

# 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

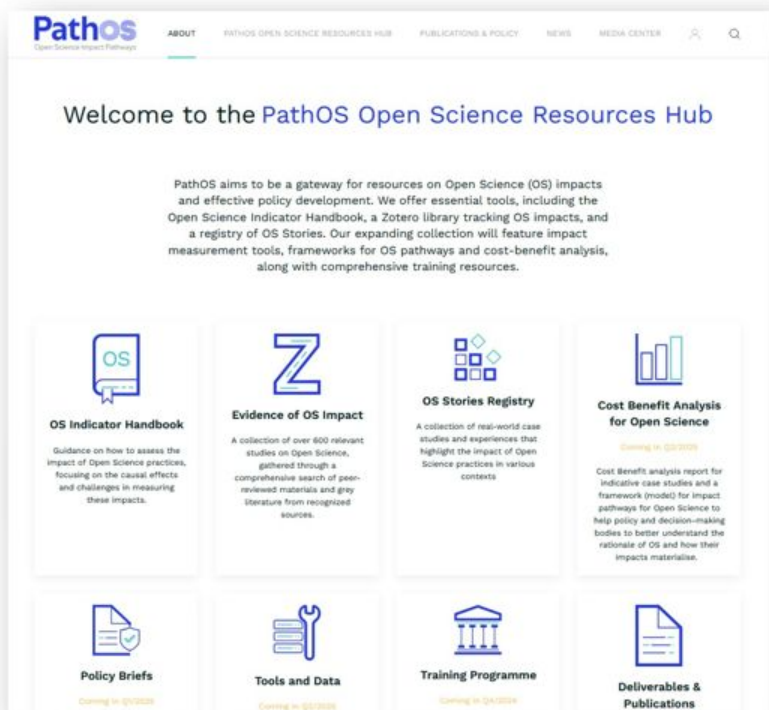
“PathOS was a Horizon Europe project that aimed to collect concrete evidence of Open Science effects and to study the pathways of Open Science practices, from inputs to outputs, outcomes, and impacts, including the consideration of enabling factors and key barriers.”

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/>



## Frameworks

01

### OS Impact Pathways

- Cost-Benefit Analysis for OS

## Handbook of OS Indicators

02

- Indicator "Recipes"
- Tools and Datasets

## Literature Insights & Registry

03

### Lit Review on OS impacts

- Online registry of OS stories

## Case Study Deep Dives

04

- OS impact assessments, Causality focus
- Cost-Benefit evaluations (*Elixir UniProt & RCAAP case studies*)

## Training & Engagement

05

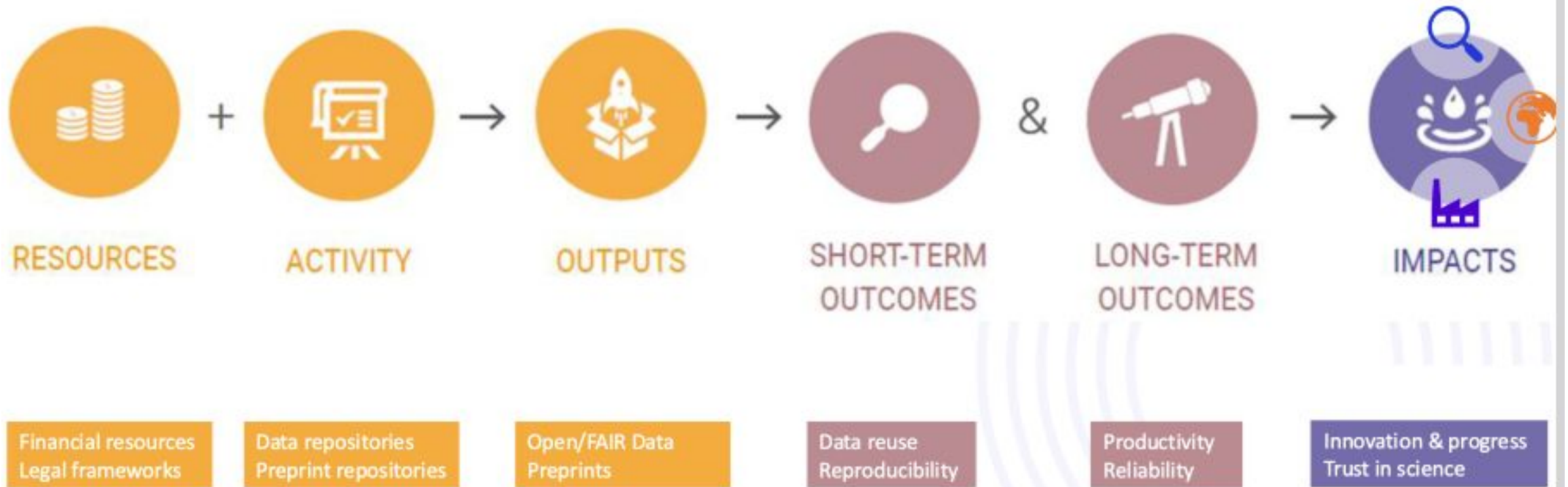
- Training for policy-makers & research administrators

## Recommendations

06

- Guidelines and best practices
- Project-derived insights

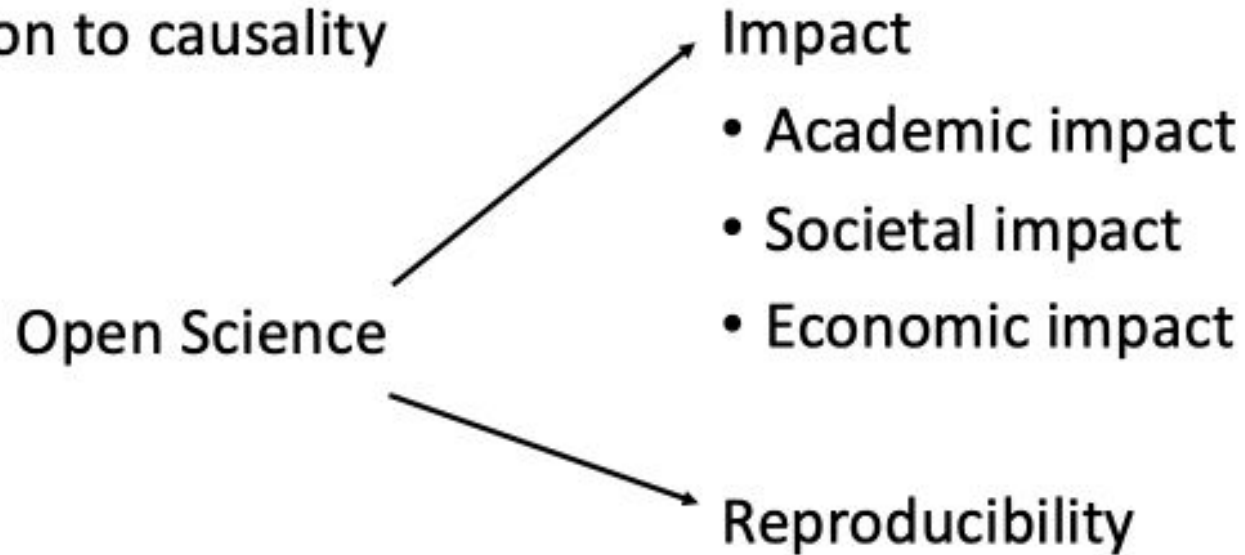
# Open Science Impact Pathways





# Overview sections

Introduction to causality





# 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 I, Ross-Hellauer T. 2025 The academic impact of Open Science: a scoping review. *R. Soc. Open Sci.* 12, 241248.

<https://doi.org/10.1098/rsos.241248>

Received: 18 August 2024

Accepted: 20 January 2025

**Subject Category:**  
Science, society and policy

**Subject Areas:**  
e-science

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

**Author for correspondence:**  
Thomas Klebel  
e-mail: [thobek@science-center.at](mailto:thobek@science-center.at)

## The academic impact of Open Science: a scoping review

Thomas Klebel<sup>1</sup>, Vincent Traag<sup>2</sup>, Ioanna Grypari<sup>1,4</sup>, Lennart Stoy<sup>5</sup> and Tony Ross-Hellauer<sup>1</sup>

<sup>1</sup>Open and Reproducible Research Group, Knowledge Research Center, Graz, Austria  
<sup>2</sup>Centre for Science and Technology Studies (CWTS), London University, London, The Netherlands  
<sup>3</sup>Adesso Research Center, Munich, Germany  
<sup>4</sup>OpenMIR, Athens, Greece  
<sup>5</sup>Technopolis Group, Brussels, Belgium

TK, 0000-0002-7311-4751; VT, 0000-0002-2570-3879;  
IG, 0000-0002-3435-1591; LS, 0000-0002-3817-9875;  
TRH, 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, scrutinized 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), OpenFAIR Data (OFAIR), 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 OFAIR. Key areas of impact studied are citations, quality, efficiency, equity, reuse, ethics and reproducibility, with most 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:

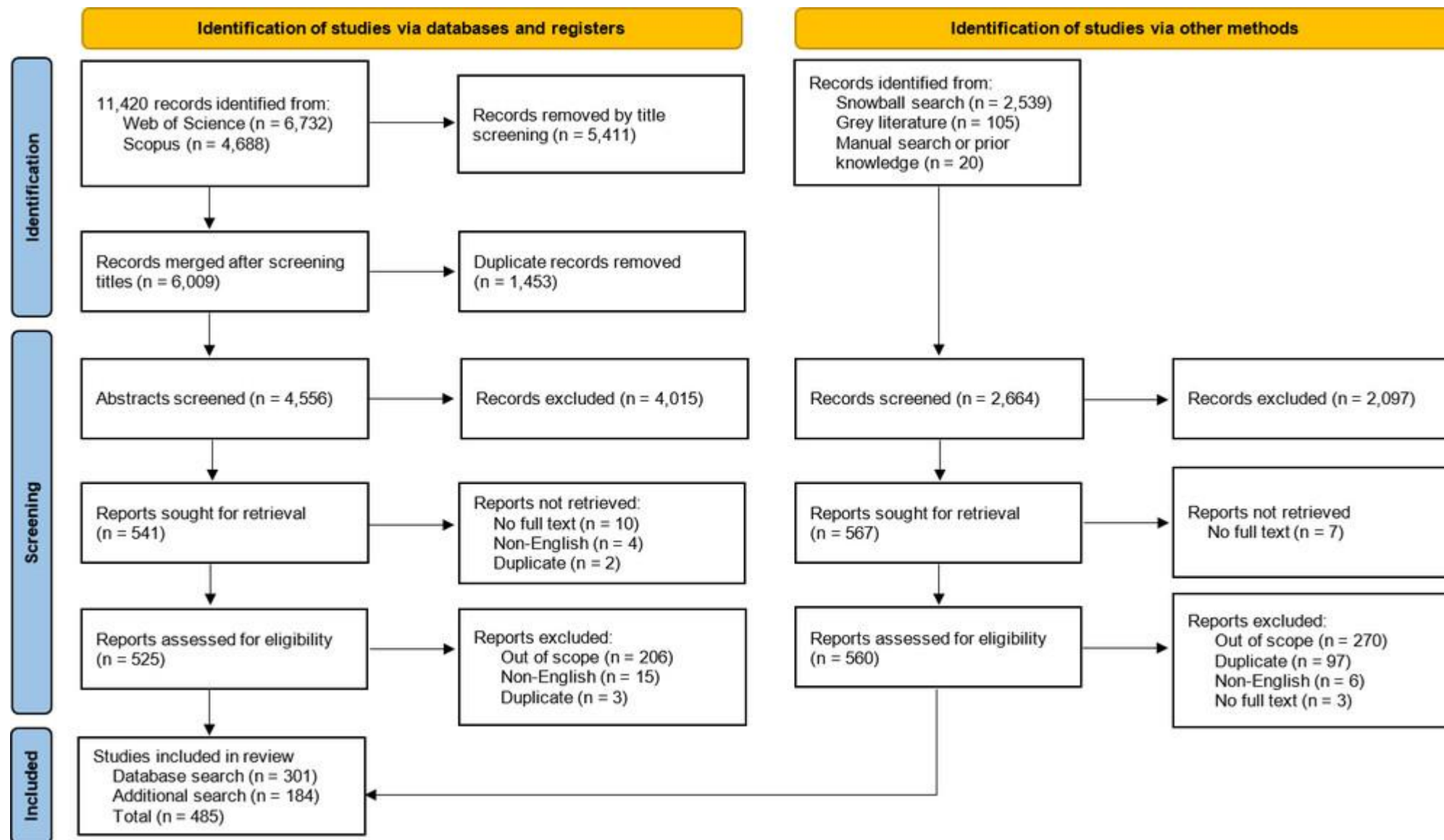
- What types of positive or negative, direct or indirect academic impact are observed?
- What kinds of mechanisms produce them?
- What specific enabling and/or inhibiting factors (drivers and barriers) are associated with these impacts?
- What knowledge gaps emerge from this analysis?

PathOS

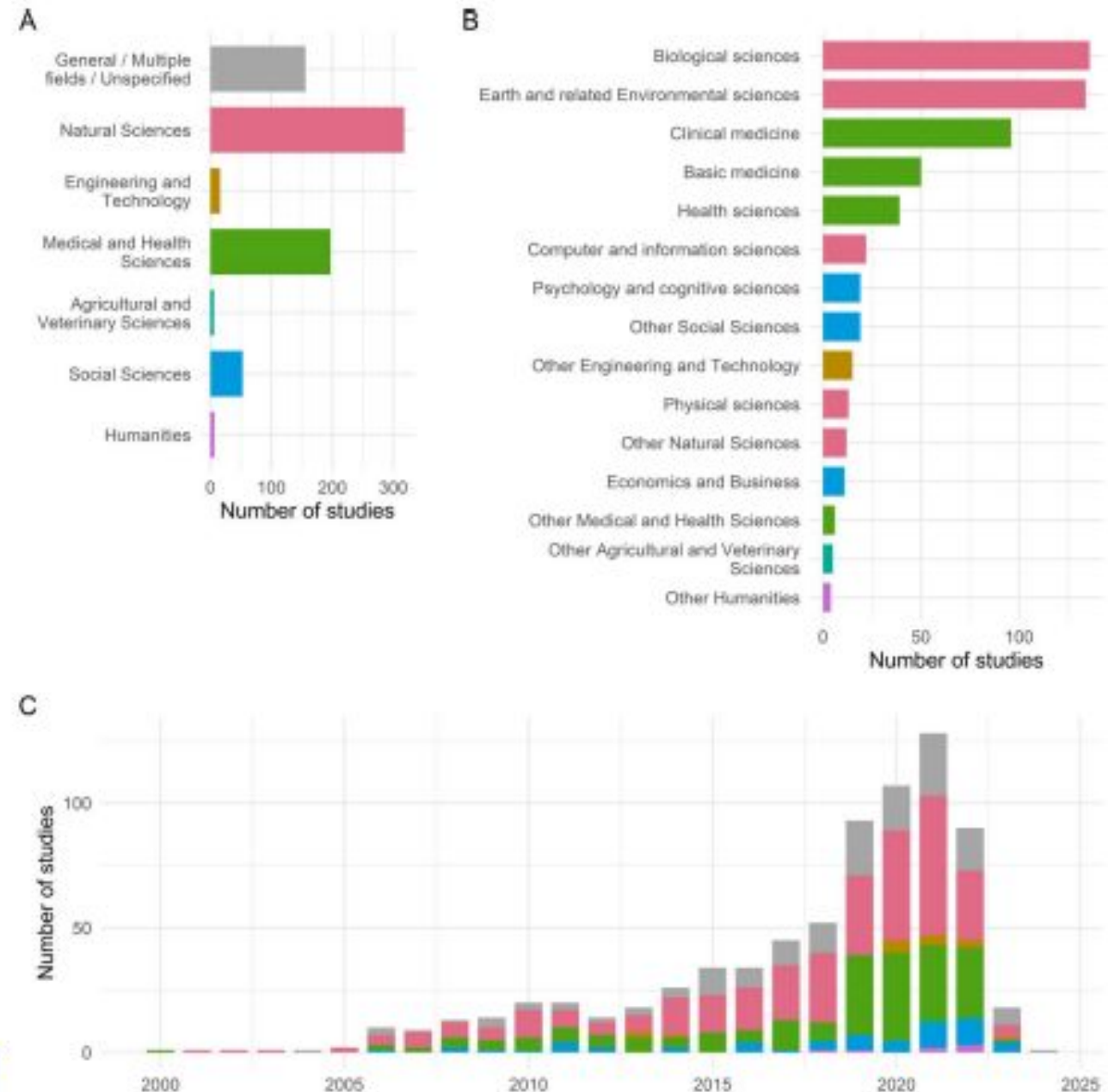


Funded by  
the European Union

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

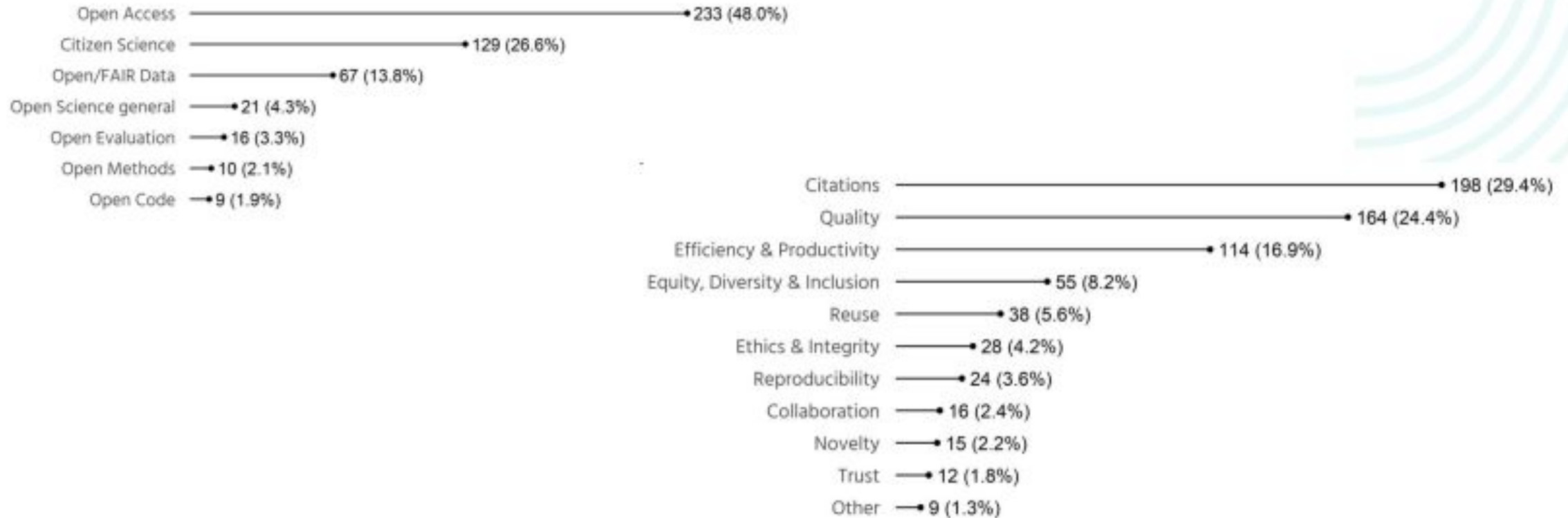


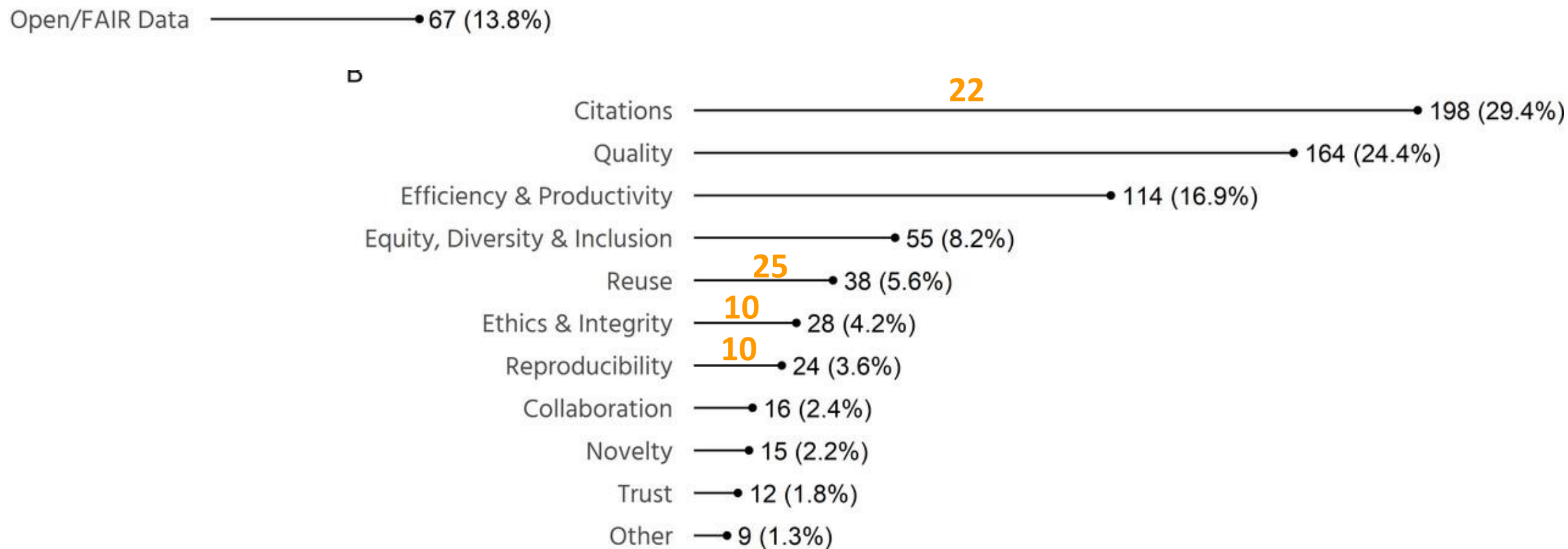
# Disciplinary coverage





# 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 COSTS OF OPEN SCIENCE

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

## TYPICAL BENEFITS OF OPEN SCIENCE

- **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**

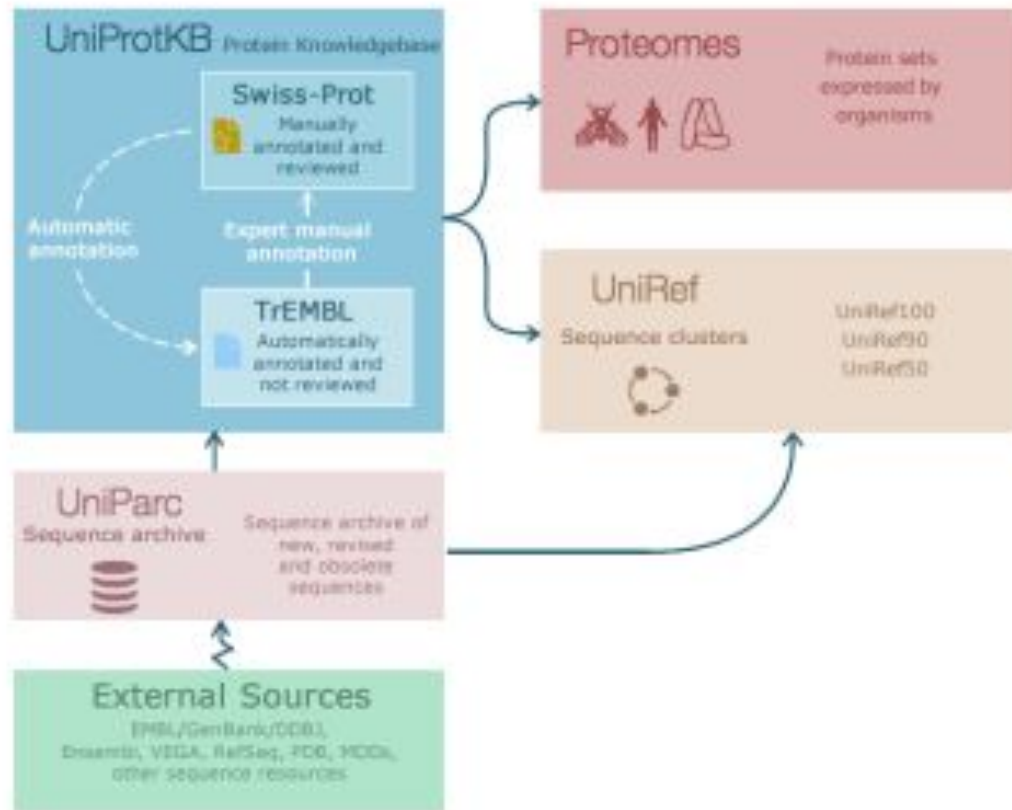






- **OS initiatives are inherently collaborative**, so it's essential to clearly define the roles and contributions of each partner to properly scope the analysis.
- **Ask: What if the OS resource didn't exist?** This reveals its true value and what would be lost without it. Modular scenarios can help 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

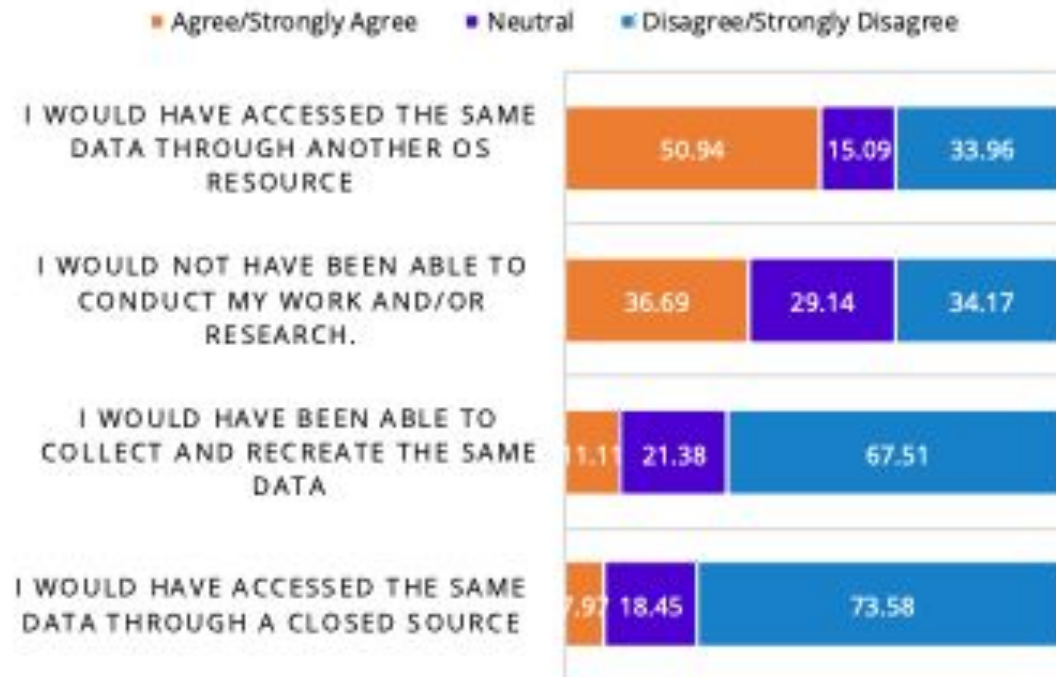


Costs and benefits related to the development of the original resources cross-referenced in UniProt



## (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?