OPEN ACCESS

CrossMark

Grant Proposal

Reference implementation for open scientometric indicators (ROSI)

Christian Hauschke‡, Simone Cartellieri‡, Lambert Heller‡

‡ Technische Informationsbibliothek (TIB) – German National Library of Science and Technology, Hannover, Germany

Abstract

Within the project "Reference implementation for Open Scientometric Indicators" (ROSI), new assessments and visualizations of conventional and alternative metrics (altmetrics) will be developed and their effect on researchers will be investigated. For this purpose, a reference implementation based on the open source research information system VIVO will be developed in which various metrics are combined with data from different openly licensed sources. In order to develop the requirements of the target groups, surveys are going to be conducted to investigate the effect of scientometric indicators on scientist's and their expectations regarding those indicators. The objectives of the project are firstly to evaluate the scientometric needs and concerns of the target groups, and secondly to implement a usable reference implementation of a toolset that reflects the results of the study and that enables transparent, license-free, flexibly adaptable analysis of the output of researchers, contributors and organisations.

Keywords

open science; scientometrics; bibliometrics; research information

Short description

The conventional evaluation of the impact of scientific publications is carried out using metrics such as the Journal Impact Factor or the h-index. The representation of high or low impact using a simple indicator has been a controversial issue for some time and causes dissatisfaction among researchers. The use of metrics such as the Journal Impact Factor (JIF) for research evaluation is problematic because it is calculated on the basis of licensed data from Clarivate Web of Science and can hardly be reproduced or varied by the end user (Herb 2016, PLoS Medicine Editors 2006). However, in addition to such proprietary metrics (including Elsevier Scopus, Google Scholar, and ResearchGate), data is increasingly available under free licenses, including those from CrossRef Event Data, Wikidata, and Microsoft Academic Search (Das 2015, p. 67).

Within the ROSI project new assessments and visualizations of freely licensed conventional and alternative metrics (altmetrics) will be developed and their effect on the target groups research and science administration will be investigated. Not only metrics are undergoing a change, the evaluation of the influence of a publication by measuring its impact in the exclusively scientific field itself is also being called into question (Bornmann and Haunschild 2016). The project is oriented towards "Leiden Manifesto for Research Metrics" (Hicks et al. 2015), which criticizes the state-of-the-art of research evaluation and sets out ten principles for improving the current practice. The ROSI project therefore also addresses the heterogeneity of research products beyond the classical journal article (research data, research software, and blog postings etc.), the roles of researchers and contributors with regard to these different products, as well as different references depending on product type and role (e.g., reviews, re-use in the case of research data or software, or derived works).

Within the framework of the project, information should be collected from freely accessible sources such as those mentioned above on the basis of personal, document or organisation identifiers. This is usually done via interfaces explicitly provided for this purpose, e.g. the CrossRef Event Data API. The information collected in this way can be displayed at different aggregation levels and thus enable evaluations at the three levels named: person, document or organization. In addition to conventional publication types such as books or articles, the definition of the term document also includes other object types such as scientific software and research data, which will become increasingly important in the scientific discourse (Peters et al. 2016, Smith et al. 2016). For this purpose, a reference implementation for the community-based open source research information system VIVO will be created, which combines various metrics with data from open sources such as CrossRef Event Data API, or Wikidata (cf. Wilsdon et al. 2015).

The project will make use of qualitative methods. We will analyze the current use of indicators, evaluate the requirements and needs of the scientific community, and examine their repercussions. In order to determine the requirements of the target groups, interviews with scientists from different disciplines are conducted in a first step in order to evaluate different views on scientometric indicators. The knowledge gained is used for an initial draft

of the reference implementation. In the further process, the perception of this reference implementation is evaluated in focus groups and the implementation is adapted. In these focus groups, the selection of individual indicators, the method of aggregation, and different ways of visualization are discussed. Diffenrent focus groups are addressed to achieve a domain-specific representation of the scientometric indicators. In this way, the handling of new object types such as software or research data can also be evaluated. The reference implementation is adapted to their needs through feedback from the target groups in an iterative process. Through free licensing of the initial data and the software created in the project and the detailed documentation of the results, the knowledge gained and also the implementation is going to be easily transferable to research-related platforms.

The objectives of the project are firstly the evaluation of the scientometric requirements and concerns of the target groups and secondly the implementation of a reference implementation of a toolset which reflects the results of the study and which enables a license-free, flexibly adaptable analysis of the output of researchers, contributors and organisations. This includes the user-friendly integration of bibliometrics into researcher profiles according to the needs of scientists.

Related projects

The DFG-funded project Linked Open Citation Database (LOC-DB) of the Hochschule der Medien (HdM) Stuttgart and the Christian-Albrechts-Universität zu Kiel deals with the development of practical tools and processes in the field of linked data technologies for the participation of libraries in an open, distributed infrastructure for the indexing of citations.

The Initiative for Open Citations (I4OC) is an association of scientific publishers, scientists and various organisations in the field of "Openness". It tries to increase the availability of freely licensed data. Founding members include Datacite, OpenCitations, PLoS and Wikimedia.

At the Centre for Science and Technology Studies at Leiden University, the Science and Evaluation Studies research group analyses the impact of various research evaluation practices. One goal of the group is to develop a deeper empirical understanding of the effects of evaluation practices on academic knowledge production. With its research results, it aims to contribute to a responsible handling of research results and metrics.

The NISO Alternative Assessment Metrics (Altmetrics) Initiative tried to identify different perspectives and requirements for dealing with new metrics in the scientific field. For this purpose, the existing metrics were documented and examined with regard to data quality problems.

VIVO is a community-based open source software for displaying researcher profiles and research output on the Net. VIVO offers storage, editing, search, browsing and visualization of scientific activities. VIVO supports visualized evaluations and thus the presentation of scientific outputs, and relies on linked data, and open standards. The

ontologies (especially ontologies C4O and CiTO for mapping bibliometric information) are used to describe scientific persons, projects, and publications. In addition, the integration of an Altmetrics widget is possible.

The ROSI project aims to bring together the findings of the projects described here in the field of scientometric indicators and to evaluate the implications for the people involved in research output in order to derive practices that take into account the interests of scientists. It will be shown how the free availability of publication and bibliometric data from different sources as well as methods (in the form of the VIVO software, its ontologies, and extensions) makes it possible to represent research impact in a very flexible and with a wide variety of configurations. On this basis, a broader and systematic comparison can be made with the requirements of the scientists themselves.

Work packages

The work programme is divided into 5 work packages and will run for twenty-four month (Fig. 1). Work packages 1 to 3 mainly deal with quantitative methods: selection of data sources for scientometric information, data aggregation and integration, and development of reference implementation based on VIVO. In work package 4 qualitative procedures are carried out, whereby the use and the evaluation standards of scientometric indicators are considered and the effects on the scientific community are evaluated by questioning experts. In work package 5, a handbook on how to deal with this data will be prepared on the basis of the findings and the project results published.

WP	PM1*	PM2*	Year 1				Year 2			
			Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
1	3	4	■	■						
2	2	6		■	■					
3		12			■	■	■	■	■	■
4	15		■	■	■	■	■	■		
5	4	2							■	■
Sum	24	24								

Figure 1. [doi]

PM1 and PM2: Person months of the first or second project member; Q: quarter.

The TIB is responsible for project management and public relations.

Work package 1: Data sources for scientometric information

Work package 1: Data sources for scientometric information

Objective: WP 1 serves to view and evaluate possible data sources for scientometric information.

Tasks: This requires extensive and in-depth research of the sources relevant for the scientific reference framework (e.g. CrossRef Event Data API or LOC-DB) and evaluation according to criteria such as openness, availability, sustainability, as well as content-related and technical suitability.

Work package 1.1: Research and evaluation of relevant data sources

In this sub-task, data sources are tested for their suitability in terms of content in relation to various scientific communities (e.g., engineering sciences, humanities). Only open data sources are taken into account.

Work package 1.2: Research and technical evaluation of relevant data sources (computer scientist)

In this subtask, potential data sources are tested for technical suitability. This includes criteria such as the existence and type of interfaces, data formats and the type of licensing.

Work package 2: Data aggregation and integration

Goal: WP 2 comprises the conception of workflows for harvesting and aggregation of the data sources found to be suitable, including the modelling of the data.

Tasks: In this WP, the collection and consolidation of the data is conceived, which ultimately results in merging the data from the sources selected in WP 1 into VIVO, the target system for the reference implementation. For this purpose, existing ontologies such as CiTO must be examined, checked for their suitability for mapping the data sources determined in WP 1 and extended by missing elements for mapping the scientometric data on the basis of the needs determined in the project.

Work package 2.1: Formulation of initial requirements

In this sub-task, existing ontologies are examined in alignment with the needs of scientific users, checked for their suitability in terms of content for mapping the identified data sources and supplemented by missing elements.

Work package 2.2: Data aggregation and integration

In this sub-task, the collection and consolidation of data is technically designed in order to merge the data from the sources selected in WP 1 into VIVO. The technical suitability for the integration of existing ontologies is tested and the integration is implemented. Based on this, a data model is designed.

Work package 3: Development of the reference implementation

Objective: WP 3 includes the development of a reference implementation in VIVO based on the collected data sources and the data model determined in WP 2 for the presentation and exploration of the scientometric data.

Tasks: The implementation is a process of mutual influence with WP 4, findings from WP 4 are transferred into the reference implementation in an iterative development process. In addition to the development of a first draft based on the literature and the findings of the individual interviews, the findings obtained in WP 4 must be translated into requirements and the software adapted accordingly. This includes, among other things, the integration of various data sources, but the continuous improvement of indicator visualization, usability and user experience (UX) of the application as a whole. Special attention is also paid to the ease of reusability of the developed application.

Work package 4: Practical use of indicators, assessment standards and repercussions

Objective: WP 4 enriches the results of the quantitative study with qualitative aspects by interviewing experts and conducting workshops with focus groups. It looks at the current use of indicators, evaluates the evaluation criteria and needs of the scientific community, and examines their repercussions.

Tasks: This work package is divided into three sub-tasks and comprises the preparatory and accompanying investigation of the impact of bibliometric visualisations and aggregations on scientists. Individual interviews and workshops with focus groups are used to ask them about their current handling of scientometric data. In addition, requirements and wishes in this regard are to be determined. During the project, the qualitative aspects of the reference implementation developed in WP 3 will be examined in more detail and the fulfilment of the requirements determined will be evaluated.

Work package 4.1: Conceptual design and implementation of individual interviews

This sub-task investigates usage practices and perception of scientometric information. It comprises the conceptiual design and application of partially standardized individual interviews with scientists from various domain-specific communities and functions, such as doctoral students, professors and research assistants.

In preparation for the interviews, a literature study will determine which indicators are currently widespread in which areas of application, and how their use and public presentation is evaluated by scientists. The indicators are clustered according to various criteria (practical relevance, dissemination, reservations, etc.). Based on this, hypotheses and questions are created for an interview guide.

This guide will be subjected to a pre-test with scientists from the TIB's collaborations (e.g. from the TIB's collaboration with universities in Hannover) and adapted. Thanks to its role as a university library and the world's largest special library for science and technology, and being a member of the Leibniz Association, a variety of contacts from various disciplines can be drawn upon to ensure that the interviewees are sufficiently diverse. The interviews are conducted by telephone or video conference due to the high time utilization of the target group. The statements from the interviews are systematically evaluated, thematically arranged, conceptualized and compared with the initial hypotheses.

Work package 4.2: Workshops with focus groups

This sub-task comprises the conceptiual design, implementation and evaluation of workshops with focus groups to evaluate the prototype created in WP 3. For this purpose, focus groups of 5 to 8 people are formed. Three different groups are going to be adressed: doctoral students/early researchers, professors, and research assistants. The groups consist of representatives from various disciplines (humanities, engineering, mathematics and natural sciences, social sciences).

Participants will be recruited from the TIB's scientific environment. The workshops serve to test the prototype, to check the acceptance of the visualized scientometric information, to prioritize different functions and to evaluate the usability of the application. The workshops are recorded for evaluation purposes. The results are documented and prepared for the further development of the prototype. The workshops take place at close intervals to allow the results of all focus groups to be incorporated into the prototypes.

Work package 4.3: Final evaluation of the prototype

This work package includes the final evaluation of the prototype in individual interviews with persons from the functional groups and disciplines outlined in WP 4.2. The interviews examine to what extent the reference implementation meets the needs and wishes or reservations that have been raised. The examination refers to content aspects (regarding indicator selection, visualization) and technical aspects (usability, user experience).

Work package 5: Project documentation and preparation for reuse

Objective: Documentation and publication; project communication

Tasks: In WP 5, all work results are documented for easy reuse and the resulting source code and anonymized raw data are published. In addition, a handbook will be prepared.

Work package 5.1: Preparation of a recommendation

In this sub-task a recommendation on the handling of scientometric data on the basis of the knowledge gained with regard to the requirements of scientists for the presentation of scientometric information is presented.

Work package 5.2: Publication of source code incl. documentation

In this sub-task the results of the work are documented and the resulting source code as well as anonymized raw data are published. This includes the reference implementation in VIVO as well as documentation for system-independent reuse of the reference implementation by third parties.

Notes on the implementation of research data management

The project will produce various research data. These are primarily the data from the individual interviews and the focus group surveys. Research data is published in accordance with the TIB Research Data Policy. Data is anonymized and stored in the most open, standardized formats possible and, as far as possible, also published.

Information on application opportunities and re-use

The results of this project can be widely re-used and applied for different purposes.

First, the VIVO-based reference implementation is available for free reuse. The reference implementation can be re-used in the context of research information systems, research profile services or discovery systems. Another technical exploitation of the project results is the re-use of the ontologies created or expanded in these or similar services.

The project handbook (see WP 5) on the handling of scientometric data will document the knowledge gained and processed for subsequent use and provides a guideline for the handling of scientometric data and indicators. Among other things, the TIB itself offers a comprehensive range of advice for publishing researchers on Open Access, rights and opportunities for authors, into which the handouts will also flow.

Another core element to reuse the scientific data of this project is the publication in peer-reviewed journals. Knowledge transfer is also realized via project communication (public relations) and multipliers. These are going to be selected from the expert surveys and the VIVO community. The TIB is an active member of the VIVO community and organizes VIVO workshops and participates in conferences and in technical communication on VIVO in general. Other starting points include events such as the annual conference of research advisers in Germany, at which the TIB had already held a heavily attended workshop in 2015, and the workshop as part of the LOC DB project (see invitation in the Letter of Intent of the Mannheim University Library).

The DINI AG FIS is also an important hub for the introduction of research information systems in Germany.

The transfer of scientific results also opens up perspectives for follow-up projects.

Funding

Funding program

Quantitative Wissenschaftsforschung

Grant title

Referenzimplementierung für offene szientometrische Indikatoren

Hosting institution

Technische Informationsbibliothek (TIB) – German National Library of Science and Technology

Conflicts of interest

The authors declare that they have no conflict of interest.

References

- Bornmann L, Haunschild R (2016) Does evaluative scientometrics lose its main focus on scientific quality by the new orientation towards societal impact? Scientometrics 110 (2): 937-943. https://doi.org/10.1007/s11192-016-2200-2
- Das AK (2015) Open Access Research Impacts. In: Satija MPMS (Ed.) Introduction to Open Access. UNESCO, Paris. URL: http://eprints.rclis.org/24906/1/LS1.5%20Open%20Access%20Research%20Impacts.pdf [ISBN 9789231000744].
- Herb U (2016) Impactmessung, Transparenz & Open Science. Young Information Scientist 1: 59-72. URL: https://yis.univie.ac.at/index.php/yis/article/view/1420/1380
- Hicks D, Wouters P, Waltman L, Rijcke S, Rafols I (2015) Bibliometrics: The Leiden Manifesto for research metrics. Nature 520 (7548): 429-431. https://doi.org/10.1038/520429a

- Peters I, Kraker P, Lex E, Gumpenberger C, Gorraiz J (2016) Research data explored: an extended analysis of citations and altmetrics. Scientometrics 107 (2): 723-744. https://doi.org/10.1007/s11192-016-1887-4
- PLoS Medicine Editors (2006) The Impact Factor Game. PLoS Medicine 3 (6): e291. https://doi.org/10.1371/journal.pmed.0030291
- Smith A, Katz D, Niemeyer K, Working Group FSC (2016) Software citation principles. PeerJ Computer Science 2: e86. https://doi.org/10.7717/peerj-cs.86
- Wilsdon J, Allen L, Belfiore E, Campbell P, Curry S, Hill S (2015) The Metric Tide. Report of the Independent Review of the Role of Metrics in Research Assessment and Management. Higher Education Funding Council for England, 163 pp. [ISBN 1902369273] https://doi.org/10.13140/RG.2.1.4929.1363