# Open Access Indicator for 2016

## Part 1

### Overview of data foundation, processes and output

Version 1, 10 April 2018

## 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. Consequently, the OA Indicator for 2016 is calculated early March 2018 in order to accommodate a full year embargo period also for publications from December 2016. In practice this means that publications from January 2016 could have embargo periods all the way up to 24 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

Adam Baden/Hanne-Louise Kirkegaard
Danish Agency for Science and Higher Education
Ministry of Higher Education and Science
Bredgade 40
DK-1260 København K
Email: aba@ufm.dk/ hki@ufm.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   Which types of Open Access are measured?

The Open Access Indicator for 2016 measures three 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]

3. *Articles and conference contributions, which are available in Open Access from other recognized Open Access repositories (Green OA).*

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 other recognized Open Access repositories, they will be credited just like all other cases belonging to type 2 or 3 above.

## 3   What is the data foundation?

- Metadata about the year's publications are collected from each of the local research databases of the 8 Danish universities. This constitutes the basic data of the OA Indicator.

- The result of the year's Danish Bibliometric Research Indicator (BFI)[5] is imported, in order to ensure that the OA Indicator may reuse the results from the BFI duplicate- and conflict handling. Duplicates are caused by two or more universities collaborating on a publication, which is consequently collected from more than one research database. Conflicts occur when duplicate registrations from more than one university disagree on details that require unambiguity at national level.

- International metadata for dedicated golden Open Access journals is imported from the Directory of Open Access Journals[6]. This enables the OA Indicator to evaluate Open Access type 1 above (Golden OA)

- A Danish authority list of recognized and compatible Open Access repositories other than the Danish University research databases. This enables the OA Indicator to evaluate Open Access type 3 above (Green OA)

- International metadata about the publishers' and journals' policies wrt. Green OA is imported from the Sherpa/Romeo[7] database. This enables the OA Indicator to establish the Open Access potential of articles in subscription journals. I.e. how many

---

[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/

articles could become Open Access via the universities' research databases and other recognized OA repositories without breaking publisher policies?

- A Danish authority list of journals with very long embargo periods, i.e. more than the 12 months accepted by the EU.

## 3.1 Recognized and compatible Open Access repositories

The OA Indicator measures Green OA from two types of repositories:

**The universities local research databases or OA repositories**
- AU http://vbn.aau.dk/en/
- AU http://pure.au.dk/portal/en/organisations/8000/publications.html
- CBS http://research.cbs.dk
- DTU http://orbit.dtu.dk
- ITU https://pure.itu.dk/portal/
- KU http://research.ku.dk/find-a-researcher/
- RUC https://forskning.ruc.dk
- SDU http://findresearcher.sdu.dk/portal/

**Other recognized and compatible OA repositories**
The following criteria should be met by the repository to be accepted on the authority list of external Open Access repositories:

- **Technical solidity**
  The repository is in stable operation with good up- and response times
- **Sustainability**
  The repository seems organizational and financial sustainable and can be expected to work for many years to come.
- **Credibility**
  The repository is dedicated to Open Access dissemination of scientific publications.
- **Contact**
  The repository has visible contact information and responds to mail inquiries within a reasonable time.
- **Accessibility**
  The repository's publications are freely available without any access barriers in the form of registration requirements or similar.
- **Link quality**
  The repository displays full texts via direct links: If a human or a machine (robot) follows the link (in a browser etc.), the actual text is downloaded directly.

The authority list is updated yearly based on input from the Danish universities. During a hearing period, the universities may suggest new repositories to be added to the list. The suggested repositories are evaluated in accordance with the six criteria and, if these are met, the repositories may be added to the list.

The list is available for download on the OA Indicator website.

## 3.2 Journals with very long embargo periods

The authority list of journals with embargo periods exceeding 12 months is updated annually. Embargo lists are gathered from five major publishers in addition to which the universities may, in a comment period, suggest other journals with very long embargo periods. The suggestions are evaluated, after which they are added to the authority list.

The list is available for download on the OA Indicator website.

## 4  How does the process flow – in brief?
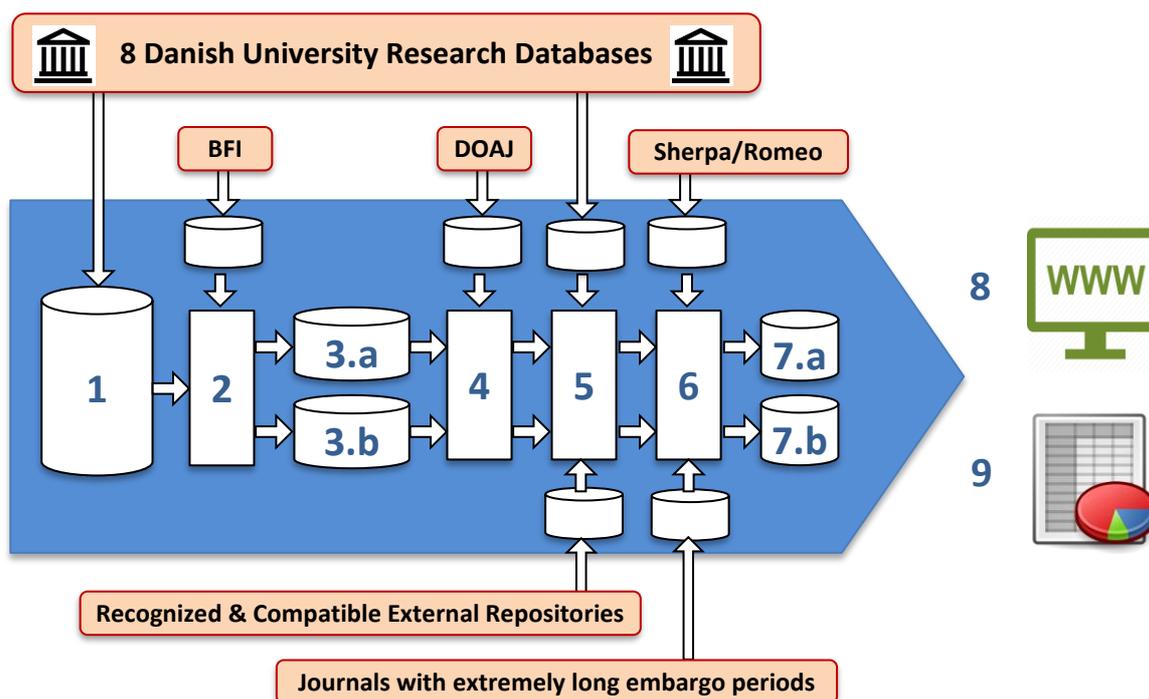


1. Publication metadata is collected automatically from the universities.

2. A subset corresponding to the definition of the OA Indicator's target field is isolated and two versions of the data set for further analysis are produced.

3.a The version "**target field with duplicates**", which will contain duplicates when a given publication has authors from more than one university, as it will have been collected more than once. This version of the target field is used as basis for calculations dealing with individual universities.

3.b The version "**target field without duplicates**", where deduplication of records has been carried out, among others using data from BFI (the Danish Bibliometric Research Indicator). This version is used as basis for calculations on the national level and by main research area.

4. Check whether the article is published in a real Gold OA journal. Here data from DOAJ (Directory of Open Access Journals) is used to ensure that the journal is a fully dedicated Gold OA journal. DOAJ, furthermore, provides information about journals

that charge a publication fee, normally referred to as an APC (Article Processing Charge). This information for the indicator's target field may be found in a downloadable spreadsheet on the OA Indicator website.

5.  Check whether the article may be downloaded from a Green Open Access repository
    - either a Danish university's research database
    - or a recognized external OA repository (on the authority list of such)

6.  Check whether the article is published in a journal with Green OA potential. Here data from the Sherpa/Romeo-database is used. In addition a list of journals with extremely long embargo periods is consulted in order not to claim Green OA potential for those with more than 12 month embargo.

    This results in a statistical dataset in two parts:

7.a University level statistics calculated using "**target field with duplicates**"

7.b National level statistics calculated using "**target field without duplicates**".

8.  The result is communicated via web-pages of the Danish Open Access Indicator and …

9.  … via spreadsheets, which may be downloaded from the Danish Open Access Indicator. Also the underlying publication data may be downloaded as spreadsheets. The OA Indicator for 2016 for the first time offers a downloadable spreadsheet with a subdivision of the realized OA category:
    - Green OA: Publication with OA via a university's local research database
    - Green OA: Publication with OA via a recognized external repository
    - Golden OA: Publication in a journal registered in DOAJ as charging  APCs
    - Golden OA: Publication in a journal registered in DOAJ as not charging  APCs

## 4.1  Quality assurance process

The results of the OA Indicator have been subjected to the following quality assurance measures:

- **Data Foundation.** The collected data has been tested in order to make sure the data foundation has been harvested correctly and in accordance with the established process. The test is based on a random sample, representative across universities.

- **Downloaded fulltext files.** The test focus on downloaded files that appear to deviate from registered metadata; deviations in page number, very small files etc.

- **Random sample.** A random sample of 5% from the total set of realized Open Access potential, from each university, has been inspected with the aim of validating the overall data quality.

**Note**:
Sherpa/Romeo data is used to determine whether a publication that hasn't been OA realized, should be classified as having unused (yellow) or unclear (red) OA potential.
In connection with the quality assurance of the OA Indicator 2016 results, a potential shortcoming has been detected in the downloaded and applied data from Sherpa/Romeo.

It turns out that the online service of the Sherpa/Romeo website provides information on more journals than are covered by the Sherpa/Romeo database, which may be downloaded by the OA Indicator and others. This is caused by the online service integrating supplementary information from two other online databases in the search results (DOAJ and ZETOC).

This discrepancy seems to have the effect that the OA indicator's red category appears a bit larger than it should and the yellow category a bit smaller. The green category (realized OA) is not affected by the discrepancy. Whether this is an important problem and whether it could be remedied, awaits a further dialogue with Sherpa/Romeo.