

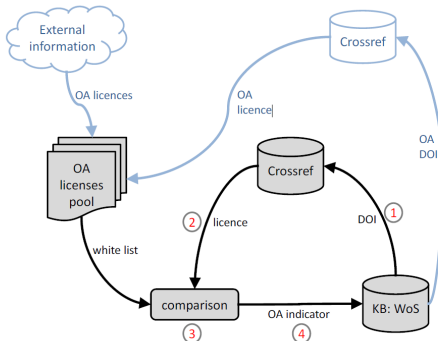
“Merits and Limits”
Applying open data to monitor Open Access
publications in bibliometric databases
OA in Crossref & Unpaywall

Aliakbar Akbaritabar,
Stephan Stahlschmidt

ISSI and STI 2019, Rome

- Use Crossref and Unpaywall information and detect OA
 - including: Gold, HiddenGold, Hybrid, Probable Hybrid and Delayed OA
- Updated report: **10.31235/osf.io/sdzft**
(<https://osf.io/preprints/socarxiv/sdzft/>)

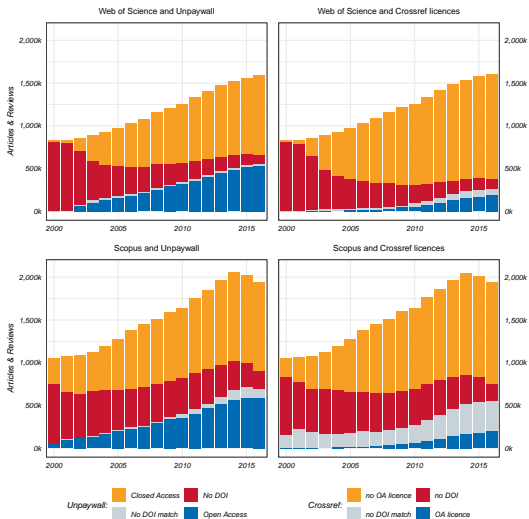
- In-house WOS/Scopus of 2017
- In-house Crossref snapshot (April 2018)
- In-house Unpaywall snapshot (April 2018)
- Schema for Crossref procedure



- ISSN in DOAJ and/or ROAD \implies **Gold OA**
- All publications had OA licences, ISSN not in DOAJ or ROAD \implies **Hidden Gold OA**
- An issue had at least one *non-OA* publication + one or more OA pubs \implies **Hybrid OA**
- An issue did not have a *non-OA* publication + one or more OA pubs + pubs w/ no licence \implies **Probable Hybrid OA**
- If delay-in-days > 0 in all above cases \implies **Delayed OA**
- ISSN not in DOAJ and/or ROAD, number of pubs in issue = *non-OA* \implies **Closed Access**
- None of the above \implies **Not available (NA)**

- DOI is *not* as standard as it is hoped
- Duplicate DOIs
- Non coherent *upper/lower* case usages
- An example:
 - *Exact matching AR/RE of WOS w/ Unpaywall* \Rightarrow 11,346,682 (57.03%)
 - *Lower case DOI* \Rightarrow 13,886,618 (69.8%)
 - Duplicated DOIs removed, increased matches even further
- WoS (Scopus), between 2000 and 2016:
 - 63 (913) journals **without an ISSN**
 - 23,000+ (140,000+) articles and reviews **without a unique DOI** (excluded from our results)
 - 5,700,000+ (7,400,000+) pubs **without any DOI**
 - 2,700,000+ (4,500,000+) pubs with a DOI which can't be resolved on Crossref

Crossref/Unpaywall matched WOS/Scopus



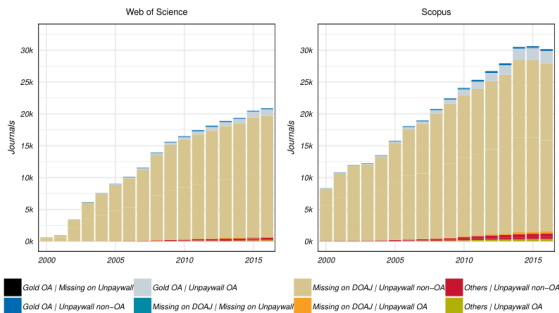


Figure 1: Journals indexed in WOS and Scopus matched with Unpaywall database and crosschecked the ISSN with DOAJ (Gold OA) between 2000 and 2016

Table 2: Number of licences per DOI found in Crossref for articles and reviews indexed in either WOS, Scopus or both between 2000 and 2016

Number of licences per DOI	Frequency of DOIs	Percent
0	6,571,079	42.74
1	8,143,752	52.97
2	655,729	4.26
3	3,472	0.02
4	17	0.00
6	1,152	0.01

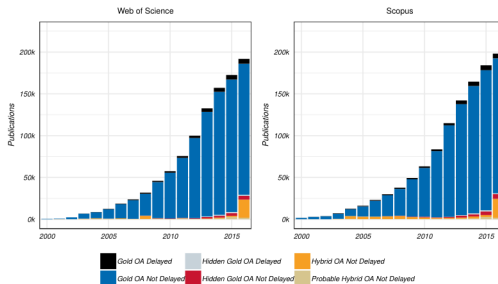
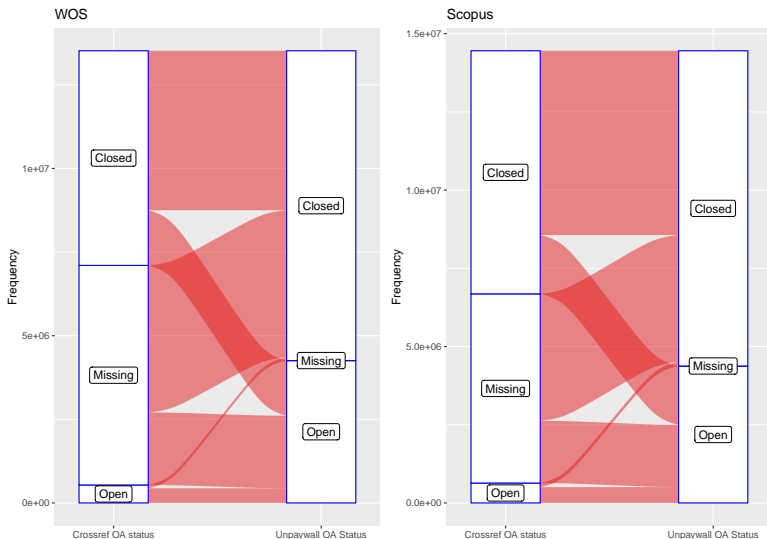


Figure 2: Count of gold and hybrid OA publications between 2000 and 2016 based on Crossref licence information, DOAJ and ROAD)

Crossref/Unpaywall vs. WOS/Scopus





- *WOS/Scopus*: **Incoherent** metadata in each database (duplicate DOIs, lower/upper case, no DOI)
- *Unpaywall/Crossref*: **Incoherent** metadata over databases (same DOI, different OA status)
- Good opportunity of mixing/matching/merging different databases
- Unpaywall has a higher coverage that goes beyond Crossref

Thanks for your attention!