Understanding Society
THE UK HOUSEHOLD LONGITUDINAL STUDY

Understanding Society, Processing Data, Challenging DDI

A presentation to 3rd Annual European DDI Users Group Meeting (EDDI11): DDI - The Basis of Managing the Data Life Cycle, Gothenburg. 5-6 Dec 2011.

Randy Banks (randy@essex.ac.uk)
ISER, University of Essex
Colchester
UK
CO4 3SQ

An initiative by the Economic and Social Research Council, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by the National Centre for Social Research.

Objectives

- Case Study – Understanding Society (UKHLS)
  - ➢ Capture and reuse of metadata
  - ➢ Important, but limited use of DDI-Codebook
- Question
  - ➢ ISER is long-term supporter of DDI and objectives, e.g. SRN (2009)
  - ➢ UKHLS should have provided unique opportunity to implement DDI-Lifecycle
  - ➢ *Why has ISER not done so?*
- Generalise lessons learned and consider challenges they raise for DDI-Lifecycle

Understanding Society (UKHLS)

- Household panel study designed to be largest of its kind
- Conducted by Institute for Social and Economic Research (ISER)
- Core funded by Economic and Social Research Council (ESRC) and Government’s Large Facilities Capital Fund
  - ➢ Largest single investment by ESRC
  - ➢ Additional funding from government departments
- Grant awarded April 2007. Fieldwork commenced Jan 2008.
- Replaces British Household Panel Study (BHPS)
  - ➢ 18 waves between 1991 and 2008.
  - ➢ BHPS sample incorporated into UKHLS at wave 2

BHPS v UKHLS. Similarities

- Annual household panel
  - Panel of individuals in changing household context
  - All household members may be followed subject to following rules
- UK-wide coverage
  - England, Scotland, Wales (Great Britain). Northern Ireland
- ISER manages project and data post-field, but competitively contracts questionnaire implementation and fieldwork to external provider
  - BHPS - GfK NOP (GB) and NISRA (NI)
  - UKHLS – NatCen, in consortium with NISRA

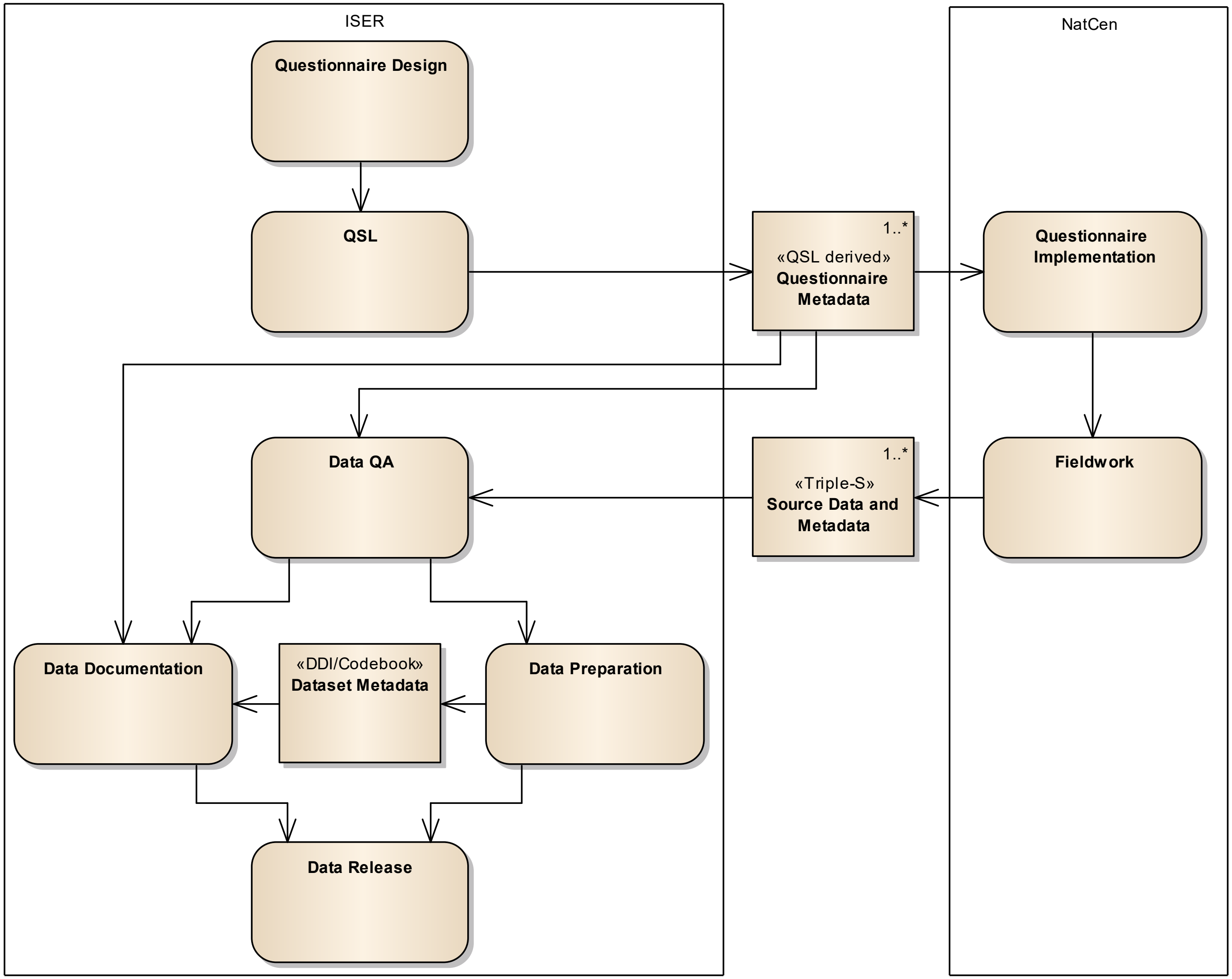
BHPS v UKHLS. Differences

	BHPS	UKHLS
Surveys	1	2 (Innovation Panel and mainstage)
Initial Sample Size (Target)	5,000	40,000 (In total)
Sample composition (mainstage)	General population	General population + ethnic minority boost + BHPS (at wave 2)
Fieldwork	4 months, September to December	24 months, January to December (mainstage), April (Innovation Panel)
Sample allocation	Block	Monthly
Coverage at start	Great Britain	United Kingdom
W1 lead time (excluding pilots)	28 months	8 months (to Innovation Panel)
Biomarker collection?	No	Yes (mainstage, as of wave 2)
Funder expectations	High	Higher

Data Processing. Challenges

- Hire new staff
- Complete re-write of data processing and sample management systems to meet new and, compared to BHPS, more extensive requirements of UKHLS
    - Maximise metadata capture for process control and documentation
- Establish working relationship with new fieldwork agency
- Bring BHPS to a close and incorporate sample
- Maintain sanity
- *etc*

Data Lifecycle. The short version.



Metadata Creation. ‘QSL’

- Questionnaires specified using in-house ‘Questionnaire Specification Language’ (QSL)
  - See Costigan and Elder (2003) for importance of and difficulties in achieving specification of electronic instruments
  - Modelled (loosely) on Blaise, but CAI-independent
  - As ISER competitively contracts questionnaire implementation, can’t guarantee the system(s) to be used
  - Semantics based on language of questionnaire designers
  - Plain-text, procedural for development speed
  - Modular, use of inheritance for development efficiency
- QSL scripts parsed and translated to XML and then repurposed as required
  - For consultation, specification, documentation, process control
  - Could generate (draft) code, but not yet
- Gradually came on stream as of Mainstage wave 1
- Continuing ‘work-in-progress’
  - Additional functionality incorporated according to need and practicality(not necessarily in that order)
- Examples at http://iserwww.essex.ac.uk/home/randy/ddi/events/2011-12-05-eddi11/

Data Exchange. Triple-S

- Triple-S is simple XML-based data and metadata exchange format
- Originated in the market research community
- Provides interface between fieldwork agency and in-house systems
- SPSS outputs from field agency are transformed into Triple-S using third-party, open-source converter
- In-house transformation script transforms (and enhances) Triple-S metadata into schema and import commands for loading into SIR/DBMS

Metadata Transfer. DDI-Codebook

- Transfer dataset metadata from data processing to documentation system
- In-house application exports SIR/DBMS metadata into (slightly tweeked) DDI 2.1 format
  - ‘long’ and ‘short’ labels in <recgrp>
  - ‘units’ attributes for string variables
  - Conditional interpretation of ‘fileid’
  - Generation of <recgrp>s when documenting rectangular files
  - List of ‘keyvar’s rather than single value

Why not DDI-Lifecycle?

- Originally, a matter of timing
  - DDI 3.0 published in April 2008; UKHLS started April 2007
  - Had to create our own questionnaire metadata definition and capture tool
- Subsequently, a matter of doing what had to be done as quickly and efficiently as possible in response to rapidly-changing circumstances and requirements
  - DDI lifecycle has high cost of entry with both long learning and implementation curve
  - Needed operational systems quickly
  - Made use of what was available
    - Triple S is relatively simple by design and had existing SPSS conversion tools
    - Could've used Triple-S to transfer dataset metadata, but already had alpha version of SIR to DDI 2.1 converter – could bring it up to operational status relatively quickly

DDI-Lifecycle for the future?

- No concrete plans as yet
- Two main strategic questions
- Where and how to start?
  - ➢ DDI-Lifecycle will have to be gradually integrated into operational systems
    - – lack of tools and guidance as to how best to integrate into ongoing system
    - – DDI tends to be presented as an all-or-nothing package, e.g. section 9 in (DDI, 2009): ‘Step-by-Step Sequence to Create a DDI File for a Simple Instance’
- How to make the business case?
  - ➢ Entry level costs will be high and front-loaded
    - – funding is severely constrained in current economic climate
  - ➢ Benefits will be long term
    - – difficult to persuade business leaders in the abstract when second best is ‘good enough’

Lessons for DDI-Lifecycle

- To (mis-) quote von Moltke:
  - No DP strategy survives the first encounter with the data
  - DP strategy is a system of expedients
  - *Must be able to incorporate components of DDI-Lifecycle as required and into on-going systems*
- Business requirements have only local plateaus
  - Will want more, more quickly and at lower cost
  - *Costs of implementing DDI-Lifecycle must be reduced*
- DDI Lifecycle may be the best standard for metadata capture and exchange, but not the only one
  - Decision to implement it depends on factors other than intrinsic merits
  - *Decision makers must be sold on DDI-Lifecycle's Unique Selling Point*

DDI-Lifecycle. Challenges

- Technical
  - More training materials
  - Guidance on how DDI components can be independently used and integrated into on-going systems
    - Enable practitioners to start with a snack rather than a 10-course meal
    - address not only ‘best’, but ‘actual’ practice
  - More and more flexible functional tools
    - to support disaggregated use
- Business
  - Sell the lifecycle *concept*
    - DDI as the best platform for achieving integration
  - ... to data producers
    - DDI initiated by archives; DDI/Codebook reflects downstream positioning
    - DDI/Lifecycle must be embedded at the point of production
  - ... focussing on funders and research leaders
    - Follow the money!

References

- BHPS - http://www.iser.essex.ac.uk/bhps.
- P Costigan and S Elder. ‘Does the Questionnaire Implement the Specification? Who Knows? ‘In R Banks *et al* (eds). *The Impact of Technology on the Survey Process. Proceedings of the ASC’s Fourth International Conference on Survey and Statistical Computing.* ASC. 2003
- DDI (2009) - Technical Specification User Guide. Part II, Version 3.1. Data Documentation Initiative (DDI), Oct 2009
- SRN (2009) – Banks et al. *A Feasibility Study to Investigate Integrated Survey Data Collection, Fieldwork Management and Survey Data Processing Systems for Longitudinal Studies. Final Report.* (http://www.surveynet.ac.uk/sms/srn_objective_5_final_report.pdf)
- Triple-S - http://www.triple-s.org/
- Understanding Society - http://www.understandingsociety.org.uk/