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
Version 4: pms

Edit Code Scheme

Id: 7

Label
PMS: Q14: Mothers father's occupation

Codes

Drag and drop to change order of codes

<table>
<thead>
<tr>
<th>Values</th>
<th>Category label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>Self-employed, not employing others</td>
</tr>
<tr>
<td>Y</td>
<td>Employer</td>
</tr>
<tr>
<td>Y</td>
<td>Employee not supervising others</td>
</tr>
<tr>
<td>Y</td>
<td>Employee supervising others</td>
</tr>
<tr>
<td>Y</td>
<td>Self-employed</td>
</tr>
</tbody>
</table>

New code

select a category...

Update Code scheme

Show | Back
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
### PAST OBSTETRIC HISTORY

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.

<table>
<thead>
<tr>
<th>Pregnancy Number</th>
<th>Date of Delivery</th>
<th>Sex</th>
<th>Birth Weight</th>
<th>Place of Delivery</th>
<th>Outcome of Delivery</th>
<th>Complications of Pregnancy</th>
<th>Method of Delivery</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Livebirth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Month of Pregnancy Changed</th>
<th>Number per day smoked after change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gave up</td>
<td>0</td>
</tr>
<tr>
<td>Cut down</td>
<td></td>
</tr>
<tr>
<td>Increased</td>
<td></td>
</tr>
</tbody>
</table>
Any questions?

www.cls.ioe.ac.uk
www.closer.ac.uk