Jisc

# Open data about research management

## A landscape review

Published February 2025

# Contents

# Open data about research management
# A landscape review

**Report commissioned by**:

Jisc
4 Portwall Lane
Bristol
BS1 6NB
**jisc.ac.uk**

**Contact**:

Victoria Moody
Director, higher education and research
**victoria.moody@jisc.ac.uk**

**Report authors**:

Victoria Moody, Rob Johnson, Angharad Roberts, Andrea Chiarelli, Frances Palmer, Enric Fuster, Nicandro Bovenzi

**research-consulting.com**

Contact:

**rob.johnson@researchconsulting.com**

Report published: January 2025

# Executive summary

## Executive summary

**Jisc commissioned Research Consulting, working with SIRIS Academic, to coordinate a landscape review of open data produced in the process of research management and administration in the UK, and of the platforms that aggregate and give access to these data.**

We wanted to develop a high-level understanding of the scope, extent and potential of open data about research management and administration, and to identify ways to improve the landscape and explore the potential for open data as a sector asset to support efficiency and insight.

## Project scope

The scope of this work includes open data, or data that could potentially be open, about research from across the research lifecycle, from hypothesis definition to research impact, including all aspects of the research management and administration landscape. It excludes research data that is produced or collected in the course of a research project.

The data in scope are not personally, commercially or otherwise sensitive or disclosive. The data may comprise aggregated, anonymised or linked data from a range of sources and, when combined in aggregated, linked form, the data would also not be personally, commercially or otherwise sensitive or disclosive.

The project is a preliminary review that can provide a foundation for future focus. It does not propose infrastructure solutions or specific mechanisms for interoperability but it does describe some of the challenges in the current landscape and identify potential for improvement.

It identifies ways to think about improving how open data about research management and administration are managed and used as a sector asset. It does not comment on the collection methods or approaches, frequency of collection, quality (except around systemic areas such as meeting open criteria, metadata or Application programming interface (API) provision) or utility of the data in scope.

The approach and scope set out a first step to support further enquiry. They do not include a more comprehensive and programmatic focus on developing the opportunities and specifying actions. The landscape of open data about research management and administration is beginning to emerge and this work presents a current snapshot.

# High-level findings

## Available data are comprehensive with many areas of excellent practice

The **Open Data Institute** summarises open data (as) "data that anyone can access, use or share".

The landscape of data about research management and administration is comprehensive. It is an important resource for analysis and insight into the research and innovation system.

Significant amounts of diverse data are created as a result of research management and administration, many of which are available, or could be available, as open data.

Some is well managed and accessible under open licences or conditions. Some may be available but not accessible consistently as open data; other data has potential to be available as open data but is not accessible.

## Key actors producing data about research management and administration

Key actors within the UK research landscape include funding agencies such as **UKRI** and the **Scottish Funding Council** as well as higher education institutions (HEIs), research commissioners such as the **National Institute for Health and Care Research (NIHR)**, infrastructure investments, the public sector, independent and private research organisations, commercial organisations, member organisations such as **Universities UK (UUK)**, **GuildHE** and **MillionPlus**, professional bodies such as **ARMA** and charity sector bodies including **Wellcome**, **Jisc** and **HESA**.

## A complex landscape

Research infrastructure investments, public sector and independent and private research organisations, member bodies, professional groups and commercial organisations and charitable sector organisations independently produce data and intersect with government, funders and higher education institutions to produce data. The four nations of the UK have different processes for managing research.

The complex landscape includes data that can duplicate or consolidate existing resources, data collection activities and data service platforms, in some cases overlapping with other data and adding to complexity and bureaucracy.

Access routes are mediated in many different ways and this can reduce the data's quality and interoperability. Metadata availability and alignment between data sources is patchy and rare. Application Programming Interfaces (APIs) exist in some areas to enable programmatic data analysis but they are not consistent or their standards are not always defined.

We have explored open data about research management and administration using a Jisc-defined, six-tier typology to review the data types, which range from open data to internal-only data.

The typology is as follows:

- **Fully open data** that meet open data criteria and in some cases elements of the **FAIR Principles**
- **Partially open data** that fall short of meeting open data criteria
- **Publicly available data** that may be embedded in individual webpages or within proprietary formats and platforms
- **Report/platform-mediated data**, where data may be held as tables in reports, or search-mediated data in platforms that do not offer a full download of the data associated with an open data licence
- **Managed access data** requiring subscription or registration for access
- **Data from internal systems not open beyond the organisation that produces it**, which could potentially be open

Participants in this project said there is an appetite to make more data about research management and administration open sustainably.

However, they noted difficulties, including the diversity of systems in use in the sector, the associated complexity of research management and administration and wider challenges such as financial constraints; all represent barriers to change.

## Data sources and platforms

We identified a longlist of 118 datasets with some accessibility and categorised the characteristics of 76 of them in terms of the of openness of the data sources.

The presentation of these data sources illustrates the complexity and fragmented nature of the landscape of open (or potentially open) data about research management and administration.

Our report shows how both open government and open science drivers can intersect as factors that contribute to data openness at different parts of the research lifecycle, and where gaps remain.

## Opportunities for efficient and innovative approaches to open data about research management

There are opportunities to extend open data about research management and administration to reduce cost and bureaucracy, enable better insight and analysis and improve the potential for innovation. Developing a system-wide approach could remove duplication and complexity. The opportunities include:

### Demonstrating the potential and the data in scope

- Build a narrative in favour of open data about research management and administration and widen the focus to engage commercial, public sector and other research organisations and the four nations of the UK, and move almost open data sources to fully open (confirm the potential)
- Audit data that is potentially in scope, assess the potential for open data approaches and map current access and licensing routes (understand extent)
- Develop a baseline of research management and administration data sources with potential for availability as fully open data (confirm the outputs)

### Digital and technology enablement

- Assess the resources needed to sustain selected data service platforms, including technical input and maintenance, data management, aggregation or augmentation of data sources, troubleshooting and enquiry management (understand cost)
- Make data sources which are usually only made available in combined form available as open data (via publication schemes or information asset registers (IARs) for sustainable access and to support efficiency and innovation, reducing the need to develop and maintain several separate often complex and resource intensive data service platforms across the reseatch landscape which don't interoperate (address bureaucracy)
- Develop a route for gradual open licensing of publicly available data that is report/platform-mediated and establish low-burden, common standards for metadata and APIs so a range of APIs can coexist interoperably with open data (unlock interoperability)

### Optimisation through existing open data policy and infrastructure

- Engage with relevant UK Government departments to understand the potential for alignment with established policy routes that support open public sector data, for example **data.gov.uk** (build on established routes)
- Expand HEIs' and public sector organisations' use of publication schemes so key actors can share the data and information they generate efficiently and sustainably, using Information asset registers (IARs) to support common areas for focus across organisations (use what works)
- Review potential for more open data that could be characterised as official statistics (assess what needs to be added to the official record)

## Acknowledgements

We have listed the focus group and interview participants in Appendix A, and we thank all those who took time to engage with the project.

# 1. Background

**This report provides an overview of open data about research management and administration in the UK, and looks at how open and accessible it is**

We have focused on activity and policy developments by the higher education and research sectors as well as the UK Government and international actors focused on open data, aiming to stimulate an appetite for more sustainable open data pipelines.

We wanted to find out how complex and resource-intensive the routes to open data about research management and administration are, and how they can be improved.

Our methodology included interviews, focus groups and desk research.

## Introduction

We identified opportunities to capitalise on the potential of open data about research management and administration to reduce bureaucratic burdens and costs, and improve insight, analysis and innovation, by exploring better support for its management, access and utility.

We wanted to find areas for developing aligned, stable and sustainable approaches to open data provision across the research landscape, and so support cost efficiency, reduced bureaucracy and enhanced innovation and insight.

So we commissioned **Research Consulting** to work with **SIRIS Academic** on a landscape review of open data that is produced in the process of research delivery, management and administration and of platforms that aggregate administrative data about research.

This work is timely and we intend to capitalise on various data-focused initiatives that impact research management and administration data.

These include the UK Government response to **The nurse review of the research, development and innovation organisational (RDI) landscape** to take an evidence-based, data-driven approach to evolving the RDI landscape, including establishing a new, **£10 million Metascience Unit** and its response to the independent **review of research bureaucracy** (Tickell Review). This has tasked Jisc with bringing actors together from the higher education and research sector to strengthen the integration of digital infrastructure, data and information across the system.

A range of additional initiatives and policy developments support a timely focus on open data about research management and administration:

- The UK Government plans to implement a **National Data Library** intended to "bring together existing research programmes and help deliver data-driven public services, whilst maintaining strong safeguards and ensuring all of the public benefit"
- The **UK Metascience Unit**, which will design and run experiments testing the effectiveness of research funding processes, oversee metascience grants programmes and disseminate metascience findings and insights. It aims to help shape the UK Government's research and development strategy
- The associated potential for AI applications to support the ethical and effective use of data about research management in support of the ambitions of the **AI Safety Institute**
- The **UNESCO Recommendations on the Ethics of Artificial Intelligence** noted that "member states should promote open data. In this regard, member states should consider reviewing their policies and regulatory frameworks, including on access to information and open government to reflect AI-specific requirements and promoting mechanisms, such as open repositories for publicly funded or publicly held data and source code and data trusts, to support the safe, fair, legal and ethical sharing of data, among others"[1]
- **The Barcelona Declaration on Open Research Information** (2024) emphasises the importance of research organisations, funders and those assessing research adopting an approach of openness by default, with a focus on the underpinning infrastructure, services, systems and collective action to enable openness of research information
- The potential for expansion of the field of data about research management and administration further into the research landscape to cover, for example digital research infrastructure for effective planning, management and sustainability: **Mapping 'federation journeys' for optimising the UK digital research infrastructure**.

## What do we mean by research management and administrative data?

Research management and administrative data are produced as a result of the management tasks and operations conducted by HEIs and research organisations, commercial suppliers, funders and public and charitable sector organisations in the UK and internationally. They include information that is systematically collected and maintained for research management and administration.

## Open research and open government initiatives offer a range of policy approaches for considering open data about research management and administration

In academia, the **open science** (or the more inclusive "open research") agenda seeks to make research more accessible, inclusive and equitable. It has focused on the openness of outputs from the research process, such as peer-reviewed journal articles, books, research datasets and software.

1 **UNESCO Recommendation on the Ethics of Artificial Intelligence**, p30, para 75

The **FAIR Principles** provide a framework to improve the Findability, Accessibility, Interoperability and Reuse of digital assets including:

- Machine-readable metadata to enable the automatic discovery of datasets and services (supporting findability)
- Ongoing access to metadata (even if the data are not available) and standardised protocols for accessing the data (accessibility)
- Shared vocabularies and languages to facilitate interoperability
- Quality, provenance, community standards and licensing to optimise reuse of the data

Additional initiatives offer an underpinning policy focus for developing open data about research management and administration, focusing on low burdens, better access to insight and trust and a strong research culture, including:

- Promoting **reproducibility and research integrity**
- Fostering a healthy **research and innovation culture**
- Furthering the **trusted research** agenda

## Open data has a relatively long history in the UK Government and the wider public sector

The UK Government and public sector have developed various policy and legislative initiatives focused on supporting transparency and accountability in public administration. These are five examples:

### The Open Government Licence and emergence of open data portals

Building on 20 years of effective open data initiatives in government and the public sector – including the establishment of the Open Knowledge Foundation in 2004, the introduction of the **Open Government Licence** in 2010 for government publications and data and the establishment of the **London Datastore** in (add), **data.gov.uk** open data portal in 2020 and the **Open Data Institute** in 2012 – the UK Government was a signatory to the international **Open Data Charter** and Principles in 2015.

### The right to reuse public sector information

**The Re-use of Public Sector Information Regulations 2015** (RPSI) cover the accessibility and reuse of any information created by central or local government, or any other public body[2]. The regulations exclude, however, "educational and research establishments including organisations established for the transfer of research results, schools and universities (except university libraries)". Accessing information within the higher education and research public sector context is covered by the **Freedom of Information Act (FoIA) 2000**.

2 Public sector information in this context means information produced as part of a public task, which in turn refers to a public body's core role and functions as defined in legislation or established through custom and practice.

## Freedom of Information Act

The FoIA establishes a right of access to all types of recorded information held by public authorities, including HEIs and funders like UKRI and others that count as public bodies. The FoIA does not apply to funders that are registered charities, however in the interests of open research, charitable funders often adopt open practices concerning corporate information, aligning with their leadership roles in research culture, research integrity and open science.

The FoIA requires every public authority to have a publication scheme to promote openness and accountability. The Information Commissioner's Office (ICO), which regulates the FoIA, offers a **definition document for universities and other HEIs**, essentially a model publication scheme to adapt and publish. It recommends that datasets are made available for reuse under the Open Government Licence in line with the **2018 Freedom of Information Code of Practice**. Some data about research are covered in the model publication scheme[3]. More data is often made available by higher education institutions outside of the publication scheme on institutional websites or via other platforms and services that aggregate data. Publication schemes are also widely deployed to reduce the volume of requests under FoIA legislation by making a range of not personally, commercially or otherwise sensitive or disclosive data available to answer commonly asked questions submitted under the legislation.

## IARs

IARs are lists of personal and non-personal information assets held by a public sector organisation. It is important to know and fully understand what information is held in order to protect it and to exploit its potential. IARs mainly operate as internal-only registers to ensure good records management and information security but they may indicate potential data about research management and administration that could be anonymised, aggregated, standardised and made available as open data.

## The Digital Economy Act 2017

With the aim of extending access to administrative government data for research, the **Digital Economy Act 2017** states that deidentified data held by a public authority in connection with the authority's functions may be disclosed to another person for the purpose of research, providing that there is no disclosure of information about a particular person's identity (or information that could allow a particular person's identity to be deduced).

Policy and legislation therefore support a focus on open data about research management and administration, and provide established pathways for easy, routine, sustainable access. While much data is already accessible to varying degrees, the goal now is to explore sector-wide, scalable approaches that optimise open data with minimal burden.

Approaches to open data from both a government and a research context have relevance to this project, see table 1.

3 The scheme requires institutions to publish: "high level information about research funding from public sector sources and research funding from commercial sources where appropriate" together with their research policy and strategy and information on publicly funded research outputs and data.

**Table 1**: Key commitments on open data with relevance to data about research management and administration

| Year | Context | Publication |
|---|---|---|
| 2004 | UK; legislation; open government | Freedom of Information Act (FoIA) |
| 2010 | UK; open government | Introduction of the Open Government Licence for government publications and data |
| 2010 | UK; open government | Creation of data.gov.uk open data portal |
| 2015 | International; open government | Open Data Charter: developed by governments, civil society and experts as "aspirational norms for how to publish data". The six principles are:<br>1. Open by default: a fundamental shift in how government behaves regarding information<br>2. Timely and comprehensive: ensuring relevance and quick and complete publication<br>3. Accessible and usable: machine readable, easy to find data, and providing files in the right formats<br>4. Comparable and interoperable: "the more quality datasets you have access to, and the easier it is for them to talk to each other, the more potential value you can get from them". Use agreed standards to facilitate this<br>5. For improved governance and citizen engagement: transparency improves public services and accountability<br>6. For inclusive development and innovation: economic benefits of access to open data, in efficiencies or innovative problem solving |
| 2015 | UK; legislation; open government | Re-use of Public Sector Information (RPSI) |
| 2016 | International; open research | FAIR Guiding Principles for scientific data management and stewardship |
| 2016 | International; open research | Initiative for Open Citations (I4OC) |
| 2017 | UK; legislation; open government | Digital Economy Act |
| 2018 | UK; open government | Freedom of Information Code of Practice |

| Year | Context | Publication |
|---|---|---|
| 2022 | International; AI | UNESCO Recommendation on the Ethics of Artificial Intelligence: Proposes values and principles for use of AI.<br>Values:<br>• Respect, protection and promotion of human rights and fundamental freedoms and human dignity<br>• Living in peaceful, just and interconnected societies<br>• Ensuring diversity and inclusiveness<br>• Environment and ecosystem flourishing<br>Principles:<br>• Proportionality and do no harm<br>• Safety and security<br>• Right to privacy and data protection<br>• Multi-stakeholder and adaptive governance and collaboration<br>• Responsibility and accountability<br>• Transparency and explainability<br>• Human oversight and determination<br>• Sustainability<br>• Awareness and literacy<br>• Fairness and non-discrimination |
| 2024 | International; open research | Barcelona Declaration on Open Research Information: Commits signatory organisations "that carry out, fund and evaluate research" to:<br>1. Make openness the default for the research information we use and produce<br>2. Work with services and systems that support and enable open research information<br>3. Support the sustainability of infrastructures for open research information<br>4. Support collective action to accelerate the transition to openness of research information |

## Defining open data relating to research management and administration

Building on the **Open Data Institute**'s definition "open data is data that anyone can access, use or share", open data are therefore broadly understood as data that are:

- Openly accessible
- Exploitable
- Editable
- Shared by anyone for any purpose
- **Licensed under an open license**[1]

Data – including data about research management and administration, government and public sector data – may exhibit a range of characteristics and degrees of openness.

The **Open Data Handbook** cites Tim Berners-Lee when describing "the five stars of open data". These "five stars" include:

1. Data that is available on the web under an open licence for use by anyone, sometimes accompanied by a non-commercial use restriction and a requirement to **ShareAlike** and credit the data creator
2. Structured data that is organised and formatted to make it easily readable and understandable by both humans and machines, for example, human-readable language,
3. Data that use a non-proprietary file format – ie a proprietary software package is not needed to analyse the data, for example comma-separated values (csv) format, which stores tabular data in plain text
4. Data that use Uniform Resource Identifiers (URIs) as identifiers; and data that use **open standards from W3C** and the URI, a unique ID for linkages
5. Data that provide links to other data sources (linked data)

Licences are available for open data. As we've noted, the National Archives launched the **Open Government Licence (OGL)** in 2010 to support reuse of government and other public sector information. Many publicly funded producers of data about research management and administration may make use of the OGL for their open data.

For the purposes of this project the OGL is the main open licence type in scope for assessing data sources where they are public sector information.

In cases where all five of these characteristics are present, the data is described as "fully open". Where only some are present we describe the data as "partially open".

In cases where data are made available by non-public sector organisations – for example charities or commercial bodies – under a different open data licence than the OGL but still meeting the five stars criteria, we note them as fully open.

Given that our focus is on data about research management and administration we also consider the FAIR criteria, which support the use of additional open data quality domains. They may be helpful in assessing:

- Digital object identifiers (DOIs) as the persistent identifier of the data source
- Open, standards-based interoperable metadata
- Standards-based open APIs

The data sources may therefore be categorised as:

**Table 2**: Data types and descriptions

| Data types | Descriptions |
|---|---|
| **Fully open data** | Data meeting the "five stars" of open data (also data made available with open, standards-based interoperable metadata, standards-based APIs and in some cases DOIs) |
| **Partially open data** | Data exhibiting a range but not all of the "five stars" of open data and/or FAIR criteria |
| **Publicly available data** | Data that is accessible but embedded in individual webpages or within proprietary formats and platforms, and which are often formed from data from a range of sources, or within a specific proprietary system (eg data presented interactively as a dashboard) |
| **Report/platform-mediated data** | Data held as tables in reports, or search-mediated data in platforms that don't offer a full download of the data associated with an open data licence |
| **Managed access data** | Data requiring subscription or registration for access but that are not exempt, excepted or otherwise sensitive |
| **Internal access-only data** | Data only available within an organisation but not exempt, excepted or otherwise sensitive |

## Confirming the scope of open data for research management and administration

The scope of this work includes open data about research from across the research lifecycle, from hypothesis definition to research impact, including all aspects of the research management and administration landscape. It excludes research data i.e. data produced or collected in the course of a research project.

Data in scope are therefore not personally, commercially or otherwise sensitive, which may include security-sensitive data or inappropriate content. The data may comprise aggregated, anonymised or linked data from a range of sources. Data that are restricted or held in secure settings or data otherwise exempt or excepted for the purposes of the FoIA (includes exemptions relating to the **General Data Protection Act 2018** or **Environmental Information Regulations 2004** [EIR]) are not in scope.

## A challenging sector context and the opportunity for collective efficiencies and simplified approaches to research management

This report is a first step in exploring shared efficiencies and simplified approaches to research management.

The next steps will be identified by the UK Government's Bureaucracy Review Reform and Implementation Network (BRRIN) data and digital subgroup, which was commissioned by the Department for Science, Innovation & Technology to implement the digital recommendations in the **independent Review of research bureaucracy**. BRINN is coordinated by Jisc and includes representation from all the UK's devolved nations.

The subgroup will consider the potential for developing approaches to the operation of the research management ecosystem that offer more efficient, sustainable, low-bureaucracy routes to more accessible data for better analysis and innovation.

The project describes the landscape of open data about research management and administration.

We have provided a broad landscape review and explored the balance between open and closed data within the UK higher education sector. We haven't commented on the collection methods or approaches, frequency of collection, quality (except around systemic areas such as meeting open criteria, metadata or API provision) or value of the data in scope.

The project is a preliminary review to provide a foundation for future focus. It does not propose infrastructure solutions or specific mechanisms to support interoperability or reuse, but it does describe challenges in the current landscape and identify potential for improvement.

We wanted to identify opportunities for the sector to consider, focused on optimising management and use of open data about research management and administration as a sector asset.

The approach and scope set out an initial step supporting further enquiry towards a more comprehensive and programmatic focus on developing the opportunities and specifying actions ahead of moving to any implementation plans.

The landscape of open data about research management and administration is beginning to emerge and this work offers a current snapshot of it.

## 1.1 Methodology

Our preliminary analysis of the open data includes the systems, platforms, services and projects that provide it, the frequency and methodology of data creation, the provenance and type of data, how sustainable the provision is and the means of aggregation.

The overarching objectives are:

- To describe the landscape of open data about research management and administration in the UK
- To identify efficiencies
- To identify opportunities to improve the landscape

### Methodology and approach: focus groups and interviews

Our review took three primary methodological approaches. The first two involved qualitative stakeholder engagement:

- Three focus groups with a total of seven stakeholders from **Jisc** and **UKRI**
- Seven individual interviews with stakeholders from across the research landscape including representatives of **ARMA**, Jisc, UKRI, **CRAC**, **Wellcome** and multiple higher education institutions

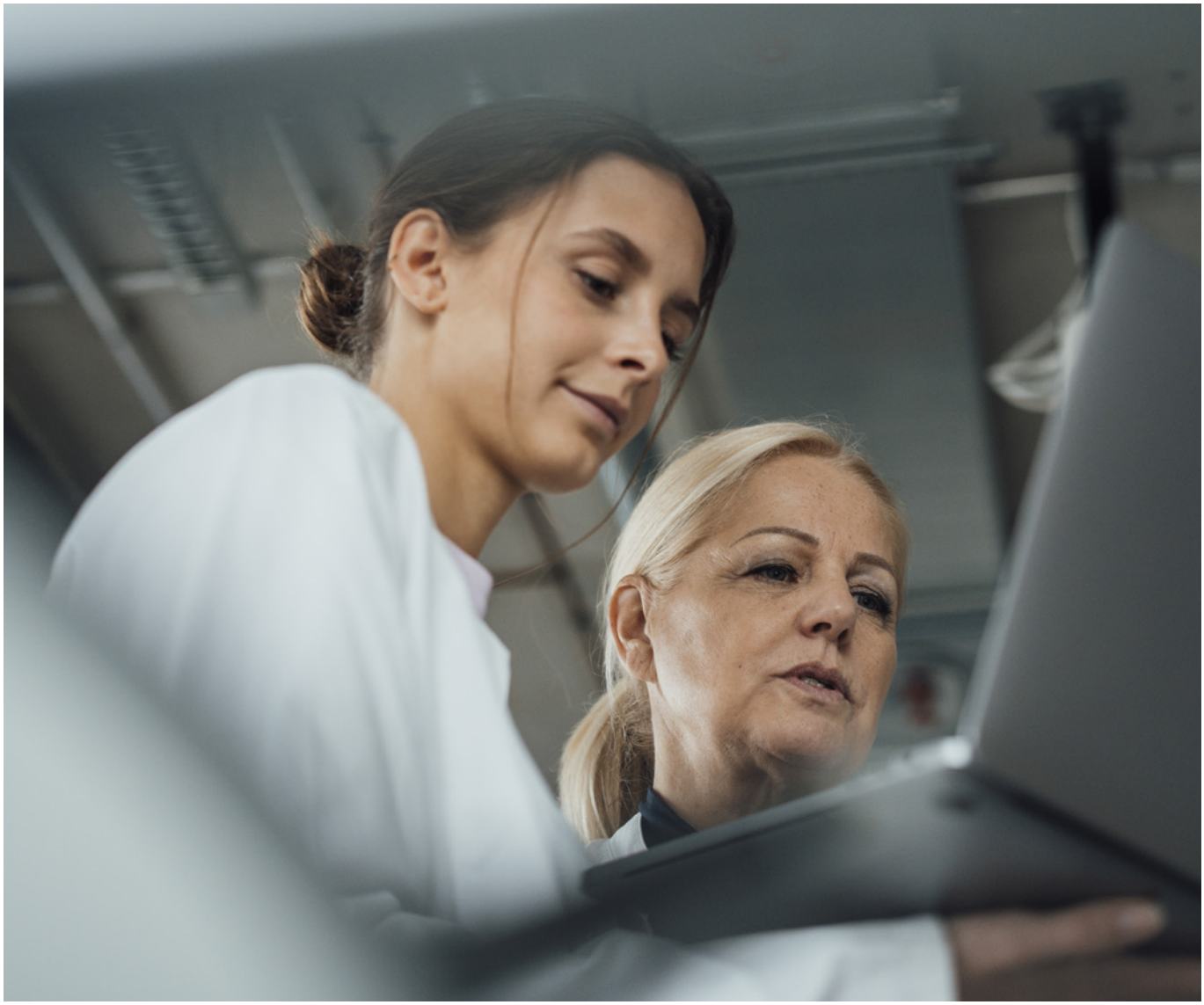
Focus group and interview participants are listed in Appendix A. We have included their anonymised quotations throughout the report to illustrate specific points and insights.

### Methodology and approach: desk research and review of example data sources

We also undertook desk research on open or potentially open data sources in the research landscape, building on previous work by Research Consulting and SIRIS Academic, and on the Jisc team's sector knowledge. The interview and focus group conversations contributed more potentially relevant datasets, allowing us to create a longlist, see table 3 below.

**Table 3**: Longlisted data sources

2021 all REF results data
2021 REF comparative data
2021 REF Environment submission downloads
2021 REF Environment submission supporting data
2021 REF Impact submission
2021 REF Impact submission - tags
2021 REF Output submission
360Giving
Academia.edu
Altmetrics
ARI: Areas of Research Interest
arXiv
Association of the British Pharmaceutical Industry (ABPI)
Athena SWAN
BASE
Be Part of Research
Classification of the Functions of Government (COFOG)
ClinicalTrials.gov
Common Aggregation Hierarchy (CAH)
Community Research and Development Information Service (CORDIS)
Companies House
Contracts Finder
CORE
CRIS (Internal)
CrossRef Event Data
Crossref metadata search
Culture, Employment and Development of Academic Researchers Survey (CEDARS)
Datacite
Department for the Economy topics (N. Ireland)
DOAB
DOAJ
DORA signatories list
EDI data from institutional reports
Eduroam
Episciences
ESAC Transformative Agreement Registry
ETER - European Tertiary Education Registry
ETHOS
EUA Public Funding Observatory
Europe PMC
European Patent Office - Open Patent Services
Finance (Internal)
Find a Tender
Gateway to Research
Global Research Identifier Database
Government Grant Information Service (GGIS)
Grants and grant management (Internal)
HDRUK gateway
HE-BCI Data
HECoS (Higher Education Classification of Subjects)
HEFCW Circulars (Wales)
HESA estates management
HESA finance Data
HESA staff data
HESA students data
HR data (internal)
HR Excellence Award Reports
HRCS Health Research Analyses
HRCS Health Research Classification System
InfraPortal
Institutional policy documents eg intellectual property; research strategy; research integrity; Open Access; Open Data; EDI; Consultancy; HR policies
IPO patents journal
IRUS
ISRCTN registry
JACS (Joint Academic Coding System)
Janet Network: customer service data
Janet Network: security data
Janet Network: site list
Je-S (FEC)
Jisc equipment data
Jisc licence subscriptions manager
Jisc Netsight (Janet Network: core infrastructure)
Jobs.ac.uk
JUSP
KEF metrics
Konfer
Lens
Metascience Unit
NIHR awards and success rates
Octopus
OECD
OpenAIRE Explore/Graph
OpenAlex
OpenAthens
OpenCitations Index/Meta
OpenDOAR
Open Policy Finder
ORCiD
Palomera
PGR data (Internal)
PIDINST
PlumX
Postgraduate Research Experience Survey (PRES)
Race Equality Charter
Repository (internal)
Research Council awards and success rates
Research England funding allocations
Researchfish
ResearchGate
ROR
Royal Society list of public and non-profit research organisations
Scottish Funding Council publications (Scotland)
Sector-level research infrastructure reports
Semantic Scholar
Shibboleth
Snowball metrics
Tenders Electronic Daily
The Funding Service
Titles and Publications Data
TRAC sector analysis
UCISA
UK data service
UKCRC Clinical Trial Units
UKPRN codes
UKRI diversity data for funding applicants and awardees
Wellcome grants awarded
Zenodo



Next, we considered each data source in turn, attributing a priority level describing how directly relevant each one is to the process of research delivery, management and administration and assessing its relevance to the UK research landscape specifically.

We described the longlisted data sources using the fields as outlined in table 2 on page 12.

**Table 4**: Overview of datasets by data provider and level of openness

| | UK Government (incl devolved administrations) | UKRI | Other UK funding bodies | HESA | Jisc | Other UK bodies | International (funders and open infrastructure) | Research performing organisations |
|---|---|---|---|---|---|---|---|---|
| **Fully open data** | | | 360Giving<br>NIHR open data | Staff<br>Finance<br>Estates<br>PGR students<br>HE-BCI | | | OpenAlex<br>Research Organization Registry (ROR)<br>DOAB | |
| **Partially open data** | Areas of Research Interest (ARI) Database<br>Government Grant Information Service<br>Contracts Finder<br>Find a Tender | Gateway to Research<br>REF submissions and results<br>Research Council success rates<br>Research England funding allocations<br>KEF (England) | Wellcome grants awarded<br>Health Research Classification System (HRCS)<br>HRCS analysis | Higher Education Classification of Subjects, Common Aggregation Hierarchy | Equipment Data service<br>Open policy finder | Intellectual Property Office: patent journal | ORCID<br>CORDIS<br>Tenders Electronic Daily (TED)<br>Europe PMC<br>Crossref metadata search | Repository (open access publications, theses, open data, metadata) |
| **Publicly available data** | SFC publications<br>Dept for the Economy (NI)<br>TRAC (Office for Students) | UKRI diversity data<br>InfraPortal | | | | Race Equality Charter<br>Ad hoc reports<br>KONFER<br>Jobs.ac.uk | | Researcher profiles |

| | UK Government (incl devolved administrations) | UKRI | Other UK funding bodies | HESA | Jisc | Other UK bodies | International (funders and open infrastructure) | Research performing organisations |
|---|---|---|---|---|---|---|---|---|
| **Report/ platform -mediated data** | HEFCW circulars<br>Ad hoc reports | Ad hoc reports | | | | Athena Swan | | Institutional policies<br>Research integrity statements<br>Annual reports |
| **Managed access data** | | Metascience Unit (with DSIT)<br>Je-S / TFS | | | IRUS<br>JUSP<br>Janet Network<br>Jisc licences | Vitae: CEDARS<br>PRES<br>UCISA<br>CORE | Lens<br>Datacite | Repository (embargoed and sensitive publications and data) |
| **Internal only** | | | | | Janet site list<br>Non-sensitive aggregated data on cyber security<br>Jisc provision to public and private sector research establishments | | Researchfish | CRIS data<br>Finance<br>HR staff data<br>PGRs<br>Grant management data<br>Equipment data |

The fields form columns in a prototype spreadsheet list, which were assessed for potential categorisation of their level of openness, except where fields were not applicable or where relevant information was not available.

In this report, we present high-level findings from the review of datasets and discuss a small selection of datasets that exemplify each of the levels of openness within the proposed typology.

Both the prioritisation of the datasets and the application of the typology of openness involved some subjective interpretation, with potential for further iteration and refinement as part of any future data audit.

**Table 5**: Approach to data categorisation for priority data sources (source: Siris Academic, 2024) as expanded 2024

| Dimension | Attribute |
|---|---|
| **Type of source** | URL to sources |
| | Primary source vs. aggregated data |
| | Sole source or composed from a range and list of sources |
| | Master data management: responsible organisation(s) for the collection, maintenance and updating of the data |
| **Format and collection method** | File format |
| | Collection method |
| | Variable labelling |
| | Code lists and controlled vocabularies |
| **Data coverage and completeness** | Time coverage |
| | Frequency of data update and/or release/count of unit of time/cumulative/point in time |
| | Geographical/institutional coverage |
| | Data granularity |
| **Accessibility, interoperability and integration** | Data access format |
| | Terms of use/licences |
| | API: custom or standard |
| **Data content** | Type of data (microdata, aggregated data...) |
| | Types of entity |
| | Collection methodology (if the source is primary) |
| | How data and metadata is collected (if the source is not primary) |
| | Most relevant exposed metadata |
| | Categorical metadata (which ones and how they are produced) |
| | Known metadata transformation and processing performed by the source |
| | Persistent identifiers (PIDs) used in the metadata |
| **Potential for text mining** | Textual fields |
| | Quality of textual information (if possible to assess) |
| | Languages (English, multilingual, etc) |
| **Interoperability** | API: custom or standard |
| **Potential for metadata enrichment** | Metadata useful for external linking |
| | Other datasets of interest |
| **Management** | Represented in/or route to IAR |
| | Available in publication scheme |
| | Retention and disposal schedule |
| **Additional information** | Known limitations and comments |
| | Additional links |

# 2. The current landscape

**This section describes the current landscape of open data about research management and administration, or data with the potential to be open. We have identified the key actors and considered dataflows between them. Note that some of the categorisation fields we used to analyse relevant datasets involved subjective interpretation.**

> "I think that's the problem with the landscape of current datasets… it feels as if it should be penetrable… it's quite frustrating for the researcher, it's almost like you're looking at a frosted window. You know there's the truth there on the other side."

## 2.1 A complex landscape

The landscape of data about research management and administration is rich and comprehensive, and it is an important resource for analysis and insight into the research and innovation system. Significant amounts of high-quality, well-used, critical data are created as a result of research management and administration, and many are available in some form.

But the landscape is also complex, with a mix of discrete and overlapping data sources covering particular areas of activity. Collectively, the key actors generate data on a range of elements of the research and innovation system including research staff and students, grants and their application and management, equipment and facilities, publications, impact, knowledge exchange, outputs of commercialisation activity and infrastructure, together with international registries and infrastructures (some hosted in the UK).

It is clear that more could be made available as open data and moved to sustainable, standards-based data maintained by the data creator organisation and made available under open licences for reduced cost, burden and complexity. Doing so would enhance innovation opportunities.

## Key actors within the UK research landscape producing data

Key actors within the UK research landscape include funding agencies such as **UKRI** and the **Scottish Funding Council** as well as higher education institutions (HEIs), research commissioners such as the **National Institute for Health and Care Research (NIHR)**, infrastructure investments, the public sector, independent and private research organisations, commercial organisations, member organisations such as **Universities UK (UUK)**, **GuildHE** and **MillionPlus**, professional bodies such as **ARMA** and charity sector bodies including **Wellcome**, **Jisc** and **HESA**.

Flows of resource within this landscape are mirrored by data flows in the system. We present a view of the flow of data from research funders, commissioners and charity sector bodies to HEIs and research organisations and vice versa (in the form of managing and monitoring research funding, research assessment, commercialisation and impact).

Much of the connectivity in the landscape centres on HEIs. Data collection and management at the system level is undertaken via Jisc/HESA which as the **Designated Data Body** collects, assures and disseminates data about higher education (HE) in the UK on behalf of its statutory customers, working with **HE providers** in each of the UK's four nations.

## We identified a longlist (Table 3) of data sources relevant to the research management and administrative landscape

It included data relating to:

- Research funding, finance and awards
- The performance of research (eg lists of research units; lists of clinical trials units)
- Staffing and research students
- Commercialisation
- People, culture, environment – including equality, diversity and inclusion considerations within HEIs
- Research infrastructure and equipment
- Resource use
- Systems use
- Publication of research – publishing platforms, overlay journals and usage monitoring
- Research outcomes, impact and the evaluation of research
- Context for procurement decisions
- Classification schemes and registries

From these we prioritised 76 in-scope resources for fuller examination, description and categorisation against the typology of openness.

## The landscape has discrete clusters of datasets covering particular areas of activity

Our research shows a lot of effort is involved in creating and curating data sources in isolation, and the resulting sources may either duplicate existing ones, overlap with others or simply hinder the data from being accessed or used to its full potential.

Figure 1 on page 22 below presents an initial view of some relevant data sources.

The datasets are shown based on their granularity (y-axis, a continuum ranging from individual-level data to sector level) and their position in relation to the inputs, activities and outputs of research (x-axis).

We have included a broad description of the areas of focus of the inputs, activities and outputs of research covering the research lifecycle from hypothesis definition to outcomes and impact.

## Extensive data is available on the financial inputs into the research base

As recipients of public or charitable funds funding agencies are required to be transparent and accountable about how they use the money. Public sector funders have obligations under Freedom of Information laws outlined in section 1.1 above; while charitable funders are expected to meet the reporting requirements of the Charity Commission (in England and Wales), Office of the Scottish Charity Regulator in Scotland or the Charity Commission for Northern Ireland.

Data on funding allocations and grant awards is available on the web from key funders such as **UKRI**, the **Scottish Funding Council** (SFC), **Medr**, **Department for the Economy** (Northern Ireland) and the **National Institute for Health and Care Research** (NIHR), with many charitable funders such as **Wellcome** also making this information publicly available.

## Financial data focused on research cost recovery are also available

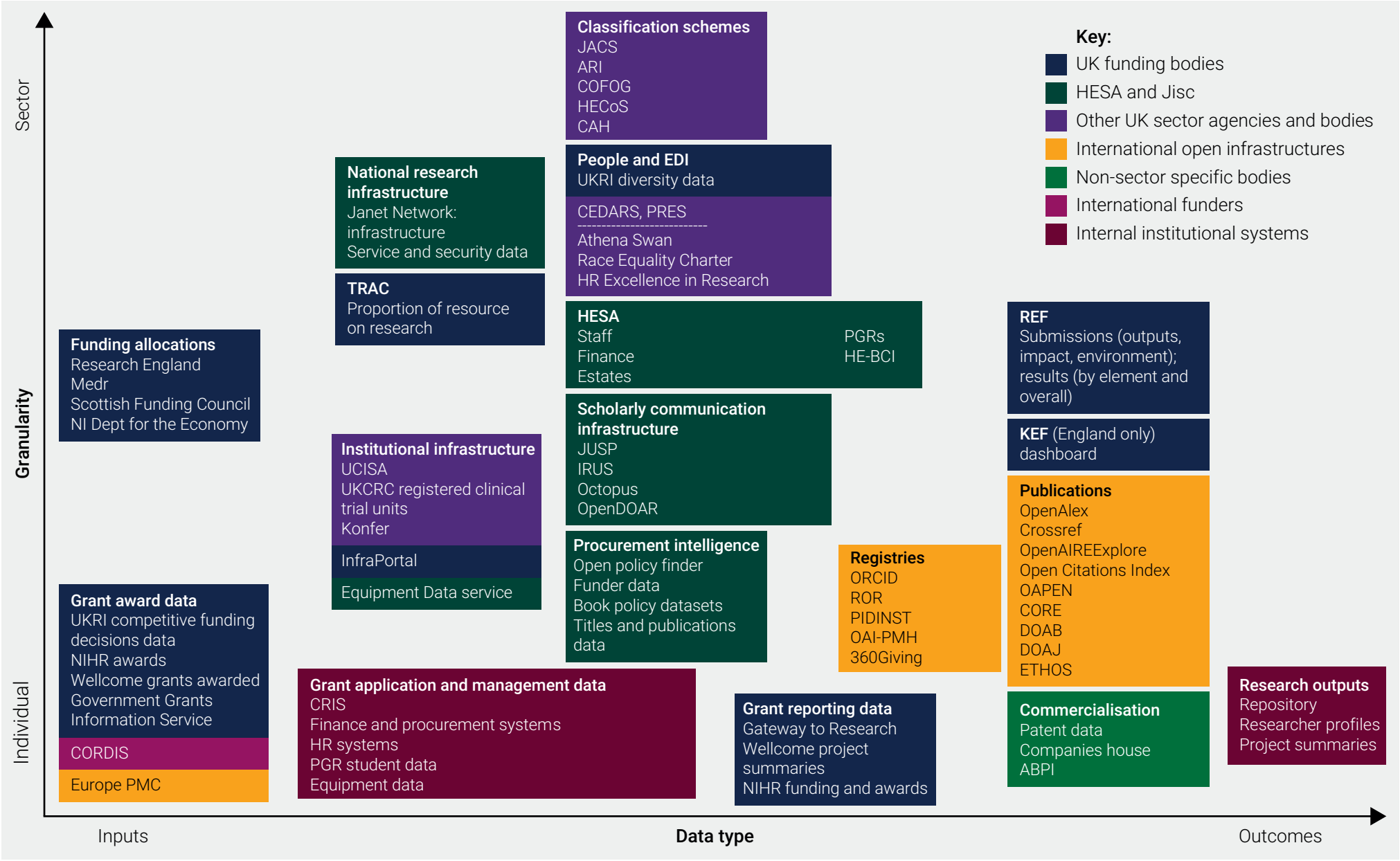
The **Office for Students** (OfS) **Transparent Approach to Costing** (TRAC), introduced in 1999, is an activity-based costing system that associates costs to different activities within an HEI. In 2005, TRAC full economic costing (fEC) for costing individual research projects was introduced to enable HEIs to improve recovery costs of research.

TRAC for **research includes any costs associated with delivering research** (as defined by the **OECD's Frascati Manual**) – fieldwork, project management, conference attendance, supervision of postgraduate research students etc. Costs are analysed across eight research sponsor types including institution own-funded, training and supervision of postgraduate research students, research councils, other government departments, European Union government bodies, UK charities, industry/commercial grants and contracts.

## Several platforms aggregate and publish grants data from multiple funders

Aggregated data on grant funding across the whole of government (including research and innovation funding) is made available annually through the **Government Grants Information System**, while a significant number of public and charitable funders, including the Department for Science, Innovation and Technology, publish open, standardised grants data through **360 Giving**. **Grant finder** holds data on grants from 35 UK and European funders in the biomedical sciences, including both public and charitable bodies, while the European Commission's **CORDIS** platform provides comprehensive information about EU research and development projects.

**Figure 1**: Overview of the landscape of data sources (positioning is indicative and each axis reflects a continuum)



## HESA hosts data about research staffing, resourcing and business interaction

The Higher Education Statistics Agency (HESA) as a Designated Data Body is internationally recognised as a high-quality source of data on educational institutions and students,[4] and its datasets on staff, estates, finances and postgraduate research students provide contextual information about the UK research landscape, while the **HE Business and Community Interaction (HE-BCI)** survey yields more in-depth insights into universities' knowledge exchange activities. Institutions submit aggregated data on an annual basis. The four core areas of data focused on research management and administration are as follows:

- Data on research income and expenditure by source and cost centre
- Institution-level data on collaborative research with industry
- Data on research degree students
- Data on staff with contracts that include responsibility for research

> "Everything in the publicly available data line, if it just had an open licence on it would move up. That's fairly straightforward to do."

## UKRI offers combined data platforms on the recipients and operation of grant funding

Significant development has been undertaken over recent years to integrate data across UKRI's nine constituent research councils within a single data warehouse. It has enabled UKRI to bring together a range of internal data sources into a "conformed data resource" that represents a compilation of separate data sources and is made openly available via the **Gateway to Research** (GtR) portal. GtR is the largest single source of data on UK publicly funded research, spanning projects, publications, people, organisations and outcomes.

UKRI derives this conformed data source with data from **HESA**, the **Office for Students** and a range of proprietary data sources including **Researchfish**, **Dimensions**, **Beauhurst** and **Overton**.

Data accessible through GtR is made available via advanced search functionality and programmatically through an API.

## Data on research infrastructure is coordinated centrally

**InfraPortal: The UK's Research and Innovation Infrastructure Portal** is a UKRI-funded website that contains information on hundreds of research and innovation infrastructures available to UK researchers and innovators. The platform focuses on major equipment, resources such as collections, archives or scientific data, e-infrastructures such as data and computing systems and communication networks. It is populated by the research and innovation community.

## Jisc holds data on foundational digital infrastructure and research communication

At Jisc, we hold extensive data on traffic through the **Janet Network** together with related service and security data. We also hold data on the cyber security landscape, which may be aggregated at a high level of abstraction, and on our provision to public and private sector research establishments. We host the **Equipment Data** service and maintain a number of datasets and services in the field of research communication, including the open policy finder, **Journal Usage Statistics Portal (JUSP)**, **IRUS-UK** and data about licensing activity and usage.

4 See for example Tomczyńska et al (eds.), 2023, **Information technology systems that support science and higher education**, National Information Processing Institute, Warsaw.

These services play a crucial role in aggregating data from institutions and publishers to meet specific use cases, primarily for the benefit of academic libraries. Many of the data are internal but there is scope for a focus on data that can be made open, including highly aggregated data on cyber security threats, network use and licensing.

## Funders coordinate data on the outcomes of research

Data on research outcomes has become more accessible in recent years. The Research Excellence Framework (REF) from 2014 **results** and **impact case studies** and REF 2021 **results** and **impact case studies** generated a number of significant datasets covering outputs, environment, impact, quality profile and comparative data.

The data offer a rich evidence base for further analysis and exploration, such as the **REF outputs analysis: maximising the use of REF data** looking at submitted outputs, and **data enhancement and analysis of the REF 2021 impact case studies,** which give a UK-wide perspective on impact demonstrated in REF 2021 (and previously for 2014), as well as other work reviewing impact in **Scotland** and in **Wales**.

Funder-mandated moves towards open access and open science have fostered the development of a large number of international open infrastructures for research publications. Particularly notable is **OpenAlex**, a bibliographic catalogue of scientific papers, authors and institutions accessible in open access mode.

> "If REF at some point said … we're only going to go for open data sources, that would start to move the conversation."

## Higher education institutions collect extensive data to support a range of sector-wide data requirements

The most granular data on research is often held by HEIs. HR data, finance systems, grant management and current research information systems (CRISs) hold dynamically updated local data that underpin both day-to-day operations and periodic applications, submissions and reporting of aggregated data to external stakeholders. Institutions may also publish some of this data in the form of reports or data on institutional websites. Data about research management and administration are not standardised within HEIs, nor are the systems that collate, generate and report them or coordinate their management. This stifles the potential to publish aggregated, anonymised open data as standard.

## Sector and membership bodies collect a range of mission-focused data

Sector and membership bodies, including **CRAC-Vitae**, **Advance HE**, **ARMA, NCUB**, **UCISA** and the **UK Committee on Research Integrity**, collect and aggregate research management and administration data for a wide range of purposes including supporting research careers, promoting researcher and research professional development, enabling equality, diversity and inclusion, monitoring systems usage and promoting research integrity.

The data, enhanced or complemented by other data through these mechanisms, is sometimes made available to members as proprietary data for their management and processes, though not intended to be open data.

> "[As the Tickell Review said] institutions themselves create their own bureaucracy. We put processes and procedures in place that actually make these things difficult. I think when you've got so many different datasets, different research groups going, 'we need to do it this way for this one and we need to data input this' that's where your time goes."

## Government bodies collect data on research and development

**Estimates of research and development** performed and funded by businesses, higher education, government, UKRI and non-profit organisations are collected by the **Office for National Statistics (ONS)**.

## International NGOs and non-profits

Various international organisations, such as the **OECD, via its Data Explorer**, generate data on topics like international collaboration, government allocations for R&D and infrastructure. Additionally, other international registries and data sources, some of which are hosted in the UK, primarily focus on scholarly communication. For example:

**ORCID**, which stands for Open Researcher and Contributor ID, is a global, not-for-profit organisation sustained by fees from its member organisations. It provides a free, unique PID for individuals to use as they engage in research, scholarship and innovation activities, and the ORCID Registry is open and searchable.

**ROR** is a global, community-led registry of open persistent identifiers for research organisations and includes IDs and metadata for more than 110,000 organisations.

## Commercial providers make some data available openly

Activities of some commercial providers directly intersect with (and make use of) open data sources – for example, developing platforms for serving data collected by funders or using open data for proprietary data platforms such as Digital Science's Dimensions product, which uses data from Gateway to Research. In other cases, commercial systems underpin core research administration activities, including CRISs (Pure, Worktribe, Symplectic), repository management or researcher publications and collaboration intelligence (Scopus, Web of Science).

Commercial administrative systems and tools, such as finance, procurement and HR systems, also underpin internal research management processes.  Academic social networks such as ResearchGate and Academia.edu harvest data on scholarly outputs to build and enrich researchers' profiles.

Key stakeholders in the landscape of open data about research administration and management are also involved in large-scale procurement of commercially provided data solutions, such as the tools to support bibliometric analysis for the 2021 REF.

Commercial providers also make some data they commission or collect available openly, for example the **open data created by Elsevier research and development teams**.

## 2.2 Flows of data within the landscape

### Data follows resource flows in the research administration landscape

Each data source may also have dependencies on (or relationships to) other data sources. The flow of resource from research funders to HEIs and outcomes reporting to sector agencies and funders is mirrored by data collection across the sector.

Figure 2 on page 27 below shows a simplified, indicative view of data flows to HEIs and then on to other organisations for reporting. The diagram distinguishes between frequencies of data flows (periodic or dynamic updating) and mechanisms of data exchange (correspondence, institutional submission, survey, database use or application). It does not depict data flows within HEIs.

Much of the connectivity in the landscape centres on HEIs and research performing organisations. They receive inputs in the form of funding allocations and awards for specific research projects, undertake the full range of activities associated with doing research and then generate outputs, including publications, impact, commercialisation and reports to funders.

Research management and administration data flows within HEIs and research performing organisations are periodic and collected in response to specific reporting, policy or financial requirements.

Core data sources are managed within proprietary systems (for finance, HR, grants and contracts management). The data often include confidential, sensitive or personally identifiable information and operate as part of the IAR of the HEI or organisation.

The data sources require anonymisation, aggregation and extraction to prepare data submissions and reports to other sector stakeholders. Examples include funding and financial flows and those related to staffing, students, research outputs and outcomes, commercialisation, IP and sensitive research.

Reporting requirements can take different forms, including annual reports or returns from institutions to funders, regulatory bodies and sector agencies. They may be collected in response to a specific requirement or on an "as available" basis. In some cases, such as the REF process, longer timeframes are used, covering comprehensive aggregations of activity undertaken over several years and census point in time data.

Annual reports and returns, or submissions-based exercises such as the REF, generally require a high degree of repurposing of data, data cleaning and transformation and fresh data entry, rather than a seamless transfer of data from internal to external systems. Some CRIS providers offer elements of seamless workflows for completing REF (and other) submissions.

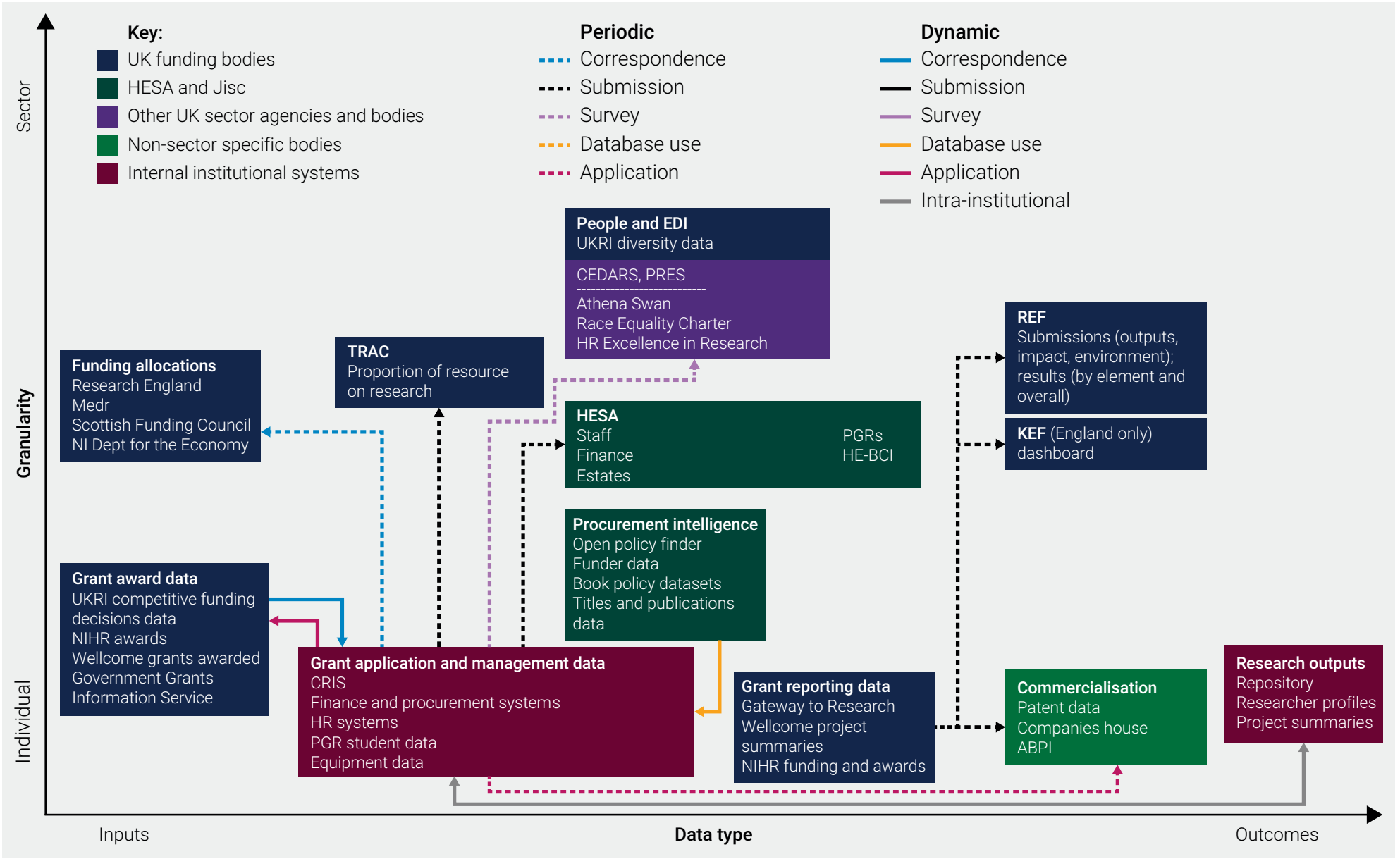
There may be potential in understanding how HEIs and research performing organisations can convene an aligned, low-burden baseline of data about research management and administration that can supply a range of reporting requirements more effectively and reduce burdens in proliferated and fragmented data recreation. Open data approaches have merit for further exploration.

### Platform interchanges

Numerous platforms have been developed to make data available with varying levels of complexity and interoperability; sometimes giving access to a single data resource or a combination of complex and non-standardised ranges of data, sometimes offering search-mediated access and not offering data download or requiring registration to access. Few meet the five stars of open data.

**Figure 2**: Data flows in the landscape.

This is an indicative, simplified view of a selection of relationships between data sources and institutional datasets. The colour of the lines indicates the nature of the relationship (e.g. correspondence, surveys, submissions) and the line style indicates frequency (dashed line = periodic updating and solid line = dynamic updating).



Platforms often serve data that are combined from a range of other data sources, in some cases a mix of proprietary and accessible data, or that rely on user-defined contribution.

Taken collectively, the platform mix and proliferation supporting access to data about research management and administration provided by public sector organisations add to complexity, sustainability, resource and technical overheads, which all need capacity and capital to maintain.

In some cases, a single platform may be developed to give access to a single data source. In others, data, which may otherwise be open, are coded or engineered within a platform that exceeds the level of technology needed to access the data effectively. It could, in some cases, cost more to give access to the data than it did to create the data.

There is potential for cost assessment over the platform landscape to understand the costs that could be taken into account when assessing the potential for a lower burden of flow from creation to access.

> "I think you've got a lot of reports that support understanding of the sector-level view. I'm less aware of regular feeds of data that get updated in the same way that, say, a HESA dataset will be updated. [A lot of things] are sitting in reports as opposed to live datasets – and they're not datasets at all. They're tables in reports."

## 2.3 Implications for researchers, project leads and research management professionals

### Limited interoperability in the landscape contributes to an increased burden on researchers, project leads and research management professionals

Although this project focused on sector stakeholder and institutional perspectives, lack of openness or integration in the data landscape of research management and administration data has significant direct implications for researchers, project leads and research management professionals.

For example, reporting on research outcomes may be more burdensome for researchers and research management professionals if the project lead is treated as the primary source of information, rather than simply reusing information already available in institutional systems.

Conversely, the number and range of reporting avenues adds to the research management burden through complexity and lack of alignment in approach, leading to duplicated, similar or slightly different requirements that could be assimilated.

The time has come for a research management and administration data baseline of sustainable data sources focused on established routes for reuse.

The UK Government's **review of research bureaucracy** notes a shift among funders (including UKRI, NIHR and Wellcome) from older platforms for managing applications and grant information. It identifies the potential opportunity to: "think collectively about how to connect these new systems to enhance the delivery and management of research, support future innovation and growth and reduce the bureaucratic burden on researchers."

As much data about research administration and management flows to, from and within HEIs and research performing organisational systems, these – as well as funder and reporting systems – are key elements in either adding to or potentially reducing burden for researchers and research management professionals.

> "What we'd like to do is push really, really hard on ORCIDs and on linked open repositories… to try to get people to put their stuff in open repositories so that we can access it, rather than us asking people for stuff."

# 3. Assessing levels of openness

**As our findings show, the data sources about research management and administration vary in their level of openness within a complex landscape. We saw various approaches that were often siloed as well as repeated data collection and management approaches**

Using the categorisation developed and presented in table 2 above (and again as table 4 below) we assessed selected data sources and associated platforms for the extent of their openness. Here, we talk about a few examples of data sources exemplifying the six levels of openness.

**Table 4**: Data types and descriptions

| Data types | Descriptions |
|---|---|
| **Fully open** | Data meeting the "five stars" of open data (and also data made available with open, standards-based interoperable metadata, standards-based APIs and in some cases DOIs) |
| **Partially open** | Data exhibiting a range but not all of the "five stars" of open data and/or FAIR criteria |
| **Publicly available** | Data that are accessible but embedded in individual webpages or within proprietary formats and platforms, and often formed from data combined form a range of sources or within a specific proprietary system (eg data presented interactively as a dashboard) |
| **Report/platform-mediated** | Data held as tables in reports, or search-mediated data in platforms that don't offer a full download of the data associated with an open data licence |
| **Managed access** | Data that require a subscription or registration for access but are not exempt, excepted or otherwise sensitive |
| **Internal access only** | Data that are only available internally within an organisation but are not exempt, excepted or otherwise sensitive |

## 3.1 Fully open data

### Fully open data that conform to criteria established by open government initiatives and can also draw on initiatives developed in an open research context

As we described above, in this project's context fully open data conform with criteria established by open government initiatives and can draw on initiatives developed in an open research context. In public sector organisations the open data licence refers to data under an OGL; non-public sector organisations may use another open data licence.

HESA's open data focus on the inputs and outputs of research (funding, publications, evaluation submissions) and cover statistics about the activities of research management and administration including data covering **staff**, **finance**, **estates** and the **Higher Education Business and Community Interaction Survey (HE-BCI)**.

Data about grants from charitable funders aggregated by **360Giving** offers search-mediated data but also open data within **its data registry**.

The data sources most closely associated with open research publications and data offer fully open data access routes, including **OpenAlex**, which describes academic entities and the connections between them, offering to **export results under an open data licence**.

## 3.2 Partially open data

### Significant numbers of datasets are available as partially or near fully open data

Many core data sources for research administration and management are available as partially open datasets. This includes **REF 2021** submissions and results and the API that enables programmatic analysis of the REF submissions, as well as full download and search options, **Research England funding allocations** and contract finding services (such as **Contracts Finder** for government and agency tenders, or **Find a Tender** for high-value public sector contracts).

Data relating to project awards tend to be characterised by partially open data approaches. This route includes some of the data in **Gateway to Research** covering UKRI-funded research grants and **Research England knowledge exchange framework data**. Institutions may also present their research publications, theses and research data in partially open ways.

Following the five stars (plus) of open data, it appears the elements that are less likely to be found in these partially open datasets are URIs linking to other datasets.

Using proprietary formats (including Microsoft XLSX format spreadsheets) is another area where data may not fully meet the five stars criteria. Often, it would be fairly straightforward to move many of these partially open data sources to fully open data very quickly. Doing so could demonstrate the potential and build a narrative in favour of open data about research management and administration as supporting an interoperable, insightful sector.

> "Funder data: making it clear that it can be openly licensed so we can ingest it easily, third party, vendors can ingest it into their systems without worrying, assuming it's got a CC BY [licence] on it. I'd say that's where I'd start because that's where the biggest value would come from."

## 3.3 Publicly available data

### Data may also be publicly available in other formats

Data may be embedded in individual webpages or within proprietary formats and platforms. These data are publicly available but they are more challenging to access than openly available data, and this limits opportunities for reuse or linking.

They include some data sources relating to inputs into research, such as **Scottish Funding Council (SFC) funding allocations**, as well as data relating to resource use and activities, for example **sector-level TRAC data**. In some cases proprietary formats like Excel files are used; in others the data may be embedded in interactive proprietary platforms, such as a Tableau dashboard.

Data made publicly available on HEI websites via publication schemes may also include data about research management and administration.

> "There's a general tension between the need for collective action or collaborative efficiencies, and the fact of competition with each other to a certain extent… there isn't necessarily the desire to be completely open."

## 3.4 Report/platform-mediated data

### Report/platform-mediated data include tables held in reports and search-mediated data in platforms that don't offer a full download of the data associated with an open data licence

These data sources may be held as tables or appendices embedded within a report narrative, often in pdf format, or mediated by search functionality that doesn't meet the five stars standard of open data in terms of offering structured data under an open licence.

In some cases, they may be ad hoc reports or policy documents, or they may form part of a series of publications that could potentially be made available in formats better suited to sharing or reuse.

Institutional policy documents are often published in this way. These include those made available through Freedom of Information publication schemes – such as those relating to intellectual property, research strategy, research integrity, open access, EDI and other topics central to the management and administration of research. Some regular institutional accountability reports, like annual statements on research integrity, follow structured templates shaped by sector stakeholders' requirements. These stakeholders could perhaps influence moves towards more open formats for these reporting activities.

> "So you have things like a large-scale review, which gives you a snapshot at a particular time, or maybe a snapshot on an annual basis, but those are sitting in reports as opposed to live datasets, or even just tables in reports."

## 3.5 Data available with managed access

### Users of data available with managed access must subscribe or register to access it

Some data sources require subscription or registration even though the datasets may have potential to be made available as open data, either as a whole or in part. Managed access may be down to reasons of security or confidentiality, or it might be to enable more customised/personalised interactions with the data.

Managed access sources include those for surveys regarding people's experience of research environments, such as **Culture, Employment and Development of Academic Researchers Survey (CEDARS)** and the **Postgraduate Research Experience Survey**, as well as the **Universities and Colleges Information Systems Association (UCISA)** systems survey and **CORE**, which aggregates publications from institutional repositories and journals. These sources provide internal insights about institutional activities and performance and offer some opportunity for comparisons with other institutions.

This data category may also include infrastructures underpinning research, such as the **Janet Network**, aggregated and non-sensitive data about the cyber security landscape, **IRUS, Jisc's institutional repository usage statistics service** and **JUSP**.

> "[Our dataset] is somewhere between publicly available data without an explicit open licence and managed access requirements. The data are openly available on the website without any access restriction, but in order to use the data via an API there's an API key that needs to be requested."

## 3.6 Internal access only data

### For practical or resource reasons institutions may adopt a closed approach to data from internal systems

Finally, organisations, institutions and sector bodies have data sources available to them internally, particularly relating to the practical operational functions of research administration and management.

Confidentiality and data security are key considerations, but anonymised and aggregated data from these internal sources may also subsequently be surfaced in reports or returns.

Core internal research management and administration data sources therefore don't yet routinely appear as open data. In many cases, lack of resource may limit options to consider making more available as open data. However, doing so would add more insight into the landscape of open data about research management and administration.

> "There will always be a different framing around the different use cases data is required for, but that shouldn't be a barrier to us having more consistent ways of capturing the data so that it can be repurposed in that different framing in different ways."

# 4. Improving the landscape

## About this section

**Opportunities to improve the landscape of open data about research administration and management include encouraging wider sharing, making publicly available or report-locked data open, facilitating access to disaggregated datasets, and optimising API provision. This section explores those opportunities and we have also provided relevant case studies of international data initiatives.**

> "It's about the way in which all this data is commonly presented. It should be possible for the software to pull it through, so we don't lose institutional independence – they can access the data that is presented for them and then we can access it back. I would say that'd be a more pragmatic way of thinking about it."

## 4.1 A significant opportunity

### The open data culture offers a significant opportunity within UK research management and administration

The current landscape of access to data about research management and administration is a product of the size, diversity and decentralised nature of higher education in the UK, together with the lack of a policy imperative in favour of open administrative data for research.

HEIs and sector bodies are not actively encouraged to make data about research management and administration open; data that are made available for regulatory and compliance purposes are not always easily discoverable or reusable. Furthermore, funding agencies and research policymakers tend not to hold their own administrative processes (and commissioned consultancy) to the same standard as those applied to academic researchers when it comes to the openness, discoverability, reusability and preservation of their work.

The volume of open data about research management and administration falls some way short of what other public sector organisations in the UK have achieved and what UK stakeholders require, as well as what other European countries have achieved, at least in part (outlined in Appendix B).

There is a clear and important opportunity to optimise open data about research management in the UK, and to simplify and stabilise the data as open data with the benefit of common standards for metadata and APIs.

### There is appetite to go further

Our interviews indicated that key stakeholders and institutions are interested in making more data more openly available. But there are also tensions in terms of the diversity of the landscape within the sector and the wider challenges (including financial constraints) facing HEIs.

There is potential to build a narrative in favour of wider open data relating to research management and administration, drawing on both open government and open science approaches and noting the potential intersection of both at points in the research lifecycle.

> "I think there's a desire to make more of our data open… we've probably traditionally taken quite a sort of risk-averse approach to sharing data, maybe opting to keep it closed – it's something that we want to look at really in terms of opening up and sharing data and information."

## 4.2 Encouraging wider sharing

### Encouraging a general presumption towards open data

To date, data about research management and administration have fallen between the two stools of open government and open science. Both of these perspectives on open data start with a core principle of "open by default", which is not yet apparent in approaches to data about the management and administration of research, given the overall focus on open access in the research sector.

With the renewed focus in the sector on open research information, exemplified in the **Barcelona Declaration on Open Research Information**, there are opportunities to encourage wider sharing of institutional data sources where these can be made available in a transparent and interoperable way. Funders and sector bodies can promote the efficiency potential by developing more examples of structured, reusable formats and by exploring ways to use institutional data to reduce the burden on researchers, deliver efficiencies or streamline resource and data flows.

> "It would transform things if universities would universally share some data from their systems."

## 4.3 Reducing complexity

### Moving towards open data

Data that are currently partially open – locked in reports, or in other ways publicly accessible but not fully open – can become fully open data when access, curation, licensing and linking are improved.

This suggests simplifying the presentation in some cases (for example, providing open data instead of – or in addition to – interactive dashboards). Where institutional accountability returns are already provided in highly structured templates, requiring these in a non-proprietary format that is more suited to reuse, linking and interoperability may shift them towards partially or fully open data. Cost-effective open data could offer significant efficiencies in planning for (and being available to) system-wide models for insight and analysis.

As we noted earlier, our project doesn't explore the data lifecycle of research management and administration or comment on the relative merits of the platforms and access points that make these data available. But we could speculate that it can cost far more to serve the data than to create them, given the number of highly engineered and siloed data service platforms involved.

Viewed from a macro perspective the current landscape could obscure the data within a complex research management data infrastructure and limit insight and innovation. Operating more cohesively would benefit all parties.

### Supporting the metascience agenda

In line with the growing interest in metascience there is scope to encourage good practice in relation to outputs from evaluation studies and similar projects. This would involve making underlying data sources openly available and requiring their deposit in recognised repositories, which may help to open up opportunities for reuse.

> "If we're saying that these evaluations are really research on research, we should be applying the same standards that we apply to our grant funded research…ask people to have a data management plan and try and make as much of the research data open as possible. Yet when it comes to our own evaluations, we quickly move on as soon as the report is published."

## 4.4 Augmenting data

### Transparency and linking can also support augmentation of datasets

Characteristics of open data, such as transparency, linking and conforming to standards, can make it easier to augment datasets.

The underlying reliability, currency and quality of the data also need checking and assurance, and open data approaches can create a sustainable and efficient route to data quality.

> "The more open data, the better, but with the caveats that quality and integration are really important to make it usable. [This means] thinking about it as an ecosystem and how certain services integrate together, whether that's PIDs or APIs or how services can make it easier for data to be accessible."

### Technical solutions including use of URIs and linked data have a role to play

Technical solutions, including use of accurate and robust metadata, URIs and effective approaches to linked data, will be important. URIs and linked data currently appear to be particular gaps for data that are partially but not fully open.

Using identifiers can help to track the journey of data related to research projects and activities through the research lifecycle.

> "The biggest barrier to taking [a past project] any further was the lack of any kind of metadata standard for the data that we were trying to collect."

## 4.5 Optimising metadata and API provision

The number and range of platforms serving data that are open or accessible without access management processes add complex technical management overheads. Metadata varies in quality and there are some excellent examples of metadata standards that could be used more widely; the quality of APIs is also variable.

Common metadata standards  would both enhance discoverability and support an interoperable system of research management and administration analysis, supporting the development of effective and well-managed AI innovation as well through enabling effective APIs.

The UK Government **advocates "API first" design**, where the API is developed before the rest of the service and it is the first interface for the data. This means services can be built around the API.[5]

User and business requirements should be the starting point for designing an API, with appropriate and proportionate design for the data source being used.

> "We've got what we have on the website [but] actually we would like to have an API that people could just pull up our data from."

## 4.6 Opportunities for improvement

There are opportunities to extend open data about research management and administration to reduce cost and bureaucracy, enable better insight and analysis and improve the potential for innovation. Developing a system-wide approach could remove duplication and complexity. The opportunities include:

### Demonstrating the potential and the data in scope

- Build a narrative in favour of open data about research management and administration and widen the focus to engage commercial, public sector and other research organisations and the four nations of the UK, and move almost open data sources to fully open (confirm the potential)
- Audit data that is potentially in scope, assess the potential for open data approaches and map current access and licensing routes (understand extent)
- Develop a baseline of research management and administration data sources with potential for availability as fully open data (confirm the outputs)

5 Central Digital and Data Office (2022), **API technical and data standards**.

## Digital and technology enablement

- Assess the resources needed to sustain selected data service platforms, including technical input and maintenance, data management, aggregation or augmentation of data sources, troubleshooting and enquiry management (understand cost)
- Make data sources which are usually only made available in combined form available as open data (via publication schemes or IARs) for sustainable access and to support efficiency and innovation, reducing the need to develop and maintain several separate often complex and resource intensive data service platforms across the reseatch landscape which don't interoperate (address bureaucracy)
- Develop a route for gradual open licensing of publicly available data that is report/platform-mediated and establish low-burden, common standards for metadata and APIs so a range of APIs can coexist interoperably with open data (unlock interoperability)

## Optimisation through existing open data policy and infrastructure

- Engage with relevant UK Government departments to understand the potential for alignment with established policy routes that support open public sector data, for example **data.gov.uk** (build on established routes)
- Expand HEIs' and public sector organisations' use of publication schemes so key actors can share the data and information they generate efficiently and sustainably, using IARs to support common areas for focus across organisations (use what works)
- Review potential for more open data that could be characterised as official statistics (assess what needs to be added to the official record)

# 5. Conclusions

**Through a mixture of desk research, interviews and focus groups we have identified the characteristics of the landscape of open data about research administration and management**

It is complex and fragmented.

We have found that a greater focus on open data about research management and administration can reduce inefficiencies and burdens for researchers, research managers and project leads.

Our key findings:

## The landscape is complex, with low levels of openness and sustainability

The landscape of research administration data is complex and fragmented, with many discrete clusters of data and silos of information. As a result the interoperability of datasets relating to research administration and management is low.

## The platform landscape adds to the complexity

The various platforms' diversity adds to the complexity and cost. It should be possible to take a systems-level overview to reduce bureaucratic burdens and consider whole system costs so efficiencies can be realised.

## Moving available data to open data is a quick win

More open data about research could be available if the data were given an open licence. In particular, in the case of public sector organisations' data, this simple change could result in far more open data about research management and administration.

## Data about inputs and outputs of research appear to be more openly available than data about internal processes, activities and experiences

Data relevant to the management and administration of research are most open when they relate to the inputs into research (funding allocations, grant awards) and the outputs of research (publications, data arising from research, impact).

The openness of inputs to the research system can be loosely attributed to drivers for open government, while the openness of scholarly outputs and associated metadata is primarily a consequence of the open science movement and associated investments in open scholarly infrastructures.

## Metadata and API standards are available and could be used more

The research management sector has a range of metadata and API standards that could be repurposed for open data.

## Duplicating data creates inefficiencies

Data follows flows of resource in the research administration landscape – but, often, it needs to be duplicated within these flows. For example, between applicants and funders, funders and HEIs and, later, from HEIs to funders and other bodies for reporting or evaluating research.

HEI data about research administration and management is central to data flows, but currently appears not to be available as open data by default. Developing more open HEI data could bring benefits in reducing costs and reducing the reporting burdens on individual researchers, research managers and project leads.

## Lack of URIs, linked data or reliance on proprietary formats limits the potential for data to be fully open

As part of the desk research for this project we compared in-scope datasets with standards suggested for fully open data. It appears that lack of URIs, absence of linked data, or reliance on proprietary formats currently limit the potential for these datasets to be fully open.

The gaps also limit the discoverability and interoperability of data sources.

## Much publicly available and report/platform-mediated data could be open

There is a significant amount of publicly available data relevant to research administration and management with the potential to be open. This may be material locked in reports – for example, where data tables are embedded in a pdf document – embedded in a webpage or in a proprietary format or platform.

## UK (and international) government exemplars have much to offer in terms of existing policy approaches

It would be beneficial to work closely with UK Government departments to develop trusted, effective routes to open data about research management and administration.

# Appendix A: Project contributors

**Many thanks to all those who contributed to this project.**

**Table A1**: Project contributors

| Organisation | Role |
|---|---|
| Jisc | Director of product – research management |
| Jisc | Head of research and innovation – customer development |
| Jisc | Director of networks |
| Jisc | Head of data and intelligence |
| Jisc | Senior business intelligence analyst |
| UKRI | Head of open research |
| UKRI MRC | Chief data officer |
| University of Exeter | University librarian |
| Wellcome Trust | Head of data and digital |
| ARMA, University of Bristol | Director of strategy, information and operations |
| University of Lincoln | PVC research and knowledge exchange |
| Vitae | Head of policy and advocacy |

Jisc
4 Portwall Lane
Bristol
BS1 6NB
0203 697 5860
**help@jisc.ac.uk**
**jisc.ac.uk**

Open data about research management: A landscape review
February 2025