Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Deep Dive into bibliometric analysis and impact assessment for scientific contributions in life science

Giuseppe Defazio
University of Bari Aldo Moro
Vincenzo Angiulli
University of Bari Aldo Moro
Chiara Bruno
Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies
Alessandro Cestaro
Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies
Marinela D'Ambrosio
Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies
Ernesto Picardi
University of Bari Aldo Moro
Graziano Pesole
University of Bari Aldo Moro
Bruno Fosso
bruno.fosso@uniba.it
University of Bari Aldo Moro
Francesca De Leo
Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies

software

Keywords: Bibliometrics, Impact, Research Infrastructure

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11725654/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Deep Dive into bibliometric analysis and impact assessment for scientific contributions in life science

Authors

Giuseppe Defazio[1†], Vincenzo Angiulli[1†], Chiara Bruno[2], Alessandro Cestaro[2,#], Marilena D'Ambrosio[2,##], Ernesto Picardi [1,2], Graziano Pesole [1,2,*], Bruno Fosso[1,*], Francesca De Leo[2,*]

1 Department of Biosciences, Biotechnology and Environment – University of Bari Aldo Moro

2 Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies – National Research Council

Current Address: Prebiomics Srl, Via della Cooperazione 53, 38123, Trento (TN), Italy.

Current Address: Laboratory of Industrial Hygiene and Toxicology, Interdisciplinary Department of Medicine-University of Bari Aldo Moro

† These authors contributed equally to this work.

* Correspondence: graziano.pesole@uniba.it (GP); bruno.fosso@uniba.it (BF); francesca.deleo@cnr.it (FdL)

Keywords

Bibliometrics, Impact, Research Infrastructure,

Abstract

Bibliometric analysis assesses scientific productivity through published literature, helping to evaluate research contributions. Well-curated databases such as OpenAlex, Scopus and Google Scholar support citation tracking, but tools to aggregate, represent and assess data at the level of large organisations, such as Academia, Research Infrastructure, or domain-specific cumulative scientific impact remain limited.

To address this gap, we present REIMPACT, a Python-based tool that performs a systematic assessment of primary and secondary bibliometric impact for RI services, Academia, and authors. In this work, we present two use cases to demonstrate the feasibility of REIMPACT in different application contexts.

Introduction

The European Commission (EC) has designated Research Infrastructures (RIs) as the backbone of the European Research Area (ERA), recognizing their strategic role in advancing scientific excellence, innovation, and international competitiveness (*The European Strategy on Research and Technology Infrastructures - Research and innovation*, 2025). Designed to provide access to cutting-edge technologies, high-quality data, and specialized expertise (Directorate-General for Research and Innovation, European Commission, 2019), RIs foster multidisciplinary collaboration and accelerate scientific discovery. Within this framework, the European Strategy For Research Infrastructure (ESFRI) Roadmap 2021 established a strategic landscape for a coherent, sustainable, and competitive European RI ecosystem—encompassing 63 RIs, comprising 41 Landmarks and 22 Projects (*ESFRI Roadmap*, 2021). This framework is currently evolving through the ESFRI Roadmap 2026 process, which places increased emphasis on long-term sustainability, digital integration, and European technological sovereignty (ESFRI, 2026). While strengthening the integration, accessibility, and resilience of RIs has been a long-standing priority, a primary policy challenge lies in systematically evaluating their multidimensional impact. The OECD Reference Framework (OECD, 2019) and European RI policy initiatives laid the groundwork for developing robust indicators, monitoring practices, and evidence-based evaluation frameworks (ESFRI, 2019; Kolar *et al.*, 2023; De Leo *et al.*, 2025; Garrard *et al.*, 2025). Today, these impact monitoring frameworks gain renewed strategic significance within the evolving ERA Policy Agenda and ongoing discussions surrounding the ERA Act, where demonstrating the scientific, technological, and socio-economic value of RIs is essential to justify long-term investments, support deep-tech innovation, and reinforce Europe's scientific sovereignty (Science Europe, 2025).

Within this landscape, ELIXIR represents a highly relevant case study for impact assessment in the life sciences. ELIXIR is an ESFRI Landmark Research Infrastructure dedicated to Life Sciences data and operates as a distributed infrastructure connecting national Nodes and European services. Its mission is to integrate and coordinate databases, software tools, standards, training activities, and computational resources to provide sustainable and interoperable services for the life science community.

ELIXIR holds a distinctive position because its contribution is not limited to the provision of physical facilities, but is strongly centred on digital resources, data stewardship, interoperability, and service provision. For this reason, ELIXIR is closely aligned with the implementation of the FAIR principles, (i.e. Findable, Accessible, Interoperable, and Reusable), which are central to enhancing the accessibility, reuse, and long-term value of life science data and services (Wilkinson *et al.*, 2016;

Directorate-General for Research and Innovation, European Commission, 2018). While *FAIRification* increases the potential reach and reuse of scientific outputs, it also requires impact assessment methods able to capture how FAIR resources, services, and data infrastructures contribute to research production, collaboration, and innovation (Mons *et al.*, 2017). This shift aligns with broader policy efforts within the publishing community to move "beyond traditional authorship" (Shillum, 2026), establishing frameworks that properly recognize non-traditional scientific contributions, such as data stewardship, software maintenance, and peer review.

To strengthen accountability and communicate the value of its distributed services, ELIXIR has developed an Impact Toolkit based on impact pathway logic (i.e. *activities* → *outputs* → *outcomes* → *impacts*) (Smith *et al.*, 2020; De Ruyck *et al.*, 2023). This approach links activities, outputs, outcomes, and impacts, helping RIs and scientific institutions to describe how their services and resources generate value over time. The Toolkit translates this logic into a standardised yet adaptable framework, including methodologies, indicators, and guidance designed to help ELIXIR Nodes and services monitor and communicate their contribution to scientific advancement, innovation, and societal benefit. Among the service-level indicators included in the Toolkit, the “Uptake/Visibility” category comprises metrics such as citations of a service or its associated DOI, dataset downloads, and references in scientific publications. These indicators provide measurable evidence of how scientific products are accessed, reused, and acknowledged, thereby connecting operational performance with broader scientific and societal impact.

Building on this framework, Martin *et al.* (2021) described how ELIXIR fosters collaborative innovation through its community-driven model, aligning scientific outputs and services with long-term sustainability and measurable impact indicators. Complementarily, De Leo *et al.* (2025) introduced a text-mining-based methodology to quantify the ELIXIR scientific legacy by tracing publication networks and evaluating the visibility of infrastructure-related outputs. Together, these contributions provide a conceptual and methodological foundation for evaluating impact across the broader European life science ecosystem.

Further, the analysis of peer-reviewed publications remains one of the most direct ways to evaluate bibliometric impact, as publications provide tangible evidence of the recognition and reuse of scientific outputs within the research community (Garfield, 2006; Bornmann and Daniel, 2008; Leydesdorff and Milojević, 2015). In this context, bibliometrics, which has been defined by Broadus (1987) as *“the quantitative study of physical published units, or of bibliographic units, or of the surrogates for either”*, represents a consistent way to measure RIs impact. Over the past three decades, bibliometric analysis has evolved from a set of techniques for counting and describing scholarly publications into a consolidated methodological framework for examining how knowledge

is produced, disseminated, reused, and acknowledged. Because it links research outputs to measurable traces of scholarly communication, bibliometrics is particularly relevant for assessing the medium- and long-term scientific and technological impact of complex organisations, such as Research Infrastructures, universities and public research bodies.

The transition to digital scholarly communication and the ubiquity of DOIs have greatly enhanced the traceability of research outputs. At the same time, the rapid expansion of digital publishing has introduced significant challenges related to research integrity, including the proliferation of paper mills, retracted literature, and artificial citation inflation (SMT, 2024). In response, institutional funders and publishers are increasingly adopting automated screening systems—such as author verification and integrity tools—to ensure transparency and compliance across scholarly outputs (Digital Science & Research Solutions Inc., 2026). These integrity concerns further emphasize the necessity for transparent, reproducible, and verifiable bibliometric assessment tools.

In this perspective, citation analysis provides quantifiable evidence of how frequently and in which contexts research outputs are cited, offering insights into their influence, visibility, and contribution to ongoing scholarly discourse (*Citation Analysis in Research Evaluation*, 2005). However, the increasing volume and diversity of life science outputs require reliable, scalable, and user-friendly tools able to support bibliometric impact assessment across different levels of analysis, from individual researchers and research groups to institutions and Research Infrastructures.

To address this need, we developed REIMPACT (Research Evaluation for IMPACT), a Python-based tool designed to automate the retrieval and analysis of citation data from large bibliometric databases, including OpenAlex and Scopus. OpenAlex is an open and comprehensive catalogue of the global research ecosystem, providing extensive bibliometric data (Priem *et al.*, 2022), while Scopus offers a broad database of peer-reviewed literature and citation metrics. Both databases have been selected among available services (e.g. Google Scholar and Web of Science) for this purpose by considering the availability of API services and analysis reproducibility (Culbert *et al.*, 2025). REIMPACT enables users to quantify the influence and visibility of publication sets through a versatile and scalable approach applicable to Research Infrastructures, individual researchers, departments, research groups, and institutions.

Finally, to demonstrate the applicability of REIMPACT, this paper presents two use cases: first, the bibliometric impact assessment of ELIXIR services provided by the Italian Node; and second, the citation-based assessment of 24 ELIXIR Head of Nodes (HoN). Together, these use cases show how REIMPACT can support RI accountability, service monitoring, and researcher-level assessment through reproducible bibliometric workflows.

Methods

Impact pathway logic for ELIXIR-IT service assessment

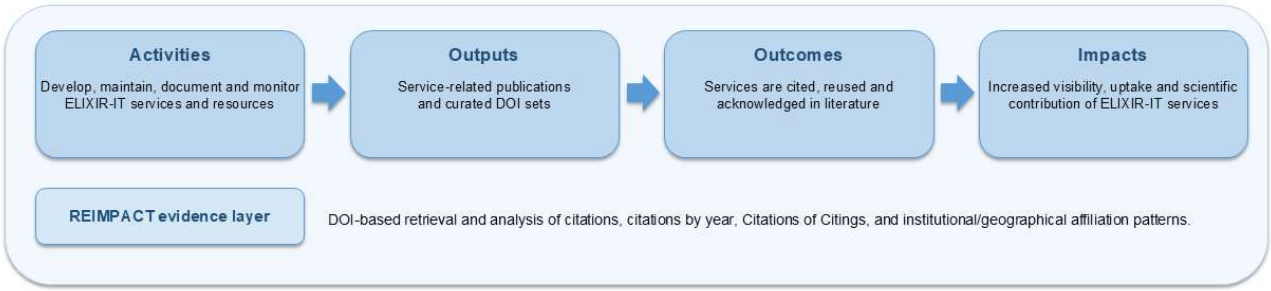
The methodological framework of this study is based on an impact pathway logic adapted from the ELIXIR Impact Toolkit. This approach was used to define the object of assessment, identify the expected chain of value generation, and select the bibliometric evidence needed to evaluate the scientific visibility, uptake, and reuse of ELIXIR-IT services.

In this framework, the object of assessment is not REIMPACT itself, but the set of ELIXIR-IT services provided by the Italian Node. Service-related activities include the development, maintenance, documentation, update, and monitoring of ELIXIR-IT services. These activities generate, or are associated with, identifiable scholarly outputs, mainly publications describing, updating, or documenting the services. When these publications are linked to persistent identifiers such as DOIs, they can be used as primary bibliometric objects for impact assessment.

The impact pathway links service-related activities to outputs, outcomes, and impacts. In the ELIXIR-IT service use case, outputs are represented by DOI-based publications associated with ELIXIR-IT services. Outcomes are captured through citations, reuse, acknowledgement, and visibility within the scientific literature. Broader impacts are interpreted as evidence of scientific legacy, bioinformatics resource uptake, research dissemination, and contribution to scientific excellence.

REIMPACT was used to operationalise this pathway by generating bibliometric evidence for selected Uptake/Visibility indicators. These include citations to service-related publications, citation trends, Citations of Citings, and the institutional and geographical distribution of authors contributing to citing publications. In this way, REIMPACT provides quantitative evidence that supports the assessment of how ELIXIR-IT services are accessed, reused, and acknowledged by the scientific community.

The impact pathway logic and the corresponding indicators used for the ELIXIR-IT service assessment are summarised in Figure 1.



List of indicators (2023) used for the ELIXIR-IT service impact case

ESFRI objective	Indicator title	Short description	ELIXIR impact area
1. Enabling scientific excellence Also: collaboration in Europe; optimising data use; international cooperation	• Number of mentions/name/accession number in literature of ELIXIR-IT services listed on bio.tools • Cumulative number of ELIXIR services/resources categorised • Authors contributing to cited publications	• Resource usefulness and contribution to enabling science • ELIXIR scientific legacy as shown by citations • Extent and utilisation of ELIXIR bioinformatics service portfolio	• Bioinformatics resource uptake • Scientific legacy • Research efficiency • Benefits derived from working together

Figure 1: Impact pathway used to frame the ELIXIR-IT service impact assessment. Service development and monitoring activities generate outputs that can be linked to DOI-based evidence; REIMPACT operationalises this bibliometric evidence layer to support uptake/visibility indicators for ELIXIR-IT services.

REIMPACT implementation

Here we introduce REIMPACT, a Python-based tool wrapped around the *PyAlex* and *PyBliometrics* libraries to efficiently query the OpenAlex and Scopus databases via their respective API (Application Programming Interface).

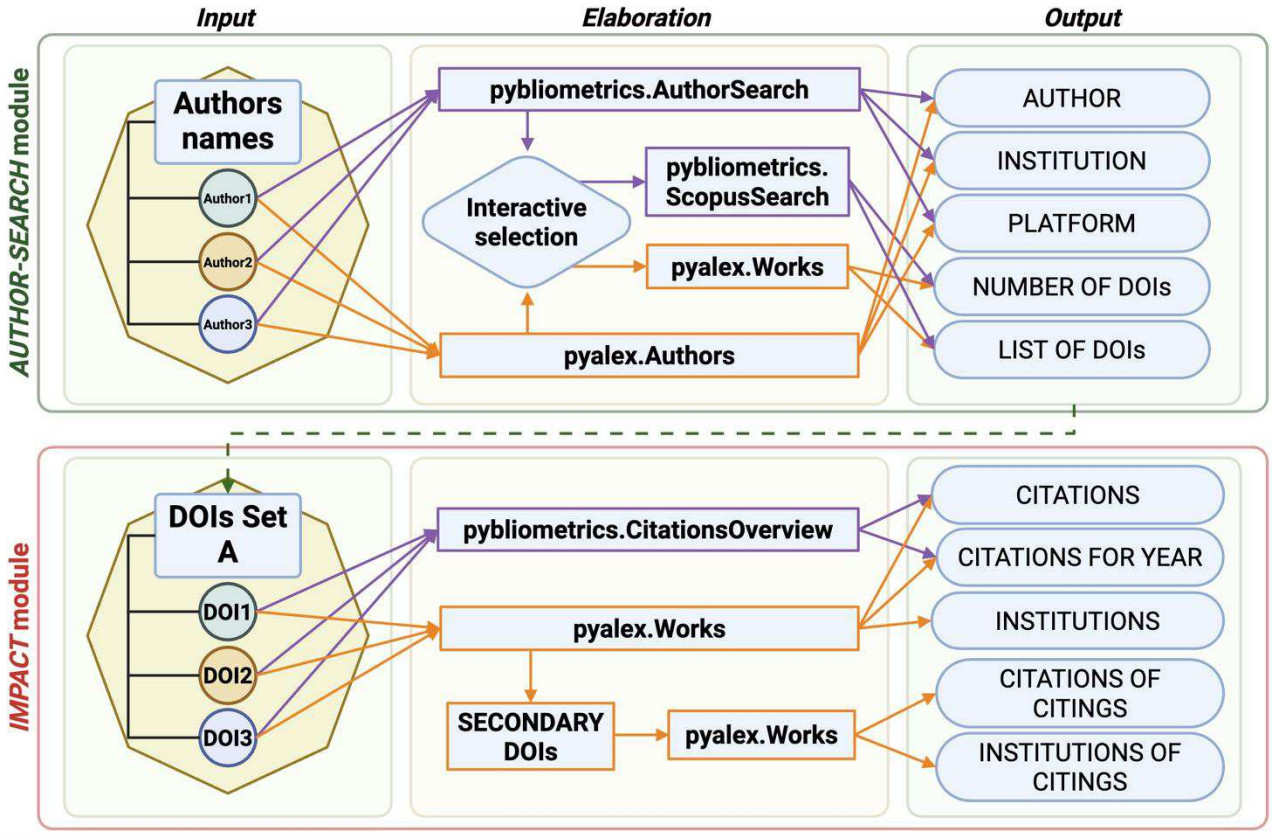


Figure 2: REIMPACT main framework.

The author-search module (green outline) retrieves DOI sets for one or more authors from OpenAlex and Scopus through interactive disambiguation, whereas the impact module (red outline) processes DOI sets to evaluate their impact. The elaboration step is carried out using PyAlex and PyBliometrics Python packages and is summarized in the central block of each module. The corresponding outputs are reported in the right-sided blocks. The output of the author-search module can also be used as input to the impact module.

Figure 2 presents the overall architecture of REIMPACT, which embeds two main modules: *author-search* and *impact.* The first one systematically retrieves all DOIs associated with one or more authors from both OpenAlex and Scopus. The *impact* core module is designed to assess the bibliometric impact of RI services, researchers and, groups from large institutions (e.g., RI, universities). All these entities may be associated to DOIs sets collected according to investigation or user criteria (e.g. DOIs set for all publications related to a specific database implementation)

author-search module

The *author-search* module accepts either a single author name or a list of authors from an Excel file, and retrieves the corresponding DOI sets from OpenAlex and/or Scopus after interactive authors disambiguation. In the implementation, author identification is performed through the PyAlex *Authors* and the PyBliometrics *AuthorSearch* objects, while DOI retrieval is carried out through PyAlex *Works* and, on the Scopus side, through PyBliometrics *ScopusSearch* object. The module supports both aggregated and separate reporting when multiple candidate profiles or affiliations are selected, allowing robust handling of career transitions, joint appointments, and metadata ambiguity. The module generates an Excel output file for each source database collecting author/s name/s, affiliation/s, source platform/s, number of retrieved DOIs, and the complete comma-separated list of DOIs for each processed author.

Impact module

The *impact* module takes as input a table that associates the names of RI services, researchers or groups from large institutions (e.g., RI, universities) with publications selected according to user defined criteria via DOIs sets, referred to as *primary DOIs*. For each primary DOI listed in the input table, REIMPACT creates a query instance using both PyAlex and PyBliometrics. In the source code, these instances are represented by the *Works* object for PyAlex and *CitationOverview* object for PyBliometrics.

The number of citations and the DOIs of citing papers (i.e., *secondary DOIs*) are used to further querying the databases, thereby obtaining the citation of related works. The sum of citations for the primary DOIs represents the primary impact of the DOIs set, while the sum of citations for the secondary DOIs represents the secondary or broader impact, defined as *Citations of Citings*.

In contrast, Scopus does not provide a scalable or straightforward way to retrieve the full list of citing works, and their citation counts for a given DOI. Therefore, OpenAlex is used exclusively for this aspect of the analysis, ensuring completeness and reproducibility.

As illustrated in Figure 3, *DOIs set A* is composed of DOI 1, DOI 2 and DOI 3 which have 3, 0 and 2 citations, respectively. DOI 1 is cited by 3 papers identified by secondary DOIs (DOI I, DOI II, and DOI III), whose citations are summed together with those of any secondary DOIs for DOI 2 (which, in this example, are zero since DOI 2 has no citations) and DOI 3, to calculate the secondary impact.

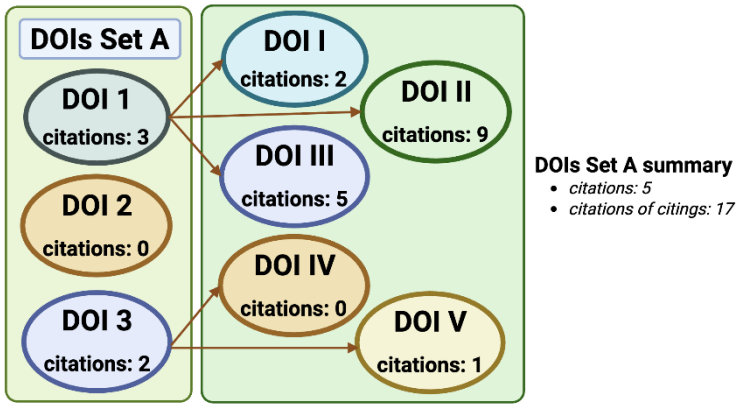


Figure 3: Representation of the rationale underlying the computation of "Citations" and "Citations of Citings".

Furthermore, for each assessed DOIs group, primary DOIs are used to query OpenAlex and Scopus to retrieve the distribution of citations per year over the last 20 years, regardless of publication date. Similarly, information regarding the institutions affiliated with the authors of both primary and secondary DOIs is collected and aggregated by service and country.

The output of *impact* module is a multi-sheet Excel file that includes:

I. The *summary* sheet which reports citation counts from both *Scopus* and *OpenAlex*, as well as *Citations of Citings* counts;
II. The *institutions* sheet, which indicates institutions and countries of the authors associated with the primary DOI;
III. The *institutionsOfCitings* sheet details the primary and secondary DOI linked to each DOIs set, along with the corresponding institutions and countries.

Additionally, a *DOIs sets* directory is generated, containing dedicated CSV files with bibliometrics statistics for each DOIs set under investigation over the last 20 years.

REIMPACT usage

REIMPACT is a user-friendly command-line tool designed to operate primarily on input tables in Excel (.xlsx) format. For the *author-search* module, when used in multi-authors mode, the table must contain only the column Author with an author per row.

For the *impact* module, the table must contain four columns in the following reported order:

I. *DOIs set name:* each row reports the name of the DOIs set to assess (e.g. authors, RI services);
II. *Applicant:* institution(s) e.g. responsible for maintaining or managing the RI service;
III. *Platform:* name of RI platform, University dept or another valuable label for this analysis;
IV. *DOIs set:* comma-separated list of DOIs corresponding to publications selected according to user/analysis criteria (e.g. papers where the RI service has been described or updated).

Alternatively, the input table can be the output of the *author-search* module, enabling a researcher-centric impact assessment workflow.

REIMPACT availability

REIMPACT is available at https://github.com/gdefazio/REIMPACT and implemented as a Conda package (https://anaconda.org/gdefazio/reimpact). We also provided the Conda package implementation for PyAlex and PyBliometrics to manage the installation of the REIMPACT Conda package.

Use cases data

To simulate a real case usage of the REIMPACT tool and highlight its usefulness in the field, we focused on the 2021 ELIXIR-IT Service Delivery Plan (SDP), a list of mature services that the Italian bioinformatics community shares within the European context. Information about services providing institutions and the corresponding ELIXIR-IT platforms (Data, Tools, Compute, Interoperability) was directly obtained from the SDP. The corresponding DOIs sets were identified through a manual query of Europe PMC (EPMC), using the service name as a keyword in titles and abstracts. Finally, service name, providing institution, platform, and DOIs were used to populate the *impact* module input table. Furthermore, an author-based use case is presented here using an Excel file containing the names of the 24 ELIXIR Head of Node (HoN) The *author-search* module was employed to retrieve all their associated publications, and these data were then used as input for *impact* module.

Results

Operationalising the impact pathway logic through REIMPACT

The results presented in this section show how REIMPACT operationalises the impact pathway logic described in the Methods for the ELIXIR-IT service use case. In this framework, ELIXIR-IT service-related activities are linked to scholarly outputs through DOI-based publications associated with each service. These publications represent the measurable outputs of the pathway. Their citations, citing publications, institutional affiliations, and geographical distribution provide evidence of outcomes related to visibility, reuse, and acknowledgement within the scientific community. Finally, broader bibliometric evidence, including Citations of Citings and the international distribution of citing authors, is used to explore the wider scientific reach of ELIXIR-IT services.

Accordingly, the Results are organised around the progressive steps of the pathway: first, the construction of the service-related DOI dataset; second, the measurement of citation-based outcomes; third, the analysis of institutional and geographical patterns of reuse; and fourth, the extension of the same workflow to an author-based use case.

ELIXIR-IT use case dataset composition

The analysis utilized a comprehensive dataset consisting of 65 distinct ELIXIR-IT services linked to 126 unique DOIs. Among these, 41 (63%) services were associated with multiple DOIs, while the remaining 24 (37%) were linked to a unique DOI. Each service belongs to an ELIXIR-IT platform category such as Compute, Data, Tools, and Interoperability. Tools is the most represented ELIXIR-IT platform with 33 services (51%), followed by Data (19 services, 29%), Compute, and Interoperability (4 services for each, 6%) and 4 services belonging to both Compute and Tools platforms. About ELIXIR-IT service deployers, the most relevant ones are National Research Council (CNR, *Consiglio Nazionale delle Ricerche)* and University of Bari (UNIBA) with 17 services each, followed by University of Padua (UNIPD) (14 services) and University of Bologna (UNIBO) (12 services). Other providers are below 10 services. In the impact pathway logic, these DOI sets represent the scholarly outputs associated with service-related activities, such as service development, update, documentation, and dissemination.

REIMPACT results on ELIXIR-IT use case dataset

In the retrieval performed on January 12th, 2026, a total of 24,920 and 30,969 citations were retrieved from Scopus and OpenAlex, respectively. Output file generated with this analysis are fully available

in the GitHub repository. To simplify visualization, citations were grouped by ELIXIR-IT platform as shown in Figure 4.

A ranking of the top 10 most-cited ELIXIR services was generated representing *Scopus*, *OpenAlex* and *Citation of Citings,* also, for the total of services included in the input table for ELIXIR-IT (Figure 5). Overall, *Citation of Citings* sum for all services is 1,328,300 and 380,250 are referred to the second most cited services MINT (Figure 5). Within the pathway logic, Citations of Citings provide additional evidence of broader scientific reach, as they capture the citation influence of the literature that reuses or acknowledges ELIXIR-IT service-related outputs.

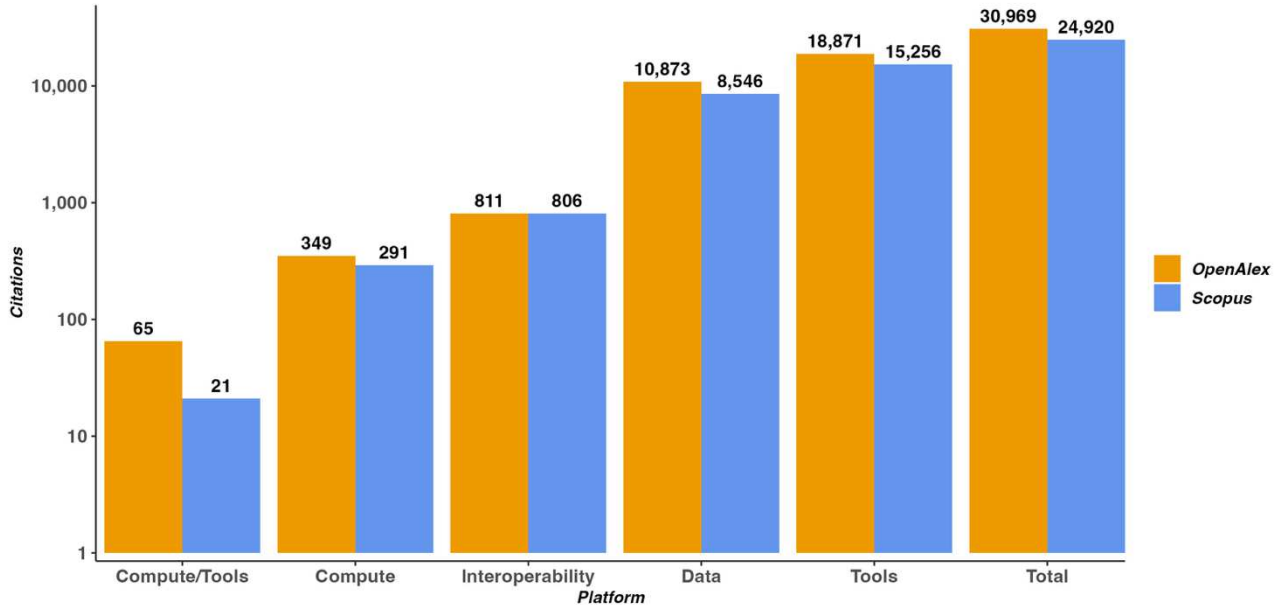


Figure 4: Total citations grouped by ELIXIR-IT Platform

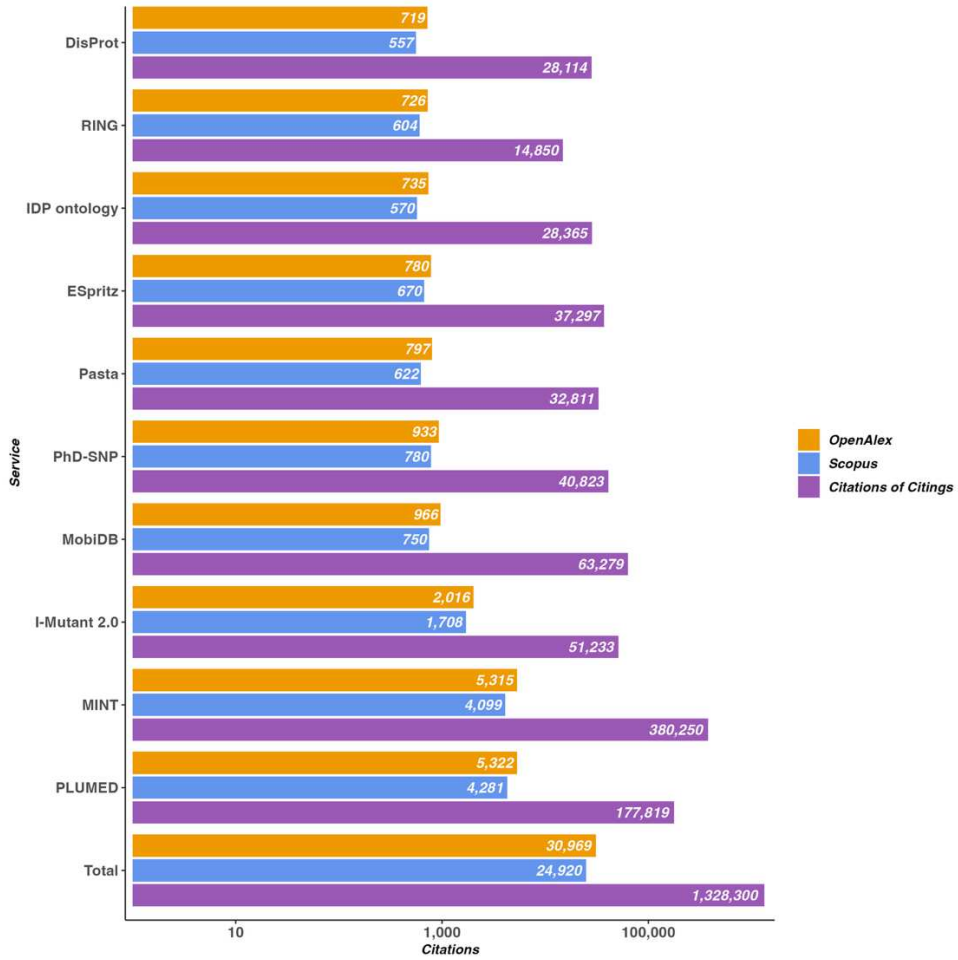


Figure 5: Top ten ELIXIR-IT Services according to OpenAlex. OpenAlex, Scopus citations, and Citations of Citings are reported as absolute values. Also, sums from all the services are reported as Total on x-axis.

REIMPACT explores also country distribution related to affiliations of authors contributing to papers identified by primary DOIs. The dataset includes affiliations from 28 countries. Italy has the highest number of author affiliations (246), followed by Hungary (HU, 30) and Argentina (AR, 28) (Figure 6B). Figure 6A shows contributions not only from Europe but also from North and Central America (e.g. Canada, Mexico), Asia (e.g. Russia) and Australia.

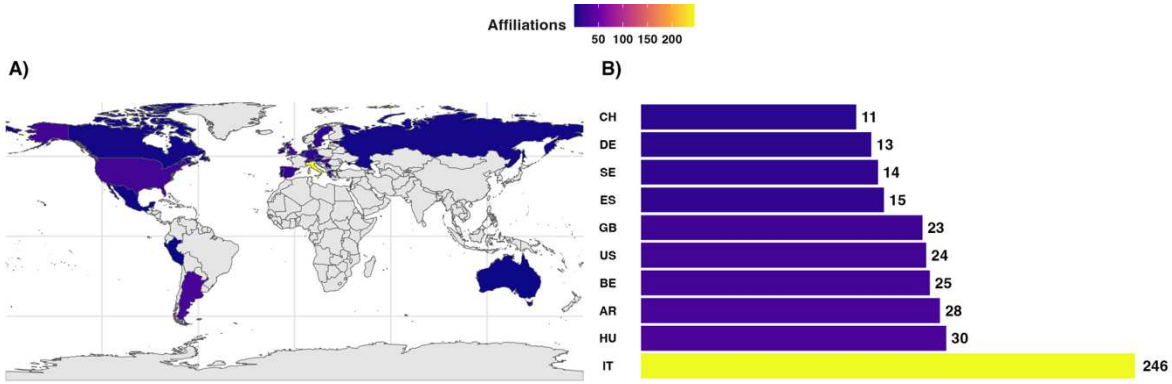


Figure 6: Geographic location of affiliations for authors contributing to primary papers. A: Choropleth colored according to the number of affiliations from countries; B: bar plot representing the 10 countries most contributing to ELIXIR-IT services (log10

transformed x-axis). Country abbreviations are reported according to ISO A2 standard (IT: Italy, HU: Hungary, AR: Argentina, BE: Belgium, US: United States of America, GB: United Kingdom, ES: Spain, SE: Sweden, DE: Germany, CH: Switzerland).

Nonetheless, REIMPACT monitoring of institutional affiliations for authors of articles identified by primary DOIs allows to trace interactions among ELIXIR-IT members. Considering each single article as a link, it is possible to plot the interactions among ELIXIR-IT members through a cord-graph (Figure 7).

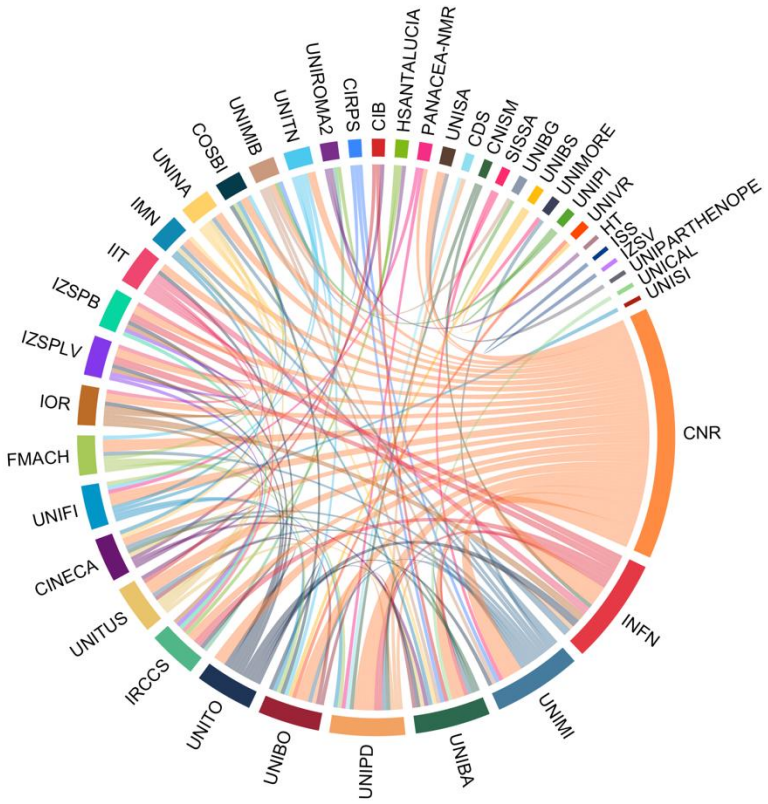


Figure 7: Chord-graph representing interactions among ELIXIR-IT members related to publication and updates of ELIXIR-IT services. Extended names are reported in Supplementary Table 1.

Also, geographic distribution of institutions affiliations for citing articles (i.e. secondary DOI) authors are summarized in Figure 8A. The United States of America (US) leads with the highest number of author affiliations (15,800), followed by China (CN) and IT with 9,116 and 6,880, respectively (Figure 8B). In the impact pathway logic, this evidence contributes to documenting the transition from outputs to outcomes and broader scientific reach.

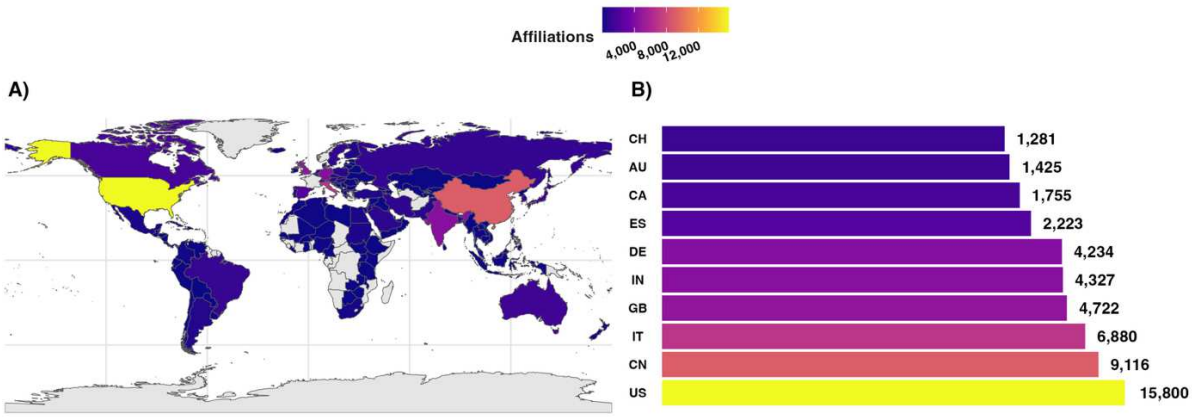


Figure 8: Geographic location of affiliations for authors contributing to secondary papers. A: Choropleth colored according to the number of affiliations from countries; B: bar plot representing the 10 countries most contributing to secondary papers (log_{10} transformed x-axis). Country abbreviations are reported according to ISO A2 standard (US: United States of America, CN: China, IT: Italy, GB: United Kingdom, IN: India, DE: Germany, ES: Spain, CA: Canada, AU: Australia, CH: Switzerland).

REIMPACT results on author search output

REIMPACT author search module on the 24 ELIXIR Head of Node names was performed on September 7th, 2026. Following are show the results for the five HoN with the highest number of associated DOIs, retrieved by interrogating both OpenAlex and Scopus. Full HoNs results are available in Supplementary Table 2 and 3.

Overall, 6,906 and 3,900 DOIs were retrieved using OpenAlex and Scopus, respectively. According to OpenAlex Balázs Győrffy (HU), Alfonso Valencia (ES), Graziano Pesole (IT), Carole Goble (GB) and, Bengt Persson (SE) are the five HoN with the highest number of associated DOIs with 1,301, 812, 593, 587, 501, respectively. On the other hand, Scopus top five was Alfonso Valencia (ES), Balázs Győrffy (HU), Graziano Pesole (IT), Carole Goble (GB) and Tim J.P. Hubbard (EU) with 582, 437, 404, 297, 255 associated DOIs, respectively.

REIMPACT was applied to the Scopus-derived DOIs for all HoNs resulting in substantial citation metrics demonstrating the scientific impact of their work (Supplementary Table 2 and 3). The sum of OpenAlex, Scopus and Secondary citations for all HoN are 521,914, 378,028 and 30,595,312, respectively. The five most cited HoN according to OpenAlex are summarized in Table 1.

Table 1: REIMPACT main module results for the most cited ELIXIR HoNs. (EU: Europe, ES: Espana, IT: Italy, HU: Hungary, GB:United Kingdom; according to ISO A2 standards)

Author	Node	Number of DOIs	OpenAlex	Scopus	Citations of Citings
Tim J.P. Hubbard	EU	255	127,458	87,723	10,291,438
Alfonso Valencia	ES	582	102,321	73,244	6,720,228
Graziano Pesole	IT	404	38,915	28,996	2,365,028
Balázs Győrffy	HU	437	38,793	29,827	1,041,936
Carole Goble	GB	297	38,304	27,288	1,162,467

Results show as Tim J.P. Hubbard is the most cited HoN according to both OpenAlex and Scopus data. Nonetheless, also secondary citations consistently positioned Tim J.P. Hubbard as the top of the chart with more than 10 million citations, with about 1 million more secondary citations than the the 5th (i.e. Carol Goble).

These results highlight the varying but substantial impact of each researcher's product, with secondary citations substantially exceeding primary citations across all three authors, illustrating how their work serves as a foundation for subsequent works.

Further, by retrieving the national affiliation of HoN coauthors, 103 countries were annotated. US was the country with the highest number of coauthors (i.e. 6,834) followed by GB, DE, ES and IT (i.e. 4,176, 2,917, 2,078, 1,863, respectively).

At the same time, REIMPACT was also able to retrieve the geographical origin of authors citing primary DOIs and 211 countries were observed. The US were the most represented followed by CN, BG, DE, ES with 328,499, 142,549, 95,618, 77,417, 46,490 citing coauthors, respectively.

Discussion

In this study, we present REIMPACT, a novel Python-based tool designed to support bibliometric impact assessment through the automated retrieval, integration, and analysis of citation data. The tool was developed to address the need for scalable and reproducible approaches for evaluating the scientific visibility and uptake of RI services, with a specific focus on the ELIXIR-IT service landscape. By integrating data from *OpenAlex* and *Scopus*, REIMPACT provides a practical framework for collecting and comparing bibliometric evidence across different sources, enabling the assessment of direct citation impact, broader citation influence, and institutional and geographical patterns of reuse. The ELIXIR-IT service use case demonstrates how REIMPACT can operationalise the impact pathway logic described in this study. In this framework, ELIXIR-IT service-related activities are linked to DOI-based scholarly outputs, whose citations and citing publications provide measurable evidence of visibility, reuse, and acknowledgement within the scientific community. The results show that REIMPACT can support this assessment by transforming service-associated DOI sets into quantitative indicators, including citation counts, citation trends, Citations of Citings, and affiliation-based information. This makes the tool particularly relevant for RI monitoring, where impact is not limited to publication productivity, but also includes the visibility, uptake, and long-term scientific contribution of services, datasets, tools, and digital resources. The analysis of ELIXIR-IT services showed that the Tools and Data platforms accounted for the largest share of citations,

suggesting that these platforms play a central role in the scientific visibility and reuse of ELIXIR-IT outputs (Figure 4).

At service level, resources such as MINT and PLUMED emerged among the most impactful services according to OpenAlex, Scopus and Citation of Citings (Figure 5). These results highlight the value of analysing impact not only at the aggregate RI level, but also at the level of individual services and platforms. Such granularity can help RIs identify highly visible resources, monitor the contribution of different service domains, and provide evidence for reporting, strategic planning, and stakeholder communication.

The geographical and institutional analyses further illustrate the potential of REIMPACT to move beyond citation counts alone. For primary DOI records, Italy emerged as the main contributing country, reflecting the national origin and development context of ELIXIR-IT services. However, the analysis of secondary DOI records showed a broader international distribution of citing authors, with the United States, China, and Italy among the most represented countries (Figure 8). This distinction between the production network and the reuse network is important for RI impact assessment, as it shows how services developed within a national Node can generate visibility and uptake across the international scientific community.

The use of both OpenAlex and Scopus proved useful for obtaining a more comprehensive view of bibliometric impact. OpenAlex generally provided broader DOI coverage and enabled the retrieval of citing works required for the computation. This makes it particularly suitable for exploring broader citation-based influence and downstream reuse. Scopus, by contrast, offered more curated bibliographic records and, in the author-based use case, provided higher citation counts and more granular institutional information for some records. These differences should not be interpreted simply as inconsistencies, but rather as evidence of the complementary nature of the two databases. Combining them allows users to compare coverage, identify metadata discrepancies, and better understand source-specific strengths and limitations. The ELIXIR HoNs author-based use case demonstrated that REIMPACT can be applied beyond RI service assessment. By retrieving and analysing DOI sets associated with individual authors, the tool can support researcher-centred, group-level, departmental, and institutional bibliometric analyses. However, results at the individual level should be interpreted with caution. Citation-based indicators provide evidence of visibility, acknowledgement, and reuse, but they do not fully capture scientific quality, individual contribution, societal value, or the broader context in which research outputs are produced.

These findings should also be interpreted in light of broader developments in scholarly communication infrastructure. Recent analyses have highlighted the increasing strategic importance of persistent identifiers and interoperable metadata in supporting more transparent and inclusive

research evaluation, particularly as scholarly contributions extend beyond traditional authorship roles and article-based outputs (Northover, *et al.*, 2025). In this context, the results reinforce the need for infrastructures capable of capturing a wider range of contributor roles and research relationships.

The observed limitations in attribution and visibility are also consistent with wider concerns about the discoverability and verification of scholarly contributions across fragmented information systems. Tools designed to support author identification and publication claiming continue to evolve in response to these challenges, underscoring the importance of accurate linkage between researchers, outputs, and affiliations. More broadly, the growing discussion around persistent identifiers points to the need to move beyond a narrow conception of scholarly credit based exclusively on authorship. ORCID and related infrastructures are increasingly presented as enablers of a more comprehensive representation of research activity, including contributions such as data curation, peer review, infrastructure work, and other forms of scholarly participation (Meadows, 2015). From this perspective, our findings support ongoing efforts to develop research assessment approaches that better reflect the diversity of contributions involved in contemporary science.

For this reason, REIMPACT should be considered as a tool for supporting evidence-based assessment, rather than as a standalone instrument for evaluating research performance. A further strength of REIMPACT is its alignment with FAIR and impact assessment principles. By relying on persistent identifiers such as DOIs, the tool enables traceable and reproducible analyses of research outputs and their citation networks. This supports transparency in monitoring activities and can help RI communicate their value to funders, stakeholders, and scientific communities. At the same time, the approach depends on the availability and quality of bibliographic metadata, the correct association between services and DOI-based outputs, and the coverage of the selected databases. These aspects represent important limitations and should be considered when interpreting the results. Overall, REIMPACT provides a flexible and scalable approach to bibliometric impact assessment. It enables RIs and academic organisations to transform DOI-based publication sets into structured evidence of scientific visibility, reuse, and broader citation influence. In doing so, it contributes to the development of transparent and reproducible monitoring practices that can support accountability, strategic decision-making, and the demonstration of public value.

Conclusion

REIMPACT is a Python-based tool designed to support bibliometric impact assessment through the automated retrieval and analysis of citation data from OpenAlex and Scopus. In this study, we applied

REIMPACT for two use cases: the assessment of ELIXIR-IT services and the analysis of publication sets associated to 24 ELIXIR HoNs.

The ELIXIR-IT use case showed how REIMPACT can operationalise an impact pathway logic by linking service-related activities to DOI-based outputs, citation-based outcomes, and broader evidence of scientific visibility and reuse. The ELIXIR HoNs author-based use case demonstrated that the same workflow can be adapted to other assessment objects, including researchers, research groups, departments, and institutions.

By providing citation counts, Citations of Citings, temporal trends, and institutional and geographical information, REIMPACT offers a scalable and reproducible solution for monitoring research impact. The tool can support RIs and academic organisations in reporting activities, strategic planning, stakeholder communication, and evidence-based impact assessment.

REIMPACT is distributed as **free software,** easy-to-use command-line tool and Conda package, with input data, output files, and scripts made available to facilitate reproducibility and reuse.

Data Availability

Data (REIMPACT input/outputs) and scripts for use cases analysis are available here https://github.com/gdefazio/REIMPACT/tree/master/manuscript

Authors Contribution

Conceptualization, F.dL, B.F., G.P., and G.D.; methodology, G.D., A.V., and B.F.; validation, G.D., and A.V.; formal analysis, G.D., and A.V.; data curation, G.D., and A.V.; writing—original draft preparation, G.D., A.V., F.dL, M.D. and B.F.; writing—review and editing, G.D., A.V., C.B., A.C., M.D.,E.P., G.P, B.F., and F.DL.; funding acquisition, G.P and F.dL. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

Vincenzo Angiulli is a PhD student within the European School of Molecular Medicine (SEMM).

This work (or Part of this work) was granted by the European Commission – NextGenerationEU, Project ELIXIRxNextGenIT, "ELIXIR x NextGenerationIT: consolidation of the Italian Infrastructure for Omics Data and Bioinformatics", project code n. IR0000010.

References

Bornmann, L. and Daniel, H. (2008) "What do citation counts measure? A review of studies on citing behavior," *Journal of Documentation*, 64(1), pp. 45–80. Available at: https://doi.org/10.1108/00220410810844150.

Broadus, R.N. (1987) "Toward a definition of 'bibliometrics,'" *Scientometrics*, 12(5–6), pp. 373–379. Available at: https://doi.org/10.1007/BF02016680.

Citation Analysis in Research Evaluation (2005). Berlin/Heidelberg: Springer-Verlag (Information Science and Knowledge Management). Available at: https://doi.org/10.1007/1-4020-3714-7.

Culbert, J.H. *et al.* (2025) "Reference coverage analysis of OpenAlex compared to Web of Science and Scopus," *Scientometrics*, 130(4), pp. 2475–2492. Available at: https://doi.org/10.1007/s11192-025-05293-3.

De Leo, F. *et al.* (2025) "Using text-mining to measure the scientific impact and legacy of ELIXIR, a distributed research infrastructure for life science data," *F1000Research*, 13, p. ELIXIR-1547. Available at: https://doi.org/10.12688/f1000research.158264.3.

De Ruyck, K. *et al.* (2023) "ELIXIR's Impact Toolkit," 11 September. Available at: https://doi.org/10.7490/f1000research.1119614.1.

Digital Science & Research Solutions Inc. (2026) *Dimensions Author Check for Government & Funders*, *Dimensions*. Available at: https://www.dimensions.ai/sector/government-and-funders/dimensions-author-check-for-government-funders/ (Accessed: October 26, 2026).

Directorate-General for Research and Innovation European Commission) (2018) *Turning FAIR into reality: final report and action plan from the European Commission expert group on FAIR data*. Publications Office of the European Union. Available at: https://data.europa.eu/doi/10.2777/1524 (Accessed: March 22, 2026).

Directorate-General for Research and Innovation (European Commission) (2019) *Research infrastructures make science happen*. Publications Office of the European Union. Available at: https://data.europa.eu/doi/10.2777/446084 (Accessed: March 22, 2026).

Directorate-General for Research and Innovation (European Commission) (2020) *Supporting the transformative impact of research infrastructures on European research: report of the High Level Expert Group to assess the progress of ESFRI and other world class research infrastructures towards implementation and long term sustainability*. Publications Office of the European Union. Available at: https://data.europa.eu/doi/10.2777/3423 (Accessed: March 22, 2026).

ESFRI (2019) *Monitoring of Research Infrastructures Performance*. Available at: https://www.esfri.eu/monitoring-research-infrastructures-performance-december-2019 (Accessed: December 17, 2024).

ESFRI (2026) *ESFRI Roadmap 2026 | www.esfri.eu*. Available at: https://www.esfri.eu/roadmap-2026 (Accessed: October 26, 2026).

ESFRI Roadmap (2021). Available at: https://roadmap2021.esfri.eu/ (Accessed: April 17, 2026).

Garfield, E. (2006) "The History and Meaning of the Journal Impact Factor," *JAMA*, 295(1), pp. 90–93. Available at: https://doi.org/10.1001/jama.295.1.90.

Garrard, C. *et al.* (2025) "Fostering and sustaining collaborative innovation: Insights from ELIXIR Europe's life science Communities." F1000Research. Available at: https://doi.org/10.12688/f1000research.168288.1.

Kolar, J. *et al.* (2023) *ESFRI Policy Brief on Assessment of Impact of Research Infrastructures*. Zenodo. Available at: https://doi.org/10.5281/zenodo.8091633.

Leydesdorff, L. and Milojević, S. (2015) "Scientometrics," *International Encyclopedia of the Social & Behavioral Sciences*. Elsevier, pp. 322–327. Available at: https://doi.org/10.1016/B978-0-08-097086-8.85030-8.

Martin, C.S. *et al.* (2021) "Demonstrating public value to funders and other stakeholders—the journey of ELIXIR, a virtual and distributed research infrastructure for life science data," *Annals of Public and Cooperative Economics*, 92(3), pp. 497–510. Available at: https://doi.org/10.1111/apce.12328.

Meadows, A. (2015) "Peer Review - A Discussion," *ORCID*, 29 September. Available at: https://info.orcid.org/peer-review-a-discussion/ (Accessed: October 28, 2026).

Monitoring of Research Infrastructures Performance, December 2019 | www.esfri.eu (no date). Available at: https://www.esfri.eu/monitoring-research-infrastructures-performance-december-2019 (Accessed: April 17, 2026).

Mons, B. *et al.* (2017) "Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud," *Information Services and Use*, 37(1), pp. 49–56. Available at: https://doi.org/10.3233/ISU-170824.

Northover, R. *et al.* (2025) "Trusted Identity in Academic Publishing - Part 2: The Researcher Identity Verification Framework."

OECD (2019) *Reference framework for assessing the scientific and socio-economic impact of research infrastructures*. OECD Science, Technology and Industry Policy Papers 65. Available at: https://doi.org/10.1787/3ffee43b-en.

Priem, J., Piwowar, H. and Orr, R. (2022) "OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts." arXiv. Available at: https://doi.org/10.48550/arXiv.2205.01833.

Science Europe, S. (2025) *SE Response to ERA Act Call for Evidence*, *Science Europe*. Available at: https://scienceeurope.org/our-resources/se-response-to-era-act-call-for-evidence/ (Accessed: October 26, 2026).

Shillum, C. (2026) "Beyond Authorship: Partnering with the Publishing Community to Recognize Peer Review Service," *ORCID*, 15 September. Available at: https://info.orcid.org/beyond-authorship-partnering-with-the-publishing-community-to-recognize-peer-review-service/ (Accessed: October 26, 2026).

Smith, A. *et al.* (2020) "ELIXIR Position Paper on Suitable Business Models for Open Data," *F1000Research*, 9(65), p. 65. Available at: https://doi.org/10.7490/f1000research.1117782.1.

SMT (2024) "Trusted Identity in Academic Publishing," *STM Association*, 31 October. Available at: https://stm-assoc.org/document/trusted-identity-in-academic-publishing/ (Accessed: October 26, 2026).

The European Strategy on Research and Technology Infrastructures - Research and innovation (2025). Available at: https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/european-strategy-research-and-technology-infrastructures_en (Accessed: October 15, 2026).

Wilkinson, M.D. *et al.* (2016) “The FAIR Guiding Principles for scientific data management and stewardship,” *Scientific Data*, 3(1), p. 160018. Available at: https://doi.org/10.1038/sdata.2016.18.

Supplementary

Supplementary Table 1: Name and Abbreviation for institutions reported in Figure 6.

National Research Council	CNR
Istituto Nazionale di Fisica Nucleare	INFN
University of Milan	UNIMI
University of Bari Aldo Moro	UNIBA
University of Padua	UNIPD
University of Bologna	UNIBO
University of Turin	UNITO
Istituti di Ricovero e Cura a Carattere Scientifico	IRCCS
Università degli Studi della Tuscia	UNITUS
Cineca	CINECA
University of Florence	UNIFI
Fondazione Edmund Mach	FMACH
Istituto Ortopedico Rizzoli	IOR
Istituto Zooprofilattico Sperimentale del Piemonte Liguria e Valle d'Aosta	IZSPLV
Istituto Zooprofilattico Sperimentale della Puglia e della Basilicata	IZSPB
Italian Institute of Technology	IIT
Mario Negri Institute for Pharmacological Research	IMN
University of Naples Federico II	UNINA
The Microsoft Research - University of Trento Centre for Computational and S ystems Biology	COSBI

University of Milano-Bicocca	UNIMIB
University of Trento	UNITN
University of Rome Tor Vergata	UNIROMA2
Centro Interuniversitario di Ricerca Per lo Sviluppo sostenibile	CIRPS
Consorzio Interuniversitario per le Biotecnologie	CIB
Fondazione Santa Lucia	HSANTALUCIA
Interuniversity Consortium for Magnetic Resonance	PANACEA-NMR
University of Salerno	UNISA
Città della Speranza Foundation	CDS
National Interuniversity Consortium for the Physical Sciences of Matter	CNISM
Scuola Internazionale Superiore di Studi Avanzati	SISSA
University of Bergamo	UNIBG
University of Brescia	UNIBS
University of Modena and Reggio Emilia	UNIMORE
University of Pisa	UNIPI
University of Verona	UNIVR
Human Technopole	HT
Istituto Superiore di Sanità	ISS
Istituto Zooprofilattico Sperimentale delle Venezie	IZSV
Parthenope University of Naples	UNIPARTHENOPE
University of Calabria	UNICAL
University of Siena	UNISI

550

Supplementary Table 2: For each ELIXIR Head of Node, the name (Author), the affiliation (Affiliation) and, the number of retrieved DOIs (Number of DOIs) by using OpenAlex are shown.

Supplementary Table 3: For each ELIXIR Head of Node, the name (Author), the affiliation (Affiliation) and, the number of retrieved DOIs (Number of DOIs) by using Scopus are shown.

Supplementary File 1: REIMPACT main module results about DOIs reported in Supplementary Table 3 and associated to ELIXIR HoNs. This table represents the summary of results reporting HoNs associated DOIs and primary citations obtained interrogating OpenAlex and Scopus and Citations of Citings.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarytable1.xlsx
- Supplementarytable3AuthorsDOIsScopus.xlsx
- SupplemetaryTablesCaptions.docx
- Supplementarytable2AuthorsDOIsOpenAlex.xlsx
- Supplementarytable4REIMPACTwScopusDOIs.xlsx