From Open-data to Open-source Reporting: Introducing the BiSO

Romain Thomas, Henri Bretel, Delphine Le Piolet, Laïli Rahimie

FROM OPEN-DATA TO OPEN-SOURCE REPORTING: INTRODUCING THE BISO

Romain THOMAS
DiBISO - Department of Libraries,
Information and Open Science
Université Paris-Saclay
91400 Orsay
France
contact@romainthomas.net

Henri BRETEL
DiBISO - Department of Libraries,
Information and Open Science
Université Paris-Saclay
91400 Orsay
France
henri.bretel@universite-paris-saclay.fr

Delphine LE PIOLET
DiBISO - Department of Libraries,
Information and Open Science
Université Paris-Saclay
91400 Orsay
France

Laïli RAHIMIE
DiBISO - Department of Libraries,
Information and Open Science
Université Paris-Saclay
91400 Orsay
France

October 29, 2025

ABSTRACT

In this paper, we present the Biso, an automated report created to help Université Paris-Saclay research units with monitoring Open Science policy compliance in their output. It includes, first the main goals of the project, then the guidelines followed during software development, finally a brief overview of the data sources and some of the metrics and visualizations. Since the project's aim is to provide a flexible infrastructure for automated reporting from open sources, including but not limited to the BiSO, the paper explains how the technical infrastructure can be used in the future for other report types.

***K*eywords** Bibliometry · Reporting · HAL · OpenAlex · ScanR

1 Introduction

1.1 Institutional context

Université Paris-Saclay is one of the largest research communities in France. With about 230 research units, totalling more than 8000 researchers and 13000 scientific publications per year, Université Paris-Saclay wanted to provide a formal support service to help the community with Open Science. This service, named the Research and Open Science Referents Network (3RSO), was created in 2020. It takes place in the Department of Libraries, Information and Open Science (DiBISO). It is embodied by library staff members, named "referents", who are involved in disseminating Open Science best practices among the research units.

1.2 Operational context

In 2024, library teams prepared lists of verified publications for each research unit, comparing these lists with existing HAL data in order to help researchers to prepared the quinquennial national research evaluation program, which took place shortly after. Support teams provided research units with added value services, including identifying publications,

verifying metadata and helping with data compliance. Researchers who benefited from those services expressed their wish to be accompanied on a more regular basis.

Since the service received positive feedback, we created an automated reporting tool which would help with the University Open Science policy without monopolizing too much of the library team's working time. We wanted to:

- Promote the Open Archive culture among all the research units
- Foster rich, well attuned, curated metadata to enhance research work visibility
- Strengthen connections between the library and the research staff

The BiSO and all associated tools are the result of us thinking all of these aims directly at a large scale: to complement already existing scripts that we used to help with works identification and deduplication [1], we created a flexible backend application to automatically produce any type of reports, of which the BiSO is the first.

Building up from this first use case of the modular framework, we created a prototype for a report type on publications and partnerships. This report type, named PubPart follows the double objective of analyzing publications for potential partnerships.

2 Method

First, we established a list of requirements:

- the code and code dependencies must be open-source
- the data must be open
- code must be as reusable as possible for future other types of reports

This led us to choose a modular approach. We then identified the key components needed to generate reports and how they had to interact with each other. Figure 1 describes the architecture of the project. We first focused on the main three elements: the plotting library [1], the LaTeX template [2] and the reporting library [3], which is used to join the plots and template in order to create a LaTeX project. The API and web application were later developed to allow non-technical personals to easily generate reports.

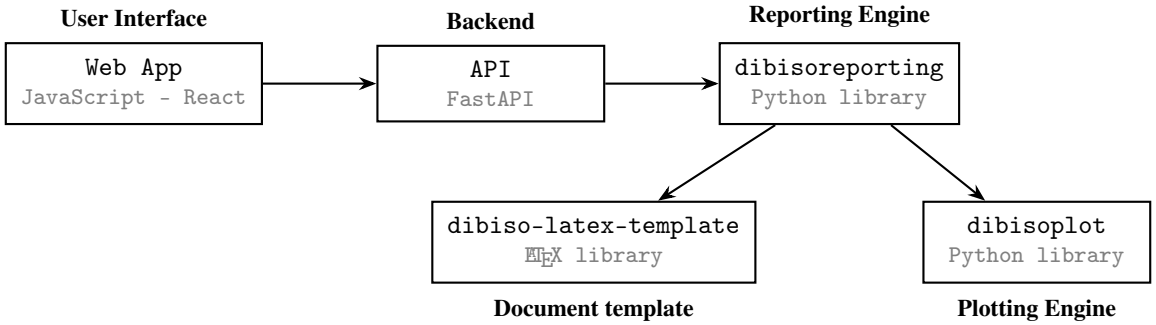

Figure 1: Dependencies graph

Each of the three main elements is structured with submodules: at the time of writing, there are submodules for BiSO and PubPart, and we plan to create other types of reports, i.e., new submodules. The submodule structure diagram is located in figure 2.

Plots are generated by the Python library `dibisoplot`: this library fetches the data from various API and returns a figure either as a Plotly figure object for plots or as a text string for tables. The LaTeX template contains a folder with the code and assets (logo and icons) of the template, which is used with a LaTeX `tex` file importing the template and the figures and tables. This "`main.tex`" file is the one we compile to generate reports. Another LaTeX `tex` file is present to compile the bibliography (the list of publications underlying each report). The `dibisoreporting` Python library generates the plots and table with `dibisoplot` and save them as `pdf` and `tex` files respectively. The output of `dibisoreporting` is a folder which contains everything required to compile the report: `pdf` files for the figures, `tex` files for the tables, LaTeX template and `main.tex` and `biblio.tex` file to compile. The report generation process is shown in figure 3.

[1] `https://github.com/hbretel/simple-hcc`

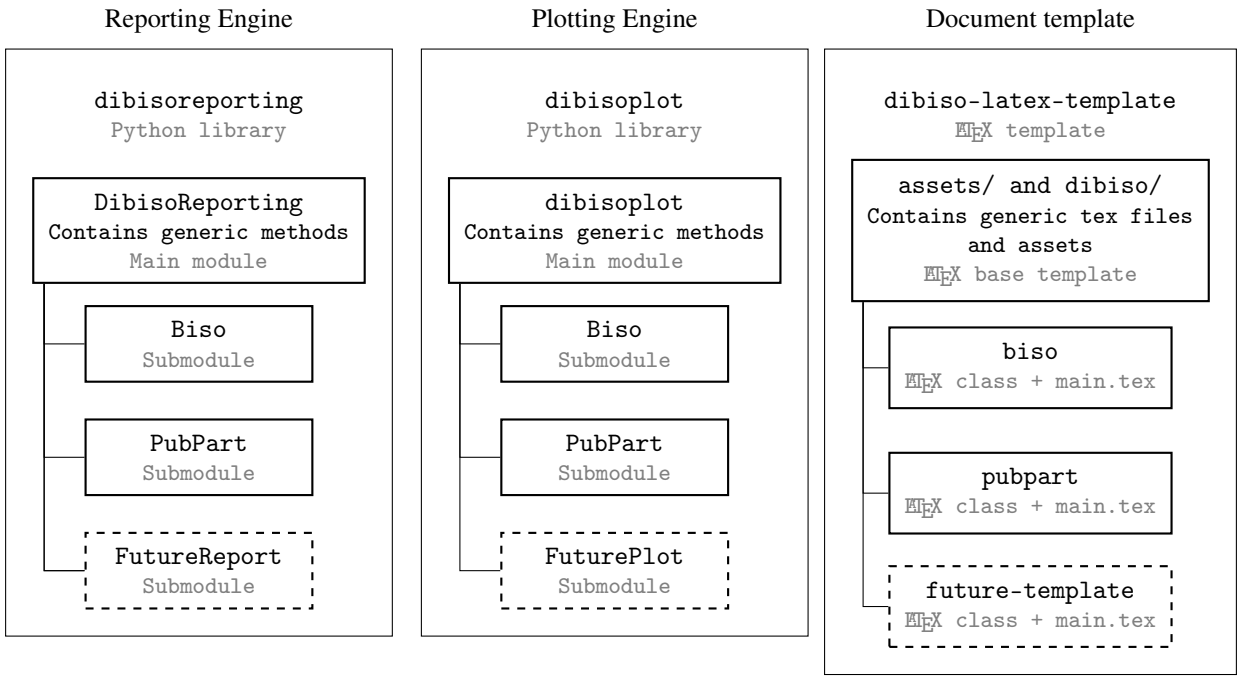

Figure 2: Structure of the submodules of `dibisoreporting`, `dibisoplot`, and `dibiso-latex-template` [xshift=1cm]

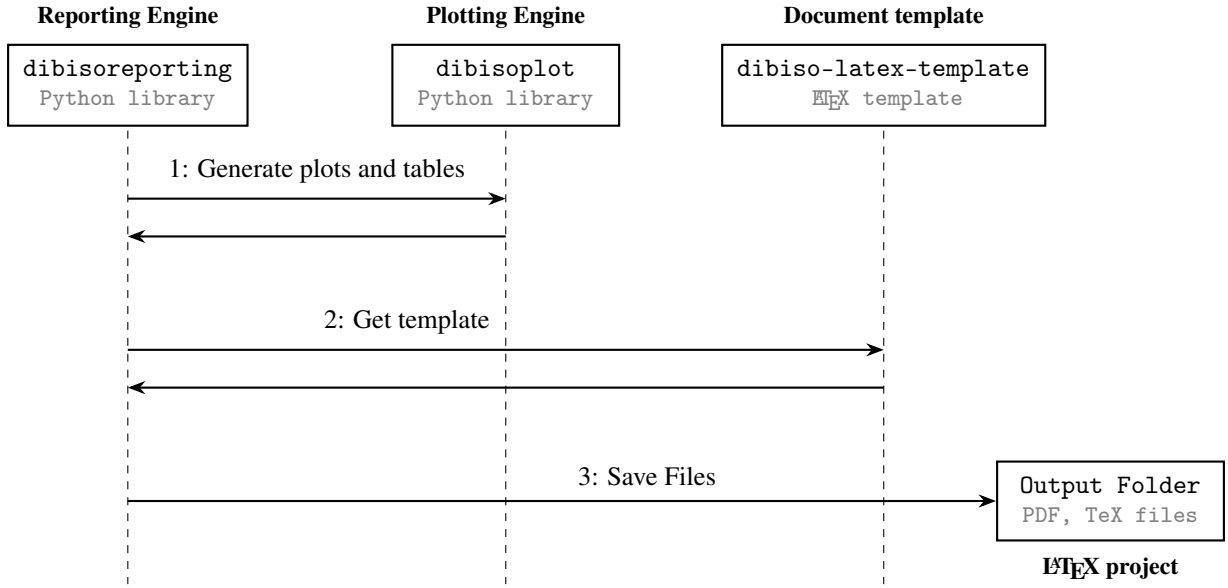

Figure 3: Sequential diagram of the report generation process

3 The BiSO

The BiSO is a report produced annually for each research unit supported by the Open Science teams of the Université Paris-Saclay libraries. Prepared in collaboration with the units, it is based on open data, mainly coming from the HAL repository [4] but also from OpenAlex[5], scanR and the French Open Science Monitor. The BiSO presents indicators such as publication types, open access rates, Article Processing Charges and collaborations. Primarily intended for Université Paris-Saclay units, the report aims to support the development of open science practices.

The figure 4 is an example of BiSO report for Université Paris-Saclay. This report comes from the documentation example and contains partial data as the number of works in the collection exceed the configurable maximum number of entities to fetch from the API.

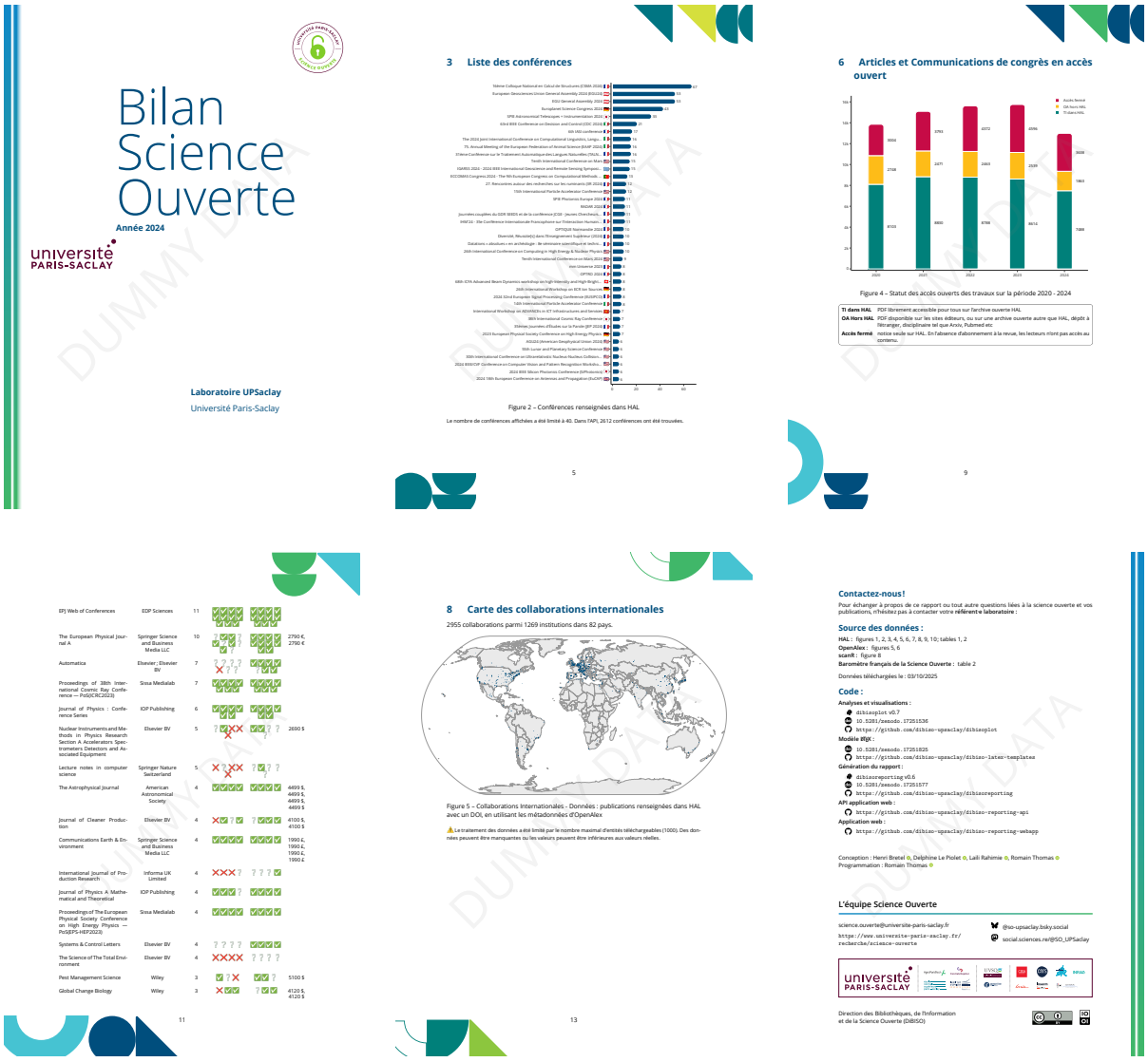

Figure 4: Example of a BiSO report

3.1 Data sources

This section only applies to the first type of report we implemented, which is name BiSO (*Bilan Science Ouverte*), which translates literally to "Open Science checkup". However, open data sources provide enough metadata for other types of reports. To generate the BiSO reports, we currently use the following 4 sources:

HAL An open archive that provides an API with metadata for millions of publications, supported by French universities and research organizations. HAL contains metadata about articles, conferences, posters and all kind of output types produced by research activities. Each work is identified by an internal HAL ID, but external IDs such as DOI for works and ORCID for people are often provided.[2]

[2] `https://hal.science/`

OpenAlex A free, comprehensive repository that provides an API with metadata for hundreds of millions of publications, supported by the nonprofit OurResearch. It aggregates articles, preprints, books, datasets, and more from 250,000+ sources. Each work is linked to disambiguated authors, institutions, topics, SDGs, and citation data. DOI, ORCID and ROR are provided for the disambiguated entities, allowing to link data between OpenAlex and other databases. OpenAlex offers full exports, API access, and even a complete dataset download—all under a CC0 license for unrestricted reuse.[3]

scanR A comprehensive open-data platform dedicated to mapping France's research and innovation ecosystem. Developed under the Ministry of Higher Education and Research's open science strategy, it aggregates and interconnects five core datasets: research structures, publications, authors, funding, and patents. Powered by Elasticsearch 8, its API is available upon request, while its full JSON dataset is publicly available to download (under the Etalab 2.0 license).[4]

French Open Science Monitor A national observatory launched by the French Ministry of Higher Education and Research to monitor and evaluate the progress of open science in France. A yearly JSON dataset is publicly available to download and an ElasticSearch index is hosted by the scanR development team.[5]

3.2 Description of some metrics and visualizations

Each metric uses the works in a HAL collection as a starting point: each research laboratory has a HAL collection which contains all works produced by their researcher.

Conferences list: Shows the top conferences in which the unit's researchers published. Queries the name of the conferences, the count and the country from HAL, and convert the country string to an emoji to include it after the conference title.

Journals table: Uses the DOI and HAL ids from the HAL collection to query the works from the French Open Science Monitor. Extracts the journal, publisher names, open access status in the journal (i.e. whether works are *Gold Open Access*) and in a repository (i.e. if works are *Green Open Access*).

International collaborations map: Uses DOI from the HAL collection to query the works from OpenAlex and get collaborating institutions' OpenAlex ids and co-authored works counts. Also uses the countries and GPS coordinates metadata from OpenAlex institutions.

References

[1] R. Thomas. *dibisoplot*. Version v0.7. Oct. 2025. DOI: 10.5281/zenodo.17251537. URL: https://github.com/dibiso-upsaclay/dibisoplot.

[2] R. Thomas, H. Bretel, D. Le Piolet, and L. Rahimie. *DiBISO LaTeX templates*. Version v0.8. Oct. 2025. DOI: 10.5281/zenodo.17276757. URL: https://github.com/dibiso-upsaclay/dibiso-latex-templates.

[3] R. Thomas. *dibisoreporting*. Version v0.7. Oct. 2025. DOI: 10.5281/zenodo.17258060. URL: https://github.com/dibiso-upsaclay/dibisoreporting.

[4] CCSD. *HAL*. 2025. URL: https://hal.science/.

[5] J. Priem, H. Piwowar, and R. Orr. *OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts*. 2022. DOI: 10.48550/arXiv.2205.01833. arXiv: 2205.01833 [cs.DL]. URL: https://arxiv.org/abs/2205.01833.

[3] https://openalex.org/

[4] https://scanr.enseignementsup-recherche.gouv.fr/

[5] https://barometredelascienceouverte.esr.gouv.fr/