025).

Using weighted average of usage proportions and licence costs across professionals and students, the **estimated replacement cost** was calculated to approximately **£750**. This figure is likely conservative, as it reflects the cheapest available licence types and assumes a high proportion of student users.

Estimating R users

The R Consortium, a non-profit organisation that supports the R language, and Oracle have independently estimated that there are "**over 2 million**" R users worldwide. Oracle published this in 2014 and since then R users have grown. These estimates are often cited in online blogs and websites, though the specific methodology behind the figure is not transparent. Nevertheless, in this study, **the demand for R (measured in terms of the number of users) is estimated to be about 2 million.**

Other public sources indicate a wide range, from 760,000 to 5 million users, reflecting different assumptions and data sources (See Table 10)

Table 10 Estimates of global R users

Estimate	Approximate Users	Source
Lower Bound	760 K	GitHub R User Community
Lower-Mid Bound	2 M	R Consortium and Oracle
Mid-High Bound	5 M	Stackoverflow and CRAN downloads

The most conservative estimate comes from GitHub's R User Groups, reporting 766,364 active members across 890 groups in 86 countries. This number likely underrepresents the global R community, as it only includes users affiliated with a user group on GitHub. Many R users may not participate in such groups or be a part of other forums making this a strict lower bound.

The upper bound of 5 million is derived from two indicators:

Stack Overflow: Stack Overflow, one of the largest community forums for seeking help with R, has 512,313 questions tagged "R". Applying "90-9-1" rule for online participation (where only 1% of users actively contribute content) suggests a ser base of at least 5 million. This is a conservative estimate, as it excludes users who ask about R under other related tags (like ggplot2, dplyr, etc.) or who use other platforms entirely.

CRAN Downloads: The "tidyverse" package, a popular collection of R packages, alone has 88 million downloads from a single CRAN mirror. Assuming a conservative ratio of 1 in 20 downloads representing a unique user implies a user base of around 4.4 million. This estimate does not include downloads from other CRAN mirrors or users who rely on other R packages, which would increase the total further.

- Masuadi, E., Mohamud, M., Almutairi, M., Alsunaidi, A., Alswayed, A. K., Aldhafeeri, O. F., & Almutairi, M. B. (2021). Trends in the usage of statistical software and their associated study designs in health sciences research: a bibliometric analysis. Cureus, 13(1).

An estimate of 20 downloads per user is chosen based on trends seen in the literature. The pattern of R downloads and the R package downloads is similar and upwards[101]. There is an upward bias in downloads due to updates and dependency downloads which inflate download counts[102]. A lot of users are more likely to only download a package a few times, 20 downloads per user is probably a conservative estimate.Combing the estimate from both Stackoverflow and CRAN downloads, which reveal very similar figures, the upper bound is taken to be approximately 5 million users.

Estimating benefit

The economic benefits of R is calculated by multiplying cost of alternative sources with the number of users. However, we recognise that the willingness to pay would not be the same for all users and a many would not be able to switch to paid alternatives. To account for this, a substitution effect framework is applied, varying the proportions of users who might switch. Three substitution scenarios are considered:

- High substitution (50%)
- Medium substitution (30%)
- Low substitution (10%)

Robust empirical literature on substitution effects in open-source software is limited. While some studies exist in related areas such as "freemium" business models (where core features are free and other features are paid) and time-limited free trials, these models are structurally different from open-source software and not directly comparable.

Table 11 below presents economic benefits using the substitution effect framework. Higher substitution rates yield greater economic benefit of approximately **£748 million**. The benefits are slightly slower with substitution effect of 30%: approximately **£449 M.** The least benefits, about **£150 M**, is estimated when substitution effect is 10%. Applying very conservative assumptions also yield large economic benefits. It reinforces the importance of both access and usage of R.

Table 11 – Total benefit estimates in different scenarios

Substitution Effect	Economic Benefits
Substitution: 50%	748.4M
Substitution: 30%	449.1M
Substitution: 10%	149.7M

Source: Technopolis (2025)

Conclusions

This case study set out to estimate the economic benefits of R, an open-source statistical software, by applying a non-market valuation approach. As traditional market-based evaluation methods do not capture the true value of freely available software, the analysis focused on estimating consumer surplus. It is the difference between what users are willing to pay and the actual cost (in this case, zero).

[101] What makes an R package popular? Zhang, 2021

[102] What makes an R package popular? Zhang, 2021

The willingness to pay was estimated at £750 per user, using a cost-replacement method based on comparable proprietary software. The number of R users was estimated to be approximately 2 million. Recognising that not all users would be willing or able to purchase commercial alternatives, a substitution framework was introduced to model varying levels of demand. Three scenarios, low (10%), medium (30%), and high (50%) substitution, were constructed to reflect different user behaviours and contexts.

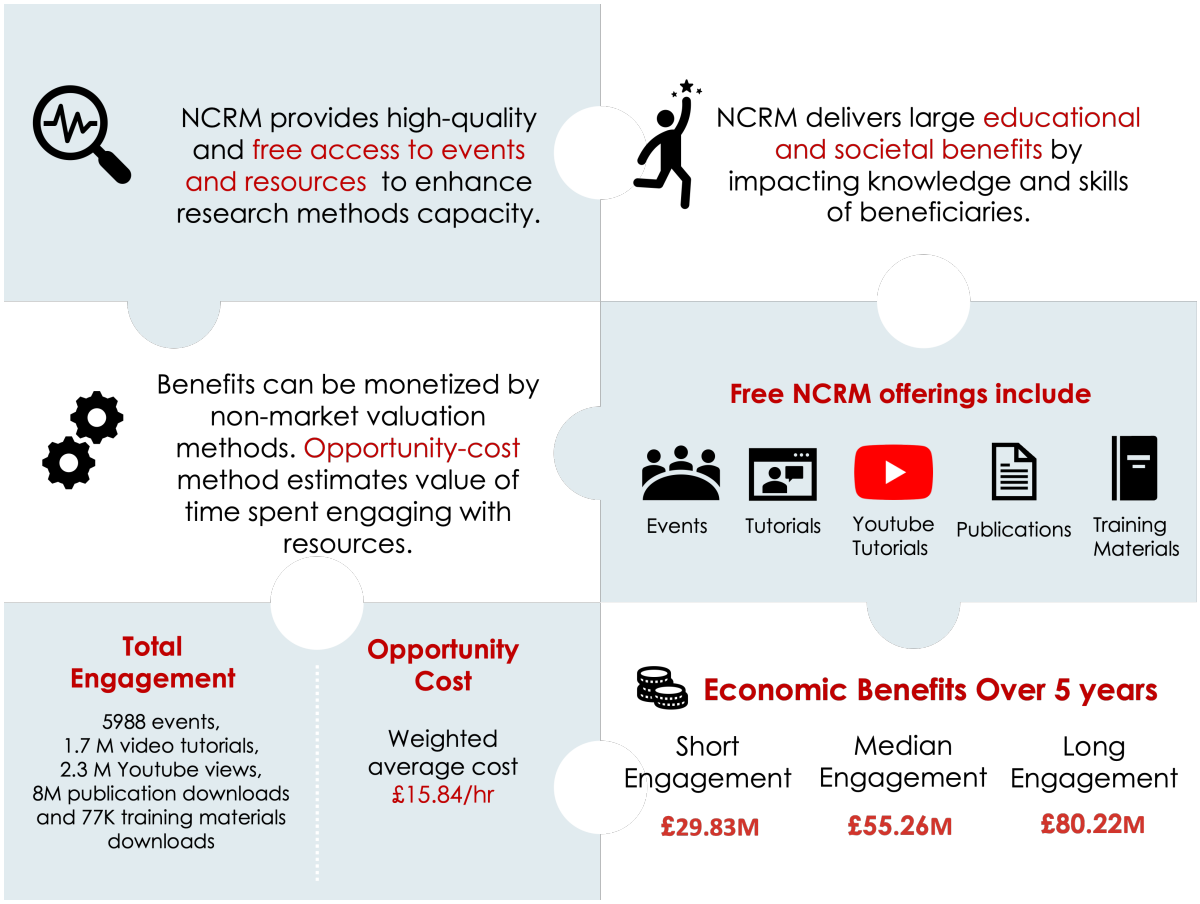
Across all scenarios, the estimated economic benefit is substantial, ranging between £150-£748 millions of pounds. The value of the case study is not just in the estimated benefits, but also in building a case for open-science. Open Science, such as OSS, drive innovation, collaboration and accessibility. Ultimately, this case study illustrates that even conservative estimates of R's replacement value underscore the broader economic value of Open Science.

C.5 National Centre for Research Methods (NCRM)

Headline Findings

The National Centre for Research Methods provides high-quality and free access to events and resources to enhance research methods capacity in the UK. NCRM's activities and resources have a strong positive impact on knowledge and skills of beneficiaries. It delivers large educational and societal benefits. The easy access to resources by NCRM exemplifies principles of Open Science.

Using opportunity-cost method, a non-market valuation tool, annual economic value of benefits derived from NCRM's activities are estimated to between £ 6M – 16M. Between 2020 and 2025, total estimated economic benefits are in the range of approximately £30M - £80M. This is likely to be a conservative estimate, and the benefits are believed to be greater.



NCRM

What is NCRM?

The National Centre for Research Methods (NCRM) is a UK-based organisation that provides training, resources, and support to enhance research methods capacity across the social sciences. Established in 2004 and funded by the Economic and Social Research Council (ESRC), NCRM serves as a central hub for disseminating best resources and practices in research methodology. Now in its fourth phase of funding (which began in 2020), it focuses on meeting the diverse and changing needs of social science researchers (and other types of researchers) through informed pedagogy and high-quality training and capacity building provision.

NCRM is led by a consortium of three leading universities: the Universities of Southampton, Manchester and Edinburgh. This core team is supported by a network of strategically selected Centre Partners, who contribute subject-specific expertise and help deliver training across disciplines and sectors.

Who are the beneficiaries?

The Centre's beneficiaries are drawn from a broad spectrum of backgrounds. While the majority are academic researchers working within the social sciences, NCRM also serves professionals from the public, private, and voluntary sectors. Approximately 20% of participants come from disciplines outside the social sciences, such as medical sciences, and around 14%

are from non-academic settings[103]. Over time, NCRM has seen increasing engagement from individuals working in government, policy, and community organisations.

What are the main activities?

NCRM's training and capacity building activities are diverse and inclusive. They include short courses, seasonal schools, bootcamps, webinars, conferences, and annual lectures. The Centre also produces a breadth of online resources, such as videos, podcasts, toolkits, and guides, that are freely accessible to the public.

In addition, NCRM facilitates collaborative events such as Innovation Fora, which bring together researchers from different sectors and career stages to explore methodological advancements. It actively supports and builds Methodological Special Interest Groups (MSIGs) and Communities of Practice to enable collaboration and knowledge sharing. Doctoral Training Partnership Training Network (DTP-TN) and Data Resources Training Network (DR-TN) are two such examples[104]. The activities are also relevant to evolving research landscapes; f or example, a recent TCB activity on Responsible AI was held.

A distinctive feature of NCRM's work is its accessibility. NCRM maintains a low fee structure and provides bursaries, enabling participation from participants may not be able to attend, particularly those from smaller institutions. Beyond its formal training courses, NCRM provides a wide range of events and resources that are entirely free to access. This includes many capacity building events, lectures on research methods, online publications, video tutorials, guides for trainers and many informative videos and podcasts.

How is NCRM delivering impact?

NCRM's activities and resources have a strong positive impact on knowledge and skills of beneficiaries. The new or improved skills of beneficiaries, which includes but is not limited to research, training, teaching, supervision, methods development, also have a further positive effect on their work and on others they engage with[105]. Since beneficiaries belong to different sectors, including academics, private, public, and volunteer, NCRM's activities and resources are also shaping the methodological landscape across and beyond social sciences. The activities and resources are high-quality, well taught and impactful.

How does NCRM support Open Science?

NCRM exemplifies the principles of **Open Science**. Its high quality and freely available events and resources align with broader Open Science values. The Centre's predominantly online delivery model ensures inclusive access, regardless of geographical or institutional constraints. Through its efforts to increase capacity by identifying shared research or training interests, knowledge exchange, and developing new networks and collaborations, NCRM demonstrates a strong ethos of openness, transparency, and collaboration which are hallmarks of Open Science in practice.

Approach to estimating economic benefits

Which activities are included in Open Science benefit analysis?

This study estimates the economic value generated by NCRM's Open Science initiatives by focusing on a defined set of freely available resources and activities. While the list below is exhaustive for the purposes of the analysis, it does not represent the full breadth of NCRM's

[103] Impact Assessment Report, NCRM, 2025

[104] Impact Assessment Report, NCRM, 2025

[105] Impact Strategy Framework, NCRM, 2022

work. Training programmes, although central to NCRM's mission, are low-cost rather than free and therefore excluded from the analysis. Additionally, some freely available resources, such as podcasts, have been excluded due to limitations in available data. The activities included in the analysis are:

- **Events** (both online and in-person focused on capacity building, networking, lectures, workshops)
- **Online Tutorials** (structured video series with supporting materials explaining various research methods.)
- **YouTube videos** (tutorials delivered through one or more videos, recordings of lectures and workshops)
- **E-publications** (freely accessible digital publications on research methods and related topics)
- **Training guides** (resources for trainers)

What is the methodological approach?

This analysis employs the **opportunity-cost method** to estimate the economic value of NCRM's open-access resources. Engagement is valued according to the opportunity cost of participants' time, reflecting their willingness to forgo alternative activities such as paid employment (professionals) or education (students).

The economic benefit of this engagement can be estimated using two key components:

- Total engagement (the cumulative time participants spend interacting with freely available resources)
- Opportunity cost of time (The value of participants' time, estimated based on alternative productive activities)

The analysis draws on data from NCRM's internal records and website analytics. It focuses on engagement from 2020 onwards, due to data availability.

Estimating Economic Benefits

Estimating engagement

This analysis focuses exclusively on NCRM activities that are freely available and accessible. These include online and in-person events, video tutorials hosted on the NCRM website and YouTube, and downloadable resources such as e-publications and training materials for educators. The total engagement across these offerings is summarised below in Table 12.

Table 12 Total engagement with NCRM's freely available offerings

Offering	Engagement (between 2020-2025)
Events (both online and offline)	5988 participants (319 in-person and 5669 online); Event durations ranged from 2 hours to 2 days (up to1 6 hrs)
Video tutorials	1,713,700 views, totalling 155900 hours of watch time
Youtube videos	Approximately 2.3 M views, totalling 216,100 hours of watch time
E-publications	8,272,162 downloads
Materials for trainers	77,184 downloads

Source: NCRM data

In the table above, the absolute and **total time of engagement with videos** (both from the website and Youtube) is incorporated into the economic model. Engagement with video tutorials and YouTube content is measured using platform analytics, which track actual watch time. This data provides a direct and reliable estimate of how long users spent engaging with the material. Because it reflects real-time usage, no further adjustment is made in the economic model.

For events, engagement is estimated based on the **scheduled duration of each session**. Events are grouped into three categories: short (2 hours), median (6 hours) and long (16 hours) engagement. These estimates are conservative. In reality, participants may spend additional time preparing for or traveling to events, especially in-person ones. However, due to limited data on these ancillary activities, only the core event duration is used in the model.

Downloads present a more complex challenge for estimating engagement. Absolute totals are likely to be overestimates of actual engagement because every download is not a meaningful interaction with content. Participants might download resources but may fail to engage with the material. The quality of engagement may also vary with some participants skimming through and others reading in more detail. To account for this uncertainty, **three scenarios are constructed**: short engagement (25% of downloads are read), median engagement (50% of downloads are read) and long engagement (75% of downloads are read).

Using existing literature, we assume participants spend **59.7 mins** per download[106]. This is based on an international study on reading patterns of researchers. A third of participants in the study were social scientists and two-thirds worked in academia. Other participants were researchers from other disciplines and non-academic sectors such as industry, non-profit and government. Given the international scope and relevance of the study population, this estimate is considered appropriate for NCRM's audience. Moreover, since NCRM resources often contain multiple articles or extended content, the 59.7-minute estimate is likely conservative.

Estimating opportunity costs

The beneficiaries of NCRM's freely available resources and events are broadly classified into two groups: students and professionals. According to survey data from a training event, 56.6% of participants were students, while 43.4% were professionals. Although detailed demographic information is not available for users of other NCRM offerings, this proportion is assumed to apply across all categories of engagement for the purposes of this analysis.

To estimate the economic value of engagement, we apply the opportunity cost method, which requires assigning a monetary value to the time spent by participants. For professionals, the opportunity cost is based on the median hourly wage in the UK, reflecting the value of time they could otherwise spend in paid employment.

For students, the opportunity cost is also considered in terms of foregone earnings. However, students typically have less work experience and may be more likely to engage in part-time or lower-wage employment. Therefore, their opportunity cost is conservatively estimated using the UK minimum wage.

[106] Seeking, Reading, and Use of Scholary Articles: an International Study of Perceptions and Behaviour of Researchers, Tenopir et al. 2019

Using the assumed 56.6% student and 43.4% professional split, a **weighted average opportunity** cost of **£15.84 per hour** is calculated[107] [108]. This figure is applied across all NCRM offerings to estimate economic benefits, with one exception.

For resources designed specifically for trainers, it is assumed that all users are professionals. This is based on the nature of the materials, which are intended for individuals delivering research methods training, and are therefore more likely to be accessed by professionals working in academic or applied research settings. Accordingly, the professional wage rate is used exclusively for this category.

Estimating benefit

Using this framework, the total economic benefit is calculated by multiplying the estimated time spent engaging with each resource by the corresponding opportunity cost. Since engagement varies across users and offerings, three scenarios are constructed to reflect different levels of interaction: short, median, and long engagement. Based on these scenarios, the estimated economic benefits generated by NCRM's Open Science activities between 2020 and 2025 are as follows:

- **Short engagement**: £29.83 million
- **Median engagement**: £55.26 million
- **Long engagement**: £80.22 million

These figures represent minimum estimates, given the conservative nature of the modelling. For example, time spent preparing for events or re-watching videos. Additionally, the model does not account for low-cost trainings and qualitative benefits such as improved research capacity or long-term career development. As such, the actual economic value of NCRM's contributions is likely to be significantly higher than the figures presented here.

Conclusions

This case study shows that NCRM generates high economic value through its Open Science activities. The economic benefits range from **£29.83 M to £80.22 M** between 2020 and 2025, depending on the level of engagement. This is equivalent to annual benefits of £6M - 16M. The figures do not capture the full breadth of NCRM's activities and consequently underestimate the economic benefits. The findings affirm the significance of investment in Open Science.

Next steps for NCRM

The findings of this study rely heavily on NCRM's internal data and web analytics. It underscores the importance of engagement data in estimating economic benefits. In the future, NCRM could collect more granular data about participants (such as demographics), their engagement (including new offerings such as podcasts) and post-engagement outcomes (skill development).

Given that trainings are a key offering by NCRM, it could consider a shift from low-cost model to open-access format. Since the trainings are often by external guests, it can be hard to cover their fees. However, NCRM could advocate for increasing funding under Open Science initiatives. These findings can support those discussions and advocate for greater Open Science strategies.

---

[107] Employee earnings in the UK, ONS, 2024

[108] What is the real Living Wage? Living Wage Foundation

technopolis
group

www.technopolis-group.com