



Collections Development Policy

External

18 February 2021

Version: 07.00

T +44 (0)1206 872001

E collections@ukdataservice.ac.uk

ukdataservice.ac.uk

Contents

1. Scope	2
2. Definition of Terms	2
3. Roles and Responsibilities	3
4. Policy	4
4.1. Background.....	4
4.2. Scope of Collections	4
4.3. Curation Categories	5
4.4. External Influences	5
4.5. Principles of Collections Development	6
5. Procedures for Acquisition	7
6. Methods of Acquiring Data Collections	7
7. Selection and Appraisal Criteria	8

1. Scope

The Collections Development Policy provides an overview of the selection and appraisal criteria applied to the UK Data Service data holdings. It demonstrates a commitment to aligning with and meeting the data needs of our major stakeholder communities, other relevant organisational policies (both internal and external), national and international standards and legislative requirements.

Successful implementation of this Policy will ensure that only the most relevant and highest impact data collections are selected for ingest and curation, ensuring that resources are concentrated on ingesting collections that are either high-use or at risk of loss.

This Policy directly informs the key service requirements of the UK Data Service and significantly impacts on the UK Data Archive's Preservation activities.¹

2. Definition of Terms

Acquisition

The process related to acquiring materials for inclusion in the collection.

Archiving

A curation activity that ensures that data are properly selected, stored, can be continuously access, and that its logical and physical integrity is maintained over time, including authenticity and security.

Digital curation

The activity of storing, managing, presenting and promoting the use of data from its point of creation to ensure it is FAIR (Findable, Accessible, Interoperable, Reusable). Dynamic datasets will require continuous updating or enrichment to keep them fit for purpose. Higher levels of curation also involve maintaining links to contextual materials, such as published documentation and other related resources.

Data holdings

The whole of the UK Data Service's collections.

Designated user community

The UK Data Service's designated user community is made up of social science and related data users within Higher and Further Education in the UK, though best efforts are made for the wider user community. All users are expected to have a basic understanding of social science methods and techniques relevant to the data collections being accessed.

¹ <https://dam.data-archive.ac.uk/controlled/cd062-preservationpolicy.pdf>

Data Owner

The rights holder in any data collection. This is not the same as the Data Producer or Depositor.

Data Producer

An individual involved in the creation of a data collection, who may also be the Data Owner.

Depositor

A Data Depositor is an individual or organisation who is named on a licence as having sufficient responsibility to grant particular rights on behalf of a data collection. The Depositor may be the principal investigator, creator or the copyright owner of a data collection, but does not have to be. In Open Archival Information System (OAIS) terminology the term producer is used in a similar sense.²

Ingest

The processes related to receiving information from an external source and preparing it for archival storage and access.

Licence Agreement

Used in this document as a generic term covering a range of different licences between an institution and the Data Owner.

Preservation

An activity within archiving in which specific items of data are maintained over time so that they can still be accessed and understood through changes in technology. Unlike 'archiving', preservation does not include maintaining the logical and physical integrity of a data object.

Selection

The process of identifying materials to be made accessible and in many cases to be preserved for long-term use.

3. Roles and Responsibilities

Service Director

Responsible for updating the Policy in line with current UK funder priorities, emerging research themes and user needs.

Collections Development team

Responsible for implementing the Policy through appraising, negotiating, licensing and acquiring data for delivery or resource discovery for the UK Data Service.

UK Data Service Advisory Committee

The UK Data Service Advisory Committee is responsible for this Policy. This responsibility will be restricted to determine the strategic development of the Collection.

The UK Data Service will review the Policy on an annual basis and present it to the Advisory Committee on a periodic basis (every 2 years). The Advisory Committee will:

- Ensure governance arrangements are correct;
- Ensure scope of materials is correct;
- Ensure selection and appraisal arrangements are fit for purpose;
- Make recommendations for acquisitions and determine (additional) priorities for the following

² <http://www.oais.info>

- period;
- Provide advice surrounding the strategic implementation of the Policy;
- Help develop a methodical approach to identifying new data sources and research areas with specific data needs;
- Work with data funders/suppliers to provide access to new sources;
- Increase the scope and research value of the UK Data Service collections by developing new relationships to facilitate data access.

Data Appraisal Group (DAG)

Responsible for reviewing offers of data and third party data sources and deciding the appropriate pathway for a data collection, in line with the Policy.

UK Data Service

The UK Data Service will:

- Maintain and review this policy and associated legal/rights management framework to ensure both are fit for purpose;
- Monitor and amend internal processes to accommodate and incorporate these changes;
- Implement the UK Data Service Collections Development Policy;
- Report to the UK Data Service Advisory Committee biannually on how the Policy is being implemented;
- Report to the UK Data Service Advisory Committee on the use of any licensed data collections for which a subscription/fee is paid;
- Develop a methodical approach to identifying new data sources and research areas with specific data needs;
- Work with the UK Data Service Advisory Committee and data funders/suppliers to provide access to new sources;
- Design a formal complaints procedure in case of data owners and producers objecting to the decisions of the Service about the relevance of a data collection to this Policy;
- Work with members of the UK Data Service Advisory Committee to increase the scope and research value of the UK Data Service collections by developing new relationships to facilitate data access.

4. Policy

4.1. Background

The Policy demonstrates a commitment to aligning the data needs of our major stakeholder communities: data users, data owners and data funders and ensuring that those needs can be met by the data service infrastructure provided by the UK Data Service. The Policy is intended to be flexible enough to meet the challenges and changing demands from these communities while ensuring alignment with resource provision for data services and with relevant organisational policies (both internal and external), national and international standards and legislative requirements.

The Policy shall be responsive in both meeting the demands of changes in technology and resulting new types of data. The UK Data Service will advance the mission of the UKRI ESRC by investigating and encouraging the deposit of new types of data and new resources in the collection in response to current and emerging research themes and user needs.

4.2. Scope of Collections

The UK Data Service collection is a wide-ranging resource of great value to research and teaching communities across and beyond the social sciences in the UK and internationally. Core data holdings are comprised of large-scale government and ESRC-commissioned survey and aggregate data, as well as other data created with UKRI funding.

UK Data Service acquires data to meet these three central purposes:

- *Potential secondary use and analysis for research:* to enable researcher access, asking new

questions, undertaking a restudy or follow-up study, or undertaking methodological work, to materials originally created to inform and support research, or to new and novel forms of data not directly collected for academic purposes but of interest to the social scientist;

- *Teaching and learning use*: to provide access to data that have been created or repurposed specifically for teaching purposes to enhance use of the collection and support skills development;
- *Replication and validation of research*: to provide researchers with access to original research, which allows checks for consistency, validation of sources and validation and replication of research outputs, both published and unpublished.

4.3. Curation Categories

The policy of the UK Data Service will be to treat collections in four broad categories:

1. *Data collections selected for long-term curation*. These data will have long-term secondary analysis potential. These collections are likely to be made available for download, or accessible via online access tools;
2. *Data collections selected for “short-term” management*. These collections will not (initially) be retained for long-term preservation, rather they will be backed-up (i.e., bit-level preservation only), made accessible and discoverable through online access tools or repository software (ReShare);³
3. *Data collections selected for ‘delivery’ only*, e.g., where data from third parties are accessed via APIs/web services and delivered to end users via a UK Data Service interface. Issues such as level of trust in owner, what documentation/metadata are required, and how rights/registration are handled would need to be agreed;
4. *Data collections selected for “discovery” only*. These collections will not be brought formally into the holdings of the UK Data Service, they will exist only in other (institutional) repositories, but the UK Data Service will harvest (or in exceptional circumstances, create) metadata records to allow these data collections to be found more easily.

Data collections may initially fall into categories two or three, but be moved to category one at a point of time in the future.

4.4. External Influences

The primary policy driver for the UK Data Service’s Collections Development Policy is the ESRC Research Data Policy.⁴ This policy demonstrates the ESRC’s commitment to the provision of access to data as the primary asset for economic and social research regardless of its underlying access criteria. The principles are guided by the earlier OECD’s report Principles and Guidelines for Access to Research Data from Public Funding and the UK Strategy for Data Resources for Social and Economic Research.⁵⁶

With the creation of the unified UK Research and Innovation (UKRI), cross-funder working has already led to greater opportunities for data users. Examples are opening up wider opportunities to access key administrative, biomedical and medical cohort data, encouraging planning for providing access to ‘Emerging’ forms of data, and enabling access to data from Low and Middle Income Countries via the Global Challenges Research Fund (GCRF).

The policy is also informed by the ONS Research Data Access Strategy and the 2017 Digital Economy Act Research Code of Practice and Accreditation Criteria that make a commitment to the provision of access to ONS microdata for statistical research.⁷⁸ This strategy is also enshrined in the Concordat between the ONS

³ These include UK Data Service online data access tools, Nesstar, QualiBank and UK.Stat

⁴ <https://esrc.ukri.org/funding/guidance-for-grant-holders/research-data-policy>

⁵ <https://esrc.ukri.org/files/research/uk-strategy-for-data-resources-for-social-and-economic-research>

⁶ <https://www.oecd.org/sti/inno/38500813.pdf>

⁷ <https://cy.ons.gov.uk/aboutus/transparencyandgovernance/datastrategy/datapolicies/onsresearchanddataaccesspolicy?uri=aboutus/transparencyandgovernance/datastrategy/datapolicies/onsresearchanddataaccesspolicy>

⁸ <https://www.gov.uk/government/consultations/digital-economy-act-part-5-data-sharing-codes-and-regulations/research-code-of->

and the UK Data Service. The UK Census 2021/22 programmes further offers the opportunity to provide access to enhanced data products under this policy for the research community.⁹

There are many external policies, which may indirectly impinge on the implementation of this Policy. These include legislation, e.g., the Census Act, 1920; the Statistics and Registration Service Act, 2007, §39(5); The Digital Economy Act, 2017, international standards, e.g., ISO 27001; and HMG Infosec documents.

4.5. Principles of Collections Development

The Policy is founded on the following principles:

1. Advancing the strategic objectives of the ESRC and, in future, the UKRI;
2. Contributing to the successful delivery of the ESRC Research Data Policy;
3. Contributing to the successful delivery of the UK Strategy for Data Resources for Social and Economic Research and subsequent post-2018 plans;
4. Supporting the designated user community by providing access to a wider range of data sources and resources, regardless of access method (i.e., Open, Safeguarded, Controlled);
5. Ensuring that data collections brought into the service are of the highest quality;
6. Promoting and supporting high-quality basic, strategic and applied research in the social sciences;
7. Promoting teaching and learning use of data and other resources to encourage secondary usage;
8. Strengthening existing partnerships with data owners and producers whilst establishing new collaborative partnerships and creating new opportunities.

Commentary

Principles 1 and 2: The UK Data Service as a data service provider is mandated by the ESRC through its Research Data Policy to provide advice and guidance to ESRC award holders on implementing data management and sharing plans. UK Data Service will provide guidance to award holders on issues relating to confidentiality, security and ethics in data management and sharing and will offer guidance on providing additional information to research participants on the purposes and benefits of data archiving. We will also work with the ESRC/MRC Cohort Resources Facility (CLOSER) with population health data.

Principle 3: A key driver of the 2013-18 UK Data Forum's five-year plan, *UK Strategy for Data Resources for Social and Economic Research*, is the provision of both good scientific training and the 'breadth and depth' of readily available data resources, aspects which will be reflected in this policy. The UK Data Service acknowledges stakeholder investment and looks to this Strategy and other related documents to provide guidance for the strategic planning, thematic drivers and further development of the collection. This plan, together with the influential 2013 OECD report on *Understanding the Human Condition*, will support the expansion of the existing materials to encompass new sources of data and new data types according to stakeholder need and current research demand.¹⁰ The UK Data Service will work with the ADR-UK to support their mission to enable access to the UK's administrative data assets for research, and the UK Census bodies to support access to enriched census data research products.

Principle 4: This work continues to support the long established principle and commitment of the ESRC to ensure that publicly funded research data are a public good, produced in the public interest and should be openly available to the maximum extent possible. The UK Data Service will strive to maximise the access to any data it collects while meeting legislative and legal requirements.

Principles 5, 6 and 7: The UK Data Service will ensure the authenticity, reliability and logical integrity of all resources entrusted to our care while providing usable versions for research and teaching or learning in perpetuity. The UK Data Service will ensure that its user community has access to the materials it requires by continuing to maintain existing resources, and the currency of its major data series.

Principle 8: The UK Data Service will work with European partners to provide and facilitate access to international data in both quantitative and qualitative forms. We will work to facilitate new collaborative partnerships and explore opportunities to contribute to the development of federated resourced discovery

[practice-and-accreditation-criteria](#)

⁹ <https://www.ons.gov.uk/census>, <https://www.scotlandscensus.gov.uk/2021> and <https://www.nisra.gov.uk/statistics/census/2021-census>

¹⁰ OECD (2013) New Data for Understanding the Human Condition: International Perspectives <http://www.oecd.org/sti/sci-tech/new-data-for-understanding-the-human-condition.pdf>

efforts and methods and tools for data curation and exploration.

5. Procedures for Acquisition

The UK Data Service implements the following steps for data acquisition:

Identification: involves identifying potential acquisitions and seeking information about the contents of data collections from a variety of sources: links with data producers, via user groups and targeted consultations, workshops, seminars and via publications. Selection criteria for ingest are shown in section 4;

Negotiation: includes evaluating the data collections suitability for inclusion, advice and guidance about the Service, levels of access, risk in data and user terms and conditions, advice and discussion about ethical and legal aspects, for example rights management including copyright and IPR together with information about data protection, extent of data documentation, and data transfer methods;

Guidance for depositors: The Service provides advice and guidance on the benefits of and procedures for depositing data collections, including a published series of Guides to Good Practice to support depositors and Research Data Management Guidance, as well via the successfully published 2019 second version of the Sage Handbook, *Managing and Sharing Research Data: A Guide to Good Practice*; ¹¹

Legal framework: UK Data Service provides a legal framework under which data collections are deposited and researcher access is managed. A Licence Agreement confirms the rights and obligations of parties and offers an opportunity for depositors to specify the conditions under which access to third parties may be given. An End User Licence reflects these terms for the End User;

Administration: The UK Data Service maintains records of acquisitions and potential acquisitions, recording details of the contact name, address, high-level collection description and status of any negotiations undertaken. In addition the UK Data Service and ESRC have a collaborative working space which allows us to contact potential depositors once an ESRC award has been granted to answer any queries they may have and to provide advice on procedures to ensure the smooth accession of their resulting research outputs into the collection;

Producer accounts: The UK Data Service provides an online system for depositors to submit data, metadata and agree appropriate data sharing licences.

6. Methods of Acquiring Data Collections

The governance model under which the UK Data Service operates will demonstrate effective collaborative working. This relationship ensures that data generated from publicly funded research are deposited, preserved and made available for re-use. All researchers receiving ESRC awards are contractually obliged to offer their resulting data, or metadata in cases where this isn't feasible, for deposit.

The UK Data Service will continue to contribute to support robust data policy and sharing for other research funders, including the MRC and Wellcome; and also to collaborate with European and global partners to maintain existing, and to facilitate new, data exchange agreements between researchers and research bodies.

The UK Data Service will continue to maintain and to strengthen its relationship with the Office for National Statistics (ONS), and other data-generating government departments where appropriate. The intention is to safeguard and extend the collection of major series materials through knowledge transfer via bespoke training and the preparation of data management and deposit guidelines for central government departments.

The UK Data Service will work with others, for example public and private sector organisations to extend the available data for policy-related research and open innovation. This will include directly sourcing data via APIs and web services.

¹¹ <https://uk.sagepub.com/en-gb/eur/managing-and-sharing-research-data/book262873>

7. Selection and Appraisal Criteria

The Collections Development Policy is supported by selection and appraisal criteria that are both transparent and accountable to enable monitoring of processes and planning. Materials are rigorously reviewed to determine if they may be viably managed, preserved and distributed. The criteria used to review materials consider:

- *Relevance to the remit:* The content of the materials meet the ESRC's current strategies and are within the 'Scope of Collections' (section 1.2 above). This includes being high quality, authoritative, reliable and up-to-date with good temporal and spatial coverage, with long and consistent time series (as appropriate), and strong opportunities for comparable research;
- *Scientific or historical value:* The materials are considered important resources either for research purposes and/or meet current demands in research trends and/or there is research and/or teaching activity in the subject area covered by the data;
- *New sources or types of data:* The materials fill a gap in the collection, may supplement areas of the collection, meet new research, such as interdisciplinary, drivers and/or have been requested by users;
- *International value.* The materials are considered to be important for international research;
- *Uniqueness/risk of loss:* The materials form a rare and complete resource of value to the research community and/or are at risk of loss if not acquired; data collections which are costly to reproduce or where continued access to data is threatened; data are not generally available in any other form e.g. paper;
- *Usability/redistribution/operational benefits:* The materials have been well prepared according to data format/documentation and ethical/legal guidelines and fall within Scope; accession to the UK Data Service makes the resources more accessible. Where data is made 'Open' by the data owners, the UK Data Service may continue to make accessible versions of the data because it fits well with the portfolio of other data;
- *Replication Data and Resources:* Materials and related outputs required to reproduce published or unpublished research, or undertake a restudy.

Where materials fall outside of scope, efforts are made to determine whether another archive, data centre or institutional repository is more appropriate for dissemination, curation or preservation of these data.

There are also criteria for not accepting data:

- *Legal and ethical issues:* there are insurmountable rights management issues e.g. consent, IPR, copyright and Data Protection issues which are unable to be satisfactorily resolved and where full use of the data would not be possible without infringing legislation;
- *Sample size:* there are problems with the sample size, the data are very localised or are of peripheral interest for social scientists;
- *Lack of sufficient contextual materials to enable re-use.*

The ESRC and the Host Organisations of the UK Data Service have invested heavily in the Service and through services provided in the past have provided researchers with access to internationally enviable and long-running resources such as *Understanding Society* and the *British Birth Cohort Studies* and through its collaboration with government departments, to a long series of nationwide Household Surveys. In addition, it has provided access to high quality ad-hoc surveys from respected research teams and from government and charities. Additionally, in the past, ESRC has provided funds to make available aggregate data from international bodies such as OECD, IMF and World Bank; much of which is now open data, and no longer requiring bespoke purchase and licensing.

This unique and rich collection benefits from extension into new areas for example: administrative data, commercial data, transactional data, geo-referenced data and economic data. More recent funding from the EPSRC has enabled the curation and delivery of large-scale UK smart meter and related household contextual data.

A Collections Development Selection and Appraisal criteria document sits alongside this Policy to enable implementation of the Policy suggestions or materials identified for inclusion in the collection can be

recorded, and their ingest managed in a transparent and accountable manner.¹²

¹² <https://ukdataservice.ac.uk/media/455175/cd234-collections-appraisal.pdf>