

Documentation guide

When archiving research data, the data should be documented on three levels: Project level, dataset level and variable level. **Project level** documentation describes the aim of the study, who is responsible for the project and the methods applied. At the **dataset level** you should provide an overview of the different files and how they relate to each other. Documentation at the **variable level** must describe the data in a way that makes it understandable to outsiders. You should provide an overview and explanation of all variables and codes applied.

Project level:

Title

Context – Project summary (purpose, research questions, existing research etc.), state the name of the person and/or institution responsible for the project, who collected the data, and any partners or funding institutions.

Data collection methods– Sampling, procedures, scales, software and tools used, instruments, geographical coverage, time periods, data types, and use of secondary data.

Access – information on who should be given access to data from the project and when the data should be made available

Example:

Title	Level of living 2009 – Cross sectional study – Working conditions
Producer	Statistics Norway
Copyright	Copyright © 2009 Statistics Norway
Abstract	Statistics Norway's (SSB) Level of living - cross sectional study (with rotating themes) have as its primary objective to, over time, cover all important issues concerning the level of living. The topics have rotated in the following way: Working conditions in 1996, 2000, 2003 and 2006; Living conditions, community and leisure in 1997, 2001 and 2004; Health in 1998, 2002 and 2005. The main topic of the Level of Living Survey 2008 is working conditions.
Date of collection	Start: 2009-06-11 End: 2010-01-09
Time period covered	2009-2010
Country	Norway
Geographic coverage	Country
Time method	Cross-sectional survey
Kind of data	Survey data
Data collector	Statistics Norway
Sampling	Persons in Norway, from the age of 16 years. Gross sample: 20460 persons Net sample: 12255 persons Response rate: 59,9%
Access	Data are made available for research purposes given a written application.

Dataset level:

Structure – File name and format, description of different cases (time, place, sample) and file content, as well as the relationship between the different files.

Version control – Version number and date. The name of responsible person.

Personal identification – Documentation on any permissions to store personal data or an explanation on how the data has been anonymized (deleted or recoded variables, generalized meaning of detailed text – for example by changing a doctors specific area of expertise to a broader medical field, use of pseudonyms, eliminating outliers etc.)

Example:

Format	Name	Size kB	Version	Date created	Date modified	Description
PDF	Dataset_documentation.pdf	1403	02.01	01/06/2007	20/01/2008	Dataset Documentation (variable list, derived variables, variables used in report tables.)
SPSS	VisionScreenResults_00_05	625	00.05	04/07/2007	25/11/2007	Results data of 120 Vision Screen Test carried out in 5 nurseries in Essex during June 2007

Variable level

(For structured/quantitative data):

Names, labels and description of variables, records and their values. Questionnaires and/or codebooks.

Explain the applied codes, values, classifications, terminology and abbreviations. Describe codes and reasons for missing values and constructed variables.

Example :

Datasett: Survey on work, lifestyle and health, 2010					
Variable group	Name	Lable	Description	Values	Forklaringer
Labour and employment	v009	q9 Permanent or temporary employee	Question: Do you have a permanent or temporary contract at your current job?	1 – Permanent position 2 – Temporary hired 8 – Do not want to answer 9 – Don't know	
	v010	q10 Contracted working hours per week	Question: What is your contracted working hours per week? Do not include overtime or absence due to illness etc. State working hours in full hours.		

Health	v095	q33 Assessment of own health	Question: How would you assess your own health?	1 – Very good 2 – Good 3 – Neither good or bad 4 – Bad 5 – Very bad 8 – Do not want to answer 9 – Don't know	
	v096	q34 Someone close to IO that IO can talk to	Question: Except from family members, do you have anyone close to you that you can talk to in confidence?	1 – Yes 2 – No 3 – Don't know 8 – Don't want to answer	IO = Interview object