The Advantages and Disadvantages of Three Software Solutions for the Development of a DDI-based Survey Catalogue

Arianna Caporali (Ined DataLab)

arianna.caporali@ined.fr

12th Annual European DDI User Conference (EDDI20)

30 Nov-2 Dec 2020, Online (France)

License: CC BY 4.0 (exceptions: see last slide)

The French Institute for Demographic Studies (INED)

- A public research institute specialized in population studies structured in research units and support departments

- 8 main research themes: couple, family & sexuality; life expectancy, mortality & health; international migrations, discrimination & integration; housing, environment & territory; gender inequalities; the stages of the life cycle, ageing; the world's populations

- Carrying out of data collections (surveys, contextual databases...)

- Created in 2019 to centralise two types of activities:

  1. Development & maintenance of aggregated and contextual databases

  2. Survey data documentation, dissemination & access requests
     - INED surveys
       → INED Nesstar Catalogue
       → PROGEDO-Quetelet Diffusion (French Data Archive) data access platform
     - Generations & Gender Programme (GGP) Nesstar Catalogue

Nesstar catalogues

http://nesstar.ined.fr/webview

- Heterogeneity of themes and years (oldest in the 1940s)
- Target: the largest research community

- Target: GGP users' community
- Longitudinal panel surveys in Europe and beyond on family and life course trajectories

https://www.ggp-i.org/data/browse-the-data/

A four-step study to choose a substitute for Nesstar

1. Identify the characteristics the new tool should meet

2. Identify possible tools to test

3. Evaluate the selected tools against key characteristics

4. Implement and test the tools (only partially accomplished)

1. Looked-for characteristics

- The new tool should
    - Be compatible with Data Documentation Initiative (DDI) standard
    - Be harvestable by the Consortium of European Social Science Data Archives (CESSDA) catalogue
    - For INED surveys only, be compatible with new Quetelet-PROGEDO Diffusion data portal
    - Target the largest research community
    - Installable on INED servers
    - Maintain the strengths of Nesstar
    - Improve the weaknesses of Nesstar

1. Looked-for characteristics
1.1 Nesstar features to maintain/ improve

Features to maintain	*Features to improve*
• Compatible with DDI • Visualization of metadata about fieldworks • Variable tabulations and graphs • Data analysis functionalities (e.g., linear regression and cross-tabulation) • *Nesstar Publisher*, a user-friendly DDI files editor	• Outdated technology • Poor search functionality • Targeted towards a specialized research community

2. Tools considered

- We have considered three tools:

  1) Dataverse (https://dataverse.org/)

  2) NADA Microdata Cataloging Tool (https://nada.ihsn.org/)

  3) Colectica Portal (https://www.colectica.com/)

3. Evaluate the selected tools

3.1 Dataverse…

Pros	Cons
• DDI compliant	• Needs to be customized
• Open source tool	• Needs a well established IT infrastructure
• Supported by Harvard University and CESSDA	• No data analysis
• User-friendly interface and search functionality	• No DDI files editor (Variable metadata can only be edited through the API or an external tool)
• Considered for implementation by French archive (PROGEDO Quetelet Diffusion)	
• Visualization of variable distributions with an external tool	

3. Evaluate the selected tools
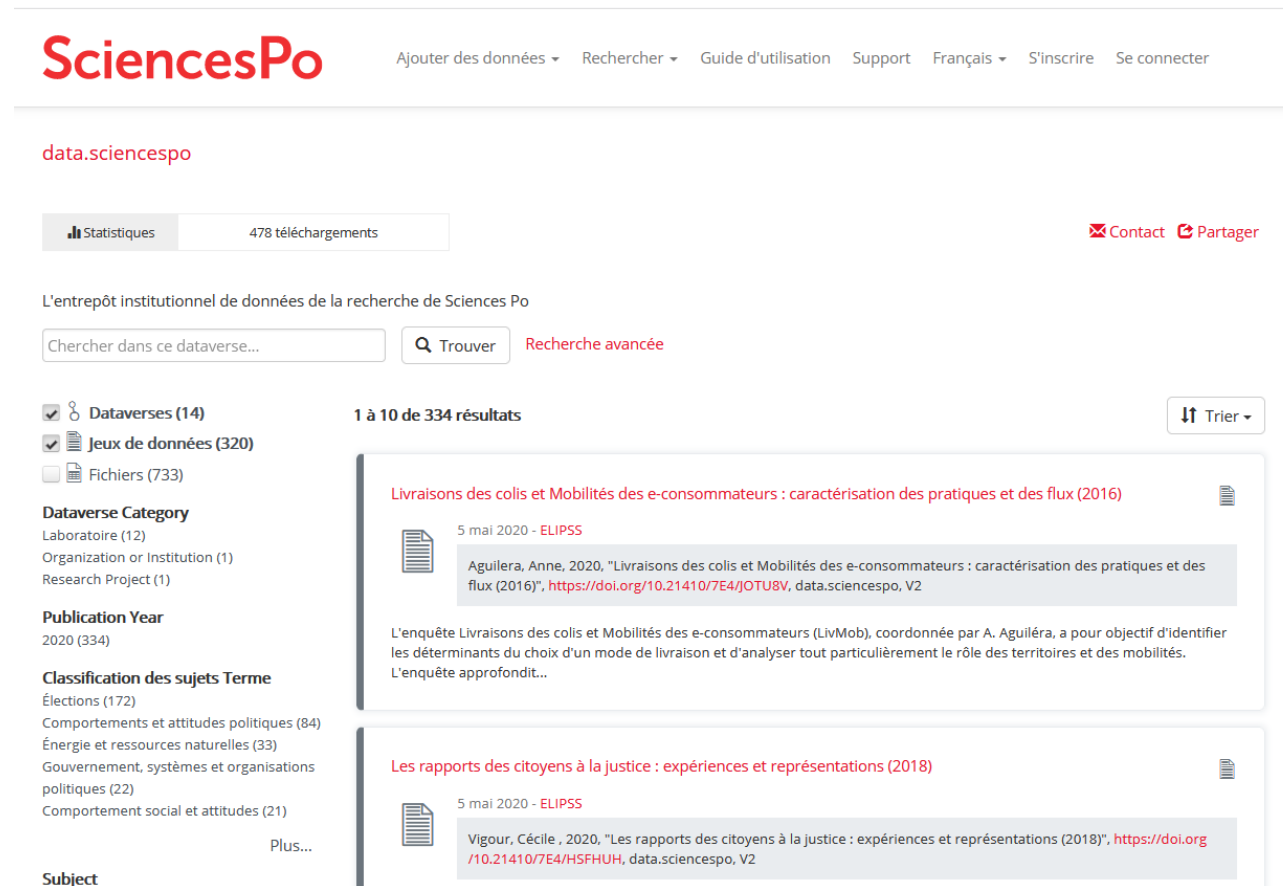
...3.1 Dataverse...

Looked-for characteristics	Evaluation
• Be compatible with DDI	+
• Be harvestable by CESSDA catalogue	+
• Be compatible with Quetelet PROGEDO Diffusion data portal	?
• Target the largest research community	+
• Installable on INED servers	?
• Maintain the strengths of Nesstar (e.g. variable tabulation & data analysis)	+/-
• Improve the weaknesses of Nesstar (e.g. search functionalities)	+

3. Evaluate the selected tools
...*3.1 Dataverse - example*

The Dataverse installation at Sciences Po, Paris:
https://data.sciencespo.fr/dataverse/sciencespo

3. Evaluate the selected tools
3.2 NADA...

Pros	Cons
• DDI compliant	• No data analysis possible
• Free of charges	• No DDI files editor
• Already installed at INED	
• Visualization of variable distributions	
• User-friendly interface and search functionality	
• Compatible with Nesstar DDI version	

3. Evaluate the selected tools
...*3.2 NADA...*

Looked-for characteristics	*Evaluation*
• Be compatible with DDI	+
• Be harvestable by CESSDA catalogue	?
• Be compatible with Quetelet PROGEDO Diffusion data portal	?
• Target the largest research community	+
• Installable on INED servers	+
• Maintain the strengths of Nesstar (e.g. variable tabulation & data analysis)	+/-
• Improve the weaknesses of Nesstar (e.g. search functionalities)	+

3. Evaluate the selected tools

...*3.2 NADA – examples*

The NADA installation at INED:
https://nada.web.ined.fr/index.php/catalog

View of a GGP variable in NADA
(offline test)

3. Evaluate the selected tools
3.3 Colectica Portal…

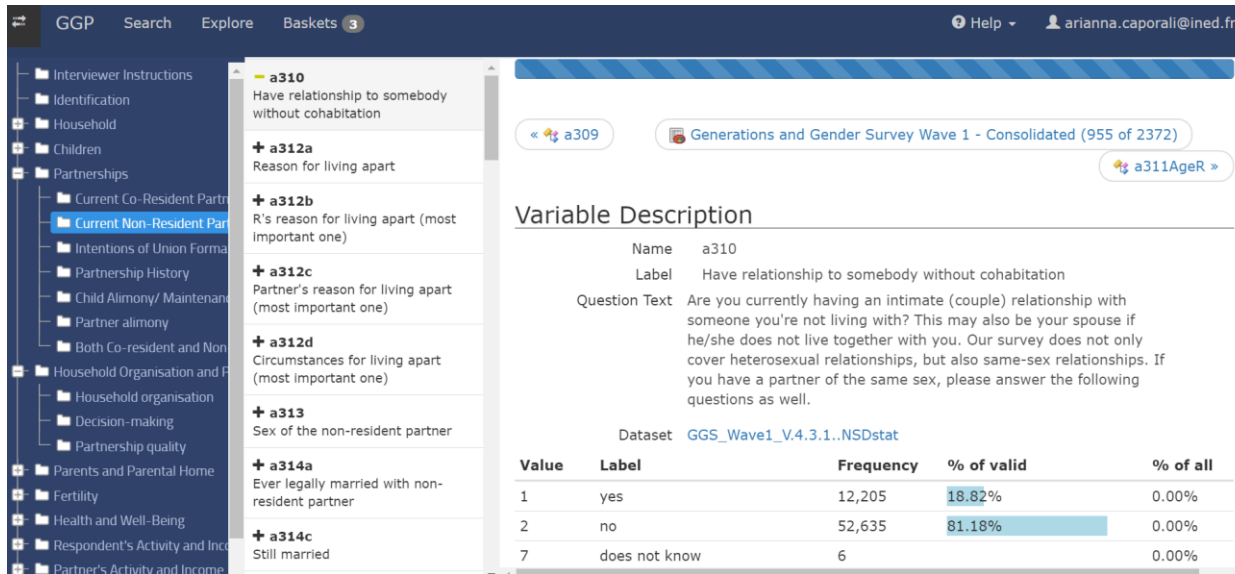| Pros	Cons

• Compatible with DDI version 3.2
  → Comparisons between variables over time for longitudinal surveys such as GGP
• Comes with a DDI files editor, *Colectica Designer,* with interesting features for GGP
  → Conversion of Blaise files to DDI
• Installed on local servers or on a cloud
• Allows to browse variable distributions
• Training workshops for users
	• Commercial product; not open source
	• Features not strictly necessary for simple surveys
	• Resources necessary to take advantage of DDI 3.2 features
	• No data analysis

3. Evaluate the selected tools
...3.2 Colectica Portal...

Looked-for characteristics	Evaluation
• Be compatible with DDI	+
• Be harvestable by CESSDA catalogue	+
• Be compatible with Quetelet PROGEDO Diffusion data portal	?
• Target the largest research community	-
• Installable on INED servers	?
• Maintain the strengths of Nesstar (e.g. variable tabulation & data analysis)	+/-
• Improve the weaknesses of Nesstar (e.g. search functionalities)	+

ined

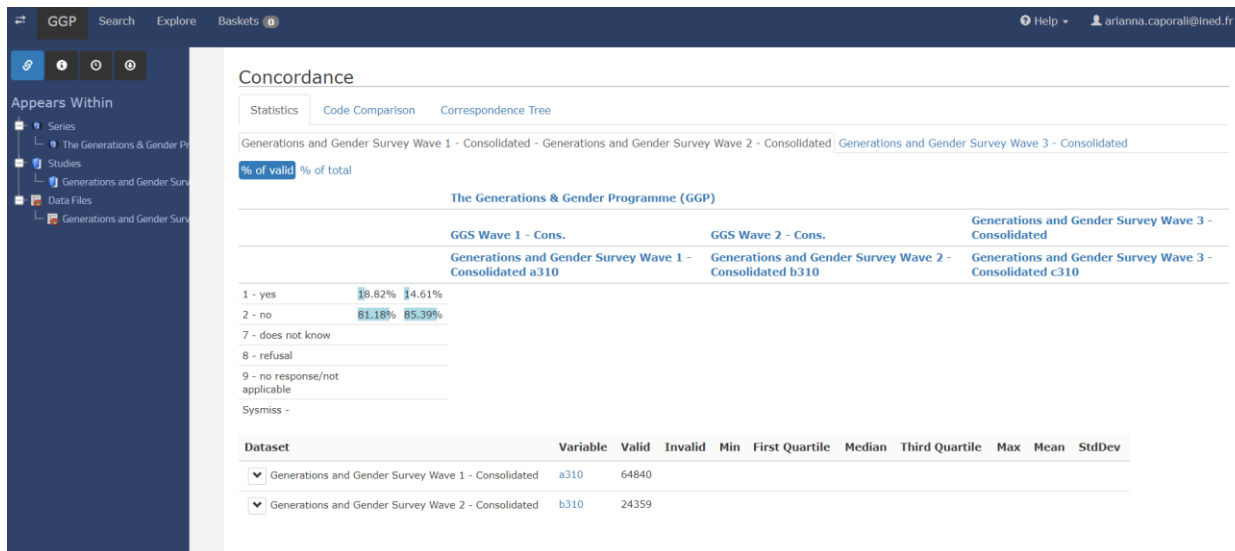# 3. Evaluate the selected tools
...*3.2 Colectica Portal – examples*

View of a GGP variable description (test in a cloud hosted solution)

Comparison of GGP variables a310 (wave 1) and b310 (wave 2)

4. Implementation and test of the tools

- Dataverse

  → Test expected next year

- NADA Microdata Cataloging Tool

  → need to test if harvestable by other data catalogues

- Colectica

  → Designer: test is ongoing for GGP surveys with good results

  → Portal: test done only with the cloud hosted solution with good results

Conclusion

- There is no perfect substitute for Nesstar

- INED survey catalogue will be based on Dataverse or NADA

- We will use a different tool for GGP → Colectica

- *What criteria should guide the choice of a software for a DDI-based survey catalogue* ?

    → contextual criteria:

    ➢ Available resources

    ➢ IT infrastructure

    ➢ Target audience

    ➢ Specific needs (e.g., longitudinal surveys)

License

- This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/ or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA (CC BY 4.0)
- The license does not apply to the following logos:

  - INED logo

  - Ined DataLab logo

  - GGP logo

  - SciencesPo

  - DEMOSTAF logo

Thank you!

Arianna Caporali

arianna.caporali@ined.fr