

Universität Bielefeld

Institute for Interdisciplinary
Studies of Science (I²SoS)

INT@CT
0@analytics

ISSN-Matching of Gold OA Journals

(ISSN-GOLD-OA)

- User Documentation -

Version 0.5
October 20, 2016

Michael Wohlgemuth
Christine Rimmert
Matthias Winterhager

Bielefeld University
Institute for Interdisciplinary Studies of Science (I²SoS)
www.uni-bielefeld.de/bibliometrie

1. General remarks

Purpose of the dataset

ISSN-GOLD-OA provides a matching list of ISSN for Gold Open Access (OA) journals. The intention was to compile a matching table as complete as possible by using different publicly available sources. The dataset can help to clear various ISSN-related issues in bibliometric studies on Gold OA.

Definition of Gold OA Journal

According to Éric Archambault et al. (2014, following Peter Suber 2012) a Gold OA Journal is a journal offering immediate cover-to-cover open access, provided by a publisher, sometimes with paid for publication fee.^{1,2} Neither *Delayed OA*, *Hybrid OA* nor *Green OA* is considered for the ISSN-GOLD-OA dataset.

Data sources

The Directory of Open Access Journals ([DOAJ](#)) is a most popular source for determining the OA status of journals and in fact up to now the vast majority of bibliometric studies dealing with Gold OA journals did rely solely on DOAJ as a source. To overcome blind spots and data quality problems within DOAJ, a number of further resources have been evaluated with regard to coverage, traceability, validity and timeliness of data. Finally three additional sources were selected as offering relevant amounts of valuable information for the ISSN-GOLD-OA dataset:

- [Directory of Open Access scholarly Resources](#)³ (ROAD)
- [PubMed Central](#)⁴ (PMC)
- [Open APC initiative](#)⁵ (OAPC)

To ensure the processing of exclusively Gold OA journals, the respective data flags are used for selection of records from PMC ('open_access='All') and OAPC ('hybrid=FALSE'). Any available ISSN is included, irrespective of the type (print, electronic, etc).

For the purpose of matching and merging data entries with different versions of ISSN from the selected sources a special table of linking ISSN ([ISSN-L](#)) is used, as provided by the ISSN International Centre.⁶

For the current version of the ISSN-GOLD-OA dataset the five sources have been used with their latest data status available on August 01, 2016, as shown in table 1.

1 Archambault, E. et al. (2014). [Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels—1996–2013](#). Deliverable D.1.8. (2014 Update). Version 11b.

2 Suber, P. (2012). [Open Access](#). Cambridge, Mass.: MIT Press. ISBN 9780262517638.

3 <http://road.issn.org/en/contenu/download-road-records> (last visited 2016-10-15)

4 <https://www.ncbi.nlm.nih.gov/pmc/journals/?format=csv> (last visited 2016-10-15)

5 <https://github.com/OpenAPC> (last visited 2016-10-15)

6 <http://www.issn.org/services/online-services/access-to-issn-l-table/> (last visited 2016-10-15)

Source	Status
DOAJ	2016-08-01
ROAD	2016-07-07
PMC	2016-07-31
OAPC	2016-08-03
ISSN-L	2016-07-30

Tab. 1: Processed versions of data sources

Journal – entity/historiography

The ISSN-L table is used as a reference to define the entity 'Journal' (i.e. the collocation of several ISSNs for a dedicated journal): two different ISSNs point to the same journal, if and only if they are linked to the same ISSN-L.

Exhaustive historiography for Gold OA journals (covering structural changes and previous OA status reports) is not intended here. A journal is selected as Gold OA if it shows up in any of the sources (under the conditions and data status mentioned above). Historical states are not taken into account, because an OA start date is often missing in the records. Moreover, a switch of a journal toward Gold OA status is usually not restricted to the current and future volumes, but extended to its backfiles. The *Scientific Bulletin of Naval Academy* for example is recorded in DOAJ with 2011 as “Year Open Access Content Began”⁷, whereas the OA status of this journal is actually covering volumes earlier than 2011⁸.

Enrichment by ISSN-L

In many cases the sources show incomplete records, missing particular ISSNs of a journal, e.g. only the print ISSN is given although the journal carries a separate electronic ISSN too. Missing ISSNs of this kind are complemented into the ISSN-GOLD-OA dataset in the following way: if any ISSN for an ISSN-L exists in any of the OA sources (under the conditions and data status mentioned above), then every other ISSN corresponding to this ISSN-L is captured as well.

Manual data cleaning

To ensure that the additional information taken from ROAD, PMC and OAPC will really lead to improved data quality against DOAJ, manual cleaning steps are performed. For PMC and OAPC, each additional ISSN candidate is checked for its Gold OA status (via web page of publisher). For reasons of economy, these checks for ROAD are performed only for ISSNs which show up in Web of Science.

7 <https://doaj.org/toc/2392-8956> (last visited 2016-10-15)

8 <https://www.anmb.ro/buletinstiintific/eng/ahive.html> (last visited 2016-10-15)

2. Explanation of data fields

Column	Comment
ISSN	ISSN
ISSN_L	Linking ISSN (ISSN-L)
TITLE	Journal title
TITLE_SOURCE	Source of journal title
ISSN_IN_DOAJ	1 if ISSN exists in DOAJ, else 0
ISSN_IN_ROAD	1 if ISSN exists in ROAD, else 0
ISSN_IN_PMC	1 if ISSN exists in PMC, else 0
ISSN_IN_OAPC	1 if ISSN exists in OAPC, else 0
JOURNAL_IN_DOAJ	1 if journal exists in DOAJ, else 0
JOURNAL_IN_ROAD	1 if journal exists in ROAD, else 0
JOURNAL_IN_PMC	1 if journal exists in PMC, else 0
JOURNAL_IN_OAPC	1 if journal exists in OAPC, else 0
ISSN_IN_WOS	1 if ISSN exists in Web of Science, else 0
JOURNAL_IN_WOS	1 if journal exists in Web of Science, else 0
ISSN_IN_SCOPUS	1 if ISSN exists in Scopus, else 0
JOURNAL_IN_SCOPUS	1 if journal exists in Scopus, else 0

Tab. 2: ISSN-GOLD-OA dataset fields

ISSN:

Every available ISSN is captured from the sources, regardless of its type (e.g. print ISSN, electronic ISSN). A formal validity check (last digit as check digit) is performed for each selected ISSN, invalid ISSNs are rejected.

ISSN_L:

“The linking ISSN, or ISSN-L, is a specific ISSN that groups the different media editions of the same serial publication. A single ISSN-L is designated for all media of a serial publication, irrespective of how many there are. A serial publication is associated with a single ISSN-L.” (ISSN International Centre 2016)⁹ Therefore the ISSN-L provides a mapping of one or more ISSNs for a given journal and can be used as a unique identifier for journals. Any ISSN in the ISSN-L table is assigned to precisely one ISSN-L.

A few ISSNs that show up in DOAJ, ROAD, PMC or OAPC can not be found in the ISSN-L table at all (despite having passed the validity check). As no assignment to an ISSN-L is available for these cases, the corresponding records in the ISSN-GOLD-OA dataset are left to *null* values in column ISSN-L. A potential lack of completeness is accepted here in

⁹ <http://www.issn.org/understanding-the-issn/assignment-rules/the-issn-l-for-publications-on-multiple-media/> (last visited 2016-10-15)

favor of insuring strict concordance to the official linking list of the ISSN International Centre.

For internal purposes only, an additional column ISSN-IDENT (not included into the regular version of the dataset) is introduced. Its values, in a first step, are set to ISSN-L for every record which actually has already a valid ISSN-L assignment. In these cases the value for column ISSN-IDENT is directly drawn from column ISSN-L, thus both columns have exactly the same value. In a second step, for the rest of the records, ISSN-IDENT is set manually, taking into account information from web pages of the journal publisher. If for a given ISSN any other known ISSN of the same journal has been assigned to an ISSN-L, this value is adopted. If none of the known ISSN of the journal is assigned to an ISSN-L at all, then the maximum ISSN of the journal is taken as ISSN-IDENT for all of its ISSNs.

In fact the figures in tables 3 and 4 in the following chapter are calculated using this approach, as it allows to include every ISSN available in any of the used OA sources.

TITLE:

Different sources can show different titles for the same journal. For the ISSN-GOLD-OA dataset, only one title per journal is included. It is captured from the source mentioned in column SOURCE_TITLE. Where possible, the title is taken from DOAJ, otherwise from ROAD, PMC or OAPC (in this priority sequence).

TITLE_SOURCE:

OA source from which TITLE is captured.

ISSN_IN_<OA-SOURCE> – flags:

Every OA source has its own flag to indicate whether (=1) or not (=0) the ISSN exists in the source.

JOURNAL_IN_<OA-SOURCE> – flags:

Every OA source has its own flag to indicate whether (=1) or not (=0) the journal exists in the source. A journal is defined here as an ISSN-L. If any of the ISSNs belonging to a specific ISSN-L (not necessarily the one mentioned in same row) exists in the source, the value is 1 here, otherwise 0.

ISSN_IN_<WOS/SCOPUS>:

Like the OA sources, Web of Science (WoS) and Scopus have separate flags of the same kind, indicating whether (=1) or not (=0) the ISSN exists in these databases.¹⁰

JOURNAL_IN_<WOS/SCOPUS>:

Like the OA sources, WoS and Scopus have separate flags of the same kind, indicating whether (=1) or not (=0) the journal (as defined by ISSN-L) exists in these databases.

¹⁰ Check of existence in WoS/Scopus done as up-to-date on August 22, 2016.

3. Statistical overview

Coverage of OA sources

The number of distinct Gold OA ISSNs and journals (ISSN-L) in total and per OA source as well as the respective coverage in Web of Science (WoS) and Scopus is shown in table 3.

	# ISSNs	# ISSNs in WoS	# ISSNs in Scopus	# Journals	# Journals in WoS	# Journals in Scopus
Total	30.527	2.527	5.958	20.450	1.672	4.130
ROAD	22.751	1.895	4.337	15.926	1.299	3.099
DOAJ	12.823	1.612	3.855	9.366	1.178	2.866
PMC	2.266	704	1.503	1.547	504	1.065
OAPC	1.007	496	677	828	431	582

Tab. 3: Number of Gold OA ISSN and Journals in OA Sources

Each ISSN and journal can be covered in one or more of the OA-sources. Numbers are calculated after manual data cleaning as mentioned above. Only ~42% of all ISSNs are covered in DOAJ. If restricted to the ISSNs that appear in Web of Science, DOAJ covers ~64% of them (Scopus: ~65%). Thus a substantial loss of information is supposed to affect bibliometric studies of OA which rely solely on DOAJ.

Out of the total number of ISSNs, 204 show up exclusively in OAPC (26 of them appearing in Web of Science and 40 in Scopus). These 204 ISSNs refer to 134 distinct journals (19 out of them covered in WoS and 30 in Scopus).

Gain of ROAD/PMC/OAPC against DOAJ

The number of dedicated Gold OA ISSNs and journals missing in DOAJ but covered in one or more of the other sources is shown in table 4. This reveals the potential amount of information loss if ROAD, PMC and OAPC would be neglected.

	# ISSNs	# ISSNs in WoS	# ISSNs in Scopus	# Journals	# Journals in WoS	# Journals in Scopus
Total	17.704	915	2.103	11.084	494	1.264
ROAD	15.861	746	1.752	10.783	441	1.145
PMC	680	179	415	390	90	248
OAPC	275	59	91	187	41	66

Tab. 4: Number of ISSN and Journals in ROAD/PMC/OAPC not in DOAJ

Funding acknowledgement

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) [grant number WI 3919/1-1].