70 Years of UK Birth Cohort Data into DDI Lifecycle?

Claude Gierl

Jon Johnson

www.ioe.ac.uk

Contents

Centre for Longitudinal Studies

The new CLOSER Project

The Input Metadata

Our experience so far:

    Pre-processing

    The Metadata Editor

    Colectica

The benefits of DDI3.2

Centre for Longitudinal Studies

Funded as an ESRC Resource Centre since 2004

Manages three of the UK's national birth cohort studies
    1958 Birth Cohort (National Child Development Study)
    1970 Birth Cohort
    Millennium Cohort Study

Anonymised data available via the UK Data Service
www.esds.ac.uk

CLOSER(Cohort and Longitudinal Studies Enhancement Resources)

Launched 2012, £5 million, 5 Years

9 Participating Studies:

Hertfordshire Cohort Study, 1946 British Birth Cohort,

1958 British Birth Cohort, 1970 British Birth Cohort, ALSPAC,

Southampton Women's Study, Understanding Society,

Life Study, Millennium Cohort Study

A Unified Search Platform

Search and discovery, study comparison

The Input Metadata

250 survey instruments (90% paper)

250,000 data variables

A variety of collection methods (paper, phone, CAI...)

Linked data

30% DDI2.0 xml from ESDS for higher level elements

Metadata Management Scripts

For non-paper inputs and where we have partial electronic data

Suite of scripts written in Python

Parsers

SPSS Interface

Pre-processing

Filters

Quality Control

Pre-populate the database

Metadata Editor

For the capture of paper records

In-house software application

Subset of DDI3 mapped to about 25 database tables (sqlite3)

Ruby on Rails user interface (web-based)

Exports DDI3 compliant files
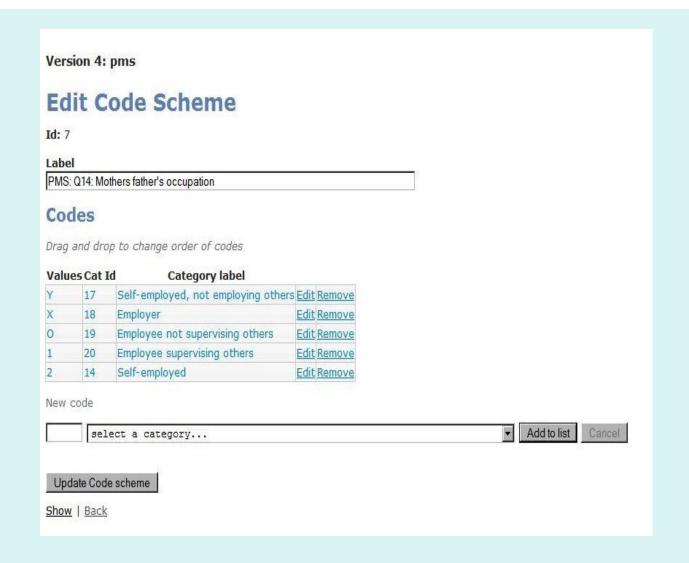
Simple to use, can be extended

Version 4: pms

Questionnaire

- DDI3 Instance: show | edit | xml
- Spine view
- Table of all control constructs
- Tables by control construct type:
    - Statements
    - Sequences
    - Loops
    - Conditions
    - Question Constructs
        - Universes
        - Question items
            - Response Domains by type:
                - Response Domains (Text)
                - Response Domains (Numeric)
                - Response Domains (Datetime)
                - Response Domains (Codes)
                    - Code Schemes
                        - Categories

8

Version 5: ncds_biomed

Spine

- 1 *sequence:* TopSequence
  - 2 *sequence:* Feed Forward
  - 3 *sequence:* Introduction
    - 20 *question:* qstart1
    - 21 *question:* nurdate
    - 22 *question:* machdate
    - 23 *question:* dateok
    - 24 *question:* strtime
    - 25 *question:* dob
    - 26 *question:* page
    - 27 *question:* name
    - 28 *question:* psex
    - 29 *question:* dmdisab
    - 30 *question:* dlastint
    - 31 *question:* drugdone
    - 32 *question:* bcmsex
    - 33 *question:* bbdat2
    - 34 *question:* cmname
    - 35 *question:* recname
    - 36 *question:* special1
  - 4 *sequence:* Explanation and Consent To Measurements
    - 37 *question:* measq
    - 38 *question:* will1
    - 39 *question:* cidiwill

Using Colectica

Overall metadata management

Input: **DDI3 from the metadata editor**
            Text such as variable-to-question link tables
            Data (SPSS)

SDK scripts for restructuring

Generate codebooks and documentation

The benefits of DDI3.2

Categories: concepts

Code List Schemes

Missing values

Complex Questions: old questionnaires, PMS examples

SECTION III.—Please note carefully; the information in this section is to be got from records if at all possible. If this is not possible, get details from mother.

PAST OBSTETRIC HISTORY.—Exclude present pregnancy.

24 Has the patient had any previous pregnancies (including miscarriages)? Yes _____ Y
No _____ X

If "Yes" please give details below, taking the pregnancies in order of occurrence (the earliest first). Record twins as two separate births.

Pregnancy Number	Date of Delivery		Sex		Birth Weight	Place of Delivery		Outcome of Delivery								Complications of Pregnancy						Method of Delivery				
								Livebirth																		
	Month	Year	Male	Female		Domiciliary	Institutional (including nursing homes)	Alive now	Died 28 days or later	Died 7-27 days inclusive	Died under 7 days	Stillborn	Miscarriage	Ectopic Pregnancy	Toxaemia	A.P.H.	Other complications	No complications at all	Not known whether any complications or not	Spontaneous	Forceps	Caesarean	Others	Method not known	
1			Y X		___lbs. ___ozs.	0	1	2	3	4	5	6	7	8	Y X 0		1		2	3	4	5	6	7	
2			Y X		___lbs. ___ozs.	0	1	2	3	4	5	6	7	8	Y X 0		1		2	3	4	5	6	7	
3			Y X		___lbs. ___ozs.	0	1	2	3	4	5	6	7	8	Y X 0		1		2	3	4	5	6	7	
4			Y X		___lbs. ___ozs.	0	1	2	3	4	5	6	7	8	Y X 0		1		2	3	4	5	6	7	
5			Y X		___lbs. ___ozs.	0	1	2	3	4	5	6	7	8	Y X 0		1		2	3	4	5	6	7	
6			Y X		___lbs. ___ozs.	0	1	2	3	4	5	6	7	8	Y X 0		1		2	3	4	5	6	7	
7			Y X		___lbs. ___ozs.	0	1	2	3	4	5	6	7	8	Y X 0		1		2	3	4	5	6	7	
8			Y X		___lbs. ___ozs.	0	1	2	3	4	5	6	7	8	Y X 0		1		2	3	4	5	6	7	
9			Y X		___lbs. ___ozs.	0	1	2	3	4	5	6	7	8	Y X 0		1		2	3	4	5	6	7	
10			Y X		___lbs. ___ozs.	0	1	2	3	4	5	6	7	8	Y X 0		1		2	3	4	5	6	7	

13

23 Did the patient smoke as many as one cigarette a day during the 12 months before the start of this pregnancy? If so, how many per day during that period?

Did not smoke as many as 1 per day................. Y

Number smoked per day in that period.........

If smoked one or more per day

(a) Did the patient change her smoking habits during this pregnancy? Record any changes in the table below—and month of pregnancy change made.

No change .... Y

	Month of Pregnancy Changed	Number per day smoked after change
Gave up		0
Cut down		
Increased		

14

Any questions?

www.cls.ioe.ac.uk

www.closer.ac.uk

Institute of Education
University of London
20 Bedford Way
London WC1H 0AL

Tel +44 (0)20 7612 6000
Fax +44 (0)20 7612 6126
Email info@ioe.ac.uk
Web www.ioe.ac.uk