Open Access Indicator for 2013

Part 1

Overview of data foundation, processes and output

Version 1, 15 March 2016

0   Contents

1 Preface

The National Steering Group for Open Access[1] has proposed the Danish Agency for Science, Technology and Innovation and Denmark's Electronic Research Library, to develop a Danish Open Access Indicator. The intention is to support the implementation of the national Open Access strategy[2] - cf. the strategy's statement on monitoring: "*The implementation of Open Access is to be monitored on an ongoing basis to ensure that all parties make a maximum effort to develop and disseminate free accessibility to Danish research findings.*"

The Open Access Indicator is calculated once per year with the target field: *Scientific and peer reviewed articles and conference contributions in journals and proceedings with ISSN.*

In the context of Horizon 2020[3], EU requires that Open Access be established within at most 6 months after publication for the areas of science, technology and health and within at most 12 months for the social sciences and humanities. This delay is caused by many journals maintaining so-called embargo periods, where they exclude researchers from establishing Open Access to the articles before the end of the embargo period.

As the OA Indicator is calculated once annually for all publications within its target field, it is designed to accept a one-year delay in Open Access to the publications. The OA Indicator for 2013 is calculated early January 2016 at the same time as the OA Indicator for 2014, so the indicator from the start could provide a view of the development over time. In practice this means that publications from January 2013 could have embargo periods all the way up to 36 months and still be credited by the OA Indicator.

The description of the Open Access Indicator is organized in two parts:

- Part 1: Overview of data foundation, processes and output
- Part 2: Technical description of data foundation, processes and output

    Note: In Part 2, the technical description, the notion of the indicator's "target field" is expressed using the term "set of scoped records".

Queries regarding the indicator may be directed to

> Jonas Bak/Hanne-Louise Kirkegaard
> 6th Division: Research Policy
> Danish Agency for Science, Technology and Innovation
> Bredgade 40
> DK-1260 København K
> Email: jonb@fi.dk/ hki@fi.dk

---

[1] http://ufm.dk/en/research-and-innovation/cooperation-between-research-and-innovation/open-access
[2] http://ufm.dk/en/research-and-innovation/cooperation-between-research-and-innovation/open-access/Publications/denmarks-national-strategy-for-open-access
[3] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

2 Overview

2.1 Which types of Open Access are measured?

The Open Access Indicator for 2013 measures two types of Open Access:

1. Articles and conference contributions, published in dedicated Open Access journals and thus available in Open Access as soon as they are published (Golden OA)

2. Articles and conference contributions, which are available in Open Access from the universities' own research databases or Open Access repositories (Green OA)[4]

The Open Access Indicator for 2015 will include an additional type:

3. Articles and conference contributions, which are available in Open Access from other recognized Open Access repositories (Green OA). The introduction of this OA type has awaited an update of the universities' local research database systems, which provide the indicator with metadata on publications.

The OA Indicator does not measure so-called Hybrid OA, where an article is published in a journal only accessible to subscribers, and where an additional fee is paid to release the individual article in OA to non-subscribers. However, if such hybrid OA articles are made available from the universities' research databases or Open Access repositories, they will be credited just like all other cases belonging to type 2 above.

2.2 What is the data foundation?

- Metadata about the year's publications are collected from the research databases of each of the 8 Danish universities. This constitutes the basic data of the OA Indicator.
- The result of the year's Bibliometric Research Indicator[5] is imported. This enables the OA Indicator to handle duplicates caused by two or more universities collaborating in writing a publication, which is consequently collected from more than one research database.
- International metadata for dedicated golden Open Access journals is imported from the Directory of Open Access Journals[6]. Thus the OA Indicator may handle Open Access type 1 above (Golden OA)
- International metadata about the publishers' and journals' policies wrt. Green OA is imported from the Sherpa/Romeo[7] database. Thus the OA Indicator may establish the Open Access potential of articles in subscription journals. I.e. how may articles could become Open Access via the universities' research databases and OA repositories without breaking publisher policies?
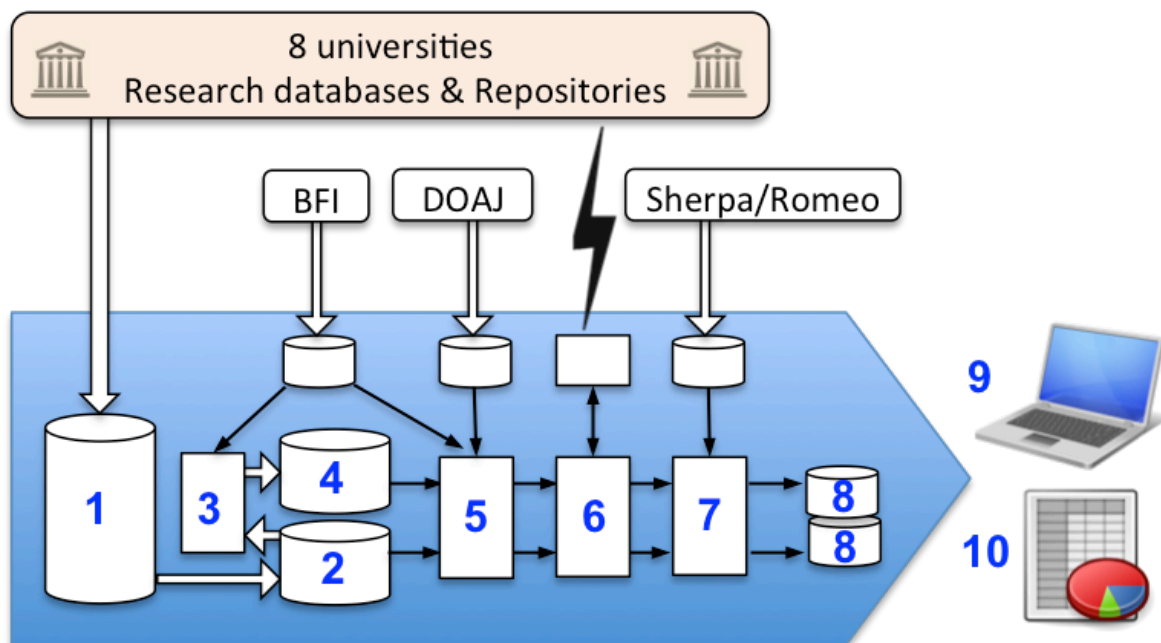
---

[4] Danish research databases (Current Research Information Systems) are primarily metadata registries of the university's publications, but they may also perform the task of being "repositories" for the publications in full text. In other cases, the universities use other dedicated systems as Open Access repositories.
[5] http://ufm.dk/forskning-og-innovation/statistik-og-analyser/den-bibliometriske-forskningsindikator
[6] https://doaj.org
[7] http://www.sherpa.ac.uk/romeo/

2.3 How does the process flow – in brief?

1. Publication metadata is collected automatically from the universities.

2. A subset corresponding to the definition of the OA Indicator's target field is isolated and form the "**target field with duplicates**", as publications with authors from more universities will be collected more than once. This version of the target field is used as basis for calculations dealing with individual universities.

3. Deduplication of records from "target field with duplicates" using data from BFI (the Bibliometric Research Indicator).

4. This results in "**target field without duplicates**", which is used as basis for calculations on the national level and by main research area.

5. Check whether the article is published in a dedicated and scientific Gold OA journal. Here BFI-data as well as data from DOAJ (Directory of Open Access Journals) are used to ensure that the journal is scientific and a fully dedicated Gold OA journal.

6. Check whether the article may be downloaded from a university's research database or OA repository (Green Open Access)

7. Check whether the article is published in a journal with Green OA potential. Here data from the Sherpa/Romeo-database is used.

8. This results in a statistical dataset in two parts:
   (1) University statistics calculated using "**target field with duplicates**"
   (2) National statistics calculated using "**target field without duplicates**".

9. The result is communicated via web-pages of the Danish National Research Database and …

10. … via spreadsheets, which may be downloaded from the Danish National Research Database. Additionally, the underlying publication data may be downloaded as spreadsheets.