Enhancing the quality of metadata: harmonising with DDI-Lifecycle
Alina Danciu, Alexandre Mairot

To cite this version:
Alina Danciu, Alexandre Mairot. Enhancing the quality of metadata: harmonising with DDI-Lifecycle. 8th Annual European DDI User Conference (EDDI16), Dec 2016, Cologne, Germany. hal-02874798

HAL Id: hal-02874798
https://sciencespo.hal.science/hal-02874798
Submitted on 3 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License
Enhancing the quality of metadata: harmonising with DDI-Lifecycle

EDDI16 – 8th Annual European DDI User Conference
December, 6 2016, Köln

Alina Danciu, Alexandre Mairot
Summary

- Including DDI-Lifecycle specifications in our practices
  - Dealing with DDI-Lifecycle
  - A new approach including DDI-Lifecycle

- Harmonisation
  - Data and metadata
  - Procedures and organisation

- Feedback on the harmonisation process
Dream big and make it happen. But how?
Integrating DDI-Lifecycle with existing data systems

- Versioning
- Uniqueness of entries
- Granularity
A new approach including DDI-Lifecycle

- 2014 : How to integrate the DDI-Lifecycle specifications in our dissemination practices?
- 2015 : How about the creation of our metadata system?
- 2016 : Process of building a metadata database
Data/metadata harmonisation

- Two types of data in the datasets: string and numeric
  - Date format: string with ISO8601 format (yyyy-mm-dd)
- Six types of missing values

<table>
<thead>
<tr>
<th>CodeList</th>
<th>Label</th>
<th>Code</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not applicable</td>
<td>6..</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Refuse</td>
<td>7..</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Don’t know</td>
<td>8..</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>No answer</td>
<td>9..</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Filter error /missing data</td>
<td>99..</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Not surveyed</td>
<td>99..6</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

- Recoded response items are written in parentheses

**ea14_D4_rec**: Number of rooms

1. 1 room
2. 2 rooms
... ......
7. 7 rooms
8. 8 rooms
9. 9 rooms and more
99. No answer
999. Missing data
9996. Not surveyed

**ea14_A2A**: Household’s total income

1. Less than 400 €
... ...
12. 4 000 to 5 999 €
13. 6 000 to 9 999 €
14. 10 000 € and more
... ......

**ea14_A2A_rec**: Household’s total income

3 (Less than 800 €)
... ...
12. 4 000 to 5 999 €
13 (6 000 € and more)
... ......
Variable harmonisation across studies

- **Variable id**: intpol
- **Variable label**: How interested in politics
- **Variable urn**: urn:ddi:fr.cdsp:Variable.intpol:1

<table>
<thead>
<tr>
<th>Study</th>
<th>Variable</th>
<th>&lt;l:VariableName&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image de la science 2001</td>
<td>int</td>
<td>ids01_INT</td>
</tr>
<tr>
<td>Panel électoral français 2002 vague 1</td>
<td>xq201</td>
<td>pef0201_XQ201</td>
</tr>
<tr>
<td>Enquête interrégionale des phénomènes politiques 1998</td>
<td>q29</td>
<td>eipp98_Q29</td>
</tr>
<tr>
<td>Enquête interrégionale des phénomènes politiques 1999</td>
<td>q26</td>
<td>eipp99_Q26</td>
</tr>
<tr>
<td>Enquête annuelle ELIPSS 2013</td>
<td>i1</td>
<td>ea13_I1</td>
</tr>
<tr>
<td>Enquête DYNAMOB vague 1 (2013–09)</td>
<td>intpol</td>
<td>dn01_INTPOL</td>
</tr>
<tr>
<td>Enquête DYNAMOB vague 2 (2013–12)</td>
<td>intpol</td>
<td>dn02_INTPOL</td>
</tr>
</tbody>
</table>

- **QuestionReference urn**: urn:ddi:fr.cdsp:QuestionItem.intpol:1
- **QuestionText**: How interested would you say you are in politics? Are you?
Variable harmonisation across waves

- **Variable id**: g5a_rec (recoded variable)
- **Variable label**: Foreign language spoken
- **Wave**: 2013 (ea13), 2014 (ea14), 2015 (ea15), 2016 (ea16)
- **Variable categories updated for each wave (dataset + documentation):**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Arabic</td>
<td>1 Arabic</td>
<td>1 Arabic</td>
<td>1 Arabic</td>
</tr>
<tr>
<td>2 German</td>
<td>2 German</td>
<td>2 English</td>
<td>2 English</td>
</tr>
<tr>
<td>3 English</td>
<td>3 English</td>
<td>3 German</td>
<td>3 German</td>
</tr>
<tr>
<td>4 Portuguese</td>
<td>4 Portuguese</td>
<td>4 Italian</td>
<td>4 Italian</td>
</tr>
<tr>
<td>5 Italian</td>
<td>5 Italian</td>
<td>5 Portuguese</td>
<td>5 Portuguese</td>
</tr>
<tr>
<td>6 Spanish</td>
<td>6 Spanish</td>
<td>6 Spanish</td>
<td>6 Spanish</td>
</tr>
<tr>
<td>7 Other</td>
<td>(Other)</td>
<td>(Other)</td>
<td>(Other)</td>
</tr>
<tr>
<td>66 Not applicable</td>
<td>66 Not applicable</td>
<td>66 Not applicable</td>
<td>66 Not applicable</td>
</tr>
</tbody>
</table>
Procedures and organisation

- DDI-Lifecycle profile defined in 2015
- DDI-Codebook template linked to DDI-Lifecycle profile
- Rules of metadata versioning:
  - Same data in several datasets ➞ Same metadata
  - Same variables in several datasets ➞ new version for the metadata were different
Some thoughts on the harmonisation process

- For datamanagers
  - Re-use of metadata
  - Quality and reliability of the metadata
  - Rethink the documentation process and workflow

- For users
  - Data more discoverable and usable
  - Visualisation of all the waves including the question and download the ones of interest
  - Access conditions tend to be more flexible

http://quanti.dime-shs.sciences-po.fr/

alina.danciu@sciencespo.fr
alexandre.mairot@sciencespo.fr