

National Bibliographic Database for Research Outputs in the Republic of Moldova: Overview and Review

Irina Cojocaru

Information Society Development Institute
E-mail irisha.cojocaru@gmail.com

Gheorghe Cuciureanu

Information Society Development Institute
E-mail gheorghe.cuciureanu@idsi.md

Igor Cojocaru

Information Society Development Institute
igor.cojocaru@idsi.md

Lately, an increasing number of countries have set up national bibliographic databases, which aim to achieve more comprehensive coverage of the national research outputs. These databases both increase the visibility of national publications and compensate the limited coverage of traditional citation databases, such as Scopus or Web of Science. National bibliographic databases are being increasingly used by a wide variety of stakeholders for numerous purposes, ranging from basic information retrieval and research outputs overview to bibliometric analysis, evaluation and research funding.

The Republic of Moldova set up its national bibliographic database – National Bibliometric Instrument (IBN) in 2011. Nowadays, IBN includes over 113.000 articles published in national scientific journals, as well as conference proceedings from the Republic of Moldova starting from 1993 to date, with over 90% of publications indexed by Google Academic. IBN is used for the evaluation of national scientific journals and provides valuable insights on the overall evolution and trends of national research outputs, but is not being used for research funding.

This paper aims to describe the history, actual structure and user possibilities of IBN. It also aims to analyse IBN design, organisation, maintenance and usage from the perspective of the Manual of good practices regarding national bibliographic databases, developed in the context of the 'European Network for Research Evaluation in the Social Sciences and Humanities' ENRESSH. Finally, the paper intends to provide a set of recommendations for IBN reengineering and improvement.

Keywords: national bibliographic database, research output, publications, National Bibliometric Instrument, design, maintenance, use

1. Introduction

Research is not one-dimensional: it is a complex process and each project has its own unique features. Research organizations also vary greatly: they teach as well as research; their research may be analytical, applied, collaborative, societal or industrial; and their activity is spread across many disciplines, each with its own academic characteristics (Adams et al. 2019). Therefore, the task of capturing the entire palette of research outputs is quite challenging.

The most common type of research outputs – scientific publications, are generally captured by bibliographic databases and further on used to derive various metrics. These metrics are often derived as a result of bibliometric explorations and aim to satisfy the expectations of different stakeholders, ranging from governments and research funding organisations to research performing organisations, business, civil society, researchers themselves and eventually, the public at large (Wilsdon et al. 2015). The aims of these metrics and consequent evaluation vary from research evaluation and performance-based research funding to policymaking, quality assurance, promotion and increased visibility, benchmarking, as well as information retrieval by any interested parties.

Over the past 15 years, Web of Science, Scopus and Google Scholar have been the three most important multidisciplinary bibliographic data sources, providing metadata on scientific publications and on citation links between these. They have been recently supplemented by Microsoft Academic and Dimensions. But all of the above databases share a common drawback: their insufficient coverage of the scientific literature across subject disciplines and in languages other than English, especially in terms of social sciences and humanities, where books and national journals are rarely indexed by bibliographic databases, or computer science and engineering sciences, where most publications tend to appear in conference proceedings.

Lately, an increasing number of countries have set up national bibliographic databases, which aim to achieve more comprehensive coverage of the national research outputs. These databases both increase the visibility of national publications and compensate the limited coverage of traditional bibliographic and citation databases. National bibliographic databases are being increasingly used by a wide variety of stakeholders for numerous purposes, ranging from basic information retrieval and research outputs overview to bibliometric analysis, evaluation and research funding.

The Republic of Moldova set up its national bibliographic database – National Bibliometric Instrument (IBN, Instrument Bibliometric National) in 2011. Nowadays, IBN includes over 117.000 articles published in national scientific journals, as well as conference proceedings from the Republic of Moldova from 1993 to date, with over 90% of publications indexed by Google Academic.

This paper aims to present an overview of the National Bibliometric Instrument, its design, organization and applications, as well as review is based on the Manual of good practices developed in the framework of the ENRESSH project.

2. Literature review

The various uses, limitations and issues of bibliographic databases, both multidisciplinary and specialised, national or regional, citational or providing access to full-text have been widely explored over time. Some national bibliographic databases have undergone really impressive evolutions. Russian Index of Science Citation launched in 2005, became the basis for the Russian Science Citation Index launched on the Web of Science platform in 2016 (Moskaleva et al. 2017). The Chinese Science Citation Database created in 1989, was the first non-English database available within Web of Science, managed by Clarivate Analytics in partnership with the Chinese Academy of Sciences (Clarivate 2021). China's own citation database, which reflects the characteristics of domestic activities, was developed in order to improve the performance of retrieving citations, searching information and evaluating science and technology in China (Jin & Wang 1999). According to Packer (2000), SciELO (Scientific Electronic Library Online) is a bibliographic database, digital library, and cooperative electronic publishing model of open access journals initially created to meet the scientific communication needs of developing countries (Latin America, Spain, Portugal, the Caribbean and South Africa) and provides an efficient way to increase visibility and access to scientific literature. Nowadays, the SciELO Citation Index is part of the Web of Science platform.

A study carried out in 2016 identified 21 national bibliographic databases in Europe (Sile et al. 2017, 2018) which are typically used to report on, monitor, and evaluate research activities within one national context. Consequently, VIRTAs Publication Information Service, launched in Finland in 2016 as an advanced decentralized solution to integrate institutional data at the national level, has been analysed in terms of its potential for extension and use at the European level (Puuska et al. 2018).

"Bibliographic database" or "database for research output" can be defined in several ways, depending on the level of granularity. The Dictionary for library and information science defines the bibliographic database as a computer file consisting of electronic entries called records, each containing a uniform description of a specific document or bibliographic item, usually retrievable by author, title, subject heading or keywords (Reitz 2004). According to the International Encyclopedia of Information and Library Science, a bibliographic database is a database of bibliographic records, an organized digital collection of references to published literature, including journal and newspaper articles, conference proceedings, reports, government and legal publications, patents, books, etc. (Feather & Sturges 2003). For the purpose of this paper, we'll adopt a broader definition of bibliographic database (Sile 2018) as a structured set of bibliographic metadata (e.g. title, publication type, year, author) needed for the calculation of the most basic indicator of research output, namely the number of publications.

3. Methodology

National Bibliometric Instrument IBN (www.ibn.idsi.md) is one of the key components of the national research e-infrastructure. Since its set-up in 2011 it was envisaged as an open, flexible and scalable system, aiming to satisfy various user needs (Cojocaru et al. 2010). At the moment, IBN is the biggest Open Access electronic library of articles published in national scientific journals, as well as conference proceedings from the Republic of Moldova starting from 1993 to date. IBN is the final outcome of the project "Development of the National Bibliometric Instrument", carried out during 2010-2011 by the Information Society Development Institute - ISDI and is being continuously upgraded and updated. IBN is not a citation database, it is basically a full-text library and it has been funded from governmental sources.

In terms of comprehensiveness, IBN includes data only on articles in national scientific journals and conference proceedings, excluding monographs, book chapters, doctoral thesis or other types of research output. In January 2021, IBN includes data on 121 scientific journals, 55 of them accredited in compliance with the Regulation for evaluation, classification and monitoring of scientific journals, approved in 2018. The classification of scientific journals started in 2008, with the adoption of the first regulation. IBN stores over 117.000 publications, out of which over 41.000 are publications in conference proceedings, in compliance with the Regulation on the recognition, classification and recording of scientific events, approved in 2018 and over 2.700 are publications of Moldovan authors in international journals and proceedings. Since its set up in 2011, the system has been accessed by over 650.000 visitors, cumulating over 35 mln views and over 572.000 downloads. During COVID pandemic, the number of database visitors has doubled from 146.558 in 2019 to 298.820 in 2020, taking into account that the national scientific community accounts roughly 3.000 researchers.

The functionalities provided by IBN include information search and retrieval, management of records, such as journals, proceedings and publications, export of publication metadata in various formats (CERIF, BibTex, DublinCore, DataCite, CrossRef), analytical features and a big number of various metrics at publication, journal, proceedings, author and organisation levels, application and evaluation forms for journals accreditation.

IBN uses are derived from the functionalities it offers and includes evaluation and funding allocation purposes, namely journal accreditation and ranking; facilitating access to information by linking data to full texts, storing full texts, assisting in information search and retrieval, as well as promoting research and its visibility; research and monitoring, as IBN data are used in bibliometric research and publication activity monitoring. IBN is interoperable with EXPERT online, the national proposal submission and evaluation system (data on researchers and projects), as well as the database of the national State Agency on Intellectual Property (data on patents).

In terms of internationalization and visibility, IBN is included in OpenDOAR (Directory of Open Access Repositories) and ROAR (Registry of Open Access Repositories) since 2018. Since 2016, IBN is ranked by the WEBOMETRICS platform. Its "Web Ranking of World repositories" in September 2020 positioned IBN among the top 30 national repositories in the world, rising from its 131 position in 2016. Over 90% of IBN publications are indexed by Google Academic, thus granting the much-needed international visibility to a large number of national publications.

The Manual of good practices for the maintenance of national bibliographic databases for research output has been developed in the context of the ENRESSH "European Network for Research Evaluation in the Social Sciences and Humanities". It is an informative resource, meant to encourage discussions for people who are involved with the implementation and maintenance of national bibliographic databases for research output, including developers, database administrators, policymakers, librarians, researchers and other interested parties (Sile et al. 2019).

The manual highlights 30 issues (not all of them of equal importance) relevant for database design, organization, maintenance and usage. All of the described issues are applicable to generic databases or bibliographic databases for any knowledge domain. The considerations that guided the development of the manual include the principles for research evaluation activities, as stipulated in Leiden Manifesto and San Francisco Declaration on Research Assessment, as well as the FAIR data principles.

The manual is based around 8 themes, namely *Design, Data collection, Organisation, Research output types, Vocabularies, authority control and identifiers, Quality control, Data use, Transparency and sustainability*. For each theme, there are a number of statements which are considered good practices, followed by the benefits of implementing such recommendations and accompanied by a brief description of practical hints for implementation.

This paper reviews IBN design, organisation, maintenance and usage from the perspective of the Manual of good practices regarding national bibliographic databases, aiming to come up with several recommendations for IBN reengineering and improvement.

4. Results and discussions

What follows is a series of general insights for each of the 8 themes and 30 issues of the Manual of good practices applied to IBN.

4.1 Design

When designing IBN in 2010, driven by the goal of improving the quality of national publications and increasing their visibility, ISDI adopted a bottom-up approach. It tried to draw on expertise from all relevant knowledge domains, as the project included a mixed team of researchers from both STEM and social sciences and humanities, representatives of policymakers and research administrators, as well as developers, database administrators and experts in research information systems. Over time, IBN has been continuously updated and improved. Therefore, ISDI team was complemented by researchers with expertise in library practices, in cataloging and research administration, all due to the new, emerging purposes of the database. Although initially envisaged as a bibliographic database of research publications from national journals, in time its purposes extended to evaluation and monitoring, namely journal accreditation and ranking, as well as bibliographic reporting, especially at journal level and bibliometric research and monitoring.

To ensure that IBN can fulfill its purposes and to benefit interoperability, the database data model was based on CERIF (Common European Research Information Format) and later on, the data structure has been partially adapted to the use of Dublin Core standard for metadata and bibliographic records. Based on the envisaged purposes of the database, taking into account its functionality, performance and maintainability, as well as prior experience of ISDI staff with certain technologies, the choice of building a new system was made. IBN was developed in Drupal CMS 7.7 & database on MySQL 5.5.68.

4.2 Data collection

In terms of systematical data collection, at the moment, the scope of IBN is limited to two output types from all national research organisations and higher education institutions, who are usually the editors of scientific journals and conference proceedings. There are agreements only with several editorial boards of national scientific journals, the majority of data being identified and covered by the ISDI staff responsible for IBN operation and maintenance. It should be noted, that to be recognized as scientific, both journals and conference proceedings must provide their publications to IBN, according to national regulations. Although manual input is a time-consuming process, and a small proportion of data already exists in 13 institutional repositories, data are reported manually by the database staff, using a mixed data collection approach: data are retrieved from publishers, institutional repositories, authors and commercial databases. IBN is updated continuously: new records are added and existing records are edited on daily basis.

4.3 Organisation

Since the set up of IBN, ISDI has made efforts to collaborate with a wide palette of relevant stakeholders, contributing to the usability and quality of the database. ISDI collaborates actively with university libraries, scientific secretaries of research organisations and vice-rectors for research from HEIs, as well as editorial boards of scientific journals and interested researchers. IBN is actively promoted via various channels, such as ISDI and IBN websites, scientific events, training workshops, individual discussions, promotional videos in public transport, printed flyers etc.

As ISDI is the only organization responsible for the maintenance of IBN, some of the roles and responsibilities regarding database flows are defined internally within ISDI, to increase operational efficiency and minimize errors.

Although the use of IBN is stipulated in two regulations of the National Agency for Quality Assurance in Education and Research (regarding evaluation, classification and monitoring of scientific journals and regarding the recognition, classification and recording of scientific events), as well as the funding contract for national research projects, stipulating the requirement to publish the research outputs of national publicly-funded project in Open Access in IBN, it is not otherwise embedded in the national legal framework. Partially this is due to its original bottom-up design process, but also due to the lack of a performance-based research funding system, including a mechanism for e-infrastructure funding, that would recognize its value. It in turn leads to the lack of commitment for the maintenance of the database at the government level, which is a constant threat for IBN existence.

4.4 Research output types

IBN has been envisaged initially to include articles from national research journals, to be extended in 2017 to include publications from conference proceedings. It should be noted that distinct characteristics of research output in different academic disciplines are not taken into account, but based on previous research of ISDI staff on the identification of relevant national research output types (Cuciureanu et al. 2018), in 2021 it is envisaged to carry out public consultations involving various stakeholders to establish types and categories of research outputs to be used consistently for various purposes (reporting, evaluation etc.) at the national level. In consequence, to facilitate more uses of the database, it is foreseen to extend IBN to include other research output types, as doctoral theses, monographs and book chapters, but only if continuous funding sources are identified at the national level.

4.5 Vocabularies, authority control and identifiers

Since its set up in 2011, aiming to provide a functional database and accurate data, IBN has strived to maintain up-to-date authority lists of journals (based on ISSN) and later on of conference proceeding, of authors and organisations, as each of these entities is represented in the database as a record with a persistent identifier. For instance, the author record contains name variants, while organization record – the time frame for information validity.

To enhance database functionality, since its inception terms from standardized vocabularies have been used, such as ISO 639 and ISO 3166 vocabularies for languages and countries, OECD FORD for scientific fields of publications and publication types recorded in IBN have been based on CASRAI typology, adjusted to the national situation.

Recently, in order to increase interoperability and research visibility, consistent efforts are being made by IBN to use international persistent identifiers, such as ORCID (3.4% of authors registered IBN have ORCID), DOI (2.6% of publications stored in IBN have DOI). It should be noted that over 53% of journal publications recorded in IBN are categorized based on UDC. The low use of ORCID and DOI is due to the fact that both of these identifiers have just recently been adopted by the national academia, there is no national agency assigning DOI, mostly because it is not yet a requirement for the national publications.

4.6 Quality control

ISDI is solely responsible for IBN maintenance, with manual input of data, therefore duplicate records are extremely rare, as deduplication is performed manually, at the data input stage. Author profiles in IBN are interconnected with various author IDs (Scopus Author ID, Researcher ID, Google Scholar ID), facilitating author disambiguation. The small size of the national scientific community also makes deduplication an easier task. One of IBN shortcomings is the lack of detailed guidelines for metadata input, which is performed manually and sometimes results in inaccurate and inconsistent data, especially regarding scientific fields for publications, because there are many multidisciplinary journals. On the other hand, automated data validation procedures (e.g. data field errors) during data input have been implemented in IBN since its set up. But due to the fact that IBN provides access to the full text of publications, missing data can be easily completed and the inaccuracies in metadata can be completed and validated, as data are matched with actual publications and also with the Scopus database. Still, this is a very resource-intensive task. A detailed statistics portal has been added to IBN additionally to the regular access to the database, allowing early detection of systematic errors.

4.7 Data use

In terms of data access, IBN basically allows for data to be searched (including advanced search features), browsed (with filtering and sorting features) and downloaded using the user interface of the database. All data are available under the Attribution-NonCommercial 2.0 Generic license. There is a procedure for data access by editorial boards, as they can edit data about their journals, submit and receive evaluation applications for journals accreditation, as required by the national regulation. There are also several journals, that export data from IBN via web services to their websites. IBN is being extensively used by the National Agency for Quality Assurance in Education and Research when awarding and confirming scientific and didactic degrees, as well as in the framework of the national Smart Specialisation strategy development, funded by the European Commission.

Another shortcoming of IBN is that, despite its export features in various formats for individual publications, data cannot be accessed in bulk, there is neither API access nor access to bibliographic data through metadata harvesting protocols. At the same time, IBN has enabled the crawling of bibliographic records by web search engines, it has also specifically implemented guidelines for Google Academic, thus leading to over 90% of IBN data being indexed in the world's largest academic database.

4.8 Transparency and sustainability

IBN has been continuously improved through the years in alignment with users' needs, mainly based on their feedback. It has established several channels for this purpose, creating a designated email address available on IBN website, which lately has been more actively accessed by users, especially to report errors, notify about missing data and plagiarism claims and provide suggestions for improvement. Another feedback channel is personal meetings with different user groups, aimed at gathering feedback. Although IBN provides on its website information about its purpose and envisioned uses, as well as user guides (<https://ibn.ISDI.md/en/about-IBN>), these documents are not available in English, don't clearly state IBN limitations and are not up-to-date.

Data provenance procedures are absent at the system level, while at the record level are ensured by the logging system of the database. Aiming to maintain IBN relevant and up-to-date, it is being continuously adapted to follow research practice and technological changes, to some extent.

5. Conclusions

This paper aims to analyse IBN design, organisation, maintenance and usage from the perspective of the Manual of good practices regarding national bibliographic databases. Out of 30 good practices presented in the manual, 16 are fully implemented, 9 are partially implemented in IBN and 5 are absent. The themes which are still challenging and require a sustained effort for implementation refer to data collection procedures (manual data input), research output types (IBN only includes two types of research outputs without distinctions related to specific academic disciplines) and data use (no possibility to access data in bulk, due to the lack of both an API and implemented metadata harvesting protocols). We believe that implementing these recommendations and also fully applying partially implemented features, if relevant and feasible for IBN based on its current status in the national research e-infrastructure, would result in an up-to-date, efficient, usable comprehensive and accurate national bibliographic database, that can facilitate multiple uses. The big challenge is to ensure IBN sustainability, under lack of clear commitment for the maintenance of the database at the government level.

References

- Adams, J., McVeigh, M., Pendlebury, D., & Szomszor, M. (2019). *Profiles, not metrics*. Clarivate Analytics. <https://clarivate.com/webofsciencgroup/campaigns/profiles-not-metrics/> (accessed 22.01.2021)
- Cojocaru, I., Cuciureanu, G., Moraru O. (2010). Instrumentul Bibliometric Național – sistem informatic performant, deschis, flexibil, scalabil [National Bibliometric Instrument - high-performance, open, flexible, scalable information system]. *Intellectus*, 2, 44-55. https://ibn.idsi.md/ro/vizualizare_articol/90079
- Clarivate Analytics. (2021). *Chinese Science Citation Database*. <https://clarivate.com/webofsciencgroup/solutions/webofscience-chinese-science-citation-index/> (accessed 22.01.2021)
- Cuciureanu, G., Țurcan, N., Ungureanu, E., Roșca, A., Cojocaru, I., Holban, I., Minciună, V., Cujba, R., Guzun, M., Cojocaru, I. Coșuleanu, I., Grecu, M., Cernăuțeanu, V., Botnaru, V. & Rusu, A., (2018). *Tipuri de rezultate ale activităților științifice [Types of scientific activity outputs]*. Institutul de Dezvoltare a Societății Informaționale.
- Feather, J., & Sturges, P. (2003). *International Encyclopedia of Information and Library Science* (Second ed.). London: Routledge. p. 127. ISBN 0-415-25901-0
- Instrumentul Bibliometric Național. (2021). <https://ibn.idsi.md/> (accessed 22.01.2021)
- Jin, B. & Wang, B. (1999). Chinese science citation database: Its construction and application. *Scientometrics* 45, 325–332. <https://doi.org/10.1007/BF02458440>.
- Moskaleva, O., Pisyakov, V., Sterligov, I., Akoev, M. & Shabanova. S. (2017). *Russian Index of Science Citation: Overview and Review*. International Society for Scientometrics and Infometrics. https://dspace.spbu.ru/bitstream/11701/8579/1/Paper200_RISC.pdf
- Packer, A. (2000). *SciELO - a Model for Cooperative Electronic Publishing in Developing Countries*. D-Lib Magazine. Vol. 6 no. 10. ISSN 1082-9873
- Puuska H.-M., Guns, R., Pölönen, J., Sivertsen, G., Mañana-Rodríguez, J., & Engels, T.C.E. (2018). *Proof of concept of a European database for social sciences and humanities publications: Description of the VIRT-A-ENRESSH pilot*. Helsinki: CSC & ENRESSH, <https://doi.org/10.6084/m9.figshare.5993506>
- Reitz, J. (2004). "bibliographic database". *Dictionary for Library and Information Science*. Westport, Connecticut: Libraries Unlimited. ISBN 1-59158-075-7.
- Sîle, L., Guns, R., Ivanović, D., Pölönen J., & Engels T.C.E. (2019). *Creating and maintaining a national bibliographic database for research output: manual of good practices*. ENRESSH & ECOOM: Antwerp. <https://doi.org/10.6084/m9.figshare.9989204>
- Sile, L., Pölönen, J., Sivertsen, G., Guns, R., Engels, T., Arefiev, P., Dušková, M., Faurbæk, L., Holl, A., Kulczycki, E., Macan, B., Nelhans, G., Petr, M., Pisk, M., Soós, S., Stojanovski, J., Stone, A., Šušol, J. & Teitelbaum, R. (2018). *Comprehensiveness of national bibliographic databases for social sciences and humanities: Findings from a European survey*, *Research Evaluation*, 27 (4). 310–322. <https://doi.org/10.1093/reseval/rvy016>
- Sîle, L., Guns, R., Sivertsen, G., & Engels, T.. (2017). *European Databases and Repositories for Social Sciences and Humanities Research Output (Version 2)*. figshare. <https://doi.org/10.6084/m9.figshare.5172322.v2>
- Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J. & Johnson, B. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. HEFCE. <https://doi.org/10.13140/RG.2.1.4929.1363>.