sciecominfo
Nordic-Baltic Forum for Scientific Communication

DANISH OPEN ACCESS BAROMETER:
Mapping Open Access to Danish research and creation of an online prototype for automated open access monitoring.

Mikael K. Elbæk

Introduction

Monitoring a phenomenon has two remarkable effects: 1. It enables us to understand its properties and interact with the object in an informed way. When it comes to open access this means that we can understand the direction of a development; is it growing, diminishing or simply stagnant. Knowing this is a key component if you want to make strategic decisions for open access; where are we? where are we going? and which measures are working and which are not. The second effect (an interesting fact about social phenomenon such as publishing), is that when something is being monitored it tends to stimulate that which is being monitored[1].

We (the project group behind the Danish Open Access Barometer) believe that open access is important and that if we want to further the progress of open access, monitoring and taking informed decisions and stimulate the progress is one of the key components.

Danish Open Access Barometer was a one year project funded by DEFF (Danish Electronic Research Library). The project started in February 2013 and was concluded in March 2014. The project partners were: Copenhagen University Library (CULIS), Roskilde University Library, University of Southern Denmark Library and the project was led by the Technical Information Center of Denmark, Technical University of Denmark.

Background

Analysis looking at the current state of open access is a growing sub-topic of bibliometrics and webometrics[2]. In our preliminary study we made a review of the current literature[3]. However, it was the Swedish project *OA-publicering vid svenska lärosäten - en kartläggning 2011*[4] that was our biggest inspiration throughout the project, and the two projects did cooperate with each other over the course of the projects. The projects had a similar scope to map open access nationally - and hence had to deal with the local context of systems and processes.

Aim of the project

The Danish Open Access Barometer had two main goals:

1. To map the state of open access to peer reviewed research articles from Danish universities in 2011

2. To create a prototype of a web based and automated Open Access Barometer.

The optimal vision for the project was to create an attractive looking web interface that would monitor open access to the research literature output from Danish research institutions and that this should be done in an automated fashion without or as little human interference as possible. Even though Denmark

[1] Taylor, Winslow Frederick (1913). The Principles of Scientific Management, Harper 144 S.

[2] Examples are research by Bo-Christer Björk, Laakso, M and Solomon, D http://www.openaccesspublishing.org/ and Canadian Science-Metrix: http://www.science-metrix.com/

[3] Review of current literature made by the project: https://infoshare.dtv.dk/twiki/bin/view/OpenAccessBarometer/LitteraturTilInspiration

[4] OA-publicering vid svenska lärosäten - en kartläggning 2011" :http://www.kb.se/openaccess/nyheter/2014/OA-publicering-vid-svenska-larosaten/

has a fairly mature infrastructure of CRIS systems and repositories (all universities and public research institutions at present report their research in the Pure CRIS system), we anticipated that at present it would not be possible to get a complete status of the current open access position in Denmark at the various research institutions. It was foreseen that focussing on these two complimentary projects would generate knowledge about the current possibilities and what must be done to achieve the optimal vision: just concentrating on one of the two aspects mentioned above would not be sufficient.

Method

The main focus for the methodology has been the repeatability of the method. In the methodology presented here we will mainly describe the data sources used and their limitations[5]. All methods have been presented at the project's workshops and sent to all stakeholders identified by the project. All preliminary and final results were also sent to the participating institutions to review and comment on the results.

Why BFI data?

The reason for choosing BFI data was two fold: 1. The data includes peer reviewed articles from all Danish universities and therefore can answer what the state of open access is, being a subset of publications from universities and which are the main object in focus of open access policies. 2. the data is well defined through negotiations between universities and the relevant government ministry, and all universities have incitements for delivering a high level of metadata quality and completeness, as this influences the allocation of money.

Open access definition

The Danish Open Access Barometer's short definition of open access is: "Open Access is research literature that is published on the Internet, either in an open archive and/or in an Open Access journal, in a way that enables public access. "[6]

Data sources

The data sources and the conditions applying to them are summarized on the following page.

Processing the data

Data sources were as far as possible harvested through the APIs of the data sources. The main BFI dataset was however only available by request and delivered on a CD-ROM in Excel (.xls) format by the Agency for Research and Innovation. All data was collected in one CSV-file using scripting. Everything has been documented on the project GitHub[7].

Review / evaluation

Methods, partial and final results were sent to all stakeholders that were monitored in the project. Most importantly the chance was given to review the final result and this increased the total open access percentage from 11 % to 21 %.

The formula on how to review the results was send to all stakeholders and they were given three weeks to provide their feedback.

[5] The complete analysis using Excel is decribed and documented on the project wiki (it can be downloaded under "Dokumenter"): https://infoshare.dtv.dk/twiki/bin/view/OpenAccessBarometer/WebHome#Dokumenter

[6] Open Access Barometer open access definition: https://infoshare.dtv.dk/twiki/bin/view/OpenAccessBarometer/ProjektDelEt#Definition_af_Open_Access

[7] Open Access Barometer GitHub: https://github.com/dtulibrary/oa_barometer

Data import direct from sources	Auto	Manual	Comments
1. BFI dataset from 2012 (2011 publications)	No	Yes	The National Bibliometric Indicator (BFI) dataset has been the main source for the project and other sources are connected to this using unique identifiers like ISSN, DOI, and Pure UUIDs. Provided on a CD-ROM.
2. CRIS (Pure) from Danish universities	Yes	-	Metadata from Pure CRIS' is included indirectly via metadata from BFI and the National Danish Research Database.
3. Institutional Repositories, where Pure is not used for the deposition of full texts and open access information	No	-	Insitutional repositories are not included - because there currently is no simple way to link records in the BFI data set to complimentary datasets and at the same time to maintain the goal of automatic processes. Missing ID's such as ISSN, DOI etc. * As with Institutional repositories subject repositories are not included*
4. Subject specific repositories (Arxiv, PudMed etc.)	No	-	Is included and matched and linked via ISSN
5. Open access journals (via DOAJ)	Yes	-	Is included and matched and linked via ISSN
6. SHERPA/RoMEO	Yes	-	Not possible because of missing metadata*
7. Delayed Open Access	No	-	Not possible because of missing metadata*
8. Hybrid open access - single articles that are paid for open access in a journals that is otherwise toll access.	No	-	(*) Sources that could not be included in the automated process but are directly or indirectly included in the manual review.

Analysis

In this article we present an overview of the results. In the original report we did several breakdowns on the statistics to compare universities, research areas and publication channels. In the follow chart we present three main factors:

- Open Access potential (green)
  - Out of the total number of publications how many could have been open access, either because they are either published in open access journals or because they are published in journals that allow deposit into a repository e.g. the green colour code in SHERPA/RoMEO.
- No-Open Access potential (red)
  - Articles that are not published in open access journals, or no deposit is allowed in a repository.
- Realised Open Access (blue)
  - Shows how much was open access after the review of the results.

The top-performing universities, IT-University and Technical University of Denmark, are both single-faculty universities in technical domains: Computer Science and Engineering. Universities performing less well tend to be universities that are strong in humanities and the social sciences[9].

Open access mapping: an analysis of Danish universities BFI publications from 2011

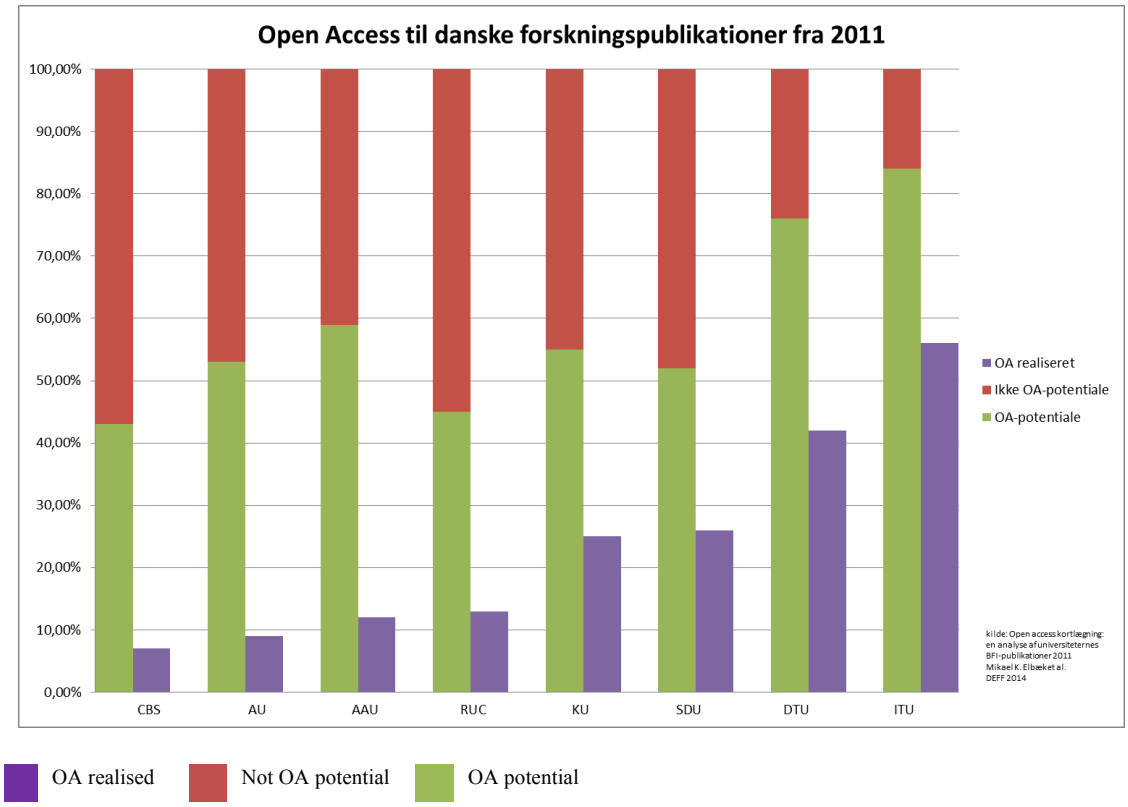


Figure 1: Open Access to Danish peer review articles 2011[8]

[8] Source: Open access-kortlægning: en analyse af universiteternes BFI-publikationer 2011, Mikael K. Elbæk et al., DEFF 2014.

[9] It is important to note that University of Copenhagen only provided a random sample in their review of the result. This may skew the final result.

Number of publications with or without OA-potential and realised OA

Universitet	OA-potential	No OA-potential	OA realised	Total (B+C+D)	Total (actual)
Copenhagen Business School	148	230	27	405	403
Aarhus University	1920	2055	417	4392	4392
Roskilde University	123	210	50	383	383
Aalborg University	586	517	150	1253	1253
University of Copenhagen[9]	29	33	18	80	73
University of Southern Denmark	1506	1966	1094	4566	4134
Technical University of Denmark	891	538	947	2376	2263
IT-University	15	7	25	47	45
Total	5218	5556	2728	13502	12946

Prototype of the Danish Open Access Barometer

The specification of the Danish Open Access Barometer was developed as a working prototype. In the following we will describe the functional requirements the project team identified. This specification is not an exclusive list of requirements but intended as a starting point to showcase some of the possibilities using data and tools readily available. Following the presentation we will present some actual screen shots from the prototype, which is also accessible here: URL: http://unstable.openaccessbarometer.cvt.dk/barometer/frontpage

Username: oab
Password: chee1Kee

It is required that the open access barometer will illustrate:

- Actual open access share compared to total production.
- Open access potential of total production.
- Development of the open access percentage over time.
- Most popular journals and their open access options.

Parameters for analysis that the open access barometer provides:

- Institutions (universities)
- Areas of expertise (The four main research areas of the BFI : sciences, social sciences, humanities and medicine)
- Open Access publication / Not Open Access publication
- Open access types (green vs. Golden OA (blue , green, yellow , white))
- Journals (top 20 of most used journals)

Parameters not currently possible with the current data sources, but which potentially should be possible:

- Publication year, which makes it possible to observe trends over time (the project had only data from 2011)
- Departments / Faculties / etc, which makes it possible to see smaller units such as departments, centres and research groups and their open access performance
- Single author open access performance - e.g. using ORCID identifiers
- Continuous and automated update of the Open Access Barometer.

Out of Scope

The following parameters are omitted:

- Discovery, Data consumption, Person information.
- Other publication types, including non-peer-reviewed research contributions, pre -prints, working papers, student essays, books and book contributions, etc.

OA Barometer definition reflects, on the one hand, the open access demands of the funders and

universities, which require only open access to peer-reviewed scientific articles.

Responsive design and platforms
In the development of Open Access Barometer an emphasis on responsive design has, as far as possible, be incorporated. The Open Access Barometer should be available regardless of platform and format and be adapted to different platforms such as PC , tablet or mobile phones, or Windows , MacOS, Linux , etc.

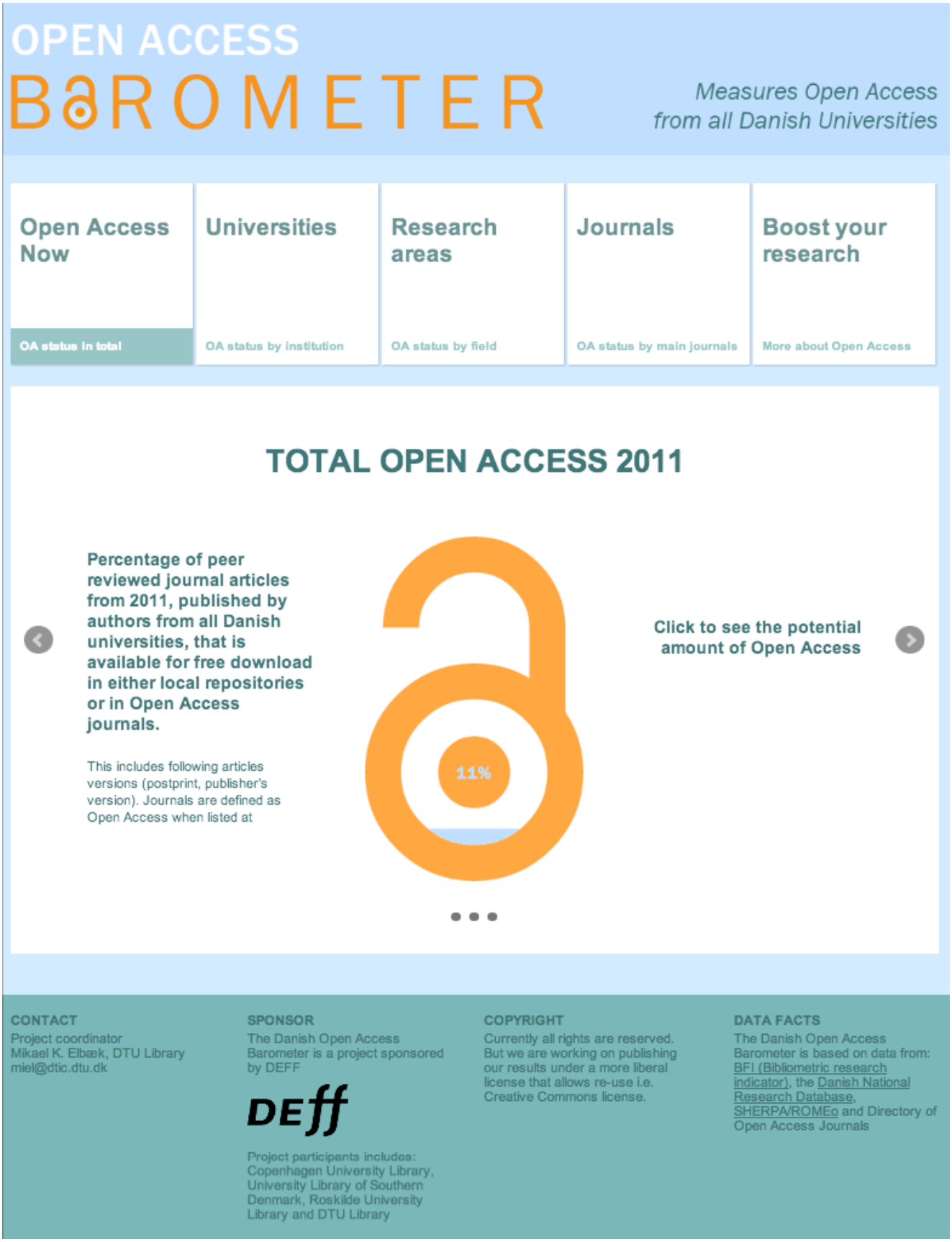
Technology
The emphasis is on maximum transparency and development in a form that matches the open access agenda that the Open Access Barometer is intended to promote.

Therefore, the Open Access Barometer is developed on an open source platform, and data and code is accessible to the public and reusable under a non-restrictive license.

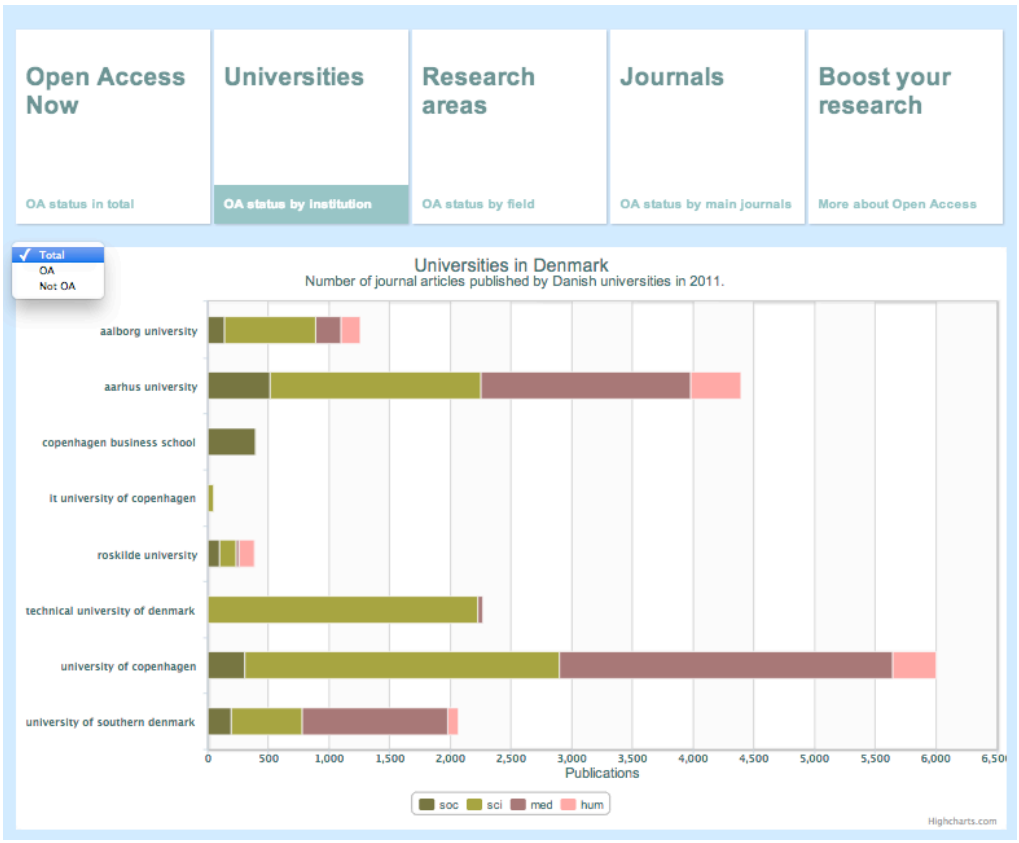
The entire Open Access Barometer database, data processing and web interface is documented in an open GitHub :
https://github.com/dtulibrary/oa_barometer

To illustrate the Open Access barometer, screenshots from the main areas of the prototype are shown on the following pages:

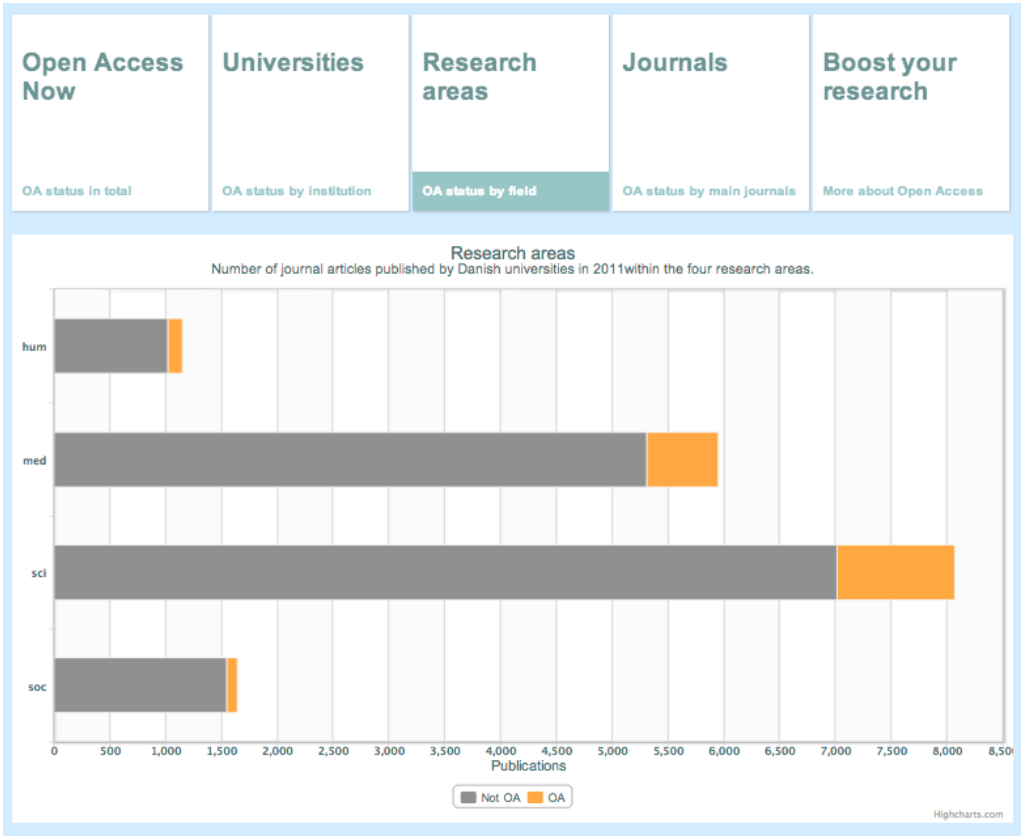
Frontpage


OPEN ACCESS
BAROMETER
Measures Open Access from all Danish Universities
Open Access Now
Universities
Research areas
Journals
Boost your research
OA status in total
OA status by institution
OA status by field
OA status by main journals
More about Open Access
TOTAL OPEN ACCESS 2011
Percentage of peer reviewed journal articles from 2011, published by authors from all Danish universities, that is available for free download in either local repositories or in Open Access journals.
This includes following articles versions (postprint, publisher's version). Journals are defined as Open Access when listed at
11%
Click to see the potential amount of Open Access
CONTACT
Project coordinator
Mikael K. Elbæk, DTU Library
miel@dtic.dtu.dk
SPONSOR
The Danish Open Access Barometer is a project sponsored by DEFF
DEFF
Project participants includes: Copenhagen University Library, University Library of Southern Denmark, Roskilde University Library and DTU Library
COPYRIGHT
Currently all rights are reserved. But we are working on publishing our results under a more liberal license that allows re-use i.e. Creative Commons license.
DATA FACTS
The Danish Open Access Barometer is based on data from: BFI (Bibliometric research indicator), the Danish National Research Database, SHERPA/ROMEo and Directory of Open Access Journals

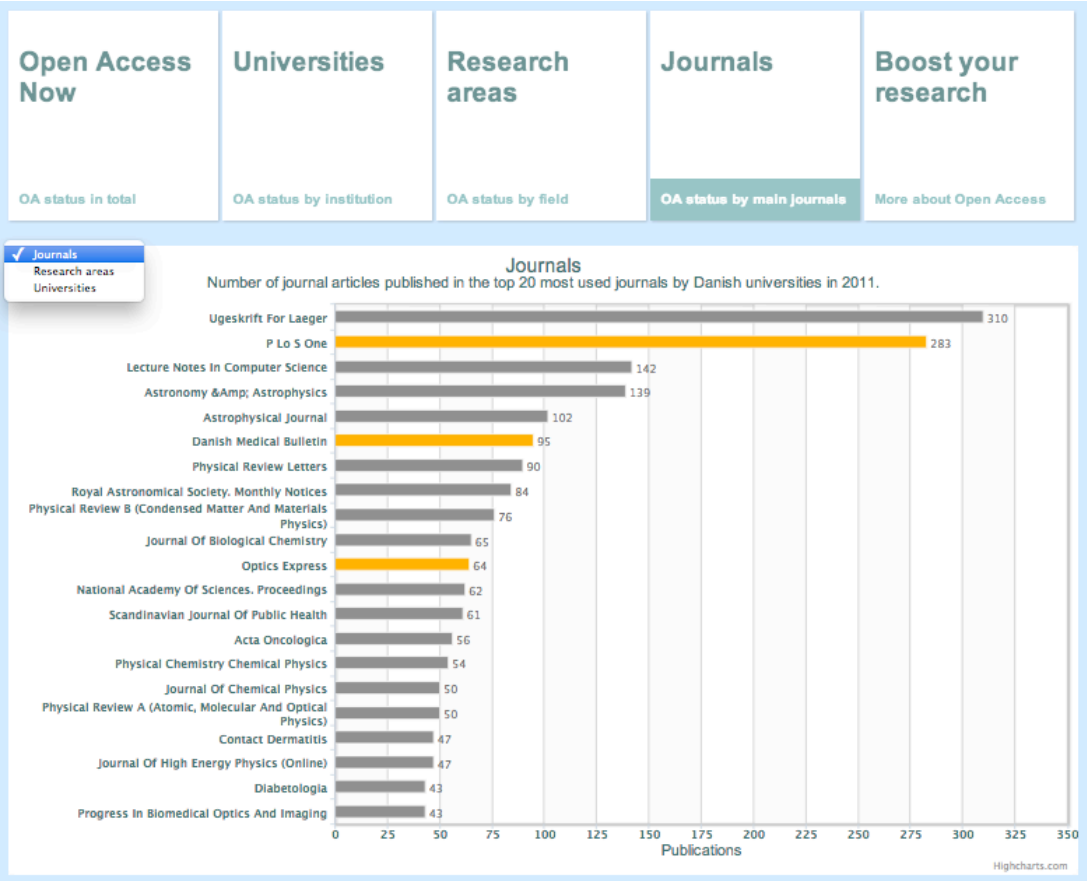
Universities



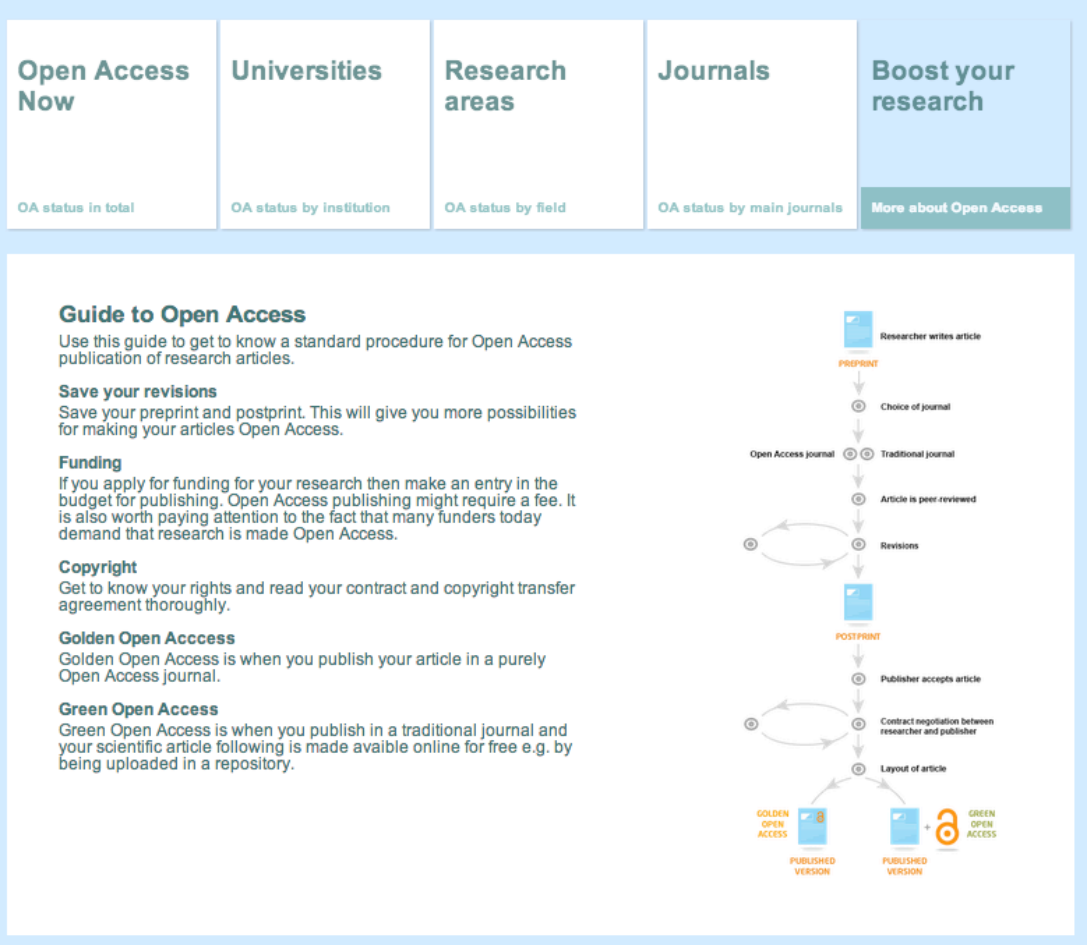
Research areas



Top Journals



Boost your research



Perspectives and concluding remarks

The results of this projects show that open access numbers for Denmark are very much in alignment with the numbers found in our neighbouring countries.

Something more exciting in the eyes of the author is that there is already great potential in making visualisations of open access performance, as demonstrated with the prototype presented here. However, in the current setting there is a need for further improvements, in particular of metadata which can be obtained from the Pure repository system and uniform registration practices at the universities.

In regards to link research funding to publications and thereby the ability to be able to measure the grant holders' ability to live up to their open access requirements, there is also a need for the research funders to start giving their grants unique ID's and even better, to start sharing metadata in a formalised way about grants and projects[10].

With regards to Pure metadata there is a need to provide metadata about open access and open access licenses such as Creative Commons, in order to derive the open access status of publication records without necessarily having the full text in the Pure CRIS.

In the DEFF project "Forskningsdokumentation og kommunikation" there is a task to redevelop the Danish National Research Database, which includes the implementation of elements that have been identified within this project.

Literature

Elbæk, MK (2013). Open Access policies and the supporting infrastructure: status in Denmark, ScieCom Info 9 (2)

Elbæk, M. K., Hansen, T. J., Larsen, A. V., Melbye, A. T., Schelde, S. L., & Dorch, S. B. F. (2014). *Open Access kortlægning 2011: en analyse af universiteternes BFI-publikationer 2011*. Danmarks Forskningsbiblioteksforening.

Fathli, M., Lundén, T., & Sjögårde, P. (2014). Open Access-publicering vid svenska lärosäten: en kartläggning av året 2011.

Taylor, Winslow Frederick (1913). The Principles of Scientific Management, Harper 144 S.

Mikael K. Elbæk, Technical Information Center of Denmark, Technical University of Denmark

[10] As propsed by the author in: Open Access policies and the supporting infrastructure: status in Denmark MK Elbæk ScieCom Info 9 (2)