Enhancing bibliometrics with cris-systems

Lars Wenaas[1] and Jens Aasheim[1]
[1] Norwegian Agency for Shared Services in Education and Research (Sikt), Norway
lars.wenaas@sikt.no, jens.aasheim@sikt.no

Abstract

A national cris alone can be used for many bibliometric analyses but comes with natural limitations, especially citation-based analyses. However, a combination of a national cris and a bibliometric database like OpenAlex can utilize the best of both worlds; the accuracy on affiliation data in a cris and citation-data from a bibliometric source. This has been piloted in Norway, and an integration between the two systems show a big potential for wider application and more accurate results, all in the spirit of open science and the Barcelona-declaration.

1 Introduction

A cris-system serves many useful purposes and functions as an inventory of scientific production. Cris-system can also be important building blocks in open science-structures in alignment with the Barcelona-declaration on Open Research Information.

While cris-system are most often implemented on the institutional level, Norway has a national cris-system. The Norwegian Research Information Repository (NVA) keeps track of all scientific publications published by Norwegian publicly financed institutions from 2011 to date. Data from the Norwegian cris serves as a foundation for many useful tasks besides keeping track of what Norwegian institutions publish, particularly gauging the level of open access over time. NVA-data is also underpinning publish-and-read-agreements, where negotiations with publishers are supported and influenced by data on which journals Norwegian researchers prefer to read and publish in. Cris-systems are suitable for supporting a wide variety of task by default; however, many analyses also require additional data to produce insights. For example, tracking open access and supporting publisher negotiations also relies on Unpaywall data as well as data from current publisher agreements and COUNTER statistics.

Cris-systems neither supports traditional bibliometric analysis involving citations out of the box. To accommodate bibliometric analysis involving citations, cris-data must be enriched with bibliometric data. This has been piloted in Norway, where Sikt – Norwegian Agency for Shared Services in Education and Research, maintains responsibility for both NVA and the National Infrastructure for bibliometrics (NIB). The pilot shows promising results drawing on the strengths of both systems.

The purpose of this paper is to outline how data from a national cris and data from traditional bibliometric sources can be joined to accommodate enhanced bibliometric analysis, and what benefits and obstacles lies ahead.

2 Background

There has always been a close relationship between the Norwegian national cris-system and open science. The newly launched NVA has recently replaced the former system CRIStin (Current Information System In Norway), where one of the predecessor's functions was to accommodate green open access, by letting the users use the system as an 'upload-hub for fulltext and at the same time populate institutional repositories with metadata from the cris [1]. The new system is strengthening the bonds to open science even more by merging the former functions of CRIStin with repository-functionalities in a single system covering 160 research institutions and approximately 80 institutional repositories in a consolidated system of data. The overarching principles of open science has thus been guiding the national cris-system from the beginning.

Open science, with its many branches, is now also covering research information through the Barcelona-declaration on Open Research Information [2]. The declaration states that research evaluations and policy decisions should be made based on transparent evidence and to be accessible by those being assessed. While the declaration, also signed by euroCRIS, covers cris-systems, its main aim is to address the use of proprietary systems of bibliometric databases as for example Scopus and Web of Science (WoS). One system that aligns with the principles of the Barcelona-declaration is OpenAlex, a promising source of bibliometric data, based on the now deprecated Microsoft Academic Graph, which is maintained by OurResearch and free to download and use for everyone [3].

OpenAlex has recently been included in the portfolio of NIB, which is a Norwegian infrastructure launched in 2019 to offer bibliometric data to Norwegian research institutions. Until 2025, NIB was limited to WoS-data only, which was offered to members of NVA exclusively, even then with limitations. However, with recent developments in bibliometric sources, OpenAlex has been added as an expansion. While OpenAlex technically is currently only available in NIB with status as a test version, it has become an important asset as a strategic step in branding NIB as an open science-infrastructure.

3 The potential in a partnership between NVA and NIB

Research evaluations have been greatly influenced by bibliometrics with easily comparable (although not always understandable) numeral metrics in rankings and comparisons. Particularly analyses with emphasis on citations, the main currency in a researcher's career, are heavily debated, and the subject of the Leiden-manifesto, which issues strong warnings on the role of bibliometrics in research evaluations [4]. Even so, bibliometrics has acquired an important and popular role in different kinds of research evaluations, which is why NIB is facilitating access to data. While bibliometrics is a research discipline in its own right, the average user of NIB is typically performing tasks in the administrative leg of the institution, using analysis in internal affairs on site.

NIB is offering both WoS and OpenAlex as a part of a multi-source-strategy, where the same analysis in principle can be performed based on different sources, and thus give two different perspectives on the same phenomena. Every analysis is influenced by the source of data, and in any case, neither source is able to represent a complete and accurate picture. For example, WoS is indexing a limited number of journals, and is further lacking non-English journals and particularly research within the social sciences and humanities [5]. While being a promising addition to

proprietary and expensive sources, OpenAlex is known for misclassified document types, which is important as letters to editor, book reviews and other material may be classified as regular research articles and thus skew analyses [6]. Both sources are also known for issues with affiliations and addresses, although the curated WoS is usually considered stronger than OpenAlex. Even with limitations, both sources support a wide range of bibliometric analysis, and access to comparative analysis using multiple sources may help improve the quality and address some of the shortcomings of relying only on a single source.

Cris-systems, especially national ones like NVA can also function as a data-source for a wide range of bibliometric analysis, particularly analyses of production. NVA, as cris-systems in general, usually do not keep track of citations, making them less suitable for these purposes. On the other hand, NVA is particularly strong on affiliations and institutional addresses. This is due to NVAs task as an institutional tool for reporting production levels to the ministry of Education and Research. This reporting, which is mandatory for all publicly financed research institutions in Norway, creates an incentive for the institutions to maintain accuracy of correct affiliation data for their publications.

Some of the limitations in either NVA or NIB can be remedied by joining data from both systems, particularly utilizing high-quality affiliations in NVA, making it possible to know the exact baseline number of records that belong in an institutional publication portfolio. A bibliometric analysis based on WoS or OpenAlex can then be validated against a verified number of records that optimally should be part of the analysis.

4 Piloting the potential

While NVA naturally comes with a traditional web interface both for viewing and managing publications, NIB currently only offers rudimentary access through SQL to perform queries and fetch datasets. This, together with complicated data models and a lack of off-the-shell analytical tools translates to a significant hurdle for usage. Even so, the main challenge for the average user is perhaps not on the technical side; bibliometrics is complicated and requires a lot of know-how and methodological knowledge. To help lower the bar, NIB pre-calculates several relevant indicators, including Source-Normalized Impact per Paper (SNIP) and percentiles, which are recalculated on a regular basis when data is updated. However, the whole process remains a bumpy road for users lacking extensive technical skills.

This led Sikt to develop an analytics pilot in the fall of 2025, with the aim of investigating the possibilities of simplifying the process from data to analysis. An important idea was to make use of NVA's enhanced affiliation data to improve accuracy of datasets for individual institutions. The scope of the pilot thus goes beyond merely combining datasets, but to relieve the whole process of analysis. To give a brief overview of the pilot, the procedures are implemented in R and can in principle be executed without prior knowledge. Mandatory steps are setting a few parameters to authenticate and set scope to limit the datasets according to the user's intent, for example by setting a time-period and what institution should be the object of analysis. Such a pilot can be built in a variety of ways but would in all implementations have to deal with some important initial steps. These cover 1) matching NVA-data with NIB-data, 2) simplifying the data model with the most relevant data from both sources and 3) merging the datasets and prepare for analysis.

The key aspect is to synchronize the datasets, matching as many records as possible between the systems. This is a routine performed based on DOI-matching every time the database is updated, a procedure that includes data-enhancement with regex-expression of DOI's in both NVA and the two databases. In case a DOI is missing in NVA, efforts are made to identify existing DOI's by querying Crossref. The matching-procedure is rather successful, with a score of approximately 94 % for recent

core-publications (defined as journal articles reported to the ministry the previous three years) in OpenAlex, and similarly 84 % in WoS based on DOI matching alone. A huge benefit from a bibliometric point of view, is that a user will have information on how many publications are matched and thus how many are excluded from a specific analysis. This has a big impact on the validity of analyses, especially citation-based.

The pilot currently offers nine demo-analytics that can be executed on a user-defined dataset. These are ranging from simple inventories of production to the mean number of citations within each open access-flavour and can be explored with data from either WoS or OpenAlex. However, the main feature of the pilot is that the user can define a dataset limited to a specific institution and/or time-period, extract the underlying data and analyse the data in whatever tool the user prefer.

5 Discussion and conclusion

The pilot was made available for Norwegian institutions early in 2026 and was met with both curiosity, encouragement and interest. While it is too soon to pass an evaluation of the pilot at this point, it seems clear from a theoretical perspective alone, that bibliometrics and research evaluation can benefit from using structured cris-data to gain more accuracy in analysis.

References

[1] Wenaas, L., Karlstrøm, N., & Vatnan, T. (2012, June 9). From a national CRIS along the road to Green Open Access – and back again: Building infrastructure from CRIStin to Institutional Repositories in Norway. https://dspacecris.eurocris.org/handle/11366/116
[2] Barcelona Declaration on Open Research Information. (n.d.). Barcelona Declaration on Open Research Information. Retrieved 17 February 2026, from https://barcelona-declaration.org/
[3] OpenAlex: The open catalog to the global research system | OpenAlex. (n.d.). Retrieved 17 February 2026, from https://openalex.org/
[4] Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Ràfols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. Nature News, 520(7548), 429. https://doi.org/10.1038/520429a
[5] Chavarro, D., Ràfols, I., & Tang, P. (2018). To what extent is inclusion in the Web of Science an indicator of journal 'quality'? Research Evaluation, 27(2), 106–118. https://doi.org/10.1093/reseval/rvy001
[6] Haupka, N., Culbert, J. H., Schniedermann, A., Jahn, N., & Mayr, P. (2024). Analysis of the Publication and Document Types in OpenAlex, Web of Science, Scopus, Pubmed and Semantic Scholar (arXiv:2406.15154; Version 1). arXiv. https://doi.org/10.48550/arXiv.2406.15154