

# Qui Bono?

## Cumulative advantage in open access publishing

Open Repositories: 7<sup>th</sup> June 2022

---

David Pride, Matteo Cancellieri, **Petr Knoth**

Big Scientific Data and Text Analytics Group  
Knowledge Media Institute, The Open University  
United Kingdom



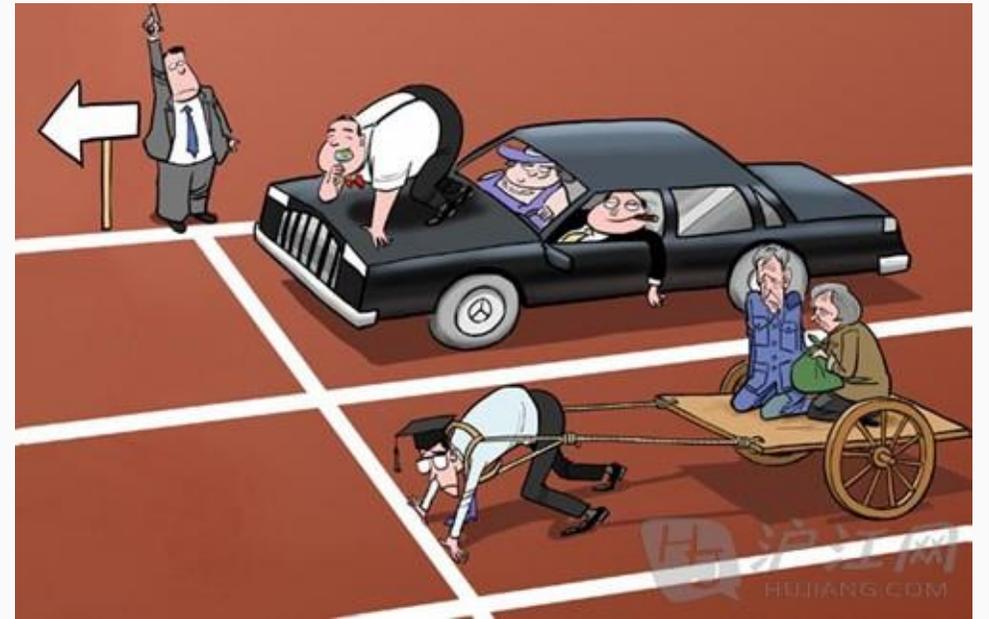
# The ON-MERRIT project

- EU H2020 project: October 2019 - March 2022
- Methods: Sociological, bibliometric and computational approaches
- Ensure that Open Science & RRI interventions contribute to a more equitable scientific system
- Distribution of rewards based on merit rather than privilege



# Premise

The premise of the *Matthew Effect*<sup>1</sup>  
*Those who begin with advantage  
accumulate more advantage over  
time and those who begin with  
disadvantage become more  
disadvantaged over time.*



<sup>1</sup>Merton, Robert K. "The Matthew Effect in Science: The reward and communication systems of science are considered." *Science* 159.3810 (1968): 56-63.

# Motivation and research questions

Identify whether barriers related to accessing research literature, such as being located at an institution with limited access to non-OA literature, affect the citation behaviour of scholars.

**Question 1:** *Do scholars located in less developed or at less prestigious institutions rely (**consume**) more OA because their access to subscription literature is limited?*

**Question 2:** *Do those who benefit from OA also **produce** more OA or are **production and consumption** independent?*

# Related work

- Recent work by Huang et al. (2020) investigated the production of OA literature around the globe based on institutions present in the THE rankings
- They found that, in 2017, the 100 top-ranked universities made 80–90% of their research publications OA.
- This figure is slightly higher than the result when looking at the data from the Leiden Dataset which suggests this figure is around 70%.

# Methodology

- Large-scale quantitative global study at the level of institutions
- 219m papers / 44k institutions / 190 countries
- OA status for each paper
- Additional segmentation based on economic and institutional ranking



# Data

## Data Sources

Source	Type
Microsoft Academic Graph	Paper metadata (authors, institutes, papers)
Unpaywall / CORE Discovery	Open Access status of each paper
Times Higher Education	Times World University Rankings (WUR)
CWTS Leiden	Leiden Ranking
World Bank	GDP / Economic indicators by country

# Data

## MAG Data Statistics - 219m papers from ~44k institutions

 **Author** - Individual author of a publication.

 **Institution** - Author institution is the institution the author was affiliated with at the time of publishing the publication.

 **Paper** - Publication title.

 **Journal** - Name of a scholarly journal.

 **Topic** - Research area, as identified by publisher keywords and Microsoft Academic algorithms.

 **Conference** - A venue where research is presented.

Microsoft Academic Graph metadata entities

# Terminology

Terminology	Description
<b>Production</b>	The publication of research papers by an entity (continent, country, institute)
<b>OA Production</b>	OA research papers produced by an entity.
<b>OA Production Rate</b>	The proportion of OA research papers produced by an entity.
<b>Consumption</b>	The use of a research paper by an entity as measured by citations in that entity's publications.
<b>OA Consumption</b>	The use of an OA research paper as measured by citations in that entity's publications.
<b>OA Consumption Rate</b>	The proportion of OA research papers used by an entity. In our work, we use as 'evidence of use' the act of citing OA research literature in manuscripts produced by an entity.

# Measuring production and consumption

Proportion of all OA scientific articles **produced** by an entity:

$$oa\_production / total\_production$$

Proportion of all OA scientific articles **consumed** by an entity

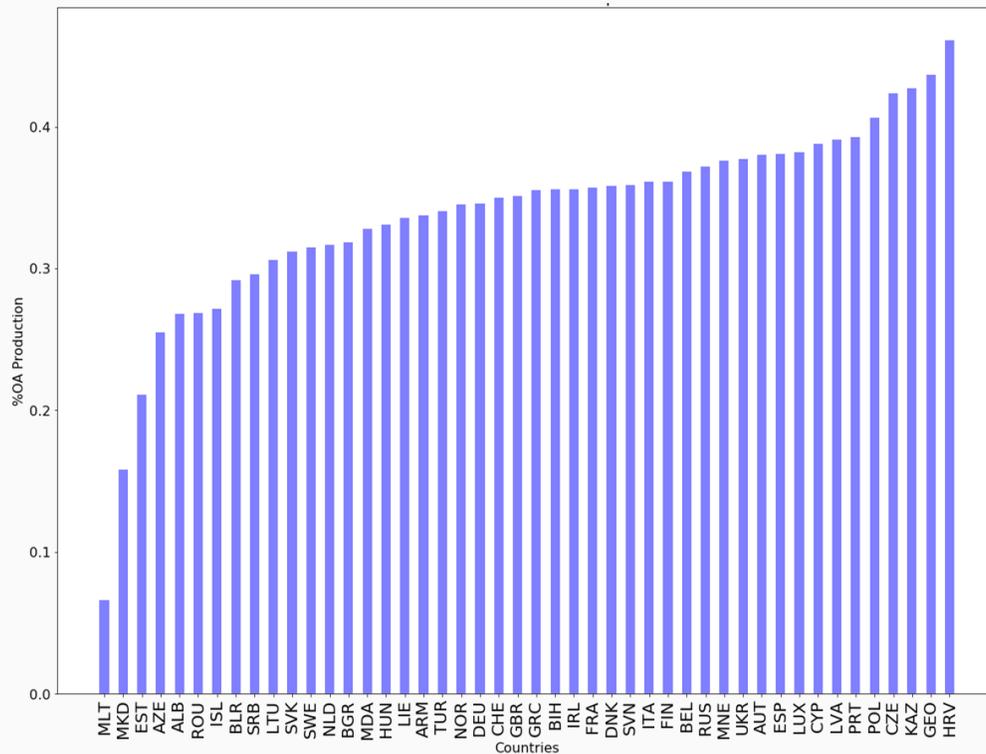
$$oa\_consumption / total\_consumption$$

Net OA Score for each entity

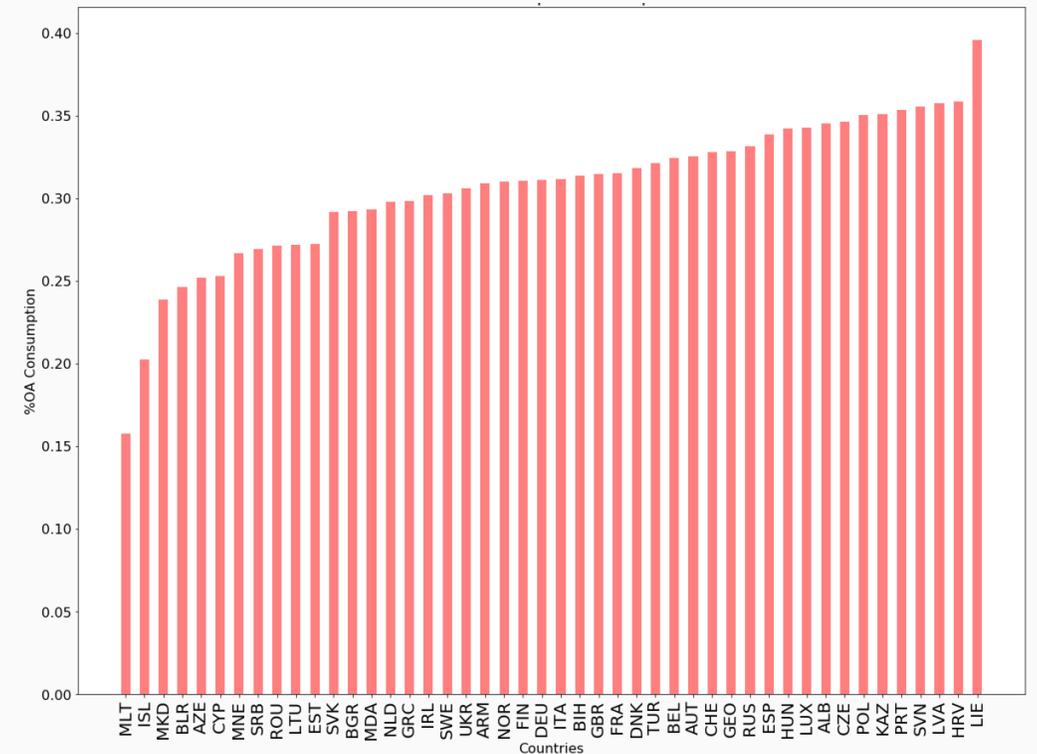
$$oa\_consumption / total\_consumption - oa\_production / total\_production$$

# Results

## Production and Consumption of OA papers in Europe 2016-2020



*Proportion of OA papers **produced** per country for Europe*

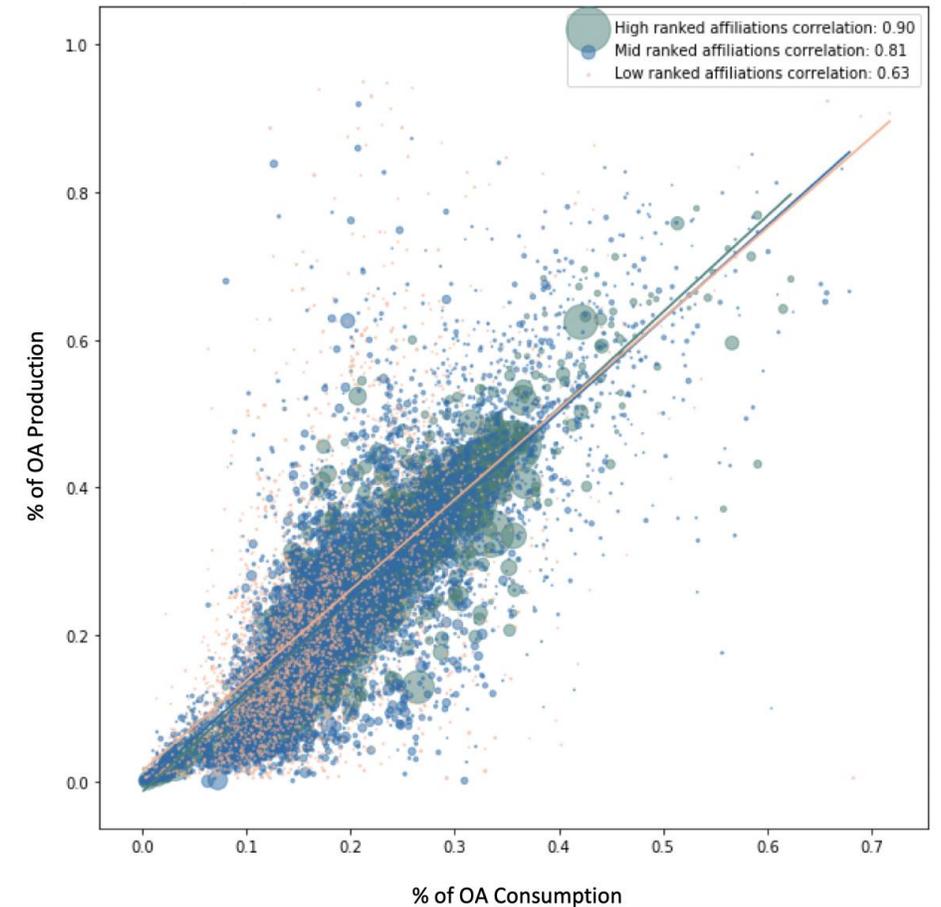


*Proportion of OA papers **consumed** per country for Europe*

# Results

## Correlation between OA production and consumption by institutional rank

- Strong correlation between the rates of OA production and consumption at institutional level
- Stronger correlation for higher ranked institutes.

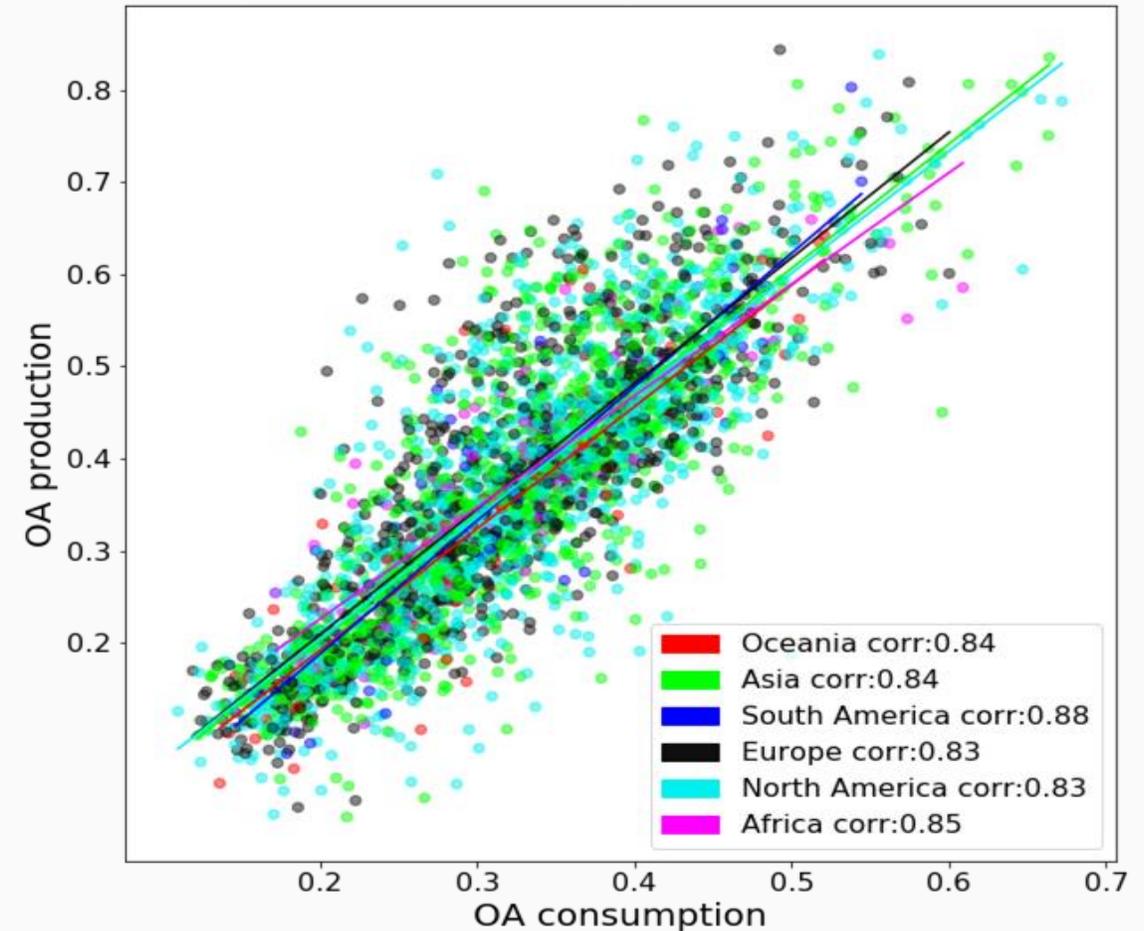


*Correlation between OA production and consumption ranking data from Times World University Rankings*

# Results

## Correlation between OA production and consumption

- Statistically significant correlation between the rates of OA production and consumption at the institutional level **across continents.**
- Little cross-continental difference
- Correlation has increased since 2010

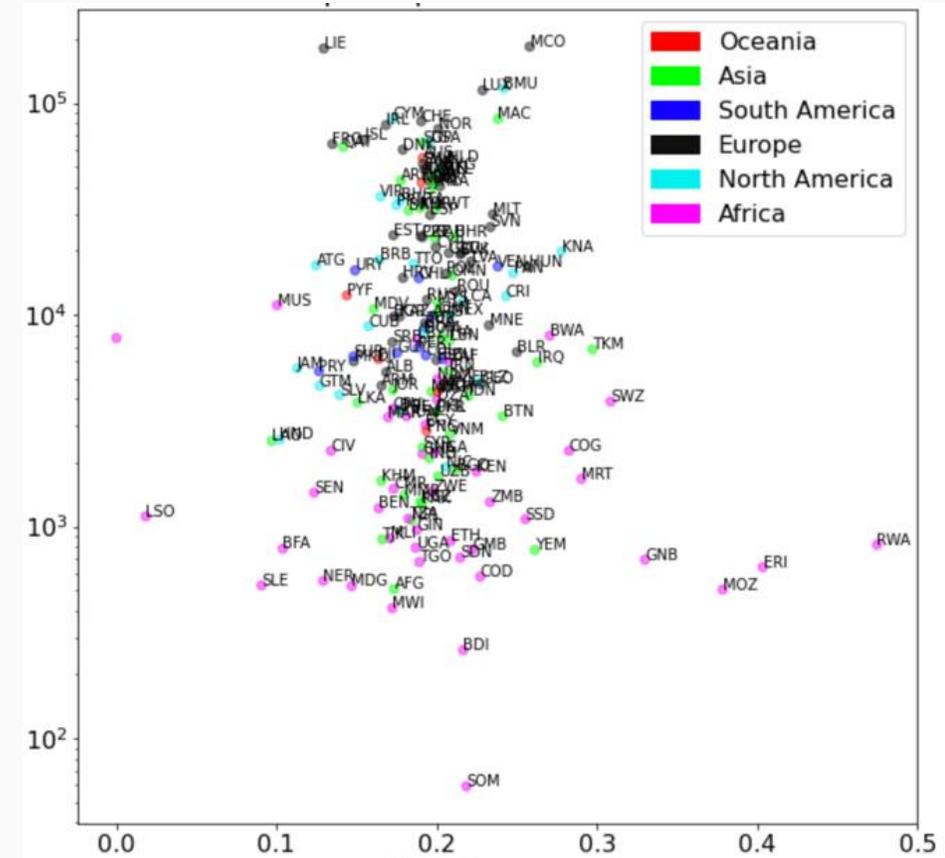


*Correlation between OA production and consumption by country and grouped by continent 2016-2020. (n=190, r=0.84)*

# Results

## Correlation between OA production and GDP per capita

- No correlation between the rates of OA production and GDP at the **country / continental level**
- Differences are at the institutional level
- Each continent has highly funded / ranked institutes.

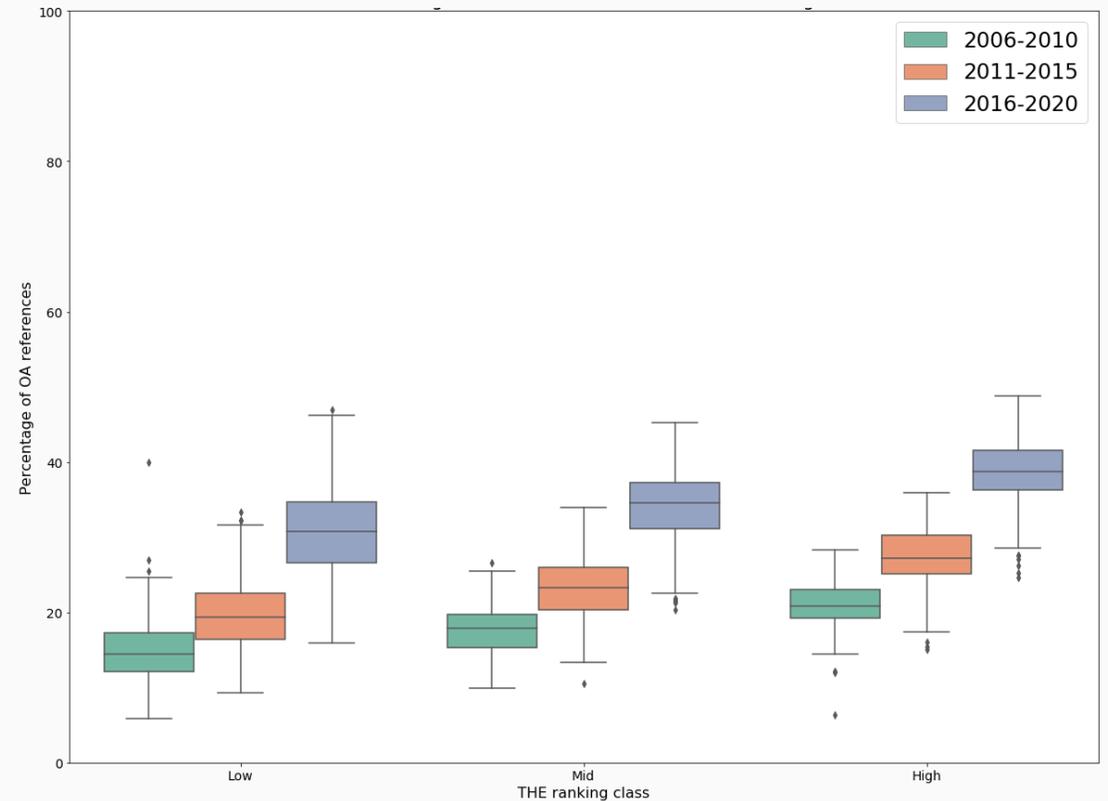


Correlation between OA production and country GDP per capita, grouped by continent 2016-2020. ( $n=190$ ,  $r=-0.02$ )

# Results

## Analysis of OA consumption and institutional rank

- Higher ranked institutes both produce and consume more OA
- Stronger correlation for higher ranked institutes.



*OA consumption vs institutional ranking using data from Times World University Rankings*

# Discussion

---

- Higher ranked / better funded institutes more able to capitalise on OA
- Benefitting from OA vs driving OA forward
- Analysing pre-Sci-Hub and post-Sci-Hub OA consumption rates, contrary to our intuition, we observed no “Sci-Hub effect” at a country level
- Our consumption measure quantifies how much one cites rather than how much one downloads

# Contribution

---

- Little correlation between GDP per capita and OA production and consumption
- More highly ranked institutions, when using THE rankings, are marginally greater producers and greater consumers of OA than lower-ranked institutions.
- OA adoption is not (or no longer) divided on the rich country vs. poor country perspective, but rather on the prestigious vs. not-prestigious institutional axis.

# Thank you!

---

<https://core.ac.uk>

