



# Research Data Management and Sharing Policy

<b>Prepared by:</b>	Open Research Action Group (ORAG)
<b>Approved by:</b>	University Senate on the recommendation of the University Research & Knowledge Exchange Committee (RKEC).
<b>Route to approval:</b>	Recommended for onward consideration and approval by: <ul style="list-style-type: none"><li>• Open Research Group (ORG) on 16 December 2022;</li><li>• Research Sub-Committee (RSC) on 17 January 2023; and</li><li>• RKEC on 7 February 2023.</li></ul>
<b>Date of approval:</b>	Approved by University Senate on 29 March 2023.
<b>Date effective:</b>	30 March 2023.
<b>Version number:</b>	2.1
<b>Summary of changes:</b>	<p>The following policies have been updated and merged into a single policy:</p> <ul style="list-style-type: none"><li>• Research Data Policy (v1.0 19 November 2013); and</li><li>• Research Data Deposit Policy (v1.0 25 November 2014).</li></ul> <p>A footnote was added to section 3.1.5 (v2.1 18 January 2024)</p>
<b>Review process:</b>	To be reviewed annually by the ORAG.
<b>Related policies and procedures:</b>	<p><b>Sector/Funder:</b> <a href="#">Concordant on Open Research Data</a> (July 2016) <a href="#">UKRI Open Access Policy</a> (Feb 2023) <a href="#">UKRI Individual Funder Data Policies</a></p> <p><b>University of Strathclyde:</b> <a href="#">Strategic Plan (2020-2025)</a> <a href="#">Research Code of Practice</a> (Oct 2023) <a href="#">Institutional Rights Retention Policy for Research Publications</a> (Jan 2024) <a href="#">Code of Practice for Postgraduate Research Students</a> (Sept 2023) <a href="#">Integrated Workflow for Publications and Datasets</a> (Aug 2021) <a href="#">Code of Practice on Investigations Involving Human Beings</a> (Feb 2017) <a href="#">Intellectual Property and Commercialisation Policy</a> (Mar 2023), which includes a specific <a href="#">IP Policy for Undergraduate Students and Postgraduate Taught Students</a></p>

## **Table of Contents**

1. Introduction .....	1
2. Principles of Research Data Management and Sharing .....	3
3. Policy requirements .....	3
4. Operational requirements .....	5
5. Rights associated with research data .....	11
Appendix A – Glossary of Terms .....	12

# 1. Introduction

The University of Strathclyde (hereafter 'the University') believes that the highest standards of research quality and integrity are facilitated by the application of good practice in research data management and sharing (RDMS). The creation of and provision for data resources (i.e. digital data/files and analogue data) at research institutions is central to maintaining a progressive research environment. Research funders are also setting increasingly ambitious research data management and sharing expectations for Researchers and their institutions. This is in furtherance of their broader open research goals, particularly in promoting superior transparency of research findings (e.g. verification, reproducibility and replicability). Such a trend has therefore become reputational and is reflected in a growing number of scholarly journals and communities of practice, many of which mandate openness and sharing of any data accompanying publications.

Research data processed by University Researchers (as defined in 1.2) is expected to be managed to a high standard throughout the research data lifecycle in support of the [University's Strategic Plan](#); to avoid data loss, contractual and/or data protection breaches; and to widely share the evidence which underpins research, for societal and public benefit. The University supports Researchers to achieve excellence through the provision of systems, infrastructure, tools, training and guidance via our cross-functional [Research Data Management and Sharing \(RDMS\) Service](#).

## 1.1 Purpose of the policy

This policy provides key definitions, principles, policy and operational requirements of research data management and sharing along with an explanation of rights relating to research data. The purpose of this policy is to ensure that Researchers manage their research data effectively, enabling them to:

- Comply with ethical, legal, funder and other requirements in relation to data and data management.
- Preserve eligible data for reuse within and outside of the University (as appropriate).
- Demonstrate the integrity of their research.

This Policy aligns with wider [requirements and policies relating to research data management and sharing](#). This includes: University research policies (e.g. the [Research Code of Practice](#) and the [Code of Practice for Postgraduate Research Students](#)); external research-related concordats and agreements to which the University is committed (e.g. [Concordat to Support Research Integrity](#) and [Concordat on Open Research Data](#)); and funder and publisher requirements (e.g. [UKRI research data policies](#) and [Open Access Policy](#)).

## 1.2 Key definitions

Key definitions for the terms 'Researcher', 'research data' and 'research data management and sharing' are provided here. A full glossary of terms is provided in [Appendix A](#).

### **Researcher**

Throughout the policy, the term 'Researcher' is used to refer to:

- (i) University staff (whether holding employment or worker contracts) involved in the creation, collection, generation or processing of research data.
- (ii) Postgraduate Researchers (PGR) involved in the creation, collection, generation or processing of research data.
- (iii) Individuals who do not fall within (i) and (ii) above but who are otherwise associated with the University and are authorised to use the University's name, facilities and/or services when involved in the creation, collection, generation or processing of research data (such persons would include, for example, those holding visiting, honorary or emeritus status).

Where expectations and requirements for Researchers differ by the role they hold (e.g. Principal Investigator, Postgraduate Researcher (PGR), or Supervisor) this is noted in Section 3. Nonetheless, each Researcher has a responsibility for the research data they are collecting and processing.

This policy does not apply to Undergraduate Students and Postgraduate Taught Students. The Intellectual Property issues associated with involving these students in research projects are addressed in ANNEX B of the [Intellectual Property \(IP\) & Commercialisation policy](#).

### **Research data**

The definition of 'research data' for the purpose of this policy is digital or analogue (i.e. represented in a physical way) information that is created, collected, and/or reused to produce, validate and/or enrich research findings.

'Research data' therefore includes:

- Data collected through any methodology (e.g. by experimentation, observation, modelling, interview or other methods)
- Data obtained from existing sources
- Quantitative data
- Qualitative data
- Raw or primary data (e.g. direct from measurement or collection)
- Data derived from primary data for subsequent analysis or interpretation (e.g. as an extract of a larger dataset)
- Data derived from existing sources where the rights may be held by others.

### **Research data management and sharing**

'Research data management and sharing' refers to the collection, organisation, storage, sharing and preservation of data throughout a research project and forms a key component of ensuring research integrity.

'Research data management and sharing' covers:

- Initial planning, routine processes for data (file) management and organisation (e.g. back-up processes; folder and file-naming systems; version-control strategies; quality control measures; data documentation).
- Consolidation and incorporation of any contractual matters relating to the research (e.g. collaborative research terms stipulated in a collaboration agreement, Material Transfer Agreement or similar).
- Observance of appropriate ethical, legal and professional frameworks, obligations and standards, including conditions of ethical approval.
- Long-term archiving, removal, restriction and/or sharing of data.

## **1.3 Scope of the Policy**

This policy applies to all research data processed by Researchers in the course of their activities at the University. The specific requirements for Researchers will differ according to the nature of the project and whether the project is funded. Research data subject to funder, regulatory and contractual requirements must be managed according to those requirements and any exceptions should be approved by the sponsor, funder or relevant authority.

This policy outlines expectations on Researchers regarding best practice in research data management and sharing. Aspects of the policy which are mandatory are differentiated using the word 'must'. The guidance associated with this policy details best practice for data storage, sharing and the preservation of data and provides examples of what is appropriate, proportionate and realistic.

## 2. Principles of research data management and sharing

In recognition that not all research data can be shared openly, the University's approach to assessing the suitability of data for sharing is that data should be 'as open as possible and as closed as necessary'. This means that data should be 'open' in order to foster reusability and to accelerate research, but 'closed' when required by legal, ethical, security and/or commercial constraints or if considered necessary to safeguard the safety and wellbeing of those involved in the research. In line with the University's strategic commitment to progress towards FAIR open data policies, whenever possible data and/or descriptive metadata produced by the University should be:

- **Findable:** easy to find for both humans and computers.
- **Accessible:** accompanied by clear instructions for access and authorisation.
- **Interoperable:** can be integrated with other data and/or work in conjunction with applications or workflows for analysis, storage and processing.
- **Reusable:** suitably described to allow further use, released with a clear and accessible data usage licence.

These [FAIR data principles](#) are regarded as a key element of good research data management and sharing practice by offering both Researchers and research institutions a standardised framework for the creation and stewardship of research data.

## 3. Policy requirements

When conducting research, Researchers are responsible for managing the data they create, collect and process. In general, overall responsibility for data management rests with the Principal Investigator of a project or study. Similarly, Postgraduate Researchers are responsible for managing the data they create, collect and process, with support and guidance to be provided by Supervisors. However, the Primary Supervisor has overall responsibility for ensuring that all research is organised and undertaken within a framework of best practice, as stated in the [Code of Practice for Postgraduate Research Students](#).

All Researchers can access support in relation to compliance with external and University research data management and sharing requirements, including the requirements in this Policy, from the [Research Data Management and Sharing \(RDMS\) service](#). Support for ethical matters can be sought from the [University Ethics Committee](#); support for contractual matters can be sought from [Grants and Contracts Team in RKES](#).

### 3.1 Researchers should ensure research data is managed responsibly and must adhere to ethical, legal, funder and other mandated requirements.

- 3.1.1 All Researchers are expected to familiarise themselves with University research data management and sharing requirements and solutions, as well as disciplinary tools and infrastructure, including all costs, terms, and conditions governing their research. This includes being aware of and adhering to all relevant legislation and policies concerning data protection, copyright and Intellectual Property (IP) as outlined in Section 4.
- 3.1.2 Principal Investigators must be aware of University and funder data requirements/terms and conditions of award; any project-related contractual and/or ethical agreements, and organise and direct data management activities, systems, and services accordingly. Principal Investigators should ensure all Researchers in the project team are aware of research data management and sharing policies and data requirements, as well as any contractual agreements which may impact on research data (for example, where there is a confidentiality agreement, which would necessitate a restriction or embargo to the release of data).
- 3.1.3 Postgraduate Researchers are expected to be familiar with relevant, University and funder data policies and requirements, and best practice in their discipline.

- 3.1.4 All Supervisors of Postgraduate Researchers are expected to provide support and guidance to supervisees. The Primary Supervisor must ensure supervisees are aware of research data management and sharing policies and data requirements, as well as any contractual agreements which may impact on research data (for example, where there is a confidentiality agreement, which would necessitate a restriction or embargo to the release of data).
- 3.1.5 As a means to ensure the responsible management of research data, all student (PGR) and staff research projects are required to have a Data Management Plan (DMP) in place prior to the commencement of research (termed the 'Provisional DMP')<sup>1</sup>. This applies to all research projects regardless of whether there are funder data requirements/terms and conditions of awards. The Provisional DMP should document key decisions concerning data collection, storage, sharing and preservation (including what is and is not appropriate to share), and be kept up to date during the course of the project. As outlined in Section 4, in the interests of transparency and to enable monitoring of research data management and sharing practice, a copy of both the Provisional DMP and the version in place at project completion (termed the 'Final DMP') should be uploaded to the [Data Management Plan inbox](#). These DMP documents will be retained for the same period as the associated research data (see section 4.6). If there is no associated research data, the documentation will be retained for three years following project completion to inform policy and guidance.

### **3.2 Research data should be made openly available where possible, in a timely and responsible manner and any restrictions should be minimised and justified.**

- 3.2.1 All Researchers should seek to make data 'as open as possible, as closed as necessary', respecting any necessary restrictions on research data sharing due to legal, ethical, security and/or commercial constraints or safety concerns, and support the preparation and preservation of data in the most appropriate manner. Where appropriate, consent for data sharing should be sought from research participants at the outset, and considered as part of the ethical approval process.
- 3.2.2 Principal Investigators are responsible for assessing the suitability of data for public sharing, in line with any external and University research data management and sharing requirements. It is the responsibility of the Principal Investigator to ensure that data is prepared and preserved via data deposit in Pure or an external repository subject to the appropriate restrictions, license and re-use agreements. As outlined in Section 4, where an external repository is used, a metadata record must be created in Pure.
- 3.2.3 For Postgraduate Research projects, decisions about the suitability for sharing/publication, and future reuse of research data, should be taken jointly between the Postgraduate Researcher and the Primary Supervisor, in consideration of any ethics and/or project collaboration agreements and commercialisation issues, as well as funder data policy/requirements and University policy. The Primary Supervisor is ultimately responsible for ensuring that data is prepared and preserved via data deposit in Pure or an external repository subject to the appropriate restrictions, license and re-use agreements, though it is expected to be the Postgraduate Researcher who completes these tasks. As outlined in Section 4, where an external repository is used, a metadata record must be created in Pure.

### **3.3 Researchers should ensure that appropriate records are kept to demonstrate the integrity of their research.**

- 3.3.1 Proper research data management is a key part of ensuring the integrity of the research record. All Researchers should maintain a robust and reliable audit trail, which can be followed to establish the authenticity of any discovery or invention and to defend the results of genuine research endeavour.

---

<sup>1</sup>Doctoral research students are not expected to have a Data Management Plan (DMP) in place prior to commencing their studies but they should implement a project DMP in line with Departmental requirements for PGR progression review, or, just prior to commencing the collection/generation of research data.

- 3.3.2 Principal Investigators are responsible for ensuring that record-keeping is sufficient to demonstrate the quality, consistency and authenticity of the research. Examples of good record-keeping in research include retaining master copies of data, and applying a systemic version-control strategy for subsequent versions (iterations); appropriately marking-up revisions – without deleting content; developing a Data Management Plan to document ongoing decisions about data (file) management; and employing robust quality-control procedures.
- 3.3.3 For Postgraduate Research projects, data management decisions must be recorded in a Data Management Plan (see Section 4), as stated in the [Code of Practice for Postgraduate Research Students](#). It is the responsibility of the Postgraduate Researcher to maintain records in order to demonstrate the integrity of their research to their Supervisory Team, to examiners and to the wider academic community once published.
- 3.3.4 For Staff research projects which are collecting, generating and/or using data, data management decisions must be recorded in a Data Management Plan in accordance with University/Funder policies (if applicable) and sector best practice.

## 4. Operational requirements

The following operational requirements, which flow from the policy requirements in Section 3, are outlined to support Researchers to meet the principles and requirements related to research data management and to ensure research data are managed and disseminated in accordance with good research practice.

### 4.1 Research projects working with research data are expected to create a project Data Management Plan (DMP) and data management costs should be factored into project applications as part of the grant submission process.

- 4.1.1 In accordance with University/funder policies and sector best practice, all student (PGR) and staff research projects are required to have a DMP. As outlined in 3.1.5, a Provisional DMP should be in place prior to project commencement and updates should be made as required over the course of the project (treating it as a ‘living’ document), resulting in a Final DMP on project completion. The DMP should document key decisions concerning data collection, storage, sharing and preservation as appropriate to the project and in accordance with applicable funder data requirements and/or contractual/ethical agreements. In the interests of transparency and to enable monitoring of research data management and sharing practice, a copy of the Provisional DMP should be uploaded to the [Data Management Plan inbox](#) prior to project commencement and the Final DMP (reflecting any decisions/updates made during the project) should be similarly submitted once the project is complete.
- 4.1.2 If the required resources to ensure effective research data management and sharing are not already available, these should be costed into project applications as appropriate. Researchers should consider costs that will be directly incurred, for example, researcher time to prepare data for preservation. Some research funders will cover expenses and costs relating to data management activities, if fully justified at application stage; for example, UKRI guidance states that ‘Support may be requested for any costs related to research data management which will be incurred during the project and not be covered by directly allocated or indirect costs.’<sup>2</sup> Guidance on costing research projects can be sought from the [Costing and Pricing Team](#) in RKES.
- 4.1.3 Researchers can use the University’s [generic DMP template](#), which contains local information and prompts. Tailored templates, such as those that adhere to funder requirements, can be found on [DMP Online](#). When applying for funding grants and bids, Researchers are likely to be asked to structure the DMP to fit the funder or scheme-specific template.

---

<sup>2</sup> See further information on [directly incurred costs in UKRI funding](#).

- 4.1.4 Researchers can upload draft DMPs to the [DMP Inbox](#) for review and feedback. To assist Researchers with preparing a plan, there are also Data Management Plan tips on the RDMS website; [training is available](#); and one-to-one meetings can be arranged to discuss data management planning for individual projects (requests should be made via email to [researchdatapoints@strath.ac.uk](mailto:researchdatapoints@strath.ac.uk)).

## **4.2 Arrangements for the long-term storage of large datasets should be factored into data management planning and costed into grant applications where appropriate.**

- 4.2.1 A decision on the long-term storage of large datasets will be assessed when offered for deposit. Datasets are considered large if they cannot be adequately dealt with by traditional data processing applications and infrastructure due to their size and complexity.
- 4.2.2 If the research outputs from a funded project are expected to include very large datasets, this should be identified and costed at the DMP and grant application stage. This information will support RDMS Service decisions on storing and archiving large datasets.
- 4.2.3 Very large data can be deposited in Pure on condition that it is appropriately packaged and documented. No individual files should exceed 20GB. FAIR consideration for re-use should ensure the data is accessible from the Pure Knowledgebase. Extremely large datasets may be held off-line in exceptional circumstances, in bespoke storage locations operated by the RDMS Service.
- 4.2.4 Storing the files in such bespoke storage locations is not conducive to FAIR data principles, negatively impacting the findability and accessibility of the data. Planning at an early stage with the DMP will mitigate this possibility including any compliance requirements pertaining to data accessibility, and identify cases where the size of the data means that such bespoke storage locations become necessary.
- 4.2.5 Projects must contact RDMS support to discuss the deposit of very large data, ideally at the DMP and deposit stages.

## **4.3 Rights, licences, ownership and ethical considerations related to research data should be recognised and agreed at the earliest opportunity within a project.**

- 4.3.1 Where research involves collaboration with external parties/partners, Strathclyde Researchers must seek advice from [Grants and Contract staff in RKES](#) to arrange an appropriate contract and/or collaboration/consortia-type agreement that clearly records the rights each party possess and permissions each party provides to ensure research data can be used for current and future research.
- 4.3.2 Data should be deposited with a clear license on the reuse of data. License options are available when finalising deposits in Pure. [CC-BY](#) is the preferred license for open data, which allows data users to copy, modify and distribute data in any format for any purpose, including commercial use, provided they acknowledge the data creators and give appropriate credit, provide a link to the license, and indicate if changes were made.
- 4.3.3 Not all research data are suitable for open/public sharing, for example, research data containing protectable intellectual property or 'personal data' about research participants. In such cases, Researchers are encouraged to contact RDMS staff to discuss the most appropriate options for the data. Licence and re-use agreements can be tailored according to project requirements.

## **4.4 Research data must be securely stored and managed throughout the time period they are retained.**

- 4.4.1 The University will identify **active storage systems**, which are suitable for storage of data during the active phase of research (while a project is ongoing), and **long-term preservation systems**, which are suitable for storage of data after a project has completed. Decisions on which storage systems are considered suitable for each phase will be taken by the University's RDMS staff, and kept under

regular review. These decisions will be taken in consultation with IS Infrastructure staff, and communicated to researchers.<sup>3</sup>

- 4.4.2 Whilst a project is ongoing, research data should be securely stored on resilient, robust [University active storage systems](#), and managed in line with the associated DMP. However, data should not be stored long-term on these systems. Data/files held on active storage platforms, such as Departmental/School i: drives, are not subject to the same preservation measures as they are when deposited in Pure (or another data repository). Moreover, data/files held in storage, linked to an individual's DS login account (e.g., H: drive; Strathcloud personal folders; OneDrive for Business, etc.) will be deleted if/when the staff member/student leaves/graduates from the University.
- 4.4.3 Following project completion, data should be archived for long term accessibility, curation and preservation in a suitable long-term preservation system. When data are deposited in Pure - as a dataset - they are catalogued, associated with metadata, assigned a persistent link/DOI, and are subject to preservation activity by Library staff, and thereafter retained for a minimum period of 10-years. Completed datasets deposited in Pure are additionally processed using the Archivematica Digital Preservation system.
- 4.4.4 Once the project has completed, and the completed data has been stored on a suitable long-term preservation system, the data on the active storage system must be reviewed and deleted if it is no longer required on that system. Decisions around deletion of data from the active store must recognise the significant ongoing financial cost to the University, and the environmental impact, of retaining data in the active store unnecessarily.
- 4.4.5 Sensitive data, for example, research data containing protectable intellectual property or 'personal data' about research participants, can be stored in Pure with the appropriate confidentiality restrictions, for example, by restricting visibility of the complete dataset, or applying an embargo to specific files. Restrictions should be minimised by any means possible, for instance by providing accessible information about how the data was collected, whether it can be accessed on a one-to-one basis under a non-disclosure agreement (NDA), or whether a subset of the data can be made available). Guidance on handling and preparing sensitive data is available from the [RDMS Service](#).

#### **4.5 Research data and accompanying metadata and documentation must be registered in the University's research information system, Pure.**

- 4.5.1 To support University monitoring and reporting on the research data produced by its Researchers, and to ensure the data are Findable, dataset records must be created and ultimately finalised in Pure following project completion or at the point of publication if this occurs during the project. As a general rule, data should be deposited in Pure within 3 months of project completion, unless necessary restrictions apply.
- 4.5.2 In cases where a dataset is deposited with a trusted external data repository, for example, a subject-specific repository or one operated by the lead institution on a project, Researchers must create a dataset record in Pure, with metadata and a persistent identifier, such as a Digital Object Identifier (DOI), directing to the source of the externally deposited dataset. There are numerous data repositories, some domain-specific, others which are advocated by a funder or publisher. Researchers should check if external data repositories specialising in their research area already exist and prioritise deposit in these repositories as they are designed for their disciplines and datasets. Assistance on locating such resources and evaluating their suitability against FAIR principles, University policy and any project-specific agreements can be accessed from the [RDMS Service](#).

---

<sup>3</sup> Current examples of suitable **active storage systems** include: departmental i:drives, Teams and SharePoint; and examples of suitable **long-term preservation systems** include: Pure (to be preserved in Archivematica), external repositories, and bespoke storage locations operated by the RDMS service for storing very large datasets.

4.5.3 As far as possible, data deposited externally should be openly accessible. If external data are deposited in a restricted format and the dataset is at risk of deletion, Researchers should offer this data for deposit in Pure for preservation purposes.

- 4.5.4 In cases where the research is sensitive and/or confidential, a Researcher may request that the data (and descriptive metadata) are embargoed or restricted (for instance, to allow time to publish research findings, or comply with contractual obligations). Decisions to approve requests will be made by the RDMS Service in the first instance, although responsibility for ensuring deposited datasets with sensitive or confidential data are restricted in Pure lies with the Researcher.
- 4.5.5 Each dataset in Pure should have an assigned contact responsible for dealing with any enquiries relating to the dataset. The contact person will normally be the Principal Investigator. If the Principal Investigator leaves the University, a new contact person should be identified before the departure date. If no new contact person is identified, it will default to the relevant Head of School /Department.
- 4.5.6 Where a dataset record draft is started but not finalised, it can remain as an 'entry in progress' record for a maximum of 24 months, after which point the record will be deleted. Where a dataset record is required to remain at 'entry in progress' for longer than 24 months, an appropriate maximum retention period should be noted in the dataset description.
- 4.5.7 Dataset records must be related to any other relevant content in Pure including publications, projects, activities and impacts.

#### **4.6 Research data that contribute to published research findings, or that have acknowledged long-term value, must be retained for a minimum of 10 years, or longer if stipulated by the funder or advised by discipline convention.**

- 4.6.1 Research data that contribute to published research findings must be deposited as in the requirement above and must be retained in order to demonstrate the integrity of the research and any published research findings. For guidance on how research data deposit activities align with the publication timeline, see the [RDMS Integrated Workflow](#).
- 4.6.2 Research data that underpins research findings must be preserved for a minimum period of 10 years, from date of last access/use. In some disciplines, such as clinical trial data, the advised retention period may be 20 years; for environmental data, this may be indefinite.
- 4.6.3 Some funders require that data with acknowledged long-term value also be preserved and remain accessible and usable for future research<sup>4</sup>. Data with acknowledged long-term value should be preserved for a minimum period of 10 years, from date of last access/use, as above.
- 4.6.4 A minimum retention period longer than 10 years may be specified by third parties for specific types of research output. Researchers can check the required retention period with the [RDMS Service](#).

#### **4.7 Documentation accompanying research data should be sufficient to understand, analyse, replicate, and re-use this data.**

- 4.7.1 Descriptive Metadata should be sufficient to ensure discovery, even in cases where public access to the data is restricted or has been removed.
- 4.7.2 Documentation may be contained in many research outputs, including survey questionnaires, codebooks, standard operating procedures, informed consent forms, processing scripts, software code, workflows and other research outputs. Documentation should comply with the guidelines appropriate to the discipline in which the research is performed, and the type of research data produced.

---

<sup>4</sup> For example, see [UKRI Principles on Research Data](#).

- 4.7.3 At a minimum, a Readme file describing the uploaded dataset files, file naming conventions, and folder structures must accompany finalised dataset deposits in Pure. It is the responsibility of Researchers to provide sufficient description in their metadata. Datasets with inadequate metadata may be returned to Researchers for further detail.
- 4.7.4 In cases where a research project/study involves external collaboration, Researchers must consult the relevant collaboration agreement for details of agreed publication processes/protocol. In the absence of a formal collaboration agreement, decisions concerning the sharing of data and publication of results should be documented in the Data Management Plan prior to project commencement; any subsequent changes to the plan should be made with the consent of all parties concerned.

#### **4.8 Published research outputs should include information on how to access deposited research data and other supporting material.**

- 4.8.1 Researchers must include a Data Access (Availability) Statement in their Accepted Author Manuscripts (AAM). Data Access Statements in publications - 'even where there are no data associated with the article or the data are inaccessible' - are a requirement of UKRI policy<sup>5</sup>, and many publishers also ask for them.<sup>6</sup>
- 4.8.2 Within the Data Access Statement, the dataset should be cited using a persistent identifier such as a Digital Object Identifier (DOI), which will be provided during the dataset deposit process in Pure. Data Access Statements provide transparency and improve the Accessibility of the data - in keeping with FAIR principles. Moreover, their inclusion alerts readers to the existence and location of data, subsequently contributing to dataset usage/downloads, which will be a factor in assessing the overall impact of the research.
- 4.8.3 Researchers should refer to the [University Research Publications Guidance](#) which advises on key aspects of preparing a research publication, including Data Access Statements. Further information is available on the [Data Access Statements](#) webpage.

---

<sup>5</sup> 'Peer-reviewed research articles submitted for publication on or after 1 April 2022, which acknowledge funding from UKRI or any of its councils, include a Data Access Statement, even where there are no data associated with the article or the data are inaccessible', [UKRI Open Access Policy](#),

<sup>6</sup> Some examples of publishers who expect data statements in publication manuscripts are: PLOS, [Data Availability](#); and Springer Nature, [Data Availability Statements](#).

## 5. Rights relating to research data

Ownership is a critical consideration in the management and sharing of research data. All right-holders should be identified at the earliest opportunity so that appropriate action can be taken to uphold their rights in accordance with relevant legislation and policies including the [UK General Data Protection Regulation \(UK GDPR\)](#); [UK Copyright law](#); and the University's [Intellectual Property \(IP\) & Commercialisation Policy](#). Besides Researchers, typical holders of rights to research data are collaborators, research participants, existing data owners and the University (as employer). The following examples illustrate some of the considerations that relate to these:

- **Collaborators:** Where research projects/studies involve collaboration with partners, industry, NHS etc., in most cases collaboration and contractual agreements will be drawn up to manage and document these partnerships. These legal agreements and contracts must be factored into project data management planning so that data are managed in line with agreed processes and licence agreements, and to ensure that data, which may, for example, not be suitable for public disclosure are not released inadvertently. The project Principal Investigator is responsible for appraising supervisees about any contractual agreements and the University's Research and Knowledge Exchange Services (RKES) should be consulted in matters relating to legal, ethical, or commercial rights. If a formal collaboration agreement is not required, decisions regarding the storage and sharing of data should be documented in the Data Management Plan prior to project commencement, with any subsequent changes made with the consent of all parties concerned.
- **Research participants:** Where research has involved the collection and/or processing of personal data, data protection legislation and ethical requirements must be complied with to ensure that the participants' rights are respected. Where a research project/study has involved collecting data from people, for example, from focus groups or interviews, and there is a plan to publish large extracts of participants' words, appropriate transfer of copyright must be in place.<sup>7</sup>
- **Existing data owners:** Where a project/study has used existing or secondary data sources, Researchers would not be expected to deposit existing/secondary data sources in Pure (unless there is a project agreement to state otherwise) - as they belong to the original data creators/owners. Instead, Researchers should include details of any published existing data sources used in the research by including the hyperlinks to the sources, initially in the DMP, and in the metadata - i.e., when describing the dataset. This way anyone wishing to scrutinise the research, would know where, and how, to access the existing/secondary data sources used, and then they too could arrange access for themselves. New data, derived from analyses of existing/secondary sources, can be deposited in Pure (if appropriate).
- **The University:** In most cases, the IP generated by University Staff and Postgraduate Researchers in the course of their employment is the property of the University (making research data a University-owned asset). For this reason, research data must be deposited into the institutional repository, Pure, at the end of the project or when a Researcher leaves employment and/or study with the University.<sup>8</sup>

---

<sup>7</sup> For data collected via interviews that are recorded and/or transcribed, the Researcher holds the copyright of recordings and transcripts but each speaker is an author of his or her recorded words in the interview (Padfield, T (2010) Copyright for archivists and records managers, 4th ed., London: Facet Publishing)., UK Data Service, [Copyright — UK Data Service](#)

<sup>8</sup> Researchers should consult the University's [Intellectual Property \(IP\) & Commercialisation policy](#) for full details. Research students should also consult the [Code of Practice for Postgraduate Research Students](#) for specifics of how this policy applies to them.

## Appendix A – Glossary of Terms

<b>Archivematica</b>	An 'open-source digital preservation system'. <a href="#">Archivematica</a> is used by Library staff to make a preservation copy of the datasets deposited in the University of Strathclyde's institutional data repository within <a href="#">Pure</a> .
<b>Data Access (Availability) Statement</b>	<p>A statement, provided within a publication manuscript, which confirms the existence of and access conditions to, research data underpinning and/or associated with said publication.</p> <p>Often a requirement of funder and publisher policy, these statements provide clarity and transparency, and aid data sharing and data discoverability. Their inclusion within a manuscript ensures compliance with funder data policies/expectations.</p> <p>Further information and example statements are available from <a href="#">the Data Access Statements web page</a>.</p>
<b>Data Management Plan</b>	<p>A written document which describes the data that will be collected or generated during a research project and sets out a detailed plan for how the data will be managed throughout the project and what will happen to it after the project completes. The plan will include information about formats, volume, documentation, storage, sharing, and destruction of data.</p> <p>A 'Provisional DMP' is the version in place at project commencement and a 'Final DMP' is the version in place at project completion.</p> <p>A DMP should be a useful document which helps researchers understand how best to ensure data is findable, accessible, interoperable, and reusable throughout, and beyond, the project lifecycle. Further information and templates are available from <a href="#">the DMP web page</a>.</p>
<b>Data Owner</b>	A person or organisation who owns a data/dataset, as outlined in the related contract and/or licence agreement.
<b>Data Creator</b>	A person/s who created/generated the data (files, etc.)
<b>Deposit</b>	<p>To upload (completed) research data, accompanied with descriptive metadata, in a research data repository (e.g., <a href="#">Pure</a>) from where it can be catalogued, preserved, and made openly available or restricted – where appropriate.</p> <p>Further information on depositing data is available from <a href="#">the Data Deposit web page</a>.</p>
<b>Digital object identifier (DOI)</b>	A unique character string which identifies the location of a journal paper or dataset and can also be used as a persistent hyperlink in a related publication's <a href="#">data access statement</a> . Further information is available from the <a href="#">Get a digital object identifier web page</a> .
<b>Documentation</b>	Information which documents and accompanies research data; related to methodology, documentation should be sufficient to allow others to scrutinise research findings and enable (or contribute to) reproducibility of research. See <b>Metadata</b> entry below, for more info.
<b>Embargo</b>	As it relates to research data, an embargo is a restriction applied to data (files) to delay the public release of said data for a period of time, as befits the discipline and related publication policies /processes.
<b>FAIR Data Principles (also 'FAIR')</b>	A set of principles, created by FORCE11, they provide guidance on how research data can be made <b>F</b> indable, <b>A</b> ccessible, <b>I</b> nteroperable, and <b>R</b> eusable. Further information can be found on <a href="#">the FORCE11 website</a> .

<b>Intellectual Property (IP)</b>	<p>'Patents, utility models, rights to inventions, copyright and related rights (including moral rights), trademarks, rights in designs, rights in get-up, business names and domain names, software algorithms and code (as a special case of copyright), data, databases, rights to use and protect the confidentiality of confidential information (including knowhow and trade secrets), semiconductor topography rights and specialist types of IP protection such as plant breeders' rights'.</p> <p>For further information read the University's <a href="#">Intellectual Property &amp; Commercialisation Policy</a>.</p>
<b>Licence</b>	<p>An arrangement or agreement, in which a person or party grants permission to another person or party.</p> <p>In the case of data, a license may be used by a licence holder (Data Owner or custodian) to grant permission to another person or party to use the data, subject to compliance with terms &amp; conditions of use, as outlined in the licence agreement.</p>
<b>Metadata</b>	<p>Can be described as 'data about data'; often intended to be machine-readable, it may include the following:</p> <ul style="list-style-type: none"> <li>• Purpose</li> <li>• Date, data were generated</li> <li>• Person generating data</li> <li>• Software version</li> <li>• Equipment used</li> <li>• Equipment settings</li> <li>• Sample ID</li> </ul> <p>It may also be in the form of a Readme or text file, which documents /accompanies the data collected/generated. The types and levels of metadata &amp; documentation will often vary between disciplines. Matters such as access restrictions (e.g., where data are restricted due to privacy or confidentiality clauses) are also a form of metadata.</p>
<b>Open Data</b>	<p>'Data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike', as defined in the <a href="#">Open Data Handbook</a></p>
<b>Personal Data</b>	<p>According to the <a href="#">Information Commissioner's Office (ICO)</a> 'Personal data only includes information relating to natural persons who:</p> <ul style="list-style-type: none"> <li>• can be identified or who are identifiable, directly from the information in question; or</li> <li>• who can be indirectly identified from that information in combination with other information'.</li> </ul>
<b>Pure</b>	<p>A Current Research Information System (CRIS) produced by Elsevier. The platform is used by the University of Strathclyde and its Researchers, for the recording of research information (including, researcher profiles; outputs; projects; datasets; equipment; activities; etc.). Datasets which are deposited in Pure are then processed and stored as preservation versions in Archivematica.</p> <p><a href="#">Pure</a> enables each individual researcher to view all the current data held about their research activity in one place. For guidance on using Pure, see <a href="#">the RKES Portal</a>.</p> <p>The public portal of Pure is called <a href="#">KnowledgeBase</a>; the portal is searchable across multiple web search engines and enables the dissemination and sharing of research information and datasets.</p>

<b>Research Data</b>	‘The evidence that underpins the answer to the research question, and can be used to validate findings regardless of its form (e.g. print, digital, or physical)’, as defined in the <a href="#">Concordat on Open Research Data</a> .
<b>Research Funder</b>	A funding council, body, or charity, who provides financial support to Higher Education Institutional and Researchers, to help them to conduct research/research projects.
<b>Research outputs</b>	Refers to items that are produced as an outcome of research and can take many forms. In <a href="#">Pure</a> and <a href="#">KnowledgeBase</a> , output types include journal articles, conference papers, chapters, books, reports, posters, theses, non-textual forms (e.g. software, artefacts, exhibitions, compositions) and patents. Although datasets are a form of output, in Pure/KnowledgeBase, there is a distinct category for <a href="#">Datasets</a> .
<b>Researcher</b>	For the purposes of this policy, the term ‘researcher’ is used to refer to: <ul style="list-style-type: none"> <li>(i) University staff (whether holding employment or worker contracts) involved in the creation, collection, generation or processing of research data.</li> <li>(ii) Postgraduate Researchers (PGR) involved in the creation, collection, generation or processing of research data.</li> <li>(iii) Individuals who do not fall within (i) and (ii) above but who are otherwise associated with the University and are authorised to use the University’s name, facilities and/or services when involved in the creation, collection, generation or processing of research data (such persons would include, for example, those holding visiting, honorary or emeritus status).</li> </ul>
<b>Restricted</b>	Restricted data are data which are not suitable for public sharing. Restrictions may be based on issues around commercial confidentiality, or safe-guarding concerns; where applied, they should be justified and made clear.
<b>Retention Period</b>	As it pertains to research data deposited in Pure, this denotes the period of time a dataset will be retained in this platform. The default retention period at the University of Strathclyde is 10-years, although retention periods will be informed by relevant funder policies and dataset download/usage stats.
<b>Sensitive Data</b>	Refers to data that required additional protection. This could be special category data, which is personal data that reveals or concerns certain types of data ( <a href="#">see ICO website for list</a> ), or commercially sensitive data, e.g. data which contains protectable intellectual property (See <b>Intellectual Property</b> entry above).