

# Data documentation and metadata for data archiving and sharing

Data Management and Sharing workshop  
Vienna, 14-15 April 2010

# Why document data?

- enables you to understand/interpret data
- needed to make data independently understandable
- ensures informed and correct use, reduces chance of incorrect use/misinterpretation
- if using your data for the first time, what would you need to know?
- UKDA uses data documentation to:
  - create/supplement catalogue record for dataset
  - create user guide(s) and data listing for dataset
  - ensure accurate processing and archiving

# What is data documentation?

## 1. Wider contextual information about project and data

- background, project history, aims, objectives, hypotheses
- publications based on dataset

# What is data documentation?

## 2. Data collection methodology and processes

- sampling
- data collection process - fieldwork, interviewer instructions
- instruments used - questionnaires, showcards, interview schedules
- temporal/geographic coverage
- data validation - cleaning, error-checking
- derived variables - compilation
- weighting: factors and variables, weighting process
- secondary data sources used

# What is data documentation?

## 3. Info on dataset structure

data files, relationships between files, records, cases...

## 4. Variable-level documentation

labels, codes, classifications, missing values, derivations, aggregations

# What is data documentation?

## 5. Data confidentiality, access and use conditions

- anonymisation carried out
  - aggregation, banding, coding and top-coding, disclosure control
  - editing of sensitive material in interview transcripts
- consent conditions/procedures
- access or use conditions of data

# Metadata

- Standardised structured core data documentation for:
  - bibliographic reference
  - resource discovery (searchable)
- International standards/schemes
  - Data Documentation Initiative (DDI)
  - Dublin Core
  - ISO19115
  - Metadata Encoding and Transmission Standard (METS)
  - General International Standard Archival Description (ISAD(G))

EXIT

EXIT

Description - Windows Internet Explorer  
 EN English (United Kingdom) Microphone Tools Handwriting Drawing Pad

http://dapc70.essex.ac.uk/testWebSite1/abstract.aspx

Google

Bookmarks 218 blocked Check AutoLink AutoFill Send to Settings

Description

Home | Contact | UKDA | Help

# UKDA-store

UK Data Archive

- Description
- Subject categories
- Contributors
- Methodology \ Sampling
- Coverage
  - Non-survey Data
  - Publications
- FORMATS
  - Preview/Submit
  - Logout

**\* Title of the data collection**

Warmwater Fish Production as a Niche Production and Market Diversification Strategy for Organic Arable Farmers

**\* Main topics (Provide the main topics covered in the data collection.)**

food

**\* Abstract\***

Provide a brief summary (max. 300 words) of the main aims and objectives of the research project or alternative process e.g. administrative function, from which the data collection arose.

The research developed a small scale, warmwater production system for growing the fish, tilapia, as a diversification strategy for UK agricultural and arable farmers. A multidisciplinary approach incorporating Marketing, Public Health, Aquaculture and Entrepreneurial perspectives shaped the research which developed a 'green' production system for this fish. The system utilises on-farm resources and

**Related data collection**

If the data collection is derived from or is closely related to others, list details. If any of these data collections are available from the UKDA, AHDS History or ESDS Qualidata, indicate the study number (if known).

**Edition/extract/version (Is this a new edition, extract or special version of the data collection?)**

Yes. Provide details below.  No

**URL**

Is there a web site containing information relevant to the data collection that can be linked to?

http://www.aqua.stir.ac.uk/Systems/tilapiaProject.htm

Internet 100%

# How to create data documentation?

- Researchers can embed certain documentation in data files:
  - quantitative data: variable/value labels; worksheet information; table relationships and queries in relational database; GIS data layers/tables
  - Examples:
    - SPSS: variable attributes documented in Variable View (label, code, data type, missing values)
    - MS-Access: variable descriptions and attributes documented in Design View; relationships
    - ArcGIS: shapefiles (layers) and tables in geodatabase; metadata created in ArcCatalog.
    - Excel: base worksheet data-related documentation.
  - qualitative data/text documents: interview transcript speech demarcation; respondent details; context in document headings
- Additional documents: final report, published reports, user guide, working paper, publications, lab books

# How to create data metadata?

- Metadata: through data deposit forms/tools
- DDI documentation derived from software packages using DDI tools

<http://tools.ddialliance.org/>

- <http://www.ddialliance.org/codebook/index.html>
- metadata tools: <http://tools.ddialliance.org>
- German Institute for Educational Progress (IQB) – educational data codebooks [www.iza.org](http://www.iza.org)

# TIPS

- Good metadata enables easy discovery of data  
provide meaningful information (titles, descriptions, abstract, keywords)
- Submitting good documentation and metadata enables data archives/data centres to create rich and detailed metadata
- Good data documentation leads to good data re-use
- Detailed yet comprehensive documentation
- Where possible integrate within dataset

# UKDA metadata

- UKDA collects and creates structured metadata for each archived dataset
- created during ingest data processing (Data-level metadata)
  - data dictionaries, format transfer, data listing, ingest processing details and information gathered in 'readme' file for users
- Catalogue record and keyword index (mix of Study-/Data-level metadata - 'Catalogue metadata'. Also contains 'Administrative metadata', such as access conditions, date of publication, etc.)
  - data deposit form
  - keyword index covers data elements and concepts
  - international standards: DDI, METS, ISAD(G), TEI
  - standardised elements + controlled vocabularies = consistent search and retrieval
  - sufficient information for users to decide if the data suitable
  - information on the provenance of a dataset
  - record of publications

# Providing good documentation

- quality of the information provided by the data creator determines ease of discovery and appropriate re-use
  - comprehensive and comprehensible documentation and metadata
  - complete the deposit form as fully as possible
- contact the UKDA if not sure what to produce or provide:
  - see advice on our Managing and Sharing web pages:  
<http://www.data-archive.ac.uk/sharing/metadata.asp>
  - contact [acquisitions@esds.ac.uk](mailto:acquisitions@esds.ac.uk)

# Recap – why document data?

- enables you to understand/interpret data
- needed to make data independently understandable
- ensures informed and correct use, reduces chance of incorrect use/misinterpretation
- if using your data for the first time, what would you need to know?

# Examples

- English Longitudinal Study of Ageing (ELSA) – very large study
- Quantitative dataset – depends on size and scale
  - Health Survey for England (HSE)
  - BHPS provides link to documentation site
  - smaller scale study, less documentation
- Qualitative dataset – depends on size and scale
  - data listing, interview schedules, methodology

# ELSA documentation

## Documentation:

Format	Name	Size in Kilo Bytes	Description
EXCEL	<a href="#">5050 Wave 1 Financial Derived Variables Relationships.xls</a>	55	Wave 1 Financial Derived Variables Relationships
EXCEL	<a href="#">5050 Wave 2 Financial Derived Variables Relationships.xls</a>	60	Wave 2 Financial Derived Variables Relationships
EXCEL	<a href="#">5050 Wave 2 Pension Grid Corresponding Variables.xls</a>	177	Wave 2 Pension Grid Corresponding Variables
EXCEL	<a href="#">5050 Wave 3 Financial Derived Variables Relationships.xls</a>	63	Wave 3 Financial Derived Variables Relationships
PDF	<a href="#">5050 Applying for ELSA Genetics Data.pdf</a>	124	Guide to Application for use of ELSA Genetic Data
PDF	<a href="#">5050 ELSA Wave 3 Life History Documentation.pdf</a>	2156	Wave 3 Life History Documentation
PDF	<a href="#">5050 Index File User Guide.pdf</a>	469	ELSA Index File User Guide
PDF	<a href="#">5050 User Guide for Geographical Variables.pdf</a>	193	User Guide for Geographical Variables
PDF	<a href="#">5050 Wave 0 User Guide.pdf</a>	487	Wave 0 User Guide
PDF	<a href="#">5050 Wave 1 Technical Report.pdf</a>	525	Wave 1 Technical Report
PDF	<a href="#">5050 Wave 1 User Guides and Documentation.pdf</a>	3255	Wave 1 User Guides and Documentation
PDF	<a href="#">5050 Wave 2 Technical Report.pdf</a>	657	Wave 2 Technical Report
PDF	<a href="#">5050 Wave 2 User Guides and Documentation.pdf</a>	5939	Wave 2 User Guides and Documentation
PDF	<a href="#">5050 Wave 3 Phase 2 User Guides and Documentation.pdf</a>	3673	Wave 3 Phase 2 User Guides and Documentation
PDF	<a href="#">ELSA data and documentation currently available from UKDA.pdf</a>	162	ELSA Data and Documentation Currently Available
HTML	<a href="#">UKDA Study 5050 Information.htm</a>	49	Study information and citation

# Quantitative study

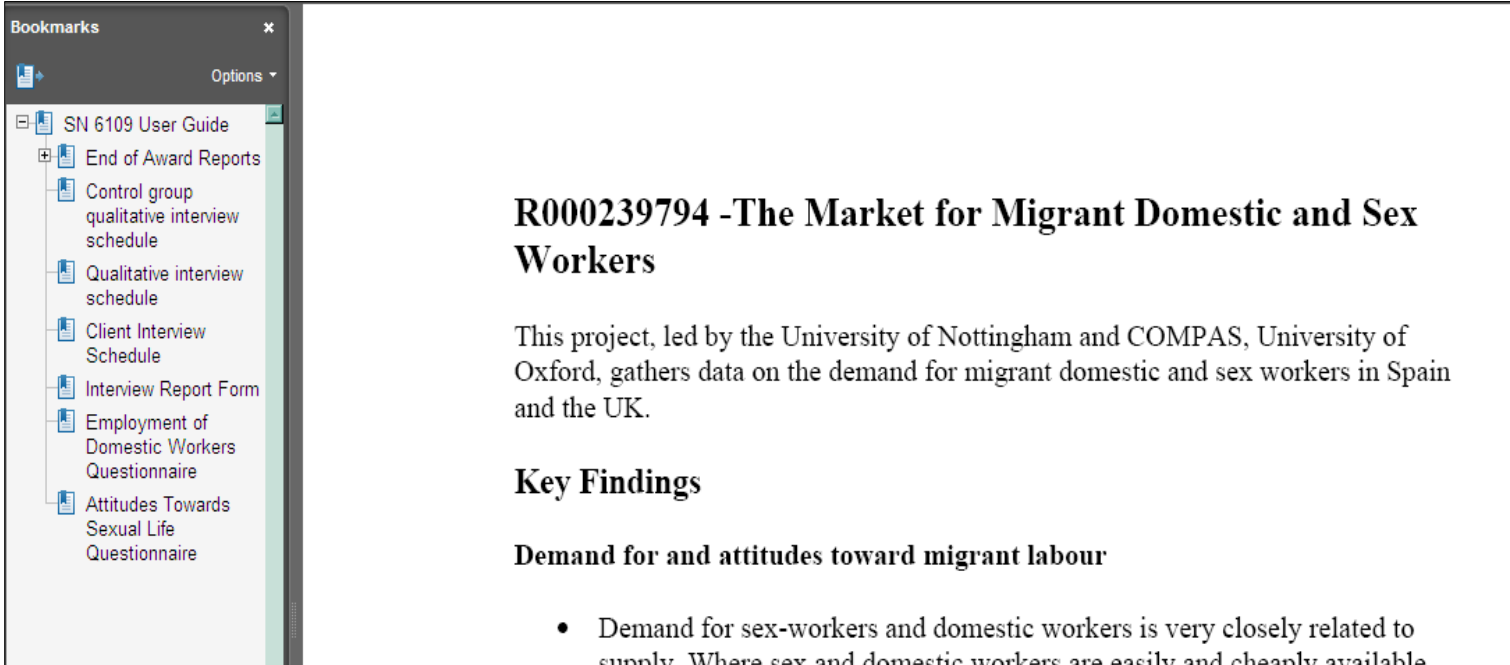
- smaller-scale study - user guide may just contain survey questionnaire, methodology information
- example from HSE 2007 – documents separated, bigger study

## Documentation:

Format	Name	Size in Kilo Bytes	Description
PDF	<a href="#">6112datadocs.pdf</a>	808	Lists of Variables and Derived Variables
PDF	<a href="#">6112interviewingdocs.pdf</a>	1655	Questionnaires, Showcards, Coding Frames and Consent Booklet
PDF	<a href="#">6112supportingdocs.pdf</a>	1573	Interviewer, Nurse, Coding and Editing Instructions
PDF	<a href="#">6112userguide.pdf</a>	288	User Guide
HTML	<a href="#">UKDA Study 6112 Information.htm</a>	25	Study information and citation

# Qualitative study 1

- User guide contains variety of documents



The screenshot shows a document viewer interface. On the left is a 'Bookmarks' sidebar with a tree view containing the following items:

- SN 6109 User Guide
  - End of Award Reports
    - Control group qualitative interview schedule
    - Qualitative interview schedule
    - Client Interview Schedule
    - Interview Report Form
    - Employment of Domestic Workers Questionnaire
    - Attitudes Towards Sexual Life Questionnaire

The main document area displays the following content:

## R000239794 -The Market for Migrant Domestic and Sex Workers

This project, led by the University of Nottingham and COMPAS, University of Oxford, gathers data on the demand for migrant domestic and sex workers in Spain and the UK.

### Key Findings

#### Demand for and attitudes toward migrant labour

- Demand for sex-workers and domestic workers is very closely related to supply. Where sex and domestic workers are easily and cheaply available

# Qualitative study 2

- Data Listing

Study Number 6109 1 of 8  
 Market for Migrant Domestic and Sex Workers, 2002-2006  
 O'Connell Davidson, J.

Name	Age	Gender	Nationality	Income	Marital Status	Ethnicity	Interview	Language
Pep	36	Male	Spanish/Catalan	Medium	Single	White	Face-to-face transcript	English translation
Jordi	34	Male	Spanish/Catalan	Medium	Single	White	Face-to-face transcript	English translation
Erik	38	Male	French	Medium	Divorced (co-habiting)	White	Face-to-face transcript	English translation
Enrique	35	Male	Spanish/Catalan	Medium/high	Married	White	Face-to-face transcript	English translation
Sergio	31	Male	Spanish	Medium	Single	White	Face-to-face transcript	English translation
~	20	Male	British (Guernsey)	Middle	Single	White	Notes	English
Mr and Mrs		Male and ~ Female	~	~	Married	~	Face-to-face transcript/notes	English
K and D		Male and ~ Female	~	~	Married	~	Notes	English
	27-40	Female		High	Married	White	Face-to-face transcript	English (some Spanish)