Redefining open - national open access monitoring in Finland

EuroCRIS Membership Meeting, Espoo, 28.5.2019

Jyrki Ilva <jyrki.ilva@helsinki.fi>

Persistent address: http://urn.fi/urn:nbn:fi-fe2019052817430

Towards open access

- Open access a political goal on the European level
    - EU (2016): All scientific publications open by 2020
    - The funders' Plan S initiative (2018) tries make this happen
- Similar developments in Finland
    - A special coefficient (1.2x) for OA publications added into the funding model of the universities (from 2021)
    - National open access strategy under planning
- Growing need for reliable data on the number and share of OA publications

Carlos Moedas
@Moedas
Follow

1/4 Agreed by all #EU #science ministers. Scientific papers #freetoread by 2020 #openaccess

In dramatic statement, European leaders call for 'immediate' open access to ...
Observers are skeptical goal can be achieved
sciencemag.org

RETWEETS 111
LIKES 59

7:28 PM - 31 May 2016

National data collection in Finland

- The Finnish OA monitoring efforts are integrated with the national data collection conducted for the Ministry of Education and Culture
    - In most cases the publication metadata is collected from the CRIS 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
- The data is collected 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

Type of Publication	▲
A Peer-reviewed scient…	209,374
A1 Journal article (r…	141,546
A2 Review article, Lit…	8,277
A3 Book section, Ch…	25,484
A4 Conference proc…	34,067
B Non-refereed scientifi…	37,901
C Scientific books (mono…	7,745
D Publications intended…	64,377
E Publications intended …	34,381
F Public artistic and desi…	2,754

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
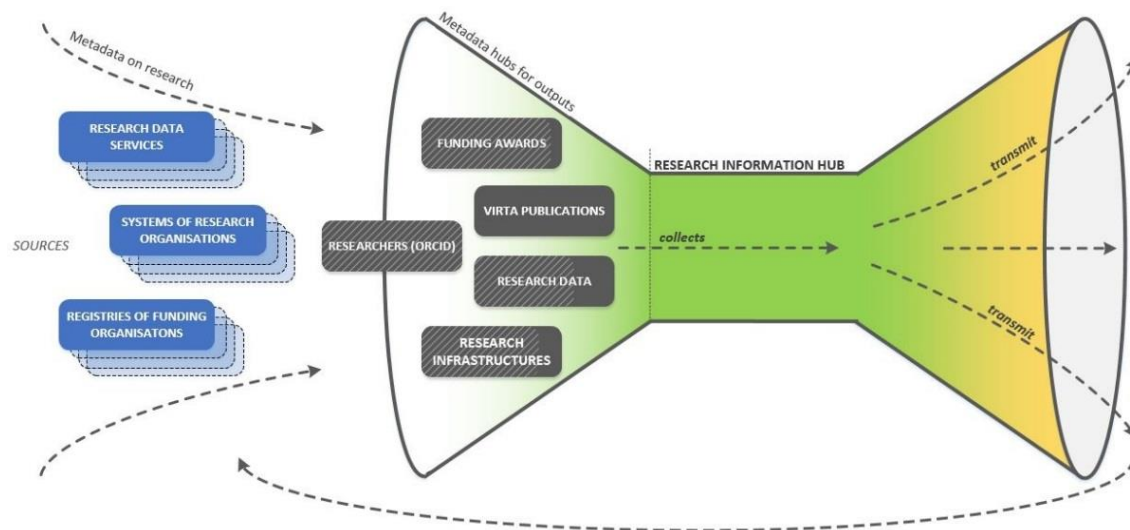
What qualifies as open access?

- 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 used in data collection

- The publication should be 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 (= no "delayed OA")

- Depositing the publication into an institutional or subject-based repository is OK (= uploading the publication into a personal/project web site or to commercial networking services like ResearchGate or Academia.edu does not qualify as open access)

- The OA version of the publication must be peer-reviewed to qualify (= an unrefereed pre print version is not enough)

OA status in the national data collection

- The current OA definitions and categories in use since 2016
    - There is now uniform data from three years (2016-2018)
- 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 (preferably DOIs, Handles or URNs) should be reported for each version

Year of Publication	▲
2018	✓

Open Access	▲
Green OA	10,518
Gold OA publication channel	6,490
Other OA availability	2,790

Issues with quality of the data

- 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
  - The organizations are using different methods in collecting and verifying the data
  - Some of the organizations have been struggling with the identification of OA publication channels vs. hybrid/other OA
  - Delayed OA is hard to identify afterwards
  - 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

Open access in different publication types

The percentage of OA publications at Finnish universities in 2018,
by publication type, data collected on April 2, 2019

Legend: OA publication channel, Other OA, Green OA+
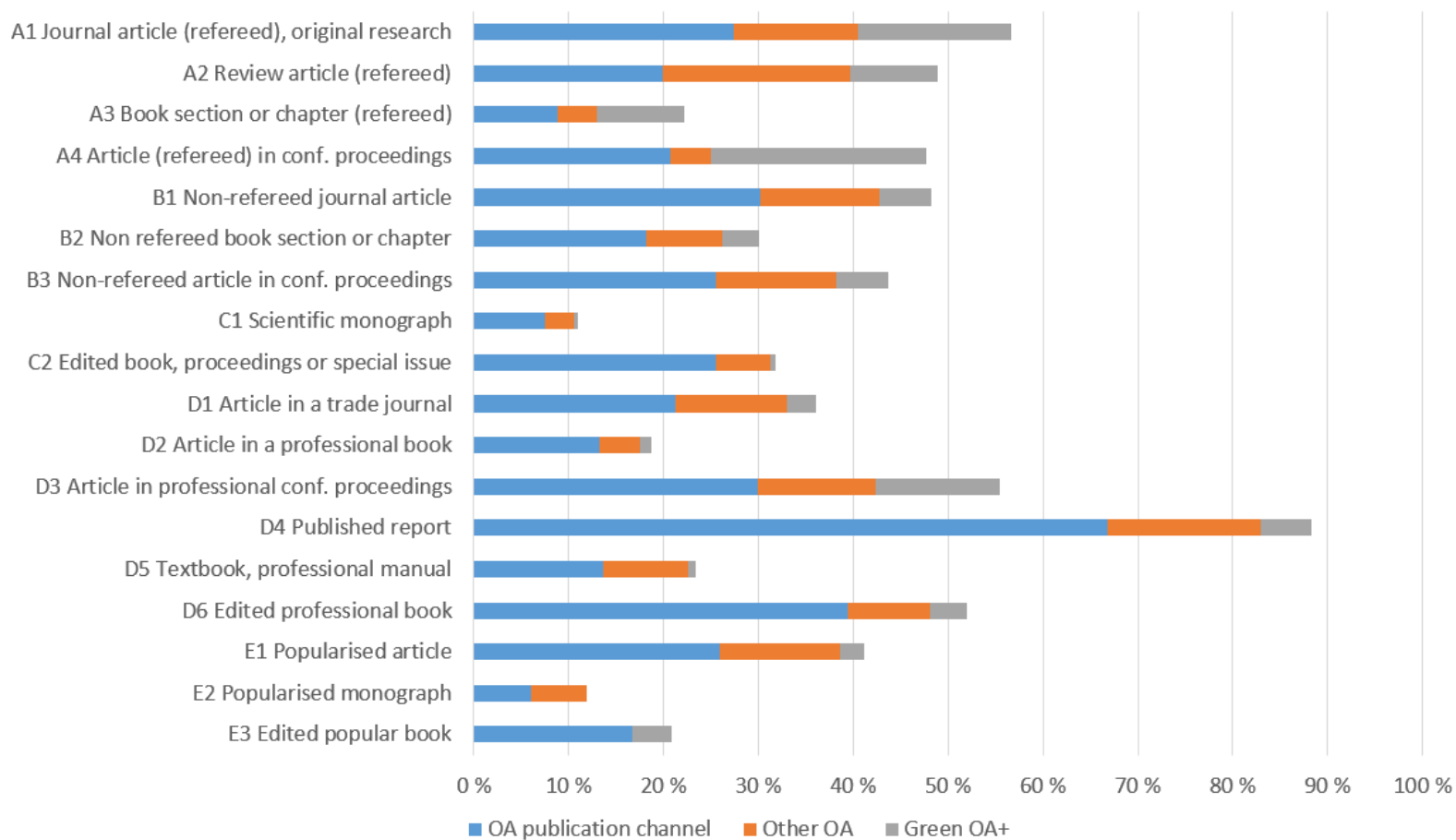
OA seems to be increasing year by year

The share of OA among the peer-reviewed articles and book chapters (A1-A4) at Finnish universities 2016-2018, data collected on April 2, 2019
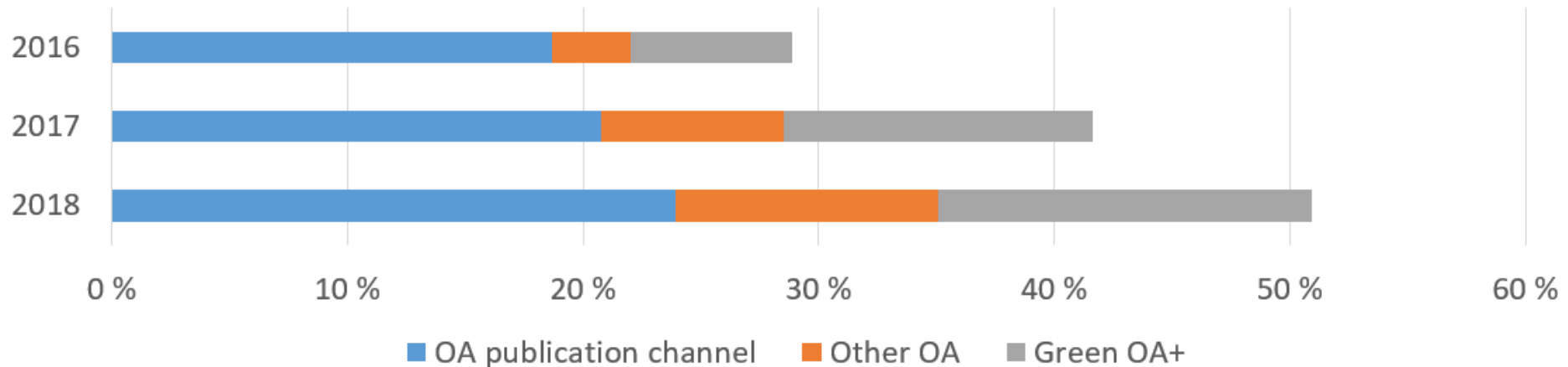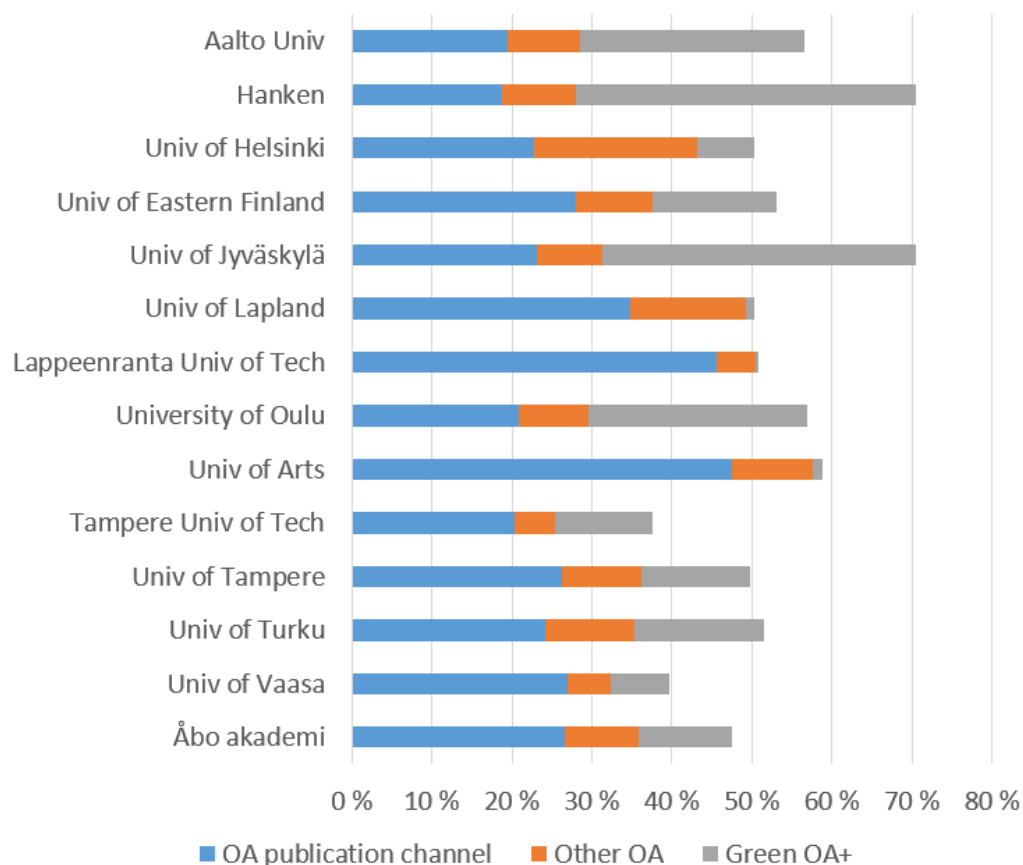

- According to the data reported by the universities, the share of OA among the peer-reviewed research articles grew from 41,7 % in 2017 to 50,9 % in 2018
- At the universities of applied sciences the amount OA grew from 56,2 % in 2017 to 61,4 % in 2018

Different routes to open access

- Large differences in the total share of OA and the share of different types of OA between universities

- 31,7 % of the peer-rewiewed articles openly available on publishers' platforms

- Lots of overlap between gold/hybrid and green OA

- Green OA added 15,8 % to the total amount of OA

The share of OA among the peer reviewed articles and book chapters (A1-A4) at each Finnish university in 2018, data collected on April 2, 2019

Aalto Univ
Hanken
Univ of Helsinki
Univ of Eastern Finland
Univ of Jyväskylä
Univ of Lapland
Lappeenranta Univ of Tech
University of Oulu
Univ of Arts
Tampere Univ of Tech
Univ of Tampere
Univ of Turku
Univ of Vaasa
Åbo akademi

0 %   10 %   20 %   30 %   40 %   50 %   60 %   70 %   80 %

■ OA publication channel   ■ Other OA   ■ Green OA+

OA in the main research fields in 2018

Peer-reviewed journal articles (A1-A2) at the Finnish universities in 2018, percentage of OA by field of science, data collected on April 2, 2019
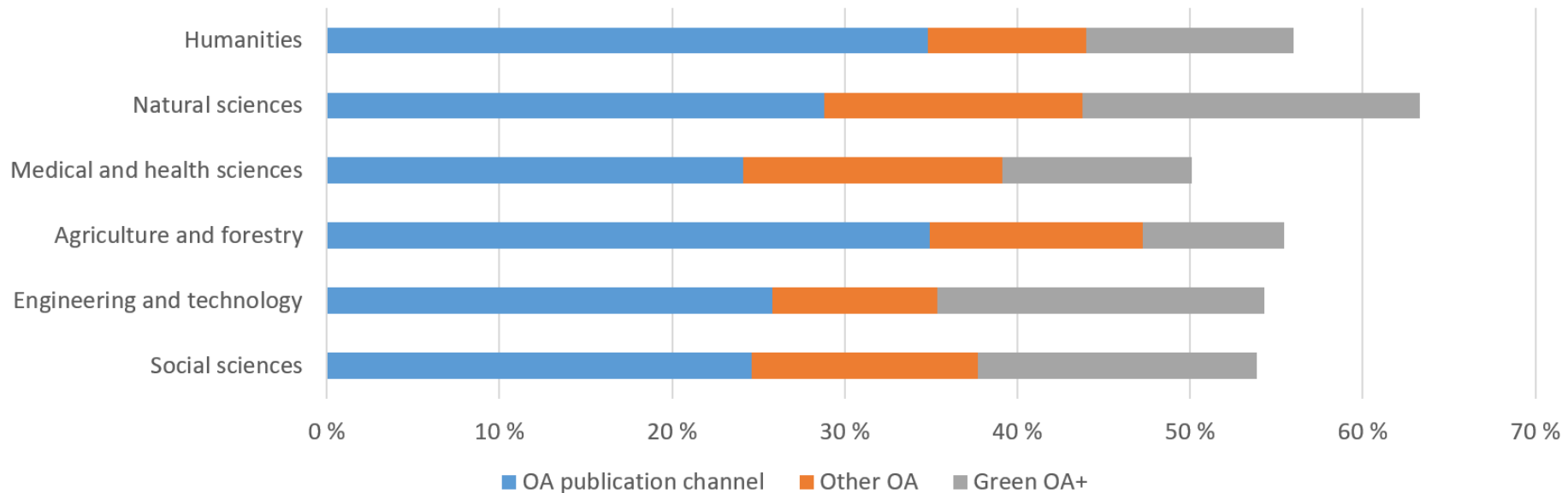
Growth continues afterwards

- The share of OA continues to grow after the original reporting year
  - Of peer-reviewed articles produced by the universities in 2016 37,5 % now openly available (originally 29,7 %)
  - 46,9 % of the 2017 articles now openly available (orig. 41,7 %)
- Obvious explanations for the continued growth of green OA
  - The end of embargo periods opens up new publications
  - It's fairly common to self-archive older publications
- But there are also changes in the comprehensiveness and quality of the information
  - The share of gold and hybrid/other OA has also grown after the original reporting year in many organizations
  - Changes in the methods of data collection – for example the use of tools like Unpaywall – the most likely explanation

New political guidelines

- [Plan S](#) is setting new goals which are likely to have an effect on OA monitoring efforts
    - Aims to get rid of green OA embargos and hybrid OA
    - At least the preliminary Plan S guidelines demand the use of certain Creative Commons -licences (mainly CC-BY = "libre OA")
    - Calls for a cap for the size of APC payments

cOAlition S

Making Open Access a reality by 2020

A DECLARATION OF COMMITMENT BY PUBLIC RESEARCH FUNDERS

http://scieur.org/coalition-s

S = science, speed, solution, *shock*

New OA monitoring tools

- [Unpaywall](#) has been most popular of the new OA monitoring tools
  - Fast and relatively reliable tool for checking the OA status of publications (as long as they have a DOI)
  - Data can be utilized via an API
- The methods used by Unpaywall seems to be having an effect on the commonly used definitions of OA
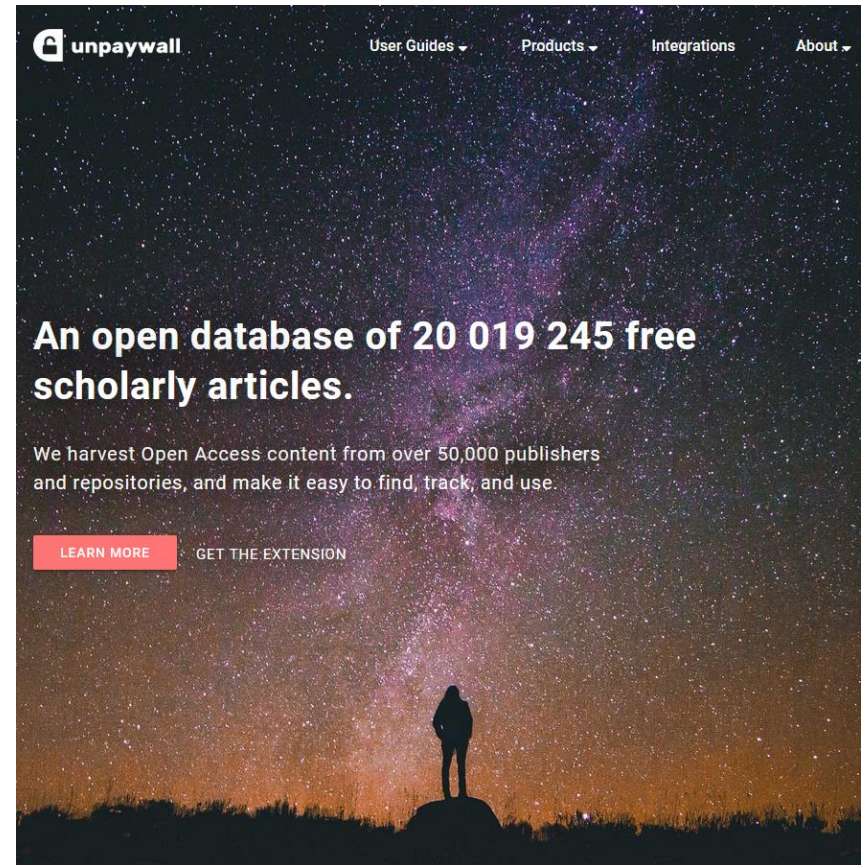
Changing categories

- A 2018 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
  - Compatible with the methods used by Unpaywall
- 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

THE NATIONAL LIBRARY OF FINLAND

From categories to 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
- The obvious downside is that there may needt to be quite a few OA fields per publication, at least in some cases

- oa status of the publication
  - publisher's version
    - current oa status
    - oa status of the publication channel
    - license of the publication
    - apc paid (in euros)
    - persistent url
  - self-archived version
    - yes / no
    - version
    - license (if differs)
    - embargo end date
    - persistent url
  - pre print version
    - yes / no
    - license (if differs)
    - publication date
    - persistent url

Enhancing the Virta data model?

- A working group on the development of the Virta data model proposed new enhancements to the OA fields (January, 2019)
    - Adoption of an attribute-based model for OA status information
    - Separating the current OA status of publisher's version from the OA status of the publication channel (enables the collection of data on delayed OA)
    - Enabling the collection of license information
    - Additional fields for information concerning the self-archived version (embargo end date, version information)
    - Enabling the collection of data on pre prints
    - Fields for collecting information on APC payments (a FinELib project funded by the Ministry is currently working on this on a national level)

Thank you!