Data Documentation Glossary for the Social Sciences

This document is useful when you are entering a dataset or a data file in order to deposit your data. It defines and explains the categories of controlled vocabulary lists used in the SWISSUbase metadata schema for the social sciences. The categories were chosen based on the set created by the Data Documentation Initiative (DDI), an international standard for describing statistical and social science data. Thus, the definitions below are mainly taken from the DDI website (http://www.ddialliance.org/controlled-vocabularies).

The categories defined in this document are part of three lists:
• Time method (creating/editing dataset)
• Analysis unit (creating/editing data file)
• Mode of data collection (creating/editing data file)

Time method

Cross-section
Data collected by observing subjects within the study period, without regard to changes over time. May include more than one collection event. Analysis of cross-sectional data often consists in comparing the differences and similarities among subjects.

Longitudinal
Data collected repeatedly over time to allow studying change in a population. At least some of the questions or modules are repeated over waves.

Time Series
Data collected repeatedly over time to study change in observations. These are typically "objective" measurements of phenomena that can be observed externally, as opposed to attitudes/opinions or feelings. Examples may include economic indicators, vital statistics, etc.

Other
Use if the time method is known, but not found in the list.

Analysis unit

Individual
Any individual person, irrespective of demographic characteristics, professional, social or legal status, or affiliation.

Couple

Two people related by marriage, domestic partnership or concubinage, whether they live together or not.

Family

Two or more people related by blood, marriage (including step-relations), adoption or fostering and who may or may not live together. For example, used when researching the extent to which people provide support and assistance for their relatives.

Household

A person or a group of persons who share the same dwelling unit and common living arrangements. These common living arrangements may include pooling some, or all, of their income and wealth, and consuming certain types of goods and services collectively, mainly housing and food.

Group

Two or more individuals assembled together or having some unifying relationship.

Organization

Any kind of formal administrative and functional structure - includes associations, institutions, agencies, businesses, political parties, schools, etc.

Geographic Unit

Any entity that can be spatially defined as a geographic area, with either natural (physical) or administrative boundaries.

Housing Unit

A housing unit is a house, an apartment, a mobile home, a group of rooms, or a single room that is occupied (or if vacant, is intended for occupancy) as separate living quarters. Separate living quarters are those in which the occupants live and eat separately from any other persons in the building and which have direct access from the outside of the building or through a common hall.

Building

Any construction serving as shelter and separate, for example a church, a factory, an administrative building, etc.

Event / Process

Any type of incident, occurrence, or activity. Events are usually one-time, individual occurrences, with a limited or short duration. Examples: criminal offenses, riots, meetings, elections, sports competitions, terrorist attacks, natural disasters like floods, etc. Processes typically take place over time, and may include multiple "events" or gradual changes that ultimately lead, or are projected to lead, to a particular result. Examples: court trials, criminal

investigations, political campaigns, medical treatments, education, athletes' training, etc.

Text Unit
Books, articles, any written piece/entity.

Time Unit
Any period of time: year, week, month, day, or bimonthly or quarterly periods, etc.

Other
Use if the unit of analysis is known, but not found in the list.

Mode of data collection

Face-to-face interview (CAPI, CAMI, PAPI, etc.)
Data collection method in which a live interviewer conducts a personal interview, presenting questions and entering the responses.

Telephone interview (CATI, etc.)
Interview administered on the telephone.

Self-administered questionnaire: Paper
Self-administered survey using a traditional paper questionnaire delivered and/or collected by mail (postal services), by fax, or in person by either interviewer, or respondent.

Self-administered questionnaire: Computer-assisted (CASI, ACASI, etc.)
Computer-assisted self-interview (CASI). Respondents enter the responses into a computer (desktop, laptop, tablet, etc.) by themselves. The administration of the questionnaire is managed by a specifically designed program/application but there is no real-time data transfer as in CAWI, the answers are stored on the device used for the interview.

Self-administered questionnaire: Web-based
Data are collected using a Web questionnaire.

Transcription
Capturing information in writing from a different source or from a different medium. For transcribed interviews or observations, it is recommended to document the primary mode of collection, using one of the interview or observation terms.

Compilation / Synthesis
Collecting and assembling data from multiple, often heterogeneous sources that have one or more reference points in common, and at least one of the sources

was originally produced for other purposes. The data are incorporated in a new entity. For example, providing data on the number of universities in the last 150 years using a variety of available sources (e.g. finance documents, official statistics, university registers), combining survey data with information about geographical areas from official statistics (e.g. population density, doctors per capita, etc.), or using RSS to collect blog posts or tweets, etc.

Observation
Observation that is conducted in a natural, non-controlled setting with or without interaction between the researcher and his/her subjects. Research method that involves collecting data as they occur (for example, observing behaviours, events, etc.), without attempting to manipulate any of the independent variables.

Self-administered writings and/or diaries
Narratives, stories, diaries, and written texts created by the research subject.

Focus group
A group discussion on a particular topic, organized for research purposes. The individuals are selected with relevance to the topic, and interaction among the participants is used as part of the method.

Experiment
Research method involving the manipulation of some or all of the independent variables included in the hypotheses.

Content coding
As a mode of secondary data collection, content coding applies coding techniques to transform qualitative data (textual, video, audio or still-image) originally produced for other purposes into quantitative data in accordance with pre-defined categorization schemes. Example: coded articles.

Recording
Registering by mechanical or electronic means, in a form that allows the information to be retrieved and/or reproduced.

Simulation
Modelling of real-world processes, events, or systems, often using computer programs. For example, a program modelling household consumption responses to indirect tax changes; or a dataset on hypothetical patients and their drug exposure, background conditions, and known adverse events.

Summary
Presentation of information in a condensed form, by reducing it to its main points. Example: Abstracts of interviews or reports are used as data rather than the full-length interviews or reports.

Aggregation

Statistics that relate to broad classes, groups, or categories. The data are averaged, totalled, or otherwise derived from individual-level data, and it is no longer possible to distinguish the characteristics of individuals within those classes, groups, or categories. For example, the number and age group of the unemployed in specific geographic regions.

Measurements and tests

Assessing specific properties (or characteristics) of beings, things, phenomena, (and/ or processes) by applying pre-established standards and/or specialized instruments or techniques.

Other

Use if the mode of data collection is known, but not found in the list.