CATALOGUING PROCEDURES AND GUIDELINES

PUBLIC

7 JUNE 2016

T +44 (0)1206 872001
E sharonb@essex.ac.uk
www.data-archive.ac.uk

UK DATA ARCHIVE
UNIVERSITY OF ESSEX
WIVENHOE PARK
COLCHESTER
ESSEX, CO4 3SQ

WE ARE SUPPORTED BY THE **UNIVERSITY OF ESSEX**, THE **ECONOMIC AND SOCIAL RESEARCH COUNCIL**, AND THE **JOINT INFORMATION SYSTEMS COMMITTEE**

Contents

Scope

Each data collection (a.k.a. 'study') contained within the UK Data Archive's Discover catalogue has an individual catalogue record. The aim of the catalogue record is to provide users with enough information to enable them to decide if a particular study will be suitable for their requirements. This document outlines the procedures for the creation of catalogue records for the Discover catalogue. It does not cover the addition of index keywords from the Humanities and Social Science Electronic Thesaurus (HASSET) to the catalogue record, which is outlined in the *Indexing Procedures* document (CD215).

Roles and Responsibilities

The Ingest Services team use the catalogue input programs application to create new catalogue records, or make changes to an existing record. Members of the team are responsible for creating or amending catalogue records according to the policy laid down in this document. The Ingest Services team are also currently responsible for maintaining the Names Authority List associated with the catalogue input programs. The Thesaurus team is responsible for the addition of new index terms to the Humanities and Social Science Thesaurus (HASSET).

1. Standards used for ingest cataloguing and indexing

Cataloguing policy at the Archive is based on the rules laid down in the following international archival standards:

AACR2 – Anglo-American Cataloguing Rules, Second Edition, 2002 Revision, 2005 Update. AACR2 has been superseded by RDA (Resource Description and Access - http://www.rda-jsc.org/rda.html, adopted by the Library of Congress during 2013. At present, the catalogue input programs will continue to use AACR2 terminology.

ANSI/NISO Z39.50-2003, maintained by the Library of Congress:

http://www.loc.gov/z3950/agency/Z39-50-2003.pdf

American Psychological Association (APA) Style Guide: http://www.apastyle.org/index.aspx

Code of Practice for the Identification of E-books and Digital Content, BIC (Book Industry Communication)

DCMI – Dublin Core Metadata Initiative: http://dublincore.org/

DDI – Data Documentation Initiative: http://www.ddialliance.org/

Information and Documentation – Guidelines for bibliographic references and citations to information resources (BS ISO 690:2010)

Metadata Encoding and Transmission Standards (METS): http://www.loc.gov/standards/mets/

NCA (National Council on Archives) Rules for Construction of Personal, Place and Corporate Names, 1997: https://www.nationalarchives.gov.uk/documents/information-management/naming-rules.pdf

OAI – Open Archives Initiative: http://www.openarchives.org/

SSD – Standard Study Description Scheme, agreed by CESSDA members in the early 1970s

Structured Vocabularies for Information Retrieval (BS 8723-1 and BS 8723-2)

XML – Extensible Markup Language http://www.w3.org/XML/

2. Catalogue records policy

Each data collection accessioned to the UK Data Archive requires a catalogue record to be created as part of the ingest processing procedure.

3. Catalogue record numbering information

While

Individual study catalogue records

Consecutive 4-digit study numbers are automatically allocated by the catalogue input programs when a new record is created. Currently, standard individual study numbers have four digits, start at 0001 and currently run up to 7400 and over.

Generic group numbering

Generic numbers have five digits, and currently begin with '33'. Generic groups include one or more 'chick' individual studies, and generic catalogue records include information common to the survey over time. At present, they run from 33001 onwards, e.g. *SN 33004 National Child Development Study*. When a new generic is created, the catalogue input programs automatically allocate the next '33###' number.

Series numbering

The Discover catalogue also includes series records, used to group together data series which may include more than one generic group. For example, the *Labour Force Survey* (LFS) includes 11 generic groups, which cover different types of data compiled from the LFS data that are collected by the Office for National Statistics. Series records are identified by 6-digit series numbers beginning with '2' for example SN 2000026 *Labour Force Survey and* SN 2000002 *Annual Population Survey*.
Neither the series number or the short series description metadata are currently created within the catalogue input programs. Where a new series is required, or a new generic needs to be added to an existing series,

the User Support and ADM teams should be consulted.

ReShare records (formerly ESRC DataStore)
Catalogue records also exist in Discover for studies held within the ReShare data repository (formerly ESRC DataStore) rather than the formal collection. They are identified by a 6-digit number beginning with ‘8’, for example SN 800595 *Climate Science in Urban Design:a historical and comparative study of applied urban climatology*
Neither the ReShare 6-digit number or the catalogue record metadata are currently created within the catalogue input programs.

Older study numbers
The method used at the Archive to allocate numbers to catalogue records has changed over time. Other five-digit study numbers also exist within the catalogue for some very old individual studies (which in some cases may also be members of generics). In these cases, the numbering normally denotes the year of the study concerned, e.g. ‘53001’ was a survey conducted in 1953, 56001-3 were conducted in 1956, etc. This year numbering system runs right up to 1972, with 72001-2, after which the 4-digit consecutive system was adopted. The complete list of these ‘year’ numbers is as follows:

53001
56001-3
57001-4
58001-4
59001-4
60001-10
61002-7
62001-5
63002-8
64001-27
65001-31
66001-42
67001-36
68001-24
69001-40
70001-16
71001-7
72001-2

There also exist within the catalogue two very old individual studies numbered 98000 and 99001. The reason behind the format of these numbers is not clear.

4. Virtual catalogue records

A virtual catalogue record (VCR) may be created to provide a metadata record for a data collection that is held elsewhere than the UK Data Archive. Examples include Census data, international macro data held at Manchester, Eurobarometer, European Social Survey and some qualitative history data collections. Virtual catalogue records are only created for special purposes. Discover has the ability to include records created outside the catalogue input programs, for example series and ReShare records (see above), which was not possible with previous versions of the Archive catalogue. Therefore, the need for VCRs has reduced accordingly. On the rare occasions that a VCR is to be created, it should bebased on the originator’s/data collection holder’s information, and should retain the original spelling, even if this does not comply with the Archive’s normal style guide.

5. Useful reference sources for cataloguing

UK Charity Commission: http://www.charitycommission.gov.uk/

Environmental Information Data Centre: https://gateway.ceh.ac.uk/

ESRC Research Catalogue: http://www.esrc.ac.uk/impacts-and-findings/research-catalogue/

ESRC web portal: http://www.esrc.ac.uk/

Getty Thesaurus of Geographic Names Online:
http://www.getty.edu/research/tools/vocabularies/tgn/index.html

The National Archives Crown Bodies list:

http://www.nationalarchives.gov.uk/information-management/our-services/uk-crown-bodies.htm
Oxford English Dictionary (OED) through Albert Sloman Library (Essex login required):
http://serlib0.essex.ac.uk/record=e1000098~S5

Universities and Colleges Directory: http://www.universitiesuk.ac.uk

University of Essex style guidance: www.essex.ac.uk/marketing/guidelines/

6. General cataloguing guidelines

6.1. Punctuation

Please avoid using the ampersand sign (&) as it causes an XML parser error in the catalogue input programs, meaning the record cannot be published. The '&' sign is found most frequently in organisation names, URLs and bibliographic references (e.g. journal names), so please take particular care with these.

6.2. HTML

The UK Data Archive is working towards making the catalogue records W3C compliant. When entering free text or hyperlinks HTML tags should be used.

6.3. Generic/chick information

As noted in section 3 above, generic numbers (five digit numbers beginning 33xxx) are used to create an umbrella record to bring together studies that form a series or discrete entity. The information below provides some further detail about the construction of generic records.

It may be obvious from the outset that some studies should be part of a generic, for example, the *Labour Force Survey*, or the *International Passenger Survey*, because the studies are repeated at frequent intervals. For others, it might not be obvious when the first survey is booked in that it is going to form part of a generic, or that the Archive will acquire more of the series in the future. In this case, a generic record will be created at a later date when the second study arrives. This will mean that some information may need to be reallocated from the 'chick' (see below) to the generic record, and be redrafted to provide contextual information.

The individual studies that belong to a particular generic are referred to as 'chicks'. The generic record should be used to hold information that is common to ALL of the chicks. Both generic and individual chick information is displayed. Some elements may have information held in the generic that is relevant to all chicks, as well as additional specific information held in an individual chick. In a standard catalogue record, the generic information in each section is displayed first and the chick information second.

Where identical information is added to one section of both the generic and chick records by mistake, it will display in the catalogue record as a double entry, i.e. as repeated information. It is therefore important to check the display of the full catalogue record (with generic information included) after cataloguing to ensure that the whole entry makes sense when read as one and no elements are repeated in error.

The following elements within a generic catalogue record will *not* display in the catalogue record of any individual chick:

- Title
- Alternative Title
- Date of Deposit
- Availability
- Contact

- ID No (five-digit SN of generic)
- Distributor
- Date of Distribution
- Version
- Time Period
- Date of Fieldwork

If a field entry has previously been relevant to all chicks in a series, and therefore put into the generic, but is no longer applicable to the latest study in the series (i.e. a new chick), the entry should be removed from the generic and put into all relevant chick records instead. This may happen, for example, where a Principal Investigator or Depositor has changed, or the name of a government department has changed.

Generic records are a legacy feature of the Archive's cataloguing system. They were created for use at a time when the size of the tables being stored on servers was a consideration. This approach may be revised in the future. However, for maintenance it is much easier to amend, for example, a typographical error in the generic, or add extra information relevant to all studies in the series in one place, rather than amend many instances individually within a large series.

For guidelines on generic titles, see the Main Title section of this guide.

6.4. Use of note fields

Many sections have a coded list where one or more items can be selected. As well as this they also have a general note field for explanatory text or qualifying information. Certain fields have recommended standard text to use to record missing information – this is shown in the relevant sections as well as in the summary of standardised wording for use in the UK Data Archive catalogue record.

6.5. Standardised wording for use in the catalogue record

Certain sections in the catalogue record input form must be completed, but do not currently have standard categories for selection. In order to ensure standardisation of terminology within the catalogue, please use the terms quoted in the relevant sections of the document below for:

- Weighting
- Sampling
- Kind of Data
- Spatial Unit
- Copyright

In the case of major series studies, every effort should be made to verify the information required and add it to the catalogue record instead, of using the standardised wording such as **No information recorded**. If unsure how to proceed, please seek advice from the Data Curation Manager.

Catalogue input programs: guidance for specific sections

7. Input program sections currently in use

Study Number (SN)	Mandatory
Title	
Main Title	Mandatory
Sub-title/Note field	Optional (no longer routinely used but may be present in legacy catalogue records)

Alternative Title	Optional
Series Title	Optional
Project Number	Optional
Data Type	Optional (but must be completed)
Distribution	
Distributor	Mandatory
Availability	Mandatory
Contact	Mandatory
Date of Deposit	Mandatory
Date of Distribution	Mandatory
Subject Category	Mandatory
Depositor	Mandatory
Principal Investigator	Mandatory
Data Collector	Optional
Original Data Producer	Optional
Sponsor	Optional
Research Initiator	Optional
Grant Number	Optional
Other Acknowledgements	Optional
Abstract	Mandatory if nothing under Main Topics
Main Topics	Mandatory if nothing under Abstract
Coverage	
Time Period	Mandatory if nothing under Dates of Fieldwork
Dates of Fieldwork	Mandatory if nothing under Time Period
Country	Mandatory
Town/Village	Optional
Region/County	Optional
Spatial Unit	Mandatory
Observation Unit	Optional
Kind of Data	Optional
Universe sampled	
Locations of Units of Observation	Mandatory
Population	Mandatory
Methodology	
Time Dimensions	Mandatory
Sampling Procedure	Mandatory
Number of Units	Optional
Method of Data Collection	Mandatory
Weighting	Mandatory
Control Operations	Optional
Level of Processing	Mandatory
Data Sources	Optional
Language of Written Material	Mandatory
Related Studies	Optional
References and Publications by PI	Optional
References and Publications by Others	Optional
Administrative Notes	Optional
Access	
Access Conditions	Mandatory
Access Code	Mandatory
Recommendations	Optional
External Note	Optional
Originating Archive	Optional
Date of Release	
Date of Edition	Mandatory
Copyright	Mandatory
Digital Object Identifier (DOI)	Required for new study/new edition releases

8. Title

This section covers: Main Title; Sub-title/Note field; Alternative Title; Series Title; Project Number.

N.B. There is another title element within the DDI – *Parallel Title*. This is used when the title is translated into another language. The Archive does not currently use this element.

8.1. Main Title

This is a mandatory element. The title must be input in a uniform format. This is to ensure that each title is unique (an exception is *Sub-Title* – see below). When a title is input, the program will check to see if the title already exists in the catalogue. Entry of a duplicate title will result in an error message. Duplicate titles would cause confusion to external users and problems with internal administration. By keeping to consistent rules, it is easy to identify updates, new editions, re-deposits.

Take the title from the deposit and review forms, following the rules below. Any queries about a title should be discussed with the Data Curation Manager in the first instance.

8.1.1. General title rules

Titles should be in mixed case with all significant words having an upper case initial letter (title case). An exception to this rule is when words which would normally not be capitalised (prepositions, pronouns etc.), but these do need to be capitalised if they are a component of the project title or acronym, e.g.:

> "Changes **Around** Food Experience: Impact of Reduced Contact with Food on the Social Engagement and Well-Being of Older Women, 2007-2008"

The project title and acronym in this case is Changes Around Food Experience (CAFE)

Articles (The, A, An) at the beginning of the title should be omitted, e.g. "Quarterly Labour Force Survey" NOT "**The** Quarterly Labour Force Survey";

Do not include acronyms within the title. These should be entered into *Alternative Title* unless the acronym is held to be an essential part of the project/study title, e.g. "EU Kids Online: Enhancing Knowledge Regarding European Children's Use, Risk And Safety Online, 2010" where 'EU Kids Online' was the registered trademarked name of the project.

Do not include a personal or organisation name or acronym in the title unless necessary to avoid confusion with a similar title (or confusion between two different entities containing some similar elements (e.g. **NHS** North of Tyne - health authority, North of Tyne - geographic area).

8.1.2. Title elements

Sub-titles should be separated from the main title by the use of a colon and space, e.g.:

> Families and Social Mobility: a Comparative Study, 1900-1988 (note that articles are not omitted in the sub-title following the colon);

When adding a generic title, where possible, a series start date should be added. The title should also be in the singular if each chick has the same core title, even if there are a number of variants. For example, Farm Business Survey, 1982- (which has the same title for each chick), or General Household Survey, 1971- (which also includes chicks with the qualifier "Surrey Files"). If there are a range of different titles within a generic, the title should be plural, e.g. British Election Studies, 1969-.

8.1.3. Title hyphenation

If the title includes hyphenated compounds, please check the Oxford English Dictionary (OED) Online whether the compound necessitates a hyphen before applying the hyphenation rules. See http://serlib0.essex.ac.uk/record=e1000098~S5 (available to University of Essex users only).

Always capitalise the **first element** of any compound in a title. Capitalise any subsequent elements unless they are articles, prepositions or co-ordinating conjunctions (and, but, for, or, nor) unless the article forms an 'active' part of a project acronym (see the 'Changes Around Food Experience (CAFE)' example above, though that title is not hyphenated

Non-capitalised examples include:

1970 British Cohort Study: Ten-Year **Follow-up**, 1980

1970 British Cohort Study: Thirty-Four-Year **Follow-up**, 2004-2005

Subsequent elements that should be capitalised include nouns, adjectives, verbs, adverbs etc.:

National Child Development Study: **37-Year** Sample Survey, 1995

Scottish **School-Leavers** Survey: Cohort Three, Sweeps One to Four, 1999-2006

Survey of **Self-Funded** Admissions to Care Homes, 1999-2000

DE/PSI National Survey of **Long-Term** Unemployment, 1980-1981

Some titles include three-part hyphenated compounds which require a mixture of the two rules, e.g.:

British Election Study, 2005: **Face-to-Face** Survey

Adult **Off-the-Job** Skills Training, 1986 Sample

If the first element of the compound is a prefix or combining form that could not stand by itself as a word ('anti', 'pre', etc.), do not capitalise the second element unless it is a proper noun (i.e. a noun that is always written with an initial capital letter, for example a person's name or a place name) or a proper adjective (for example, a word denoting a nation or ethnic group, such as Japanese or Aboriginal):

Non-industrial Civil Service Staff Statistics: British Isles, 1939-1999

Trans-Atlantic Slave Trade: a Revised and Enlarged Database, 1500-1867

In the case of hyphenated written numbers with two elements, do not capitalise the second element, e.g.:

Twenty-first Birthday

However, if there are additional hyphenated elements following a written number, and the rules demand the capitalisation of these elements, then the written number should also be capitalised:

1970 British Cohort Study: **Twenty-Six-Year** Follow-up, 1996

1970 British Cohort Study: **Twenty-Nine-Year** Follow-up, 1999-2000

8.1.4. Title dates

Where a date or date range should be given at the end of the title, to give a clear indication of the study's time coverage. If the depositor's specified study title in the deposit form contains the date within the title, this should be moved to appear at the end, unless by doing so the sense of the title is lost. Unless it is absolutely unavoidable, the title should not start with the date, as this may affect catalogue search indexes.

If there is no date given in the data collection deposit form then the date should be the date of coverage of the data, or the date of fieldwork. For quantitative data collections they are often the same, but historical or qualitative collections may be different. In case of query, seek advice from the Data Services Manager.

The format of the date should follow consistent rules, as explained below.

Single year
Title, comma, space, year, e.g. Survey of Mortgage Lenders, 1996

Single month within single year
Title, comma, space, month, space, year, e.g. ONS Omnibus Survey, June 2004

Spanning months within single year

Title, comma, space, start month, space, hyphen, space, end month, comma, space, year, e.g.

Quarterly Labour Force Survey, March - May, 1998

Double months within a single year when there is an 'and' referring to the module months

Title, comma, space, month, space, and, space, month, comma, space, year, e.g.:

ONS Omnibus Survey, Smoking Module, October and November, 2001

(N.B This is mostly the case for ONS Omnibus/Opinions surveys, in order to enable searching for particular months.)

Multiple months within a single year with an 'and' referring to the module months

Title, comma, space, month, comma, space, month, comma, space, month, space, and, space, month, comma, space, year, e.g.

ONS Omnibus Survey, Attendance and Participation in the Arts Module, July, September, October and November, 2001

ONS Omnibus Survey, Changes in Family Circumstances Module, March, April, June, July, August and September, 2002

Spanning months covering two years
Title, comma, space, start month, space, year, space, hyphen, space, end month, space, year, e.g. Quarterly Labour Force Survey, December 1993 - February 1994

Multiple months in multiple years when there is an 'and' referring to the module months

Title, comma, space, month, space, year, space, and, space, month, space, and, space, month, space, year, e.g.:

ONS Opinions Survey, Drinking Module, April 2008 and April and May 2009

ONS Omnibus Survey, Tobacco Consumption Module, October, November and December 2005 and January and February 2006

ONS Omnibus Survey, Tobacco Consumption Module, September, October, November and December 2006 and January, February, March, April, June and July 2007

ONS Opinions Survey, Non-Resident Parental Contact Module, September, October, November and December 2007 and January and February 2008: Special Licence Access

Span of years

Title, comma, space, start year, hyphen, end year, e.g. Survey of English Housing, 1997-1998

Spans of several years

Title, comma, space, start year, slash, end year, hyphen, start year, slash, end year, e.g. Universities' Statistical Record, 1972/73-1993/94: First Destination

Multiple, non-consecutive years/spans

If two years: Title, comma, space, year, space, and, space, year, e.g.

> DSS/PSI Programme of Research into Low Income Families, 1996 and 1998

If more than two years: Title, comma, space, year, comma, space, year, space, and, space, year, e.g.

> British Election Panel Study, 1983, 1986 and 1987

If two spans: Title, comma, space, year, hyphen, year, space, and, space, year, hyphen, year, e.g.

> Great Britain Historical Database: Health and Health Care Data: Mortality Statistics, 1851-1973 and 1981-1992

If a mixture: Title, comma, space, year, space, and, space, year, hyphen, year, e.g.

> Dissension in the House of Commons, 1924 and 1929-1931

Non-consecutive months
Title, comma, space, month, space, and, space, month, space, hyphen, space, month, comma, space, year, e.g.

> ONS Omnibus Survey, Attendance and Participation in the Arts Module, July and September - November, 2001

The only exceptions to the above are if the data collection comes from another archive and therefore already has a 'published' title. In this case use the existing title even if this deviates from UK Data Archive rules, e.g.

> World Values Surveys and European Values Surveys, 1981-1984, 1990-1993, and 1995-1997

(in the above example a comma would not normally be inserted before the 'and'.)

Approximate dates

Use the abbreviation c. for *circa*, e.g. Crop Yields and Animal Carcass Weights in England, c.1700-1914

8.2. Sub-title/Note field

This field was previously used as a note field, but is referred to as 'Sub-title' in the DDI. **It is no longer in use at the Archive**. Sub-titles are not currently added to new catalogue records. All the information required should be entered into the Main Title, unless an Alternative Title or a Series Title exists (see below).

However, if a sub-title should be required, please add it to the end of the main title after a colon, as for the following legacy example:

> Universities' Statistical Record, 1972/73-1993/94: First Destination

Sub-titles currently in legacy catalogue records should however remain in place for the moment.

The sub-title element was an optional field, which was used to indicate the specific contents of one particular data collection where the title was common to more than one collection (e.g. the Universities' Statistical Record example above, or where a note to explain the file content was felt necessary, for example an indication of the subject content or geographic coverage. However, because sub-titles are not currently show in the results when catalogue searches are conducted, the decision was taken to avoid using sub-titles in new catalogue records and instead to indicate the particular subject or geographical information in the main title.

8.3. Alternative Title

This is an optional field. It is used when a study is also known by another title than the main one. The field is normally used for series acronyms or where project titles differ from the deposited data collection title.

Example of use of alternative titles:

Scottish Social Attitudes Survey, 1999
Alternative Title: Scottish Parliamentary Election Study, 1999

Where the alternative title is an acronym:

Social Change and Economic Life Initiative Surveys, 1986-1987
Alternative Title: SCELI

One or more alternative titles may be used, separated by semi-colons if required. This usually occurs where there is both a known alternative project title and an acronym, e.g.:

Adult Psychiatric Morbidity Survey, 2007
Alternative titles: APMS; Adult Psychiatric Morbidity in England, 2007

National Survey of Third Sector Organisations, 2008
Alternative titles: NSTSO; State of the Sector Panel Survey

An alternative title may also be used where a series name has changed over time, e.g.:

Eurobarometer Survey Series, 1974-
Alternative Title: Euro-Barometer Survey Series

(In this case the entry is in the generic. It allows for retrieval of all Eurobarometer studies regardless of whether a search is done on the old or new title.)

The following examples shows the use of both an Alternative Title and a Series Title (see below):

1970 British Cohort Study: Birth and 22-Month Subsample, 1970-1972
Alternative Title: 1970 Birth Cohort: Birth and 22-Month Subsample, 1970-1972; British Births
Series: (1970 British Cohort Study)

Studies accessioned from the Rural Economy and Land Use (RELU) programme

The RELU programme, funded by the ESRC and the Natural Environment Research Council (NERC) ran from 2004-2013. The programme funded interdisciplinary research covering social, economic and environmental consequences of environmental change in order to inform future policy and practice with choices on how to manage the countryside and rural economies. Several studies from the RELU programme were acquired for the Archive's collection, with some implications for cataloguing procedures. With regard to study titles, iff the project title as used in the RELU Knowledge Portal at http://relu.data-archive.ac.uk/datasets.asp is significantly different from the Main Title (except for the date), then use Knowledge Portal project title as an alternative title. The RELU data collection's processing plan may specify a preferred title.

Alternative title display
Alternative titles in the generic do not display in the chick catalogue records, so as a rule alternative titles will be added to the chicks. However, there may be exceptions in the case of very large existing series. Also, in some circumstances, considerable legacy work would be required to remove an alternative title from an existing generic and put it in the chicks. These cases would be considered on an individual basis by the Data Curation Manager before any legacy work is undertaken.

8.4. Series Title

This is an optional field. It is no longer used for new data collections, unless they are already part of a generic group with an existing series title, where it should be used for consistency. For data series where the titles for individual studies differ, or where there are several generics within a series, a series title was used to pull them together and facilitate retrieval. This function is now done by the major studies subject category (see below). The only Series Title options that are currently available may be selected from a drop-down list of existing series titles. Where it exists, the series title is displayed as part of the catalogue record and in the HTML citation file, whereas the generic title is not.

Examples of series with an existing series title include the *National Child Development Study* and the *Labour Force Survey*.

8.5. Project Number

This is an optional text field and is no longer frequently used. **Do not confuse Project Number with Grant Number**.

Legacy studies may include a project number with separation of letters and numbers by a semi-colon. This was automatically generated by a previous UNIX-based version of the catalogue input programs. Examples include “MORI ; 201” and “P ; 126”.

Start here

8.6. Data Type

For the UKDS Discover catalogue, released with the UKDS website in April 2013, a ‘Key Data’ search facet was created - see http://ukdataservice.ac.uk/get-data/key-data.aspx

To create the facet, a ‘Data Type’ tag was allocated to each study in the catalogue during the development phase of Discover. Now that Discover is live, one or more Data Type tags must be added to each new study catalogue record from the following tick box list:

- Administrative data
- Business microdata
- Census data
- Cohort and longitudinal studies
- Cross-national survey data
- Experimental data
- Geospatial data
- Historical data
- International macrodata
- Other surveys
- Qualitative and mixed methods data
- Teaching data
- Time series
- UK Survey data

Normally, the Data Type is assigned at pre-processing check stage when the study number is allocated. However, once processing is complete and final cataloguing takes place, extra Data Types may be added. Where the study concerned is part of a generic, Data Type tags should be added to individual chick records rather than the generic record. This provides flexibility where the generic may contain different datasets, e.g. where a time series dataset is included as well as standard survey data (such as the General Household Survey (GHS)) or a qualitative study alongside survey data (such as the National Child Development Study (NCDS)).

Retrospective Data Type tagging work was done automatically for the Discover catalogue and is not generally undertaken by the Ingest Services team; tags are only added for new studies, from SN 7260 onwards. However, if during routine legacy cataloguing work (e.g. a new edition) it is apparent that another Data Type tag should be added or an allocated Data Type is wrong, additions/replacements should be made. Should the need for significant re-tagging arise (for example the addition of a new Data Type, ADM should be consulted so that the work can be programmed.

9. Distribution

This is a mandatory field. This section covers:

- Distributor (must be in each individual chick record, not the generic)
- Availability (a mandatory field. It is automatically generated from the Distributor code)

- Contact
- Date of Deposit (a mandatory field, automatically generated when a new study number is generated – internal display only).
- Date of Distribution. (a mandatory field, automatically generated when the record is released. It is displayed under *Date of Release.* Any subsequent edition dates are automatically generated when the new edition box is checked, which should be done immediately prior to re-release in order to reflect the correct edition date.)

In the Discover catalogue record these sections appear at the bottom of the record, in the *Administrative and Access Information* section. The *Distributor* automatically generates the *Availability* and the *Contact* information. Where the record is a Virtual Catalogue Record (VCR) for data not held at the Archive, there is a tick box to select and the option to add free text information about availability, for example an address or a URL.

10. Subject Category

This is a mandatory field. Subject categories are used:

- to define the overall subject content of the data at study level (whereas keywords are assigned at question/variable level);
- to identify and retrieve data that cover a particular subject area
- to link studies related to the same series, but not necessarily in the same generic (Series Title may previously have done this). This includes for example, derived data collections, such as teaching datasets created from a larger study by researchers who are not part of the original study team.
- for internal performance figures and collection management, e.g. to identify how many studies have been acquired over the year for particular subject areas.
- to identify useful studies for a particular aim, e.g. subject theme workshop.

One or more subject categories may be selected from a controlled list.

ONS Opinions and Lifestyle
For ONS Opinions and Lifestyle data collections, an MS Excel file exists which contains most, but not all, subject category levels that have been assigned to Opinions and Lifestyle modules in the past. This file allows the cataloguer to use subject categories consistently and saves searching for/identifying which subject categories have been previously assigned to modules. If a new Opinions and Lifestyle module that has not been asked in the past has been deposited, the subject category information assigned to that module should be recorded in the Opinions and Lifestyle subject category Excel file for future reference. Please consult the Senior Data Curation Officer for further information.

10.1. New subject categories

Occasionally, a new subject category may be required, for example if a new major data series is acquired, or new data collections cover a substantially new subject field. It should be noted that the subject categories correspond to the CESSDA Topic Classifications. New subject categories are only added when the need arises, i.e. when the Archive actually receives studies that cover the new subject (this ensures that users do not search on a newly-added subject category which yields no search results).
New subject categories should be proposed only when studies are received that cover concepts/subjects that are not covered in the existing list of subject categories. If HASSET keywords exist that cover the concept, the proposed subject categories should match these. Discussions regarding new categories should involve the Thesaurus team and the Data Curation Manager or another member of the Ingest Services team, to ensure that the balance of the collection is maintained.
Any new subject category added will requires legacy cataloguing work. Unlike HASSET keywords, subject categories currently have no usage scope notes detailing date of introduction as an index term, so all

relevant existing studies will also need to be re-assigned to the new subject category. The amount of legacy work required and resources needed should be a consideration in the decision to create a new subject category.

CESSDA Topic Classifications and subject categories

If the decision is made to add the new subject category, the Thesaurus team will check whether this subject category is included in the current CESSDA Topic Classifications. If it is not included there, it should be considered whether the new subject category fits under one of the existing subject categories as a sub-category, or whether it needs to be a new standalone category. All Archive subject categories are mapped to the CESSDA subject categories, which are in use in the CESSDA data portal (depending on future developments).

Requesting the subject category

When discussions are complete, a JIRA Helpdesk request under the ‘Ingest Programs' project should be made, detailing the exact wording of the new subject category and whether it is a new standalone category or a subcategory under an existing one. The request is thus automatically assigned to the lead developer for the project. If the new subject category is not currently included in the CESSDA subject categories list, a request should also be added to the 'Subject Categories' component of the 'CESSDA Portal' project in JIRA, requesting that the new term is mapped to CESSDA subject categories .

Mapping to HASSET

All existing subject categories have also been mapped to HASSET. The Thesaurus team will check that any new subject category is also included in HASSET as a keyword index term. If not, a new keyword must be added to HASSET, in consultation with the Thesaurus Team.

11. Names section (Depositor, Principal Investigator, Sponsor, etc.)

Names information

All names associated with the creation of the data (with the exception of *Other Acknowledgements)* are drawn from a controlled vocabulary Names Authority List Each time a new name is created, it is added to the relevant controlled list in a standard format. All ‘names' sections in the catalogue record (including *Depositor*) may have multiple entries.

Name entries will be in the form of either an organisation, or an individual's name followed by his/her affiliated organisation. In rare cases an individual may not have an affiliated organisation; since there is no scope in the catalogue input programs for this, a request should be made for the entry to be added by ADM. However, please consult the Data Curation Manager before doing this.

In the catalogue input program Names Authority Lists, existing names may have a certain level of colour-coding to denote status. ‘Current' names have no background colour. A name with pink background colour is a non-preferred version of the name. A name with yellow or beige background colour means an old (out of date) version of the name, for example where the person concerned has moved to another institution and their affiliation has changed, or the name of an organisation has changed (although this may be used if an old study has been deposited and the older organisation name is chronologically correct for that study). Queries should be referred to the Data Curation Manager.

Organisation names - formatting

Government departments in countries other than the UK should have the country name included in the first level (organisation name) and separated from the department name by a full stop, e.g.

United States. Department of Commerce

For guidance on the official name of a country, please use the Getty Thesaurus of Geographic Names Online www.getty.edu/research/conducting_research/vocabularies/tgn/

For government departments of the devolved governments of Scotland, Wales and Northern Ireland, add the location as a qualifier in brackets after the department name (unless the location is already part of the department name), e.g.

Department of Culture, Arts and Leisure (Northern Ireland)

Local authorities in the UK should include the location as part of the first level but without a full stop, e.g.:

Essex County Council

Other organisations outside the UK should have the organisation name as the first level. If clarification is needed then the country or city name should be added in brackets after the name, e.g.

Cornell University (Ithaca)

Attwood Statistics N. V. (Rotterdam)

Emory University (Atlanta, Georgia)

Saar Poll Limited (Tallinn, Estonia)

The clarification in brackets should not be added when the organisation name is unambiguous, because the location is included in the name, e.g.

New York University

The UK Data Archive follows the *Anglo-American Cataloguing Rules, 2nd edn*. (AACR2) for the creation of the names authority lists, apart from names in foreign languages where National Council on Archives (NCA) *Rules for the Construction of Personal, Place and Corporate Names* are followed – see below).

For non-English names, the NCA 2.3C *Special rules for names in foreign languages* are followed (these are based on AACR2 rules, but may be easier to use).

Depositors may use abbreviated or even incorrect names for their organisations/departments so it is important to verify the information if there is any doubt. The name may already exist in the Names Authority List, so always check first before making a new entry. When creating a new entry, the AACR2 or NCA rules quoted above should be followed.

NCA rules: http://anws.llgc.org.uk/ncarules/title.htm

(Note: AACR2 has been superseded by RDA (Resource Description and Access - http://www.rda-jsc.org/rda.html, which will be adopted by the U.S. Library of Congress during 2013. At present, the catalogue input programs will continue to use AACR2 terminology.)

Changes over time

Be aware that the name of an organisation, especially a government department, may change over time. Care should be taken that the name is chronologically correct for the date that the study took place, and a new entry should be created to accommodate this if possible. It is possible that an organisation name may change between the time the study is conducted and its subsequent deposit. In that case, the Depositor name should be correct for the time of deposit (see *Depositor* section) even if it then differs from the Principal Investigator name – a note may be added to the *Other Acknowledgements* section for clarification.

The main things to remember in this process are:

- verify information;
- check existing entries;
- follow rules for new entries – if in doubt consult the Data Curation Manager
- if a department/organisation name changes, DO NOT amend the entry – create a new entry and request the addition of a cross-reference (please consult the Data Services Manager).

If the names of departments/organisations have changed over time for a long-running series, it may be useful to add a brief history under ‘Other Acknowledgements’. For example, in the Quarterly Labour Force Survey generic, the following text appears:

> *Other Acknowledgements*:
> Until 5 July 1995, the Labour Force Survey (LFS) was the responsibility of the Employment Department. On that date, the Employment Department was abolished and responsibility for the survey was passed to the Central Statistical Office (CSO). In April 1996, the Central Statistical Office and the Office of Population Censuses and Surveys amalgamated to become the Office for National Statistics (ONS).

Names in foreign languages

When recording personal and/or organisation names in foreign languages, please do not use any special characters apart from Latin characters, as they may not appear properly in our systems.

If an organisation name is in any language other than English, and the depositor has specified the local language name of the organisation, please use this local language name as the main entry. If the organisation also has an official name in English, the English name should be added to the Names Authority List as a synonym.

If an organisation name is in German, and uses hyphens between words, the name should be without the hyphens (which may hamper search retrieval).

11.1. Depositor

This is a mandatory field. The *Depositor* is the person or organisation that has deposited the study. Usually, but not always, the depositor is the signatory on the Licence Form.

Affiliation details for the *Depositor* should reflect the affiliation of the depositor **at the time the data collection was deposited**.

11.2. Principal Investigator

The *Principal Investigator* (PI) is the lead researcher for the study, usually the grant holder. Several names may be listed on the deposit form where the whole team is credited; the names should be added in the order given on the deposit form.

Organisation name/affiliation details for the *Principal Investigator* should reflect their affiliation at the time the study was conducted. If the *Depositor* (see section 10.1 above) and *Principal Investigator* are the same person/organisation, but the name (or affiliation for an academic depositor) has changed since the study was conducted meaning that the new arrangements apply at the time of subsequent deposit, the chronological rule should still be followed. Explanatory notes may be added to the *Other Acknowledgments* section.

11.3. Data Collector

This is an optional field. The *Data Collector* is the person(s)/organisation(s) who has collected the data. This may be a survey company working on behalf of the PI or Sponsor of the research.

11.4. Original Producer

This is an optional field. For studies based on data previously collected (e.g. derived data collections based on government data, census data, etc.), Original Producer is used to record who was responsible for the original data on which the later study is based.

11.5. Research Initiator

This is an optional field. *Research Initiator* is no longer used. If necessary, a note may be added to the *Other*

Acknowledgements field.

11.6. Sponsor

This is a mandatory field. The *Sponsor* is the person/organisation(s) who has provided funding for the study.

12. Grant Number

This is an optional field. *Grant Number* may be listed on the data review forms, deposit forms, or in the documentation. In the case of Economic and Social Research Council (ESRC)-funded studies, it may also be obtained from the online ESRC Research Catalogue.

12.1. ESRC award grant numbers

ESRC award numbers **must** be formatted in a specifi way, as the ESRC use an automatic harvester to link their award pages with Archive catalogue records. The correct presence/absence of hyphens and spaces is essential for the harvester, especially as the grant number field in the catalogue input programs is necessarily a free text field, since larger studies may have more than one sponsor (see below). The most common formats for an ESRC award number are as follows:

RES-123-45-6789

(Occasionally '-A' is included at the end of the number, e.g. RES-123-45-6789-A. This is allocated if the Principal Investigator changes or moves institution midway through a project.)

Where a hyperlink is provided in the processing notes to the ESRC award page, the link must be checked to ensure it works, and the ESRC grant number should then be formatted exactly according to the link, irrespective of how it may have been written on forms or in the documentation. For example, if the link to ESRC award page is http://www.esrc.ac.uk/my-esrc/grants/RES-062-23-2997/read , the correct format for the grant number will be RES-062-23-2997 .

If there is no link given in the processing notes, the online ESRC Research Catalogue at http://www.esrc.ac.uk/impacts-and-findings/research-catalogue/ must be searched to ensure that the number is entered in the correct format.

12.2. Multiple grant numbers

The grant number does not display adjacent to its sponsor organisation, but may be added in the grant number text field. However, if one of the sponsoring organisations is the ESRC, and multiple grant numbers are provided, **only** the ESRC number should be entered in the *Grant Number* field; otherwise, the ESRC's automatic grant number harvesting program may fail to retrieve the catalogue record. Other grant numbers may be quoted in the *Other Acknowledgements* section.

Examples:

Sponsor(s):
Joseph Rowntree Foundation
Economic and Social Research Council

Grant Number:
RES-062-23-2997

13. Other Acknowledgements

This is an optional, free-text field. It may be used for the names of persons or organisations which should be acknowledged as having had some input into the study but which do not fit into any of the other specified categories; or where it would not be sensible to add them to the names authority list, e.g. 'All District

Councils in Great Britain provided data on…..'.

It can also be used for notes to qualify or further explain any of the other 'Names' entries, as, for example, when the Depositor and Principal Investigator are the same person, but their affiliation has changed since the completion of the study. See also guidance in the *Depositor* section.

14. Abstract

This is a mandatory field if the *Main Topics* field is empty. It is a text field. The *Abstract* covers the general aims/purpose and background of the study. In general this is taken from the data collection deposit form with supplementary information taken from the accompanying documentation/reports if necessary.

It should be remembered that the abstract will generally be read on screen, therefore should be succinct; users may be referred to the documentation for more detail.

If the study is part of a generic group it is important to add some contextual or background information to describe the series. Information that is common to all studies in the generic should be added to the generic Abstract. Wave-specific information should be added to the individual chick Abstract. (In some cases, such as the *National Child Development Study* and the *1970 British Cohort Study*, a number of diverse studies are included in the group, and the generic simply forms a framework to hold the group together for searching purposes; hence the bulk of the information is held in the chicks.)

HTML tags may be used in the Abstract section: please consult the *Basic HTML Editing Reference* document for advice on tag formats.

It may also be useful to provide links to project websites and funders' award webpages (see the ESRC information under Sponsor above). These are usually added at the end of the Abstract, such as

> "Further information may be found on the ESRC's The Impact of Retail Pricing on Overeating and Food Waste award webpage."

14.1. Specific Abstract text: Qualidata and RELU

Some studies, depending on their nature, have specific text included within the Abstract:

Qualidata:
The first sentence of the Abstract should be:

"This is a qualitative data collection." or "This is a mixed-methods data collection." as appropriate.

Example:

> "
>
> "This is a qualitative data collection. The experience of terrorist attacks since 9/11 has led to an increased interest in governmental responses to unconventional political violence. To date however, very little research has been conducted on how ordinary people think about such measures, and particularly the ways in which specific anti-terrorist measures impact on people's perceptions and experiences of safety and security."

RELU:
The Abstracts for RELU studies should begin with the sentence:

> "This study is part of the Rural Economy and Land Use (RELU) programme."

The Abstract should end with directions and a link to the relevant RELU Knowledge Portal page:

“Further documentation and publications are available via the RELU Knowledge Portal’ and the relevant HTML tag and title added: 'RELU Knowledge Portal: brief study title”

Example:

> Further documentation for this study may be found through the RELU Knowledge Portal: Eating Biodiversity

If some RELU data from the study are held at the Centre for Ecology and Hydrology (CEH), the following text should also be added:

> “Ecological data of the study are archived at the Environmental Information Data Centre’, with an HTML link to the CEH Gateway at https://gateway.ceh.ac.uk/ ”

Example:

> Ecological data from this study are available at the Environmental Information Data Centre of the Centre for Ecology and Hydrology.

15. Main Topics

This is a mandatory field if there is nothing in the *Abstract* field. It is used to indicate the subject/topics covered in the data, and may simply be a list of keywords if the *Abstract* is comprehensive.

The UK Data Archive web catalogue record displays Main Topics separately from the Abstract, but in the DDI record they are the same field. It is mandatory within the catalogue input programs to have either Abstract or Main Topics fields populated.

16. Coverage

16.1. Time Period/Dates of Fieldwork

Time Period describes the time period that the data cover. *Dates of Fieldwork* refers to the time period of the **collection** of the data. At least one of the two fields must be populated.

Sometimes only the *Time Period* element will be used, for example with respect to aggregate data which are being updated at regular periods. Very often quantitative data will have the same *Time Period* as the *Dates of Fieldwork*, in which case only the *Dates of Fieldwork* need be entered. Qualitative data collections may have a different *Time Period* to *Dates of Fieldwork,* for example where life history interviews have been conducted. If there is no end date for fieldwork, enter a start date only. If the dates cannot be entered in terms of year, year/month, year/month/day, the note field should be used instead.

For several waves/sweeps of a particular survey, enter the date of the first wave as the start date of fieldwork and the date of the last wave as the end date, then use the note field to give details.

16.2. Country

This is a mandatory field. If more than one country is covered by the data, multiple entries are possible. Countries are chosen from a controlled list.

The geography authority lists are not exhaustive; additions are made as necessary, dictated by the geographic coverage of the data holdings within the Archive's collection. When a new addition needs to be made, the correct name and spelling should be verified, using the census place names or a gazetteer. The Getty Thesaurus of Geographic Names Online (TGN) at http://www.getty.edu/research/tools/vocabularies/tgn/index.html is the standard source currently in use at the Archive. .

If the new addition is a town name, it may also be necessary to verify its wider county/region and add this to the *Region/County* geographic list too. For historical studies the *Region/County* should reflect where the town/village was situated at the time of the study. If a geographical term is added, please check if an equivalent term exists in the Archive's Humanities and Social Sciences Electronic Thesaurus (HASSET) (this can be done via the catalogue input programs) and if so, the study should be indexed with the lowest-level geographic term applicable. If the term does not exist in HASSET, please request it through JIRA so that the Thesaurus team may add it.

If there is no geographical coverage for the data e.g. where the files comprise a computer program, a bibliography, etc. then the standard wording **not applicable** must be used.

If the data cover more than one country then **multi-nation** should be selected from the controlled list. It is preferable for all countries to be listed too – except where the data are world coverage or there are too many countries to list, in which case **multi-nation** will suffice. (For International Macro data, the full list of countries supplied by Manchester should be pasted into the *Population* field.)

16.3. Spatial Unit

This is a mandatory field, and every effort should be made to populate it.

The spatial unit used should be recorded in the deposit form completed by the data depositor, but if not, please look at the data and documentation files to assess what the correct spatial units may be. Suitable units may be selected from a list of tick-box options, compiled from a set of ONS geographies. If the unit required is not contained in the list, tick the box marked 'Historical/Other geography (please specify)' and copy the appropriate entry in from the latest list of UGeo-compliant spatial units available in Excel format at **\\daedalus\brm$\PFTT\Projects\p0109-ResourceDiscoveryPilot\Project Activity\Metadata\SpatialUnits.** When the catalogue input programs are updated during the NextGen project, the list will be included as automatic choices.

What if the spatial unit is not shown in either list?
If the spatial unit is not available in either the tick-box options or the list, please type the name of the spatial unit concerned in the data into the text box, then email details of the new term with the associated study number it has been used for, to the Thesaurus team for addition to the list.

Standard text

- if no spatial unit has been used, add **No spatial unit** to the 'Historical/Other geography (please specify)' text box
- where no clear information exists about spatial units, add **No information recorded**

16.4. Observation Unit

This is an optional field. Select one or more from the controlled list displayed: Individuals; Families/households; Groups; Institutions/organisations; Administrative units (geographical/political); Text units (documents/chapters/words). There is also a note field for a free-text entry.

16.5. Kind of Data

This is an optional field. Select one or more from the controlled list displayed:

Textual data; Numeric data; Alpha/numeric data; Image; Sound;

and then one or more of the subsequent options:

Individual (micro) level; Aggregate (macro) level.

Most quantitative studies will be covered by the choices available in the existing list (see above) provided in the input programs.

For qualitative data collections, please select the appropriate terms from the list below and add them to the free-text note field:

In-depth/unstructured interview transcripts
Semi-structured interview transcripts
Structured interview questionnaires
Focus group transcripts
Interview notes
Interview summaries
Interview extracts
Unstructured/semi-structured diaries
Observation field notes
Kinship diagrams
Press clippings
Photographs
Video-taped interviews
Audio-taped interviews
Minutes of meetings
Case study notes
Naturally occurring speech/conversation transcripts
Correspondence

The following definitions, provided by Qualidata, should help with choosing the most appropriate term(s) from the above list:

In-depth/unstructured interview transcripts

- Usually no questionnaire, perhaps only a list of suggested topics
- Often only a few (2-10) topics suggested, not detailed list of 20+ questions
- Sometimes several different ways of broaching a topic will be suggested
- Interviewer may respond in depth, may offer reciprocal disclosures
- Interviewees have the freedom to tell their biographical stories in their own way, although there may be some gentle guidance offered by the interviewer in order to keep the narrative going

Semi-structured interview transcripts

- Semi-structured interviewing is more flexible than standardised methods such as the structured interview or survey
- Although the interviewer in this technique will have some established general topics for investigation, this method allows for the exploration of emergent themes and ideas rather than relying only on concepts and questions defined in advance of the interview
- The interviewer would usually use a standardised interview schedule with set questions which will be asked of all respondents. The questions tend to be asked in a similar order and format in order to make a form of comparison between answers possible
- There is scope for pursuing and probing for novel, relevant information, through additional questions

often noted as prompts on the schedule. The interviewer frequently has to formulate impromptu questions in order to follow up leads that emerge during the interview

- Usually the interviewer's role is engaged and encouraging but not personally involved. The interviewer facilitates the interviewees to talk about their views and experiences in depth but with limited reciprocal engagement or disclosure

Structured interview questionnaires

- Structured interviewing involves asking each interviewee the same set of standardised questions
- The order of questioning is fixed and wording is usually specific; there is little scope for probing or deviating from the specified agenda

Focus group transcripts

- Multiple respondents with one interview
- Groups can vary; can be picked for differing viewpoints, or for key similarities

Interview notes

- Notes made by the researcher/interviewer about the interview
- Might include descriptions of the person, behaviour, setting
- Might include interviewer’s initial reactions to the interview, research ideas that were stimulated, next steps to pursue

Interview summaries

- Usually a somewhat more formal abstract of the interview, intended to capture its key ideas, themes and topics

Interview extracts

- Short quotations from an interview
- Not intended to act as an overall summary or capture key themes
- Reflects a formal interview situation
- Usually prepared in preference to a full transcription or full audio recording

Unstructured/semi-structured diaries

- There are a number of diary types used
- Research diaries – logs or work notes kept by the researcher on progress of the project,
- Time diaries – a kind of qualitative data where people record what they were doing at fixed increments of time for a given period (this method has structured elements)

Observation field notes

- Notes researchers collect during fieldwork
- Most often done when research is based in natural settings (home, workplace, etc), not the “artificial” setting of a formal interview
- Typical of anthropology and ethnographic methods

Kinship diagrams

- Any diagrams or maps of how people are related by blood, marriage, or other ties

Press clippings

- From published print sources

Case study notes

- There may be several cases being compared within a project
- Cases can be any kind of unit: individual, family, organisation, setting, or any other category deemed relevant by the researcher

Naturally occurring speech/conversation transcripts

- Naturally occurring usually used to distinguish spontaneous conversation from any conversation directed or led by the researcher. Interviews and focus groups are NOT naturally occurring
- Often, these transcripts are analysed for their detailed structure (turn-taking, pauses, interruptions, etc.) and this requires very detailed form of transcribing called Jeffersonian.

Correspondence

- Letters, and probably now email and other communications, explicitly directed to one or more recipients.

17. Universe Sampled

17.1. Location of Units of Observation

This is a mandatory field. Select one or more of: Cross-national, National, Subnational:

- If the data cover an entire country (or levels below country are not defined) choose **National**.
- If the data cover part of a country (e.g. one or more towns, or counties) choose **Subnational**.
- If the data cover more than one country choose **Cross-national**. If Cross-national is selected, either National and/or Subnational should also be added, depending on whether the data cover the whole of the individual countries or parts thereof, as above.

17.2. Population

This is a mandatory, free-text field. It is used to indicate the units or items being studied or recorded. These could be groups of people, organisations, events or items such as text units. For example: primary school teachers, industrial firms, planning applications. When compiling population information, ‘Who’, ‘Where’ and ‘When’ should be indicated within a sentence if all details are known and appropriate.

Examples from the catalogue (survey, non-survey and historical data):

- Women aged between 18 and 65 working outside of the home and earning less than £5 per hour, living in Essex in 2004-2005
- Members of three political organisations in the North West of England during 2000-2002
- Private households in the United Kingdom
- Voters in parliamentary elections in the constituency of Westminster, 1749-1820
- Selected Bibliographical Sources of British and Irish History for the period between 55 BC to present
- List of synthesised records of selected contracts for the supply of wool c. 1200-1330
- Papers generated by the original Mass-Observation social research organisation (1937 to early 1950s) and material collected continuously since 1981
- Speeches in the British Parliament, 1870-1914, and in the German Reichstag, 1871-1921
- Committee papers of the Child Poverty Action Group, from 1965 to 1998

RELU studies

Since RELU data collections frequently contain multiple diverse datasets, population should be described in detail at individual dataset level. It should be noted that ‘dataset’ here does not mean ’data collection’ or ‘study’ level, but refers to a file or group of files; for example, a survey dataset may be one quantitative file, but a set of interview transcripts may be a group of files that constitute one 'dataset' in this sense.

18. Methodology

18.1. Time Dimensions

This is a mandatory field. One or more options may be selected from a controlled list. There is also a note field for free-text entry which can be used to enter additional or qualifying information.

For studies other than cross-sectional one-time studies, the frequency (e.g. number of follow-ups, or intended) and interval (e.g. how often) should be added to the note field, e.g. 'Repeated cross-sectional study; annual'; 'Longitudinal/panel/cohort study; 5 waves'.

Some cross-sectional studies include a longitudinal element (e.g. where follow-up surveys were conducted with the same respondents). If so, include a sentence in the note field to cover this.

18.2. Sampling Procedure

This is a mandatory field. One or more options may be selected from a controlled list. There is also a note field for free-text entry which can be used to enter additional or qualifying information.

Where sampling information is complex and not covered by one of the choices listed but is detailed in the documentation, add **See documentation for details** in the note field.

Where sampling information is not covered by one of the choices listed and is not detailed in the documentation (or there is no documentation), add **No information recorded** in the note field.

18.3. Number of Units

This is an optional, free-text field.

RELU data collections
Since RELU studies frequently contain multiple diverse datasets, number of units should be described in detail at dataset level. It should be noted that 'dataset' here does not mean 'study' level, but refers to a file or group of files; for example, a survey dataset may be one quantitative file, but a set of interview transcripts may be a group of files that constitute one dataset in this sense.

ONS Opinions/Omnibus:
Number of Units information is located in the Response Rate table within the Technical Report deposited with each Opinions survey.

18.4. Method of Data Collection

This is a mandatory field. One or more options may be selected from a controlled list. There is also a note field for free-text entry which can be used to enter additional or qualifying information.

For certain newer methods of collecting data, options are not currently available in the controlled list. In these cases, please use the appropriate terms from the list below in the free-text note field:

Email survey
Content analysis
Online (web-based) survey

18.5. Weighting

This is a mandatory field, but is currently free-text and so the following standard wording should be used:

- Where no weighting has been used (for quantitative studies) add **No weighting used**
- Where information exists and is contained within the documentation, add: **Weighting used. See documentation for details.**
- Where no information exists in the documentation, but a weighting variable is included in the data file, add: **Weighting variable in data. No further information available**
- If no clear information exists within the documentation, use **No information recorded**.
- Where the study is qualitative and therefore does not use weighting, add **Not applicable**

18.6. Control Operations

This is an optional, rarely-used field. One or more options may be selected from a controlled list. There is also a note field for free-text entry which can be used to enter additional or qualifying information. This field does not display in the external version of the catalogue record.

18.7. Level of Processing

This is a mandatory field. Select the appropriate level (or 'None') from the controlled list.

19. Data Sources

This is an optional field. One or more options may be selected from a controlled list. There is also a note field for free-text entry which can be used to enter additional or qualifying information. Where many sources have been used and it is judged that the list would be too long to be practical in the study description, the entry can be reasonably generic and the user referred to the documentation for further details.

20. Language of Written Material

This is a mandatory field, with two elements: 'Study Description' and 'Study Documentation'. If the materials are in English, a tick box can be checked. Otherwise entry is by free text

21. Related Studies

This is an optional field. The appropriate generic may be selected, or other studies may be cross-referenced. See also information on generic records above.

22. References and Publications by PI/Others

These are optional free-text fields, where HTML tags may be used. See the document *Bibliographic References Format Guide* (UKDA 129) for appropriate reference formatting and *Basic HTML Editing Reference* (UKDA 098) for information on HTML tag use.

RELU studies
For RELU data collections, the publications list should begin with a reference to the RELU Knowledge Portal, where project publications are harvested weekly from the ESRC web portal, including a project-specific HTML link (with brief study title), e.g.:

> "Publications may be found via the RELU Knowledge Portal: Eating Biodiversity** ."

23. Administrative Notes

This is an optional field and is free-text. It does not appear in the external display of the catalogue record and is used used for recording information that is useful for internal administration.

24. Access

24.1. Access Conditions

This is a mandatory field. Text is automatically generated by the *Access Code.*

24.2. Access Code

This is a mandatory field. One code should be selected from a drop down list. In the case of generic groups, the access code is added to individual chicks, not the generic record, as access conditions may differ for studies within the same generic group. A current list of access codes is available in the *Discover catalogue Access Codes* document.

24.3. Recommendations

This is an optional field, used, for example, to recommend that user purchase a particular publication for use with the data collection. It is a free-text field where HTML tags may be used. See the document *Basic HTML Editing Reference* (UKDA 098) for information on HTML tag use.

24.4. External Note

This is an optional field, used, for example, to provide further information on access or availability of the data collection, such as an embargo notification. It is a free-text field where HTML tags may be used. See the document *Basic HTML Editing Reference* (UKDA 098) for information on HTML tag use.

RELU collections
Where data have been supplied to CEH, a note may be added here to alert users to additional ecological data available from the Environmental Information Data Centre of the Centre for Ecology and Hydrology, with a link to https://gateway.ceh.ac.uk/ , e.g.:

> ***"***Ecological data from this study are available via the Environmental Information Data Centre of the Centre for Ecology and Hydrology." (with link to https://gateway.ceh.ac.uk/)**.**"

24.5. Originating Archive

This is an optional field. One or more options may be selected from a controlled list. There is also a note field for free-text entry where the originating archive's study reference number should be entered.

25. Date of Release/Date of Edition

This is a mandatory field. It is generated automatically when a record is first released to the catalogue. Any subsequent edition dates are automatically generated when the new edition box is checked. This should be done just prior to re-releasing in order to reflect the correct date. New edition dates may be manually altered using the boxes provided, but this should only be done in exceptional circumstances.

26. Copyright

This is a mandatory field. It is free-text, but there is an option to select the standard text for **Crown copyright**. In the case of generic groups, the entry can go in either the chick or generic as appropriate. If it is added to the generic record then the information must be checked with each new chick record catalogued.

The copyright statement start with the word **Copyright** other than for Crown Copyright statements or **No information recorded** (see below). This is because there is no copyright heading on the title page, so the statement would be meaningless.

Crown copyright

For Crown copyright data, check the box provided to display this statement:

> **Crown copyright material is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland**

If the copyright is held jointly between Crown and other:

> **Crown copyright held jointly with the Institute for Fiscal Studies. Crown copyright material is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.**

The following link provides a useful list of Crown Bodies as well as a list of non-crown bodies commonly mistaken for Crown:

UK Crown Bodies www.opsi.gov.uk/advice/crown-copyright/uk-crown-bodies.htm

Copyright held by either an individual or a corporate body

> Copyright J. Bloggs
>
> or
>
> Copyright University of Essex. Department of Government

Copyright held jointly

> Copyright J. Bloggs and D. Smith
>
> or
>
> Copyright J. Bloggs, D. Smith, W. Green and K. Brown

Further clarification

Where there is need for further clarification, e.g. if a certain person or body holds copyright for only part of the data, then an explanatory note should be added. However, remember the copyright statement appears on the title page (which is a separate part of the catalogue record). Therefore, if adding explanatory text would look clumsy, this information can be added under *Other Acknowledgements*.

Punctuation

Copyright statements do not generally require full stops at the end of the statement. However, where there is more than one sentence, as in the joint Crown Copyright example above, all sentences/statements should end with a full stop

No copyright information

If there is no information available on copyright (which should only be the case for old studies), use No information recorded. This statement indicates that UK Data Archive do not have any information about the copyright of the particular study, but does not preclude the possibility that information may exist somewhere. The processor/cataloguer **must** endeavour to find this copyright information, especially in the case of a major ESDS series, either by reference to the documentation, study website, or by contacting the depositor directly. If in doubt, please consult the Data Services Manager.

27. Checking catalogue records

Once the catalogue record and keyword index is complete and the corresponding data collection archived, the record should undergo a proof-reading check for accuracy by a member of the Ingest Services team. However, this should not hold up release; post-release check is fine. In this case, the record should be released as normal and a Digital Object Identifier (DOI) allocated. Other staff may also be consulted for advice where they have a special interest in the data collection, for example a member of staff with qualitative expertise.

28. Releasing the record: the Digital Object Identifier (DOI)

The Digital Object Identifier (DOI) framework is an international standard for identifying objects (or information about them) in a globally unique way. Using a DOI ensures that even if the location or information about an object changes, the DOI will always link or 'resolve' to the same object. A DOI comprises a string of letters and numbers that can be used to make resources directly available over the internet. When the identifier is persistent, a citation will always resolve to the original object. During 2011, the Archive undertook a project with the DataCite organisation to introduce DOIs and improve data citation. This ensures that even if the location of the data changes, the DOI will always link to the data that was used.

Since October 2011, DOIs have been allocated to all Archive data collections apart from those with ‘virtual catalogue records’ where the data are physically held elsewhere (and thus may have an existing DOI). A DOI is included in the citation, which is supplied to users in the HTML Study Information file included in each download package, and is also viewable on the DOI ‘jump’ webpage attached to the individual catalogue record. The DOI is added via the catalogue input programs and appears at the end of the citation, similar to an ISBN/ISSN number. Once a DOI has been allocated to a study, it is incremented at times of major change to the study, for example when a new edition is released.

28.1. Adding a DOI to new studies

When the new study catalogue record is published and ready to release, the ‘Release datasets and mint DOI’ button on the published record screen is selected. For new data collections, the following screen will appear and the ‘first edition’ DOI will be allocated automatically.

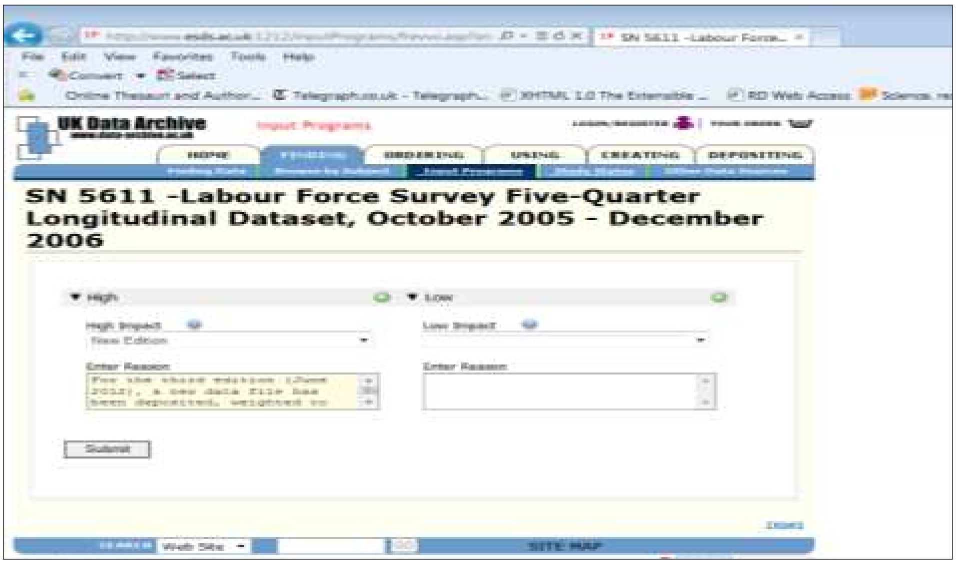
Example of first edition study citation with DOI:

> Office for National Statistics. Social Survey Division, *Labour Force Survey Five-Quarter Longitudinal Dataset, April 2010 - June 2011* [computer file]. Colchester, Essex: UK Data Archive [distributor], June 2012. SN: 7026, http://dx.doi.org/10.5255/UKDA-SN-7026-1

Further information is given on the catalogue ‘jump’ page associated with the DOI; for this example see http://www.esds.ac.uk/doi/?sn=7026 (the DOI URL included in the citation always resolves to the appropriate jump page)

28.2. New editions/high impact changes and DOI

When a new edition record is released (a ‘high-impact’ change), the DOI can be updated accordingly. When the updated catalogue record is ready to release, select the ‘Release datasets and mint DOI’ button again. When the following high-impact/low-impact screen appears, select ‘New edition’ from the high-impact change drop-down list and add the text of the catalogue record new edition statement to the ‘Enter Reason’ text box (this may be edited slightly if necessary).



For the release of the new edition, the DOI number undergoes an increment. (Note: the final version number changes to 2 in this example (although this is the third edition); this is because of the timing of the DOI project in 2011):

> Office for National Statistics. Social and Vital Statistics Division and Northern Ireland Statistics and Research Agency. Central Survey Unit, *Labour Force Survey Five-Quarter Longitudinal Dataset, October 2005 - December 2006* [computer file]. *3rd Edition.* Colchester, Essex: UK Data Archive [distributor], June 2012. SN: 5611, http://dx.doi.org/10.5255/UKDA-SN-5611-2

28.3. Low impact changes and DOI

For minor changes to the catalogue record such as the correction of spelling errors, catalogue records will need to be re-released in order for the update to show. However, because these are low-impact changes that do not substantially affect the study materials, there is no need to create a new DOI. A low-impact change can be registered using the same procedure as that for a high-impact change as described above, but this time select an option from the Low Impact change drop-down list, and enter text in the box provided, such as ‘Spelling error corrected’, as appropriate.

Publication references
Most publication references are added to generic records, which do not have a DOI. Therefore, when the publication reference has been added, the standard 'release datasets' button can be selected. For non-generic studies, add the reference as normal and then follow the procedure for low-impact changes. There is an option on the drop-down list for 'Addition of publication reference'.

[Note: Additional Archive documentation regarding DOI policy and procedures is currently in development.]

29. Editing catalogue records after release

Unlike the previous Data Catalogue on the ESDS website, where edits appeared immediately on (re)release, Discover catalogue records update overnight, so that the metadata can go through a series of automated checks. Therefore, if a catalogue record requires an edit post-release, the edits will not appear in Discover until the following day. Therefore, cataloguers must remember to revisit the Discover record to check any edits made. An internal staff version of Discover for quality assurance purposes is available at http://qa.ukdataservice.ac.uk/discover and the catalogue record should appear there c. 1 hour after release. Edits can then be made as needed, but the main Discover catalogue record should still be checked the following day. r

30. Catalogue records for specific data series

30.1. ONS Opinions and Lifestyle (formerly Omnibus)

The ONS Opinions and Lifestyle (OPN) series comprises a set of core questions asked each month the study is conducted. In addition, ad hoc modules on certain topics are added periodically. These modules are often re-run at a later date in another OPN survey.

When a new OPN study is received, it is helpful to check previous OPN catalogue records to find those where the same ad hoc modules have previously run. For example Module 363, Disability Monitoring and Module 210, Tobacco Consumption were asked in the OPN for March 2010. When creating the catalogue record for the OPN study for March 2011, when the modules were run again, the earlier study's catalogue record was mass-copied via the catalogue input programs to create the new record, meaning that the 'Abstract' information and 'Keywords' for Module 210 and Module 363 were already in situ. The information was then easily edited to suit the new study.

The following procedures can be undertaken to create the OPN catalogue record:

- Edit the study **'Title'** field from the previous 'mass-copied' record to reflect the new study title.

- Add the **'Alternative Title'** field information (ONS Omnibus Surveys; OPN) into each study chick (not the generic catalogue record).

- Editing the **'Subject Category'** field:

    - Identify the subject of each module deposited and include all relevant subject categories.

    - The following two subject categories are contained within the generic catalogue record as they relate to the 'core' demographic questions that are always asked within each module: ***Social attitudes and behaviour - Society and culture*** and ***ONS Opinions and Lifestyle Survey - Major studies***.

    - Delete subject category information relating to modules not included within the new record (i.e. from the redundant mass-copied metadata).

    - Add subject category information relating to modules where information has not been mass-copied and for new modules within the new dataset.

    - An Excel .xls file exists which contains most, but not all, subject category levels that have been assigned to Omnibus/Opinions modules in the past. This file allows the cataloguer to be consistent and saves on searching for/identifying which subject categories have been previously assigned to modules. Please contact the Senior Data Curation Officer for details.

    - If a new module that has not been asked in the past has been deposited, the subject category information assigned to that module should be recorded in the subject category Excel file for future reference.

- Editing the **'Sponsor'** field: checks should be made as follows:

    - Identify the module sponsor (this information is contained within the Interviewer Instructions within the study documentation, which lists the modules asked each month,). Always check the sponsor organization via their respective web pages as the documentation deposited is not completely reliable.

    - If there has been a change to the module sponsor/organization name, it may be necessary to go back and edit the information within released studies if the correct information has not been made available to us until sometime after the event has taken place. If this is the case, it is necessary to inform the Nesstar processing staff as edits will be necessary to study metadata within the UKDS Nesstar Catalogue.

    - Module naming convention has changed over time with titles changing and the trend for numbers evolving to include letters, for example: Module 330 Internet Access, became Module 362 e-Government, which became Module MAI Internet Access, which became Module MAZ Internet Access, which became Module MAJ Internet Access, extra component.

    - Delete any 'Sponsor' information massed across, relating to modules not included in the new study.

    - Add any 'Sponsor' information relating to modules where there is no massed across information and for new modules deposited within a new dataset.

- Editing the **'Abstract'** field:

    - Add/delete metadata information relating to the modules that make up the survey if information has been massed across from a previous study as the new study will possibly contain some but not all of the same modules. For consistency sake, a search needs to be made of previous studies containing re-used modules so that metadata can be copied into the new study Abstract resulting in consistent topic information over time.

- Editing **'Date of Fieldwork'** field*:*

  - Add in the month and year of the new study.

- Editing **'Number of Units'** field:

  - This information is located in the Response Rate table within the Technical Report deposited with each Opinions survey.