



EUCCONET Metadata Workshop Edinburgh, Scotland, 25th June, 2012

Organised by Kelly Ward, Head of Data Management, (NatCen Social Research) and Joan Corbett, Senior Data Manager (ScotCen Social Research) with support from Jackie Palmer, Data Manager (ScotCen Social Research).

1. Summary

Standards have been emerging over recent years which provide a common framework for describing and capturing metadata. Metadata provides important information enabling users to understand the origin and structure of datasets. This serves a number of purposes both for secondary analysis but also to ensure appropriate capture of the study and potential re-use of information for other studies. An example is the Data and Documentation Initiative (version 3) which aims primarily to document the whole of the survey cycle. This standard has been adapted and developed by academics and data archives; however the standard is complex and the DDI specific tools are often incorporated based on the needs of a specific project.

This workshop aimed to bring together a number of specialists who have experience in setting up and working with metadata as well as those who are just starting out. The workshop was designed to share current experiences, showcase tools with online demonstrations and understand alternative metadata models and developments by inviting presentations from a range of delegates.

A summary of each discussion is presented in section 2 below along with any useful links.

2. Study Reports

Centre for Longitudinal Studies, Birth Cohorts, Jon Johnson

Jon talked through the history of metadata focusing on what metadata is and how it helps us to understand and communicate "data about data". The questionnaire is the starting point for many of us, along with code-books and together they cross-reference to describe the output.

In the early 1990s there were 4 main tools, Excel, SPSS, Access and Word. Surveys continue to look the same, but the volume increases, the complexity increases, routing is increasingly more complex, and data usage has become

much more widespread. There is a decline however, in the researcher's understanding of the data as a separation appears between the roles of survey managers, researchers and data managers.

Tools have emerged such as the Nesstar catalogue available via the ESDS UK Data Archive. This tool helps to understand the data but many organisations are reviewing how best to capture the questionnaire content and routing. DDI is a good framework but there is currently a lack of standard tools that can be easily applied.

The Economic and Social Research Council (ESRC) and the Medical Research Council (MRC) have recently launched the initiative CLOSER – Cohorts and Longitudinal Studies Enhancement Resource – which aims to bring together some of the most important studies of people's lives in the UK. Part of this initiative will explore the need for a centralised resource that builds a world-class metadata repository.

Further details of this can be found at <http://www.closer.ac.uk>

Growing up In New Zealand, Arier Lee

Arier discussed the essential features of the survey and recognised that meta-data has two uses, data-dictionary and application. The Growing up In New Zealand study produces a data dictionary for each dataset and Arier described the process of creating this.

The team hope to build a new system that will 'talk' to the relevant data locations and will create a data dictionary before the interview rather than post interview. A Questionnaire library will offer source questionnaires for the next cohort. They anticipate that this will provide a streamlined system with a centralised repository and a full audit trail, version-controlled data dictionary. This will enable robust data cleaning and centralised tracking to minimise both errors and effort.

NatCen Social Research (NatCen), Colin Micelli

Colin demonstrated the Blaise Data Model that NatCen use on some of their large scale surveys. The tool was developed in Holland and is suitable for multi-mode, face-to-face interview and web-browser integration. Colin's demonstration noted the different components of the Blaise tool with particular emphasis on the metadata translation tool, the data viewer and the documentation tool.

Metadata Technology, Arofan Gregory

Arofan provided an overview of DDI and SDMX. The Data Documentation Initiative (DDI) started off in Data Archives as a drive to develop a system for marking up codebooks to make them machine readable. So the early steps were to define what metadata elements were needed to describe a dataset. Over the years this has extended to cover the whole 'lifecycle' – the current version of DDI is DDI3. The concept behind it is that by using a standard way of identifying each element, the metadata for a study that has been documented using DDI3 can be stored and re-used. Full details can be seen at <http://www.ddialliance.org/what>

DDI is an XML-based language and as such can be read by a great many tools and programming languages already available to developers. Software can be developed that reads the DDI and extracts the metadata for reuse, for example conversion of metadata to bespoke documentation. Other organisations have been developing ways of using DDI within their own processes.

Avon Longitudinal Study of Parents and Children (ALSPAC), Andy Boyd

Andy presented an overview of the ALSPAC study and noted that there have been 68 questionnaires between birth and age 18 to their cohort respondents. The ALSPAC team have recently adapted a new variable naming and labelling convention using a universal structure.

Their data is accessed by a range of academic researchers, survey participants, government users and the media. ALSPAC are also involved in the CLOSER initiative.

Data Archive, Jack Kneeshaw

Jack presented an overview of the ESDS UK Data Archive. Data processors play an important role in providing relevant information for a study when they complete the Data deposit form and other documentation to accompany the datasets.

The Data Archive run SPSS to DDI transformation scripts to create metadata records to the Nesstar standards. This was developed from the DDI Alliance and enables search functions at the variable level. They are currently working on translating into the DDI 3 lifecycle and adding additional keywords to enable a wider more comprehensive search function.

Further information can be found at <http://www.data-archive.ac.uk/create-manage/document/metadata>

Longitudinal Study of Young People in England (LSYPE), Alicia Heptinstall

Alicia presented the LSYPE metadata tool which exists to enable access to anonymised data from the LSYPE study and metadata such as sampling details, questionnaires, coding schemes.

Further information on this tool is available here: www.education.gov.uk/ilsype

National Education Panel Study (NEPS), Christian Matiyas

Christian provided an overview of the metadata cycle that involves a large team of survey programmers, software specialists and academics. The team at NEPS have created a relational database for a normalised, single source of information, where each object is identifiable and has a flexible query language.

Further information about NEPS can be found here: <https://www.neps-data.de/en-us/home.aspx>

3. Impact

The main aim of this workshop was to enable a range of specialists such as those working in academic research, data management and survey programming to share their own experiences of working with metadata. The day was designed as an open forum with presentations and question and answer sessions. Participants were actively engaged throughout the day and generated useful discussions to share experiences and take away new information which may be useful when exploring metadata further.

We received very positive feedback from participants about the range of presentations given and the shared experiences that they were able to take away with them. Some colleagues were at the very early stages of considering how best to use metadata and although we recognised that there was not one easy tool to use, the day enabled all of us to consider wider issues.

We hope that the day facilitated useful networking that can continue outside of the workshop and we note that participants would welcome further workshops around this increasingly important issue.

Metadata and documentation for the survey life-cycle

Programme

Monday 25th June

10:00	10:30	Registration
10:30	10:45	Welcome and Introduction
10:45	11:45	Keynote: Jon Johnson: The history of metadata (Birth cohort studies)
11:45	12:15	Arier Chi Lun Lee
12:15	12:30	Coffee
12:30	13:00	Colin Micelli, NatCen
13:00	14:00	Lunch
14:00	15:00	Keynote: Arofan Gregory
15:00	15:30	Andy Boyd, ALSPAC
15:30	16:00	Jack Kneeshaw, ESDS Data Archive
15:45	16:00	Coffee
16:00	16:30	Alicia Heptinstall, LSYPE
16:30	17:30	Keynote: Christian Matyas - German project NEPS
17:30		End
19:00		Group evening meal

Participants

First Name	Surname	Post	Organisation
Amanda	Quail	Research Analyst	Growing up in Ireland Institut national d'etudes demographiques (INED)
Sarah	Cadorel	Archivist - Webmaster	Institut national d'etudes demographiques (INED)
Karine	Lautredoux	ELFE - Growing up in France	Institute of Education
Heather	Joshi	C-Chair Eucconet	UK Data Archive
Jack	Kneeshaw	Service Manager	University Bamberg
Christian	Matyas	System Architect, Web Development	Born in Bradford
Emily	Petherick	Senior Epidemiologist	ALSPAC (University of Bradford)
Andy	Boyd	ALSPAC Data Linkage Manager	Centre for Longitudinal Studies
Jon	Johnson	Senior Database Manager	University of Bristol
Karen	Birmingham	Research Ethics Manager	Swansea University
David	Ford	Director, eHealth Industries Innovation Centre	ScotCen Social Research
Jackie	Palmer	Data Manager	Norwegian Institute of Public Health
Arild	Sunde	Programmer	NatCen Social Research
Colin	Miceli	Senior Project Programmer	NatCen Social Research
Kelly	Ward	Head of Data Management	University of Leeds
Neil	Hancock	Information Manager	Department of Education
Alicia	Heptinstall	Statistician	Growing up in New Zealand
Arier	Chi Lun Lee	Senior Biostatistician	Metadata Technology
Arofan	Gregory	Senior Partner	

