

Open Data: Is It Worth It?

The Big Question

We are told to "share our data".

But it takes time, effort, and
resources.

So, we must ask the hard question:

**Is the reward *really* worth the
effort?**

Is it worth it for an individual researcher to share their data?

- time
- effort
- resources
- ???



- citations
- collaboration opportunities
- innovation
- ???

Open Data: Is It Worth It?

Insights from the PathOS Project

Elin Kronander at the Uppsala Open Science Community 2025-11-14

We Need Evidence, Not Just Anecdotes



3 years - 2022-09 – 2025-08

“PathOS was a Horizon Europe project that aimed to collect concrete evidence of Open Science effects and to study the pathways of Open Science practices, from inputs to outputs, outcomes, and impacts, including the consideration of enabling factors and key barriers.”

My connection to the PathOS-project:

- The ELIXIR Hub was a member of the project
- I attended a few training sessions

Key Outputs

<https://pathos-project.eu/>

Welcome to the PathOS Open Science Resources Hub

PathOS aims to be a gateway for resources on Open Science (OS) impacts and effective policy development. We offer essential tools, including the Open Science Indicator Handbook, a Zotero library tracking OS impacts, and a registry of OS Stories. Our expanding collection will feature impact measurement tools, frameworks for OS pathways and cost-benefit analysis, along with comprehensive training resources.

OS Indicator Handbook
Guidance on how to assess the impact of Open Science practices, focusing on the causal effects and challenges in measuring these impacts.

Evidence of OS Impact
A collection of over 600 relevant studies on Open Science, gathered through a comprehensive search of peer-reviewed materials and grey literature from recognized sources.

OS Stories Registry
A collection of real-world case studies and experiences that highlight the impact of Open Science practices in various contexts.

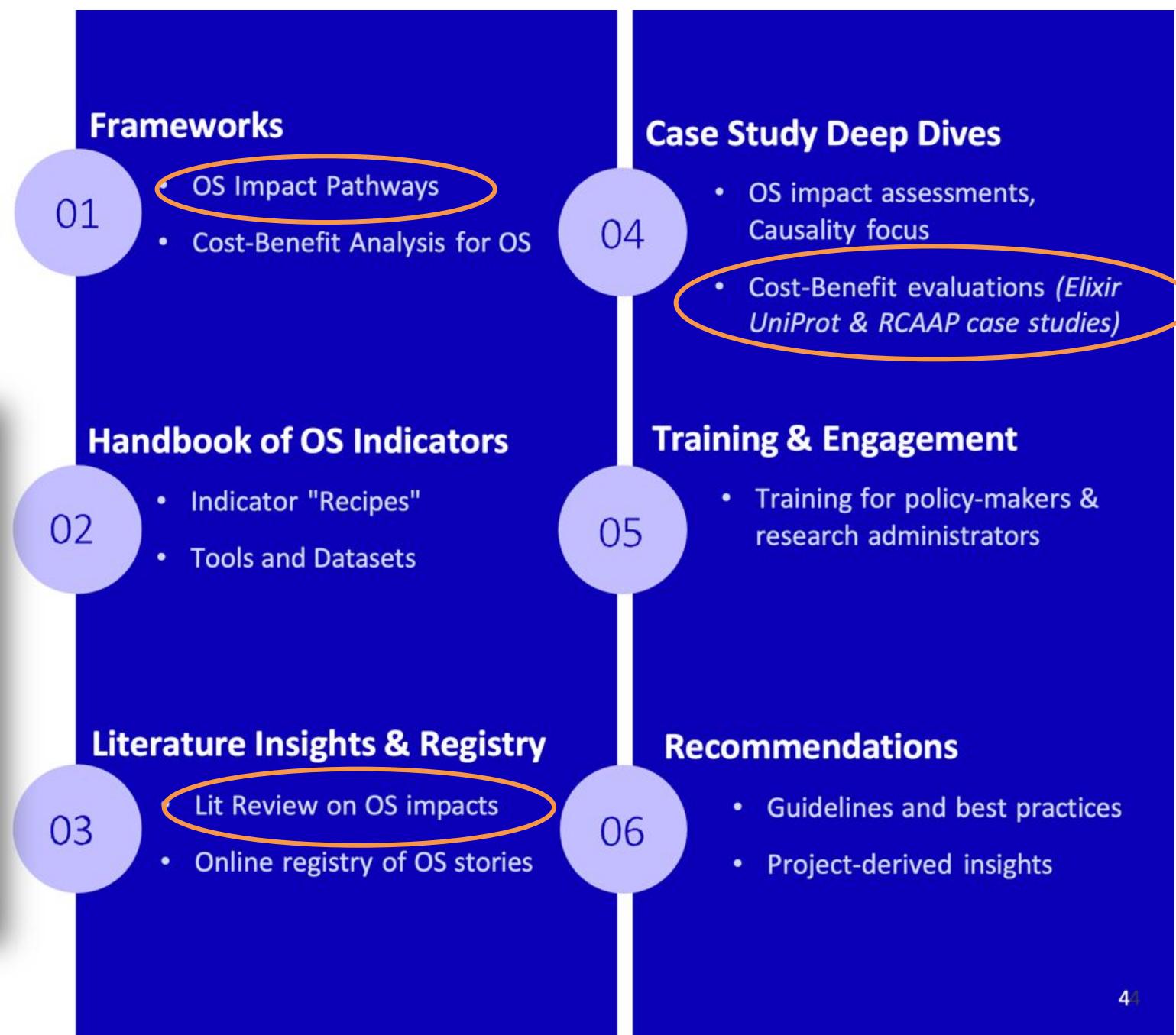
Cost Benefit Analysis for Open Science
Coming in Q3/2024
Cost Benefit analysis report for indicative case studies and a framework (model) for impact pathways for Open Science to help policy and decision-making bodies to better understand the rationale of OS and how their impacts materialize.

Policy Briefs
Coming in Q3/2024

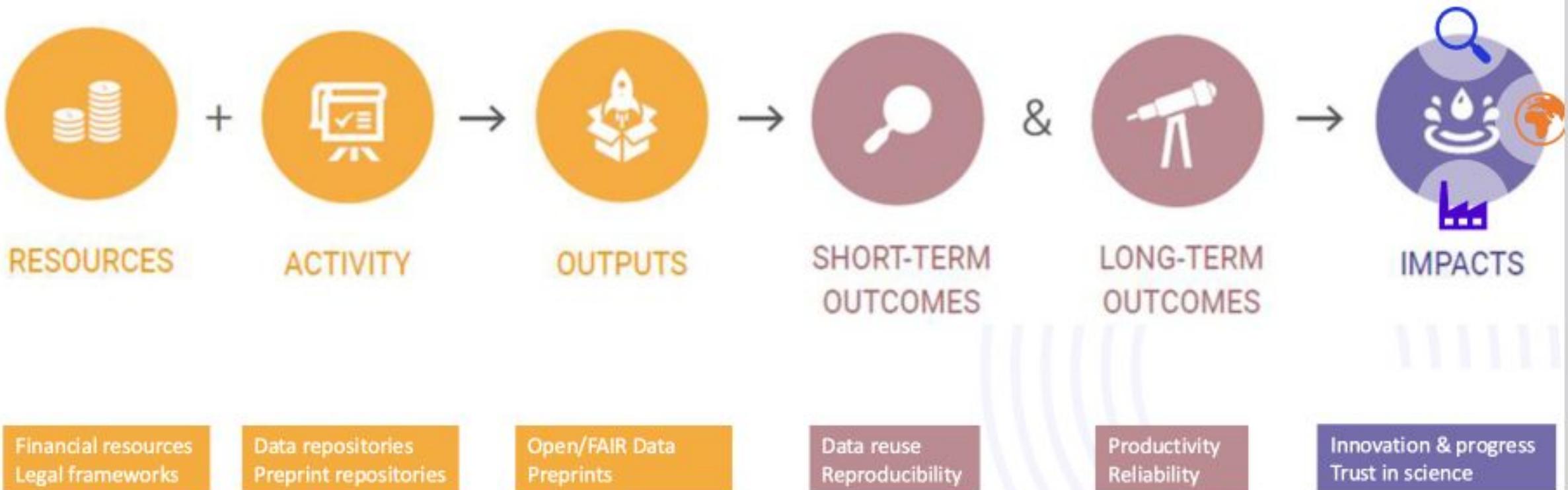
Tools and Data
Coming in Q3/2024

Training Programme
Coming in Q3/2024

Deliverables & Publications



Open Science Impact Pathways



Overview sections

Introduction to causality

Open Science

Impact

- Academic impact
- Societal impact
- Economic impact

Reproducibility

Different kind of impact

Academic Impact

Publishing

- Readership
- Citation
- Collaboration
- Diversity
- Interdisciplinarity / novelty
- Quality

Data

- Use of data in research

Software

- Use of software in research

Societal Impact

Uptake

- Societal issues
- (Social) media
- Policy
- Education
- Patient groups
- Medical practice
- Legal sector

Broader

- Scientific literacy

Economic Impact

Innovation

- Innovation output
- Science-industry collaboration
- Socially relevant products and processes

Costs & benefits

- Labour market impacts
- Economic growth
- Cost savings

Academic impact: 485 included studies

ROYAL SOCIETY
OPEN SCIENCE

royalsocietypublishing.org/journal/rsos

Review



Cite this article: Klebel T, Traag V, Grypari I, Stoy L, Ross-Hellauer T. 2025 The academic impact of Open Science: a scoping review. *J. Soc. Open Sci.* 12: 3128. <https://doi.org/10.1088/1749-4688/aca134>

Received: 16 August 2024
Accepted: 20 January 2025

Subject Category:
Science, society and policy

Subject Areas:
e-society

Keywords:
academic impact, Open Science, Open Access, FAIR Data, Citizen Science, scoping review

Author for correspondence:
Thomas Klebel
e-mail: tklebel@knoeckel.at

The academic impact of Open Science: a scoping review

Thomas Klebel¹, Vincent Traag², Joanna Grypari^{3,4},
Lennart Stoy⁵ and Tony Ross-Hellauer¹

¹Open and Reproducible Research Group, Know-Center Research GmbH, Graz, Austria
²Centre for Science and Technology Studies (CSTS), Leiden University, Leiden, The Netherlands
³Atmos Research Center, Münster, Germany
⁴OpenAIRE, Athens, Greece
⁵Teknopolis Group, Brussels, Belgium

DOI: [10.1088/1749-4688/aca134](https://doi.org/10.1088/1749-4688/aca134) | [10.5281/zenodo.1504003](https://doi.org/10.5281/zenodo.1504003)
IC: 0000-0002-7311-0751 | ET: 0000-0003-2100-3079
IE: 0000-0002-3485-1591 | LS: 0000-0002-3827-9875
TB: 0000-0001-4479-7027

Open Science seeks to make research processes and outputs more accessible, transparent and inclusive, ensuring that scientific findings can be freely shared, synthesized and built upon by researchers and others. To date, there has been no systematic synthesis of the extent to which Open Science (OS) reaches these aims. We use the PRISMA scoping review methodology to partially address this gap, scoping evidence on the academic (but not societal or economic) impacts of OS. We identify 485 studies related to all aspects of OS, including Open Access (OA), Open FAIR Data (OFA), Open Code/Software, Open Evaluation and Citizen Science (CS). Analysing and synthesizing findings, we show that the majority of studies investigated effects of OA, CS and OFD. Key areas of impact studied are citations, quality, efficiency, equity, reuse, ethics and reproducibility, with more studies reporting positive or at least mixed impacts. However, we also identified significant unintended negative impacts, especially those regarding equity, diversity and inclusion. Overall, the main barrier to academic impact of OS is lack of skills, resources and infrastructure to effectively re-use and build on existing research. Building on this synthesis, we identify gaps within this literature and draw implications for future research and policy.

Main research question:

What evidence exists in the literature regarding the effect of OS on the academic impact of research?

Secondary research questions:

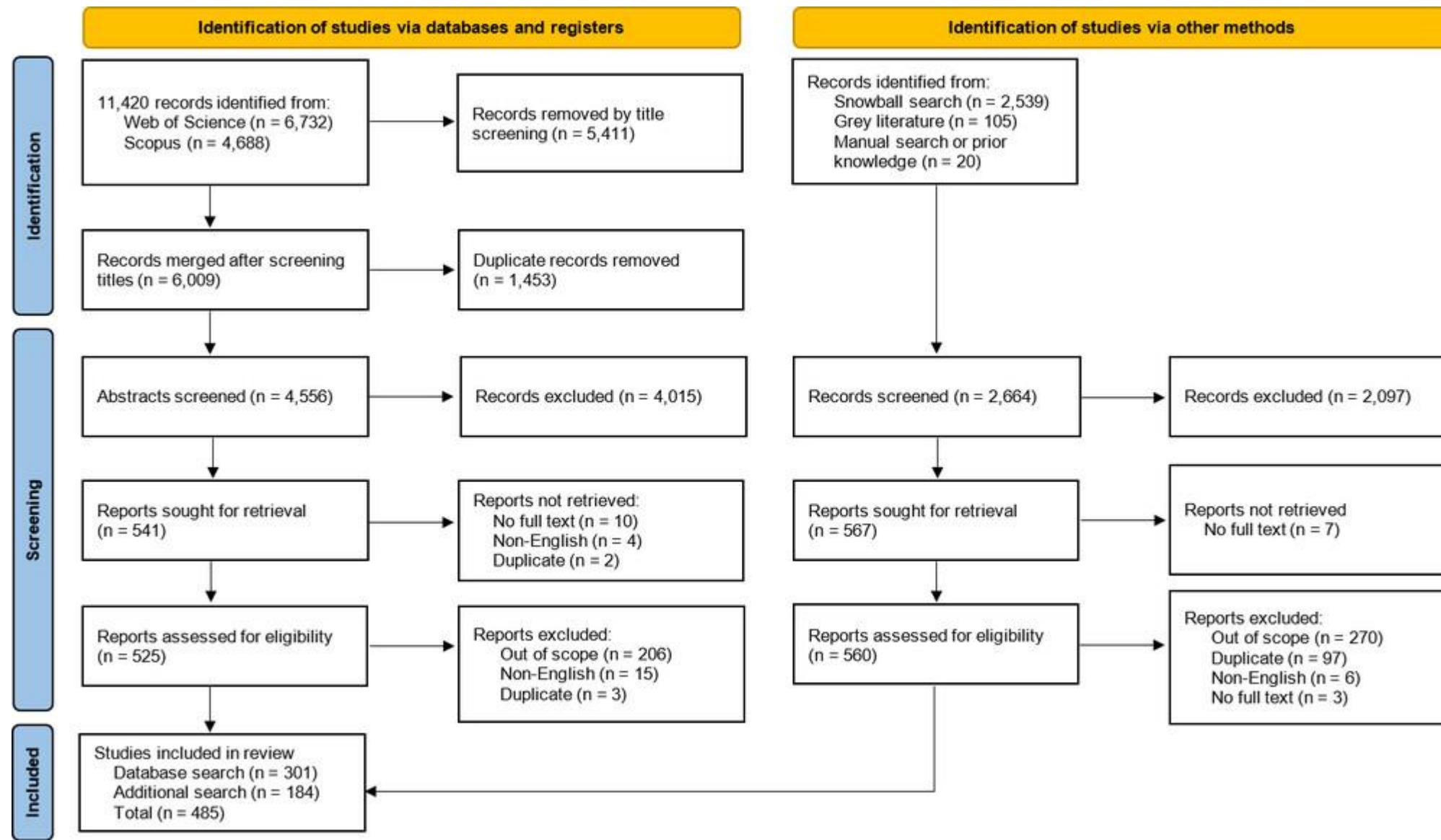
- What types of positive or negative, direct or indirect academic impact are observed?
- What kinds of mechanisms produce them?
- What specific enabling and/or inhibiting factors (drivers and barriers) are associated with these impacts?
- What knowledge gaps emerge from this analysis?

Pathos

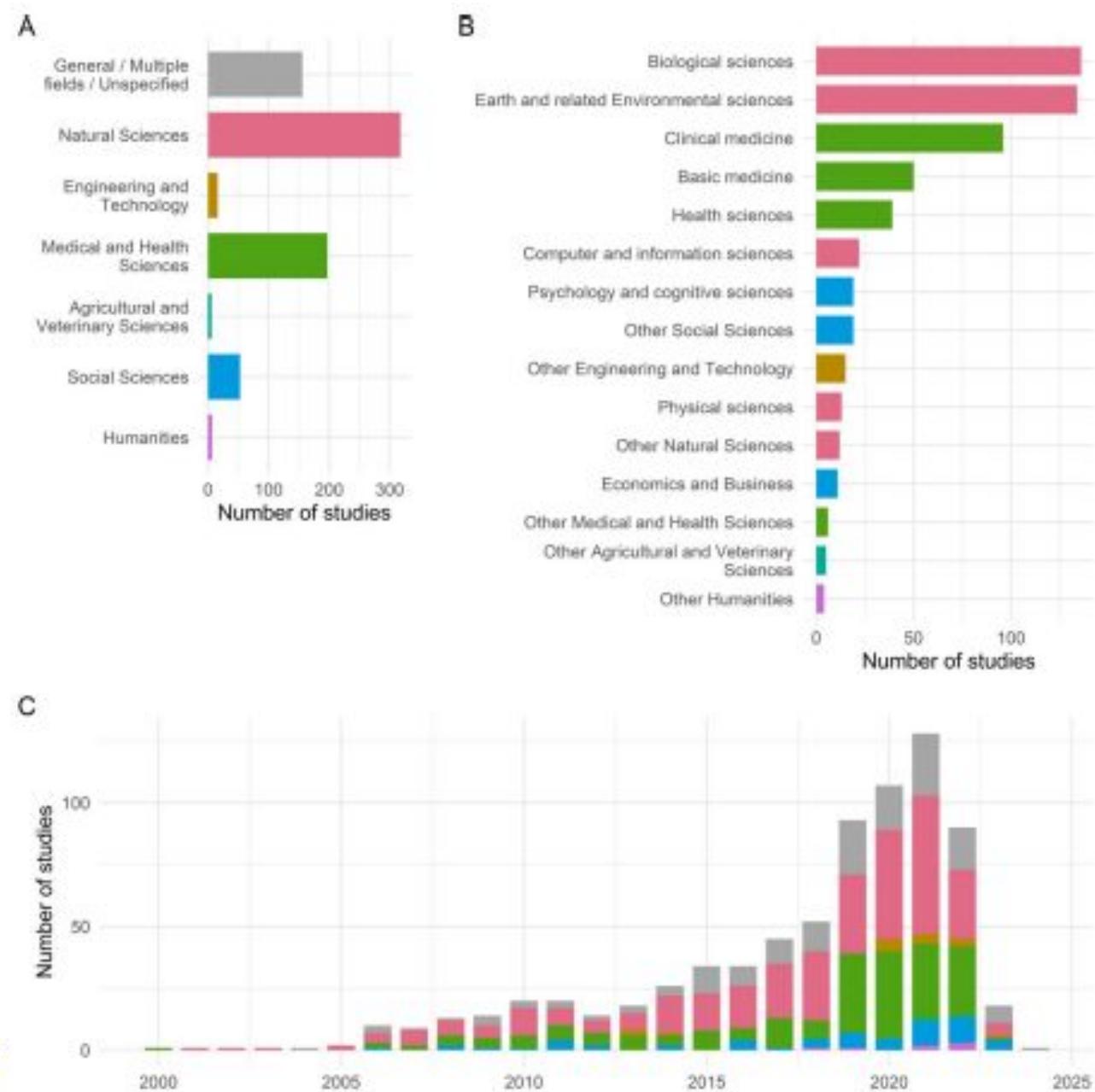


Funded by
the European Union

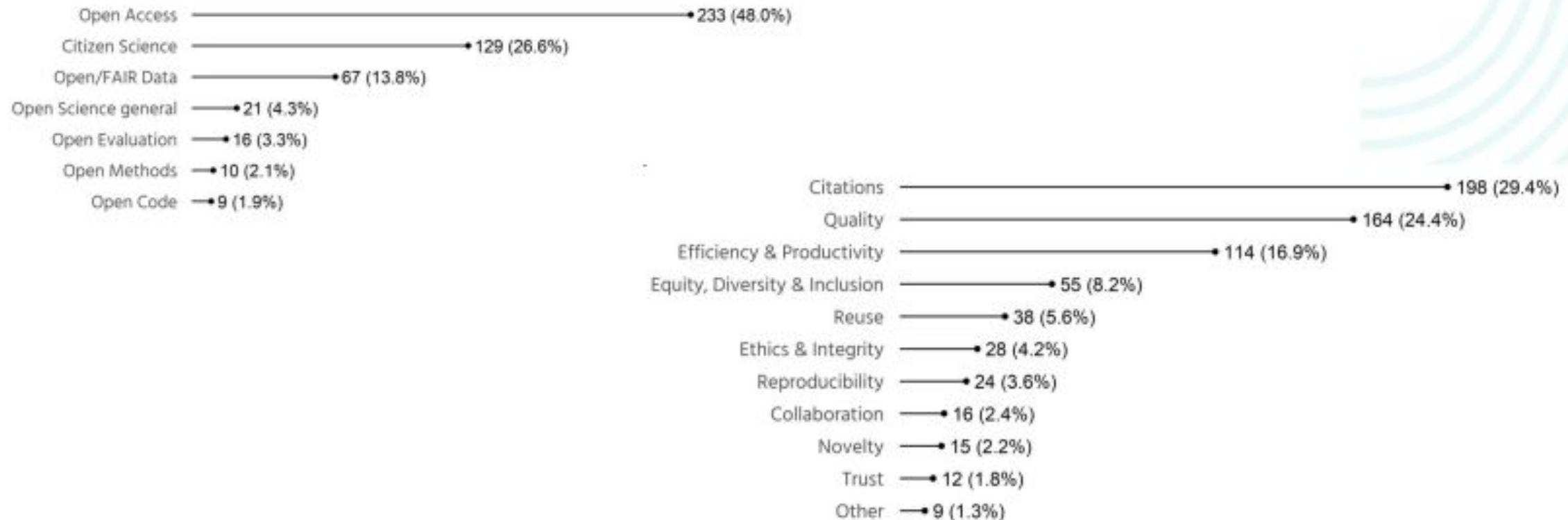
PathOS Training Session: Exploring the Academic Impact of Open Science – 17th March 2025

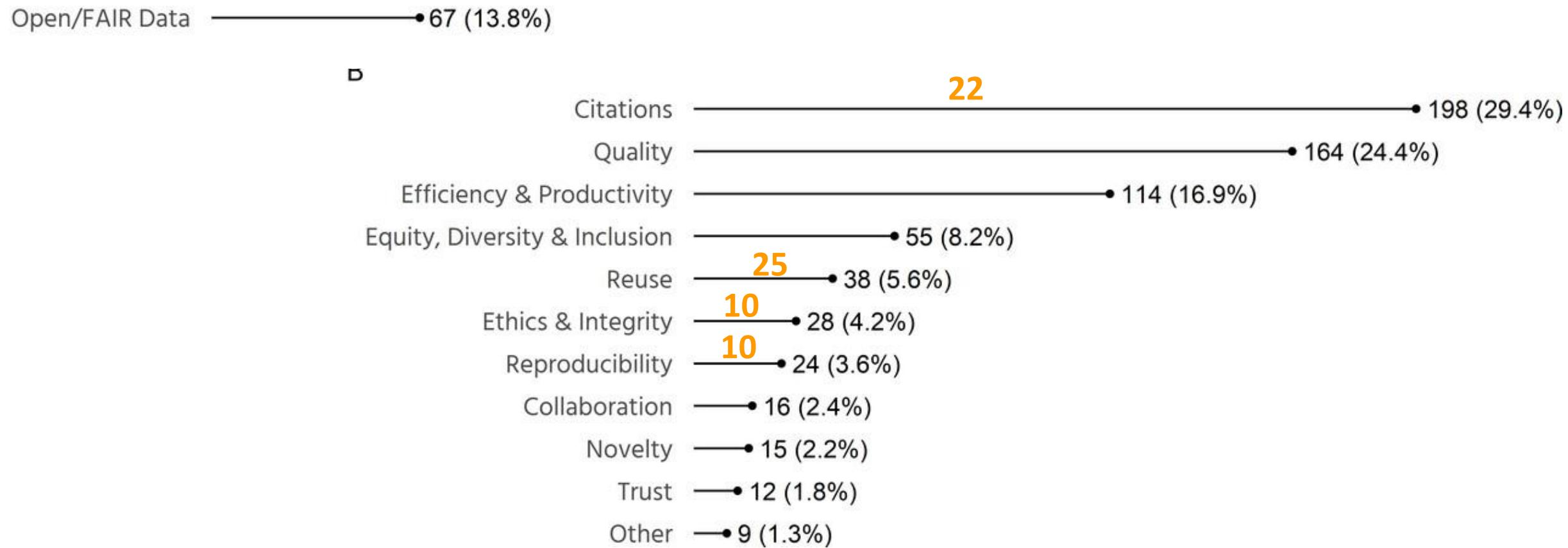


Disciplinary coverage



Types of Open Science & of impact





“Overall, few publications that are included in the review provide concrete evidence of impact of Open Data. Various publications study effects of data sharing policies, instead of effects of data sharing itself.”

Academic impact: Main findings

- **Citations**
 - Small to moderate increase from OA, Open/FAIR Data, and Open Evaluation
 - Unclear effects from Open Code, no effect from OS badges
- **Quality**
 - Neutral to moderate positive effects from Open Peer Review
 - Conflicting evidence from OA, Citizen Science neutral effect on quality given sufficient training
- **Efficiency & productivity**
 - Positive effects from Citizen Science, OA, and Open Science in general
 - Unclear effect of Open Evaluation
 - Wasted time from predatory publisher emails (OA)
- **Equity, diversity and inclusion**
 - OA leads to more diverse citations and international collaboration
 - Marginalization of those with fewer resources (OA-APC, Open/FAIR Data) or lower status (Open Evaluation)
 - Citizen Science activities focused in the Global North
- **Reuse**
 - Positive effect of Open/FAIR data
- **Reproducibility**
 - Positive effects of preregistrations and registered reports
 - No effects of Open/FAIR Data or Open Methods
- **Novelty**
 - Potentially positive effect of OS practices on rate of true discoveries
- **Ethics & Integrity**
 - Unclear impact of Open Evaluation on integrity of reviews
 - Open/FAIR data has risk of re-identifying participants
- **Trust**
 - Positive effect of OS badges on trust in results by scientists.

CBA in a nutshell



ADEQUATE TIME HORIZON

Time horizon for the analysis matching the characteristics of the evaluated intervention

INCREMENTAL PERSPECTIVE

Comparison of the intervention vs a counterfactual

SOCIO-ECONOMIC PERSPECTIVE

Use of shadow prices reflecting value to society

PERFORMANCE INDICATORS IN MONETARY TERMS

Conversion of all costs and benefits to monetary values

MICROECONOMICS-BASED

Approach grounded in microeconomic theory

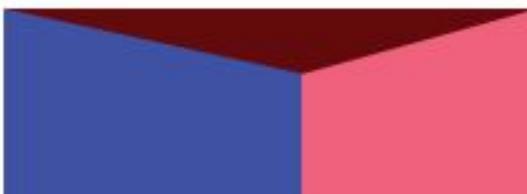
WHY A CBA FRAMEWORK FOR OS?

- OS is a **priority** in EU agenda, embedded in several programmes and initiatives (e.g. Horizon Europe, EOSC, etc) with increasing demand for **transparency and accountability**.
- Much of the existing literature has focused on the total impact rather than on the net impact, that is, the difference between costs and benefits when comparing scenarios with and without OS, potentially leading to an overestimation of benefits and underestimation of costs and implementation challenges.
- The focus has largely been on describing effects, drivers, and barriers, without providing a **reproducible framework** for estimating and consistently comparing socio-economic costs and benefits to assess the convenience of specific practice.

TYPICAL BENEFITS OF OPEN SCIENCE

TYPICAL COSTS OF OPEN SCIENCE

- **Set-up costs**
- **Operational costs**
- **Users' costs**
(e.g. time spent to become efficient users and/or for updates).



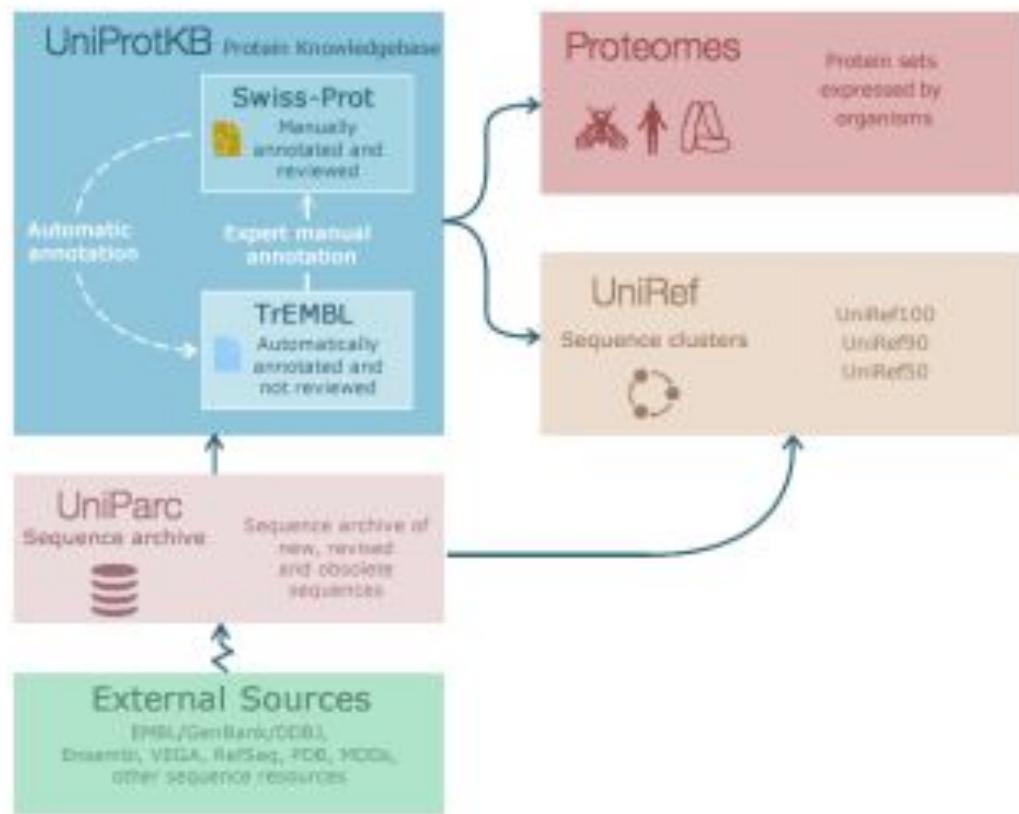
- **Saved time in data/information retrieval**
- **Time and effort saved in working activities**
(e.g. avoiding duplication)
- **Lower expenses related to data storage**
- **Reduced time and costs associated with negotiating access to alternative data sources**





- **OS initiatives are inherently collaborative**, so it's essential to clearly define the roles and contributions of each partner to properly scope the analysis.
- **Ask: What if the OS resource didn't exist?** This reveals its true value and what would be lost without it. Modular scenarios can help capture this in the OS context.
- **Measuring impact is challenging** because OS resources are freely accessible — but it's achievable, especially with AI-powered tools and methods. Tracking usage requires both granular and accurate metrics, as well as expertise in interpreting these data

UniProt – Universal Protein Resource



Key features of the database are the following:

- **Sequence Data:** UniProt contains a vast amount of protein sequence data, with annotations regarding the protein's function, domains, post-translational modifications, and more.
- **Cross-References:** The database includes extensive cross-references to other biological databases, allowing users to integrate protein data with other types of biological information.
- **Download, Search and Analysis Tools:** The data can be retrieved from the website through different download options and via programmatic access. UniProt provides various tools for searching, analysing, and visualising protein data, making it a powerful resource for researchers.
- **Completely free and accessible:** no registration is required to access the databases.

(1) Scope of the analysis

Which are **the boundaries of UniProt?**



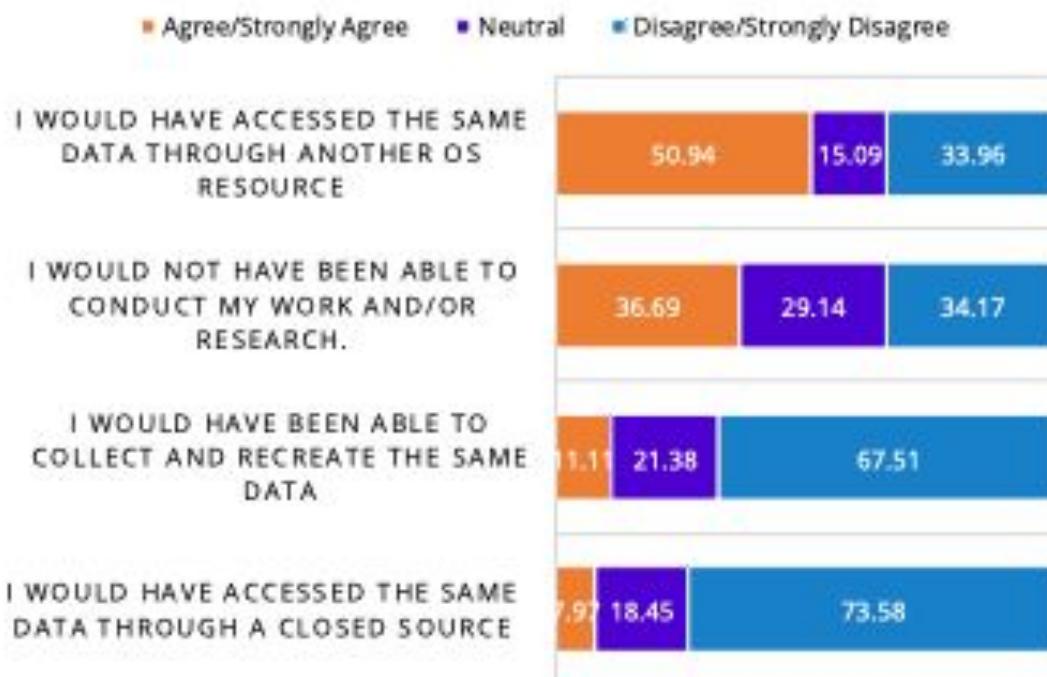
- UniProt resources provided by the UniProt Consortium
- Activities of the user community



Costs and benefits related to the development of the original resources cross-referenced in UniProt

(2) Identification of the counterfactual

CBA compares the costs and benefits of a project to what would have occurred without it



There is no single comparable alternative to UniProt.

- No mutual exclusion in alternatives → people provided inconsistent preferences
- Alternatives to UniProt depend on the field, the type of use, and resource accessed

Overview of results

Average value of the efficiency gains for all users per year:
from EUR 373 MLN to EUR 565 MLN

- Transaction cost savings
- Access cost savings
- Labour cost savings

 230-300 hours saved on average per year by each user

 Net benefit per user per year
EUR 3,500 – 5,000
the value of the time saved
per user per year

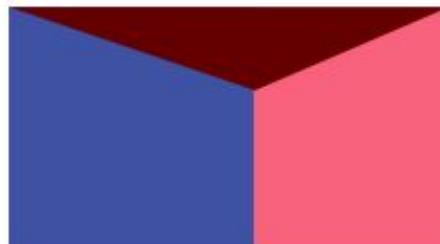
Average annual costs: EUR 14.6 MLN

- Maintenance annual costs
- Cost of sending contributions to improve protein data

TOTAL AVERAGE ANNUAL COST & BENEFITS

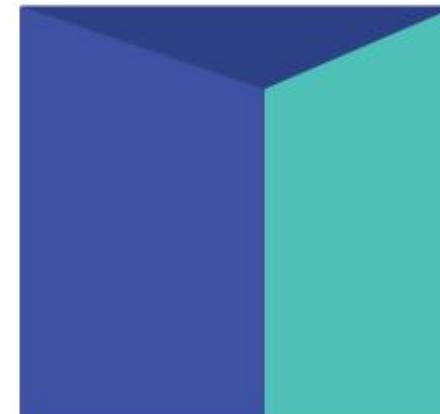
UniProt's operation depends heavily on human expertise (~70% of the expenses).

Total annual costs €14.6 million



Users benefit from time savings in accessing data and conducting research.

**Total annual benefits:
€373 million – €565 million**



POWERING DISCOVERY, INNOVATION AND ECONOMIC GROWTH



Creating new knowledge

15,200+ scientific publications mentioned UniProt and its resources.

Accelerating R&D and innovation

183,000+ patent publications mentioned UniProt and its resources.

Innovations driven mainly in **biotechnology and pharmaceuticals** fields.

UniProt fosters spin-offs and start-ups within the life sciences sector that provide expertise in data manipulation, interpretation, and analysis.

The resource accelerates the development of **new products and technologies**, for example, protein identification, drug target validation, machine learning models.

Pfizer, Genentech, Roche, Novartis, and other global leaders rely on UniProt for bioinformatics-driven discovery.

WHAT IF UNIPROT DID NOT EXIST?

- 74%** **Would not have had access to the same data from a closed source**
- 68%** **Could not have recreated the same data themselves**
- 65%** **Would have experienced slower publication timelines**
- 52%** **Said academic collaboration would have been limited**

No other resource offers UniProt's unique combination of protein data quality, coverage, and integration.

Open Data: Is It Worth It?