Open access monitoring in Finland

AT2OA workshop, Vienna, April 9, 2018
Jyrki Ilva <jyrki.ilva@helsinki.fi>
Background: Open Science and Research (1)

  - A national project funded and coordinated by the Ministry of Education and Culture
  - “The objective is for Finland to become one of the leading countries in openness of science and research by the year 2017 and to ensure that the possibilities of open science will be widely utilized in our society”
Background: Open Science and Research (2)

- A top-down initiative, with many stakeholders involved
  - Successful in raising awareness both among the researchers and decision makers
  - Led to the creation of new open science policies at many Finnish research organizations
  - Provided funding for several projects and services
  - Main focus of the initiative was on research data
- There were ambitious national targets set for open access as well: 65% OA in 2017, 75% in 2018 and 100% in 2020
National data collection

- The Finnish OA monitoring efforts are integrated with the national data collection conducted for the Ministry of Education and Culture
  - The Ministry has been collecting publication data from the universities since 2011
  - The universities of applied sciences joined in 2012, central hospital districts and some of the state research institutes in 2015
- The universities have a lot of motivation to provide quality information
  - 13% of the state funding to the universities is distributed based on the number and quality of the publications (= 200 million euros a year)
Open access and state funding?

- Would it be possible to use the open access availability of research output as a factor in the funding model for the universities?
  - Possibly starting from 2019, as the funding is based on the publication data from the previous three years
  - The committee responsible for the university funding model has discussed this, but no decision has been made
Collecting the publication data

- CSC - IT Center for Science is responsible for the actual data collection
  - Close collaboration with the research organizations, the Ministry, the Federation of Learned Societies and the National Library
- The publication metadata is usually collected from the Current Research Information System of each organization
  - An automated data collection process was launched in 2016
  - Many organizations still upload their data on annual basis - deadline at the end of March
Technical platforms

- The data is collected in XML format into the Virta publication information service built by CSC.
- The de-duplicated publication data can be browsed at Juuli, www.juuli.fi, a VuFind-based discovery interface run by the National Library.
- The data in Juuli is updated automatically from Virta each night.
OA status in the national data collection

- There used to be serious issues with the quality of the OA status data
  - Starting from 2016, improved definitions and categories
- Two fields for indicating the OA status
  - One field indicates whether the publication is OA via a Gold or a Hybrid channel
  - The second field is used for indicating whether it has been deposited into a repository (Green OA)
- URLs should be reported for each version, preferably based on DOIs, URNs, Handles
What qualifies as Open Access?

- It would be very nice to have a short and easy-to-understand definition for Open Access
- However, Open Access is actually a relatively complicated issue, with many different flavors
  - Gold OA?
  - Green OA?
  - Hybrid OA?
  - Bronze OA?
  - Gratis OA?
  - Libre OA?
  - Embargoes?
  - Versions?
OA definition

- Starting from 2016, the OA rules were made as clear as possible
  - The publication should be at least free to read (“gratis OA”)
  - Gold/Hybrid and Green Open Access are all accepted
  - Embargoes allowed for Green OA but not for Gold or Hybrid
  - Depositing into an institutional or subject-based repository is OK, but papers available at personal/project websites or at walled gardens like Research Gate or Academia.edu don’t count
  - The OA version must be peer-reviewed to qualify
Observations after two years…

- We now have two years of data (2016, 2017) that has been collected using the new guidelines.
- Although the quality of the data is better than before, it is still far from perfect.
  - Some organizations are putting more effort into it than others.
  - Some of the organizations have been struggling with the identification of OA publication channels vs. Hybrid/Other OA.
  - There are still some technical issues with de-duplication, especially with Gold and Hybrid OA.
- No data on embargoes, licenses or APC costs collected at the moment.
  - Embargoed publications can be reported as Green OA, if the final URL is already available.
OA in 2017 – some (very) preliminary data

The percentage of OA publications at Finnish universities in 2017, by publication type, data collected on 26/3/2018

- A1 Journal article (refereed), original research
- A2 Review article (refereed)
- A3 Book section or chapter (refereed)
- A4 Article (refereed) in conf. proceedings
- B1 Non-refereed journal articles
- B2 Non refereed book sections
- B3 Non-refereed article in conf. proceedings
- C1 Scientific monograph
- C2 Edited book, proceedings or special issue
- D1 Article in a trade journal
- D2 Article in a professional book
- D3 Article in professional conf. proceedings
- D4 Published report
- D5 Textbook, professional manual
- D6 Edited professional book
- E1 Popularised article
- E2 Popularised monograph
- E3 Edited popular book

OA publication channel, Other OA, Green OA+
The good news: OA is increasing

- The share of peer-reviewed OA articles at the Finnish universities rose from 28.9 % in 2016 to 39.9 % in 2017
- In the universities of applied sciences the amount OA went from 49.9 % to 56.0 %
- Most of this increase is probably based on actual change, but some of it may be explained by improvements in the quality of the data
## The 2017 data: Peer-reviewed articles (A1-A4) and OA at Finnish universities

<table>
<thead>
<tr>
<th>University</th>
<th>All publications</th>
<th>OA publication channel</th>
<th>Other OA</th>
<th>Green OA</th>
<th>OA total</th>
<th>% of OA</th>
<th>Green OA+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aalto Univ</td>
<td>3225</td>
<td>459</td>
<td>250</td>
<td>895</td>
<td>986</td>
<td>30,6 %</td>
<td>277</td>
</tr>
<tr>
<td>Hanken</td>
<td>202</td>
<td>37</td>
<td>7</td>
<td>98</td>
<td>117</td>
<td>57,9 %</td>
<td>73</td>
</tr>
<tr>
<td>Univ of Helsinki</td>
<td>7084</td>
<td>1264</td>
<td>1087</td>
<td>2423</td>
<td>2814</td>
<td>39,7 %</td>
<td>463</td>
</tr>
<tr>
<td>Univ of Eastern Finland</td>
<td>2317</td>
<td>574</td>
<td>137</td>
<td>708</td>
<td>936</td>
<td>40,4 %</td>
<td>225</td>
</tr>
<tr>
<td>Univ of Jyväskylä</td>
<td>1965</td>
<td>400</td>
<td>82</td>
<td>1209</td>
<td>1235</td>
<td>62,8 %</td>
<td>753</td>
</tr>
<tr>
<td>Univ of Lapland</td>
<td>335</td>
<td>141</td>
<td>7</td>
<td>16</td>
<td>153</td>
<td>45,7 %</td>
<td>5</td>
</tr>
<tr>
<td>Lappeenranta Univ of Tech</td>
<td>838</td>
<td>301</td>
<td>7</td>
<td>8</td>
<td>312</td>
<td>37,2 %</td>
<td>4</td>
</tr>
<tr>
<td>University of Oulu</td>
<td>2302</td>
<td>471</td>
<td>126</td>
<td>743</td>
<td>906</td>
<td>39,4 %</td>
<td>309</td>
</tr>
<tr>
<td>Univ of Arts</td>
<td>90</td>
<td>29</td>
<td>3</td>
<td>8</td>
<td>36</td>
<td>40,0 %</td>
<td>4</td>
</tr>
<tr>
<td>Tampere Univ of Tech</td>
<td>1428</td>
<td>293</td>
<td>43</td>
<td>455</td>
<td>548</td>
<td>38,4 %</td>
<td>212</td>
</tr>
<tr>
<td>Univ of Tampere</td>
<td>1955</td>
<td>497</td>
<td>152</td>
<td>716</td>
<td>893</td>
<td>45,7 %</td>
<td>244</td>
</tr>
<tr>
<td>Univ of Turku</td>
<td>3320</td>
<td>734</td>
<td>449</td>
<td>1300</td>
<td>1655</td>
<td>49,8 %</td>
<td>472</td>
</tr>
<tr>
<td>Univ of Vaasa</td>
<td>265</td>
<td>27</td>
<td>22</td>
<td>18</td>
<td>58</td>
<td>21,9 %</td>
<td>9</td>
</tr>
<tr>
<td>Åbo Akademi</td>
<td>935</td>
<td>215</td>
<td>69</td>
<td>70</td>
<td>305</td>
<td>32,6 %</td>
<td>21</td>
</tr>
<tr>
<td><strong>All universities</strong></td>
<td><strong>23006</strong></td>
<td><strong>4707</strong></td>
<td><strong>1923</strong></td>
<td><strong>7126</strong></td>
<td><strong>9186</strong></td>
<td><strong>39,9 %</strong></td>
<td><strong>2556</strong></td>
</tr>
<tr>
<td><strong>All universities of applied sciences</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Different types of open access

- Lots of overlap between Gold/Hybrid and Green OA
- Some universities systematically collect all of their Gold and Hybrid OA publications into their own repositories
- Green OA+: Additional OA provided by Green OA was about 11% of the total peer-reviewed article output in 2017
OA in different research fields in 2017

Peer-reviewed journal articles (A1-A2) at the Finnish universities in 2017, percentage of OA by field of science, data collected on 26/3/2017

- Agriculture and forestry
- Engineering and technology
- Humanities
- Medical and health sciences
- Natural sciences
- Social sciences

OA publication channel, Other OA, Green OA+
Plans for the future

- The Ministry of Education and Culture has launched a new project (2017-2020) which is developing a National Research Information Hub
  - The development of the Virta service will continue, but the new system will also cover e.g. research data and infrastructures
- The collection of OA data is one of the areas which may require further development
New developments

- The current Finnish OA monitoring guidelines were created in 2015 - there have been quite a few developments on an international level since then
  - New tools and data sources like Unpaywall now available
  - A fast and fairly reliable way to check the current OA status of a large number of publications (as long as they have a DOI)
- We haven’t yet looked at what would be the best way to utilize these tools, either on a national level or at a local level at each organization
  - Some of the organizations may be already using them
Changing categories

- A recent article by Piwowar et al. adopted a strict definition of Gold OA and introduced a new category, Bronze OA
  - Bronze OA includes both delayed OA and the OA publication channels that don’t use CC licenses
  - Not compatible with the current Finnish OA definition
  - We have a more inclusive definition of Gold OA and don’t accept delayed OA at all
Categories vs. attributes?

- It seems that the commonly used OA categories (Gold, Hybrid, Green, Bronze, Black, etc.) and their definitions are constantly evolving.
- Instead of trying to adapt to all of these changes and conflicting definitions it might be a good idea to use attributes instead of categories.
- The categories (if they are needed) could be generated from the attributes.
- But there may be a downside as well: you might end up with quite a few OA fields per publication…

<table>
<thead>
<tr>
<th>publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>publisher's version</td>
</tr>
<tr>
<td>current OA status (yes/no)</td>
</tr>
<tr>
<td>link: publication channel (OA/not)</td>
</tr>
<tr>
<td>license</td>
</tr>
<tr>
<td>delay end date (if known)</td>
</tr>
<tr>
<td>APC paid by the organization (euros)</td>
</tr>
<tr>
<td>identifier / url</td>
</tr>
<tr>
<td>pre-print version (open)</td>
</tr>
<tr>
<td>license (if other)</td>
</tr>
<tr>
<td>embargo end date</td>
</tr>
<tr>
<td>identifier / url</td>
</tr>
<tr>
<td>self-archived version</td>
</tr>
<tr>
<td>version (final draft, etc.)</td>
</tr>
<tr>
<td>license (if other)</td>
</tr>
<tr>
<td>embargo end date</td>
</tr>
<tr>
<td>identifier / url</td>
</tr>
</tbody>
</table>
Some final thoughts

- Although it is easy to add new fields into the data collection requirements, each extra piece of information has a cost
  - The collection of additional data requires resources and often also changes to the information systems used by the reporting organizations
  - If you ask the organizations to provide information there should be a strong motivation for doing it
- What is it that you are really trying to measure, and why?
  - New data sources and automatic tools provide opportunities for creating better and more cost-effective processes
  - However, you shouldn’t necessarily do things in a certain way just because it happens to be easy
Thank you!