Pathos
Open Science Impact Pathways

UNLOCKING THE INSIGHTS FROM THE PATHOS CASE STUDIES

PathOS is a Horizon Europe project aiming to collect concrete evidence of Open Science effects, study the pathways of Open Science practices, from input to output, outcome and impact, including the consideration of enabling factors and key barriers.

The PathOS project has completed 6 case studies that model Open Science pathways, measuring input indicators, costs, and outcomes. These case studies have tested and operationalised Open Science indicators, providing valuable data and expert insights from various local environments. They have also contributed to the Cost-Benefit Analysis and validated key results. Each case study tells a complete story, covering different objectives, implementation mechanisms, and actors across the research and innovation ecosystem.

Key Questions Explored in the PathOS Case Studies

Impact of Open Access Routes on Topic Persistence

Does Open Access increase the long-term influence, real-world uptake, and equity in climate–AI research?

Portuguese Repository Infrastructure RCAAP

How does Portugal's RCAAP infrastructure affect industry use, visibility, and reuse of research publications?

Effects of Data Repositories on Data Usage

How does repository type influence the reuse of research data, especially in the social sciences?

ELIXIR´s Bioinformatics Resources

How do ELIXIR's open resources fuel innovation and industry-wide benefits?

Impact of Artefact Reuse in COVID-19 Publications

Does the reusability of open COVID-19 artefacts influence downstream uptake in health and innovation?

French Open Access Infrastructure

How are French Open Science platforms used beyond academia, and by whom?

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101058728.

Pathos
pathos-project.eu

Impact of Open Access Routes on Topic Persistence

Examines how two distinct Open Access routes, self-archiving in repositories (Green Open Access) and journal-based publication with open licences (Published OA), differentially influence the thematic persistence and downstream uptake of AI and climate research in academic, industrial, and equity-related contexts.

DATA SOURCES USED

This case study uses publications and data from the OpenAIRE Knowledge Graph (for Open Access status, funding, and affiliations), Semantic Scholar (for citations), PATSTAT (for patent citations), and ROR.org (for organisation types). It focuses on AI and Climate Change publications across disciplines.

HIGH LEVEL METHODOLOGY

The methodology compares Open Access and closed-access publications using Propensity Score Matching (PSM) to isolate the causal effects of two specific Open Access routes: Green Open Access (repository-based self-archiving) and Published Open Access (journal-mediated with explicit open licences). Analyses draw on structured metadata from the OpenAIRE Graph API (for Open Access status and funding links), Semantic Scholar (for citation metrics), PATSTAT (for patent citations), and the SciNoBo platform (for field classification, citation impact, topic persistence, and Sustainable Development Goals relevance). Author gender is inferred from given names to examine gender equity. Economic and translational impact is assessed via patent links and science–industry collaboration patterns. A central focus is placed on whether different Open Access routes enhance the long-term persistence of AI-for-Climate research topics, with supporting attention to collaboration structures and gender participation.

CREATED DATA COLLECTION

Collection of AI and Climate Change publications, enriched with metadata such as research fields (FOS), citation impact (FWCI), author demographics, Open Access status, funding information, Sustainable Development Goals targets, and patent citations. This data identifies emerging and persistent topics in AI and Climate Change.

CREATED / UPDATED TOOLS

The case study led to the development of a tool for detecting persistent topics based on publication trends, citation impact, and topic longevity. Additional tools include an affiliation analysis tool for identifying organisation types and a publication-to-patent citation analysis tool to explore innovation links. Author gender classification was also applied to assess demographic patterns in research.

Academic Impact

Especially through Green Open Access routes, it significantly boosts citation impact and helps sustain the visibility of emerging AI and Climate Change research topics. Green Open Access is consistently associated with higher academic reach, stronger field-weighted impact, and markedly greater topic persistence over time. Published Open Access contributes to visibility but shows a weaker or even negative effect on long-term topic vitality.

- Citation impact
- Thematic persistence
- Diversity

Economic Impact

Green Open Access publications show stronger connections between research and industry, including higher science–industry collaboration rates and modest increases in patent citations. These findings suggest that repository-based access facilitates knowledge transfer and enhances the practical uptake of AI-for-Climate research.

- Science-industry collaboration (affiliation analysis, patent citations)

Societal Impact

Open Access articles, particularly those in the Published Open Access category, correlate with greater alignment to the Sustainable Development Goals and modest gains in gender equity, especially in increasing the share of women in senior (last-author) positions. However, Green Open Access is linked to a small decline in all-women author teams, highlighting the complex dynamics between openness and inclusion.

- Uptake in and impact on societal issues (SDG Targets)

Link to indicators from the Open Science Impact Indicator Handbook

Portuguese Repository Infrastructure RCAAP

The presented case study investigates whether the availability of Open Access publications in the Portuguese repository infrastructure, RCAAP, increases usage and collaborations between academia and industry.

DATA SOURCES USED

Data from OpenAIRE Knowledge Graph, ORBIS, and ROR.org is used to explore publications in the RCAAP (Open Access Repository of Portuguese Academic Publications). The focus is on metadata related to citation patterns, Open Access status, funding information, and the involvement of Portuguese companies in research.

HIGH LEVEL METHODOLOGY

The methodology for the RCAAP case study involves analysing citation impact and collaborations through A/B group comparisons and OA status. The study uses tools to assess citation impact and the level of cooperation in scientific research, with a focus on science-industry collaborations. The approach incorporates AI-driven, big data techniques to evaluate academic and economic impacts.

CREATED DATA COLLECTION

The collection of RCAAP and Portuguese publications is enriched with metadata on research fields (FOS), citation impact (FWCI), Open Access status, funding information, affiliation types, and Portuguese company affiliations.

CREATED / UPDATED TOOLS

The tools generate Affiliation Analysis (focusing on the type of organisation and its connection with Portuguese companies) and create Publication-Patent Citation Analysis, which together assess these publications' academic and economic impact.

Academic Impact

Identified that RCAAP Open Access publications have a strong citation impact and foster collaboration intensity, leading to improved visibility in academia.

- Citation impact
- Collaboration intensity

Economic Impact

Highlighted how RCAAP infrastructure contributes to cost savings and efficiency gains through mutualisation.

- Science-industry collaboration

Societal Impact

Not identified in this case study.

- Not applicable

Link to indicators from the Open Science Impact Indicator Handbook

Effects of Data Repositories on Data Usage

The presented case study evaluates whether data hosted in a national platform leads to higher reuse or citation frequency than data shared through alternative repositories, illuminating how localised governance can shape OS adoption.

DATA SOURCES USED

Data from sources like OpenAlex, Scopus, and Data Citation Corpus are used to investigate data mentions in publications. The focus is on metadata related to datasets, their reuse, and citation patterns, particularly how research artefacts are created or reused, with a specific focus on the social sciences, and what role data repositories play.

HIGH LEVEL METHODOLOGY

The repository case study uses citation impact, dataset reuse, and field classifications to assess the academic impact of open data. The methodology involves citation impact analysis, field of science analysis, and artefact analysis to explore the reuse of research artefacts and their impact on academic communities.

CREATED DATA COLLECTION

Collection of SSH publications, enriched with metadata on research artefacts that were reused to study data mentions in SSH generally and the accuracy of algorithmic tools for extracting data mentions.

CREATED / UPDATED TOOLS

The tool adapts SciNoBo Research Artefact Analysis to create outputs that assess the academic impact of research artefacts and their reuse across various scientific disciplines.

Academic Impact

Demonstrated that the repository where data is shared is associated with citation and dataset reuse.

- Dataset reuse

Economic Impact

Not identified in this case study.

- Not applicable

Societal Impact

Not identified in this case study.

- Not applicable

Link to indicators from the Open Science Impact Indicator Handbook

ELIXIR´s Bioinformatics Resources

The presented case study looks at how industry integrates Open Data and tools into their workflows, and whether usage analytics or patent references capture genuine economic and societal gains.

DATA SOURCES USED

Data for this case study is drawn from Semantic Scholar, EuropePMC, and PATSTAT, focusing on bioinformatics research. It includes metadata on publications, citation impact (FWCI), and reference counts. The study also looks at patents related to ELIXIR resources and digital skills in bioinformatics job vacancies.

HIGH LEVEL METHODOLOGY

This case study uses text mining techniques to extract publications and citation impact data, focusing on bioinformatics resources provided by ELIXIR. The methodology involves extracting patents and analysing economic growth, labour market impacts, and digital skills in the industry. Text mining for patent data, CBA (Cost-Benefit Analysis), including surveys and interviews to assess economic growth and cost savings for companies. The analysis of science-industry collaborations and patent citations is central to this study. AI-driven and big data techniques are employed to generate evidence of economic impact.

CREATED DATA COLLECTION

The collection of ELIXIR-supported publications is enriched with metadata such as research fields (FOS), citation impact (FWCI), and reference counts. Patents mentioning ELIXIR resources, and digital skills data from ELIXIR job vacancies.

CREATED / UPDATED TOOLS

The tools generate text mining techniques for extracting data from patents and create analysis on digital skills in the bioinformatics industry, enabling insights into the impact of ELIXIR resources on academic research and industry needs.

Academic Impact

Identified how using ELIXIR's public resources enhances citation impact and fosters interdisciplinary research.

- Citation impact
- Interdisciplinarity

Economic Impact

Highlighted the economic growth of companies through science-industry collaborations and how digital skills in bioinformatics lead to labour market benefits and cost savings

- Innovation output
- Economic growth of companies
- Labour market impacts
- Cost savings
- Science-industry collaborations

Societal Impact

Not identified in this case study.

- Not applicable

Link to indicators from the Open Science Impact Indicator Handbook

Impact of Artefact Reuse in COVID-19 Publications

The presented case study investigates whether the referencing and reusing of datasets and software in COVID-19 research publications contributed to downstream impact, including mentions in patents and clinical trials or guidelines, highlighting how practical openness may support research translation during and beyond emergency-response contexts.

DATA SOURCES USED

This case study uses enriched data from OpenAIRE Knowledge Graph, Semantic Scholar, CORD-19 (full texts), PATSTAT, and ROR.org. It focuses on COVID-19-related publications and analyses citation impact, Open Access status, research affiliations, reuse of open artefacts such as datasets and software, and patent citations. The study also integrates data from PUBMED on citations in clinical trials and guidelines to assess uptake in medical practice.

HIGH LEVEL METHODOLOGY

This case study investigates the downstream effects of Open Science practices by analysing COVID-19 research publications that created datasets or software. Rather than comparing Open Access versus non-Open Access status, it focuses on two observable behaviours: referencing external research artefacts (traceability) and the documented reuse of shared artefacts by others (reusability). A regression-based analytical design with interaction terms estimates impact while controlling for scientific quality, citation exposure, publication year, author count, and artefact volume. Academic impact is assessed using tools that analyse citation impact, field of science classification, research community engagement, and the reuse of datasets and software. Economic impact is evaluated through analysis of affiliations by organisation type (academic or industry) and tracking of patent citations. Societal impact is measured by analysing how often publications are cited in clinical trials and guidelines. The methodology combines large-scale data analysis, AI techniques, and causal inference approaches to understand the effects of Open Science.

CREATED DATA COLLECTION

Collection of COVID-19 publications enriched with metadata on research fields (FOS), citation impact (FWCI), Open Access status, affiliation types, patent citations, created/reused research artefacts, and clinical trial and guideline citations.

CREATED / UPDATED TOOLS

The case study resulted in refinements and adaptations of the SciNoBo Citance Analysis and Research Artefact Analysis tools. New tools were developed for analysing affiliation types (academic, industry), tracking citations from patents, and examining research uptake through citations in clinical trials and medical guidelines.

Academic Impact

Demonstrated that the COVID-19 research community adopted Open Science practices at scale, including structured sharing and referencing datasets and software. Papers whose artefacts were reused showed modest gains in cross-sector collaboration and interdisciplinary visibility, especially when paired with high quality and early dissemination.

- Interdisciplinary and novelty
- Reproducibility: polarity of publications (engagement and stance of the community)

Economic Impact

Revealed that artefact reuse is positively associated with industry engagement and technological uptake. Publications with documented artefact reuse received more patent citations and showed stronger links to science–industry collaboration, highlighting their role in driving innovation.

- Science-industry collaboration (affiliation analysis, patent citations)

Societal Impact

Found that artefact reuse alone did not lead to increased clinical uptake. Once factors like study quality, visibility, and timing were accounted for, reused papers showed slightly fewer citations in clinical trials and guidelines, suggesting that clinical influence depends more on research credibility and timeliness than on reuse alone.

- Uptake in medical practice (citations in clinical trials and clinical guidelines)

Link to indicators from the Open Science Impact Indicator Handbook

French Open Access Infrastructure

The case study analyses how users within academia and other social and economic sectors access the publications in the two leading French Open Science portals (OpenEdition and HAL) and assesses whether Open Access publications are accessed more frequently than closed ones across different sectors, disciplines, and countries.

DATA SOURCES USED

This case study analyses connection logs from two major French Open Science platforms, OpenEdition and HAL, covering January 2023 to September 2024. We enrich the raw logs with OpenAlex metadata on publication topics and openness status. Using the IP addresses recorded in the logs and data from Ipinfo.io, we classify access to publications by national geolocation and according to Eurostat's Statistical Classification of Economic Activities.

HIGH LEVEL METHODOLOGY

The key question this case study addresses is whether making a resource openly available increases its likelihood of being accessed. Connection logs are collected, securely stored, indexed, and enriched with multiple classifications to enable aggregated analyses. The final dataset can be explored via an online web application that computes the open access advantage for any topic, sector, platform and geolocation combination.

CREATED DATA COLLECTION

Collection of non-academic connection logs from French Open Science resources, enriched with topic metadata, openness status, and sector engagement. This data helps understand how French Open Science resources are accessed and used outside academic circles.

CREATED / UPDATED TOOLS

The tools generate IP classification (based on IPinfo, ElasticSearch and a custom Node.js script using llama3.3:latest) and adapt a middleware tool to enrich logs with openness status and scientific field of resources, providing deeper insights into the interactions between Open Science resources and non-academic users.

Academic Impact

Considering all disciplines and countries, Open Science publications have a distinctive 2.1% access advantage over closed publications.

- N/A

Economic Impact

For economic sectors other than education, the advantage of open science is more significant and amounts to 3.5%. Some sectors, like the Public administration and Professional activities, have even larger advantages, respectively 8% and 9%.

- Uptake by the economic sector

Societal Impact

Comparison between different countries reveals that an advantage for OS in many Global South and Global North countries: US (12%), India (9%), France (8%), China (8%), Canada (7%), Italy (7%), Portugal (7%), Singapore (6%), Bolivia (6%), Mozambique (6%).

- Open access advantage
- Uptake by the societal sector (topic, openness status, sector distribution)

Link to indicators from the Open Science Impact Indicator Handbook