

## A Tale of Two Societies

### Are UK and US Chemistry Publishing Diverging on Open Access?

## Colophon

- Code and data on Github: <https://github.com/Curtin-Open-Knowledge-Initiative/presentation-acs2021/>
- Presentation on Github: <https://curtin-open-knowledge-initiative.github.io/presentation-acs2021/>
- Code and presentation at Zenodo: Coming soon...
- DOI: Coming soon...

Copyright Cameron Neylon 2020-2021. This slide deck is licensed under a [Creative Commons Attribution 4.0 International License](#). Code is available under an Apache v2.0 license at Github.

## Two societies

...both alike in dignity

In FAIR[?] chemistry where we lay our scene

Different national policy environments on OA 2010-20

...and different publisher responses

A quick segue on the data

### The Curtin Open Knowledge Initiative

"Our goal is to **change the stories that universities tell about themselves**, placing open knowledge at the heart of that narrative"

## The Data

### The Data

Data is derived from the following sources

- Crossref - weekly dump via Metadata Plus program
- Unpaywall - Open Access Status data via open data dump (October 2020)
- Microsoft Academic - Affiliation and authorship data via biweekly dump
- GRID - Information on organisations via regular data dump

Data is integrated and processed via Observatory Platform, an open source workflow system developed within COKI to integrate data related to scholarly communications. The [code is available on Github](#) including the template for [SQL queries which generates the OA status](#) we use from the Unpaywall data. More detail on the OA categories is also provided on the slide below.

For this analysis, publisher is defined by the text string in the Crossref metadata, affiliation is derived from the assignment by Microsoft Academic and field is the [Level Zero field "Chemistry" from Microsoft Academic](#) as provided in the data dump used. Some comparative data for ["Materials Science"](#) and ["Biology"](#) is presented in additional slides.

## Open Access Categories

The following is derived from the template SQL query used to define the OA categories in Observatory Platform. For the most up to date information view the [file on Github](#).

### AGGREGATE UNPAYWALL QUERY TEMPLATE

This template query contains the SQL that directly interprets Unpaywall data to determine OA categories at the output level. This is therefore the canonical location for the precise definitions used for OA categories. Ideally this file should contain both the queries themselves and a description clear enough for a non-expert in SQL to understand how each category is defined.

The current categories of Open Access described in this file are:

- is\_oa: derived directly from Unpaywall
- hybrid: accessible at the publisher with a recognised license
- bronze: accessible at the publisher with no recognised license
- gold\_just\_doaj: an article in a journal that is in DOAJ
- gold: an article that is in a DOAJ journal OR is accessible at the publisher site with a recognised license (hybrid)
- green: accessible at any site recognised as a repository (including preprints)
- green\_only: accessible at a repository and not (in a DOAJ journal OR hybrid OR bronze)
- green\_only\_ignoring\_bronze: accessible at a repository and not (in a DOAJ journal or hybrid)

## Open Access Categories

Levels of open access publishing are as provided by Unpaywall for all outputs with RSC and ACS as publisher in Crossref metadata. In the Observatory Platform terminology this is "gold" open access and includes all articles published in either fully open access (as defined by indexing in DOAJ) and in mixed journals (i.e. "hybrid")

As defined in the SQL used to generate these categories.

Gold Open Access:

gold OA is defined as either the journal being in DOAJ or the best\_oa\_location being a publisher and a license being detected. This works because Unpaywall will set the publisher as the best oa location if it identifies an accessible publisher copy.

```
CASE
  WHEN journal_is_in_doaj OR (best_oa_location.host_type = "publisher" AND best_oa_location.license is not null AND not journal_is_in_doaj)
  THEN TRUE
  ELSE FALSE
END
as gold,
```

## Information

<http://openknowledge.community/dashboards/>

### What about publishers?

**What can we tell about policy  
and culture change?**

**Publisher choice in UK**

**Publisher choice in UK**

**Publisher choice by country**

**Different patterns with OA publishing levels?**

**Publisher percent of Chemistry by institution**

**Publisher percent of OA publishing in chemistry by institution**

**Parallel effects at the country level**

## Conclusions

## Conclusions

There are significant shifts in national patterns that can be associated with changes in funder policy and with the offerings of RSC and ACS

RSC took a significant lead in early open access provision for chemistry, particularly in the UK but has fallen back

National averages don't tell the full picture. Specific institutions show very different and quite specific patterns. There are differential policy effects

Recent changes are strongly driven by read and publish agreements with substantial shifts in publisher choice corresponding to introduction of deals.

There is evidence of concentration of publishing in chemistry with two large publishers taking up an increasing percentage. Should we be concerned about diversity?

## Chemistry has been following, not leading...

## Conclusions

...but maybe that is starting to change

## COKI Team

### Centre for Culture and Technology

- Cameron Neylon
- Lucy Montgomery
- Katie Wilson
- Chun-Kai (Karl) Huang
- Chloe-Brookes Kenworthy
- Tim Winkler

### Funding from

- Research Office at Curtin
- Faculty of Humanities
- School of Media, Creative Arts and Social Enquiry
- Andrew W. Mellon Foundation
- Arcadia

### Curtin Institute for Computation

- Richard Hosking
- Rebecca Handcock
- Aniek Roelofs
- Jamie Diprose
- Tuan Chien

### Educopia Foundation

- Katherine Skinner
- Rebecca Meyerson

@COKIProject - @cameronneylon

<http://openknowledge.community>

- **Subscribe to the COKI Newsletter**
- **View the public dashboards**

## Notes and further information

## Notes

Publisher percentage of chemistry is calculated based on the total count of publications assigned in Microsoft Academic to the Level 0 field of "chemistry" as the denominator with the total count of publications in Crossref where the publisher name field corresponds to the most common variant of the publisher name (*Royal Society of Chemistry (RSC)* and *American Chemical Society (ACS)* respectively).

Strictly these are not percentages and could in theory go to greater than 100%. They are also sensitive in magnitude to changes in the Microsoft field assignment. However it is a like for like comparison across the two publishers.

An analysis by journal would also be interesting but currently the metadata for journal identification is less reliable with journal names and choice of ISSN provided in Crossref metadata changing from year to year for both publishers. Future analysis deploying ISSN-L could address this.

**Publisher Count by year**

**Publisher % Green by year**

**Chemistry Count by year**

**Publishers % of Materials Field**

**Publishers % of Biology Field**

**Publisher percent of Materials Science by institution**

**Publisher percent of Biology by institution**