# 3 Documentation

## 3.1 Making sense of metadata

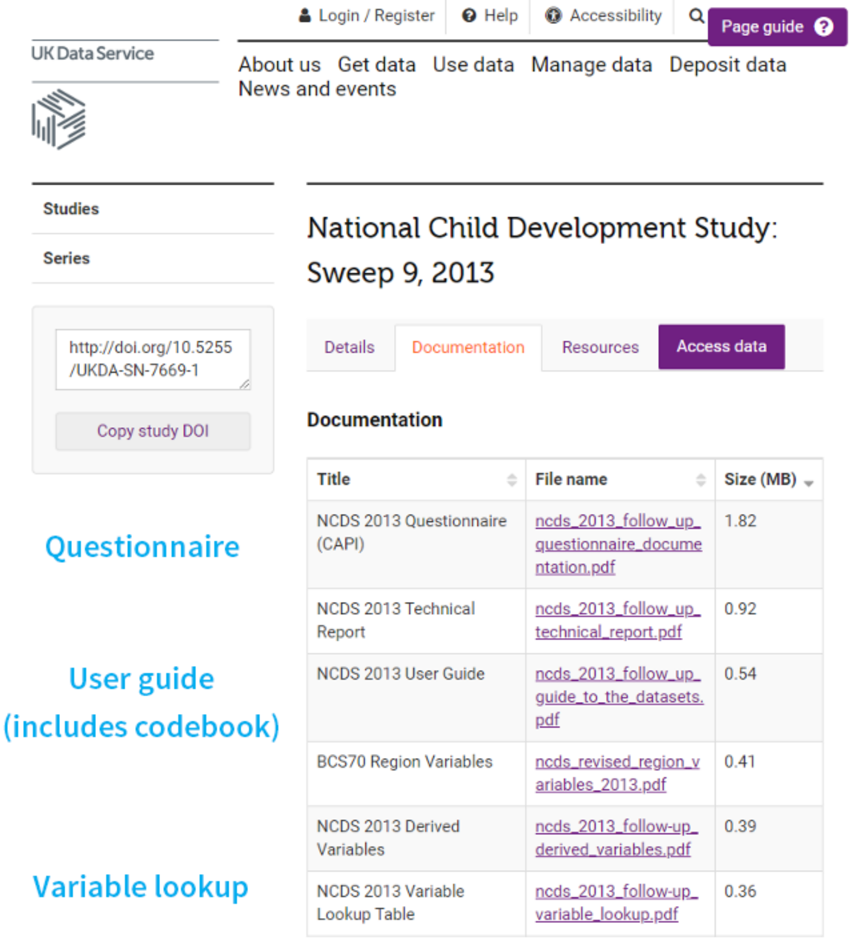
In the context of longitudinal study datasets, metadata are usually provided as a set of documents available to download alongside the data itself. This describes, for example, when a given data collection wave took place and who conducted it. It describes the coverage of the data. It describes the exact questions asked and how they fit into an overall data collection instrument.

This documentation typically includes:

- Codebook
- User guide
- Questionnaire(s)

**Suggested citation:** Kaye, N., Mills, H. & Johnson, J. (2020). *Understanding metadata*. CLOSER Learning Hub, London, UK: CLOSER

Authors: Neil Kaye, Hayley Mills and Jon Johnson



http://doi.org/10.5255/UKDA-SN-7669-1

National Child Development Study: Sweep 9, 2013

Documentation

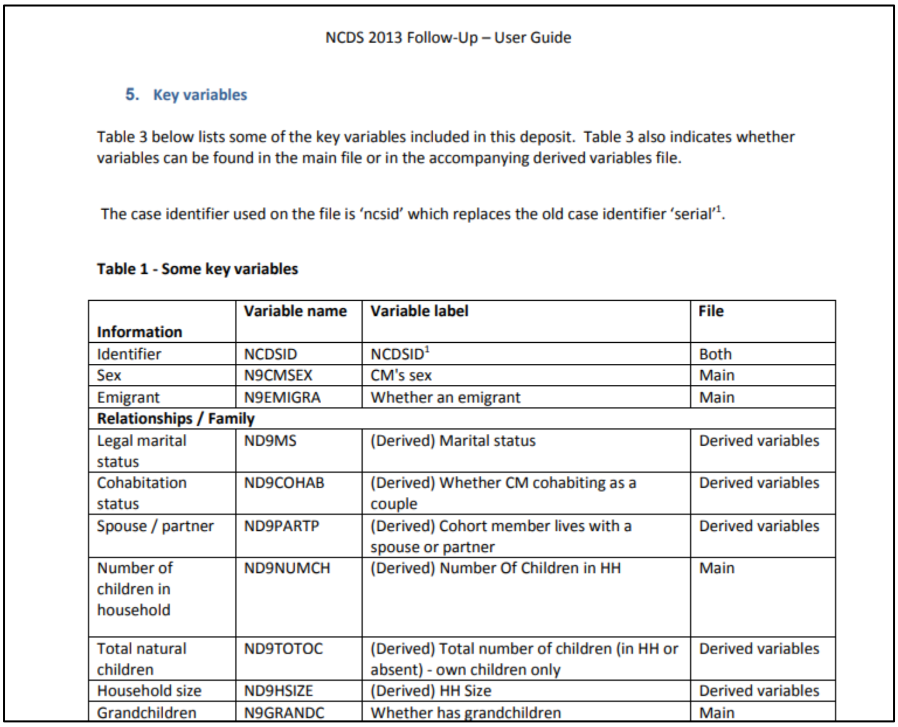
| Title | File name | Size (MB) |
|---|---|---|
| NCDS 2013 Questionnaire (CAPI) | ncds_2013_follow_up_questionnaire_documentation.pdf | 1.82 |
| NCDS 2013 Technical Report | ncds_2013_follow_up_technical_report.pdf | 0.92 |
| NCDS 2013 User Guide | ncds_2013_follow_up_guide_to_the_datasets.pdf | 0.54 |
| BCS70 Region Variables | ncds_revised_region_variables_2013.pdf | 0.41 |
| NCDS 2013 Derived Variables | ncds_2013_follow-up_derived_variables.pdf | 0.39 |
| NCDS 2013 Variable Lookup Table | ncds_2013_follow-up_variable_lookup.pdf | 0.36 |

Questionnaire

User guide (includes codebook)

Variable lookup

**Suggested citation:** Kaye, N., Mills, H. & Johnson, J. (2020). *Understanding metadata*. CLOSER Learning Hub, London, UK: CLOSER



NCDS 2013 Follow-Up – User Guide

5. Key variables

Table 3 below lists some of the key variables included in this deposit. Table 3 also indicates whether variables can be found in the main file or in the accompanying derived variables file.

The case identifier used on the file is 'ncsid' which replaces the old case identifier 'serial'[1].

**Table 1 - Some key variables**

| Information | Variable name | Variable label | File |
|---|---|---|---|
| Identifier | NCDSID | NCDSID[1] | Both |
| Sex | N9CMSEX | CM's sex | Main |
| Emigrant | N9EMIGRA | Whether an emigrant | Main |
| **Relationships / Family** | | | |
| Legal marital status | ND9MS | (Derived) Marital status | Derived variables |
| Cohabitation status | ND9COHAB | (Derived) Whether CM cohabiting as a couple | Derived variables |
| Spouse / partner | ND9PARTP | (Derived) Cohort member lives with a spouse or partner | Derived variables |
| Number of children in household | ND9NUMCH | (Derived) Number Of Children in HH | Main |
| Total natural children | ND9TOTOC | (Derived) Total number of children (in HH or absent) - own children only | Derived variables |
| Household size | ND9HSIZE | (Derived) HH Size | Derived variables |
| Grandchildren | N9GRANDC | Whether has grandchildren | Main |

*Extract from NCDS 2013 Follow-Up User Guide – codebook for key variables, p14.*

## 3.2 Where to find data and documentation

As a researcher, you will often need the accompanying documentation to make sense of the dataset, understand the variable codes, how variables have been derived and technical information on, for example, using weights.

All the necessary documentation should be available in the same place as the dataset. Data available through the UK Data Service, for example, are listed alongside their accompanying documentation.

Documentation may be available at **study-level** – including information about the research design, methodology, sampling, data collection methods and questionnaire – and/or at **data-level** – most-commonly in the form of a codebook, containing information on variable names, question texts, labels and descriptions, coding frames and missing data codes.

You can also usually view the original questionnaire or topic guide and instructions for those who administered the research instruments.

Datasets, along with their accompanying documentation, can be found through a vast number of data archives, catalogues and repositories. In the UK, the main data catalogues for biomedical and social sciences include:

- UK Data Service – https://www.ukdataservice.ac.uk/
- UK Data Archive – https://www.data-archive.ac.uk/
- Office for National Statistics - https://www.ons.gov.uk/
- Data.gov.uk – https://data.gov.uk/
- UK Biobank – https://www.ukbiobank.ac.uk/

## 3.3 Using documentation for data harmonisation

Documentation plays a key role in guiding efforts to bring together and harmonising data from different studies or time periods. Understanding how data were collected and what processing may have been applied to clean and transform the data requires access to documentation that is comprehensive and accurate.

User guides and the original questionnaires can explain why data from one source has a large amount of **missing values** on certain variables (e.g. due to survey logic).

Such documentation may offer important detail on the **meaning of labels** used within categorical variables that might otherwise superficially seem different to the labels used within data files from other sweeps or studies.

Understanding the **instructions given to respondents** may inform how we interpret the responses they gave. (e.g. a question asked directly by an interviewer and one answered on a private screen)

Details on the devices used to take **biomedical measurements** can help us figure out if and how the data can be made more comparable through established calibration steps. (e.g. the type of device used to take blood pressure readings)

Ensuring that data collection methods are well documented is important in ensuring that other studies can replicate the same methods to ensure new data collected are prospectively harmonised.

Documentation is also important in any data produced by harmonisation efforts. The decisions made over what data are included, from what studies, and how the harmonised variables are generated and defined, are all essential detail in directing researchers who want to subsequently use and analyse the newly produced harmonised datasets in their own work.

More detail on data harmonisation can be found in the 'Data Harmonisation' module.

### 3.3.1 Harmonised datasets

In the harmonised datasets produced by CLOSER's work packages, we have deposited the data at the UK Data Service, alongside user guides and the analysis script files used in generating these new harmonised data resources:

**Suggested citation:** Kaye, N., Mills, H. & Johnson, J. (2020). *Understanding metadata*. CLOSER Learning Hub, London, UK: CLOSER