

Quality Assurance and Benchmarking of Research Impact

Guidelines for the prudent use of SciVal and Incites

Version 13 September 2017

Research & Valorisation, Strategy & Quality Assurance

To be discussed by the UCW Sep 26th 2017

Content:

<i>Introduction</i>	<i>p. 1</i>
<i>Guidelines</i>	<i>p. 2</i>
<i>Appendix 1) Relevant definitions</i>	<i>p. 4</i>
<i>Appendix 2) Current tools and (expected) developments</i>	<i>p.10</i>
<i>Appendix 3) Leiden Manifesto...</i>	<i>p.11</i>
<i>Appendix 4) Model benchmark analysis (of Heymans Institute).....</i>	<i>p.14</i>
<i>Appendix 5) Expertise Network Research Analytics.....</i>	<i>p.35</i>

Introduction

The development and supply of tools to analyze research performance and impact have increased tremendously over the past years. This allows universities and researchers to perform their own, in-house analyses at a level of sophistication previously attained by external experts such as the Leiden CWTS¹ only. The two major tools currently available and used at the University of Groningen (UoG), are Clarivate's² InCites and Elsevier's SciVal, based on the citation databases Web of Science (WoS) respectively Scopus.

The increasing availability and user friendliness of these tools offers new opportunities to gain insight, to provide custom benchmarks and to provide valuable information to support both research management and PRC's (see appendix 1). Potentially, it can save time and money as well. However, this also poses new risks and challenges.

Incompetent use of tools or erroneous interpretation of results may lead to biased conclusions and advice. In-house analyses also raise the question of independence: how trustworthy are these compared to external benchmarks by independent experts?

As a **first step**, this memo proposes some guidelines for the prudent use of our current two major benchmark tools, SciVal/Scopus and InCites/WoS. **Final goal** is to develop a UoG policy and vision on the use, support and development of (current and future) benchmark tools, allowing sophisticated in-house analyses of research impact, and how this relates to our quality assessment policy and procedures.

Restrictions

A clear limitation of our current tools is that they mainly generate quantitative 'evidence' for one dimension of (academic) impact only, viz. indicators for citation impact and collaboration. However, as described in appendix 1 (Relevant definitions), new 'altmetric' tools develop rapidly and promise to provide more insight in societal impact as well. They are therefore included in our longer term goals for monitoring and testing new tools.

Likewise, the provision of qualitative evidence that may be used in SEP benchmarks as well, requires other sources and expertise than currently available.

Given the current lack of in-house experience with altmetric tools and qualitative sources, the guidelines below mainly apply to the use of SciVal and Scopus (see appendix 2 for a brief description and some recent and expected developments). In the near future, additional guidelines may be added in response to the anticipated expansion of our tool set and expertise.

¹ The Centre for Science and Technology Studies (CWTS) of Leiden University, a renowned research institute specialized in bibliometrics and frequently hired to conduct customised research impact analyses. See www.cwts.nl and www.cwtsbv.nl

² Formerly Thomson Reuters

GUIDELINES

SEP, SciVal and Scopus

Assessment of Research quality

A major principle underlying the SEP and the UoG research protocol, is that quality assessment requires independent expert review by informed peers (assigned to Peer Review Committees, PRC's). "Informed" means that their judgement must be based on both qualitative (interviews, key outputs, narratives, self evaluation texts) and quantitative (SEP tables with key figures, indicators, benchmarks) information.

Benchmarking

The benchmark forms an important and integral part of the SWOT analysis to be included in a SEP self-evaluation (see appendix 1). Our current in-house tools may provide useful information to include in the benchmark and could in principle render externally commissioned benchmarks obsolete. However, exclusive presentation of quantitative information obtained with our in-house tools might raise questions about the objectivity, reliability and suitability³ of the data. Therefore we propose the following:

Guidelines for the use of in-house benchmarks in relation to SEP assessments:

- a) Internal benchmarks may be used for SEP evaluations, but must always be validated by an external expert, when not supplemented with an independent external benchmark.
- b) The in-house tools can be used by the researchers/institutes under review to explain and respond to the outcome of an external benchmark.
- c) Before including the results of in-house analyses in a self-assessment, they should be submitted for internal peer review by RI specialists.

Guidelines for the use of SciVal and InCites for in-house benchmarking:

- d) The UoG adheres to the 10 principles propagated in the Leiden Manifesto (appendix 3), in particular:

Nr.5: *"Allow those evaluated to verify data and analysis."*

This means an obligation to check the completeness and validate the attribution of all outputs in PURE before analyzing an uploaded publication set in SciVal (and to adhere to guideline c).

Nr.7: *"Base assessment of individual researchers on a qualitative judgement of their portfolio."*

This means an obligation to:

- Discourage the sole use of citation metrics to evaluate an individual researcher, in particular when still in their early career⁴.
- Stimulate investigation of one's own micro-level field using 'fingerprinting tools' (sophisticated algorithms yielding research profiles based on full text analyses; see appendix 2 for a SciVal example), e.g. to find relevant colleagues, potential postdoc positions, etc.

³ The SEP explicitly states that a benchmark may also comprise the SEP or even be limited to qualitative information; this requires other benchmark sources.

⁴ It takes time to attain a high citation impact, as e.g. reflected in the H-index (partly an Age-index).

Development RI expertise and support

- e) To act as internal RI specialist, mentioned in guideline c), UoG staff will have to join the ENRA and participate actively in its activities. Activity and level of expertise to be monitored by the UB and R&V RI coordinators (Shaya Abdolazadeh, Jules van Rooij).
- f) Internal RI specialists are encouraged to keep up with the most relevant developments, e.g. by following relevant courses (CWTS⁵, CHEPS⁶, ESSS⁷), reading and sharing relevant literature and by active participation in relevant networks, both locally (ENRA), nationally (Research Intelligence Network Nederland, RINN, hosted by a SURF Special Interest Group) and internationally.

DRAFT

⁵ The Leiden Centre for Science and Technology Studies (CWTS) organizes several relevant courses for postgraduates and professionals, see <https://www.cwts.nl/training-education>

⁶ The Univ of Twente Department of Science, Technology, and Policy Studies, organizes an annual R&D Evaluation course, see https://www.utwente.nl/en/bms/steps/education/postgraduateprofessional/R_D_evaluation_course/#the-topic

⁷ European Summer School for Scientometrics, see <http://www.scientometrics-school.eu/about.html>

Appendix 1) Relevant definitions

Research quality

A major principle underlying the UoG research policy, is that quality assessment requires independent expert review by informed peers (assigned to Peer Review Committees, PRC's). "Informed" means that their judgement (of Quality, Societal relevance and Viability) must be based on both qualitative (interviews, key outputs, narratives, self evaluation texts) and quantitative (SEP tables with key figures, indicators, benchmarks) information.

The outcome of quantitative analyses can never replace the work of PRC's, nor does it suffice as the sole source of information for a full assessment. However, good analyses and sophisticated benchmarks do provide valuable information to support both research management and PRC's.

See also appendix 3: Hicks, D., Wouters, P., Waltman, L., De Rijcke S. & Rafols, I., 2015. *Bibliometrics: The Leiden Manifesto for research metrics*, Nature 520(7548), pp. 429-431.

Research impact

We here define impact in its broadest sense, referring to both academic and societal impact. It therefore requires an explanation of the three underlying terms:

- **Impact**: is generally defined⁸ as (having) a demonstrable effect, i.e. resulting in a perceptible change, shift or influence; in addition, the magnitude of the effect is relevant and may require quantitative indicators or proxies.
- **Societal impact**: in the UK, since the introduction of 'impact' as a separate evaluation criterion in the national research assessment (REF 2014), impact is often used in a more restricted sense, as formulated by the Higher Education Funding Council for England (HEFCE)⁹: "*The Research Excellence Framework was the first exercise to assess the impact of research outside of academia. Impact was defined as 'an effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia'.*"

Our national protocol (SEP) requires assessment of *Societal relevance* as a separate criterion (next to Quality and Viability) but does not provide a definition. However, Dutch universities have agreed upon a slightly different definition of societal impact in 2013, referred to as 'valorisation', in the VSNU 'Raamwerk Valorisatie-indicatoren'¹⁰: "*Het proces van waardecreatie uit kennis, door kennis geschikt en/of beschikbaar te maken voor economische en maatschappelijke benutting en geschikt te maken voor vertaling in concurrerende producten, diensten, processen en nieuwe bedrijvigheid.*"

Important to note here is that neither the UK nor Dutch definition is limited to economic impact; both private / commercial and governmental / not-for-profit stakeholders can be targeted, as well as the general public.

⁸ E.g. in Robin Chin Roemer and Rachel Borchardt (eds.), 2015. *Meaningful metrics : a 21st century librarian's guide to bibliometrics, altmetrics, and research impact*. Association of College and Research Libraries, American Library Association Chicago, Illinois

⁹ See <http://www.hefce.ac.uk/rsrch/REFimpact/>

¹⁰ See <http://www.vsnu.nl/valorisatie-indicatoren.html>

- **Academic impact:** we here use the definition of the Research Councils UK¹¹: “*the demonstrable contribution that excellent research makes to academic advances, across and within disciplines, including significant advances in understanding, methods, theory and application*”.

For the present purpose, the differences between the two definitions of societal impact are not all that relevant. More important is that we explicitly address academic impact as well. Besides their societal relevance, the impact researchers have on academic peers obviously remains a major focus of SEP research assessments (as reflected in the criterion Quality).

Benchmarking

In business, benchmarking is defined as¹²: “*comparing one’s business processes and performance metrics to industry bests and best practices from other companies*”. For research units, this translates to comparing an institute’s performance to that of other institutes with a comparable mission. If the mission is confined to doing fundamental, curiosity driven research, the benchmark could be limited to comparing academic impact only. However, the SEP requires assessment of *Societal relevance* for any institute, regardless of its mission. **For units with a predominantly fundamental mission, UoG policy is that sufficient effort must be invested at least in ‘outreach’**, i.e. informing stakeholders beyond academia and/or the general public in such a way that sufficient support for the research can be maintained.

The SEP requirements for the self evaluation documents to be provided to a PRC, include an analysis of Strengths, Weaknesses, Opportunities and Threats (SWOT) plus benchmark¹³:

“The SWOT analysis is intended primarily as a tool for reflecting on the research unit’s position in research and society and opportunities for development in these domains in the near future. For an academic institution this means two things: a form of benchmarking by looking at the performance of comparable units elsewhere, and an analysis of its own mission in relation to developments in the Dutch and European policy context. The unit indicates the reference points for comparison in the relevant context (international or national). There are all sorts of different SWOT analyses. At one end of the spectrum, there are analyses based on huge quantities of data; on the other, there are analyses that make use of interviews, focus groups or other more qualitative methods. Where possible and useful, a SWOT analysis may include quantitative data. Where that is not possible or useful, qualitative data will suffice. It is naturally possible to combine the two. The SWOT analysis should be supported by specific evidence.”

This quote reveals that the requested benchmark is a major component of the SWOT analysis and not necessarily confined to quantitative analyses; qualitative benchmarks are allowed as well.

Research impact metrics: we follow the distinction between ‘*traditional metrics*’, ‘*usage metrics*’ and ‘*altmetrics*’, as made in a recent EC ‘Expert Group on Altmetrics’ report¹⁴, from which the following figure and quotes were copied:

¹¹ See <http://www.rcuk.ac.uk/innovation/impacts/>

¹² E.g. in Wikipedia: <https://en.wikipedia.org/wiki/Benchmarking>

¹³ As described in SEP appendix D4, see footnote 4 for a url to the SEP

¹⁴ AJ. Wilsdon et al., 2017. *Next-generation metrics: Responsible metrics and evaluation for open science*. Downloaded from https://ec.europa.eu/research/openscience/index.cfm?pg=altmetrics_eg

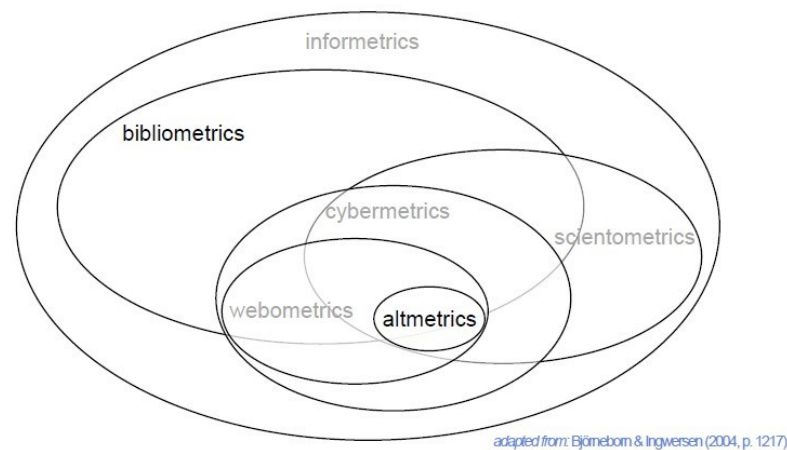


Figure 2: The basket of metrics for the evaluation of science (Haustein, 2015).

- Traditional metrics** (based on research outputs, mainly journal publications): “The two basic types of metrics are the number of publications, and the number of citations the publication receives. These measures may be aggregated at different levels. An object of evaluation can be a single publication, a researcher, a research unit, an institution, or a country. Citation and publication counts are derived from bibliometric databases (Web of Science – WOS, Scopus and to some extent from Google Scholar). More recently, bibliographic databases have also started to display citation counts of publications. The citation counts are based on the coverage of specific databases and thus vary considerably between the different data sources (Bar-Ilan, 2008).”

....

“From the raw data (publication and citation counts) more sophisticated indicators have been created, such as the Journal Impact Factor – JIF (Garfield, 1972), the h-index (Hirsch, 2005), field normalized citation indicators (Waltman & van Eck, 2013), Eigenfactor (Bergstrom, West & Wiseman, 2008), SJR (Gonzalez-Pereira, Guerrero-Bote, & Moya-Anegon, 2010), SNIP (Moed, 2010), or the newly introduced CiteScore (Elsevier, 2016). Publication and citation cultures differ greatly across disciplines and sub-disciplines, which implies a need to normalize indicators when objects are compared from several disciplines or different aggregation levels.”

....

“When conventional bibliometric indicators are used properly they provide valuable insights on the scientific impact of research publications. As already noted, in an evaluation context, best practice is for bibliometric indicators to be used together with qualitative assessment (peer review).”

- Usage metrics** (based on the number of views or downloads of an item): “Usage is different from citations, because there are many potential users (students, policy makers, the interested public) who read publications or use data without ever publishing. In addition, not everything a researcher reads is referenced in her publications. Usage based metrics, like the usage impact factor (Kurtz & Bollen, 2008) or libcitations (White et al., 2009) measure attention and uptake. Usage metrics are highly relevant for open-science, not only in terms of the usage of publications, but also for tracking non-traditional publications (posts, blogs) and for the re-use of open data or open software. Open access publishers provide usage information of individual articles (e.g. PLoS), and commercial publishers are also starting to open up. Several publishers, (e.g. Springer Nature, IEEE, ACM) display the number of downloads of the specific article from

their platform. Elsevier's Science Direct in cooperation with Mendeley provides information to researchers on the number of downloads of their publications from the Science Direct platform."

- **Altmetrics** (based on social media applications): "With the advent of Web 2.0 technologies, new possibilities have appeared for assessing the "impact" of scientific publications, not only of journal publications, but also books, reports, data and other non-traditional publication types. Altmetrics have become a means for measuring the broader societal impacts of scientific research. The idea was introduced by Neylon and Wu (2009) as "article-level metrics" and by Priem and Hemminger (2010) as "Scientometrics 2.0". The concept of alternative ways to assess scholarly activities was further developed and extended in "Altmetrics: A Manifesto" (Priem, Taraborelli, Groth & Neylon, 2010), and was named "altmetrics" a shorthand for alternative metrics. Altmetrics are based mainly on social media applications, like blogs, Twitter, ResearchGate and Mendeley.

It should be noted that there are constant changes and developments in online applications: new platforms appear (e.g., Loop, WhatsApp, Kudos), while other lose their appeal (e.g., MySpace and, according to some reports, even Facebook) or even disappear (e.g., Connotea, Delicious, ReaderMeter). There are different kinds of measurable signals on social media, e.g., likes, shares, followers, downloads, posts, mentions and comments - each indicating a different level of involvement. Thus, several categorizations of altmetrics have emerged (e.g. Lin & Fenner, 2013; Haustein, 2016).

Priem et al. (2010) emphasized the advantages of using altmetric signals – they are fast relative to citations; cover not only journal publications, but also datasets, code, experimental design, nanopublications, blog posts, comments and tweets; and are diverse, i.e. providing a diversity of signals for the same object (e.g., downloads, likes, and comments).

A theoretical framework was introduced by Haustein et al. (2016), where they define acts leading to online events that are recorded and thus can be measured. They differentiate between three levels of act categories of increasing engagement:

- Access (e.g. view metadata, accessing content, storing research object)
- Appraise (e.g. on Twitter, on a listserv, in Wikipedia, in mainstream media, in a scientific document, in a policy document)
- Apply (e.g. theories, methods, results, software code, datasets)

Currently there are three major altmetrics aggregators: Altmetric.com, PLUMx and ImpactStory each collect a slightly different set of indicators from primary sources. It should be noted that some of the data sources can be directly accessed through available APIs without the need to subscribe to the commercial aggregators. In addition, commercial aggregators often offer data for research purposes (e.g., Altmetric.com).

Article level indicators are provided by most major publishers (downloads, Mendeley readers, tweets, news mentions etc.). Most of this information is provided by the above-mentioned commercial aggregators. Springer together with Altmetric.com introduced Bookmetrix – that aggregates altmetric data from chapters within the book to the book level (Springer, 2015). Currently, this platform is only available for Springer's books, but such an indicator would be useful for open-access books as well. Author-level altmetrics are provided by Impactstory. PLUMx provides also research unit and institution level altmetrics. Datacite, Zenodo, GitHub and Figshare (and possibly other repositories) provide DOIs for uploaded data, which enables to cite data sources and to track usage, an excellent altmetric for open science."

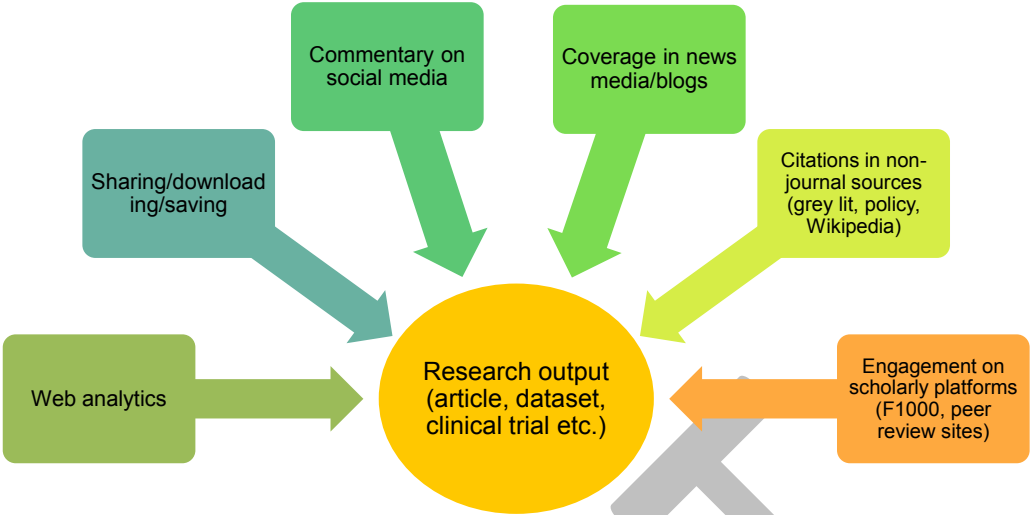


Figure from presentation Shaya Abdolazadeh 13 June 2017

DRAFT

Appendix 2) Current tools and (expected) developments

Major tools available

Web of Science and WoS based tools: **InCites**, **Essential Science Indicators (ESI)**, **Journal Citation Reports (JCR)**. Campus license, available at <http://apps.webofknowledge.com/>

< Brief descriptions of tools to be added >

Scopus and its tools: **SciVal**, **CiteScore**, **Source Normalized Impact per Paper (SNIP)** and **SCImago Journal Rank (SJR)**. Campus license, available at <https://www.scival.com/customer/authenticate/loginfull>

SciVal modules:

- **Overview**: to obtain a high-level overview of the research performance of UoG, other Institutions, Countries and Groups of Researchers.
- **Benchmarking**: to compare and benchmark UoG to other Institutions, Researchers and Groups of Researchers using a variety of metrics.
- **Collaboration**: to explore the collaboration network of both UoG and other Institutions.
- **To be added: Trends¹⁵**: to acquire insight into current scientific trends, based on text analyses of publication titles, key words and abstracts; to support research strategies, collaboration opportunities and to scout 'rising stars'.

See also: UCW presentation Shaya Abdolazadeh 20 June 2017.

THEDataPoints: Times Higher Education's performance monitoring and benchmarking platform, built upon 500,000+ 'datapoints' comprising THE's global university database. Allows benchmarking of UoG with 25 selected other Universities. Accessible to limited number of UoG users.

Modules: (<Brief descriptions to be added >)

- Data+
- Reputation

Recent and expected developments

Expansion PURE: with Award Management Module, i.e. an integration of the Application, Award and Project modules.

Expansion SciVal: with a 'Funding Discovery' tool: using sophisticated Artificial Intelligence / text analysis algorithms, the 'profile' of a call from a funding agency is matched with that of a (group of) researcher(s). So far, based on English titles and abstracts only, to be obtained from PURE.

Project proposal: 'Research Analytics Alliantie'; submitted by the University Library (UB) and the department Research & Valorisation (R&V) to (and approved by) the ICT strategy committee (june 8th) and the Board of Deans (june 7th). Aims to optimize the use and availability of research impact metrics by setting up and expanding a UoG network of experts and by developing and supplementing our present tools.

¹⁵ This year, the Trends module is expected to be replaced by a more sophisticated version, 'Topics'.