

AutoBib: Automating Bibliometric Analysis Reporting and Scientific Collaborations

Benjamin Lepourtois Adji TOURE, Ecole de Technologie Superieure Vanessa Ayotte, Ecole de Technologie Superieure

Library technician for research support services such as research data management, bibliometrics and open access.

Sonia Seck, Ecole de Technologie Superieure Mrs. Judith Boissonneault, Ecole de Technologie Superieure Prof. Lokman Sboui, École de technologie supérieure

Lokman Sboui is currently a Professor in the Department of System Engineering at École de Technologie Supérieure (ÉTS) in Montréal, Canada, where his research interests focus on energy efficiency and performance analysis of wireless communications, LEO satellites for IoT, industrial automation and digital twins, urban air mobility (UAM), cognitive radio systems, and VoIP protocols, contributing to advancements in wireless communication technologies, automation, and the development of emerging technologies that impact sectors such as transportation, telecommunications, and smart city initiatives.

AutoBib: Automating Bibliometric Analysis Reporting and Scientific Collaborations

Abstract

Understanding research impact and institutional collaborations require accurate and timely bibliometric analyses. However, manually processing data to produce bibliometric reports is often time-consuming and labor-intensive. This paper introduces AutoBib, an automated solution developed to streamline the generation of bibliometric reports. AutoBib was developed at the École de Technologie Supérieure (ÉTS) library in Montreal with contributions from students at the École Centrale de Nantes. AutoBib integrates data extraction from bibliometric databases such as Scopus and SciVal through Python libraries and features a user-friendly interface developed using a QT dialog box. Additionally, AutoBib automates MS Excel formatting and report generation in MS Word using VBA scripting. While currently optimized for Scopus and SciVal databases, AutoBib's architecture can potentially support integration with other databases to enhance interdisciplinary analysis. The deployment of AutoBib has resulted in significant time savings for the ÉTS library team, enabling the generation of in-depth reports on institutional collaborations with minimal manual effort. This paper discusses the development process, challenges encountered, and the benefits of AutoBib for bibliometric reporting efficiency. A comparative analysis with manual methods highlights its advantages and limitations, along with the potential for broader adoption in academic and research libraries.

Keywords: Bibliometric Analysis, Automated Reporting, Data Extraction, Research Impact, Academic Libraries

Introduction

Evaluating research impact and analyzing scientific collaborations are fundamental tasks that contribute to the advancement of knowledge and strategic planning in academic institutions. Bibliometric analysis is a crucial tool in this context, as it provides quantitative measures of scholarly output and influence^{1,2}. Bibliometrics uses citation counts, publication metrics, and coauthorship networks, allowing researchers and funding agencies to assess the dissemination and reception of scientific work^{3,4}.

Bibliometric analysis has evolved significantly over the past decades. Foundational works provide a concise overview of the field, offering an accessible introduction to core concepts⁵ and presenting a mathematical theory of citing that elucidates the stochastic nature of citation accruay⁶. Recent studies have examined the evolution and development of scholarly

communication, highlighting emerging trends in publication practices and citation behavior⁷. In addition, university libraries play a significant role in performing bibliometric analysis: they support researchers and administrators by providing insight into research performance, collaboration patterns, and impact metrics through different reports that can help inform strategic decisions and policy development⁸.

Traditional bibliometric analysis methods often involve manual data collection and processing, which can be time-consuming and labor-intensive⁹. The increasing volume of scholarly publications amplifies these challenges, making it difficult for researchers and institutions to keep pace with the rapidly expanding literature¹⁰. Reliance on multiple databases such as Scopus, SciVal, Web of Science, and Google Scholar introduces variability in data retrieval, complicating comprehensive analysis⁹. Although additional databases could be utilized, this study primarily focuses on data retrieval from Scopus and SciVal.

To address these challenges, user-friendly tools that automate bibliometric analysis and reporting are needed. Automation has the potential to reduce the time and effort required to generate reports, allowing university libraries and academic institutions to focus on strategic decision-making rather than data processing¹¹. Moreover, such tools should be accessible to users without programming expertise, ensuring wider adoption and functionality across disciplines.

This paper introduces AutoBib (**Auto**mation of **Bib**liometric reports), an automated tool for generating detailed bibliometric reports and analyzing scientific collaborations. AutoBib was developed at the École de Technologie Supérieure (ÉTS) library in Montreal, a french-language university. AutoBib was designed by a team of librarians, information science researchers, and software developer interns as a non-commercial initiative to automate bibliometric reporting for internal use, with potential benefits for similar institutions.

AutoBib was built to integrate data from bibliographical and bibliometric databases. Currently, AutoBib utilizes data extraction techniques from the Elsevier' Scopus and SciVal platforms, which provide access through Application Programming Interfaces (APIs) and databases. To extract the data from Elsevier's Databases, AutoBib rely on an extended version of Rose and Kitchin "pybliometrics" libraries ¹². AutoBib is also designed following best practices in object-oriented programming to ensure modularity, scalability, and ease of use. Its intuitive user interface enables users to extract and analyze data without requiring programming knowledge.

AutoBib automates tasks such as data retrieval, report formatting in MS Excel, and document generation in MS Word using Visual Basic for Applications (VBA) scripting. This significantly improves efficiency and accuracy in bibliometric reporting, especially for university libraries tasked with supporting researchers and administrators.

The contribution of this work is the provision of a novel, efficient tool for bibliometric analysis, supporting the research community, and aiding in modern scholarly communication. Automation saves time (reports are generated in minutes instead of hours) and resources, enhancing the ability to make informed decisions based on comprehensive and up-to-date data.

This paper presents the development and implementation of AutoBib, detailing how it automates

the extraction and analysis of bibliometric data. This work addresses the challenges encountered during the development process, describe the features and functionalities of the tool, and highlight its advantages over existing solutions. A comparative analysis underscores AutoBib's potential for broader adoption in academic and research libraries, particularly in facilitating engineering research evaluation.

1 Background on Bibliometric Reports

Gaining significant insights from bibliometric indicators with a very limited set of data is a delicate task. Bibliometric indicators are useful for identifying trends in the research landscape when analyzing a large set of documents¹³. However, when dealing with a smaller set of documents, metrics offer limited information, and compiling indicators may be inappropriate if the set is too limited or contains extreme values¹⁴.

1.1 Bibliometric Reports for Individual Researchers

Faculty members request information on their publications when applying for grants, awards or tenure. To meet this need, the library offers personalized bibliometric reports. These reports consist of a first section, common to all users, containing automated graphics highlighting different metrics such as number of publications by year, and article by journal impact. The second section of the report is tailored to the researcher's profile, career stage, scientific or teaching activities, and specific elements they wish to highlight.

Developing an automated solution with flexible parameters for meaningful metrics in each report generated is a priority. The library team is aware of the misuse of bibliometric indicators in research evaluation and is committed to providing contextualized data. The solution should include features to exclude or include specific document types. In certain engineering disciplines, particularly in computer science, conference proceedings are often considered significant contributions. Some argue that top-tier conferences in these fields are sometimes regarded as more prestigious than high-impact factor journals¹⁵. However, it should be noted that this perception varies by subdiscipline and regional academic practices. Additionally, it is crucial to exclude publications where the author contribution is not significant (e.g., numerous authors), choose a longer date range, or hide metrics that are not relevant to the demand. The automated solution needed to meet most of these criteria. When automation was not possible, visual marks on MS Excel or MS Word documents indicate areas where user review is strongly recommended.

1.2 Bibliometric Reports for Collaboration with Other Countries or Institutions

The second phase of the automated solution development focused on a highly requested report type: the collaboration report. Our university ÉTS aims to expand partnerships abroad and gain insight into the existing links between its research community and foreign entities. This goal is driven by strategic objectives of the institution's internationalization policy. Mapping the collaboration landscape is also useful for meeting due diligence requirements under the Canadian Policy on sensitive technology research and affiliations of concern¹⁶. Various offices frequently request collaboration data from the ÉTS library team on short notice. These requests typically do not require extensive analysis; requesters often refine their needs after reviewing initial data

extractions. Consequently, providing rapid and accurate publication metadata makes this reporting task ideal for automation.

When generated manually, the collaboration report is prepared by a single library employee. Depending on the request, the method for information retrieval and the deliverable can vary. Producing a collaboration report is not as labor intensive as other types of deliverables.

Although SciVal is used to retrieve and export bibliometric data, its user interface imposes certain limitations that slow down data collection and add extra steps to handle duplicates. The interface allows viewing and extracting publication sets only over seven fixed-year ranges. For example, assembling a complete list of institutions collaborating with ÉTS from 2014 to the present cannot be done in a single step. Users must adjust year ranges, apply filters, merge multiple exported files, and then identify and consolidate duplicate entities.

Furthermore, collecting a complete list of authors from a large publication set is cumbersome and error-prone, as SciVal supports exporting only 500 authors at a time. This requires splitting the publication set into subsets, exporting each subset, and then merging the resulting files. Even with tools such as MS Excel formulas and the Fuzzy Lookup extension*, identifying and removing duplicate author data remains time-consuming.

Requests for collaboration reports typically involve identifying faculty members, from our university ÉTS, who collaborate with various entities such as a country, an institution or a group of researchers. Analysts from the Bureau of Prevention and Security of ÉTS may also request data on potential collaborators and groups of institutions, such as those listed on the Named Research Organizations List (NRO), whose collaboration in certain research areas could pose a risk to Canada's national security ¹⁷. To secure federal funding, it is necessary to demonstrate that collaborations are not occurring in sensitive technology research areas with researchers affiliated with organizations of concern. Consequently, publication data remains a crucial source of information for analysts performing these verifications.

1.3 Automated Solution Implementation in the ÉTS Library

Requests for data analysis services related to scientific output have increased from various users within the institution, including faculty members, the Dean of Research Office, the International Relations Office, the Bureau of Partnership Initiatives, the Bureau of Prevention and Security, departments, and research groups. To improve response time to the ever-increasing demands, the ÉTS library team sought ways to optimize report production time. They were motivated to collaborate with the project developers to create a time-saving solution that also allowed them to reflect on priorities, improve deliverables, and enhance workflow. The chosen name for this project is AutoBib (Automation of Bibliometric reports).

The early involvement of library staff in the development of AutoBib facilitates its adoption, as their specific needs and work environment were taken into account. To ease the transition from a manual method of producing these reports to an automated one, training and guidance were

^{*}The Fuzzy Lookup extension is a Microsoft Excel add-in that uses fuzzy matching algorithms to identify nearduplicate text strings, even when they differ slightly in spelling or format. It helps merge datasets with inconsistent entries, thereby improving data quality.

provided from the automated solution developer, and visual aids were included within the application to assist users in querying the tool.

2 Architecture and Design of AutoBib

The development of AutoBib focused on creating a robust, simple, scalable, and user-friendly architecture. The main goal is to integrate tools for retrieving bibliometric data about researchers and institutions, process these data according to librarians' expectations, and generate a bibliometric report in MS Word format. We developed a user interface specifically designed for librarians, with design choices aimed at facilitating the use of the tool. An example of an AutoBib generated bibliographic report is shown in Figure 1.



Figure 1: Example of the bibliometric report for individual researchers generated by AutoBib (translated from French to English).

Note that the report in Figure 1 is generated by the library team as a service to visualize data of researchers. Although such report is not available in SciVal or Scopus, the underlying data originate from these two databases. Before AutoBib, the library team conducted numerous manual procedures—including issuing commands, extracting data, and performing various manipulations—to produce the graphs. The added value of AutoBib lies in its ability to automate the generation of such reports in minimal time while reducing human errors.

2.1 Modular Architecture

AutoBib adopts a modular design to simplify maintenance and future improvements. The first module retrieves bibliometric data through APIs that interact with the Scopus and SciVal databases. Specifically, HTTP requests are sent to extract the data, with error management mechanisms (including API keys and tokens) simplifying the usage process. Table 1 provides an overview of AutoBib's core functionalities, from data retrieval to report generation, summarizing the automation steps and key features:

Step	Description
Data Retrieval	Extracts metadata and metrics using Scopus and SciVal APIs
Data Processing	Uses Python libraries to clean, analyze, and aggregate data
Report	Automates output formatting in MS Word and MS Excel
Generation	using VBA macros
Customizable	Potential filtering, entity-specific analyses, and user-defined
Outputs	templates

Table 1: Overview of AutoBib's Core Functionalities

The data processing module plays a central role, as the Scopus and SciVal APIs only return raw bibliometric data. At present, it is necessary to recalculate statistics to replicate the information available on their websites. We use Pandas library in Python to process and analyze this data. The processing workflow includes identifying co-author networks, analyzing collaboration patterns, and detecting citation trends. Flexible functions allow for customized data manipulation to meet the various needs of users.

Once we process the data, we export it—preferably to an MS Word document—to allow further customization based on the specific needs of researchers or institutions. In this module, we use editable MS Excel and MS Word templates that include VBA macros (Visual Basic for Applications). These macros automate the structuring of data into professional-quality tables, graphs, and diagrams while minimizing user intervention.

2.2 Finite State Machine Implementation

AutoBib uses a Finite State Machine (FSM) approach to manage workflows while maintaining a simple software architecture. Each state corresponds to a specific step, such as configuring API keys, performing bibliometric analysis, or mapping collaborations. Transitions between states are triggered by user inputs and system validations. This approach ensures a logical progression,

minimizes errors, and enhances both the reliability and user-friendliness of the tool as shown in Figure 2.



Figure 2: Finite State Machine Representation of AutoBib

Figure 2 illustrates the state machine workflow implemented in AutoBib, highlighting its modular and logical progression through various stages of bibliometric analysis. The state machine begins at state -1, which ensures the user has entered valid API keys, tokens, and paths required to access data from bibliometric databases.

• **State 0**: This is the main decision point where the user selects the type of document to generate, either an individual bibliometric report or a collaboration report.

- **States 1–5**: These states are dedicated to individual researchers (researcher's name, ID, type of document, desired year range, type of document to highlight)
- States 11–17: These states handle collaboration reports (selection of Entity A and Entity B, specification of predefined or custom lists, entering IDs or country name, and defining the year range for analysis).

The transitions between these states are triggered by user inputs and validations. This FSM approach minimizes errors and improves the efficiency and user-friendliness of the tool.

2.3 Interface Design

The user interface of AutoBib, as shown in Figure 3, is designed using PySide6 (QT for Python) to provide an accessible and intuitive experience for users. It offers customizable options for data input and report preferences, guiding users step-by-step through the bibliometric analysis and collaboration reporting process. This interface eliminates the need for programming knowledge, making it especially suitable for librarians and administrators.

Figure 3 showcases a clean and interactive interface with the following key elements:

- **Toolbar**: The red toolbar at the top contains essential functions like navigation and API configuration, which can be adjusted for convenience.
- **Command Instructions**: The top section includes clear instructions for using the software, such as default commands and value separators, ensuring ease of use.
- **Text Prompt System**: The interface employs a text-driven input system, where users can select options or provide data through the entry bar at the bottom. This system dynamically updates based on user selections, reducing errors and enhancing workflow clarity.
- **Document Selection**: Users are prompted to choose between generating either an individual bibliometric report or a collaboration report, ensuring flexibility in report generation. Subsequent options, such as specifying entities (e.g., researchers or institutions) and their relationships, further refine the report parameters.

2.4 Functionality and Features

AutoBib is designed to simplify and automate the processes involved in bibliometric analysis and reporting. These reports fall into two categories: personalized bibliometric reports, which assess the research impact of an individual researcher over a specific period, and collaboration reports between ÉTS and various entities, such as countries or institutions. To this point, AutoBib connects to Scopus and SciVal APIs to retrieve essential data, including publication records, citation metrics, and detailed collaboration insights. Connection between AutoBib and other data sources could eventually be added to automatize other sections of requested reports. By leveraging optimized query mechanisms, AutoBib ensures accurate and efficient retrieval of relevant data, even in cases where author names or affiliations may lead to ambiguity. For instance, it intelligently handles homonymy issues by refining searches to improve relevance. AutoBib also simplifies data retrieval by allowing users to input predefined entities or validated



Figure 3: AutoBib User Interface (translated from French to English).

Scopus identifiers directly. This flexibility accommodates different workflows whether analyzing specific researchers or exploring broader institutional collaborations.

AutoBib goes beyond simple data retrieval by processing the information into actionable insights. Citation trends are analyzed over customizable time periods, allowing users to track the evolution of research impact. Users can generate detailed insights into partnerships between institutions, researchers, or countries. For example, in cases involving collaborations between ÉTS and a specific country, AutoBib provides an MS Word document containing multiple tables. These include the list of collaborating institutions, the names of the ÉTS's professors involved, other ÉTS's researchers (students, research assistants, etc.), and external authors affiliated with institutions from the specified country. The system even highlights discrepancies within the dataset, such as mismatches between columns, prompting users to review and refine the results before finalizing reports. The templates are dynamic and can be adapted to specific reporting needs, making them both practical and customizable. Additionally, users can modify configuration files, to update lists of institutions or researchers dynamically. This adaptability ensures that the software remains relevant for evolving research needs.

AutoBib also facilitates the identification of faculty members within a list of coauthors. A matching algorithm is integrated to quickly identify, within a list of authors, researchers at ÉTS. With this feature, ÉTS's faculty members collaborating with authors affiliated to an institution in a country of interest can quickly be isolated from the list of all authors. This feature is currently functional for inquiries about collaborating countries only, but it is on the list of future elements to implement to make it available for other levels of entities.

AutoBib has also positively sped up the preparation of bibliometric reports. The software generates a concise graphical analysis in just a few minutes, providing a clear and quick overview of the research impact of an individual researcher. Future developments should enable AutoBib to generates the same for a group of researchers. These reports include metrics such as the total

number of publications, the field-weighted citation impact values, the proportion of papers highly cited, the proportion of publications published in the world's top journals based on the Source Normalized Impact per Paper (SNIP) indicator, and the proportion of papers by collaboration type (international, national, institutional or single author). By integrating automated data retrieval with customizable templates, AutoBib ensures that bibliographic reports meet institutional standards while remaining adaptable for specific needs. For instance, it is possible to include or exclude conference papers in the calculation of a metric, depending on the importance of this type of contribution for a researcher or in a field of research. It is also possible to modify the associated MS Excel or MS Word templates to adjust the formatting of these reports, ensuring alignment with organizational guidelines. Deletion of information is possible if insufficient data are available for the metric to be meaningful. The high level of automation significantly reduces the time and effort typically required for manual report creation, while maintaining consistency and professionalism.

The PySide6-based design incorporates intuitive prompts and dynamic validation to ensure that users can easily input API keys, specify file paths, and define analysis parameters without encountering errors. The interface guides users through each step of the process, making complex bibliometric analyses accessible even to those with no programming experience. Error-handling mechanisms provide real-time notifications for issues such as failed API requests or incorrect inputs. Helpful troubleshooting tips are displayed alongside these notifications, enabling users to quickly resolve problems and continue their workflows without significant interruptions.

2.5 Automating Collaboration Reports

Before AutoBib, the traditional method relied on the Scopus web interface to retrieve collaboration data between individuals and the NRO list for those with matching profiles. Using a saved search query, it is quick to obtain a list of publications in the database search result format. For further investigations, with a query template pasted into the advanced search module, the analyst can check if any publications are retrieved, and consult each document details page to verify if the field of research discussed in the document is linked to a sensitive research area. The main challenge with this approach isn't the number of clicks required, but rather quickly retrieving the most useful information, such as the abstract, main keywords, and the name of the research funding organization, if mentioned. The affiliation of concern can sometimes be hard to find on the document details page. The child entity of the NRO might be listed alongside the name of a coauthor, but not the name of the parent entity. For an in-depth investigation, child entities are also included in the search strategy. When using Scopus and SciVal web interfaces to collect data, separate collaboration reports are generated for each listed research organization with a Scopus profile. The collaboration report included a list of publications (with a link to the document details page in Scopus) and a list of authors, both extracted from SciVal. To consult the authors' keywords or the name of the funding entity, the analyst had to extract the data from Scopus web interface.

Now, by consulting AutoBib, it is possible to quickly extract a list of publications whose authors are affiliated with one or more entities on the NRO list, provided that a Scopus affiliation identifier has been previously collected. The tool can be queried using a Scopus Author identifier or an author's name. AutoBib also offers the flexibility to combine various entities: researchers,



Figure 4: Comparison of Traditional vs. Automated Collaboration Report Generation Workflows.

groups of researchers, institutions, groups of institutions, and country. A simple query using the name of a researcher, an institution or a country allows for the extraction of strategic data for a single entity or for the group of research organizations on the watch list within a few seconds. Among the useful fields extracted are publications titles, authors' names and affiliations, abstract, keywords, and funding agency. Future developments should enable AutoBib to be queried using groups of institutions or groups of researchers. Figure 4 provides a high-level overview the traditional workflow versus the automated process.

3 Results and Evaluation of AutoBib

3.1 Personalized Bibliometric Report

When manually creating these types of reports and querying both Scopus and SciVal user interfaces for data extraction, the workflow typically involves at least two library employees: a library technician and a librarian. The librarian provides instructions about the parameters to use, such as which type of documents to include in the calculation of a metric and which period to focus on. The technician completes the first section of the document, which contains the automated graphs, based on these instructions. Once the initial section is finished, the librarian takes over and completes the second section which is more personalized according to the researcher's profile. This section cannot be automated. It is not uncommon for the delay between the production of both sections to be too long, necessitating a revision of the first section to ensure an up-to-date portrait.

With the automated solution, the involvement of the library technician is no longer necessary, thus freeing up their time to assist with the second section of the report. The librarian uses AutoBib to

define the parameters and can generate the first section when ready to complete the second one. Updating the first section of a report now takes just a few minutes. If metrics are needed a few weeks after the bibliometric report has been delivered, the library team can quickly provide an updated version. This capability encourages users to request reports well before submission deadlines, ensuring their metrics are current.

Reproducing the exact template from the manual bibliometric report is currently not possible. SciVal's API limitations require slight modifications to the report's content. The shorter date range available from SciVal's API reduces the level of detail in the final report. However, the gain in efficiency justifies this minor loss of completeness. Only researchers who began their careers more than 10 years ago are affected by the reduced data range.

3.2 Collaboration analysis

For investigating current collaborations, the primary advantage of automation is its ability to extract a relevant data set all at once, providing a quick overview of the nature of the collaboration between two entities of interest. The report generated with AutoBib returns an Excel table with the count by document type for the set, along with a table containing the details for the following fields: EID; Abstract, Title; Author; Author ID; Author Affiliation ID; Source Title; Year; Cited By; DOI; Author keywords; Affiliation ID; Affiliation name; Country; Document Type; Funding details; Funding texts. The purpose of this table is to provide a quick overview of the topics discussed in the set of collaborative publications. This information is used to generate a word-processing summary for institution-country collaboration reports. In addition, two separate tables are generated: one with the names of authors; publication count; author IDs and affiliation name, organized by compared entities.

In comparison, the report generated with SciVal returns two separate tables with co-authors' name, number of publications, and citations. Autobib is programmed to exclude documents with multiple authors. Contributions of researchers listed on a publication with multiple authors might be negligible, and including these publications' metadata introduce noise into the analysis. In the following example, SciVal report includes a publication co-authored by 3,627 researchers, shich explains why the list of co-authors extracted from SciVal and the list extracted from AutoBib do not display the exact same set of co-authors.

For collaboration reports that include a country entity, AutoBib provides an additional Excel sheet with a table of the institutions in the country of interest. Additionally, when collaboration reports include an ÉTS entity, an extra Excel sheet is available to isolate the data for faculty members, including their full name, total number of publications and the department they belong to. In SciVal, faculty members' names are listed along with all authors affiliated with ÉTS (students, other employees, etc.), which slows down the process of compiling a shortlist of persons of interest.

Finally, for collaboration reports that include the list of NRO and a specific author, AutoBib provides an additional Excel sheet with a list of all the researcher's co-authors and a separate table with a list of all authors from the set affiliated to an entity appearing on the list of NRO. When searching for same information using the Scopus user interface, it might take a while to isolate these precise details.

3.3 Time Efficiency: Automated vs. Manual Reporting

To measure the time savings achieved by using an automated solution instead of the traditional approach of extracting data and generating reports, two typical bibliometric report requests are completed using both methods. The tasks are performed by two individuals, and the average task time is calculated. On average, using the automated solution is found to be 12.6 times faster than completing section one of the reports manually, resulting in a time savings of 92%.

Typical collaboration report requests are also completed using both the manual method and the automated solution. For collaboration between ÉTS and a specific country during the 2018–2024 period, using AutoBib is 3.4 times faster than extracting the data manually through SciVal user interface, resulting in a time saving of 71%. For this request, only the MS Excel file is generated. A second test is performed, generating both MS Excel and MS Word reports to obtain the collaboration portrait between ÉTS and a specific country for the same period. Using the automated solution is 4.2 times faster than extracting the data through the SciVal user interface, resulting in a time savings of 76%.

A third test is conducted to measure the time required to retrieve information on a specific researcher's collaboration with entities listed on the NRO list for the 2019–2024 period. Using the automated solution is 3.7 times faster than extracting the data using SciVal, resulting in a time savings of 73%. Data extracted with the automated solution are easier to analyze, as relevant information is clearly presented and unnecessary data is omitted. A summary of the evaluation is presented in the following table:

	Manual	AutoBib	Time
Task	Method (min)	(min)	Savings (%)
Personalized Bibliometric report	38	3	92%
Collaboration Report (Excel only)	55	16	71%
Collaboration Report (Excel + Word)	88	21	76%
NRO Collaboration Analysis	11	3	73%

Table 2:	Comparison	of Manual	and Automated	Reporting	Times
	1			1 0	

4 Dsicussion and Further Development

The current version of AutoBib is currently limited to data extraction from Scopus, relying solely on its citation index. This makes it suitable for researchers or entities with a representative sample of their work indexed in Scopus but less effective for broader or interdisciplinary evaluations. Future developments could integrate complementary databases, such as PubMed or patent repositories, to provide a more comprehensive analysis. Tracking educational resources and nontraditional output could also enhance its applicability for engineering schools. Furthermore, leveraging artificial intelligence, as demonstrated by the PyBibX library, could broaden the scope of future bibliometric analyses¹⁸.

Existing commercial tools such as SciVal and InCites offer automated bibliometric reporting, but are typically tailored for English-speaking institutions. For organizations where another language

is predominant, such as ÉTS, these tools often require extensive translation efforts. AutoBib addresses this gap by offering a French interface and generating reports directly in the required language, improving efficiency and usability.

Relying on interns for the development of the automated solution has been, and is still, the main concern of the library regarding this project. During the second development phase, time constraints resulted in insufficient testing toward the end of the internship, and timely follow-ups are sometimes challenging due to the internship being held during the summer. Random errors have since been observed, particularly in larger country-level collaboration reports, reinforcing the need for a dedicated developer to ensure stability and promptly resolve bugs.

A lack of comprehensive API documentation from data source providers is another obstacle faced during the project programming phase. While commercial bibliographic databases offer cleaner and more exhaustive publications metadata, the lack of transparency and documentation on how to leverage their APIs can significantly slow down the development process. This is particularly challenging for a short-term internship project.

Despite these challenges, AutoBib has proven highly effective for most reporting tasks, significantly reducing the workload and enabling timely updates. Ensuring sustainable development and addressing identified limitations will be critical for its continued success and wider adoption.

Conclusion

AutoBib is a modular tool that streamlines bibliometric report creation with Python-based data extraction, a Finite State Machine, and VBA macros. It generates concise, high-quality reports in minutes, surpassing tools like SciVal through customizable templates, French-language output, and swift author identification. Initially, expectations were low, but staff no longer wish to revert to manual methods. The main challenge is sustaining development, ideally through regular maintenance and ongoing internships. Ultimately, AutoBib lowers manual effort while supporting timely, data-driven decisions with its efficient, user-friendly approach.

References

- [1] H. F. Moed, *Citation analysis in research evaluation*. Dordrecht, Netherland: Springer Dordrecht, 2005, vol. 9.
- [2] C. R. Sugimoto and V. Larivière, *Measuring research: What everyone needs to know*. New York, USA: Oxford University Press, 2018.
- [3] L. Bornmann and H.-D. Daniel, "What do citation counts measure? A review of studies on citing behavior," *Journal of Documentation*, vol. 64, no. 1, pp. 45–80, 2008.
- [4] W. Glanzel, "Bibliometrics as a research field : A course on theory and application of bibliometric indicators," https://www.researchgate.net/publication/ 242406991_Bibliometrics_as_a_research_field_A_course_on_theory_and_application_of_bibliometric_indicators, [Magyar Tudományos Akadémia, Kutatásszervezési Intézet], 2003, [PDF document].

- [5] M. Franceschet, "Frozen footprints," arXiv:0811.4603, 2009.
- [6] M. V. Simkin and V. P. Roychowdhury, "A mathematical theory of citing," *Journal of the American Society for Information Science and Technology*, vol. 58, no. 11, pp. 1661–1673, 2007, doi:10.1002/asi.20653.
- [7] D. Du, J. Ding, Y. Chen, and C. Liu, "The evolution and development landscape of scholarly communication based on the analysis of published articles," *Journal of Librarianship and Information Science*, vol. 57, no. 1, pp. 40–56, 2025, doi:10.1177/09610006231196336.
- [8] D. Hicks and J. Melkers, "Bibliometrics as a tool for research evaluation," in *Handbook on the Theory and Practice of Program Evaluation*. Edward Elgar Publishing Ltd., 2013, pp. 323–349.
- [9] A.-W. Harzing and S. Alakangas, "Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison," *Scientometrics*, vol. 106, pp. 787–804, 2016, doi:10.1007/s11192-015-1798-9.
- [10] W. W. Hood and C. S. Wilson, "The literature of bibliometrics, scientometrics, and informetrics," *Scientometrics*, vol. 52, no. 2, pp. 291–314, 2001, doi:10.1023/A:1017919924342.
- [11] A. F. van Raan, "Advances in bibliometric analysis: Research performance assessment and science mapping," in *Bibliometrics: Use and Abuse in the Review of Research Performance*. London, United Kingdom: Portland Press Limited, 2014, pp. 17–28.
- [12] M. E. Rose and J. R. Kitchin, "Pybliometrics: Scriptable bibliometrics using a Python interface to Scopus," *SoftwareX*, vol. 10, pp. 1–6, 2019, doi:10.1016/j.softx.2019.100263.
- [13] Council of Canadian Academies, "Informing research choices: Indicators and judgment," pp. 1–142, 2012.
 [Online]. Available: https://www.cca-reports.ca/wp-content/uploads/2018/10/Informing-Research-Choices-Full-Report-EN.pdf
- [14] Y. Gingras, *Les dérives de l'évaluation de la recherche: du bon usage de la bibliométrie*. Paris, France: Éditions Raisons d'agir, 2014.
- [15] M. Y. Vardi, "Conferences vs. journals in computing research." *Communications of the ACM*, vol. 52, no. 5, p. 5, 2009, doi:10.1145/1506409.1506410.
- [16] Minister of Innovation, Science and Economic Development Canada, "Policy on Sensitive Technology Research and Affiliations of Concern." [Online]. Available: https://science.gc.ca/site/science/en/safeguarding-your-research/ guidelines-and-tools-implement-research-security/sensitive-technology-research-and-affiliations-concern/ policy-sensitive-technology-research-and-affiliations-concern
- [17] Government of Canada, "Named research organizations," January 2024, accessed April 24, 2025, Cat. No. Iu37-45/2023E-PDF, ISBN 978-0-660-67945-7. [Online]. Available: https: //science.gc.ca/site/science/en/safeguarding-your-research/guidelines-and-tools-implement-research-security/ sensitive-technology-research-and-affiliations-concern/named-research-organizations
- [18] V. Pereira, M. P. Basilio, and C. H. T. Santos, "PyBibX a Python library for bibliometric and scientometric analysis powered with artificial intelligence tools," *Data Technologies and Applications*, vol. 59, no. 2, pp. 302–337, 2025, doi:10.1108/DTA-08-2023-0461.