



HAL
open science

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

CodeList Label	Code	Level
Not applicable	6..	1
Refuse	7..	2
Don't know	8..	3
No answer	9..	4
Filter error /missing data	99..	5
Not surveyed	99..6	6

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

- Recoded response items are written in parentheses

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 accross studies

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

Study	Variable	<I:VariableName>
Image de la science 2001	int	ids01_INT
Panel électoral français 2002 vague 1	xq201	pef0201_XQ201
Enquête interrégionale des phénomènes politiques 1998	q29	eipp98_Q29
Enquête interrégionale des phénomènes politiques 1999	q26	eipp99_Q26
Enquête annuelle ELIPSS 2013	i1	ea13_I1
Enquête DYNAMOB vague 1 (2013-09)	intpol	dn01_INTPOL
Enquête DYNAMOB vague 2 (2013-12)	intpol	dn02_INTPOL

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

Variable harmonisation accross 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):**

version 2013

1	Arabic
2	German
3	English
4	Portuguese
5	Italian
6	Spanish
7	Other
66	Not applicable

version 2014

1	Arabic
2	German
3	English
4	Portuguese
5	Italian
6	Spanish
7	(Other)
66	Not applicable

version 2015

1	Arabic
2	English
3	German
4	Italian
5	Portuguese
6	Spanish
7	(Other)
66	Not applicable

version 2016

1	Arabic
2	English
3	German
4	Italian
5	Portuguese
6	Spanish
7	(Regional language)
8	(Other)
66	Not applicable

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://cdsp.sciences-po.fr>

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

alina.danciu@sciencespo.fr

alexandre.mairot@sciencespo.fr