

RESEARCH DATA DEPOSIT POLICY

Version No.	Description	Author	Approval	Effective Date
1.1	Research Data Deposit Policy	RKES	RKEC (25/11/2014)	25 Nov 2014

11/25/2014

Version 1.1

the place of useful learning

The University of Strathclyde is a charitable body, registered in Scotland, number SC015263

[Research data](#) is valuable and because of its value should be retained so that its worth can be realised through access and reuse potentially for many years to come. This policy deals with data stored electronically (otherwise known as 'digital data'); data in physical forms is dealt with in a separate document.

As all digital data is software dependent, there is a real and inherent risk that it will become unreadable or unusable over time as software evolves and/or becomes obsolete. Storing data in a place that can attend to its digital [preservation](#) and [curation](#) reduces the risk of data becoming unusable. University of Strathclyde provides a [repository](#) for the storage and, where appropriate, sharing of [Research data](#) via the Pure system and also supports deposit in approved external repositories.

This policy outlines the requirements for the minimum data quality standards, file format specifications, deposit and licensing procedures expected for [Research data](#) deposit in Pure. Not all of these requirements may be applicable or appropriate for every research project but data should be inspected to ensure that an appropriate level of conformance is achieved.

Roles & Responsibilities

University of Strathclyde aims to support and facilitate data deposit and sharing by its [Researchers](#) generally. Projects funded via RCUK, EU and other funders are mandated to store and share their [Research data](#). [Researchers](#) with [datasets](#) derived from projects not funded by a mandating funder are also strongly recommended to deposit in PURE to protect their data.

Research data deposit in Pure does not preclude deposit with any [External data repository](#) and in many cases deposit with an [External subject based repository](#) is a mandated funder requirement. If depositing externally the [Data Depositor](#) must additionally create a [metadata](#) record in Pure with adequate documentation and links to ensure that the data remains discoverable and compliance is maintained.

As far as possible, [Researchers](#) should ensure that deposited [datasets](#) are of high quality, contain no errors and are well documented. Data content and quality are entirely the responsibility of the [Data Depositor](#) and prior to deposit data should, at a minimum, be:

- Fully virus checked
- Unencrypted
- Free from password protections
- In an uncompressed format where ever possible
- Anonymised where required

Data Integrity & Metadata

Deposit in Pure should constitute the original, complete and final copy of the [Research data](#) or a time-stamped copy of a continuing [dataset](#) at a particular point in time e.g. the end of a period of funding or on publication.

Where necessary an updated version of a [dataset](#) may be deposited. In such cases the earlier version will be removed from public view, but not deleted, and links created to clearly identify the relationships between [datasets](#).

Data will be assessed for its suitability and readiness for persistent access and re-use by [Research Data Support staff](#). Assessment will consist of a series of data and format integrity checks, it will explicitly **not** concern itself with the worthiness of the intellectual content. Where data does not conform to the minimum requirements for deposit, [Research Data Support staff](#) will engage with the researcher to identify an agreed solution for any data management related issues which could enable the [dataset](#) to be uploaded.

Metadata

To aid future access and the effective curation of [datasets](#), in addition to completion of the maximum possible number of [metadata](#) fields and deposit of the data itself in PURE, the corresponding [Data Management Plan](#) should be stored alongside links to any publications and funding awards related to the dataset. In total the information supplied should enable anyone to independently understand the data within the collection.

If the [dataset](#) is made available publicly, the [metadata](#) collected in Pure may be re-used in any medium without prior permission for not-for-profit purposes and re-sold commercially provided the OAI Identifier or a link to the original metadata record are given.

Data and File Formats

All data submitted to Pure will be assessed on an individual basis by the [Research Data Support Staff](#). In order to comply with funder recommendations and the University [Research Data Policy](#), every effort must be made to ensure deposited data remains authentic, [intelligible](#) and, where ever possible, available online for re-use. Filenames, file version control and data/folder structure will be assessed for any potential access and preservation issues.

When [Research data](#) is deposited and assessed, [Research Data Support Staff](#) will potentially create and store three different versions of the [dataset](#): [original](#), [distribution](#) and [preservation](#) copies.

Very diverse types of electronic data are produced at University of Strathclyde and it is not possible for comprehensive digital curation activities to be performed on all data types nor is it sensible to restrict the types of data that can be produced and/or stored. Therefore while all data types will be able to be deposited in Pure, a selective policy with regard to preferred file formats and their corresponding level of support will operate.

Data will be evaluated against the following criteria and assigned a corresponding level of support:

Preferred formats

Non-proprietary open source or openly documented formats which are extensively adopted and supported by a wide range of software platforms. These formats are best suited to long-term preservation and reuse and will receive [full digital preservation and curation support](#).

Acceptable formats

Proprietary formats, usually created by commercially available software, will depreciate in favour of newer versions however, they are also widely adopted and will be unavoidable for

practical reasons. These formats will receive [full preservation support but will receive more limited curation support](#) and are likely to require migration to a preferred format in order to remain sustainable.

Migration routes will be agreed in advance and published to ensure researchers are aware of what will happen to deposited data. The migration routes will be regularly updated in light of formats of incoming data deposit.

Problematic formats

Proprietary, unrecognised or unsustainable formats, which have poor documentation, are not widely adopted or supported by available software platforms. These formats will receive [“bit-level” best-effort preservation support](#), i.e., they will be stored and distributed “as deposited” with no curation activities applied to guarantee future usability.

Prioritisation

To ensure that the University of Strathclyde meets its commitments certain types of [dataset](#) will be prioritised for assessment as follows:

- [Datasets](#) associated with external funding awards where funder mandates data sharing.
- [Datasets](#) associated with publications
- Stand-alone data publications
- Other categories of [dataset](#) (uniqueness, non-replicability, scientific/historical value)

Prioritisation categories will be re-assessed at regular intervals alongside information on volume of deposits.

Format guidelines and prioritisation information will be held on the Research Data Management online support site.

Rights, Permissions & Ownership

The individual depositing [Research data](#) in Pure must be the [rights-holder](#) or have the authority of the [rights-holder](#) to do so. All contributors to the data, and their role, will be assigned in Pure as will be the designated data contact known as the [‘Data Owner’](#).

[Research data](#) must not breach any applicable law nor infringe the copyright of any person or third party. All necessary permissions must be secured for any third party data which forms part of the [dataset](#).

All legitimate restrictions on access or re-use of the [Research data](#), including embargos, must be clearly identified and, if necessary, be supported by the inclusion of relevant documentation.

[Research data](#) must be deposited in compliance with relevant funder guidelines and timescales where these apply.

Withdrawal/Take-down of Data from Public View

[Research data](#) which explicitly, or potentially, breaches University regulations, funder mandates, UK and Scottish law on:

- Research ethics
- Copyright infringement
- Disclosure of confidential or sensitive content (commercial and personal)
- Plagiarism or falsified research
- Expired funder data retention timescale

will not be accepted for sharing or will be withdrawn from public access immediately. The PURE [dataset](#) metadata record will be retained as a record. The validity and authenticity of submissions are the sole responsibility of the depositor.

In the event of Pure being removed or replaced as the University institutional repository for [Research data](#), data will be transferred to another appropriate facility and [data owners](#) notified.

License Agreements

License agreements are necessary to safeguard the rights of the University, the [Data depositor](#) and the [research data](#). A minimum of two agreements must cover all data deposited:

- [Depositor](#) Agreement: Declaration by the depositor that data is theirs & breaks no laws. Sets out what rights are granted to Pure as the holding repository. The depositor agrees to this on deposit in Pure.
- End user Agreement/Licence: Sets out what the recipient of the data can and cannot do and any incumbent responsibilities. The depositor must indicate which licence data is available under when depositing in Pure.

This policy operates in conjunction with other relevant University and funder policies:

[Research Code of Practice](#)

[Research Data Policy](#)

Definitions

Research Data: Research data is digital information created in the course of research but which isn't a published research output. Research data excludes purely administrative records. The highest priority research data is that which underpins a research output. Research data do **not** include publications, articles, lectures or presentations. Also referred to as 'data' and 'datasets'.

Authentic: Data which displays three significant provable properties – data is what it purports to be, was created by the purported person, and was created at the purported time.

Data deposit: uploading/saving of data files into a repository.

Repository: a storage location, often for safety or preservation, in this case for research data.

Researcher: in the context of this policy 'researcher' includes:

- i. University staff (whether holding employment or worker contracts);
- ii. Registered postgraduate research students;
- 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 engaging in research activities (this includes those holding visiting, honorary or emeritus status).

External data repository: a repository not owned, managed or controlled by University of Strathclyde. University of Strathclyde has only one institutional data repository which is Pure.

External subject based repository: a repository created for the collection of research data related to a particular specialism.

Approved external repositories: A list of repositories which can be used to store University of Strathclyde-owned research data will be maintained and published. Researchers will be able to deposit data in these repositories rather than the University of Strathclyde repository if this constitutes a better strategy for dissemination and reuse. In order to facilitate compliance reporting, a record must be completed in Pure for any dataset deposited in an external repository.

Data Depositor: The person who uploads data to Pure or another repository. This is normally the person designated with responsibility for data management. The person responsible for data management is the Principal Investigator unless otherwise formally delegated as per the University of Strathclyde [Research Data Policy](#).

Metadata: data or information about data which describes or contextualises the data. In this case information about datasets stored in a repository/database template. For example, an image may require metadata that describe how large the picture is, the colour depth, the image resolution, when the image was created, and other data. A text document's metadata may contain information about how long the document is, who the author is, when the document was written, and a short summary of the document.

Research Data Support staff: Specialist members of staff experienced in data management planning and able to assist researchers will all matters relevant to research data management including deposit, curation, preservation, etc. These staff are based in the Research Data Management & Sharing Team.

Data Management Plan (DMP): A formal working document which outlines how data will be handled both during the active research phase and after the project is completed. DMPs in some form are now a requirement of a research grant proposals and therefore must be addressed at the earliest phase of the research lifecycle.

Metadata: Data or information about research data that describes or identifies the research data. Information supplied to the Pure dataset template is 'metadata'.

Intelligible: Ensuring that data can be understood over time by providing adequate levels of metadata to describe not only the technical components of a resource but also its context, purpose and the processes involved

Digital preservation: An umbrella term to describe various processes to ensure digital data are made accessible and intelligible over time. At a basic level the process of storing the bits & bytes of a digital object over time, which while maintaining the existence of a digital object does not guarantee its continued usability.

Digital curation: Differs from digital preservation in that curation involves actively **managing** digital data over time. Curation requires action on the part of curators so that resources remain secure, discoverable and useable throughout data lifecycle. The active management of research data reduces threats to their long-term research value and mitigates the risk of digital obsolescence. Curation includes: selection, appraisal, preservation, disposal and transformation for, e.g., migration to an updated format.

File Versions:

Original: the dataset as originally deposited in Pure

Distribution: the dataset normalised and migrated where appropriate to a version suitable for sharing

Preservation: the dataset normalised and migrated to a format most suitable for long-term preservation.

Full Digital Preservation and Curation Support: Migration to preservation formats, creation of shareable formats and management of the data to ensure that it remains accessible and usable over time.

Full preservation support but more limited curation support: Migration to preservation formats, bit-level preservation but not all files will be able to be migrated to new versions of proprietary software indefinitely depending on licences available at University of Strathclyde and business continuation of software providers

Bit-level best-effort preservation support: preservation of the file as it was submitted. Includes maintaining onsite and offsite backup copies, virus checking, fixity-checking, and periodic refreshment by copying files to new storage media. In other words, maintaining the integrity of the original file is preserved for later dissemination

Rights-holder: The person(s)/organisation(s) who own the intellectual property within the research data.

Data owner: The named contact designated 'data owner' in Pure. This is the primary contact for the research data. A data owner must always exist for a dataset in order that there is a responsible person who can be contacted when decision, such as retention, are required. The data owner may also be a data creator and/or the rights holder. If a data owner should leave their employment at the University then Research Data Support Staff are automatically

notified by the Pure system and a relevant replacement Data owner is assigned responsibility for the dataset.

