

Pilotní ověření návrhu nové metodiky hodnocení výzkumných organizací

Pilot Test of New Evaluation Methodology of Research Organisations

**Samostatný doplňující dokument 4:
Úpravy bibliometrické zprávy**

***Background document 4:
Adjustments to the Bibliometric
Report***

Tým KA1 IPN Metodika
Martin Lhoták
Pavel Mika
Jakuba Szarzec

List of content

1. Differences in comparison to the Technopolis proposal.....	3
1.1 Labels.....	3
1.2 Format.....	4
1.3 Calculations.....	4
1.4 Bibliometrics for excellent outputs.....	4
1.5 Summary overviews.....	5
2. Problems occurring during the compilation of the bibliometric report.....	6
2.1 Registration of research unit – unclassified scientists.....	6
2.2 Mapping the records in IS R&D&I /RIV-WoS.....	6
2.3 Processing / calculating indicators and processing bibliometric reports.....	6
2.4 Distribution of data and formatting the reports.....	7
2.5 Recommendations for the bibliometric report.....	7

Tento dokument byl zpracován v rámci Individuálního projektu národního pro oblast terciárního vzdělávání, výzkumu a vývoje „Efektivní systém hodnocení a financování výzkumu, vývoje a inovací, CZ.1.07/4.1.00/33.0003“. Projekt byl realizován Ministerstvem školství, mládeže a tělovýchovy a financován prostřednictvím Operačního programu Vzdělávání pro konkurenceschopnost z Evropského sociálního fondu a státního rozpočtu České republiky.

This document has been prepared as a part of the Individual National Project for the area of Tertiary Education, Research and Development „Effective System of Research Financing, Development and Innovation, CZ.1.07/4.1.00/33.0003“. The project was realised by the Ministry of Education, Youth and Sports and financed by the Operational Programme Education for Competitiveness of the European Social Fund and state budget of the Czech Republic.

Dokument „Adjustments to the Bibliometric Report“ neprošel jazykovou korekturou.

BIBLIOMETRIC REPORT

1. Differences in comparison to the Technopolis proposal

1.1 LABELS

In comparison to the proposal provided by Technopolis, one major change was the addition of several comments providing explanations that could prevent incorrect interpretations. Further, the years of the outputs were included into all the graphs and tables (due to the fact that citation indicators were calculated only in 2009-2012).

Specifically, comments were added (in the introduction and then with individual indicators), as follows.

It is important to highlight here, that publication could be classified to more than one Web of Science fields (max 6) and those multi-fields publications are counted once in each field in the field divided indicators (such as A1, A2, A4 - A9, B1).

RD&I IS classification doesn't exactly fit the OECD classification. There are some OECD fields to which are not coupled any RD&I IS fields.

- Not coupled OECD fields:
 - 1.7 Other natural sciences
 - 2.8 Environmental biotechnology
 - 2.10 Nano-technology
 - 6.5 Other humanities
- A2 -Field "Other natural sciences" is not coupled with any RD&I IS classification field (described in the section 1.1.6).
- A4/B2 -Publication could be counted more than once in different fields (described in the section 1.1.6).
- A7 - High and oscillating values in the "Other natural sciences" field are caused by occasionally appearing highly cited articles from journals Nature and Science which are classified to this field.
- F2/G2 - Calculated using number of publications in field, which is different in this instance to F1/G1.
- F3 – note re Jimp: It is not possible to compare the number of Jimp with a number of WoS publication (from indicator B2) due to the different publication-type classification in the two data sources (WoS, RD&I IS) and due to the way how WoS publications of certain RU were identified.
- F4 –note re column with proportion: Calculated using number of publications in field, which is different in this instance to F1.

1.2 FORMAT

Further changes were made to the format of the individual tables and graphs.

In some cases, a graph was changed into a table, also, additional columns were added. Specifically, this concerns the changes stated below:

- A6/C4 - shown in table
- B2 - shown in table with numbers for RU and EvU
- D2 –shown in table including numbers
- A1, A2 – “Bases” from which the percentages were calculated were added, this Base cannot be considered a base for the individual columns in the case of WoS, due to the multidisciplinary issue.
- A9 – the proportion of national cooperation also added
- A10 – the citation impact is just a summary of all years, due to a small number of outputs within individual co-operations, decreasing it further by dividing it into years makes no sense
- B1 – numbers added alongside percentages
- B2 – percentages and numbers
- D2 – shown in table with numbers
- F1 – just one summarising table called “Research Outputs” (is not divided into scholarly/ non-traditional/non-scholarly)
- F2/G2 – a “Number for the RU” column added
- F3/4 – an “Article in a periodical, which is not in any previous group (Jother)” row added
- G1 – a number for all the years added

1.3 CALCULATIONS

Some indicators, compared to the small pilot evaluation, were calculated in a different way. The whole process of calculation and these differences were discussed during the KA1 and bibliometrics team of the provider’s consortium meetings. Changes were made in the following indicators:

- A9 – the proportion of national cooperation was also calculated due to publically unavailable data from InfoScience (national cooperation was added in the case of united outputs).
- C4 – In the case of outputs added to several fields, the highest values were not chosen, but a percentage average of the rating was calculated, determining this quartile from this average.
- D1 – When calculating the citation impact, average values for WoS fields of the given output were correctly used in the denominator, as opposed to SPE, where the RU fields were used.

1.4 BIBLIOMETRICS FOR EXCELLENT OUTPUTS

Apart from bibliometric reports, bibliometric overviews of excellent outputs which were chosen for peer-review evaluation by the referees were also prepared. These overviews contained five bibliometric indicators:

- Number of citations;
- Category expected citation – the average number of citations of documents of the same type, the year they were published in and the field (in the case of a multidisciplinary document, the average value was used);
- Citation percentile – percentile rating based on number of citations in documents of the same type, year published and field (in multidisciplinary cases, the best, i.e. lowest value is taken into account);
- Journal impact factor – average citation of the journal;
- Journal percentile – percentile rating of a journal based on Journal IF within the same publishing year and field (in multidisciplinary documents, the average value is taken into account).

These indicators and how they are calculated was determined based on consultations with the bibliometric team of the provider's consortium, members of KA4 team and also given the availability of the indicators.

1.5 SUMMARY OVERVIEWS

Summary overviews containing data based on five chosen indicators, C1, F4, C4, D1 and D2, were created for panel members' use.

2. Problems occurring during the compilation of the bibliometric report

2.1 REGISTRATION OF RESEARCH UNIT – UNCLASSIFIED SCIENTISTS

Out of 3782 scientists (listed as local science makers in the RIV register), 658 remained unclassified in any of the Research Units RUs. In some cases, the reason for the lack of classification was given (member of a different institution, former worker), however, the question remains why they were listed as local authors.

The scientists who were not classified in any of the RUs were authors or co-authors of 3984 outputs, out of which 290 outputs were not included in the evaluation process at all (bibliometric overviews). 54 of the unclassified outputs had a WoS identifier in the record, so it should have been easy to allocate them in the WoS database, moreover, certain outputs had significant citation.

2.2 MAPPING THE RECORDS IN IS R&D&I /RIV-WOS

This process is key when preparing the data. In order to achieve greater efficiency and mainly credibility, it would be useful to use a higher degree of control. Also, given the ethics of evaluation, the participants should have the option to check and confirm the data that their evaluations will be based on.

Part of the mapping process would be the verification of accuracy of this data. The verification would be made by the institutional participants of the evaluation, as well as the teams preparing it. The Pilot Testing showed that data on publications in the RIV is incomplete. The Document Object Identifier (DOI), which is a term used for electronic documents, can serve as an example. Entering these into the RIV is voluntary.

This made checking the data between RIV and WoS particularly difficult. An obligatory DOI in the case of outputs would simplify the mapping process, checking data accuracy and further processing of bibliographical and statistical data. If the DOI is stated with the publication, then the RIV should automatically appear as obligatory information. This is going to simplify finding the links between citation databases.

Furthermore, incorrect journal name registrations were found, resulting in a failure to verify the information on outputs. This could be resolved by creating an internal database (list) of resources that would automatically verify the names of sources. In some cases, the types of documents were incorrectly classified. At the same time, the RIV lacks the differentiation of printed and electronic versions of journals using ISSN.

These errors may be caused by the fact that the evaluation methodology changes every year.

2.3 PROCESSING / CALCULATING INDICATORS AND PROCESSING BIBLIOMETRIC REPORTS

Several errors occurred in the calculations during the data processing and filling in the templates of the individual reports. These were not system errors in data or algorithm calculations, but, in most cases, these were caused by human error due to the information load and new ways of processing the data. Most errors were identified when the report was published – together with the corrections, some additional comments and labels of tables and graphs were also included.

2.4 DISTRIBUTION OF DATA AND FORMATTING THE REPORTS

For practical reasons, it was necessary to make the process of redistributing information in the tables and graphs more efficient, i.e. change it into text form. Filling in information into the templates manually proved to be time consuming, error prone and made further formatting changes more difficult.

2.5 RECOMMENDATIONS FOR THE BIBLIOMETRIC REPORT

1. The bibliometric report based on the proposal of Technopolis, including the changes as proposed above, option to simplify the report (recommendation of panel members).
2. CZ indicators: to be stated for both the field in which the research unit RU is registered and the average value throughout disciplines based on the publication profile of the RU.
3. Presentation of key indicators of the bibliometric report also in form of an overview table, throughout the RUs.
4. Publications of authors from several RUs should be accounted for multiply.
5. Reflect on evaluation of social sciences and humanities outputs more deeply.
6. Expand the RIV, Czech information system for R&D by book reviews.
7. Resolve the question of taking non-scholarly outputs of applied research into account (already done in pilot testing).
8. Bibliometric reports will be generated automatically by the system and available on-line, on a website used for the evaluation.

Pilotní ověření návrhu nové metodiky hodnocení výzkumných organizací
Samostatný doplňující dokument 4
Úpravy bibliometrické zprávy

Vydává Ministerstvo školství, mládeže a tělovýchovy, Karmelitská 7, Praha 1
Individuální projekt národní pro oblast terciárního vzdělávání, výzkumu a vývoje:
Efektivní systém hodnocení a financování výzkumu, vývoje a inovací (IPN Metodika)
www.metodika.reformy-msmt.cz

Praha 2015