



## UK Data Service – Collections Development Policy

Public version  
18 March 2013  
Version: 02.00

### 1. Scope

The Collections Development Policy provides an overview of the selection and appraisal criteria applied to the UK Data Service data holdings. It outlines the principles by which the UK Data Service develops its data holdings to meet best the needs of all its stakeholder communities. It also outlines the practices by which the Universities of Essex and Manchester (as two of the hosts of the UK Data Service hold licences for these data and how these fit into the overall rights management framework for collections development. At present this policy does not outline the practices of licensing and data holding for other institutions within the UK Data Service Value-Added Services (VAS)). The 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 limited resources are not expended on ingesting potentially low-use data collections.

### 2. Policy

#### 2.1. 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.

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

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 ESRC by investigating and encouraging the deposit of new types of data and new resources in the collection.

This policy will deal with both the scope of collections and the level to which data collections are curated.

#### 2.2. Scope of Collections

The UK Data Service collection is a wide-ranging resource of enormous value to research and teaching communities across and beyond the social science discipline in the UK, EU and globally. The core data holding is comprised of large scale government and ESRC-commissioned survey and aggregate data, as well as other data created with ESRC funding.

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

- *Potential secondary use and analysis for research:* to enable researcher access to materials originally created to inform and support research, but asking new questions or undertaking a restudy or follow-up study;
- *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.

### 2.3. 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<sup>1</sup> or via repository software (currently ESRC Data Store);
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.

### 2.4. External Influences

The primary policy driver for the UK Data Service's Collections Development Policy is the ESRC's Research Data Policy.<sup>2</sup> 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. This policy is in turn guided by the central principles of the OECD's report on 'Promoting Access to Public Research Data for Scientific, Economic, and Social Development'.<sup>3</sup> Other policy drivers are the National Data Strategy,<sup>4</sup> which promotes the provision of both good scientific training and the 'breadth and depth' of readily available data resources and the National Strategic Directorate for e-Social Science.

The policy is also informed by the draft ONS Research Data Access Strategy which makes a commitment to the provision of access to ONS microdata for statistical research and in turn is allied to the ONS's Policy for Social Survey Microdata.<sup>5</sup> This strategy is also enshrined in the Concordat between the ONS and the UK Data Service. Other initiatives such as the Open Government Licence,<sup>6</sup> the proposed Open Data Corporation will impact on the scope of the policy.<sup>7</sup>

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); international standards, e.g., ISO 27001; and HMG Infosec documents.

### 2.5. Principles of Collections Development

The Policy is founded on the following principles:

1. Advancing the strategic objectives of the ESRC;
2. Contributing to the successful delivery of the ESRC Research Data Policy;
3. Contributing to the successful delivery of the National Data Strategy;
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, Licensed, Secure);
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 NDS 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 the NDS to provide guidance for the strategic planning, thematic drivers and further development of the collection. This guidance 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.

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 acquire new data sources such as transaction data, footprint data and commercial data. The UK Data Service will work to incorporate standards such as the INSPIRE Directive to extend the research value and impact of existing material. We will work to expand the collection of geo-spatial data and environmental data building upon the materials acquired from the Rural and Environmental and Land Use Programme.

### 3. Procedures for Acquisition

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

*Identification:* identifying potential acquisitions and seeking information about the contents of data collections from a variety of sources: links with data producers, via user groups, workshops, seminars and via publications. Selection criteria for ingest are shown in section 5;

*Negotiation:* the negotiation process includes, evaluating the data collections suitability for inclusion, advice and guidance about the service and levels of access, advice and discussion about the legal aspects, for example rights management including copyright and IPR together with information about data protection and secure methods of data delivery, and adequacy of data documentation;

*Guidance for depositors:* The UK Data Service will provide advice and guidance on the benefits and procedures for depositing material, including a published series of Guides to Good Practice to support depositors and Research Data Management Guidance;



*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 and 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, brief collection level 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 into the collection;

*Producer accounts:* The UK Data Service will develop a facility to upload data and documentation, and to provide a mechanism whereby data producers can view and investigate usage statistics.

## 4. Methods of Acquiring Data Collections

The governance model under which the UK Data Service will exist demonstrates 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 for deposit.

The UK Data Service will continue to contribute to European developments and 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 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. This will include directly sourcing data via APIs and web services.

## 5. 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 considers:

- *Relevance to the remit:* The content of the materials meet the ESRC's current strategies and are within the 'Scope of Collections' (section 2.2 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 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 replicate 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 resources such as *Understanding Society* and through its collaboration with government departments, to a long series of Household Surveys. In addition it has provided access to high quality ad-hoc surveys from respected research teams and from government. Additionally, the ESRC has provided funds to make available data from international bodies such as OECD, IMF and World Bank.

This unique and rich collection will benefit from extension into new areas for example: administrative data, commercial data, transactional data, geo-referenced data and economic data.

A Collections Development Plan sits alongside this Policy where initial suggestions or materials identified for inclusion in the collection can be recorded and their ingest managed in a transparent and accountable format. This will be presented to the UK Data Service Governing Board on an annual basis and reported on in intervening meetings.

## 6. Responsibilities

### 6.1. Collections Development Committee

The Governing Board of the UK Data Service are responsible for this Policy and effectively form an independent Collections Development Committee (CDC). 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 Governing Board for approval. The Governing Board will:

- Ensure governance arrangements [sections 6.1 and 6.2] are correct;
- Ensure scope of materials is correct [section 2.2];
- Ensure selection and appraisal arrangements are fit for purpose [section 5];
- Make recommendations for acquisitions and determine (additional) priorities for the following year [section 0];
- 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.

## 6.2. 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 Governing Board annually on how the Policy is being implemented;
- report to the Governing Board 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 Governing Board 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 Governing Board to increase the scope and research value of the UK Data Service collections by developing new relationships to facilitate data access.



## Definitions

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

*Archiving:* A curation activity which ensures that data is properly selected, stored, can be accessed and that its logical and physical integrity is maintained over time, including security and authenticity.<sup>8</sup>

*Curation:* The activity of, managing and promoting the use of data from its point of creation, to ensure it is fit for contemporary purpose, and available for discovery and re-use. For dynamic datasets this may mean continuous enrichment or updating to keep it fit for purpose. Higher levels of curation will also involve maintaining links with annotation and with other published materials.<sup>9</sup>

*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 HE and FE in the UK, though best efforts are made for all users. (For the purpose of this definition, quantitative and qualitative data collections created by or for historians are considered to be 'social scientific'.) All users are expected to have a basic understanding of social science methods and techniques relevant to the data collections being accessed.<sup>10</sup>

*Data Owner:* The rights holder in any data collection. Not to be confused with the Data Producer.

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

*Depositor:* A depositor is an individual who is named on a Licence Agreement as having sufficient responsibility to grant particular rights to the UK Data Archive on behalf of a data collection. The depositor may be the instigator, creator or the copyright owner of a data collection, but does not have to be. In OAIS terminology the term producer is used in a similar sense.

*End User Agreement:* The correct form for what is currently called the End User Licence.

*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. Note that preservation differs from Archiving in as much as preservation does not include maintaining the logical and physical integrity of a data object.<sup>11</sup>

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



## References

[URLs checked 18 March 2013]

<sup>1</sup> These tools include Nesstar, InFuse, dotstat, etc.

<sup>2</sup> ESRC Research Data Policy (2010) <http://www.esrc.ac.uk/about-esrc/information/data-policy.aspx>

<sup>3</sup> P Arzberger, et al, 'Promoting Access to Public Research Data for Scientific, Economic, and Social Development' Data Science Journal, Volume 3, 29 November 2004, 135-151. Available at: [https://www.jstage.jst.go.jp/article/dsj/3/0/3\\_0\\_135/\\_pdf](https://www.jstage.jst.go.jp/article/dsj/3/0/3_0_135/_pdf)

<sup>4</sup> ESRC National Data Strategy (2009) [http://www.esrc.ac.uk/\\_images/NDS%20publication\\_tcm8-5155.pdf](http://www.esrc.ac.uk/_images/NDS%20publication_tcm8-5155.pdf). This strategy is currently being revised. The new version will be entitled the UK Strategy for Data Resources for Social and Economic Research.

<sup>5</sup> Currently this is only in draft form. See: <http://www.ccsr.ac.uk/sars/events/2011-01-28/ons.ppt>. See also <http://www.ons.gov.uk/ons/guide-method/best-practice/disclosure-control-policy-for-social-survey-microdata/gss-microdata-policy.pdf>

<sup>6</sup> Open Government Licence for public sector information.  
<http://www.nationalarchives.gov.uk/doc/open-government-licence/>

<sup>7</sup> Korn and Oppenheimer (2011) Licensing Open Data: A Practical Guide.  
[http://discovery.ac.uk/files/pdf/Licensing\\_Open\\_Data\\_A\\_Practical\\_Guide.pdf](http://discovery.ac.uk/files/pdf/Licensing_Open_Data_A_Practical_Guide.pdf)

<sup>8</sup> Lord and Macdonald (2003) Data curation for e-Science in the UK: an audit to establish requirements for future curation and provision.  
<http://www.jisc.ac.uk/media/documents/programmes/preservation/e-sciencefinal.pdf>

<sup>9</sup> Lord and Macdonald (2003).

<sup>10</sup> UK Data Archive Preservation Policy (2011) <http://www.data-archive.ac.uk/media/54776/ukda062-dps-preservationpolicy.pdf>

<sup>11</sup> Lord and Macdonald (2003).