■ AN AUSTRIAN PROPOSAL FOR THE CLASSIFICATION OF OPEN ACCESS TUPLES (COAT) – DISTINGUISH DIFFERENT OPEN ACCESS TYPES BEYOND COLORS

by Patrick Danowski

Abstract: *In this article a model is described how Open Access definitions can be formed on the basis of objective criteria. The common Open Access definitions such as "gold" and "green" are not exactly defined. This becomes a problem as soon as one begins to measure Open Access, for example if the development of the Open Access share should be monitored. This was discussed in the working group on Open Access Monitoring of the AT2OA project and the present model was developed, which is based on 5 critics with 4 characteristics: location, licence, version, embargo and conditions of the Open Access publication are taken into account. In the meantime, the model has also been tested in practice using R scripts, and the initial results are quite promising.*

Contents

The classic Open Access (OA) colors Gold and Green are widely used in OA studies and also in the first monitoring activities. Still, the understanding of the colors is not as clear as expected, especially for Gold, the notion of the term can be quite diverse. Some studies label articles published in Hybrid journals as Gold OA, others like to establish a whole new category for this kind of publications. The same is true for the newly established category of Bronze OA[1].

Thinking about monitoring OA and about interoperability of different monitoring systems, these various notions of the OA definition could become a problem in the near future.

In the *Austrian Transition to Open Access (AT2OA)* project, a working group discussed how an OA monitoring can be developed. As a first step, we started to think about a controlled vocabulary for the different OA types. The main problem with such a vocabulary would be that only librarians and very OA affine users are going to understand it.

Eventually, we discarded this approach and started to look at what criteria affect the notion of OA types.

We determined five different criteria:

1. **Place of OA**
2. **License**
3. **Publication Version**
4. **Embargo Period**
5. **Conditions of OA**

Every criterion is defined by one or more metadata elements. Depending on the values, four different classes can be built for every criterion.
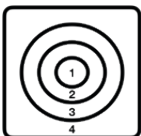
A class defines a minimum and contains always the values of the smaller classes. Class 3 contains also everything from the classes 1 and 2 but not what is defined by class 4.

1. Place of Open Access

Description: This criterion defines where the OA version is available.

Metadata fields:
- **Identifier URI** of the original version (e.g. the DOI link for a journal article),
- **Link to Open Access version(s)**

Evaluation classes:
1) The source is OA (a link to an OA version is identical with the URI)
2) An OA version is available in a repository (A link to an OA version is a link to a repository listed in the ROAR[2] or OpenDOAR[3]).
3) Other OA version, if all links to the OA versions are not fitting into 1 or 2
4) No OA

2. License

Description: License type

Metadata field:
- Link to license

Evaluation classes:
1) Open License (CC BY or CC BY-SA, CC0 or comparable license complying with the Open Definition[4])
2) Free License (Other CC license or comparably licensed)
3) Proprietary licenses (e.g. Publisher specific)
4) No license / license unknown

3. Publication version

Description: The version of the OA publication

Metadata field:
- Publication version (by using DRIVER vocabulary[5], e.g. info:eu-repo/semantics/submittedVersion)

Evaluation classes:
1) Publisher's version
2) Postprint or last submitted version
3) Preprint or first submitted version
4) Unknown

4. Embargo period

Description: Length of the embargo

Metadata fields:
– Publication Date (in ISO format)
– Embargo Date (in ISO format)

Evaluation classes:
1) No embargo
2) Up to 6 months
3) Up to 12 months
4) More than 12 months / unknown

5. Conditions of Open Access

Description: Under which financial conditions OA was realized

Metadata fields: Currently, no metadata model contains a field to store the information; one reason why a topic related vocabulary had to be developed.

Evaluation classes:
1) Free
2) Paid in OA media
3) Paid in subscription media
4) No OA

Taking the categories into account, the different OA types can be defined as follows:

Description	Place of OA	License	Publication Version	Embargo Period	Conditions of OA	Recommended writing
Gold as in the Berlin Declaration	1	1	1	1	3	(1,1,1,1,3)
Gold without Hybrid, any license	1	4	1	1	2	(1,4,1,1,2)
Gold including Hybrid, Free License	1	2	1	1	3	(1,2,1,1,3)
Bronze/ Gratis OA	1	4	1	4	1	(1,4,1,4,1)
Green, Postprints Only	2	4	2	4	1	(2,4,2,4,1)
Green any Version	2	4	4	4	1	(2,4,4,4,1)
Other OA	3	4	4	4	1	(3,4,4,4,1)

As a reminder the numbers always define the maximum of the value, smaller numbers are included with an OR condition. For Gold without Hybrid with free license the conditions would be:

– Identifier URI is part of Links to OA version(s) (1) AND
– License category is Open OR License category is free (2) AND
– Publication Version is "Publisher's Version" (1) AND
– Publication Date = Embargo Date (1) AND
– Journal listed in DOAJ (2)

How to use the definitions in studies and in monitoring activities

At the moment, the evaluated categories in the different studies are mainly described as Gold, Green and, less frequently, the category Hybrid is also used. The definitions of the colors vary depending on the study. To get a better feeling for the definitions and also to make the studies comparable, it would be advisable to add the tuple as information. The tuple can be added to the color. We evaluated gold (1,4,1,1,2), hybrid (1,2,1,1,3) and green (2,4,2,4,1).

In case, another study defines the color as gold (1,1,1,1,2), hybrid (1,2,1,1,3) and green (2,4,4,4,1), it becomes obvious why the results look totally different from the very start. If the raw data contains the information needed for all five criteria as proposed than the evaluation could be repeated with the different OA definitions.

Using a set of OA Definitions

Most monitoring approaches are not only using one OA definition they are using OA categories like gold, hybrid and green to differentiate.

In this case the "lower" definition has to take into account the higher definitions. Lets say we like to use gold (1,4,1,1,2), hybrid (1,4,1,1,3) and green (2,4,2,4.1) then our conditions will look like:

Gold
- Identifier URI is part of Links to OA version(s) (1) AND
- License type is Any (4) AND
- Publication Version is "Publisher's Version" (1) AND
- Publication Date = Embargo Date (1) AND
- Journal listed in DOAJ (2)

Hybrid
- Identifier URI is part of Links to OA version(s) (1) AND
- License type is Any (4) AND
- Publication Version is "Publisher's Version" (1) AND
- Publication Date = Embargo Date (1) AND
- Journal NOT listed in DOAJ (2) OR Payment tracked in OpenAPC[6] for Identifier

Green
- Identifier URI is NOT part of Links to OA version(s) (1) AND Links to OA version(s) lead to ROAR registered site

- License type is Any (4) AND
- Publication Version is "Post Print" (2) AND
- Publication Data >= Embargo Data (4) AND
- Journal NOT listed in DOAJ (2) AND NO payment tracked in Open-APC for Identifier

This logic has been implemented in the system. If a metadata field for the last category is defined this certainly has to be adapted.

The classification was discussed in the OA monitoring working group of the Austrian Transition to Open Access (AT2OA) project and was first presented during the workshop "Open Access Monitoring – Approaches and Perspectives"[7].

A first implementation of this concept in R is available at GitHub: https://github.com/patrickda/COAT.

Dipl. Inf. Patrick Danowski, MA
ORCID: https://orcid.org/0000-0002-6026-4409
Institute of Science and Technology Austria (IST Austria), Library
E-Mail: patrick.danowski@ist.ac.at

* Thanks to Márton Villányi, Andreas Ferus and Michael Uwe Möbius for their feedback to improve this article.
1 Hybrid and Bronze used in: Piwowar H, Priem J, Larivière V, Alperin JP, Matthias L, Norlander B, Farley A, West J, Haustein S. (2017). The State of OA: A large-scale analysis of the prevalence and impact of Open Access articles. PeerJ Preprints 5:e3119v1. https://doi.org/10.7287/peerj.preprints.3119v1 and Bosman J, Kramer B. (2018). Open access levels: a quantitative exploration using Web of Science and oaDOI data. PeerJ Preprints 6:e3520v1. https://doi.org/10.7287/peerj.preprints.3520v1
2 Registry of Open Access Repositories: http://roar.eprints.org
3 Directory of Open Access Repositories: http://www.opendoar.org
4 The Open Definition: https://opendefinition.org
5 DRIVER guidelines, version vocabulary: https://wiki.surfnet.nl/display/DRIVERguidelines/Version+vocabulary
6 GitHub, OpenAPC: https://github.com/OpenAPC/openapc-de
7 See also: Danowski P, Ferus A, Hikl A, Kaiser O, McNeill G, Reding S, Schautz M, Zarka T, Zojer M (2018). Report on "Open Access Monitoring – Approaches and Perspectives" (2-Day-Workshop, 09–10 April 2018, Vienna). Zenodo. http://doi.org/10.5281/zenodo.2415894