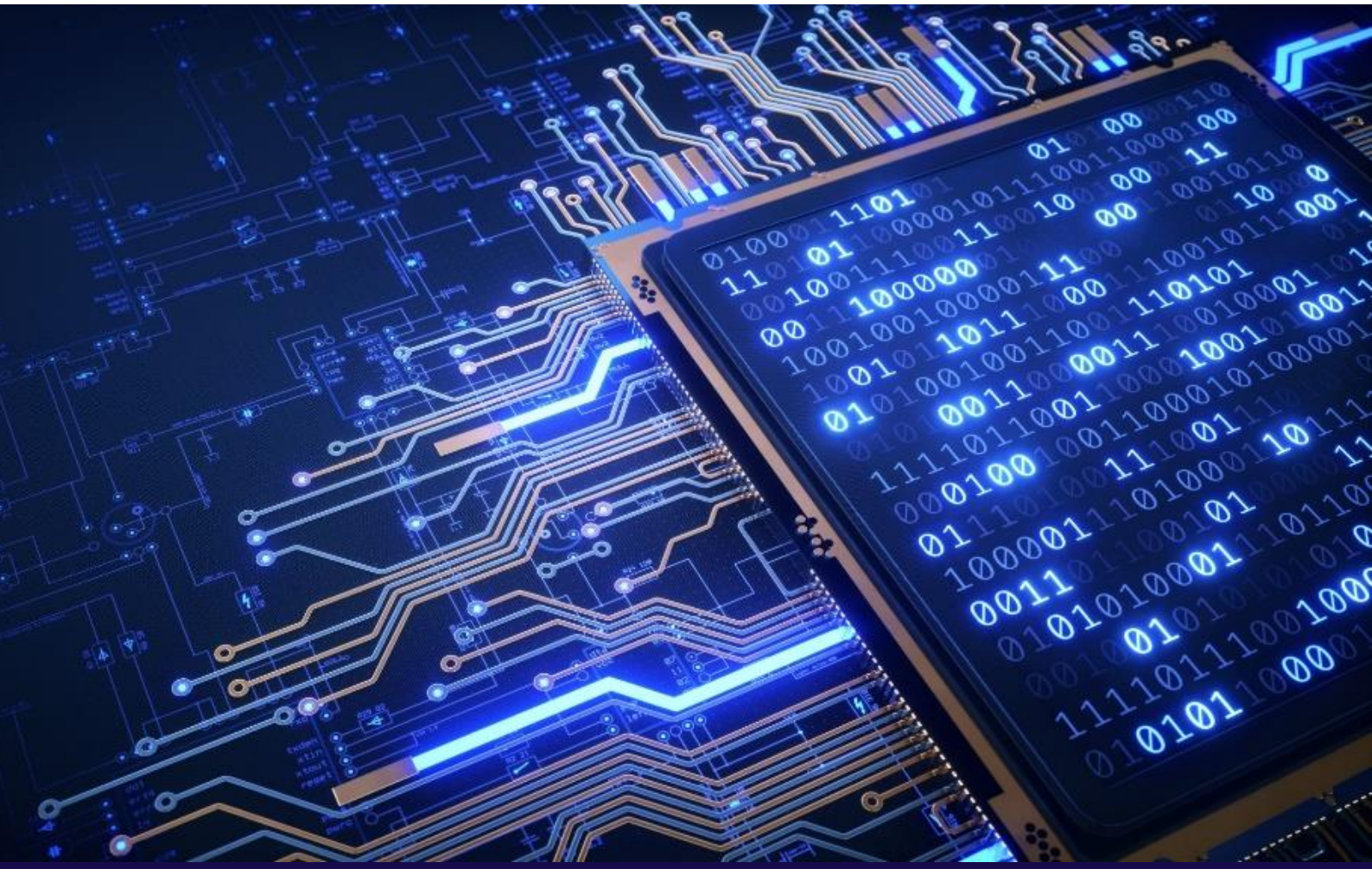


Technical report: Monitoring and evaluation of UKRI's Open Access Policy - Exploring the use of open data sources to inform baseline values

September 2024

Technical report: Monitoring and evaluation of UKRI's Open Access Policy - Exploring the use of open data sources to inform baseline values

Report commissioned by:
UKRI
https://www.ukri.org/

Contact:
Rachel Bruce, Head of Open Research
openresearch@ukri.org

Report authors:
Andrea Chiarelli, Frances Palmer
www.research-consulting.com

Bianca Kramer
sesameopenscience.org

Contact:
andrea.chiarelli@research-consulting.com

Report dated: September 2024

This work is licensed under a Creative Commons Attribution 4.0 International License.

Available on Zenodo at:
https://doi.org/10.5281/zenodo.12804855

Contents

1. Introduction 5
1.1 Background and scope of work 5
1.2 Methodology 6
2. Findings 10
2.1 Our dataset 10
2.2 Baseline data 10
2.3 Overall findings 12
2.4 Open access status by route 13
2.5 Local, national and international author collaboration 21
2.6 Citations 23
2.7 Views and downloads 28
2.8 Altmetrics 31
3. Future outlooks 34
Appendix A. Technical decision points 36

Acknowledgements

We acknowledge the support provided by the Curtin Open Knowledge Initiative (COKI) and Cameron Neylon, who provided access to a Google Big Query instance including the datasets of interest for this work as well as the environment to run the relevant code.

We also thank the UKRI team for their support and guidance throughout the project, including Claire Symeonides, Joanna Jacklin, Rachel Bruce and Sara Ball.

Glossary

Please note the following definitions, which are used throughout this report. Terms are sorted thematically.

Term	Definition
UKRI-funded	Journal articles arising from UKRI funding, identified via Gateway to Research or using Crossref Funder IDs
UK-affiliated	Journal articles with at least one UK-affiliated author, identified via OpenAlex
GtR without DOI	Outputs identified via Gateway to Research with no DOI available
GtR non-matched DOIs	Items identified via Gateway to Research with a DOI that is not matched in Crossref (either because the DOI is invalid, or it is a DOI from a different registrar, e.g. DataCite)
GtR only	UKRI-funded outputs identified via Gateway to Research but not via Crossref
GtR and Crossref	UKRI-funded outputs identified via both Gateway to Research and Crossref
Crossref only	UKRI-funded outputs identified via Crossref but not via Gateway to Research
Gold DOAJ[1]	Articles in journals included in DOAJ at time of publication, as detected by Unpaywall
Gold non-DOAJ	Articles in full Open Access journals not included in DOAJ, as detected by Unpaywall
Hybrid	Articles in subscription journals that are made free to read on the publisher's website at the time of publication with an open licence, as detected by Unpaywall
Bronze	Articles in subscription journals that are made free to read on the publisher's website without an open licence, as detected by Unpaywall; there may be a delay between publication and availability to read, and often articles can be removed unilaterally by the publisher; Bronze is not a permissible route under UKRI's Open Access policy
Bronze only	Articles available as Bronze for which the accepted or published version is not also available in a repository, as detected by Unpaywall
Bronze and Green (acc/pub)	Articles available as Bronze for which the accepted or published version is also available in a repository, as detected by Unpaywall
Green only (acc/pub) excl. Bronze	Articles for which the accepted or published version is available in a repository which are not also available as Bronze, as detected by Unpaywall
Green only (acc/pub)	Articles for which the accepted or published version is available in a repository, including those that are also available as Bronze, but excluding those that are available as Gold DOAJ, Gold non-DOAJ or Hybrid, as detected by Unpaywall
Green only (sub)	Articles for which only the submitted version is available in a repository, as detected by Unpaywall
Closed	Articles that are not free to read through any of the routes included in the analysis

Please note: the Gold and Hybrid routes, and the Green routes, respectively are named Route 1 and Route 2, respectively, in UKRI's Open Access policy.

[1] The Directory of Open Access Journals (DOAJ) is a community-curated list of open access journals, maintained by Infrastructure Services for Open Access (IS4OA). It was launched in 2003 with 300 OA journals and now lists over 20,400. Journals are listed on the DOAJ based on whether they meet a list of inclusion criteria that described transparency and best practice in scholarly publishing.

1. Introduction

1.1 Background and scope of work

Context

Research Consulting and Sesame Open Science were commissioned by UKRI to develop a set of baseline values, building on the report "Monitoring and evaluating the effectiveness of UKRI's open access policy: Principles, opportunities and challenges".

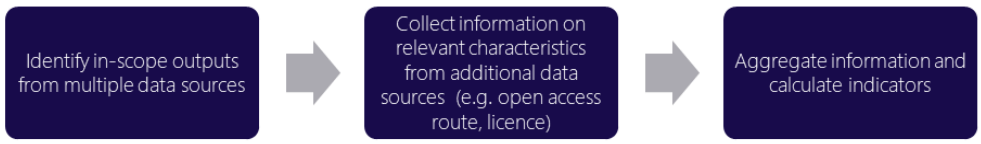
This analysis assesses the feasibility of answering a set of monitoring and evaluation (M&E) questions based exclusively on open data sources. Using open data communicates a commitment to open science and open sharing, in line with UKRI's wider open research strategy. The choice of sources presented reflects the analysis and discussion in the above-mentioned report; however, this is not meant as a formal recommendation to UKRI, since there has not yet been a full assessment of feasibility and comparison with other options (including proprietary options).

In this context, baseline values are intended as a set of indicators, in the form of counts or percentages, that reflect the state of open access publishing for a defined set of research outputs at a specific point in time. They provide a reference point against which future progress in open access publishing can be measured and evaluated.

We delivered this analysis by creating a record-level dataset that merges information from multiple open data sources. In practice, we collected a set of individual outputs within our scope of work (UKRI-funded and UK-affiliated articles) and harvested information on relevant characteristics of interest (see Figure 1). This information was then used to calculate appropriate indicators in the forms of counts or percentages.

The remainder of this document provides a detailed overview of our methodology, access to code and research data as well as a narrative description of findings.

Figure 1. High-level description of the analysis process.



Scope of work

Our analysis focused on peer-reviewed journal articles with Crossref DOIs published in the period 2012-2022. It aimed to examine the below areas, reflecting a subset of the M&E questions in the report Monitoring and evaluating the effectiveness of UKRI's open access policy: Principles, opportunities and challenges:

- levels of open access (OA) by route, licence and embargo period
- (inter)national collaborations
- citations, usage & altmetrics

Our analysis also included a split of data by publisher and research organisation, which is not covered in this report. The split of data by publisher was one of the recommended M&E questions in the report Monitoring and evaluating the effectiveness of UKRI's open access policy: Principles, opportunities and challenges. Given it sheds light on only one aspect of the open access publishing landscape, this data will be published later as part of the 2024 Light Touch Review of the UKRI open access policy

for research articles. It will be presented alongside other evidence, including stakeholder views via a survey and focus groups, enabling a more rounded picture to be drawn. The split of data by research organisation will inform UKRI's approach for monitoring compliance as part of its terms and conditions with research organisations. This is another piece of ongoing work, and UKRI will undertake further discussion with research organisations before this data is published.

The baseline data for our assessment is what we described as the 'denominator' in Annex A to our previous M&E work for UKRI. This refers to the set of journal articles that are either:

- Outputs of research funded by UKRI; or
- Outputs of research authored by an individual affiliated with a UK-based organisation.

In our record-level dataset, the baseline data is represented by each row: every in-scope journal article will have its dedicated row and a number of columns mapped to different dimensions. Further information on the data underpinning our work is available below as part of our data and code availability statements.

1.2 Methodology

Data sources

The data sources listed in Table 1 informed our analysis; we have included dataset versions, as future iterations may lead to slightly different results due to updates in the underlying information. Please note that Unpaywall data and OpenAlex data overlap to an extent, but not fully as of the date of this study (December 2023). The key reason for using Unpaywall as a distinct data source is that this includes key information (in the field 'oa_date') that underpins the analysis of embargo periods, which is currently not available via OpenAlex.

Table 1. Data sources for our analysis.

Area	Data sources	Dataset version
UKRI-funded publications	Gateway to Research	2023-11-14 (last update 2023-09)
	Crossref	2023-10-31
UK-affiliated publications	OpenAlex	2023-10-18
OA classification	Unpaywall	2023-11-27
Author affiliations	OpenAlex	2023-10-18

High-level approach and key decision points

Our overarching methodology based on the data sources outlined in Table 1 is presented in Figure 2 via a flowchart. The flowcharts are presented in human-readable format, to ensure that the process followed is clear and transparent. A more technical overview of the process is outlined as part of the code shared (see code availability statement), which includes comments.

Figure 2 is complemented by the decision points listed and described in Appendix A, which provides an overview of detailed methodological challenges and how these have been overcome.

For clarity, we highlight that wherever this document refers to 'Year', this is to indicate the year of publication.

Code availability statement

This research has been executed using the Google Big Query platform provided by the Curtin Open Knowledge Initiative (COKI).[2] Code in SQL language underpins our analysis and is available via a dedicated GitHub repository: https://github.com/bmkramer/ukri_oa_baseline

Code is annotated to enable reuse and assessment and has been split into a set of standalone portions that perform the analysis tasks required.

Data availability statement

A record-level dataset underpinning our analysis is available in csv format via Zenodo at: https://doi.org/10.5281/zenodo.12801805.

Limitations

The key limitation of this work is that we could only identify and analyse outputs that are included in the databases listed in Table 1. As a result, some publications may have been excluded, for example journal articles resulting from UKRI funding that (i) are not included in Gateway to Research; and (ii) are not attributable to UKRI funding via structured Crossref metadata.

Furthermore, the current analysis is restricted to Crossref DOIs to enable matching and retrieval of OA information via Crossref. Research outputs in Gateway to Research without Crossref DOIs are not included in our assessment, and the same applies to UK-affiliated publications in OpenAlex without Crossref DOIs.

An overall limitation of this work is that while we are presenting data as time series (for publication years 2012-2022), this reflects the current state of OA at the time of analysis, rather than at the time of publication of the research output. This is particularly relevant when embargoes (for Green OA) or 'moving walls'[3] (for the Bronze mechanism) are considered. This limitation does not apply when actual embargo times are analysed using information on OA dates, although additional limitations in this regard are noted below.

Other limitations with a minor impact on our analysis included the following:

- Green OA embargo times:
  - Unpaywall started including the OA date variable in August 2020. Therefore, zero embargo OA for repository versions can only be detected from this date onwards, and complete embargo information for the full publication year is only available for 2021 and 2022.
  - Not all records with a repository version have an OA date for that version. This mostly concerns versions in PubMed Central. In our dataset, 20% of 2021-2022

[2] The Curtin Open Knowledge Initiative (COKI), co-led by Cameron Neylon and Lucy Montgomery, seeks to be the world's leading hub for analysis and evaluation of Open Knowledge in higher education. Founded in 2017 at Curtin University in Perth, Australia, the COKI project team collaborate with national and international partners to create fresh insights into Open Knowledge practices around the world. COKI has developed the world's leading open knowledge data set, drawing together more than 12 trillion data elements, providing a comprehensive understanding of Open Knowledge practices and impact. The COKI project team has developed insights, analyses and tools which can enable universities to become Open Knowledge Institutions.

[3] The term 'moving walls' refers to the fact that publications may be made free to read only after a certain time period. The main implication of this is that publications may be listed under different OA routes over time, for example if the same analysis is run in the future.

articles that are only OA through a repository are missing the OA date, 98% of which are only in PubMed Central.

- OA licence:
  - UKRI's OA policy considers the use of the Open Government Licence (OGL) as compliant. However, Unpaywall does not identify this as a separate licence. Licence types 'public-domain', 'pd' and 'cc0' are included as compliant licences in our analysis, as are CC BY and CC BY-ND.[4]
- Hybrid OA, accepted or published version:
  - A small percentage of articles identified as Hybrid OA (OA via the publisher, with an open licence as defined by Unpaywall) are classified as 'accepted version' rather than 'published version'. In our full dataset, this concerns 4% of all Hybrid OA journal articles. For the purposes of this analysis, we have counted these articles as Hybrid OA.

We highlight that, due to the complexity of the approach and mix of data sources used, it is not straightforward to assess whether biases in the data are unintentionally affecting our analysis (e.g. underestimation of true numbers of outputs, overrepresentation of certain subjects). Although this technical report has not sought to provide an assessment regarding these potential challenges, our prior experience does not indicate obvious biases to be present in the data sources used.

[4] Please note that the OGL and CC BY-ND licences are allowed by UKRI OA policy as exceptions to the CC BY licence. An OGL licence can be used when a research article is subject to Crown Copyright, whereas the CC BY-ND licence is allowed on a case-by-case basis.

Figure 2. Identification of research outputs (denominator) and OA classification.

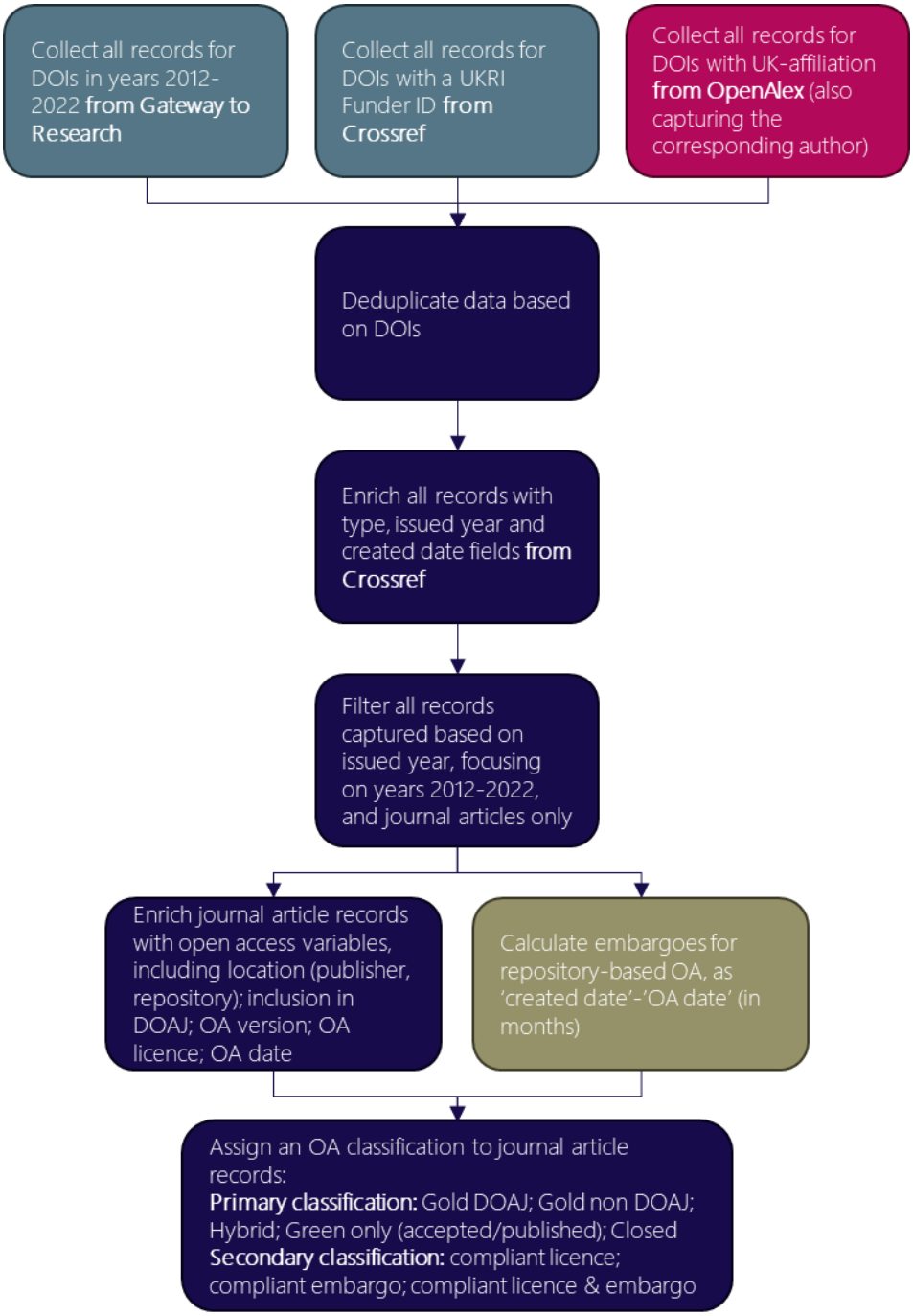
Legend:

UKRI-funded outputs

UK-affiliated outputs

All outputs

Repository OA only



2. Findings

2.1 Our dataset

Coverage

Table 2 provides an overview of all articles identified as part of our analysis, split between UKRI-funded and UK-affiliated (which includes the UKRI-funded portion).

Table 2. Number of DOIs identified for UKRI-funded and UK-affiliated articles.

Year of publication	# of DOIs – UKRI-funded	# of DOIs – UK-affiliated
2022	37,902	224,905
2021	44,122	239,538
2020	42,618	228,567
2019	40,266	207,350
2018	38,332	197,172
2017	36,179	190,921
2016	34,930	188,585
2015	32,032	179,900
2014	27,178	165,863
2013	23,893	163,259
2012	19,598	152,774

2.2 Baseline data

Approach

The following sections cover our detailed findings, presenting charts as well as tables. These are accompanied by a descriptive narrative, which covers our results as well as trends over time.

We note that the narrative does not seek to discuss the possible reasons behind the findings (e.g. policy interventions, sector developments), in recognition of the fact that scholarly publishing takes place within a complex, multi-stakeholder and international landscape. As a result, a mix of quantitative and qualitative investigations in the form of contribution analysis would be the recommended approach to further investigate and discuss why specific results might have arisen.

High-level findings – UKRI-funded articles

Findings regarding the baseline data on UKRI-funded articles are available in Figure 3. We note the following:

- A key insight from Figure 3 is that Gateway to Research does not capture all UKRI-funded publications. In 2022, the proportion of UKRI-funded articles available via Crossref was 54.0%, 37.8% of which had DOIs in both Gateway to Research and Crossref and 16.2% in Crossref alone. For those not tagged as UKRI-funded via Crossref, 41.1% were in Gateway to Research only with a Crossref DOI; 3.7% were in Gateway to Research without a DOI; and 1.2% were in Gateway to Research with a

DOI from a different registrar than Crossref (e.g. DataCite or mEDRA) or with an invalid DOI (i.e. a DOI that would not resolve).

- The total number of UKRI-funded articles in our dataset increased from 19,598 in 2012 to 37,902 in 2022, an increase of 18,304 items, or 93.4%.
- Between 2012 and 2022, the number of UKRI-funded articles available via Crossref increased from 1,402 to 21,529, a very significant increase of 20,127 items.[5]

Figure 3. UKRI-funded articles.

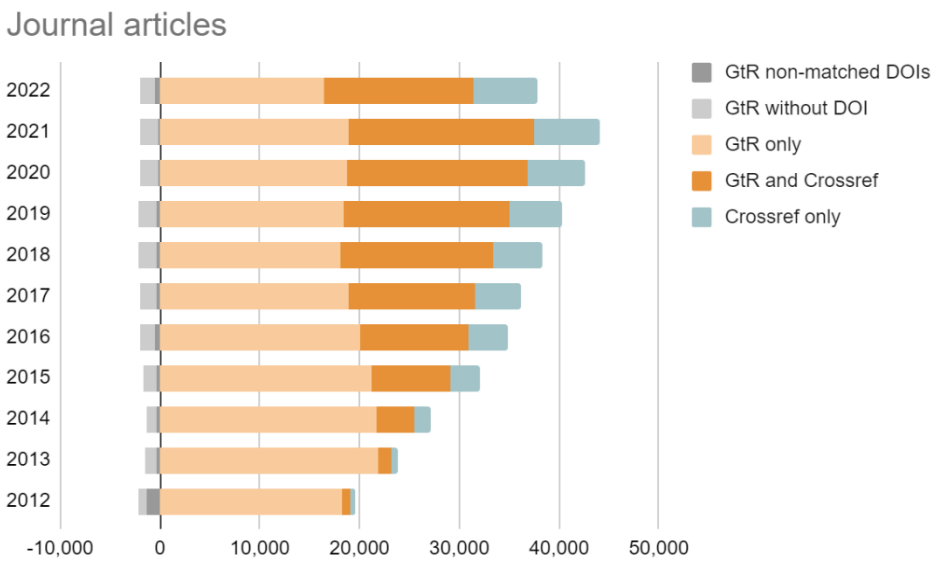
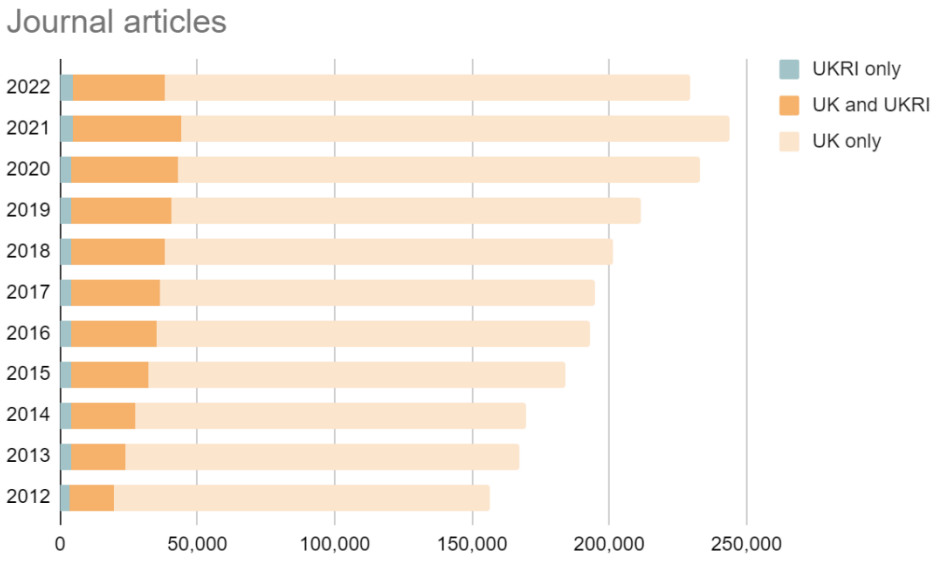


Figure 4. UK-affiliated and UKRI-funded only articles.



[5] Please note that this change is partly related with improved tagging of UKRI funding in Crossref data over time.

High-level findings – UK-affiliated articles

Findings regarding UK-affiliated articles are available in Figure 4. We note the following:

- The total number of articles in our dataset increased from 156,270 in 2012 to 229,378 in 2022, a 47% increase.
- In 2022, 83.5% of articles (191,476) in our dataset were UK-affiliated, 14.6% were both UK-affiliated and UKRI-funded (33,429) and 2.0% were articles).

2.3 Overall findings

Alignment levels across UKRI-funded and UK-affiliated articles

Findings regarding overall alignment with UKRI's requirements can be found in Table 3. It should be noted that UK-affiliated articles may be subject to expectations different from UKRI's Open Access policy, or no expectations at all. As a result, we discuss alignment with UKRI's requirements rather than compliance levels.

We note the following:

- Overall, alignment with UKRI's requirements has been growing steadily across both datasets, although UKRI-funded articles show higher proportions of compliant journal articles. This is to be expected, as grantees are held to more stringent requirements compared to the broader population of UK-affiliated articles.
- Alignment with UKRI's requirements is highest when the OA route only is considered (85.3% for UKRI-funded articles and 65.7% for UK-affiliated articles). Considering the year 2022, we highlight a lower level of alignment with UKRI licensing requirements (66.3% for UKRI-funded articles and 40.5% for UK-affiliated articles), whereas embargo periods are more often aligned with UKRI's requirements (73.8% for UKRI-funded articles and 55.3% for UK-affiliated articles).
  - From a methodological perspective, please note that stating that an article has a compliant licence also implies that it has been published through a compliant OA route (i.e. it isn't only a check of the licence type). Similarly, stating that an article has a compliant embargo, means that both the embargo period and the OA route are compliant.
- Full alignment with UKRI's requirements, including OA route, an appropriate licence and an appropriate embargo period has been improving over time. This is at 63.1% for UKRI-funded articles, but the level of alignment for all UK-affiliated articles is only at 38.7%, driven by low uptake of compliant licences.
- Monitoring alignment with specific UKRI requirements, rather than OA routes only, can provide additional detail on progress and practice change. For example, for UKRI-funded publications, alignment with UKRI's requirements for licences increased by 8.8% between 2021 and 2022, and alignment with requirements for embargo periods by 7.4%. In the same period, alignment with OA routes overall only showed a 2.4% increase.
- For the sake of clarity, we highlight that Table 3 as well as our overall analysis cover alignment with UKRI's current Open Access policy, which came into force in April 2022. For example, articles subject to the previous Research Councils UK Open Access policy would have allowed embargo periods for articles. As a result, the findings reported in this document are not implying that past articles were not in alignment with the relevant policy at the time, but simply that they are not aligned with today's UKRI requirements.

Table 3. Levels of alignment with UKRI's requirements.

Please note:

- 'Compliant licence' indicates that an article has both a compliant OA route and a compliant licence
- 'Compliant embargo' indicates that an article has both a compliant OA route and a compliant embargo period

UKRI-funded journal articles

Year	OA route only	Compliant license	Compliant embargo	Compliant license & embargo
2022	85.3%	66.3%	73.8%	63.1%
2021	82.9%	57.5%	66.4%	54.5%
2020	81.4%	52.5%	59.4%	49.5%
2019	79.3%	48.4%		
2018	80.7%	51.0%		
2017	80.6%	49.7%		
2016	79.3%	47.9%		
2015	70.5%	41.0%		
2014	62.7%	35.1%		
2013	50.8%	23.0%		
2012	41.7%	14.7%		

UK-affiliated journal articles

Year	OA route only	Compliant license	Compliant embargo	Compliant license & embargo
2022	65.7%	40.5%	55.3%	38.7%
2021	65.9%	35.1%	50.0%	33.3%
2020	63.7%	30.7%	43.2%	28.7%
2019	60.1%	26.3%		
2018	59.3%	25.9%		
2017	58.1%	24.4%		
2016	54.3%	22.4%		
2015	47.0%	20.0%		
2014	40.9%	17.1%		
2013	33.9%	12.8%		
2012	29.3%	9.6%		

2.4 Open access status by route

Approach

We note that the charts below include a detailed split between different categories of Gold and Green OA. For the purpose of the narrative, Gold and Green sub-categories have been merged, but these remain available as separate options in the underlying data.

For more information on the definitions used for OA pathways in our data, please see the Glossary at the beginning of this document.

In the next paragraphs, we have used the following approach to colour-code our findings:

- Decreasing trends are shown using a red icon: ↓
- Increasing trends are shown using green icons, as follows:
  - Increase by 0% to 33%: ↑

- Increase by 34% to 66%: ↑
- Increase by over 67%: ↑

We note that no increases over 67% have been detected as part of this work. However, the category has been retained to ensure compatibility with future monitoring work.

Publishing routes for UKRI-funded articles

Figure 5 and Figure 6 showcase all publishing routes for in-scope journal articles, including routes that UKRI does not consider compliant with its OA requirements (e.g. Bronze and only sharing non-peer-reviewed versions of a manuscript).

We note the following:

The total proportion of UKRI-funded articles (Figure 5) available via Bronze varies between 16.3% in 2012 and 2.8% in 2022, with 8.9% (2012) to 1.5% (2022) available as Bronze only (i.e. without an accepted/published version available via Green). It should be noted that these figures are affected by 'moving walls', as publications are made free to read only after a certain time period. For example, stating that 16.3% of journal articles from 2012 are available via the Bronze mechanism does not mean that these were made free to read at that point, but only that, as of today, they are free to read. Similarly, some publications from 2022 that are closed today may be available via the Bronze mechanism when checked in future years.

- The proportion of UKRI-funded articles (Figure 5) that are only available as a submitted version (i.e. the non-peer reviewed version) via Green OA varies between 11.4% in 2012 and 5.4% in 2022.
- Findings for UK-affiliated articles are shown in Figure 6. We note higher proportions being available via the Bronze mechanism for UK-affiliated articles in recent years compared to UKRI-funded articles (6.9% vs 2.3% for 2022). In contrast, for submitted versions shared via Green OA, proportions are slightly lower for UK-affiliated articles than for UKRI-funded articles across all years (4.4.% vs. 5.4% for 2022 and 5.1% vs 11.4% for 2012).

From Figure 7 onwards, all charts focus on journal articles published via OA routes that are compliant with UKRI's requirements. This means that articles that are only available via the Bronze mechanism or as non-peer-reviewed manuscript versions are excluded from the analysis. Articles that are available via the Bronze mechanism and for which the accepted/published version is also shared in a repository are included as Green OA in all subsequent charts.

Figure 5. Publishing routes for UKRI-funded articles.

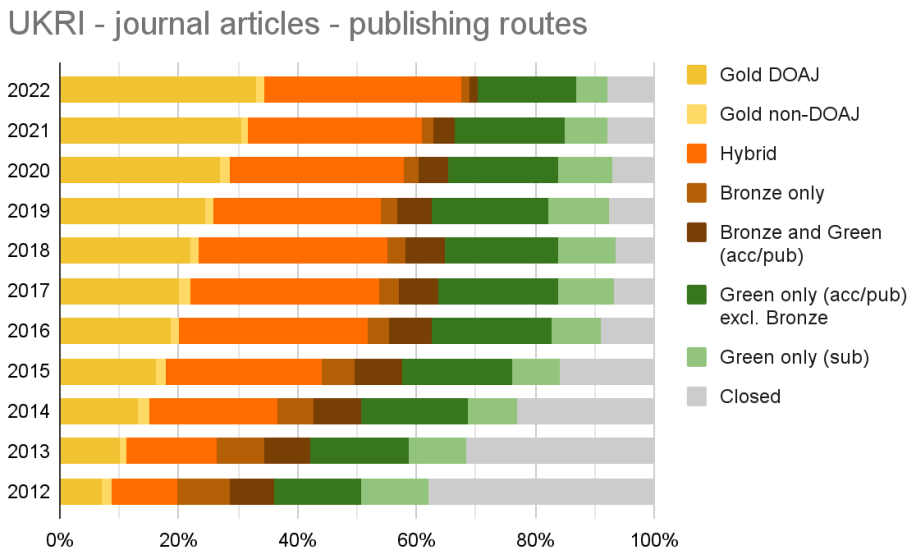
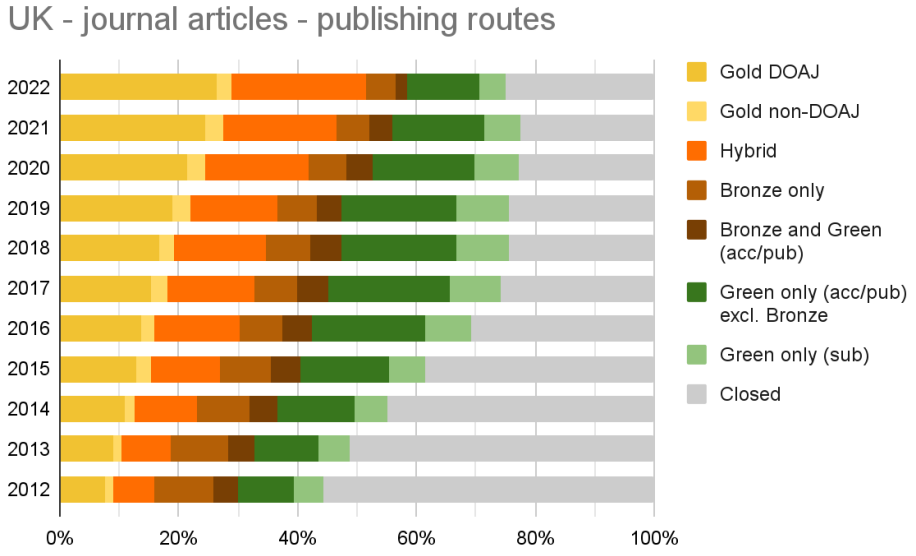


Figure 6. Publishing routes for UK-affiliated articles.



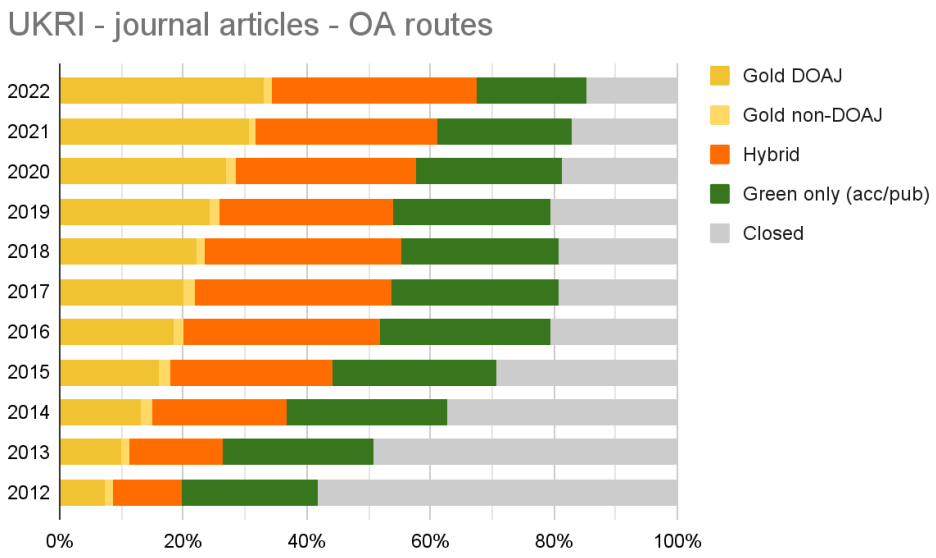
OA routes for UKRI-funded articles

Findings regarding the proportion of UKRI-funded articles available via an OA route are available in Figure 7. We note the following:

- The overall proportion of journal articles available via an OA route has increased by 43.6% between 2012 (41.7%) and 2022 (85.3%). ↑
- The proportion of journal articles available via Gold OA has increased by 25.7% between 2012 (8.6%) and 2022 (34.3%). ↑
- The proportion of journal articles available via Hybrid OA has increased by 22.1% between 2012 (11.1%) and 2022 (33.2%). ↑
- The proportion of journal articles available via Green OA has decreased by 4.2% between 2012 (22.1%) and 2022 (17.9%). ↓ However, it should be noted that these

figures are affected by longer embargoes (over 9-12 months) still in place for 2022 articles at the time of this study.

Figure 7. OA routes for UKRI-funded articles.



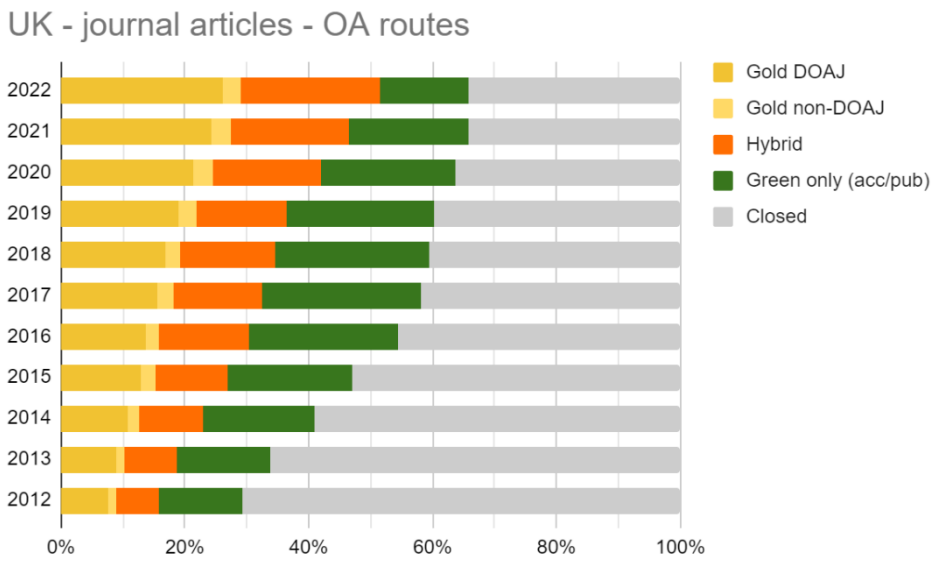
OA routes for UK-affiliated articles

Findings regarding the proportion of UK-affiliated articles available via an OA route are available in Figure 8. We note the following:

- The overall proportion of journal articles available via OA has increased by 36.4% from 2012 (29.3%) to 2022 (67.7%). ↑
- The proportion of journal articles available via Gold OA has increased by 19.9% from 2012 (9.0%) to 2022 (28.9%). ↑
- The proportion of journal articles available via Hybrid OA has increased by 15.6% from 2012 (6.8%) to in 2022 (22.5%). ↑
- The proportion of journal articles available via Green OA has increased by 0.8% from 2012 (13.5%) to 2022 (14.3%). ↑ As highlighted in the previous paragraph, it should be noted that these figures are affected by longer embargoes (over 9-12 months) still in place for 2022 articles at the time of this study.

A growing trend is seen within both UK-affiliated and UKRI-funded articles, however the proportion of articles made available via an OA route was consistently higher for UKRI-funded authors than for the UK overall, with a difference of 19.6% between the 2022 figures. This difference is for the largest part attributable to Hybrid OA (10.7% difference), with Gold OA and Green OA accounting for 5.3% and 3.6% of the difference, respectively.

Figure 8. OA routes for UK-affiliated articles.



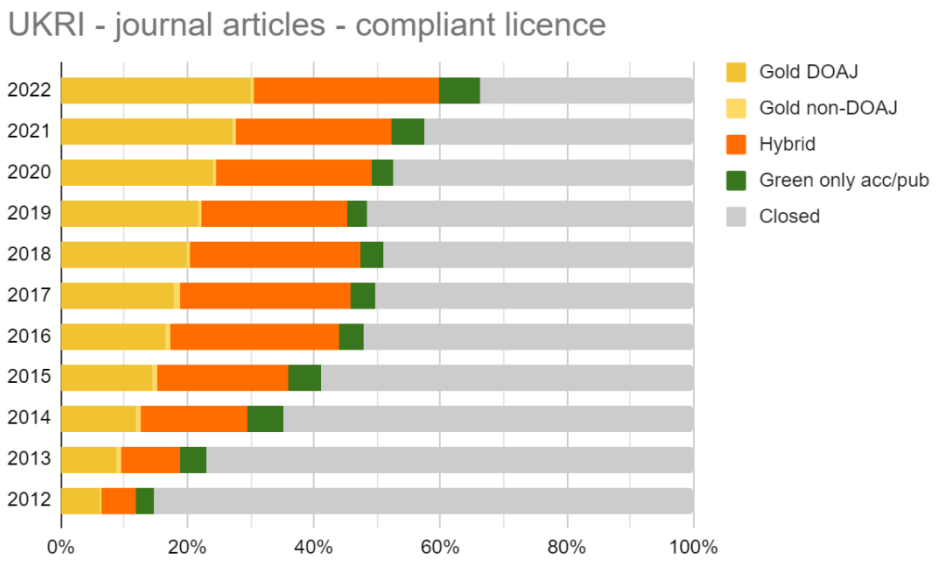
OA status for UKRI-funded articles with a compliant licence

Findings regarding UKRI-funded articles available via OA with a compliant licence are available in Figure 9. For the purposes of this work, a compliant licence was defined as one of: CC BY, CC BY-ND, CC0/PD.

We note the following:

- The overall proportion of journal articles with a compliant licence available via OA has increased by 51.6% from 2012 (14.7%) to 2022 (66.3%). ↑
- The proportion of journal articles with a compliant licence available via Gold OA has increased by 24.2% from 2012 (6.4%) to 2022 (30.6%). ↑
- The proportion of journal articles with a compliant licence available via Hybrid OA increased by 23.8% from 2012 (5.5%) to 2022 (29.3%). ↑
- The proportion of journal articles with a compliant licence available via Green OA has increased by 3.6% from 2012 (2.8%) to 2022 (6.4%). ↑

Figure 9. OA status for UKRI-funded articles with a compliant licence.



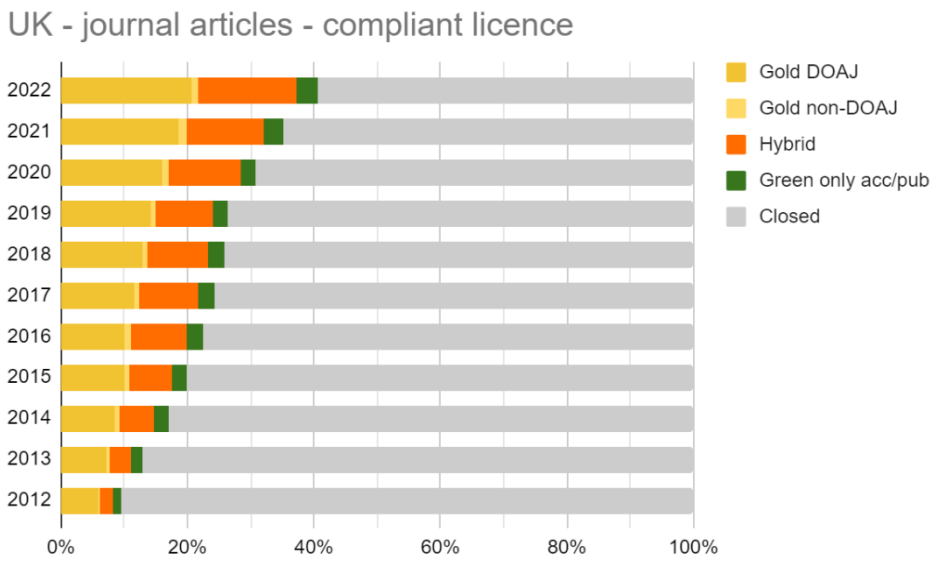
OA status for UK-affiliated articles with a compliant licence

Findings regarding UK-affiliated articles with a compliant licence made available by an OA route can be seen in Figure 10. We note the following:

- The proportion of UK-affiliated articles with a compliant licence made available via OA increased by 30.9% from 2012 (9.6%) to 2022 (40.5%). ↑
- The proportion of articles with a compliant licence available via Gold OA has increased by 15.5% from 2012 (6.2%) to 2022 (21.7%). ↑
- The proportion of articles with a compliant licence available via Hybrid OA increased by 13.2% from 2012 (2.1%) to 2022 (15.4%). ↑
- The proportion of articles with a compliant licence available via Green OA has increased by 2.1% from 2012 (1.3%) to 2022 (3.4%). ↑

Although a growing trend is seen in both Figure 9 and Figure 10, a larger proportion of UKRI-funded articles have a compliant licence each year compared to the broader set of UK-affiliated articles. There is a difference of 25.8% between OA articles with a compliant licence in 2022, when UKRI-funded (66.3%) and UK-affiliated (40.5%) are compared. Such a difference is for the largest part attributable to Hybrid OA (13.7% difference in 2022), with Gold OA and Green OA accounting for an 8.8% and 3.1% difference, respectively.

Figure 10. OA status for UK-affiliated articles with a compliant licence.



OA status for articles with a compliant embargo period

Findings regarding articles made available by an OA route with a compliant embargo and with both a compliant licence and embargo are shown in Figure 11, Figure 12, Figure 13 and Figure 14.

We have not provided detailed commentary on these charts, as data on embargo periods for repository-based OA is only available from mid-2020 (see Limitations). This means the proportion of compliant embargo periods for repository-based OA observed for publication year 2020 will be an underestimation, and no information on embargoes for repository-based OA is available for earlier publication years.

As a result, the data on embargo periods presented below cannot be analysed as a data series as done for previous charts. The information on embargo periods pertaining to 2022, which is expected to be reliable, is discussed as part of Table 3 above.

For the sake of completeness, we acknowledge that the data pertaining to 2021 is also expected to be reliable. We have not commented on this as the difference compared to 2022 is not substantial.

Figure 11. OA status for UKRI-funded articles with a compliant embargo period.

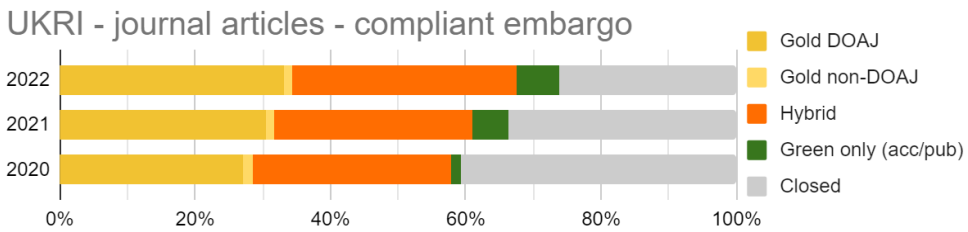


Figure 12. OA status for UK-affiliated articles with a compliant embargo period.

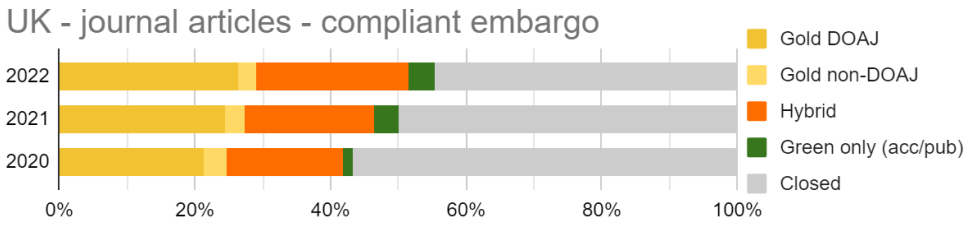


Figure 13. OA status for UKRI-funded articles with a compliant licence and embargo period.

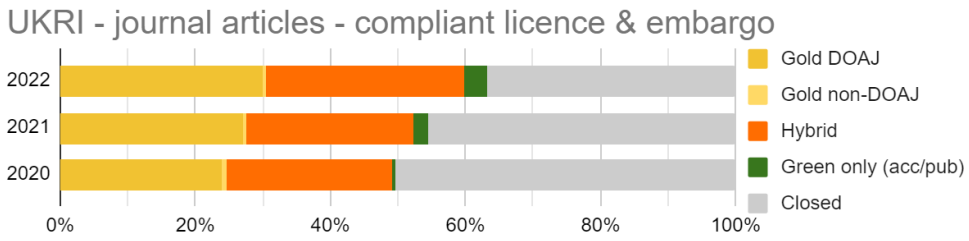
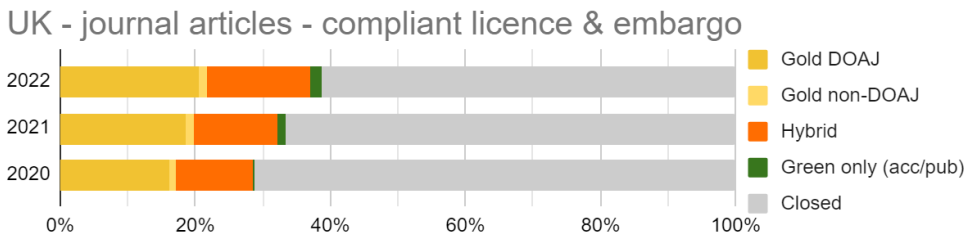


Figure 14. OA status for UK-affiliated articles with a compliant licence and embargo period.



2.5 Local, national and international author collaboration

Background

Figure 15 and Figure 16 illustrate the extent of author collaboration underpinning UKRI-funded and UK-affiliated articles, respectively.

We note the following:

- In the last five years, the majority of articles are a result of international collaboration in the case of both UKRI-funded and UK-affiliated articles. Over the whole observation window, the share of international collaborations appears to be increasing for UK-affiliated articles, whereas it appears to plateau for UKRI-funded ones.
- National collaborations form the smallest share of collaborations each year in the case of both UKRI-funded and UK-affiliated articles.
- In 2022, just under 25% of articles were the output of a single institution in the case of both UKRI-funded and UK-affiliated articles. This figure was close to 40% in 2012 and has progressively declined with the increase in international collaborations.

We highlight that the analysis of collaboration can only be carried out when author affiliation identifiers are included in OpenAlex data. As a result, 93%, of UKRI-funded articles we collected could be analysed for collaboration, and 100% of UK-affiliated articles (as the latter were collected based on presence of affiliation identifiers). In addition, in cases where not all authors of an article have structured affiliation identifiers, the results are potentially an underestimation of national and international collaborations.

Figure 15. Collaboration levels for UKRI-funded articles.

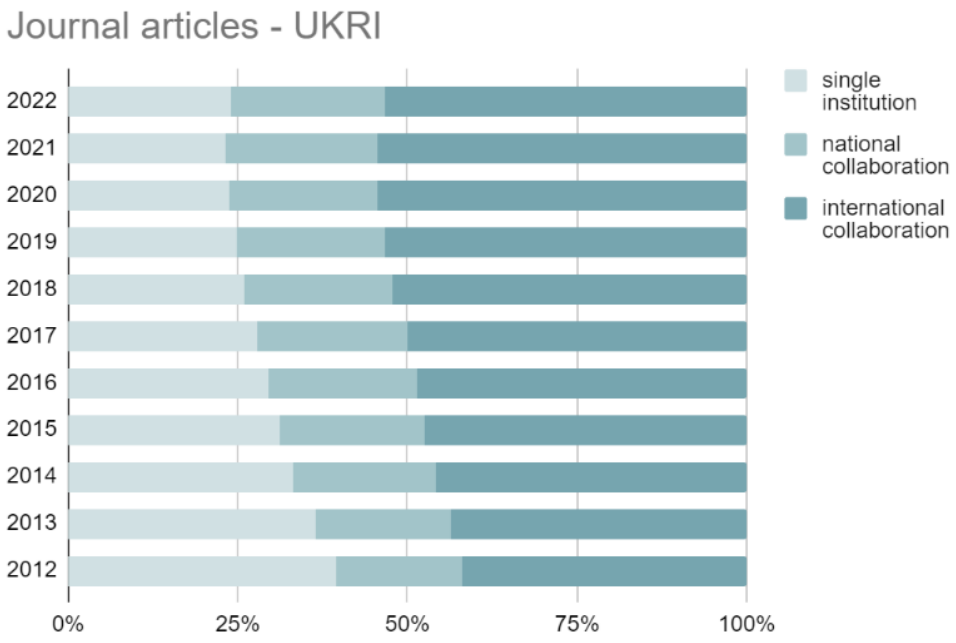
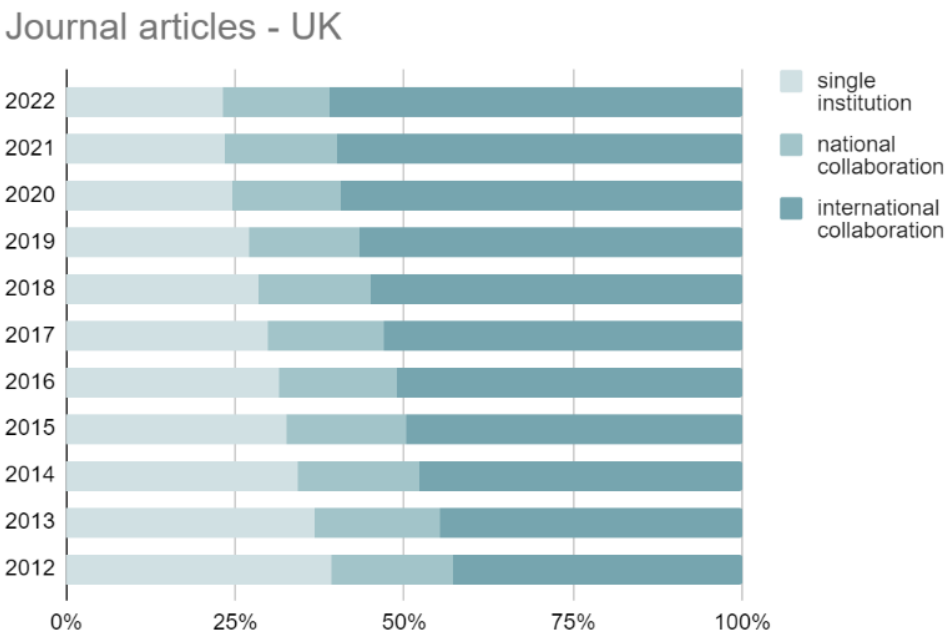


Figure 16. Collaborations levels for UK-affiliated articles.



Alignment with UKRI's requirements of outputs arising from different types of collaborations

Figure 17 and Figure 18 show the proportion of UKRI-funded and UK-affiliated articles that are aligned with UKRI OA policy requirements by type of collaboration for 2022. The highest rate of alignment with UKRI's requirements is for national collaboration in both cases (69.0% and 43.1% for UKRI-funded and UK-affiliated articles, respectively), whereas the lowest rates of alignment are for articles that are the output of a single institution (60.9% and 31.3%).

Although our research could not investigate the reasons for these results, it may be hypothesised that access to transformative agreements or OA publishing funds is likely to be higher if multiple authors from a high-income country are involved.[6] Similarly, it can be hypothesised that articles from a single institution would have more limited access to transformative agreements or OA publishing funds. This is even more likely to be the case in the arts and humanities, where collaboration is often limited and single-author contributions are more frequent,[7] which clearly limits access to transformative agreements or OA publishing funds for each article.

The lower rate of alignment for articles resulting from international collaborations compared to those resulting from national collaborations may arise from international co-authors being subjected to different OA mandates or expectations regarding OA publication. A follow-up analysis looking at corresponding authors could shed more light on the effect of the URKI policy on (international) collaborations.

[6]This consideration is exemplified in the ESAC Market Watch. However, it should be noted that some high-income countries, such as the United States, have lower rates of transformative agreement compared with European countries, for example.
[7] See, for example, Thelwall and Maflahi (2022), which highlights that social sciences and humanities present lower rates of co-authorship compared to other disciplines.

Figure 17. Alignment with UKRI's requirements in the case of different types of collaborations; UKRI-funded articles (2022).

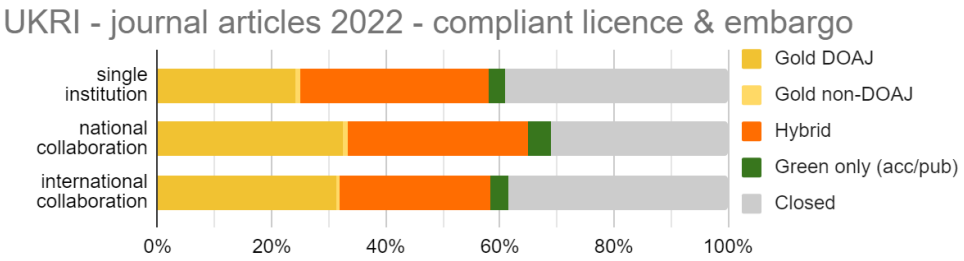
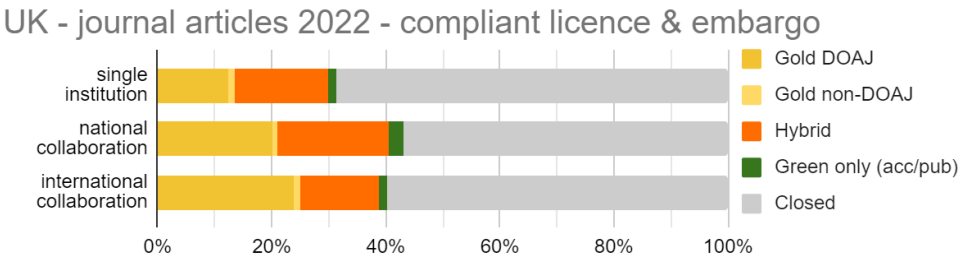


Figure 18. Alignment with UKRI's requirements in the case of different types of collaborations; UK-affiliated articles (2022).



2.6 Citations

Background

Citation information was captured via OpenAlex. Table 4 and Table 5 provide an overview of mean citation counts for UKRI-funded and UK-affiliated articles, per OA type. We note the following:

- UKRI-funded articles obtain higher mean citations across all types of OA status when compared with UK-affiliated articles.
- When looking at the full range of publishing routes (Figure 28 and Figure 29), mean citation counts are highest for the 'Bronze and Green' category for both UKRI-funded and UK-affiliated articles. followed by the 'Hybrid" category in both cases.
  - The 'Bronze and Green' category refers to articles available as Bronze for which the accepted or published version is also available in a repository. Journals occurring most frequently in this category are from the fields of Astronomy/Astrophysics and Medicine, and mean citation counts in this category are highest for journals in the Nature Portfolio and in the field of Medicine. This typically concerns articles that are available on the publisher's website without an open licence (e.g. via a moving wall), and are also made available through a repository, like PubMed Central for Medicine.
  - Compared to the "Bronze and Green' category, mean citation counts are lower for articles only available as bronze OA (without a repository copy), and articles for which the accepted/published version is only available via a repository ("Green only excl Bronze").
- When articles with a compliant licence and embargo are considered separately (Figure 21 and Figure 22), mean citations remain higher for UKRI-funded articles. The Hybrid route is the one with the highest mean citations, followed by Green, then Gold DOAJ.
  - In this breakdown, the Green route includes both articles in the previously mentioned "Bronze and Green" category (with higher median citation counts) as well as the "Green only excl. Bronze" category (with lower median citation

counts). In other words, this group includes all articles for which an accepted/published version in a repository is the only version aligned with UKRI's OA policy (irrespective of whether the article is also available as Bronze OA).

- By plotting citation trends via histograms (exploratory analysis, not shown), it can be observed that, in all cases, there is a large number of articles with low citation counts, followed by rapidly decreasing numbers of articles as the number of citations increases. The use of histograms and/or box plots is recommended for monitoring and evaluation purposes, as mean values are strongly affected by outliers in the data (i.e. small number of articles with very high numbers of citations).

Table 4. Mean citation counts for UKRI-funded articles by publishing route (2021).[8]

OA status	Total DOIs	Mean citations
Gold DOAJ	13,463	13.5
Gold non-DOAJ	472	6.1
Hybrid	12,981	17.7
Bronze only	837	12.3
Bronze and Green (acc/pub)	1,510	40.8
Green only (acc/pub) excl. Bronze	8,090	16.3
Green only (sub)	3,194	14.7
Closed	3,575	9.7

Table 5. Mean citation counts for UK-affiliated articles by publishing route (2021).

OA status	Total DOIs	Mean citations
Gold DOAJ	58,366	10
Gold non-DOAJ	7,261	4.1
Hybrid	45,518	13.4
Bronze only	13,457	5
Bronze and Green (acc/pub)	9,242	21.7
Green only (acc/pub) excl. Bronze	36,393	11.8
Green only (sub)	13,815	11
Closed	55,486	5.4

[8] This data is presented for 2021 as citations and usage grow over time. As a result, figures for 2022 were significantly lower at the time of analysis, although showing similar trends.

Figure 19. Citation distribution by OA status; UKRI-funded articles (2021).

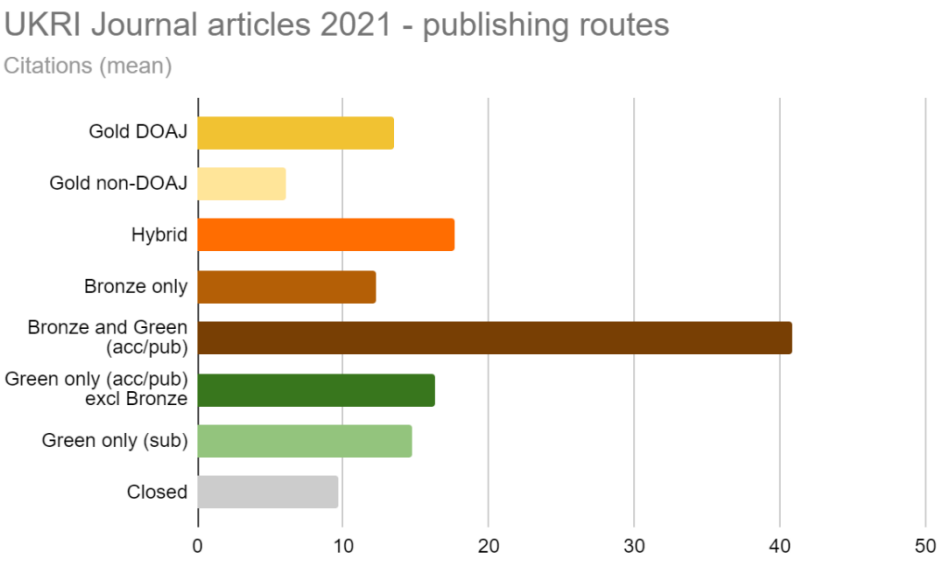


Figure 20. Citation distribution by OA status; UK-affiliated articles (2021).

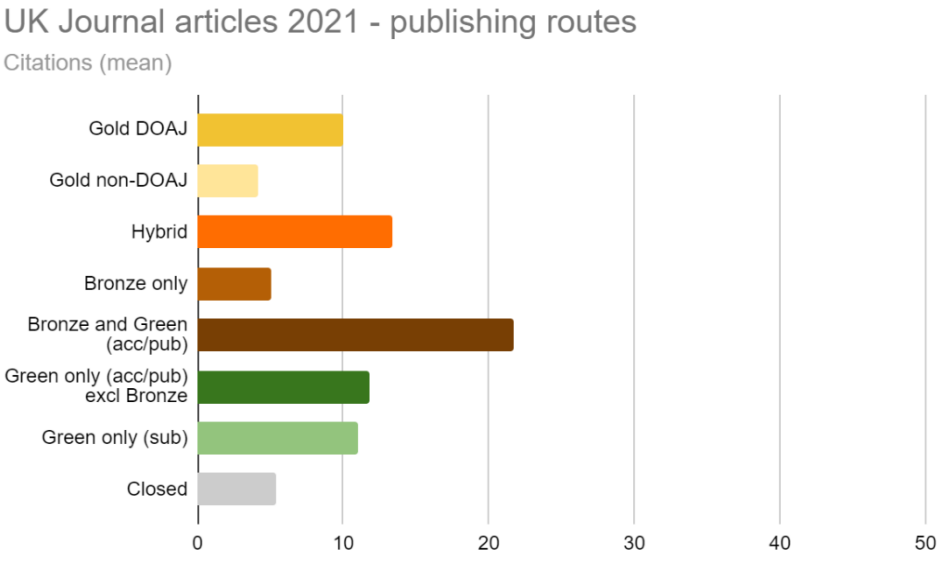


Figure 21. Citation distribution for UKRI-funded articles with a compliant licence and embargo period, by OA status (2021).

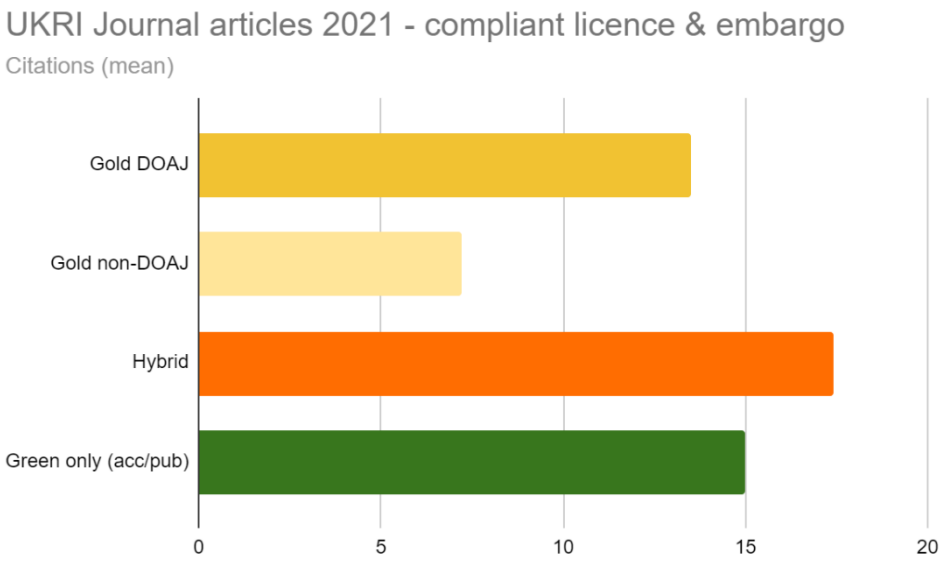
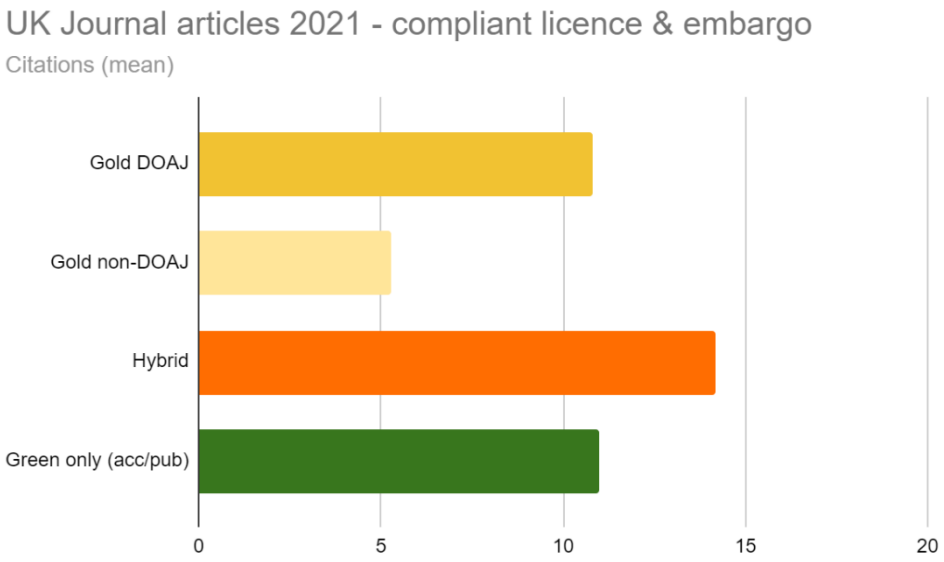


Figure 22. Citation distribution for UK-affiliated articles with a compliant licence and embargo period, by OA status (2021).



Subject classification for the analysis of citations in UKRI-funded and UK-affiliated articles with a compliant licence and embargo period

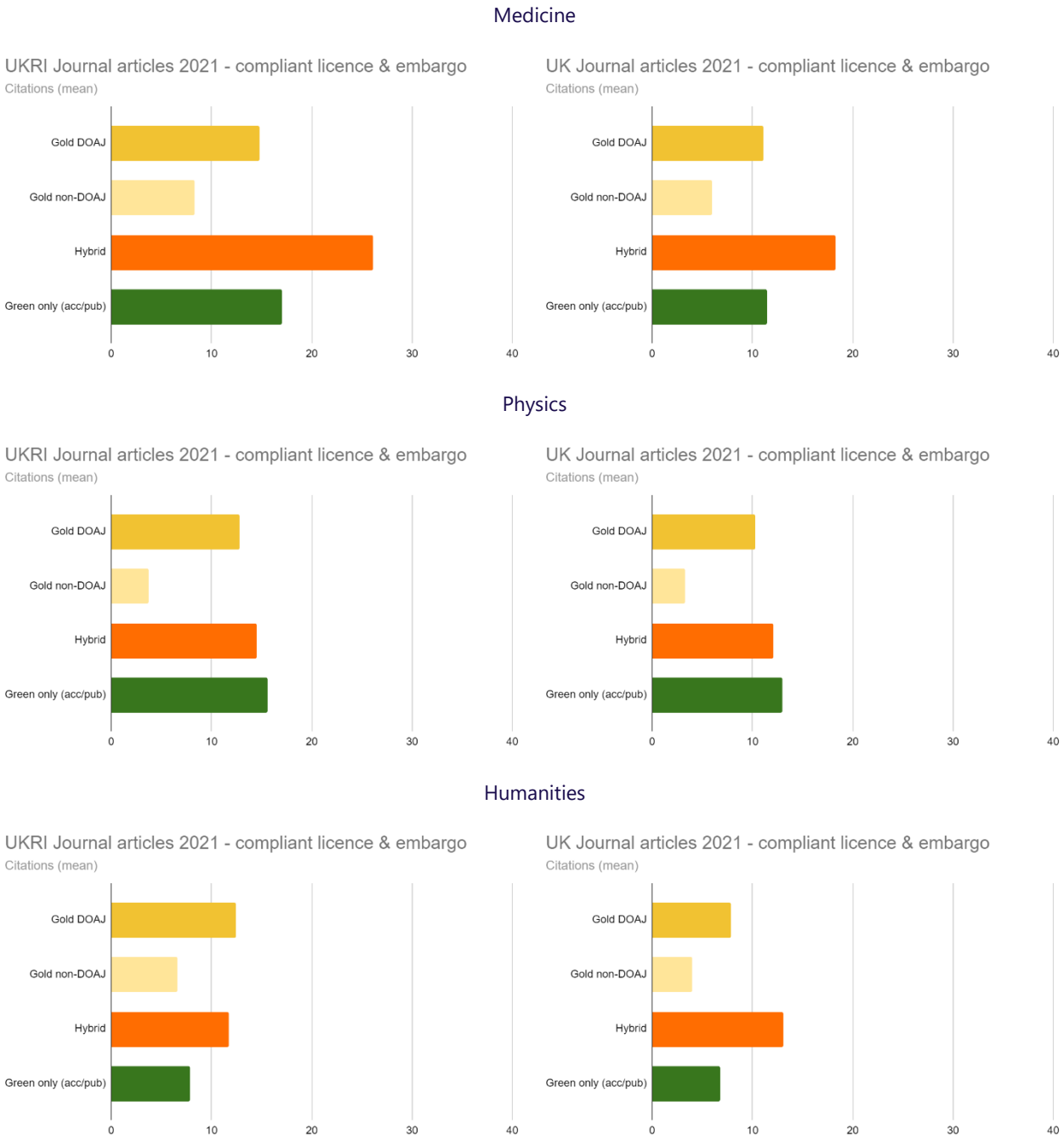
The aggregate figures regarding citations presented above do not account for potential disciplinary effects (i.e. differences in citation behaviour between disciplines). In general, this can be addressed through field-normalisation or by breaking down the data by discipline, as done here for three selected disciplines (Figure 23). Further information on the decision to break down the data by discipline is provided in Appendix A.

This analysis is enabled by 'Concepts' included in OpenAlex data, which are assigned at article level and mapped to 19 top-level Fields, one or more of which are attributed to each article. Notably, this approach is being deprecated to shift to the use of 'Topics' in OpenAlex, which similarly assigned at article level and are mapped to 4 Domains and 26 Fields (based on Scopus's ASJC categories). As a result, the findings of Figure 23 should be considered as a proof of concept of the approach only.

For UKRI-funded articles, the mean number of citations for the Hybrid and Green routes are highest in Medicine, followed by Physics and then the Humanities. Mean citations for Gold DOAJ articles are comparable across these fields.

Trends are broadly comparable for UK-affiliated articles in Medicine and Physics, but the data regarding Humanities suggests that those in Hybrid journals are more cited than other OA routes as well as when compared to UKRI-funded articles.

Figure 23. Example of subject classification for the analysis of citations in UKRI-funded and UK-affiliated articles with a compliant licence and embargo period (2021, sample fields).



2.7 Views and downloads

Background

Information on views and downloads was captured via IRUS UK, which provides institutional repository usage statistics in the United Kingdom. IRUS UK data covers two types of usage statistics: (i) downloads only (variable 'Unique_Item_Requests') or (ii) views and downloads (variable 'Unique_Item_Investigations'). Our analysis considers both views and downloads, as this offers a more comprehensive picture of usage.

We acknowledge the following limitations of this analysis:

- IRUS doesn't cover all UK repositories, and a full list of participants is available here.
- IRUS data does not cover downloads from publisher websites, which means that Gold/Hybrid usage is not fully captured as part of our analysis.

Table 6 and Table 7 provide an overview of mean download counts for UKRI-funded and UK-affiliated articles. We note the following:

- IRUS UK data only includes counts for items that have received at least one view or download in the period studied (January to December 2023). This is a significant feature of the data, as mean values do not account for articles that have not received any views or downloads.
- UKRI-funded and UK-affiliated articles obtain very similar mean usage counts across all types of OA status.
- Usage counts are highest for Green only articles in both cases. The second-highest counts are Non-OA / Closed articles, followed by Hybrid ones. This finding regarding Green OA articles is not surprising given that IRUS data focuses on institutional repositories. The significant interest in Green only versions is, however, notable, as these articles would not otherwise be available via the publisher's website without a subscription. Such a finding is an indication of the value of institutional repositories in enabling access to publisher literature.
- When the Non-OA / Closed category is broken down into the full range of non-compliant publication routes (Figure 24 and Figure 25), the impact of providing access to Green only articles is further articulated, with Green only articles showing significantly more usage than the other routes
- When articles with a compliant licence and embargo are considered separately (Figure 26 and Figure 27), mean downloads are broadly comparable for UKRI-funded and UK-affiliated articles. The Green only route is the one with the highest mean downloads, followed by Hybrid and Gold.
- By plotting citation trends via histograms (not shown), it can be observed that, in all cases except for Green only, there is a large number of articles with low download counts, followed by rapidly decreasing numbers of articles as the number of download increases. In the case of Green only articles, a peak is still observed regarding the number of articles with low download numbers; however, the number of articles with larger download counts decreased more slowly compared to other OA routes (i.e. there are more Green only articles with high download numbers compared to other OA routes). Similarly to the observations made regarding citations, we recommend that histograms and/or box plots are used for monitoring and evaluation purposes.

Table 6. Mean usage counts for UKRI-funded articles by publishing route (publication year 2021).

OA status	Total DOIs	DOIs available via IRUS UK	Mean views and downloads
Gold DOAJ	13,463	7,878	15.7
Gold non-DOAJ	472	211	17.7
Hybrid	12,981	7,719	20.7
Bronze only	837	169	10.7
Bronze and Green (acc/pub)	1510	810	20.8
Green only (acc/pub) excl. Bronze	8,090	5,699	52.6
Green only (sub)	3,194	1,343	39.2
Closed	3,575	595	22.9

Table 7. Mean usage counts for UK-affiliated articles by publishing route (2021).

OA status	Total DOIs	DOIs available via IRUS UK	Mean views and downloads
Gold DOAJ	58,366	25,495	16.7
Gold non-DOAJ	7,261	1,668	27.2
Hybrid	45,518	22,169	22.7
Bronze only	13,457	791	13.4
Bronze and Green (acc/pub)	9,242	3,119	25
Green only (acc/pub) excl. Bronze	36,393	23,477	66.5
Green only (sub)	13,815	5,334	45.5
Closed	55,486	4,563	23.9

Figure 24. Usage distribution for UKRI-funded articles with a compliant licence and embargo period, by OA status (2021).

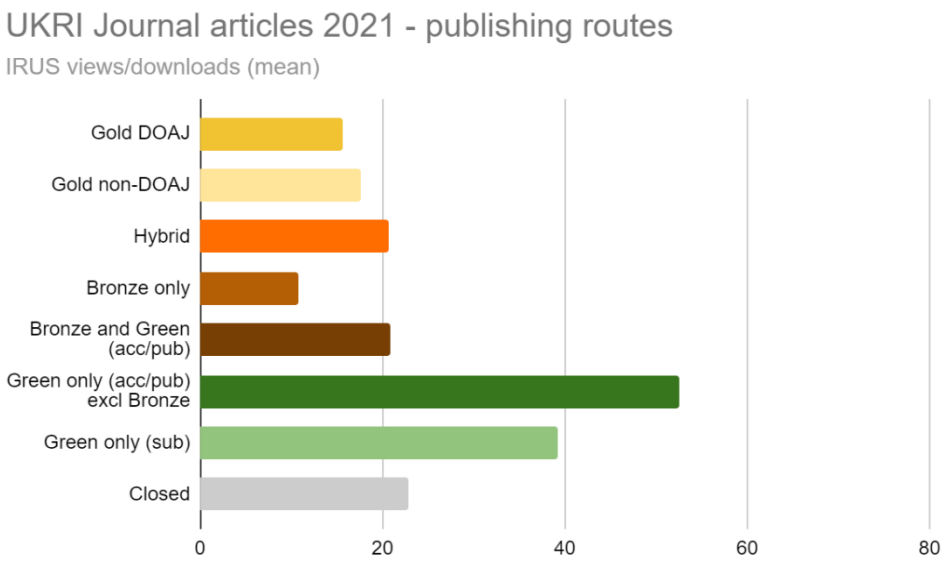


Figure 25. Usage distribution for UK-affiliated articles with a compliant licence and embargo period, by OA status (2021).

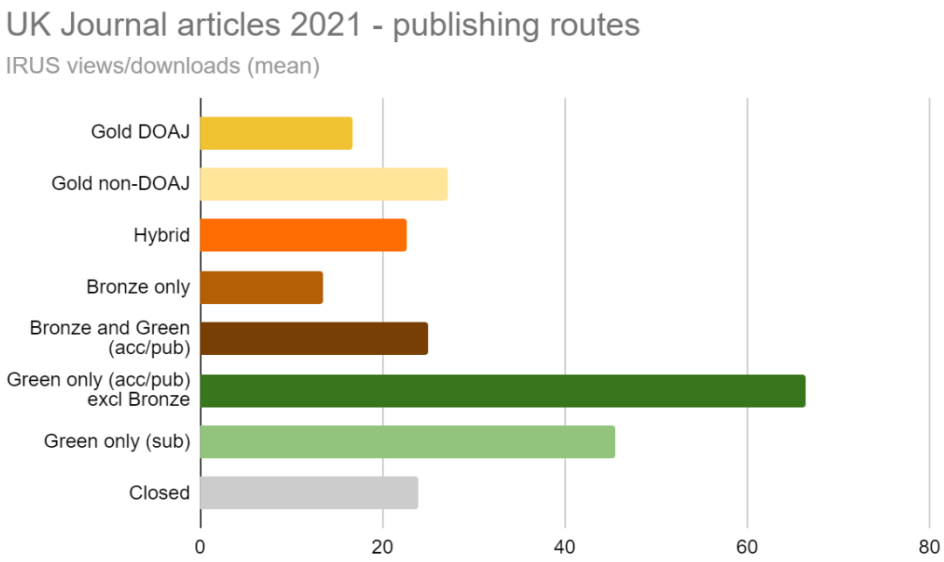


Figure 26. Usage distribution for UKRI-funded articles with a compliant licence and embargo period, by OA status (2021).

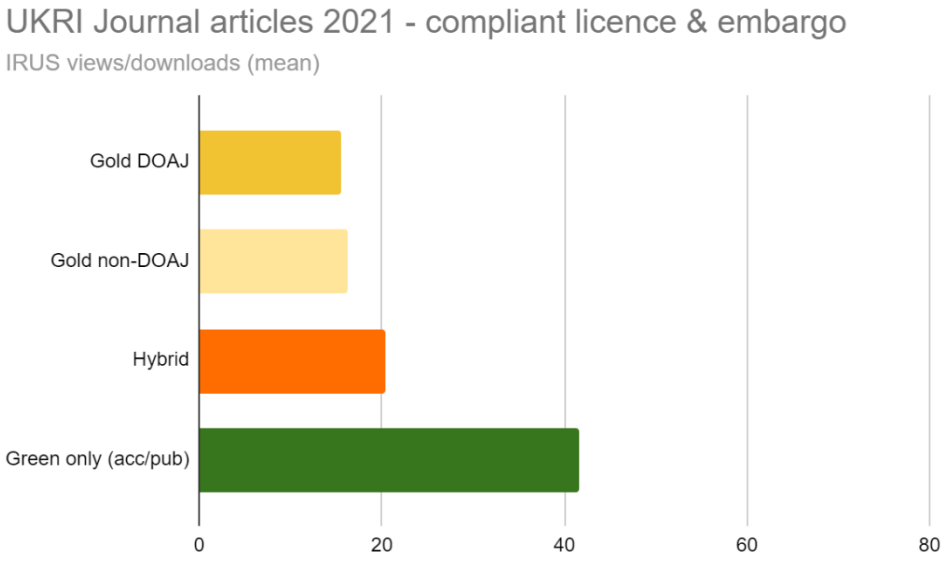
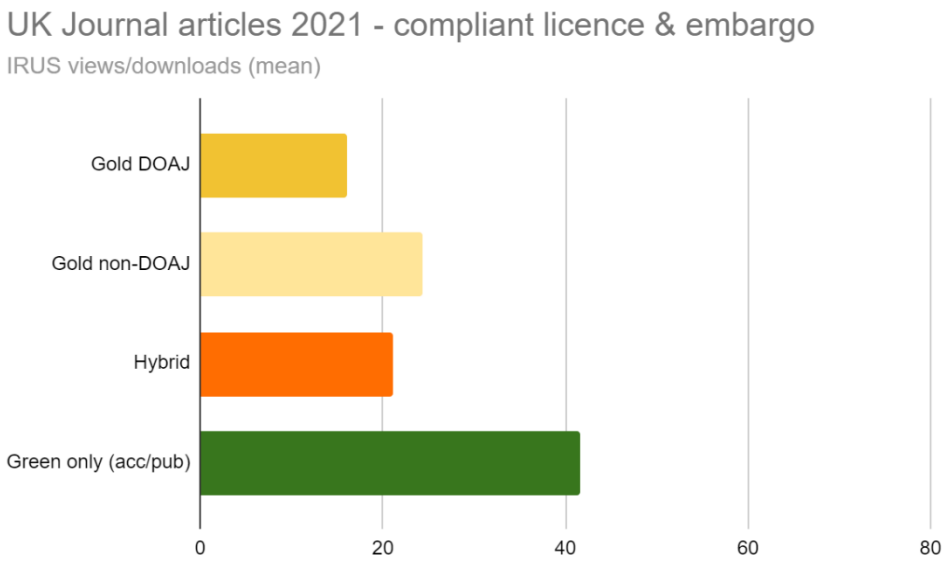


Figure 27. Usage distribution for UK-affiliated articles with a compliant licence and embargo period, by OA status (2021).



2.8 Altmetrics

Background

Information on altmetrics was captured by analysing Crossref Event Data for DOIs in our dataset. Notably, Crossref Event Data will be discontinued in the near future, to be replaced by the new Relationships API. As a result, our analysis was limited to the latest data snapshot available via COKI infrastructure and focused on interactions with the following sources:

- Twitter: DOIs discussed in Tweets
- Newsfeed: DOIs discussed in blogs and media, tracked via RSS
- Reddit Links: DOIs discussed on sites linked to in subreddits
- Wikipedia: DOIs referenced on Wikipedia pages

Information regarding interactions with other sources is also available in our full dataset and code and is not presented in this document due to the very low levels of usage detected. Furthermore, it should be noted that events from Twitter (now X) were collected until February 2023 and have now been removed from search results in accordance with Crossref's contract with Twitter. As our analysis is based on a Crossref snapshot, we are able to showcase this information, but this would not be possible via API queries run in the future. This report refers to Twitter rather than X, because the naming change occurred after the latest data snapshot.

Due to the changing nature of Crossref data regarding linkages to indexed DOIs, Figure 28 is only presented as a proof of concept of what is possible in principle. We recommend that this type of analysis is run again once the Crossref Relationships API is finalised, to better gauge performance and options available to UKRI.

Please note that our analysis refers to altmetrics with the meaning of 'alternative metrics' and not to the commercial product called Altmetric.

Analysis of altmetrics for UKRI-funded and UK-affiliated articles

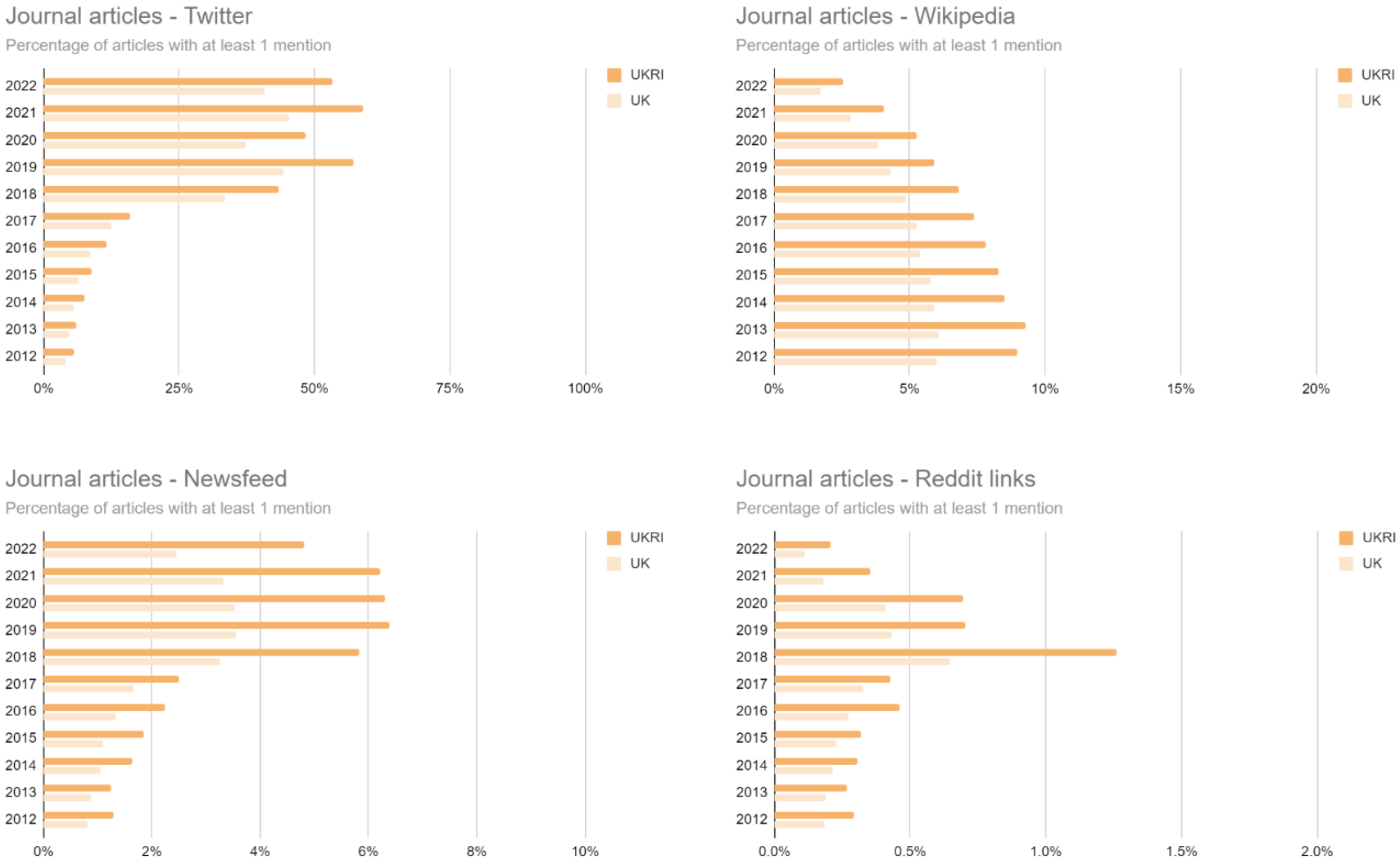
As shown in Figure 28, a higher proportion of UKRI-funded articles tend to achieve at least one mention relative to UK-affiliated articles, which is consistent with the findings reported on citations. Usage patterns over time vary based on the channel being considered, which highlights the importance of examining different types of usage.

Due to the pending deprecation of Event Data, our research has further analysed the nature of usage in each channel; however, we can observe that patterns over time appear to be internally consistent, with smooth trends in all cases. This is a positive indication that the data underpinning altmetrics analysis is likely to be of suitable quality for future use.

When the data in Figure 28 is split by publishing route (not shown), it can be observed that usage via Wikipedia and Newsfeed sources tends to rely more significantly on openly available outputs, whereas usage via Twitter has a higher proportion of links to 'Closed' articles and does not show preferential behaviours towards any given publishing route.

The above might indicate that openly available outputs are more frequently picked up by Wikipedia authors as well as journalists, owing to the possibility to engage with a full text for the purposes of writing up articles and other informational materials (e.g. internal research).

Figure 28. Example of analysis of altmetrics for UKRI-funded and UK-affiliated articles.



3. Future outlooks

The present document shows that it is currently possible to analyse a broad range of areas relating to OA publishing by the means of open data sources. The work done to date has not been compared against the results that would be obtained by using commercial databases to replace some of the data sources we have considered; however, we highlight that the availability of a record-level dataset does make this type of analysis possible.

In Table 8, we highlight a range of observations that should be considered in taking the M&E approach forward. The column 'Data source(s) used' highlights the information needed to develop our analysis:

- In the case of the first two rows (corpus of articles), this indicates the data needed to create our record-level dataset (i.e. rows in a traditional spreadsheet).
- For all subsequent rows, the column 'Data source(s) used' indicates the data that was used to extend our record-level dataset for analysis purposes (i.e. columns in a traditional spreadsheet).

Our use of COKI-provided infrastructure showcases that operationalising M&E efforts, including a high extent of automation, is indeed possible. At the same time, questions about the financial sustainability of open digital infrastructures need to be taken into account. This particularly refers to whether UKRI is looking to develop an in-house solution to deliver M&E assessment in future or to support existing infrastructures (or mixes of infrastructure).

Please note that our reflections refer to the data sources that we used at a specific point in time. We acknowledge that these data sources are subject to continuous improvements (e.g. OpenAlex is working on subjects classifications and publisher lineage; Crossref is working on relationships data). This is a double-edged sword from the perspective of UKRI's M&E approach: on the one hand, data is likely to keep improving over time, leading to more accurate data and results as well as to the possibility of answering new M&E questions; on the other hand, such variability means that the approach we have taken and the code developed will have to be adapted over time to reflect changes in the digital infrastructure and data used.

Table 8. Key observations and lessons learned.

Area	Data source(s) used	Key observations and lessons learned
Identification of the corpus of UKRI-funded articles	Gateway to Research, Crossref	• Gateway to Research data is limited to the quality and completeness of author reporting • Crossref data is limited by the quality and completeness of publisher reporting • Publisher awareness about metadata quality is increasing, with an expected positive impact on the coverage and quality of Crossref data • OpenAlex is currently looking at improving the coverage and quality of funder metadata • Further collaboration between UKRI and OA Switchboard may provide an enhanced ability to identify UKRI-funded articles; this may be considered as and additional data source to Gateway to Research and Crossref
Identification of the corpus of UK-affiliated articles	OpenAlex	• OpenAlex identifiers are used rather than affiliation text strings, thanks to the harmonisation of the data carried out by OpenAlex • The attribution of an identifier to an article relies on (i) the availability of an affiliation text string; and (ii) the quality of the parsing algorithm • OpenAlex is continually improving their approach to parsing affiliation strings to extract structured identifiers

Area	Data source(s) used	Key observations and lessons learned
OA by route, licence and embargo period	Unpaywall	• Using data from Unpaywall enables fine-grained analysis, including analysis of OA version, licence and embargo for each identified version of the article (e.g. Gold OA and a repository copy) • The availability of the variable 'OA_date' in Unpaywall enables the analysis of embargo periods • The above points enable a detailed analysis of levels of alignment with UKRI's requirements, which would not be possible using OpenAlex • OpenAlex data cannot be used directly for the analysis of embargo periods, as not all Unpaywall information is reflected on a 1:1 basis in OpenAlex; as a result, both data sources must be used • Future or full integration of Unpaywall data in OpenAlex is being considered, and this will require a further assessment of the approach we developed
(Inter)national collaborations	OpenAlex	• See 'Identification of the corpus of UK-affiliated articles' • In cases where not all authors have an institutional identifier via OpenAlex, collaboration may be underestimated
Citations	OpenAlex	• Availability of citation information in OpenAlex has directly benefited from the Initiative for Open Citations, resulting in a large proportion of publishers making references available • The subject classifications in OpenAlex enable field normalisation, which was traditionally only possible via commercial databases; this may be considered in future M&E efforts
Usage	IRUS UK	• A central database covering usage data a broad range of UK-based repositories is a significant asset and can help demonstrate the value add of these infrastructures (however, it should be noted that this is not comprehensive and only lists repositories that subscribe to the service) • Usage data are limited to views and downloads via UK-based repositories; for example, usage via repositories such as Europe PMC is not captured nor is usage via publisher platforms • The need to download the full IRUS UK dataset in small batches is very resource intensive • Data can only be downloaded for a specific period • Data can only be downloaded for articles that have seen usage in the chosen period of time • There is an opportunity to negotiate improved access to IRUS UK data (full dataset access) • COUNTER data, which is currently not publicly available, would extend UKRI's ability to analyse usage
Altmetrics	Crossref Event Data (to be replaced)	• There are no other centralised open information sources on altmetrics • The approach to accessing this data is currently being replaced, which will lead to a need to revise the analysis approach • Coverage is limited, with policy mentions being excluded and X (previously Twitter) mentions having been removed due to contractual requirements

Appendix A. Technical decision points

Area	Decision required	Implications
Publication type	For a given record, publication type can vary between different databases. For unambiguous grouping of publications by type (e.g. journal articles), the Crossref string field 'type' is used to identify publication type.	For a number of records in our baseline dataset, the publication type based on the Crossref 'type' field will differ from the publication type recorded in OpenAlex (which defined 'articles' more broadly, including conference proceedings and preprints) and/or Gateway to Research. Crossref type 'journal article' is assigned at the level of publication venue (journals) and may include non-article publications such as editorials and letters to the editor.
Year of publication	For a given record, the year of publication can vary between different databases. For unambiguous grouping of publications by year, the Crossref date field 'issued' is used, which represents the earliest of date between the fields 'published-print' and 'published-online'. For all records in our baseline dataset, this field contains at least the year part of the date.	For a number of records in our baseline dataset, the recorded year of publication based on Crossref 'issued' field will differ from the year of publication as recorded in OpenAlex and/or Gateway to Research. For the majority of cases, the difference is no more than 1 year.
Date of publication	To calculate embargo time (for repository-based OA), complete (Y/M/D) information is needed for the date of publication. The date field 'issued' (used to determine year of publication) cannot be used for this purpose, as it can contain Y, Y/M or Y/M/D information. The only date field in Crossref that contains complete (Y/M/D) information for all records is 'created', representing the date on which the DOI was first registered. This field is therefore used in calculations of embargo periods. The field 'created' cannot be used by itself to determine publication year, as this would result in the inclusion of publications that were published earlier (e.g. in print only) but for which a DOI was registered only recently.	For 13% of records in our baseline dataset, the publication year (based on date field 'issued') differs from the date of publication used to calculate embargo times (based on date field 'created'). For 80% of these cases, the difference is no more than 1 year.
Funder information	Funder information in Crossref can consist of funder names and/or Funder IDs. To enable unambiguous identification of UKRI funded publications from Crossref, only Funder IDs are used. At the time of analysis (November-December 2023), OpenAlex takes funder information only from Crossref, similarly based on Funder IDs.	UKRI-funded publications for which only funder names are included in Crossref, and which are not included in Gateway to Research, are not included in the baseline dataset. They may still be included as UK-affiliated publications.
OpenAlex ID	A small number of DOIs occur more than once in OpenAlex. To retrieve information (e.g. affiliation information) for DOIs in our baseline dataset, an unambiguous match must be made with OpenAlex IDs As a heuristic, the OpenAlex ID with the highest number of distinct affiliations is retained. When the number of distinct affiliations is equal, the record with the highest OpenAlex ID is retained, which is generally the most recently created record.	Where multiple OpenAlex records exist for the same DOI, only one is used. This may impact included affiliation information. The heuristic applied aims to ensure the most affiliations are captured.

Area	Decision required	Implications
OA Licence	All OA information is recorded at the level of the journal article, not at the record of individual versions of the article. To achieve this, information about OA licence is collected as T/F variables, indicating whether one or more version of the article have a licence meeting specific criteria.	For each journal article, it is recorded (separately for publisher version and repository version) whether there is a version with a licence, whether that licence is a Creative Commons licence and whether it is a compliant licence. At this time, information on the specific licences for each publisher or repository version is not recorded, and thus not available for further breakdowns.
OA Classification	Publications can be free to read via the publisher's website (Gold/Hybrid/Bronze OA) and simultaneously be available in a repository (Green OA). Depending on the purpose of the analysis, different decisions can be made as to how to classify these articles.	Articles available in full OA journals (Gold OA) or OA in Hybrid journals (Hybrid OA) are always classified as such, regardless of whether they are also available as Green OA. For analyses that only look at Gold, Hybrid and Green OA, (charts labelled "OA routes"), articles available as Bronze OA are included as Green OA when they are also available in a repository, together with articles that are only available through a repository. Together, this category is labelled "Green only". For analyses that include all possible pathways (charts labelled "publishing routes", articles available as Bronze OA are classified separately, and a distinction is made between articles that are only available as Bronze OA ("Bronze only"), articles that are available as Bronze OA and in a repository ("Bronze and Green") and articles that are only available in a repository ('Green only").
Field normalisation	Aggregate figures for e.g. citations, usage and altmetrics do not account for potential disciplinary effects (i.e. differences in citation behaviour between disciplines). In general, this can be addressed through field-normalisation or by breaking down the data by discipline.	Field normalisation entails comparing the number of citations (or another measure) for each article to the average number of citations (or another measure) for articles in the same discipline and publication year, determining for each article the ratio of citations received to the average number of citation received by articles in the same field in the same year ('field normalised citation ratio'). This method requires the availability of a sufficiently large corpus of articles for which field averages can be determined. For citations, this can be done using OpenAlex, combining citation information with article-level subject classification. For usage (views/downloads) and altmetrics, average values for a sufficiently large reference population cannot be determined as reliably. For this reason, we instead propose to directly calculate results by discipline where desired, as this can be done using only the records included in the dataset, without the need for an external reference population.

Research Consulting

Research Consulting

The Ingenuity Centre, University of Nottingham Innovation Park, Nottingham, NG7 2TU, UK

www.research-consulting.com

This work is licensed under a Creative Commons Attribution 4.0 International License