

LibRA: A Tool for Researcher Metrics Management

Chiara Rebuffi¹, Roberto Cavanna², Paolo Uva²

¹Scientific Direction, IRCCS Istituto Giannina Gaslini, Genoa, Italy, ²Clinical Bioinformatics, IRCCS Istituto Giannina Gaslini, Genoa, Italy

Abstract. Research conducted at the Gaslini Pediatric Hospital is primarily funded by the Italian Ministry of Health based on bibliometric indicators of scientific performance. The analysis of these metrics can be facilitated by software that automates bibliometric processes and calculations for use by administrative staff in research evaluation. To this end we developed LibRA (Library for Research Assessment), a Python-based application which leverages application programming interface (API) technology to query relevant databases (e.g. Scopus – Elsevier) and retrieve performance indicators in real time

Keywords. Research evaluation, bibliometric analysis, APIs, citation databases

1. Introduction

The Giannina Gaslini Institute is a pediatric hospital that integrates healthcare and assistance with research activities, aiming to advance biomedical knowledge. This is largely made possible thanks to funding from the Italian Ministry of Health (MoH). Research and clinical activities undergo a competitive assessment among the IRCCS based on the scientific performance for the allocation of Ricerca Corrente funds, as well as additional funding opportunities such as Ricerca Finalizzata, Piano Nazionale di Ripresa e Resilienza (PNRR), and thematic networks.

Approximately 10,000 researchers work in these institutes whose career progression is regulated by a merit-based, competitive system aligned with the standards of public healthcare research institutions.

Performance evaluation is therefore conducted at multiple levels — including the institute, research groups and individual researchers. The analysis of the parameters on which this evaluation is based plays a crucial role in the allocation of funding.

This evaluation increasingly relies on bibliometric indicators, such as the h-index and citation index which are calculated using proprietary databases selected by the MoH. Among these are databases provided by Elsevier, such as Scopus (for citation data) and SciVal (for citation analysis) (Research Intelligence, 2019). The relevant regulations specify not only the tools to be used for bibliometric analysis, but also the methodology to be followed and the thresholds to be met. For example, the use of unique author identifiers (e.g. Scopus ID) is mandatory for retrieving scientific publications. To complete the literature collection PubMed APIs were also used.

These assessments are managed by the staff of the Scientific Directions, who are involved

in the preliminary eligibility checks for funding calls, in monitoring the validated results communicated by the MoH, and in executing the internal redistribution of funds.

These processes require significant staff involvement and are highly time-consuming, as the data often undergo multiple normalization steps to ensure fairness across different research areas, researcher seniority, and other variables. Moreover, these evaluations typically rely on large-scale data extraction and manipulation which, in the absence of adequate technological support, may introduce a considerable margin of error.

To address these challenges and support the transition towards a data-driven research hospital, a web application for automating bibliometric data collection has been proposed as a pilot initiative. The hosting of LibRA on the GARR Virtual Data Center is currently under evaluation.

2. Methods

2.1 System overview

LibRA is composed of two web interfaces - Administrators and Researchers - built using Streamlit (<https://streamlit.io>), a Python-based open-source framework for creating interactive data applications. Streamlit provides the web interface, while Pandas, Pycogp2, and Requests python libraries are used for data handling, database interaction, and API calls, respectively. Configuration settings, such as API keys and parameters, are stored securely in a `secrets.toml` file. Authentication for the Researcher web application is implemented using basic session handling within Streamlit, while the Administrator interface is protected through internal network access and/or external authentication mechanisms, depending on deployment context.

The Administrator and Researcher web applications share a common backend and are connected to a PostgreSQL relational database and external bibliometric data sources via public Scopus (https://dev.elsevier.com/api_docs.html) and PubMed APIs.

An overview of the architecture is provided in Figure 1. At the core of the system is a central server that coordinates data retrieval, processing, and interaction between components.

The core workflow involves:

- Importing demographic data via Excel file
- Fetching publication metadata and citation metrics from Scopus and biomedical literature from PubMed through APIs
- Computing bibliometric metrics in Python across different timeframes. The system also detects first, last, and corresponding authorship roles to identify the contribution in each publication
- Storing metrics and raw metadata in a PostgreSQL database for reuse and comparison
- Displaying and editing information via the Admin or Researcher interface to explore and filter data, view metrics in tabular form, and export reports and publication lists in Excel format

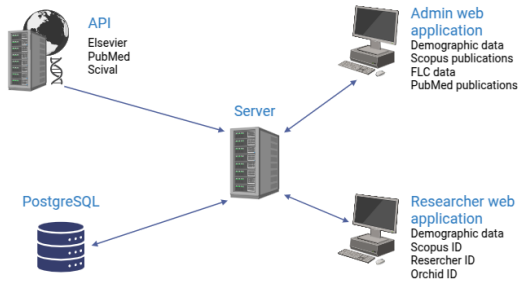


Fig. 1
Overview of
the software
architecture

3. User interface

LibRA has two web interfaces for managing and viewing bibliometric data.

3.1 Admin web application

The Admin interface is used to manage demographic data, update publication records, compute bibliometric indicators and authorship roles (first, last, corresponding) from Scopus and aggregate metrics across departments or researcher age groups. It is composed of 8 pages:

- Home. Includes several subsections:
 - Personal Data Import – Import of researcher demographic data via Excel file. Each row includes name, surname, department, Scopus ID, ORCID and ResearcherID
 - Register – Retrieval of publications via Scopus API. For each researcher, h-index, number of publications and total number of citations over three time frames (entire career, last 10 and 5 years) are computed
 - FLC from Scopus – Retrieval of First, Last, and Corresponding authors (FLC)
 - PubMed – Retrieval of publications available on PubMed
- Demographic Data. List of researchers. Can be sorted, filtered by year and exported to Excel
- Researcher Details. Metrics and PubMed publications not yet listed in Scopus. Also enables updates to demographic data
- Register. Table of researcher bibliometric metrics sorted by h-index
- Statistics. Individual and aggregated metrics (median h-index by Research Unit, including all members or under 40)
- Requests. Interface for approval of external requests to join the researcher list
- Scopus. Publications from Scopus, grouped by year and researcher, with a secondary table for missing PubMed entries to highlight gaps
- PubMed. Publications from PubMed

3.2 Researcher web application

This allows researchers to update their own demographic information, and access to the list of publications.

- Researcher Dashboard (login required). Publications (reference, Scopus Author ID, DOI, PubMed ID, citation count, and authorship role) and metrics (h-index, number of publications, citation count, and the number of publications where the author is listed as FLC). Allows updating of demographic data
- Public Request Form (no login). Researchers not in the database may request inclusion, subject to Administrator review

4. Use case

The software has been under testing for approximately one year at the Giannina Gaslini Institute. In early 2025, it was updated with the list of researchers from the Workflow della Ricerca, a web-based platform of the MoH, which includes 460 individuals working at our Institute with a valid Scopus ID. Metrics were also aggregated for each of the 68 RU. The most recent update carried out in June 2025 - four months after the previous one - required less than 5 minutes and showed changes in the number of publications or related metrics for 426 out of 460 researchers (93%).

5. Conclusions

Some limitations emerged that suggest areas for improving LibRA's performance. In particular, data collection was hindered by insufficient metadata in the source database, resulting in occasional inaccuracies in author attribution.

For example, it is not currently possible to identify shared author positions in articles because this information is not tracked by Scopus. We plan to identify possible solutions to these biases in future releases.

At the same time, outreach initiatives for researchers have been launched: our Code of Research Integrity promotes regular updating of unique identifiers to prevent information loss during data extraction due to duplicate author profiles.

Finally, we plan to add novel functionalities based on complex performance analyses from SciVal.

In our Institute LibRA has proven to significantly reduce the time required for data collection and to minimize the risk of human error. We believe that it could also be useful to other IRCCSs, because to our knowledge no tool calculates all the bibliometric indicators provided by LibRA as required by MoH (e.g. Moschini and Molinari, 2022).

6. Data Availability

The source code along with documentation and Excel template for demographic data import is available at <https://github.com/igg-bioinfo/igg-biblio> (Administrators) and <https://github.com/igg-bioinfo/igg-researcher> (Researchers).

References

Moschini U., Molinari E. (2022). Designing ecosystems to enable recognition and adoption of Open Science measures [Internet]. Genoa Open Access Week. Available from: https://openscience.unige.it/sites/openscience.unige.it/files/2022-12/07.1_GenOAweek2022_

Moschini.pdf

Research Intelligence (2019). Research Metrics Guidebook [Internet]. Elsevier. Available from: <https://brand.elsevier.com/share/nSegnbJj4TG2qKKGxrYE>

Authors



Chiara Rebuffi chiararebuffi@gaslini.org

Graduated in Archival Sciences, Documentation, and Library Science, she has been working for about 15 years in Libraries and Scientific documentation services for biomedical research institutions, currently works at the IRCCS Gaslini institute in Genoa. She is involved in the evaluation of research and researchers, both at national and international levels, and regularly uses tools, such as citation databases, to assess performance.

Roberto Cavanna robertocavanna@gaslini.org

Software developer with over twenty years of experience, including the last ten dedicated to scientific research at the IRCCS Giannina Gaslini Institute. Specialized in designing and maintaining web applications for clinical registries and research trials, he combines deep technical expertise with a strong understanding of research workflows and data integrity.



Paolo Uva paolouva@gaslini.org

Coordinator of the Bioinformatics Unit at the IRCCS Giannina Gaslini Institute in Genoa. The activities of his research group focuses on the development and application of methods for the analysis of clinical and omics data generated through high-throughput technologies. The group also works on the digitalization and automation of procedures to ensure compliance with FAIR principles, supporting the institute's transition toward a data-oriented model.