

Annex A

Data specification

Monitoring and evaluating the effectiveness of UKRI's open access policy

Annex A – Data specification

Report commissioned by:

UK Research and Innovation

<https://www.ukri.org/>

Contact:

Rachel Bruce, Head of Open Research

openresearch@ukri.org

Report authors:

Bianca Kramer, Andrea Chiarelli, Angharad Roberts, Rob Johnson, Cameron Neylon

www.research-consulting.com

Contact:

andrea.chiarelli@research-consulting.com

Report dated: September 2023



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

Available on Zenodo at:

<http://doi.org/10.5281/zenodo.7773583>

Introduction

About this document

This data specification complements the report “Designing a Monitoring and Evaluation framework for UKRI’s open access policy - Principles, opportunities and challenges”, which is available [via Zenodo](#). It provides a detailed overview of the specific M&E questions included in Research Consulting’s initial recommendations to UKRI, as well as methods to answer them and potential challenges. **Appendix A** presents a list of sources that were used, alongside stakeholder consultation, to identify a longlist of possible M&E questions, while the remainder of this document focuses on the questions that were shortlisted as described in the main report.

The data specification includes general steps to collect data from suggested data sources, where available, including examples of queries or API calls. As described in the main report, in most cases data can be collected by:

- direct download (e.g. as csv)
- programmatically (e.g. as JSON or XML object)
- by integrating multiple data sources into an infrastructure solution for analysis

The methodology for obtaining and processing data differs depending on the approach chosen. Therefore, methods for data collection and analysis provided here are indicative, but not described in detail.

Mapping of M&E questions

This document includes the description of a number of recommended datasets and research methods to answer M&E questions. Mapping between datasets and research methods to M&E questions is covered in **Table 1** below, while the detailed descriptions of each dataset and “step” of data preparation – each of which refers to specific M&E questions and involves different aspects of the data – are provided in the main body of this Annex. We note that our recommended approach has been designed in such a way that multiple questions can be answered by using each of the datasets.

Table 1. List of M&E questions and datasets generated to answer these.

#	M&E question	Focus	Dataset
1	What is the number of UKRI-funded outputs annually? How do these figures compare to outputs with an author affiliated with a UK-based research performing organisation? (by discipline, publisher)	Articles / Long-form outputs	Dataset 1, Step A
2	What is the percentage of UKRI-funded outputs compliant with UKRI’s OA policy? (by OA route, license, embargo period) How does this compare to the findings of M&E frameworks run by other funders? (approximated comparison)	Articles / Long-form outputs	Dataset 1, Step B
3	What is the share by publisher of UKRI-funded vs UK outputs? (overall and by OA model)	Articles / Long-form outputs	Dataset 1, Step C
4	What is the percentage of UKRI-funded articles published in journals under Jisc-approved transitional agreements?	Articles	Dataset 1, Step C
5	What are the OA options offered by journals in which UKRI-funded / UK-affiliated authors publish? (by discipline, journal, publisher)	Articles	Dataset 1, Step C
6	To what extent does UKRI’s OA policy affect the number of (inter)national collaborations involving UKRI-funded authors? To what extent are (inter)nationally co-authored publications compliant with policy requirements?	Landscape	Dataset 1, Step D + Qualitative research
7	What are the reasons for non-compliance with UKRI’s OA policy’s terms? (incl. technical requirements, allowed exceptions)	Articles / Long-form outputs	Dataset 1, Step E + Qualitative research
8	What is the percentage of UKRI-funded articles available as an author accepted manuscript in a repository, with a Route 2 Licensing statement?	Articles	Dataset 1, Step F
9	What is the percentage of UKRI-funded articles that include a Data Access Statement?	Articles	Dataset 1, Step F
10	What is the percentage of UKRI-funded articles in journals / repositories meeting technical standards as set out in the UKRI OA policy?	Articles	Dataset 1, Step G

Annex A – Data specification

Monitoring and evaluating the effectiveness of UKRI’s open access policy

#	M&E question	Focus	Dataset
11	How often are UKRI-funded (OA) publications cited/downloaded compared to UK-affiliated OA publications and to UK-affiliated non-OA publications?	Articles / Long-form outputs	Dataset 1, Step H
12	How often are UKRI-funded (OA) publications used/discussed (altmetrics) compared to UK-affiliated OA publications and to UK-affiliated non-OA publications?	Articles / Long-form outputs	Dataset 1, Step H
13	To what extent does OA affect the diversity of affiliation countries of authors citing published outputs , for UKRI-funded and UK-affiliated authors? (by discipline)	Articles / Long-form outputs	Dataset 1, Step I
14	What is the proportion of UKRI OA block grant funding going towards gold OA (including diamond OA), hybrid OA (via Tas) and green OA (via investment in repository infrastructure and staff)?	Landscape	Dataset 2 + Qualitative research
15	What is the (estimated) annual expenditure of institutions towards reading and publishing? (by publishing model)	Articles / Long-form outputs	Dataset 2 + Qualitative research
16	What is the number of OA publications funded via UKRI OA block grant awards/OA fund/other institutional means, and what is the estimated average article/book processing charge? (by discipline, journal, publisher)	Articles / Long-form outputs	Dataset 2 + Qualitative research
17	What do institutions expect/experience to be the main challenges/opportunities arising from UKRI’s OA policy? (incl. around EDI, career progression, research evaluation) ¹	Landscape	Qualitative research
18	What do publishers expect/experience to be the main challenges/opportunities arising from UKRI’s OA policy? (incl. around EDI, career progression, research evaluation)	Landscape	Qualitative research
19	What do researchers expect/experience to be the main challenges/opportunities arising from UKRI’s OA policy? (incl. around EDI, career progression, research evaluation)	Landscape	Qualitative research
20	What difference has access to OA outputs made for non-academic stakeholders? (e.g. industry, general public, practitioners)	Landscape	Dataset 1, Step H + Qualitative research

Data and information sharing

As noted in the main report, this work sought to assess the feasibility of answering M&E questions via the use of open data sources, in combination with proprietary options only where needed. This aimed to demonstrate that a large number of questions on the open access landscape may, in principle, be answered by the means of open data.

The prioritisation of open data sources pursued as part of this work should not be considered as a formal recommendation to UKRI, as no testing has taken place to date to compare results using different sources. As a result, different mixes of open and/or proprietary data may be most appropriate to answer the M&E questions identified.

We note that most data sources described in the present document are openly licensed and can be re-used and re-shared by UKRI. In a limited number of cases, however, this may not be possible and would require ad-hoc arrangements or the payment of licensing fees. The following data sources are likely to require additional arrangements or further consideration prior to sharing in the public domain: Jisc Article Level Metadata; DataSeer; Overton; OA Switchboard; UKRI internal data.

Definitions

Over the course of this document, we use the terms ‘Data source’ and ‘Dataset’. These are defined as follows:

- **Data source:** an existing (open or proprietary) data collection used as a primary source of information
- **Dataset:** a data collection created for the purpose of answering one or more M&E questions, by ingesting and combining data from various data sources

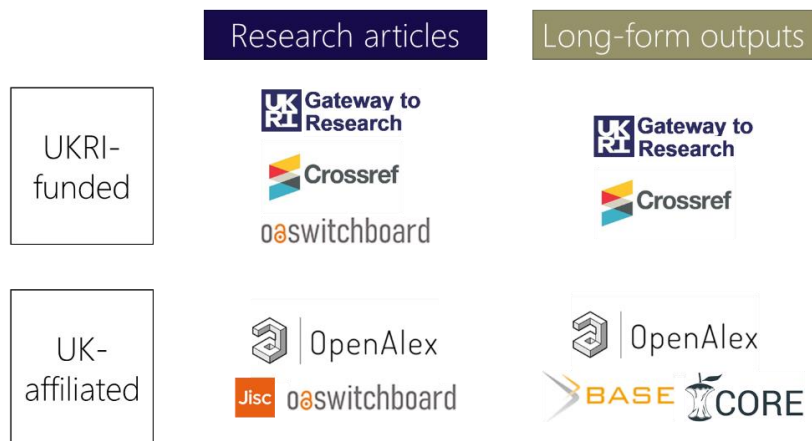
¹ Please note that more detail on the specific dimensions to be addressed in Questions 17-20 is provided in Annex A.

Dataset 1: Record-level information on UKRI-funded and UK-affiliated research outputs

This section describes the preparation of **Dataset 1**, which covers record-level information on research outputs. **Dataset 1** includes a broad range of information and is built by merging and deduplicating data from different sources. Due to its complexity, the preparation of the dataset is subdivided into different 'steps', each of which refers to specific M&E questions and draws on different data sources.

Population of UKRI-funded and UK-affiliated outputs

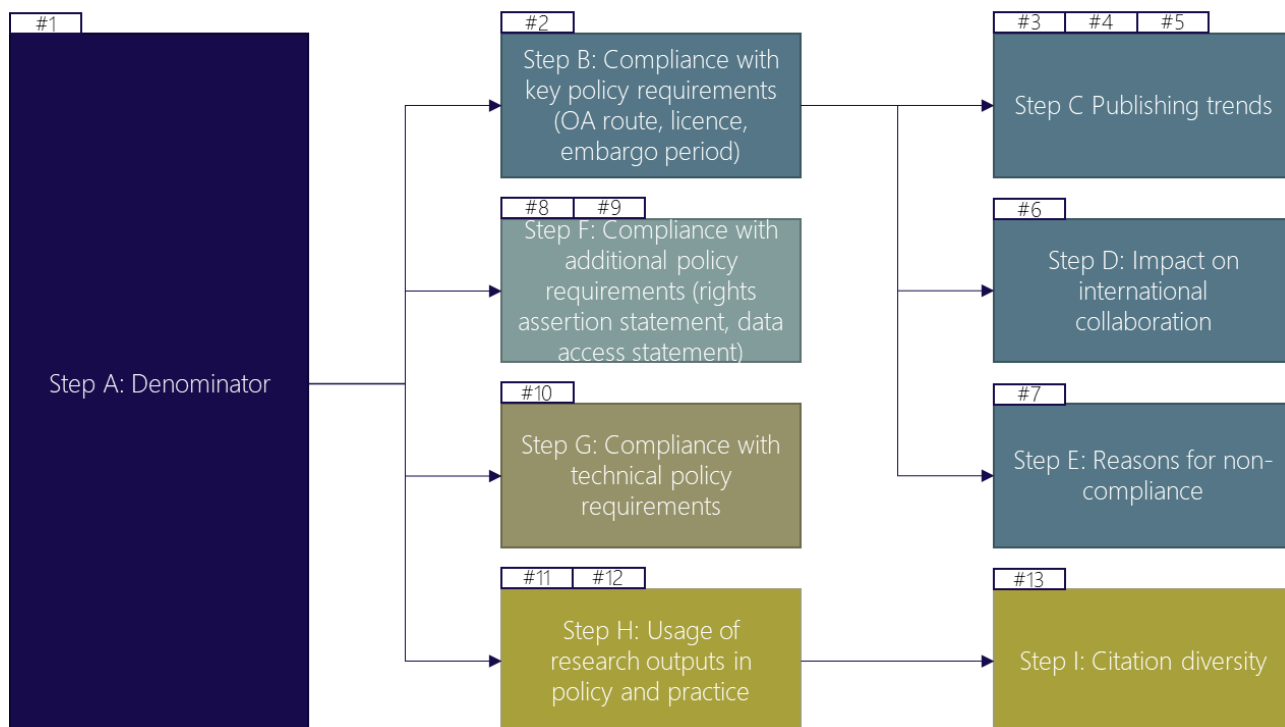
The following graphic covers the main data sources we considered for the identification of the population of in-scope outputs for the M&E framework. This is further explored in Step A below, and an overview of relevant metadata fields for Gateway to Research, Crossref and OpenAlex is available in **Appendix B**.



It should be noted that a broad range of additional (open and proprietary) data sources are discussed over the remainder of this document which are not shown in the graphic above.

Mapping questions to Steps

Given the complexity of Dataset 1, we have provided the following mapping to aid with navigation. Steps are shown in the shaded boxes, and corresponding question numbers are available above each box. Colour-coding is used to illustrate the relationship between steps and the data used.



Step A: Denominator

General information

M&E question(s)	<ul style="list-style-type: none"> #1 What is the number of UKRI-funded outputs annually? How do these figures compare to outputs with an author affiliated with a UK-based research performing organisation? (by discipline, publisher)
Purpose of the question	<ul style="list-style-type: none"> The purpose of this question is to define the set of outputs that forms the basis for the monitoring exercise (including comparison of UKRI-funded outputs and all UK outputs) and allow subsequent grouping by discipline, research performing organisation and publisher.
Source(s)	<ul style="list-style-type: none"> Articles <ul style="list-style-type: none"> Gateway to Research (GtR, for UKRI-funded research outputs) OpenAlex (for UK-affiliated research outputs + enriched data on affiliations, publishers and disciplines) Crossref (for additional research outputs based on funding data, as well as for publication date, if selected as authoritative source) Article-level metadata from Jisc and OA Switchboard, if available (for additional research outputs based on funder data and for enriched affiliation data) Long-form outputs <ul style="list-style-type: none"> Gateway to Research (for UKRI-funded research outputs) OpenAlex (for UK-affiliated research outputs, + enriched data on affiliations, publishers and disciplines) Crossref (for additional research outputs based on funding data, as well as for publication date, if selected as authoritative source) BASE, CORE (for UK-affiliated research outputs, with known limitations)
Mapping to the OA policy logic model	<ul style="list-style-type: none"> N/A (denominator)

Table specification

Variable	Data type	Sample data
Source-specific identifier	OpenAlex ID (alphanumeric), GtR Outcomeld (alphanumeric)	W3136078166, 616ef5329dd1b
Identifier	DOI or ISBN (alphanumeric)	10.26784/sbir.v5i1.312
UKRI-funded output	String or Logical	UKRI or TRUE/FALSE
UK-affiliated output	String or Logical	UK or TRUE/FALSE
Publication type	String	Article
Publication date	Date	2020-01-11
Research performing organisation(s) (ROR)	ROR ID (alphanumeric)	03yrm5c26
Journal title	String	Small Business International Review
Journal (ISSN)	ISSN	2531-0046
Publisher	String	AECA
Discipline	String	Economics

Variable	Data type	Sample data
UKRI council	String	ESRC

Data preparation

High-level overview

- Collect publication metadata for UKRI outputs and UK-affiliated outputs for a given publication year.
- Optional: add publication outputs from additional data sources with information on UKRI funding.
- Enrich resulting dataset with information on affiliations, publishers and disciplines².

Recommended approach

- Collect basic publication metadata from [Gateway to Research](#) (for UKRI outputs)
 - Select on ‘publications’, publication year 2022, relevant publication types for articles (including conference proceedings) and long-form outputs
 - See [example query](#) result
 - Download as csv or use [API access](#) for data in XML/JSON format (including ISBNs).
 - Deduplicate on Gateway to Research identifier (GtR Outcomeld) and DOI
- [Optional]: Collect basic publication metadata for UKRI-funded outputs from [Crossref](#)
 - Select on publication year, relevant publication types for articles (including conference proceedings) and long-form outputs, and Funder IDs for UKRI and UKRI councils. For an overview of UKRI-related Funder IDs, see <https://doi.org/10.5281/zenodo.5562842>
 - see [API example](#) for (2022, UKRI, journal articles).
 - Add records to existing dataset, deduplicate on DOI
- [Optional]: Collect basic publication metadata for UKRI-funded outputs from Jisc (Article Level Metadata) and/or OA Switchboard
 - Processing depending on format (csv or JSON) in which data is made available
 - Add records to existing dataset, deduplicate on DOI
- Enrich publication metadata for UKRI outputs with DOIs with affiliations, publishers and disciplines from OpenAlex
 - See example [API call](#)
 - Add variables for affiliations, publishers and disciplines to existing records by matching on DOI
- Collect publication metadata from OpenAlex (for UK-affiliated outputs), including affiliations, publishers, disciplines*
 - See example [API call](#)
 - Add records to existing dataset, deduplicate on DOI/ISBN

Potential challenges and issues

Challenge	Mitigation
Incomplete coverage	Combine data sources and, if desirable, consider licensing proprietary data
Innovate UK and Research England outputs not included in Gateway to Research	Included in Crossref data to some extent; consider feasibility and assess the need to gather additional information from Innovate UK and Research England directly to fill gaps and enhance the data available (see

² OpenAlex contains information on disciplines based on title, abstract, and title of the publication venue, using a hierarchical classification with 19 top levels (for more information see <https://docs.openalex.org/api-entities/concepts>). Our own exploration has shown that subject information is now available for >99% of records in OpenAlex.

Annex A – Data specification

Monitoring and evaluating the effectiveness of UKRI’s open access policy

Challenge	Mitigation
	further discussion on this topic in the main report); consider licensing proprietary data
Conflicting information between sources	Define priority ranking for cases where mismatches are encountered (e.g. year of publication can differ between data sources)
Deduplication of records where an identifier is not available	Employ fuzzy matching using a programmatic approach, or manual checks (not recommended due to high resource intensity)
Gateway to Research does not include ISBN as export variable	Assess the potential for ISBN to be added as variable to Gateway to Research prior to monitoring long-form outputs, or use API to access data
ISBN is not a unique identifier across versions (e.g. print book, eBook)	Employ fuzzy matching, using a programmatic approach, or manual checks (not recommended due to high resource intensity)
Limited coverage/centralised availability of book metadata (incl. affiliation and funding data)	Use repository holdings (e.g. in BASE, CORE) as proxy for affiliation Ensure this is appropriately caveated as part of the M&E methodology and narrative See next two rows for additional limitations on the usage of BASE and CORE
BASE contains duplicate records; content providers include UK publishers via Crossref	Explore feasibility of using OAI-PMH or API access to filter UK repositories (see http://oai.base-search.net/)
CORE does not display publication types	Explore feasibility of using API access for better retrieval options
Publication types are not consistent across data sources; publication types in data sources do not always correspond 1:1 with eligible publication types in UKRI’s OA Policy	Document original publication types in each data source before harmonising; choose a priority source (or sources) to be used as the reference point(s), taking into account the best match with publication types in UKRI’s OA policy; harmonise publication types across data from different data sources

Step B: Compliance with key policy requirements (OA route, licence, embargo period)

General information

M&E question(s)	<ul style="list-style-type: none"> #2 What is the percentage of UKRI-funded outputs compliant with UKRI’s OA policy? (by OA route, license, embargo period) How does this compare to the findings of M&E frameworks run by other funders? (approximated comparison)
Purpose of the question	<ul style="list-style-type: none"> The purpose of this question is to assess overall compliance with the UKRI OA policy and to identify the use of the different OA routes (full gold OA journals, hybrid journals in transitional agreements, and green OA), including in the form of year-on-year comparisons Collecting information on licence and embargo allows detailed compliance monitoring, assessing and characterising non-compliance, and monitoring trends in licence and embargo use over time (Aspects of) compliance can be further broken down by publication type, publisher, discipline and UKRI council (and research performing organisation, if desired) using the variables collected in Dataset 1, Step A.
Source(s)	<ul style="list-style-type: none"> Articles <ul style="list-style-type: none"> Unpaywall Long-form outputs <ul style="list-style-type: none"> Unpaywall, DOAB
Mapping to the OA policy logic model	<ul style="list-style-type: none"> Outputs <ul style="list-style-type: none"> UKRI funded articles are full and immediate OA UKRI funded articles meet open licensing requirements UKRI have a greater level of alignment with partners nationally and internationally Outcomes <ul style="list-style-type: none"> Research findings are open, re-usable and repeatable Increased knowledge use/exploitation from articles through improved access e.g. allowing for T&D mining

Table specification

Primary variables

Variable	Data type	Sample data
DOI / ISBN	DOI or ISBN (alphanumeric)	10.26784/sbir.v5i1.312
Open access venue	String	Publisher, Repository, NA
Open access venue source	String	Unpaywall, DOAB, BASE
Open access version	String	VOR, AAM, NA
DOAJ status	Logical	TRUE, FALSE
DOAJ APC status	String	APC, non-APC, NA
Licence	String	CC BY, NA
License source	String	Unpaywall, DOAB

Annex A – Data specification

Monitoring and evaluating the effectiveness of UKRI’s open access policy

Variable	Data type	Sample data
Open access date	Date	2022-10-22
Embargo period (months)	Integer	8
Repository	String	Nottingham ePrints
Repository source	String	Unpaywall, BASE, CORE

Secondary variables

Variable	Data type	Sample data
OA type	String	Gold DOAJ
Non-APC (diamond)	Logical	TRUE, FALSE
AAM/VOR in repository (all green)	Logical	TRUE, FALSE
AAM/PV in EuropePMC (relevant for MRC and BBSRC)	Logical	TRUE, FALSE
AAM/VOR in repository – best licence	String	CC-BY
AAM/VOR in repository - shortest embargo (months)	Integer	0
Overall compliance	Logical	TRUE, FALSE

Data preparation

High-level overview

- Collect primary variables
- Match primary variables to output records in **Dataset 1, Step A** by matching on DOI or ISBN
- Calculate secondary variables based on primary variables

Recommended approach

For each DOI in **Dataset 1, Step A**:

- Collect open access information from [Unpaywall](#)
- Search by DOI
 - See example [API call](#)
- Export as csv or use API access
- Add data to **Dataset 1, Step A** by matching on DOI
- Calculate embargo time for each DOI with repository version(s):
 - Calculate difference in days between Crossref publication data and Unpaywall is-oa date
 - Calculate embargo time in months
- Assign OA type classification based on primary variables:
 - Gold DOAJ: VOR via publisher + DOAJ journal TRUE
 - Hybrid: VOR via publisher + DOAJ journal FALSE, VOR has licence
 - [optional] Bronze: VOR via publisher + DOAJ journal FALSE, VOR has no licence
 - Green only: not gold or hybrid, AAM/VOR in repository
 - Closed: no VOR via publisher, no AAM/VOR in repository
- Assign non-APC (diamond) status based on primary variables:
 - VOR via publisher + DOAJ journal TRUE + DOAJ non-APC
- Assign overall compliance based on primary variables:
 - VOR via publisher + CC-BY or CC-BY-ND licence

Annex A – Data specification

Monitoring and evaluating the effectiveness of UKRI’s open access policy

- AAM/VOR in repository + CC-BY or CC-BY-ND licence + 0-month embargo

For each ISBN in **Dataset 1, Step A**:

- Check presence + license information in [DOAB](#)
- Search by ISBN
- Download csv or use [API access](#)
- Add data to **Dataset 1, Step A** by matching on ISBN
- Assign overall compliance based on primary variables:
 - In DOAB + any CC licence

Data analysis

Calculate percentage of compliant outputs:

- Overall compliance
- Distribution of OA types
- Distribution of licences (for evaluation also include non-compliant licences)
- Distribution of shortest embargo times (for evaluation, including non-compliant embargo times)

Aspects of compliance can be further broken down by publication type, publisher, discipline and UKRI council (and research performing organisation, if desired), and compared between UKRI-funded outputs and UK-affiliated outputs.

Potential challenges and issues

Challenge	Mitigation
Unpaywall only has information based on DOIs	N/A – Ensure this is appropriately caveated as part of the M&E methodology and narrative
Ambiguity in compliance checking	For publication year 2022, only include UKRI-funded articles after the start of the OA policy. For long-form outputs from 2024 onwards, it is not possible to tell from the data whether a contract has been signed between the author and the publisher before the start of the policy: ensure this is appropriately caveated as part of the M&E methodology and narrative. There is potential for UKRI to feed exception decisions back into the M&E dataset, but this is likely to be a manual and labour-intensive effort.
DOAB and Unpaywall data are structured differently, as they refer to different types of outputs and include widely incompatible columns of data ³	Log information from the sources in separate columns
DOAB and OAPEN do not include books in repositories	Use CORE/BASE as additional sources of repository-based OA
Temporary status of OA information (i.e. changes in OA status and their detection may vary in time)	N/A – Ensure this is appropriately caveated as part of the M&E methodology and narrative

³ Unpaywall and DOAB are completely independent data sources, one focusing on books/monographs and one focusing on articles. All columns differ and are not compatible with one another. Examples may be found as follows:

- DOAB csv export: <https://directory.doabooks.org/download-export?format=csv>
- Unpaywall data format: <https://unpaywall.org/data-format>

Annex A – Data specification

Monitoring and evaluating the effectiveness of UKRI’s open access policy

Challenge	Mitigation
Some variables can have multiple values (e.g. a record can have multiple repository versions, each with its own licence)	Collect data on all versions, then prioritise based on whether a compliant version is available
Not all variables available for long-form outputs (e.g. version, embargo time)	N/A – Ensure this is appropriately caveated as part of the M&E methodology and narrative
OAPEN currently does not offer repository function for UKRI-funded authors	Explore feasibility of this option with OAPEN
DOAB currently does not specifically track UKRI-funded OA long-form outputs	Explore feasibility of this option with DOAB
Not all open access long-form outputs may be included in DOAB	Ensure this is appropriately caveated as part of the M&E methodology and narrative

Step C: Publishing trends

General information

M&E question(s)	<ul style="list-style-type: none"> • #3 What is the share by publisher of UKRI-funded vs UK outputs? (overall and by OA model) • #4 What is the percentage of UKRI-funded articles published in journals under Jisc-approved transitional agreements? • #5 What are the OA options offered by journals in which UKRI-funded / UK-affiliated authors publish? (by discipline, journal, publisher)
Purpose of the question	<ul style="list-style-type: none"> • The purpose of these questions is to assess where UKRI-funded and UK-affiliated authors publish over time, as well as to chart the different OA models used by venues in which UKRI-funded (and UK-affiliated) researchers publish. • This can give information on how the publishing landscape develops, including consolidation or differentiation of OA models (including full gold OA and non-APC based OA) and publishers offering these models. • It can also indicate whether publishing behaviour is shaped by OA availability (including transitional agreements)
Source(s)	<ul style="list-style-type: none"> • Articles <ul style="list-style-type: none"> - Journal Checker Tool (for information on journals in transitional agreements and participating research performing organisations) - Jisc article-level data (for articles covered under transitional agreements) - OA Switchboard (for articles covered under transitional agreements) - Jisc list of transformative journals • Long-form outputs <ul style="list-style-type: none"> - Jisc data on agreements for long-form outputs subject to availability (currently some pilots for diamond open access)
Mapping to the OA policy logic model	<ul style="list-style-type: none"> • Outputs <ul style="list-style-type: none"> - Broader adoption of OA practices / publishing models for a diversity of publishers that cover UKRI authors e.g. transitional agreements • Outcomes <ul style="list-style-type: none"> - Diverse OA publishing landscape that meets the needs of the research sector

Table specification

Variable	Data type	Sample data
DOI	DOI (alphanumeric)	10.26784/sbir.v5i1.312
Journal is part of Jisc transitional agreement	Logical	TRUE, FALSE
Journal is part of transitional agreement in UK for one or more authors of the publication	Logical	TRUE, FALSE
Journal is Jisc-approved transformative journal	Logical	TRUE, FALSE

Data preparation

High-level overview

- Collect primary variables
- Match primary variables to output records in **Dataset 1, Step A** by matching on DOI

Recommended approach

- Use value for variables journal (ISSN) and publication year from **Dataset 1, Step A** to check if journal is covered under a UK transitional agreement (in [Journal Checker Tool](#)) or is a transformative journal (in [Sherpa Romeo](#))
- Additionally, use values for variable affiliation (ROR) to check if research performing organisation(s) participate in transformative journals (if any)
- Use Journal Checker Tool API to query Journal Checker Tool with ISSN and ROR, or look up combination of ISSN, ROR and publication year in Journal Checker Tool spreadsheet with information on transitional agreements

Data analysis

Using information on publisher from **Dataset 1, Step A** and information on OA type from **Dataset 1, Step B**, calculate the following parameters for both UKRI-funded publications and UK-affiliated publications:

- Frequency distribution of records per publisher overall and by OA model
- Frequency distribution of records per OA model by publisher
- Percentage of records in transitional agreements, overall and by publisher
- Percentage of records in transformative journals, overall and by publisher
- Percentage of journals offering the different OA options, overall and by publisher

Aspects of compliance can be further broken down by publication type, discipline and UKRI council (and research performing organisation, if desired) and compared between UKRI-funded outputs and all UK-affiliated outputs.

Potential challenges and issues

Challenge	Mitigation
Transitional agreements can only be used by corresponding authors, which is not information available in article metadata	Use Jisc article level metadata and/or OA Switchboard data to check if publication is published under a transitional agreement in UK
Availability of transitional agreements differs by research performing organisation	Show separate results for (representative) research performing organisations; or present two sets of results for OA availability: one with and one without Jisc deals included
Publisher stance re: Route 2 Licensing statement can affect OA availability (e.g. can Route 2 Licensing statements be used when paid OA option is available?)	Ensure this is appropriately caveated as part of the M&E methodology and narrative

Step D: Impact on (inter)national collaboration

General information

M&E question(s)	<ul style="list-style-type: none"> • #6 To what extent does UKRI’s OA policy affect the number of (inter)national collaborations involving UKRI-funded authors? To what extent are (inter)nationally co-authored publications compliant with policy requirements?
Purpose of the question	<ul style="list-style-type: none"> • The purpose of this question is to investigate and assess trends in collaboration over time and try to isolate the effect of the UKRI OA policy. • The underlying assumptions to test are a) the UKRI OA policy might limit opportunities for (inter)national collaboration due to the mandate for OA publishing; b) wider availability of UKRI-funded research might increase potential (inter)national collaborations
Source(s)	<ul style="list-style-type: none"> • Articles, long-form outputs <ul style="list-style-type: none"> - Dataset 1, Step A - Dataset 1, Step B (OA type / compliance)
Mapping to the OA policy logic model	<ul style="list-style-type: none"> • Outputs <ul style="list-style-type: none"> - Authors maintain a sufficient level of choice as to where their outputs are published • Outcomes <ul style="list-style-type: none"> - Accurate flow of article information between researchers, publishers, funders and research performing organisations - Improved transparency and efficiency of the research process - Diverse OA publishing landscape that meets the needs of the research sector • Impact <ul style="list-style-type: none"> - UKRI effectively stewards a more open, fair and transparent R&I system - Improved understanding of ourselves and the world around us

Table specification

N/A (no new variables generated)

Data preparation

High-level overview

- Use affiliation data (ROR, affiliation country) for each record (DOI/ISBN) in **Dataset 1, Step A**
- Use discipline, UKRI council as additional breakdown parameters

Recommended approach

- Create a table where each DOI/ISBN is mapped to a single ROR (e.g. if an output has three affiliations listed, there will be three entries for it, one for each ROR available)

Data analysis

- Calculate frequency distribution of research performing organisations/countries over time for UKRI-funded (OA) outputs and UK-affiliated OA and non-OA outputs
- Compare levels of compliance (overall and by OA type) for research publications resulting from collaborations with different research performing organisations/countries for UKRI-funded (OA) outputs and UK-affiliated OA and non-OA outputs.

Annex A – Data specification

Monitoring and evaluating the effectiveness of UKRI’s open access policy

- Optional: perform network analysis of collaborations, identifying:
 - most important clusters
 - strength of connections
 - developments over time

Analysis can be further broken down by discipline, UKRI council and country (if desired).

Potential challenges and issues

Challenge	Mitigation
Effect of potential confounders (geopolitical, economic)	Ensure this is appropriately caveated as part of the M&E methodology and narrative
Network/cluster analysis requires more specific data analysis expertise	Approach as research project additional to core M&E framework
Co-authorship is only one specific output measure of collaboration	Also consider looking at affiliation data from UKRI-funded applications if a broader scope is desirable (internal UKRI data). Data from unsuccessful applications is unlikely to be available for analysis at scale.

Step E: Reasons for non-compliance

General information

M&E question(s)	<ul style="list-style-type: none"> #7 What are the reasons for non-compliance with UKRI’s OA policy’s terms? (incl. technical requirements, allowed exceptions)
Purpose of the question	<ul style="list-style-type: none"> The purpose of this question is to assess reasons for non-compliance, and any differences across different dimensions, e.g. disciplines or research performing organisations. Record-level data on compliance can identify the proportion of non-compliant outputs and the areas of the policy where they are not compliant (e.g. license, embargo time). This can then be followed up in a qualitative way by a) checking for existing exceptions and b) using the quantitative data on reasons for non-compliance to inform qualitative research engaging research performing organisations and publishers (see M&E Question #17-18 and Appendix C)
Source(s)	<ul style="list-style-type: none"> Articles, long-form outputs <ul style="list-style-type: none"> Dataset 1, Step B Manual checks Internal UKRI information (No-derivatives licence exception requests) Qualitative research
Mapping to the OA policy logic model	<ul style="list-style-type: none"> Outputs <ul style="list-style-type: none"> Broader adoption of OA practices / publishing models for a diversity of publishers that cover UKRI authors e.g. transitional agreements Authors maintain a sufficient level of choice as to where their article will be published Outcomes <ul style="list-style-type: none"> Diverse OA publishing landscape that meets the needs of the research sector

Table specification

Variable	Data type	Sample data
Reason(s) for non-compliance	String	No OA version

Data preparation

High-level overview

- Identify non-compliant records from **Dataset 1, Step B**
- Identify reasons for non-compliance based on information in **Dataset 1, Step B**
- Collect additional information through manual checks (exceptions)

Recommended approach

- From **Dataset 1, Step B**, identify non-compliant records (UKRI-funded outputs only)
- For each non-compliant record, check compliance/non-compliance with individual elements of the UKRI OA policy (OA version available, license, embargo time)
- Manually check existence of exceptions (if appropriate), including:
 - Matching to **no-derivatives licence exception request forms** (internal UKRI information)
- Classification of reasons for non-compliance (including license/embargo)
 - No OA version

Annex A – Data specification

Monitoring and evaluating the effectiveness of UKRI’s open access policy

- Non-compliant license
- Non-compliant embargo time
- Valid exception (upon manual inspection)
- Classification error
- Incomplete information
- Other

Data analysis

For non-compliant records, calculate distribution of reasons for non-compliance.

Reasons for non-compliance can be further broken down by publication type, publisher, discipline and UKRI council (and research performing organisation, if desired).

This quantitative approach can provide input to further qualitative research engaging research performing organisations and publishers (see M&E questions #17-18 and **Appendix D**).

Potential challenges and issues

Challenge	Mitigation
Manual checks for existing exceptions are too resource-intensive	Restrict to random sample (stratified by publication type, publisher, discipline, research performing organisation)

Step F: Compliance with additional policy requirements (Route 2 Licensing statement, data access statement)

General information

M&E question(s)	<ul style="list-style-type: none"> • #8 What is the percentage of UKRI-funded articles available as an author accepted manuscript in a repository, with a Route 2 Licensing statement? • #9 What is the percentage of UKRI-funded articles that include a Data Access Statement?
Purpose of the question	<ul style="list-style-type: none"> • The purpose of these questions is to monitor specific aspects of the UKRI OA policy for research articles, namely the use of a Route 2 Licensing statement when using the repository route and the requirement to include a Data Access Statement. • Monitoring of the use of a Route 2 Licensing statement is useful to evaluate the uptake and effectiveness of this practice (and potentially identify publishers not accepting this). • Monitoring the presence of a Data Access statement is useful to assess uptake of open science practices beyond open access. Identifying data access statements could also allow deeper study into their nature and contents (which is beyond the UKRI OA policy)
Source(s)	<ul style="list-style-type: none"> • Articles <ul style="list-style-type: none"> - Route 2 Licensing statement <ul style="list-style-type: none"> ▪ EuropePMC ▪ CORE - Data access statement <ul style="list-style-type: none"> ▪ OA Switchboard ▪ DataSeer • Long-form outputs <ul style="list-style-type: none"> - N/A
Mapping to the OA policy logic model	<ul style="list-style-type: none"> • Outputs <ul style="list-style-type: none"> - Authors maintain a sufficient level of choice as to where their outputs are published • Outcomes <ul style="list-style-type: none"> - Broader adoption of OA across all stakeholders nationally and internationally - Diverse OA publishing landscape that meets the needs of the research sector - Research findings are open, re-usable and repeatable - Improved transparency and efficiency of the research process - Improved confidence in quality of research outputs - Less research waste and improved research integrity

Table specification

Variable	Data type	Sample data
DOI	DOI (alphanumeric)	10.26784/sbir.v5i1.312
Availability of Route 2 Licensing statement	Logical	TRUE, FALSE

Variable	Data type	Sample data
Availability of data access statement	Logical	TRUE, FALSE
Full text data access statement	String	All data created during this research is openly available from the University of Bath Research Data Archive at https://doi.org/10.15125/BATH-01069

Data preparation

High-level overview

- Collect primary variables
- Match primary variables to output records in **Dataset 1, Step A** by matching on DOI or ISBN

Recommended approach

For each DOI in **Dataset 1, Step A** (UKRI-funded outputs only):

- Check availability of Route 2 Licensing statement in [EuropePMC](#)
 - Search for key phrase in Route 2 Licensing statement, e.g. ‘For the purpose of open access’, see [example query](#)
 - Export matching records as csv, or use API access
 - Match DOIs to records in **Dataset 1**
 - Only record TRUE/FALSE, do not include all EuropePMC fields
- Check availability of Route 2 Licensing statement in [CORE](#)
 - Search for key phrase in Route 2 Licensing statement, e.g. ‘For the purpose of open access’, see [example query](#)
 - Export matching records as csv, or use API access
 - Match DOIs to records in **Dataset 1**
 - Only record TRUE/FALSE, do not include all CORE fields
- Check availability of data access statement in OA Switchboard
 - OA Switchboard data contain field with TRUE/FALSE statement for availability of data access statement
 - Data made available as csv or JSON
 - Add data to **Dataset 1** by matching on DOIs
- Check availability of data access statement in [DataSeer](#)
 - Identify data access statements for relevant set of DOIs from **Dataset 1** with DataSeer
 - Processing (TRUE/FALSE and content of data statement) depending on export format provided
 - Add data to **Dataset 1** by matching on DOI
 - Note: DataSeer can also further analyse data access statements and available data, see [collaboration PLOS – DataSeer](#) for example and more details

Data analysis

Calculate percentage of compliant research articles:

- Use of Route 2 Licensing statements in articles made OA through repository
- Availability of data access statements (articles only)

Aspects of compliance can be further broken down by publisher, discipline and UKRI council (and research performing organisation, if desired).

Potential challenges and issues

Challenge	Mitigation
Available sources might not provide good enough coverage	Text mining approach on full text articles (not recommended due to high resource intensity)
Route 2 Licensing statements may vary in wording	Sample DOIs of UKRI-funded articles made OA through repository to explore phrasing variants, add variants to query
OA Switchboard only provides T/F information on the availability of a data access statement	Text mining approach on identified full text articles (not recommended due to high resource intensity)
Detailed study of data access statements is beyond M&E framework	Identifying data access statements will make them available for studies adjacent to M&E framework

Step G: Compliance with technical policy requirements

General information

M&E question(s)	<ul style="list-style-type: none"> #10 What is the percentage of UKRI-funded articles in journals / repositories meeting technical standards as set out in the UKRI OA policy?
Purpose of the question	<ul style="list-style-type: none"> The purpose of this question is to monitor further specific aspects of the UKRI OA policy, namely technical standards for journals and repositories. Monitoring technical standards for journals and repositories as specified in the UKRI OA policy will likely constitute a separate project in connection with the UKRI’s OA Technical Standards T&F. However, the data on journals and repositories used by UKRI-funded authors collected in the M&E framework can potentially be used for this type of detailed future study. (see Appendix D) Only those standards for which metadata are currently mature enough to provide useable information are included in this data specification. (see Appendix D)
Source(s)	<ul style="list-style-type: none"> Articles <ul style="list-style-type: none"> Crossref SHERPA RoMEO DOAR Long-form outputs <ul style="list-style-type: none"> Not included as metadata not sufficiently mature
Mapping to the OA policy logic model	<ul style="list-style-type: none"> Outputs <ul style="list-style-type: none"> Journals / publishing platforms / repositories where UKRI-funded outputs are published meet a minimum level of technical standards Journals / publishing platforms / repositories have improved metadata standards Persistent identifiers are used more frequently Improved interoperability of research systems Outcomes <ul style="list-style-type: none"> Accurate flow of article information between researchers, publishers, funders and research performing organisations

Table specification

Variable	Data type	Sample data
DOI / ISBN	DOI or ISBN (alphanumeric)	10.26784/sbir.v5i1.312
Journal uses PIDs for research articles	Logical	TRUE, FALSE
Journal provides information on licence in the article metadata	Logical	TRUE, FALSE
Journal supports long-term preservation	Logical	TRUE, FALSE
Journal has openly available citations	Logical	TRUE, FALSE
Journal’s self-archiving policies are registered in SHERPA RoMEO	Logical	TRUE, FALSE
Journal supports ORCID	Logical	TRUE, FALSE

Variable	Data type	Sample data
Journal supports organisation identifier (e.g. ROR)	Logical	TRUE, FALSE
Journal supports funder identifier (e.g. Crossref Funder ID)	Logical	TRUE, FALSE
Repository is registered in the Directory of Open Access Repositories	Logical	TRUE, FALSE

Data preparation

High-level overview

- Collect primary variables
- Match primary variables to output records in **Dataset 1, Step A** by matching on DOI or ISBN

Recommended approach

For each record in **Dataset 1, Step A**, check if journal/publication venue meets technical requirements (UKRI-funded outputs only).

- Journal uses PIDs for research articles
 - TRUE if record has DOI
 - Add data to **Dataset 1** by matching on DOI
- Journal provides information on licence in the article metadata
 - Search [Crossref](#) by ISSN
 - Get proportion of current articles that have license information
 - variable “flags.deposits-licenses-current” (TRUE/FALSE)
 - variable “coverage.current.licenses-current” (percentage)
 - Determine cut-off value for TRUE/FALSE decision
 - See example [API call](#)
 - Add data to **Dataset 1** by matching on DOI
- Journal supports long-term preservation
 - Search [ISSN Keeper’s registry](#) by ISSN (this includes Portico, CLOCKSS, will include JASPER)
 - Information on preservation in ‘Archival Status’
 - See example [query results](#) or access using API (see [example](#) for all journals using CLOCKSS)
- Journal has openly available citations
 - Search [Crossref](#) by ISSN
 - Get proportion of current articles that deposit citations (NB all citations are currently open in Crossref)
 - variable “flags.deposits-references-current” (TRUE/FALSE)
 - variable “coverage.current.references-current” (percentage)
 - Apply predetermined threshold value for TRUE/FALSE decision
 - See example [API call](#)
 - Add data to **Dataset 1** by matching on DOI
- Journal’s self-archiving policies are registered in Sherpa Romeo
 - Search [Sherpa Romeo](#) by ISSN
 - See example [query result](#) or access [using API](#)
 - Add data on presence (TRUE/FALSE) to **Dataset 1** by matching on DOI
- Journal supports ORCID
 - Search [Crossref](#) by ISSN
 - Get proportion of current articles that deposit ORCID
 - variable “flags.deposits-orcids-current” (TRUE/FALSE)
 - variable “coverage.current.orcids” (percentage)
 - Apply predetermined threshold value for TRUE/FALSE decision
 - See example [API call](#)
 - Add data to **Dataset 1** by matching on DOI

Annex A – Data specification

Monitoring and evaluating the effectiveness of UKRI’s open access policy

- Journal supports organisation identifier (e.g. ROR⁴)
 - Search [Crossref](#) by ISSN
 - Get proportion of current articles that deposit ROR
 - variable “flags.deposits-ror-ids--current” (TRUE/FALSE)
 - variable “coverage.current.ror-ids” (percentage)
 - Apply predetermined threshold value for TRUE/FALSE decision
 - See example [API call](#)
 - Add data to **Dataset 1** by matching on DOI
- Journal supports funder identifier (e.g. Crossref Funder ID)
 - Search [Crossref](#) by ISSN
 - Get proportion of current articles that deposit funder information
 - variable “flags.deposits-funders-current” (TRUE/FALSE)
 - variable “coverage.current.funders” (percentage)
 - Apply predetermined threshold value for TRUE/FALSE decision
 - See example [API call](#)
 - Add data to **Dataset 1** by matching on DOI
- Repository is registered in the Directory of Open Access Repositories
 - For each repository location, search [OpenDOAR](#) by repository name
 - See example [query result](#) or access [using API](#)
 - Add data on presence (TRUE/FALSE) to **Dataset 1** by matching on DOI

Data analysis

Calculate percentage of compliant outputs:

- Meeting individual or all technical standards

Aspects of compliance can be further broken down by publisher, discipline and UKRI council (and research performing organisation, if desired).

Potential challenges and issues

Challenge	Mitigation
For a number of technical standards, data availability is currently limited	Align with UKRI's OA Technical Standards activity and supporting work by MoreBrains
A journal may not assign certain metadata (e.g. licenses, citations) for all its content	Decide on threshold value, Limit checks to specific publication types (where possible)
Journal can deposit citations in Crossref without adhering to I4OC standards	Check presence of citations in OpenCitations

⁴ Other identifiers that may be considered are Grid ID and Ringgold. ROR is an open identifier and the identifier used by journals when depositing metadata to Crossref.

Step H: Usage of research outputs in policy and practice

General information

M&E question(s)	<ul style="list-style-type: none"> • #11 How often are UKRI-funded (OA) publications cited/downloaded compared to UK-affiliated OA publications and to UK-affiliated non-OA publications? • #12 How often are UKRI-funded (OA) publications used/discussed (altmetrics) compared to UK-affiliated OA publications and to UK-affiliated non-OA publications?
Purpose of the question	<ul style="list-style-type: none"> • The purpose of these questions is to assess usage and impact of UKRI-funded OA publications, both within academia and within society and to assess impact of the UKRI OA policy compared to other OA publications by UK-affiliated authors. • For impact within society, citations in policy documents can demonstrate relevant usage. Discussions in news media and on social media can provide additional examples of usage outside academia and can help identify specific audiences, rather than be used as a mere quantitative measure. • Downloads from repositories can demonstrate the importance of repository-based access and assess the effectiveness (in terms of usage) of publications made available through repositories via Route 2 Licensing statements. Alternatively, Unsub data can give an estimate of usage of OA publications only available through repositories. • For books, several projects are currently underway that aim to collect usage information for OA books.
Source(s)	<ul style="list-style-type: none"> • Articles <ul style="list-style-type: none"> - Dataset 1, Step A - OpenAlex (for citations) - IRUS-UK (for repository downloads) - Overton (for use in policy documents) - Crossref Event Data (for e.g. Wikipedia, social media) - Lens (for patents) • Long-form outputs <ul style="list-style-type: none"> - As for articles (for research outputs with DOIs) - IRUS-UK for repository and OAPEN downloads - Data collected using <ul style="list-style-type: none"> ▪ Book Analytics Dashboard project [in development] ▪ OA Book Usage Data Trust [in development]
Mapping to the OA policy logic model	<ul style="list-style-type: none"> • Outcomes <ul style="list-style-type: none"> - Research findings are open, re-usable and repeatable - Improved discovery of research outputs • Impact <ul style="list-style-type: none"> - UKRI effectively stewards a more open and transparent R&I system - Improved understanding of ourselves and the world around us - Industrial challenges addressed - Societal challenges addressed

Table specification

Variable	Data type	Sample data
DOI/ISBN (from Dataset 1, Step A)	DOI / ISBN (alphanumeric)	10.26784/sbir.v5i1.312
Citations by year	Array of integers	2023: 6, 2024: 10
List of citing DOIs	List of DOIs (alphanumeric)	10.1007/s12022-022-09710-8
Downloads by year by source	Array of integers	2023: 100, 2024: 75
Citations in policy documents by year	Array of integers	2023: 2, 2024: 5
List of citing policy documents	List of PIDs/URLs	[no example]
News and media mentions per year by source	Array of integers	2023: 10, 2024: 5
List of news and media mentions	List of URLs	http://twitter.com/PaulKiem/statuses/832519346811310080
Citations in patents by year	Array of integers	2023: 1, 2024: 3
List of citing patterns	List of patent numbers, applicants	EP 19895501 A, SOLIDUS BIOSCIENCES INC

Data preparation

High-level overview

- For each record in **Dataset 1, Step A** (DOI/ISBN), collect usage data from selected sources
- Document both counts per year by source as well as URLs/PIDs of individual citations and mentions for further analysis.

Recommended approach

For each record (DOI/ISBN) in **Dataset 1, Step A**:

- Collect citation count from [OpenAlex](#)
- Search by OpenAlex IDs from **Dataset 1, Step A**
- Get citation count (variable: 'cited-by-count')
 - See example [API call](#)
- Add data to **Dataset 1** by matching on OpenAlex ID
- Collect citation data from [OpenAlex](#)
 - Search by OpenAlex IDs from **Dataset 1, Step A**
 - Get list of citing works with full metadata
 - See example [API call](#)
 - Add data to **Dataset 1** by matching on OpenAlex ID
- Collect repository usage data from [IRUS-UK](#)
 - Item-level view download data of repository records
 - Query by DOI or OAI
 - Export as csv or access via (closed) API
 - Add data to **Dataset 1** by matching on DOI/OAI
 - Access managed by Jisc
- Collect mentions in policy documents from [Overton](#)

Annex A – Data specification

Monitoring and evaluating the effectiveness of UKRI’s open access policy

- Search by DOI or set of DOIs
- Get list of citing policy documents, with organisations and policy areas
- Export as csv or use API access
- Add data to **Dataset 1** by matching on DOI
- Get usage data (news, social media) from [Crossref Event Data](#)
 - Search by DOI
 - Get list of events, including citations, newfeed of blogs and media, Wikipedia and social media
 - API access, see example [API call](#)
 - Add data to **Dataset 1** by matching on DOI
- Get patent data from [Lens](#)
 - Search by DOI or set of DOIs
 - Get list of citing patents, with patent owners
 - See [example query](#) for patents citing UK-affiliated research outputs, publication year 2022
 - Export as csv or JSON, or use API access
 - Add data to **Dataset 1** by matching on DOI

Data analysis

- Calculate distribution of values for each usage parameter for UKRI-funded and UK-affiliated OA and non-OA outputs
- Qualitative analysis: use mentions to identify audience groups for further impact assessment (including as target groups for qualitative research – see M&E question 20)

Analysis can be further broken down by discipline and different OA types.

Potential challenges and issues

Challenge	Mitigation
Effect of potential confounders (geopolitical, economic)	Ensure this is appropriately caveated as part of the M&E methodology and narrative
Crossref Event Data requires API access and accompanying technical expertise for data collection/processing	Approach as research project additional to core M&E framework. Consider Altmetric as an additional data source
Unclear how to match IRUS-UK records for non-DOI outputs (e.g. for books/booters)	Liaise with Jisc to establish workflow Use OAI identifier for repository content
JUSP data not yet available at record-level	Liaise with Jisc to follow progress
Most sources cannot be used for long-form outputs without DOIs	Given that this limitation cannot be fully mitigated, we recommend that UKRI limits the analysis to articles and long-form outputs with DOIs and ensure this is appropriately caveated as part of the M&E methodology and narrative
OA book usage data are still maturing; two international projects around OA eBook usage data are currently underway	Monitor project status and outcomes for when UKRI OA policy for long-form outputs comes into effect
Data for OA book usage may mature in advance of that for non-OA book usage	Monitor trends in OA book usage over time, instead of comparing OA vs non-OA book usage
The proportion of non-OA UKRI-funded outputs is likely to become very small in the future	Do not compare UKRI-funded OA and non-OA outputs, but rather compare UKRI-funded (OA) outputs to both UK-affiliated OA and non-OA outputs, to assess impact of the UKRI OA policy compared to other pressures or choices to publish OA among UK-affiliated authors.

Step I: Citation diversity

General information

M&E question(s)	<ul style="list-style-type: none"> #13 To what extent does OA affect the diversity of affiliation countries of authors citing published outputs, for UKRI-funded and UK-affiliated authors? (by discipline)
Purpose of the question	<ul style="list-style-type: none"> The purpose of this question is to monitor global reach and impact of OA publications by looking at affiliation countries of citing researchers, compared to non-OA publications in the same journals/disciplines.
Source(s)	<ul style="list-style-type: none"> Articles <ul style="list-style-type: none"> Dataset 1, Step A OpenAlex (for affiliation info of citing publications) Long-form outputs <ul style="list-style-type: none"> same, only for outputs with DOIs
Mapping to the OA policy logic model	<ul style="list-style-type: none"> Outcomes <ul style="list-style-type: none"> Research findings are open, re-usable and repeatable Improved discovery of research outputs Impact <ul style="list-style-type: none"> UKRI effectively stewards a more open and transparent R&I system Improved understanding of ourselves and the world around us

Table specification

Variable	Data type	Sample data
DOI / ISBN (from Dataset 1, Step A)	DOI / ISBN (alphanumeric)	10.26784/sbir.v5i1.312
List of citing DOIs (from Dataset 1, Step H)	List of DOIs (alphanumeric)	10.1007/s12022-022-09710-8
Research performing organisation(s) (ROR) for each citing DOI	ROR ID (alphanumeric)	03yrm5c26
Research performing organisation(s) country for each citing DOI	ISO 3166 (alpha-2)	GB

Data preparation

High-level overview

- Collect primary variables
- Match primary variables to citing DOIs by matching on DOI

Recommended approach

- For each citing DOI, get affiliations (ROR and country) from OpenAlex
 - See example [API call](#)
- Add variables for affiliations to citing records by matching on citing DOI
- Add collected data for citing records to cited DOI by matching on cited DOI

Data analysis

- For each output, calculate Shannon Entropy (or Shannon Index) and the Gini-Simpson Index (or Gini’s Diversity Index) as measures of citation diversity. Higher scores for these indices are indicators of more citation diversity

Annex A – Data specification

Monitoring and evaluating the effectiveness of UKRI’s open access policy

- Calculate median and mean diversity indexes based on different ways of grouping citation links: by research performing organisation, country, and field of research, for different OA types
- Results can be compared between UKRI-funded (OA) outputs and UK-affiliated OA and non-OA outputs
- Reference for data analysis: <https://doi.org/10.5281/zenodo.7099438>

Potential challenges and issues

Challenge	Mitigation
Effect of potential confounders (geopolitical, economic)	Ensure this is appropriately caveated as part of the M&E methodology and narrative
Calculation of diversity indexes requires more specific data analysis expertise	Approach as research project additional to core M&E framework
No citation data for long-form outputs without DOIs	Limit analysis to articles and long-form outputs with DOIs

Dataset 2: Financial information

This section describes the preparation of **Dataset 2**, which covers information on the financial costs of OA publishing. The collection of financial information draws on a smaller set of data sources but also draws heavily on **Dataset 1** in terms of identifying in-scope research outputs for which data need to be gathered.

General information

M&E question(s)	<ul style="list-style-type: none"> • #14 What is the proportion of UKRI OA block grant funding going towards gold OA (including diamond OA), hybrid OA (via TAs) and green OA (via investment in repository infrastructure and staff)? • #15 What is the (estimated) annual expenditure of institutions towards reading and publishing? (by publishing model) • #16 What is the number of OA publications funded via UKRI OA block grant awards/OA fund/other institutional means, and what is the estimated average article/book processing charge? (by discipline, journal, publisher)
Purpose of the question	<ul style="list-style-type: none"> • The purpose of these questions is to assess how costs for OA publishing under the UKRI OA policy are divided between OA models (and trends therein over time), as well as how these costs are covered (e.g. through UKRI OA block grant awards or institutional means) and how this changes over time. • This will also enable observations of whether the UKRI OA policy contributes to consolidation of spending on particular models, or vice versa, to diversification of financial models (e.g. through the Jisc OACF model).
Source(s)	<ul style="list-style-type: none"> • Articles <ul style="list-style-type: none"> - Dataset 1, Step A and Step B (output records and OA type) - Dataset 1, Step C (transitional agreements and transformative journals) - OA Switchboard (for research performing organisations or funders paying for APC, where applicable) - Jisc agreements (full OA journals, OACF) - Costs per article under Jisc agreements - List prices/OpenAPC where outside Jisc agreements - Survey/interviews of research performing organisations (on how OA block grant awards are spent) (see M&E question 17) • Long-form outputs <ul style="list-style-type: none"> - Listed prices on publishers’ websites/OpenBPC - Survey/interviews of research performing organisations (on how OA funds are spent) (see M&E question 17) - UKRI information on how OA funds are spent (note that the availability of this data will depend on how the UKRI fund for long-form outputs will be administered and the related data tracked and managed)
Mapping to the OA policy logic model	<ul style="list-style-type: none"> • Outputs <ul style="list-style-type: none"> - OA publishing is affordable for research organisations - Broader adoption of OA practices / publishing models for a diversity of publishers that cover UKRI authors e.g. transitional agreements - OA publishing models are sustainable in the long-term • Outcomes <ul style="list-style-type: none"> - Diverse OA publishing landscape that meets the needs of the research sector - Affordability and sustainability of OA publishing - Better value for money throughout the system

Table specification

Primary variables (record level)

Variable	Data type	Sample data
OA type (from Dataset 1, Step B)	String	Gold DOAJ
Covered under transitional agreement (from Dataset 1, Step C)	Logical	TRUE, FALSE
Covered under full OA agreement	Logical	TRUE, FALSE
Covered under OACF agreement	Logical	TRUE, FALSE
APC/BPC	Currency	\$1000
Source of APC/BPC info	String	Jisc expenditure information

Primary variables (level of research performing organisation)

Variable	Data type	Sample data
Total OA block grant award	Currency	£500,000
Total OA costs	Currency	£800,000
Percentage of block grant award spend by type of OA cost	Array (string, percentage)	Jisc transitional agreements, 50%
Percentage of total OA spend by type of OA cost	Array (string, percentage)	Jisc transitional agreements, 20%

Data preparation

High-level overview

- Collect primary variables, at record level and research performing organisation level.

Recommended approach

Record level, for each DOI (and ISBN for OpenBPC):

- Journals covered under transitional agreements (for affiliated authors)
 - **Dataset 1, Step C**
 - Information from Jisc article level metadata on organisations responsible for payment
 - Information from OA Switchboard on organisations responsible for payment
- Journals covered under full OA agreements (for affiliated authors)
 - Obtain information from Jisc (journal list + participating research performing organisations)
- Journals covered under OACF agreements (for affiliated authors)
 - Obtain information from Jisc (e.g. journal list + participating research performing organisations)
- APC/BPC
 - For journals in transitional agreements, obtain information from Jisc on publishing cost per article
 - For journals outside transitional agreements (including transformative journals), use publisher list price or average from OpenAPC
 - For long-form outputs, use publisher information or average from OpenBPC (publisher-level)

Annex A – Data specification

Monitoring and evaluating the effectiveness of UKRI’s open access policy

Research performing organisation level:

- Collect data via
 - UKRI: amount of OA block grant award per research performing organisation
 - Jisc: information (overall and by research performing organisation) on expenditure on transitional agreements, as well as full OA agreements and participation in OACF. For Read and Publish agreements expenditure is separated out to identify the read from the publish fee.
 - Qualitative research / survey to obtain information on OA spend from a sample of research performing organisations (see **Appendix C**)
- Classification of types of OA costs
 - Jisc transitional agreements
 - Jisc full OA agreements
 - Jisc OACF
 - Publisher agreements outside Jisc
 - APC costs paid directly (full gold and hybrid)
 - Diamond OA support directly
 - Infrastructure (repositories)
 - Administrative costs

Data analysis

- Calculate distribution of costs across OA types for overall outputs, and by discipline/publisher based on record level data
- Calculate distribution of costs across OA types for a sample of research performing organisations, based on research performing organisation level data

Potential challenges and issues

Challenge	Mitigation
Record level data on UKRI OA block grant spend no longer available	Use higher-level data from grant reporting and from survey/interviews of selected research performing organisations (see M&E question 16)
Transformative deals comprise both reading and publishing costs	Use Jisc breakdown of costs in transitional agreements
Limited information on costs charged for OA for long-form outputs	Consider survey of UKRI-funded authors of OA long-form outputs to get information on costs and how these were covered

Qualitative research

This section provides information on the qualitative exploration of M&E questions. Specific sub-questions for investigation are outlined, building on the methods and recommendations in the main report.

Qualitative and mixed-methods research

General information

M&E question(s)	<ul style="list-style-type: none"> • #17 What do institutions expect/experience to be the main challenges/opportunities arising from UKRI’s OA policy? (incl. around EDI, career progression, research evaluation) • #18 What do publishers expect/experience to be the main challenges/opportunities arising from UKRI’s OA policy? (incl. around EDI, career progression, research evaluation) • #19 What do researchers expect/experience to be the main challenges/opportunities arising from UKRI’s OA policy? (incl. around EDI, career progression, research evaluation) • #20 What difference has access to OA outputs made for non-academic stakeholders? (e.g. industry, general public, practitioners)
Purpose of the question	<ul style="list-style-type: none"> • The purpose of these questions is to study expectations and experiences around the UKRI OA policy among various stakeholders over time, including impact of OA on society (beyond academia) • A longitudinal approach allows UKRI to assess whether initial expectations (positive and negative) around the UKRI OA policy are met and to follow changes in experiences over time
Source(s)	<ul style="list-style-type: none"> • Potential formats (see Overview of qualitative research methods below) <ul style="list-style-type: none"> - Interviews - (Longitudinal) survey - Focus groups - Case studies/desk research • Target groups: <ul style="list-style-type: none"> - Research performing organisations - Publishers - Researchers - Non-academic stakeholders (e.g. industry, general public, practitioners)
Mapping to the OA policy logic model	<ul style="list-style-type: none"> • Outputs <ul style="list-style-type: none"> - Authors maintain a sufficient level of choice as to where their article will be published • Outcomes <ul style="list-style-type: none"> - Diverse OA publishing landscape that meets the needs of the research sector - More efficient workflows and reduction in bureaucracy - Affordability and sustainability of OA publishing • Impact <ul style="list-style-type: none"> - UKRI effectively stewards a more open and transparent R&I system - Improved understanding of ourselves and the world around us - Industrial challenges addressed - Societal challenges addressed

Overview of qualitative research approaches

The table below provides an overview of qualitative research methods. This information is intended to feed into UKRI’s choices of research methodologies to deliver the qualitative elements in the M&E approach, to ensure these

Annex A – Data specification

Monitoring and evaluating the effectiveness of UKRI’s open access policy

are suitable to meet the intended policy objectives (e.g. statistically representative results vs. narrative and nuanced evidence to inform storytelling). We give some specific recommendations for research methods together with suggested questions for each target group (see **Appendix C**).

It should be noted that case studies are not included as a standalone research method, as any of the below approaches can be used to inform the preparation of case studies.

Research Method	Research objectives	Key features	Practical requirements and planning
Interview	<i>Gather in-depth perspectives from individuals</i>	<ul style="list-style-type: none"> • Semi-structured format allows deep dives into specific topics • Focused on listening 	<ul style="list-style-type: none"> • 1 hour per interview • Development of interview scripts • Invitations and scheduling • Recording and transcription
Focus group	<i>Gather in-depth perspectives from groups</i>	<ul style="list-style-type: none"> • Range of participants allows for a breadth of views • Focused on back-and-forth dialogue • Allows the researcher to gauge reactions and thoughts around a topic 	<ul style="list-style-type: none"> • 1.5 hours per session • Preparation of activities and supporting materials • Invitations and scheduling • Recording and transcription
Literature review	<i>Develop an understanding of existing research on a topic</i>	<ul style="list-style-type: none"> • Broad and diverse range of perspectives captured across the same topic, based on expert views expressed in a chosen set of documents 	<ul style="list-style-type: none"> • Identification of appropriate data sources • Development of search strategy, including inclusion/exclusion criteria for sources of information
Survey	<i>Produce a snapshot of attitudes and opinions of a population</i>	<ul style="list-style-type: none"> • Qualitative and quantitative insights gathered across a broad range of topics • Broad range of stakeholder views, including potential for statistical analysis • May also include the identification of participants for potential further in-depth engagement (e.g. interviews) 	<ul style="list-style-type: none"> • Questionnaire design • Invitations and respondent management • (Optional) reward mechanisms to encourage participation
Workshop	<i>Provide space for discussion, problem solving or validation</i>	<ul style="list-style-type: none"> • Range of participants allows for a breadth of views • Focused on group exercises and activities to solve chosen problems or address issues 	<ul style="list-style-type: none"> • Up to 3 hours per session • Preparation of activities and supporting materials • Invitations and scheduling • Works best in person due to the focus on group activities

Selection of research participants

Approaches to engage contributors to qualitative research are outlined in the main report, with reference to specific M&E questions.

Additional strategies to select participants may include the following:

- Research performing organisations
 - selection based on share of UKRI-funded research outputs
 - selection based on rate of (non)compliance

Monitoring and evaluating the effectiveness of UKRI’s open access policy

- selection based on size (e.g. research intensive, small and specialist)
- participation in prior scoping research for UKRI M&E framework
- Publishers
 - selection based on share of UKRI-funded research outputs
 - selection based on rate of (non)compliance (e.g. when certain publishers have higher rates of non-compliant outputs)
 - selection based on OA models offered
 - selection based on size, discipline, publication type
 - participation in prior scoping research for UKRI M&E framework
- Researchers
 - self-selection (open call for participation, potentially limited to UKRI-funded researchers)
 - selection based on research performing organisation (when recruiting via research performing organisation)
 - selection based on research discipline
- Non-academic stakeholders (e.g. industry, general public, practitioners)
 - identification through usage data of OA outputs (patents, policy citations)
 - outreach through professional organisations (e.g. trade bodies, patient organisations)

Suggested topics and questions

- See Appendix C

Data analysis

Depending on the research methodology selected for a particular topic and population, results can be analysed and presented in various ways:

- Topic coding of free-form answers (from interviews or free-text response questions in surveys), followed by either narrative summarisation or tabulation of e.g. frequency of answers.
- Presentation of ‘vignettes’ highlighting representative opinions, or particularly illustrative examples
- Quantitative analysis of survey questions with pre-set answer options - depending on whether answer options are quantitative (discrete or continuous variables) or qualitative (nominal or ordinal variables).

Potential challenges and issues

Challenge	Mitigation
Survey fatigue	<ul style="list-style-type: none"> Limit number of questions Merge M&E approach into existing outreach activities (including surveys) Combine topics in one questionnaire Send out questionnaires on different topics to different participants Choose research methods requiring fewer participants
Selection bias	<ul style="list-style-type: none"> Adjust the target objectives (representativeness vs narrative evidence) when selection bias cannot be prevented

Appendices

Appendix A – Sources used to identify potential M&E questions

The following sources were consulted to identify a longlist of M&E questions to be subsequently prioritised, as described in the main report.

Title	Organisation	Year
UKRI OA policy	UKRI	2021
Equality impact assessment of the UKRI Open Access policy	UKRI	2021
Economic implications and benefits of updated UKRI Open Access policy	UKRI Alma Economics	2021
Monitoring the effects of Plan S on Research and Scholarly Communication	cOAlition S	2020
Monitoring the transition to open access	Universities UK	2017
Briefing paper on open access monitoring	Science Europe	2021
Collecting data on open access publications	Jisc	2022
Managing open access publication workflows and compliance	Jisc	2021
Metadata to support the UKRI Open Access policy: landscape and community readiness analysis	UKRI MoreBrains	2022
UK Concordat to Support Research Integrity	Universities UK	2019
Review of research bureaucracy	UK Government Dept BEIS UKRI	2021
Monitoring the open access policy of Horizon 2020	European Commission	2021
Monitoring open access publishing of NWO funded research	NWO CWTS	2022
French Open Science Barometer	French Ministry of Higher Education and Research	2022
Austrian Science Fund (FWF) Open Access Compliance Monitoring 2021	FWF	2022
Austrian Transition to Open Access (AT2OA ²)	Austrian Transition to Open Access 2 project	2021
DFG Funding Programme Open Access Publishing - Report about the Funding	DFG	2020
Developing a Pilot Data Trust for Open Access Ebook Usage (2020-2022)	UNT Curtin University BISG University of Michigan Educopia	2021
Book Analytics Dashboard Project (2022-2025)	Curtin University COARD OAPEN Educopia	2022
OA Book Usage Data Trust	UNT OpenAIRE OPERAS	2022
Diversifying readership through open access: A usage analysis for open access books	SpringerNature COARD	2021
COPIM project (various outputs)	COPIM	Ongoing
UKRI Gap Analysis of Open Access Monographs Infrastructure	UKRI OAPEN	2021
An Analysis of the Current Bibliographical Data Landscape in the Humanities. A Case for the Joint Bibliodata Agendas of Public Stakeholders	DARIAH-ERIC consortium	2022
Briefing Paper on Open Access to Academic Books	Science Europe	2019
A Landscape Study on Open Access and Monographs: Policies, Funding and Publishing in Eight European Countries	Knowledge Exchange	2017

Appendix B – Recommended data sources and relevant metadata fields

We recommend a variety of (mostly open) data sources for use in the M&E framework. As an example, the tables below list the relevant metadata fields to extract from Gateway to Research, Crossref and OpenAlex.

Gateway to Research

Output type	Metadata concept	Variable name	Values relevant for M&E
Articles / Long-form	Data source-specific UID (record)	GtR Outcomeld	all
Articles / Long-form	DOI	DOI	all
Articles / Long-form	UKRI council	FundingOrg	all
Articles / Long-form	Publication year	Year	all years monitored
Articles / Long-form	Publication type	PublicationType	Journal Article/Review, Systematic review, Conference/Paper/Proceeding/Abstract, Book Chapter, Book, Book edited, Monograph

Crossref

Output type	Metadata concept	Variable name	Values relevant for M&E
Articles / Long-form	DOI	DOI	all
Long-form	ISBN	ISBN	all
Articles / Long-form	Funder ID	funder.DOI ⁵	10.13039/100013266 10.13039/100014013 10.13039/100014570 10.13039/501100000265 10.13039/501100000266 10.13039/501100000267 10.13039/501100000268 10.13039/501100000269 10.13039/501100000270 10.13039/501100000271

⁵ Values for UKRI-related funder IDs taken from <https://doi.org/10.5281/zenodo.5562842>

Annex A – Data specification

Monitoring and evaluating the effectiveness of UKRI’s open access policy

Output type	Metadata concept	Variable name	Values relevant for M&E
			10.13039/501100000690 10.13039/501100006041 10.13039/501100007849 10.13039/501100009187 10.13039/501100011027 10.13039/501100012508 10.13039/501100013341 10.13039/501100013589 10.13039/501100013915 10.13039/501100014813 10.13039/501100014814 10.13039/501100018959 10.13039/501100019328
Articles / Long-form	Publication date	published	For all years monitored
Articles / Long-form	Publication type	type	journal-article, proceedings-article, proceedings, book-chapter, book, monograph, book-section, book-part
Articles / Long-form	ORCID	author.ORCID	all
Articles / Long-form	Authenticated ORCID	author. authenticated_orcid	all
Articles / Long-form	ROR	author. affiliation.id.id (where author.affiliation.id.id_type = 'ROR')	all
Articles / Long-form	License	license.URL	all
Articles / Long-form	Version to which license applies	license.content_version	vor, am
Articles / Long-form	ISSN	ISSN	all
Articles / Long-form	Publisher Member ID	member	all
Articles / Long-form	Number of references in Crossref	references_count	all

Annex A – Data specification

Monitoring and evaluating the effectiveness of UKRI’s open access policy

OpenAlex

Output type	Metadata concept	Variable name	Values relevant for M&E
Articles / Long-form	Data source-specific UID	id	all
Articles / Long-form	DOI	doi	all
Articles / Long-form	Publication year	publication_year	all years monitored
Articles / Long-form	Publication type	type	journal-article, proceedings-article, proceedings, book-chapter, book, monograph, book-part
Articles / Long-form	Affiliation – data source specific UID	authorships.institutions.id	all
Articles / Long-form	Affiliation - ROR	authorships.institutions.ror	all
Articles / Long-form	Affiliation - name	authorships.institutions.display_name	all
Articles / Long-form	Affiliation - country	authorships.institutions.country_code	all
Articles / Long-form	ISSN	host_venue.issn	all
Articles / Long-form	ISSN-L	host_venue.issn_l	all
Articles / Long-form	Publisher	host_venue.publisher	all
Articles / Long-form	Number of citations	cited-by-count	all
Articles / Long-form	List of citing works	cited_by_api_url	all
Articles / Long-form	Discipline	concepts.display_name (where concepts.level = 0)	all

Appendix C – Qualitative research questions

Background

We recommend that qualitative questions for research performing organisations and publishers are addressed via advisory input, for example by leveraging a panel of organisations who may provide input regularly (e.g. via focus groups or workshops). This could cover reasons for non-compliance (M&E question #7), the challenges and opportunities arising from UKRI’s OA policy as well as its impact on EDI considerations, career progression and researcher evaluation. We note that existing engagement mechanisms that UKRI has in place (e.g. Task & Finish, consultation or user groups) may be expanded to include input in these areas, to minimise burdens on these stakeholders. A key consideration will be to ensure representation from a diverse cohort of organisations (e.g. size, turnover, disciplinary focus).

We recommend that views from researchers on challenges and opportunities arising from UKRI’s OA policy as well as its impact on EDI considerations, career progression and researcher evaluation are sought in a more structured way given their far larger numbers compared to research performing organisations or publishers. Furthermore, in the case of researchers, it is also difficult to identify suitable individuals with protected characteristics, which indicates that casting a wide net is likely to be needed to gather the input required (in contrast, organisations can talk about their approaches to EDI in general terms). We would recommend the use of an online survey as a starting point, potentially complemented by a set of in-depth interviews depending on available resources and level of insight required. Such a survey may either be added to existing mechanisms (smaller number of questions) or be run as a standalone exercise (larger number of questions).

Regarding questions on challenges and opportunities of the UKRI OA policy, a longitudinal approach, where organisations are asked periodically (e.g. yearly) to reflect on experiences in each of the previously identified areas of challenges and opportunities, can inform additional insights. In the case of researchers, a cohort approach can be considered (where a group of respondents is followed over time), or a new representative group of researchers can be selected each year. In the first year, questions can be open-ended to identify expected challenges and opportunities rather than using pre-set answer options to prompt responses.

Research performing organisations – challenges and opportunities of UKRI OA policy

Questions

What are the biggest challenges your organisation faces / expects to face in complying with the UKRI OA policy?

What are the biggest opportunities (your organisation expects) to arise from the UKRI OA policy?

How have the UKRI OA policy and funding mechanisms affected organisational developments in support of OA? (e.g. training, support provision, research collaborations and partnerships)

To what extent is the guidance provided by UKRI on its OA policy requirements appropriate and sufficient?

What (internal or external) supporting infrastructure, services, tools and technologies are available to your organisation to implement the UKRI OA policy?

To what extent are the current funding levels appropriate to meet the UKRI OA policy requirements? To what extent, if at all, are researchers restricted in OA publishing due to limited (or not) access to funds?

How could observed reasons for non-compliance with the UKRI OA policy be addressed? (building on M&E question #7)

Selection of questions on EDI (see below)

For research performing organisations, a separate area to be addressed with qualitative research is open access spend, in relation to M&E questions #14-16. For more information on this, see the section ‘Research performing organisations – open access spend’ (p.47).

Publishers – challenges and opportunities of UKRI OA policy

Questions

What are the biggest challenges you (expect to) face in implementing the UKRI OA policy?

What are the biggest challenges reported by submitting authors needing to comply with the UKRI OA policy?

What are the biggest opportunities (you expect) to result from the UKRI OA policy?

To what extent is the guidance provided by UKRI on its OA policy requirements appropriate and sufficient?

What (internal or external) supporting infrastructure, services, tools and technologies are available to you to implement the UKRI OA policy?

To what extent has the UKRI OA policy influenced industry developments, including the development of new infrastructures and collaborations?

To what extent has UKRI’s OA policy affected your publishing business model(s)?

To what extent have UKRI’s funding mechanisms affected your publishing business model(s) and workflows?

To what extent do submitting authors report issues related to current funding levels to meet the UKRI OA policy requirements?

How could observed reasons for non-compliance with the UKRI OA policy be addressed? (building on M&E question #7)

Selection of questions on EDI (see below)

Researchers – challenges and opportunities of UKRI OA policy

Questions

What are the biggest challenges you (expect to) face in complying with the UKRI OA policy?

What are the biggest opportunities (you expect) to result from the UKRI OA policy?

To what extent do you feel you can publish in the venues you consider most appropriate for your research, when you are funded by UKRI?

To what extent, if any, do you expect the UKRI OA policy to affect your career opportunities (either positively or negatively)?

Have there been occasion(s) where you have not been able to comply with UKRI’s OA policy’s terms? If so, for what reason(s)?

To what extent do you perceive open access publications to lead to outcomes such as increased use, a quicker and more efficient research process or improvements in innovation?

Selection of questions on EDI (see below)

Non-academic stakeholders – Impact of open access

Our recommendation regarding views of non-academic stakeholders is to seek to develop narrative case studies, building on publicly available information as well as in-depth engagement with specific stakeholder groups. The recommended approach for delivery would include an investigation of quantitative data on usage of OA outputs to identify individuals and organisations to be reviewed and engaged.

Questions

What difference has access to OA outputs made for your organisation/activities? Please consider the following dimensions: increased access to knowledge, quicker access to knowledge, reduced costs for access to information, reduced R&D costs (due to more information being available), higher efficiency.

What do you consider to be the biggest challenges and opportunities related to OA?

Equality, diversity and inclusion

EDI questions considered as part of the M&E framework may cover a range of potential areas, following [UKRI's Equality Impact Assessment](#) of the OA policy:

1. Protected characteristics
 - a. Disability
 - b. Pregnancy and maternity
 - c. Race
 - d. Religion or belief
 - e. Sexual orientation
 - f. Sex (gender)
 - g. Age
2. Other EDI areas
 - a. Early career researchers (ECRs)
 - b. Those in low- and middle-income countries (LMICs)
 - c. Those with career breaks/alternative career paths
 - d. Those experiencing language barriers

We recommend that the M&E approach does not include a subset of protected characteristics that UKRI have considered should not be unduly affected by their OA policy: religion or belief; sexual orientation; gender reassignment; and marriage or civil partnership.

For most EDI questions, the chosen approach to research will strongly affect the wording, tone and number of questions asked. For example:

- in the case of surveys, screening questions may be required in some cases, to gauge whether respondents identify as being part of a protected group prior to asking relevant questions (e.g. 'Do you consider yourself to be a disabled person?': Yes/No + 'To what extent has your disability affected....?')
- in the case of interviews, focus groups, workshops or case studies, it would be necessary to identify individuals who identify as part of a protected group prior to engaging them

Notably, the recruitment of individuals who identify as part of a protected group suffers from some limitations, as it assumes that the individual already identifies by a given characteristic. This, however, is not always the case, and it is acknowledged that individuals who would come under the Equality Act definition may not, for example, identify as being disabled. Additionally, some individuals may not feel comfortable contributing from this perspective. We recommend that UKRI acknowledges these considerations as limitations as part of the M&E methodology, as fully resolving them would require disproportionate efforts compared to other M&E areas.

In the case of individuals, we recommend that all questions are asked to all contributors rather than pre-selecting or mapping questions to EDI characteristics (i.e. asking a question only to individuals with a given EDI characteristic). This will give contributors the opportunity to state whether a question might not apply to them, as appropriate, and will mitigate potential biases in questionnaire design.

We note that the questions below may be answered by a mix of research approaches and do not all need to be part of the same research exercise. The two tables below outline our recommended questions on EDI for individuals and organisations.

EDI questions for individual researchers

Questions
To what extent are you able to discuss implementation challenges with UKRI’s OA team?
To what extent has UKRI’s OA policy affected your ability to publish in your preferred publishing venue, and what (if any) implications has this had regarding your [protected characteristic]?
Did UKRI’s OA policy lead to unintended consequences that unduly affect any of your protected characteristics?
The transition to OA has been improving access to articles and long-form outputs online, without needing to attend a physical location. To what extent has this made it easier for you to read relevant materials?
If you have been on long-term leave (incl. career breaks) during the publication of a relevant research output, to what extent has this affected your ability to meet the expectations of UKRI’s OA policy?
Did you take advantage of UKRI’s policy exception that allows for extensions in these circumstances, and to what extent was this beneficial?
To what extent is UKRI’s OA policy leading towards more responsible use of metrics in research(er) evaluation?
To what extent are you familiar with the concept of OA and its practical implementation?
To what extent does your level of understanding affect your ability to meet the requirements of UKRI’s OA policy?
To what extent have you been able to access ODA project grants to publish in your desired venue?
What channels are available for you to communicate with UKRI with regard to your LMIC status?
To what extent is UKRI’s OA policy aligned with other OA requirements in your country?
To what extent is the fact that a large proportion of publications are written and shared in English a barrier when carrying out and publishing your own research?
To what extent has your ability to access translated content been affected by the licensing terms applied to UKRI-funded work?
To what extent have technological barriers prevented you from meeting the expectations of UKRI’s OA policy?
What funding mechanisms enable you to publish OA in line with the expectations of UKRI’s OA policy?
To what extent are these sufficient and in line with your needs as a researcher?

EDI questions for organisations (publishers and research performing organisations)

Questions	Key audience	Recommended characteristics
To what extent is your organisation able to discuss implementation challenges with UKRI’s OA team?	Research performing organisations	All
What support do you provide to individuals with a protected characteristic, to overcome any barriers they may experience in meeting UKRI’s OA policy?	Publishers, Research performing organisations	All
From your perspective, to what extent has UKRI’s OA policy made provision to	Research performing organisations	All

Questions	Key audience	Recommended characteristics
support individuals with a protected characteristic?		
To what extent is UKRI’s OA policy leading towards more responsible use of metrics in research(er) evaluation?	Publishers, Research performing organisations	All
To what extent are your [researchers/authors] familiar with the concept of OA and its practical implementation?	Publishers, Research performing organisations	All
To what extent does their level of understanding affect their ability to meet the requirements of UKRI’s OA policy?		

Research performing organisations – open access spend

Questions on OA spend can be asked in the form of an institutional survey, either for all research performing organisations or for a (representative) sample of these. The aim of these questions is to retrieve information on OA spend, including UKRI OA block grant awards, that cannot be retrieved via existing sources, such as regular grant reporting, Jisc article level data, OA Switchboard and Jisc data on subscriptions and expenditure.

Notably, we recommend that, as part of research performing organisations, research council institutes are considered and engaged, too, as a parallel stakeholder group. Research council institutes are subject to the OA policy, but funding for OA costs associated with their core business is dealt with separately to the block grant on a council-by-council basis. However, research council institutes may additionally receive OA block grant award if they are awarded competitive UKRI grants. This nuance is discussed in Annex A, as part of qualitative M&E questions.

Questions below build on the survey used for the Economic implications and benefits of updated UKRI Open Access policy assessment. These also include questions on cost and time spent on administering OA, which could be used to monitor administrative burden.

Questions for research performing organisations on OA spend – including allocation and use of UKRI OA block grant awards

Questions
Please estimate your library’s total allocated budget in the financial year 202X/202Y for reading and publishing (including UKRI OA block grant award and any other funding related to open access)
Please estimate, by category, the total spend by your research performing organisation in the financial year 202X/202Y on the following categories:
<ul style="list-style-type: none"> • Subscriptions • Transitional agreements with publishers (negotiated by Jisc) • Agreements with full OA publishers negotiated by Jisc • Agreements with full OA publishers outside those negotiated by Jisc • Support for non-APC/BPC OA publishing (including OACF and Subscribe2Open) negotiated by Jisc • Support for non-APC/BPC OA publishing (including Subscribe2Open) outside those negotiated by Jisc • APCs/BPCs paid directly (outside publisher agreements) • Green open access (repository and staff costs)

Annex A – Data specification

Monitoring and evaluating the effectiveness of UKRI's open access policy

Please estimate the proportion of UKRI OA block grant award spend on the following categories in the financial year 202X/202Y:

- Transitional agreements with publishers
- Agreements with full OA publishers
- Support for non-APC/BPC OA publishing (including OACF and Subscribe2Open)
- APCs/BPCs paid directly (outside publisher agreements)
- Green open access (repository and staff costs)
- Administrative costs

Please estimate the proportion of all (i.e. not UKRI-funded only) articles/long-form outputs for which researchers made use of their personal/departmental research budget to cover the cost of APCs/BPCs in the financial year 202X/202Y

Please estimate the proportion of long-form outputs for which BPCs were paid by central funds that were not covered by UKRI OA fund for long-form outputs for the financial year 202X/202Y

Appendix D – Metadata to support the UKRI Open Access policy (MoreBrains report)

MoreBrains Cooperative was commissioned by UKRI in August 2021 to carry out a landscape and community readiness analysis of metadata required to support the UKRI Open Access policy. Their final report “[Metadata to support the UKRI Open Access Policy: Landscape and community readiness analysis](#)” summarises both the availability of metadata fields in common metadata schemes and application profiles (JATS and Crossref for journals, RIOXX, DublinCore/OpenAIRE and DataCite for repositories) as well as the actual coverage of certain key metadata fields. Summary tables for both are included below as a point of reference.

An important remark in the MoreBrains report about Dublin Core is that as a schema, it is deliberately flexible in design, in order to support as wide a range of applications as possible. Consequently, without the use of an application profile like RIOXX or OpenAIRE to prescribe how the Dublin Core fields are used, consistency between repositories, and thus consistent retrieval of metadata is effectively impossible. The MoreBrains report notes that of 132 UK repositories analysed together with CORE, 65 (49%) used the RIOXX application profile and 67 (51%) used Dublin Core. For the latter, implementation of OpenAIRE guidelines was reported to be extremely variable, preventing meaningful computational assessment of metadata completeness at the time of the study.

The MoreBrains study focuses on the availability of metadata at record level and recommends working with standard organisations, publishers and universities to fill existing gaps in standards and increase uptake by publishers and repositories. This will likely be a medium- to long-term strategy. Pending this, in the current M&E framework, we have focused on those metadata elements that are relatively mature, as well as those that, while not (yet) available at record level, might be sourced at journal or repository level.

Journal-relevant metadata fields in the policy mapped to JATS and the Crossref schema

Requirement	Output type	For journals	JATS	Crossref
Article level PID (DOI, URN, Handle)	Articles	Mandatory	Yes	Yes
ORCID (all authors)	Articles	Mandatory	Yes	Yes
Authenticated ORCID	Articles	Encouraged	Yes	Yes
Licence (non-proprietary format)	Articles / Long-form	Mandatory	Yes	Yes
Preservation location (Portico etc.)	Articles	Mandatory	No	Yes
Self-archiving policy registered in Sherpa Romeo	Articles	Mandatory	No	No
Citation data according to I4OC standards ⁶	Articles	Mandatory	No	No
PID for funders	Articles	Encouraged	Yes	Yes
PID for research performing orgs	Articles	Encouraged	Yes	Yes
PID for grants	Articles	Encouraged	Yes	Yes
PID for projects	Articles	Encouraged	No	No

⁶ In the MoreBrains report, this is stated as "Citations in I4OC (<https://opencitations.net>)"

Annex A – Data specification

Monitoring and evaluating the effectiveness of UKRI’s open access policy

Journal-relevant metadata fields in the policy – current status

Requirement	Output type	For journals	Current status
Article level PID (DOI, URN, Handle)	Articles	Mandatory	Metadata fields are available and well-described with adequate adoption
ORCID (all authors)	Articles	Mandatory	Metadata fields are available but are poorly used or misused
Authenticated ORCID	Articles	Encouraged	There are no metadata fields and no mechanism to monitor levels of compliance
Licence (non-proprietary format)	Articles / Long-form	Mandatory	Metadata fields are available and well-described, but adoption needs to improve
Preservation location (Portico etc.)	Articles	Mandatory	There are no metadata fields and no mechanism to monitor levels of compliance
Self-archiving policy registered in Sherpa Romeo	Articles	Mandatory	There are no metadata fields and no mechanism to monitor levels of compliance
Citation data according to I4OC standards ⁷	Articles	Mandatory	There are no metadata fields and no mechanism to monitor levels of compliance
PID for funders	Articles	Encouraged	Metadata fields are available and well-described, but adoption needs to improve
PID for research performing orgs	Articles	Encouraged	Metadata fields are available but are poorly used or misused
PID for grants	Articles	Encouraged	Metadata fields are available but are poorly used or misused
PID for projects	Articles	Encouraged	There are no metadata fields and no mechanism to monitor levels of compliance

Repository-relevant metadata fields in the policy mapped to RIOXX version 3, OpenAIRE, and the DataCite schema

Requirement	Output type	For repositories	RIOXX	OpenAIRE	DataCite
Article level PID (DOI, URN, Handle)	Articles	Mandatory	Yes	Yes	Yes
ORCID (all authors)	Articles	Mandatory	Yes	Yes	Yes
Authenticated ORCID	Articles	Mandatory ⁸	No	No	Yes
Licence (non-proprietary format)	Articles / Long-form	Mandatory	Yes	Yes	Yes

⁷ In the MoreBrains report, this is stated as "Citations in I4OC (<https://opencitations.net>)".

⁸ In the MoreBrains report, this requirement was given as 'Mandatory' for repositories and 'Encouraged' for publishers.

Annex A – Data specification

Monitoring and evaluating the effectiveness of UKRI's open access policy

Requirement	Output type	For repositories	RIOXX	OpenAIRE	DataCite
Preservation location (Portico etc.)	Articles	Mandatory	No	No	Yes
Registered in OpenDOAR	Articles / Long-form	Mandatory	No	No	No
PID for funders	Articles	Mandatory	Yes	Yes	Yes
PID for research performing orgs	Articles	Encouraged	No	No	Yes
PID for grants	Articles	Encouraged	No	Yes	No
PID for projects	Articles	Encouraged	No	No	No
Final version or AAM	Long-form	Mandatory	Yes	Yes	No

Repository-relevant metadata fields in the policy – current status

Requirement	Output type	For repositories	Current status
Article level PID (DOI, URN, Handle)	Articles	Mandatory	Metadata fields are available but are poorly used or misused
ORCID (all authors)	Articles	Mandatory	Metadata fields are available but are poorly used or misused
Authenticated ORCID	Articles	Mandatory ⁹	Metadata fields are available but are poorly used or misused
Licence (non-proprietary format)	Articles / Long-form	Mandatory	Metadata fields are available but are poorly used or misused
Registered in OpenDOAR	Articles / Long-form	Mandatory	There are no metadata fields and no mechanism to monitor levels of compliance
PID for funders	Articles	Mandatory	Metadata fields are available but are poorly used or misused
PID for research performing orgs	Articles	Encouraged	There are no metadata fields and no mechanism to monitor levels of compliance
PID for grants	Articles	Encouraged	There are no metadata fields and no mechanism to monitor levels of compliance
PID for projects	Articles	Encouraged	There are no metadata fields and no mechanism to monitor levels of compliance
Final version or AAM	Long-form	Mandatory	Metadata fields are available but are poorly used or misused

⁹ In the MoreBrains report, this requirement was given as 'Mandatory' for repositories and 'Encouraged' for publishers.