Can we implement the Leiden Manifesto principles in our daily work with research indicators?

Report from the fifth meeting of the Danish Research Indicator Network (FIN)
January 21st, 2016, at Copenhagen University Library - Frederiksberg

FIN is a Danish network for practitioners working with research indicators, in particular bibliometric indicators. FIN aims to facilitate the exchange of ideas and experiences between peers and to provide inspiration for the daily work with research indicators. FIN is not affiliated with any institution. The participants represent different sectors: university administration, bibliometric researchers, colleges of professional education, other research institutions, central government administration, and private research foundations etc.

At this, fifth FIN meeting, the Leiden Manifesto was on the agenda. The first part of the program included presentations of two research evaluations from the University of Copenhagen and as an illustration the ten principles from the Leiden Manifesto were used to assess the implemented methodologies. It was shown that the manifesto is not straight forward to implement. The second part of the program was organized as group discussions concerning the ten principles. The participants could each attend two discussions on principles nr. 1-5 and two discussions on principles nr. 6-10. The overall question steering the discussions was: Can we implement the Leiden Manifesto principles in our daily work with research indicators? A summary of the group discussions is reported below.

"1) Quantitative evaluation should support qualitative, expert assessment"[1]
Signed up for the discussion: 16
Rapporteur: Lorna Wildgaard

The participants were in agreement that quantitative evaluation supports qualitative evaluation, but also vice versa. Several of the participants have experience with combining qualitative and quantitative methods in research evaluations. The participants gave examples of evaluations of institutions or departments where quantitative indicators were combined with self-evaluations, peer review or site visits from international panels. The latter type of evaluation can be expensive and is often time-consuming. The qualitative part includes costs for panels/peer review, and in the quantitative part, indicator reports are often done by external consultants. The evaluations are used by institutions/departments for strategic purposes.

The combination qualitative and quantitative methods can facilitate transparency and the possibility to control and if necessary criticize the indicators. In addition, areas which are not covered by reliably quantifiable indicators can be included in an evaluation, for example societal impact and public outreach. To do a balanced evaluation, it is important to know the possibilities and limitations of the indicators and to know the purpose of an evaluation. Evaluations can present both qualitative and quantitative analyses but

[1] Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics. Nature, 520(7548), 429-431. doi:10.1038/520429a.

sometimes the users only focus on the indicators. The qualitative results are typically only read if they are very concise.

Quantitative evaluations can support qualitative evaluations but it can also work the other way around. Indicators may describe an area or institution and give inspiration for a qualitative evaluation.

"2) Measure performance against the research missions of the institution, group or researcher"[1]
Signed up for the discussion: 18
Rapporteur: Marianne Gauffriau

During the discussion different perceptions of missions came up.

- Some mission statements are expressed as indicators or include predefined indicators. Examples are a university's contract with the Ministry of Higher Education and Science, key performance indicators from a higher level in the organization, improved performance compared to last year's performance, or indicators which simply can be downloaded "ready-to-use" from a database. The possibility to design indicators for evaluation of these missions is typically limited to small variations of the predefined indicators.
- On the other hand mission statements can also be very general and give little guidance on which indicators would be appropriate to evaluate the mission. It was discussed if bibliometricians could contribute to the mission design phase with advice on how indicators can be used to monitor a mission. Some participants had with success produced data to inform a mission design phase.
- Mission was also perceived in a more informal way. When a user requests a specific indicator, some participants ask for the purpose (or mission) to be able to suggest adequate indicators. To know the purpose often leads to other and more appropriate indicators then those the user originally had requested. The participants also experience that users often do not have a well-defined purpose and the purpose is accordingly defined in the dialogue with the user.
- Finally, the participants often experienced that a mission was not available or could not be formulated. The user just asked for evaluations of everything or requested the traditional, known indicators (for example the h-index). A link to a mission is never made.

"3) Protect excellence in locally relevant research"[1]
Signed up for the discussion: 8
Rapporteur: Jens Peter Andersen

Local specificities, such as language, occur in all research areas, although they may be more dominant in some parts of the humanities and social sciences. Areas which have a close connection with practice, e.g. through trade journals, typically communicate more in local language. While such publications are not typically included in bibliometric assessment, they may play a very important role locally. Examples could be public health or education.

Other research areas have limitations which obstruct traditional measurement, such as research on international, military operations, as publications may be classified.

Some research areas use very specific publishing channels, e.g. art, books, conferences.

We should strive to illustrate, e.g. visually, what we can measure and what we cannot measure to show what scientometric assessment can inform about. Ideally this is done through some form of figure relating specifically to the analysed unit, e.g. institute.

It is seen as a problem, that no proper database covers the humanities.

National expert rankings of publication channels, such as the Norwegian and Danish systems (in Denmark Bibliometric Research Indicator: BFI[2]) may to some extent protect locally relevant research, if this is included in the design.

"4) Keep data collection and analytical processes open, transparent and simple"[1]
Signed up for the discussion: 18
Rapporteur: Karen Skytte Larsen

Principle 4 is viewed upon as being a rather complicated principle to follow in practice. Especially since the terms "open, transparent and simple" can be interpreted in many ways: "What I find open, transparent and simple may not be what you find is open, transparent and simple". When doing an evaluation or analysis it needs to be defined how the terms are perceived, in order to be on a common ground. Although it is difficult, it was generally agreed upon amongst the workshop participants, that when using metrics to evaluate research we should always strive for achieving principle 4, but the level of achievement depends on the task and the situation.

Simplicity is difficult to obtain when working with data. Data is not necessarily simple, since it in many cases consists of many calculations, before the data actually come into use. As an example of an indicator which at first seems quite easy to comprehend is the h-index (and thus popular), but the calculations behind the h-index may not be quite as simple, and easy to understand for all users.

When we use data from Elsevier or Thomson Reuters it is at times unclear to us how data have been processed. Therefore already in the early stages of analyzing data we lose insight into the nature of data. And although we might be able to investigate it further, it is a very time-consuming task. Also data sets aren't static, and can change while we are doing our analyses. While working with data, which originate from external sources, can be a problem; it can also be an advantage. It was mentioned that using the name of the data set could even at times be a sign of quality, e.g. Leiden data.

One way of complying with principle 4 in practice is applying information on data and its caveats to analyses. This is done as a rule according to the participants at the workshop both when doing longer term projects, but also on delivery of small datasets and figures for management purposes. It seems to be a sort of code of conduct, but it is also in some instances used to cover people's backs.

"5) Allow those evaluated to verify data and analysis"[1]
Signed up for the discussion: 8
Rapporteur: Adrian Price

[2] http://ufm.dk/forskning-og-innovation/statistik-og-analyser/den-bibliometriske-forskningsindikator (in Danish)

All participants were in agreement that openness in data and in analysis and method are extremely important issues. They are important, as the integrity of our work is a major factor which underpins our success or failure and our reputations as bibliometricians.

Often though we are forced to rely on data which originates in other sources, for example in the international bibliometric databases, where there can be issues beyond our control. These issues can concern data quality originating from errors in names, organisations, publication channels etc. as well as errors originating in analyses and methods we have no means whatever of being able to verify. Often the tools we have access to are "black boxes" (for example Web of Science/Incites, Scopus/SciVal, Google Scholar), where often we must rely blindly on what we are served up. Nonetheless, we should be extremely critical in our use of this data and do what we can to verify its correctness, and to document any errors found.

Ethics is also an issue when we produce bibliometric analyses. Openness and verification of data through test and control must be stringently adhered to, and awareness of the fact that data is nearly always political.

"6) Account for variation by field in publication and citation practices"[1]
Signed up for the discussion: 23
Rapporteur: Marianne Gauffriau

This principle was the most popular in terms of how many had signed up for the discussion. The participants looked positively at the principle's implementation potential and gave several examples of how this could be and is already done. Three strategies were discussed: 1) To show and explain the difference between research fields. Based on this information, the users of an evaluation can interpret the results. 2) To work with field normalized indicators which make comparisons across research fields possible. 3) To present those to be evaluated for a set of indicators and let them choose the indicators which best suit their purpose.

None of the strategies were seen as straight forward. Especially field normalization was discussed. The participants saw field comparable relative numbers as attractive for many purposes but the indicators must be carefully explained to the users of an evaluation, i.e. how they are calculated, the information included and how to interpret the results. Furthermore, Danish research is in general performing above the world average regarding citation impact and thus field normalizations comparing to the world average may have little meaning in a Danish context. Finally, new methods for field normalization are introduced quite often and this may affect the results of an evaluation more than the actual performance of those being analyzed. One example is the number and definition of research fields debated in the bibliometric literature. None of the methods for field normalization seem perfect. Take the case of the Danish Research Indicator (BFI). All research areas are divided in 67 research fields each covered by an expert group who divide publication outlets in two levels, level 1 and 2, where level 2 is the more prestigious. The indicator is used to assign points to publications published in these outlets, level 2 attracting more points, and distribute a share of the basic funds to the Danish Universities based on this point system. When the system was introduced, it

was suggested to have 35 research fields and expert groups. The current structure with the 67 research fields is often criticized but no obvious alternative has been found.

"7) Base assessment of individual researchers on a qualitative judgment of their portfolio"[1]
Signed up for the discussion: 11
Rapporteur: Karen Skytte Larsen

Not many have experience with an actual qualitative judgment of portfolios. But several of the workshop participants have worked with broader impact, and the inclusion of several research outputs as opposed to only using publication activity and their impact in research evaluation.

Most participants at the workshop did not in practice regard principle 7 as a simple choice of either the use of the h-index, or the use of a broad qualitative judgment of researcher's portfolio. Some mentioned that they viewed principle 7 as related to principle 1: Quantitative evaluation should support qualitative, expert assessment.

For funding purposes researchers are often asked to give their h-index. It is usually not optional, and therefore not possible to avoid. Although many participants at the workshop where interested in supplying funders with a broader view of researchers portfolio, the system seems quite rigid. This is not always the case at research institutions. It seems that at some institutions there is focus on a broader perspective of researchers' merits. Workshop participants explained that they had experienced this to happen more often when doing internal evaluations at their institutions, than when they participate in evaluations done externally. Quite often though it is seen as a task of managers or PI's to define what kind of output will grant merit. It depends very much on the research area you do your research in; it seems more common within the applied sciences and the practice-based sciences to apply a broader perspective on research output.

It was mentioned that it might also depend on the use of a CRIS (Current Research Information Systems) system, which in some cases are designed to include information on a wide variety of research outputs, e.g. research activities such as conducting reviews or participating at conferences or workshops, or perhaps even making it possible to submit videos or performances. Such systems make it easier to implement the use of a broader perspective of research outputs for the researchers or institutions as a whole.

It was also mentioned that this trend demands analysts, who are able to detect and do qualitative judgments, or be able to employ someone who have experience with doing analyses on the basis of different types of research output. This might lead to bringing in new competencies in research evaluation.

"8) Avoid misplaced concreteness and false precision"[1]
Signed up for the discussion: 5
Rapporteur: Adrian Price

The pressure is often on when researchers and research administrators ask for the evaluation of research (and researchers) through the use of single indicators. All were in agreement that this was unacceptable practice, and some resolved the issue by always delivering several indicators and attaching explanations and words of warning in delivered reports, even though these were not asked for.

The group discussed at some length how important it also is to really think through all extrapolations from data, to be sure one does not fall into any traps, which might not be immediately visible or which have not been properly thought through. Two concrete examples were given. Claims that a (whole) faculty had substantive cooperation with, say, a specific individual country, where in reality this cooperation was through one department with large authorships with the same people, a fact which was actually enough to skew conclusions. Another example was when claims of an apparently substantial percent coverage were made in a report, which in reality was extrapolated from data based on small samples.

Discretion and awareness of factors behind the data are vital factors when analyzing and presenting bibliometric data.

"9) Recognize the systemic effects of assessment and indicators"[1]
Signed up for the discussion: 16
Rapporteur: Jens Peter Andersen

The main concern is that focus on specific activities will influence the way researchers think about these activities, e.g. by changing publication strategy towards smaller units (salami-slicing) However, there is also some concern that those activities which are not measured are not merited, e.g. editorial roles.

Governmental strategic funding influences all researchers, e.g. in the case of Denmark as a pressure to belong to the world elite. Danish experience shows that the closer the assessment gets to the researcher, the larger the effect. The example is the national research assessment (BFI), which allocates some of the basic funding to the Danish universities. However, one of these universities has adopted the model to allocate funding locally as well, which has been felt intensely by the faculty. We recommend local assessment to qualify/diversify national assessments, taking local conditions into account.

When seeking external funding, researchers will try to meet the criteria required by funding agencies, which may influence the use of e.g. the h-index or journal impact factor.

"10) Scrutinize indicators regularly and update them"[1]
Signed up for the discussion: 10
Rapporteur: Lorna Wildgaard

The participants are aware that indicators are improved/changed and do update their indicators and supplement with new indicators. Examples of areas for new indicators which can supplement the more

traditional bibliometric indicators are press cuttings and media appearances, research projects and collaborations, altmetrics, publication language and editors of journals or books.

In the discussion about updating of existing indicators, a lot of questions were raised. When an indicator is changed, evolutionary comparisons based on former evaluations is difficult, and sometimes the new version of the indicator is not even available for historic data. This can have an unfounded impact on the institutions under evaluation and it takes time to explain these factors to the users. An example is university rankings which change methodology and as a consequence the ranking order changes. The participants also discussed the importance of high quality and updated data. Flawed data will lead to unclear evaluations. The Danish Research Indicator (BFI) system is an example where the data quality is debated.

The implementation of new indicators and the updating of existing indicators/data require time and other resources, typically financial, which are not always available. Sometimes the updating of indicators is done on the fly without a formal decision or supporting documentation. An update can be initiated because a new database offers new possibilities, a new type of task must be evaluated, the needs of the users change or new users have new needs.

Rapporteurs

- Lorna Wildgaard, Royal School of Library and Information Science, University of Copenhagen
- Jens Peter Andersen, Danish Centre for Studies in Research and Research Policy, Aarhus University
- Karen Skytte Larsen, Danish Agency for Science, Technology and Innovation
- Adrian Price, Copenhagen University Library – Frederiksberg
- Marianne Gauffriau, Faculty Library of Natural and Health Sciences, Copenhagen University Library, The Royal Library

Participants

Name	Organization
Adrian Price	University of Copenhagen - SCIENCE
Anders Agerbæk Kjøller Nielsen	Novo Nordisk Foundation
Anita Elleby	Danish Defence Library Center
Anne Mette Navntoft	Aarhus University Library, Foulum
Anne Thorst Melbye	University of Southern Denmark, Library
Annette Balle Sørensen	State and University Library
Asger Væring Larsen	University of Southern Denmark
Bertil F. Dorch	University of Southern Denmark
Birger Erichsen	Novo Nordisk Foundation
Birgit Rønne	National Museum of Denmark
Christina Steensboe	Technical University of Denmark
Claus Rosenkrantz Hansen	Copenhagen Business School
Claus Vesterager Pedersen	Roskilde University Library
Dicte Madsen	Faculty Library of Natural and Health Sciences, Copenhagen University Library
Gert Vilhelm Balling	Novo Nordisk Foundation

Hans Emborg Bünemann	Faculty of Humanities, Univ. of Copenhagen
Helen Korsgaard	University of Southern Denmark
Jens Peter Andersen	Aarhus University
Karen Skytte Larsen	Danish Agency for Science, Technology and Innovation
Karen Hytteballe Ibanez	Roskilde University Library
Kasper Bruun	Royal School of Library and Information Science, University of Copenhagen
Katrine Astrup Jacobsen	University of Southern Denmark, Library
Katrine Düring Davidsen	Aarhus University Library, Campus Emdrup
Lene Kaad	Metropolitan University College
Lone Bredahl Jensen	University of Southern Denmark
Lorna Wildgaard	Royal School of Library and Information Science, University of Copenhagen
Lotte Faurbæk	Danish Agency for Science, Technology and Innovation
Maj-Britt Rosenkilde	Aalborg University Library
Marianne Nielsen	Medical Library, Aalborg University Hospital
Marianne Gauffriau	University of Copenhagen / Copenhagen University Library
Mette Fjeldhagen	Technical University of Denmark
Pernille Grøngaard	Aarhus University
Peter Olsen	Copenhagen University Library, Frederiksberg
Poul Melchiorsen	Aalborg University Library
Rikke Christensen	Novo Nordisk Foundation
Rune Kurdahl	Copenhagen University Library / Royal Library
Siri Falkenstrøm	Copenhagen University Library, Frederiksberg
Suzanne Løje	University of Copenhagen, faculty of Humanities
Svend Kragballe	University of Copenhagen
Thea Drachen	University of Southern Denmark, Library
Trine Azbi	Metropolitan University College
Ulla Jakobsen	Lundbeck Foundation