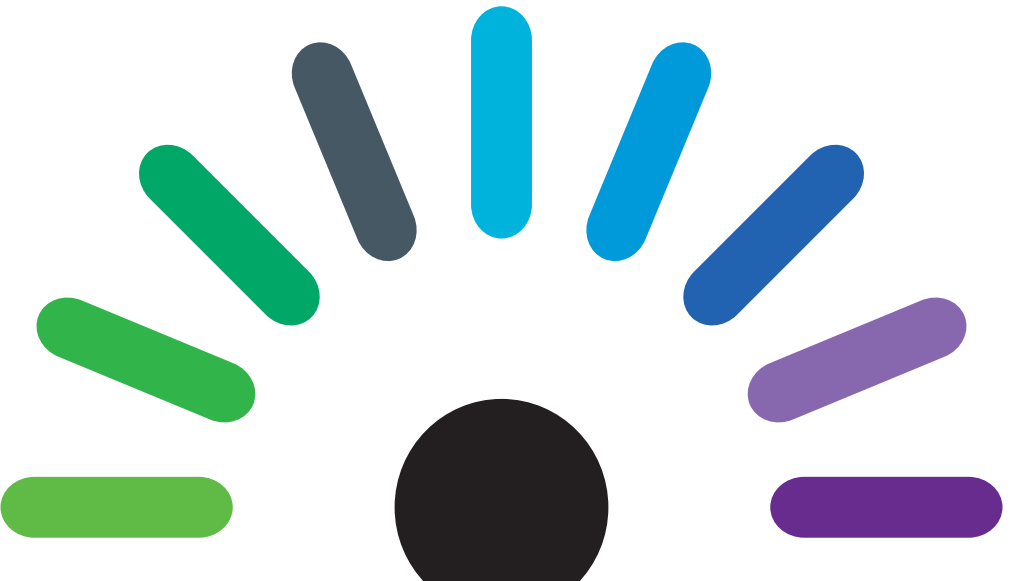


Depositing
shareable
survey data

UK Data Service







PLAN

What to do before fieldwork begins



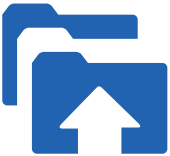
PREPARE

What to do during and after fieldwork



NEGOTIATE

Discuss deposit with us



DEPOSIT

Transfer data to us



INGEST

What we do with your data

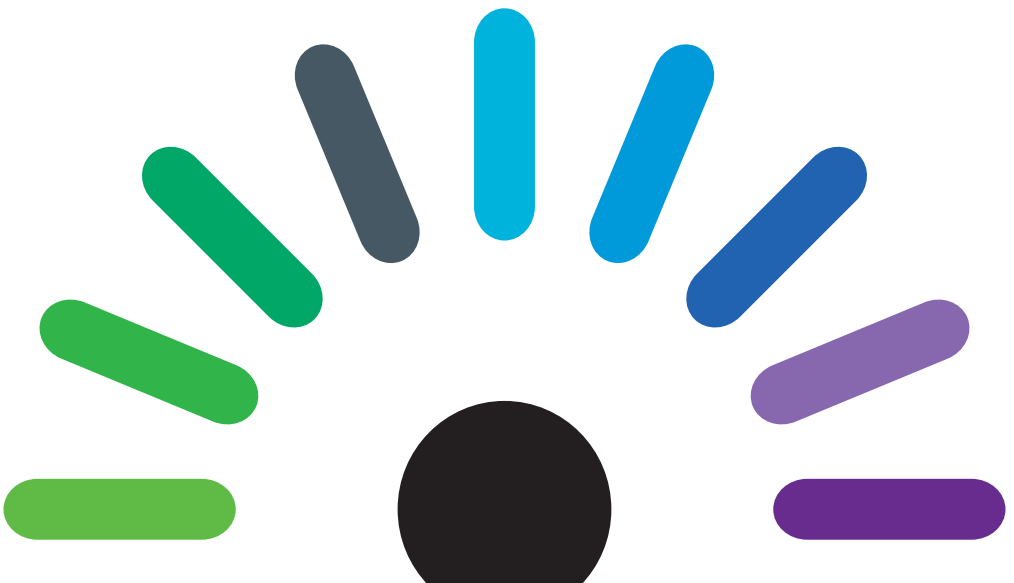
Introduction

There are clear demonstrable benefits to data owners and data creators as a result of sharing survey data. Most obvious is the contribution to the transparency agenda but also, data sharing provides an additional return on investment and demonstrates value for money through further use of data.

This guide is for survey commissioners, survey managers and contractors who are involved in producing and archiving social survey data. It is aimed primarily at UK government-sponsored operations although it is relevant to other organisations undertaking large-scale survey data collection.

The guide outlines key recommendations for depositing survey data with us, the UK Data Service, the Economic and Social Research Council's (ESRC) flagship organisation, supported to provide a data service to researchers and beyond. Using formal structures like the UK Data Service to deposit data protects data owners' and subjects' rights effectively and can lead to increased visibility of the data creator.

We can provide usage statistics and impact case studies to the data owner which offer leverage in discussions about prioritisation of future activities.



Using this guide

In writing this guidance we recognise that not all survey operations are of the same scale or equally resourced. The guide thus sets out practices at the ideal end of the spectrum, for example, those followed by long-running, large-scale continuous surveys used to produce national or official statistics. We fully appreciate that there may be some choices to be made about data preparation depending on budget and scale, and, of course, on the survey owner's own requirements.

Survey commissioners should be aware that specifying a high standard of data delivery across all projects may be desirable in generating richer deliverables and in creating a level playing field between providers, but runs the risk of increasing project costs.

Plans for depositing and providing access to data should always be considered early in the survey commissioning and design operation processes. They should include allocating appropriate resources and setting out exactly who has responsibility for each set of activities.

If it is not possible for you to follow all of our recommendations in order to offer your data to us, contact us to discuss in more detail. There is always a positive way forward.

Getting in touch early

Contact the UK Data Service Collections Development team at the earliest opportunity if you intend to deposit a new dataset. We can then discuss and agree appropriate access arrangements for these data. We are also able to provide strategic advice on the benefits of data sharing

Email: collections@ukdataservice.ac.uk



WHAT TO DO BEFORE FIELDWORK BEGINS



// Fully briefed on the requirements of delivering a dataset //

Commissioning tenders

We recommend that a paragraph like the example below, or one that is tailored to the requirements, be included in commissioning tenders and project contracts. Contractors should be fully briefed on the requirements of delivering a dataset that can be archived and shared, within the constraints of the survey's budget.

Example

Data and documentation to be delivered to the UK Data Service by [an agreed date], to [an agreed quality standard] and prepared according to UK Data Service guidelines.

Set out expectations and link to this guide, perhaps with pointers to specific pages or sections.

Tenders should indicate exactly which party is expected to take responsibility for delivering archiving-related requirements. Some survey tenders in the UK are already becoming far more explicit in their data documentation expectations.



Example

The Centre for Longitudinal Studies (CLS) currently commissions the national cohort studies: the Millennium Cohort Study; the National Child Development Study; the 1970 British Cohort Study and the Longitudinal Study of Young People in England. CLS has started to expect as a deliverable the Computer Assisted Interviewing (CAI) implementation as a Data Documentation Initiative (DDI) compliant XML file, and a file that maps the CAI question to the data variables.

Consent protocols

The form of consent provided by survey respondents should allow for archiving and sharing of data and future reuse by researchers. Consider how informed consent will be handled and discussed in cases of particularly sensitive or confidential data.

Written consent may not be required for data gathered from questionnaire surveys where no personal data are collected, or where personal identifiers are removed from the data.

As for most surveys, an information sheet should be provided to participants, making clear that consent is implied from participating in the survey or by completing the questionnaire. That the data will be deposited with the UK Data Service can also be mentioned if it is considered helpful.

Example

The UK Data Service provides examples of wording that have been used to explain the nature of participation consent in introductory letters in surveys: ukdataservice.ac.uk/manage-data/legal-ethical/consent-data-sharing/surveys.aspx

Deposit checklist

It is useful to have a checklist relating to deposit activities for each data collection. A good example of a checklist that is used by the ONS Opinions Survey team can be found at the back of this guide.



WHAT TO DO DURING AND AFTER FIELDWORK



// Give future researchers sufficient information ... reuse the data //

Preparing data files

- use consistent and meaningful file names that reflect the file content, avoiding spaces and special characters; if data are sensitive or restricted, indicate this in the file name
- use meaningful and self-explanatory variable names, codes and abbreviations
- ensure variable and value labels are complete and consistent, both questionnaire and derived variables
- ensure internal consistency checks are completed
- remove all your own temporary, administrative or dummy variables created for internal purposes/not of use to researchers
- ensure no repetition of variables, especially redundancy in derived variables
- check that the level of detail included in the data is suitable for the agreed access arrangements and licensing
- apply an appropriate level of anonymisation e.g. serial numbers anonymised so that they cannot be linked to other sources, any top coding applied, cases removed

Example

See the back of this guide for examples of typical variables that may need attention



- provide anonymised Primary Sampling Unit information if possible so that researchers can incorporate the sampling design into their analyses
- check that any textual variables included are suitable for dissemination e.g. no disclosive information or internal comments in free-text variables
- ensure consistent treatment and labelling of missing values
- include weights as variables but do not apply them in the deposited data files
- use our recommended file formats

Example

See our guidance on file formats:

ukdataservice.ac.uk/manage-data/format/recommended-formats.aspx

- if converting data across file formats, check that no data or internal metadata have been lost or changed
- finally, make sure that data are complete and try to ensure one deposit only, with any data issues resolved before deposit

Additional requirements for Nesstar preparation

Nesstar is the UK Data Service's online data browsing, analysis, subsetting and download tool that enables easy access to richly documented variables. Instant tabulation and graphing can be done. Full question text, universe and routing information is displayed alongside variable name, code values and labels, and frequencies.

nesstar.ukdataservice.ac.uk

A selection of key data, typically from government departments, are made available through our Nesstar service. These requires additional processing work to render them suitable for user-friendly online browsing.

Example of enhanced processing needed for Nessar publishing:

- variable and value labels must be clear and consistent, avoiding truncation of variable and value labels
- non-compliant characters, such as &, @ and <>, should be removed
- question text should be made available in as structured a format as possible, e.g. XML or spreadsheet



Preparing data documentation

Data documentation should give future researchers sufficient information to be able to understand and reuse the data. Consider what kind of survey, data, and other contextual documentation, can explain what the data mean, how they were collected and which methods were used to create them. For survey series, ensure that documentation refers to the current year's data.

Types of survey documentation

- survey technical report with standard headings, describing sampling, achieved sample size, fieldwork and weighting and so on – the level of detail may vary depending on the scale and resourcing of the survey

Example

User Guide for the Health Survey for England, 2011

discover.ukdataservice.ac.uk/Catalogue/?sn=7260&type=Data catalogue#documentation

- information leaflets and consent forms
- questionnaires, with universe and routing instructions
- showcards
- interviewer instructions – if these are commercially sensitive, then a summary of briefing content
- coding frames and coding instructions
- links to primary reports and publications
- information about questionnaire variables that have been removed and the reason for this e.g. confidentiality
- information about any known errors or issues in the data
- structured information for the UK Data service collection-level metadata record, using DDI-compliant metadata and controlled vocabularies, as set out in our deposit form



Example

The UK Data Services data deposit form utilises metadata from the Data Documentation Initiative (DDI), which is an international XML-based descriptive metadata standard for research data used by many social science data archives across the world.

DDI standard www.ddialliance.org/

UK Data Service deposit form: ukdataservice.ac.uk/deposit-data/support/regular-depositors.aspx

Types of data documentation

- variable list
- links from variables to questions in the questionnaire (CAI or otherwise)
- code book or data dictionary (we can also generate a data dictionary from the data files at the ingest stage)
- clear, unique definitions of variables

Example

Geographical identifiers or spatial units should be defined using:

- a unique name
- referenced definition or authority
- datestamp of when unit boundaries were defined, not when the sampling was recorded

- weighting variables described
- syntax for any derived variables
- where possible or practical, change in key content over time - questions and variables

Example

Summary of changes in topics and sampling from the User Guide for the Health Survey for England

doc.ukdataservice.ac.uk/doc/6986/mrdoc/pdf/6986userguide2010.pdf#13



DISCUSS DEPOSIT WITH US

// Categorise your data according to its impact level //



Access and licensing

- seek your department's or organisation's guidance from the division responsible for access, such as any local Microdata Release Panel, on the appropriate level of detail to be included in the data and the relevant access level - obtain legal advice where necessary
- where possible, categorise your data according to its impact level, which can then be used to determine the most appropriate access conditions
- alongside the 2014 Government Security Classification, many government departments may have their own impact level and recommendations for classification, so seek local advice on this
- it is possible to disseminate multiple versions of data under different access conditions according to our data access policy
- agree access conditions, copyright and licencing with the UK Data Service Collections Development team

Example

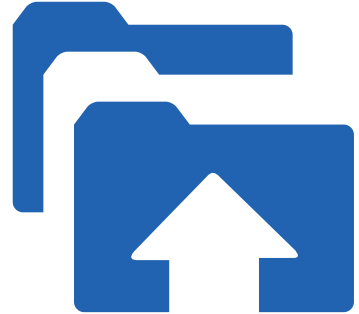
At the UK Data Service data may be licensed under one or more of these levels:

- open data – under open licence without any registration, either Open Government Licence (OGL) for Crown Copyright data or Creative Commons Attribution-ShareAlike 4.0 International Licence for other data
- safeguarded data – requiring an End User Licence, users to be authenticated, and, where appropriate, special conditions agreed to
- controlled data – requiring user accreditation and registration through training and approval by a data access committee, and users to be authenticated
- ukdataservice.ac.uk/get-data/how-to-access/conditions.aspx



TRANSFER DATA TO US

// Consult locally on your ... preferred method of data delivery **//**



Depositing data

- complete our licence agreement, to be signed by copyright holders; this is not required for ONS data as this is covered by a Concordat between the ONS and the UK Data Service
Form: ukdataservice.ac.uk/media/28102/licenceform.pdf
Guidance: ukdataservice.ac.uk/deposit-data/support/regular-depositors/licence.aspx
- complete the UK Data Service deposit form used to populate the Discover catalogue record for the dataset
Form: ukdataservice.ac.uk/media/158537/Data_Deposit_Form.docx
Guidance: ukdataservice.ac.uk/deposit-data/support/regular-depositors.aspx

Submit data and documentation files

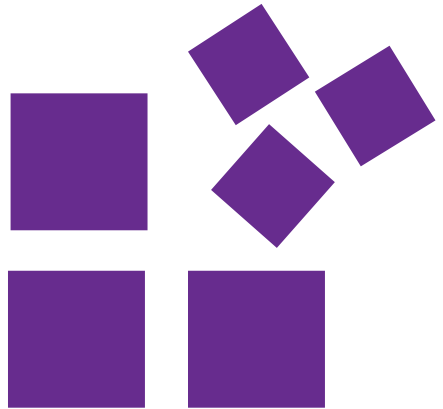
- consult locally on your department's or organisation's own preferred method of data delivery
- where possible, use the University of Essex ZendTo Service, addressing the deposit to the email account "ukdscoll@essex.ac.uk" and noting the study title or depositor surname in the dropbox description
ZendTo: zendto.essex.ac.uk/verify.php
- deliver in person or mail via portable media, e.g. CD/DVD/memory stick
- encrypt and send by courier data files that contain sensitive or personal information
Guidance: ukdataservice.ac.uk/manage-data/store/encryption.aspx
- include the completed deposit form and licence agreement with the data files or send separately by email to collections@ukdataservice.ac.uk.
- the licence agreement can be scanned or posted



INGEST

WHAT WE DO WITH YOUR DATA

// Assess the data before full processing begins //



Safeguarding your data

Since 2010, the UK Data Archive, the organisation that houses and hosts the data based at the University of Essex, has been certified under the international ISO 27001 standard for information security. A number of government departments have also carried out surveillance visits to the Archive to clarify our information security regime.

Our Information Security Management System supports all aspects of digital curation carried out in-house, such as the selection, acquisition, ingest, archiving, provision of access to data, and the management and planning of digital curation. The ISO certification allows us to handle secure data on site and supports secure remote access to these research-rich data.

In house, data are classified according to their level of detail, sensitivity and confidentiality and appropriate data handling and access safeguards are in place.

Contact the Collections Development team if your department needs more information about our accreditation and security arrangements.

In-house data review checks

Our Ingest Services team assess the data before full processing begins and makes recommendations regarding:

- confidentiality and data integrity
- file preservation and dissemination formats
- composition of documentation
- levels of validation needed

Any proposed changes are agreed with the depositor before work is carried out to prepare data for preservation and dissemination. Any changes made to the data, such as removal of a variable or editing of a variable for confidentiality purposes, are recorded in our processing notes.

In-house processing activities depend on the unique characteristics of each dataset, but in general, the following procedures are always undertaken:

- review data for disclosive information
- check for basic inconsistencies in data
- carry out data enhancement to agreed standards with UK government series data usually processed to 'A' standard

Our processing standards: www.data-archive.ac.uk/curate/archive-quality

- generate the collection in multiple file formats for dissemination and preservation
- collate and prepare user documentation as bookmarked PDF/A documents where possible
- create enhanced DDI-compliant catalogue metadata for our Discover (and for Nesstar), using information provided in the deposit form provided by the depositor and in accompanying documentation
- gather citations to related publications for inclusion in the catalogue metadata
- assign a Digital Object Identifier (DataCite DOI) for the data collection
- release' data via our Discover system, within 30 working days of the in-house ingest team receiving a dataset ready for full processing

More detailed information on data processing

Quick processing guide: www.data-archive.ac.uk/media/54770/ukda081-ds-quantitativadataprocessingprocedures.pdf

Survey processing guide: www.data-archive.ac.uk/media/54764/ukda080-ds-processingquickreference.pdf

DEPOSIT CHECKLIST

// Create a bespoke checklist //

For regular deposits of individual years or quarters of a series, we recommend that you create a bespoke checklist to follow for data you are depositing. We are happy to help create this.

Example

ONS Opinions Survey team: Internal checklist for regular deposits of the Opinions and Lifestyle Survey (OLS)

- file names are intuitive and make sense
- weight variables are included
- certain old data input errors from the past have not crept back into the data files
- the data are complete
- correct number of variables and cases in the data files
- the correct month's data are in the correctly named data files
- files deposited without the 'weight' on
- serial numbers (for EUL data) are anonymised
- the classificatory variables RAGE, DVHSize, Marital status, Relationship and Ethnicity (for EUL data) are top-coded
- cases for under 18s are removed from the smoking modules (for EUL data)
- any module variables considered disclosive are recorded or deleted (for EUL data)
- for OPN, showcards are not deposited – but a check is made in case they need to deposit a 'picture showcard' which, if necessary, will then be include

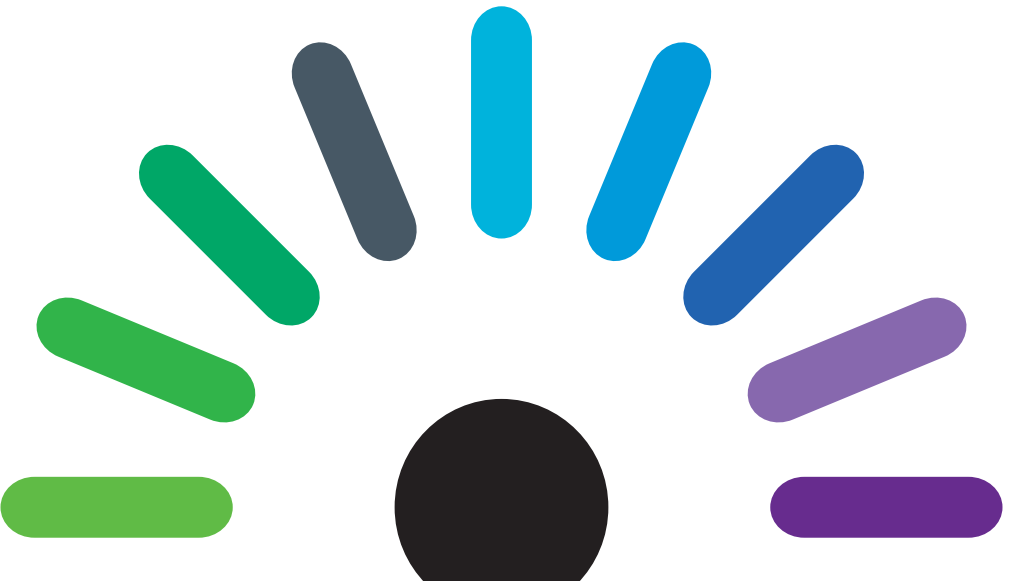
Ons opinions and lifestyle survey team pre-deposit spreadsheet

Cycle	Mod no.	Title	Client	Specs' removed	Variables deleted/amended from module	Serial numbers anonymised	Rage top-coded	Cases removed	Archive type
Aug-12									
Aug-12	MCZ	Well being	ONS	N/A	None	Yes	95+		EUL
Aug-12	M210	Tobacco	HMRC	None	None	Yes	95+	13 cases under 18's deleted	EUL
Sep-12									
Sep-12	MCZ	Well being	ONS	N/A	None	Yes	95+	None	EUL
Sep-12	M210	Tobacco	HMRC	None	None	Yes	95+	14 cases under 18's deleted	EUL

Keeping in touch

To ensure timely response, use our generic email address for correspondence with us at collections@ukdataservice.ac.uk

Do let us know if the survey manager for a survey or series changes, and do brief the new manager on these depositing requirements.



Funding provided by



With support from



E collections@ukdataservice.ac.uk
T +44 (0)1206 872572
W ukdataservice.ac.uk/deposit

UK Data Service

