



ined
INSTITUT
NATIONAL
D'ÉTUDES
DÉMOGRA
PHIQUES

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](https://creativecommons.org/licenses/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

ined **Catalogue Nesstar des enquêtes de l'INED** [Commander les données](#) / [Retour au catalogue](#)

Catalogue des enquêtes de l'INED
 Enquêtes mises à disposition
 Enquêtes non mises à disposition

Bienvenue sur le catalogue des enquêtes de l'INED

Le serveur Nesstar de l'INED permet un accès en ligne, direct et gratuit, aux enquêtes que l'Institut National d'Études Démographiques a réalisées depuis sa création en 1945 (soit un peu plus de 200 références). Aucune inscription n'est nécessaire, et il est possible pour chaque enquête de :

- Consulter la documentation liée à l'enquête
- Parcourir les données documentées
- Voir les tris à plat*
- Réaliser des tris croisés ou analyses simples*
- Télécharger l'ensemble de la documentation
- Télécharger les matériaux de l'enquête (ex : questionnaire)

Pour télécharger des jeux de données, rendez vous sur le [portail des données de réseaux Ouestlet](#) ou cliquez à tout moment de votre navigation sur le lien **"Commander les données"** (en haut à droite de chaque page).

*Les données de certaines enquêtes doivent être pondérées avant de pouvoir être utilisées. Veuillez consulter la documentation associée à chaque jeu de données pour plus d'informations sur la pondération à appliquer.

Aide à la navigation: Le serveur Nesstar de l'INED comporte deux cadres. Le cadre de gauche permet de parcourir les bases de données disponibles et le cadre de droite d'explorer et analyser les données.

Les enquêtes sont divisées en 2 catégories : "les Enquêtes mises à disposition" sont celles pour lesquelles il est possible de commander le(s) fichier(s) de données, tandis que les "Enquêtes non mises à disposition" sont celles où il n'est pas possible actuellement de commander le fichier de données (soit parce que celui-ci n'existe pas, soit parce qu'il n'est pas encore mis à disposition).

Pour rechercher un terme ou un sujet précis, utilisez les fonctions de recherche situées en haut à gauche de votre écran (sous le logo de l'INED).

Pour plus d'aide à la navigation, consultez [notre guide d'utilisation de Nesstar](#) ou cliquez à n'importe quel moment sur l'icône? située en haut à droite de votre écran.

Ce catalogue est géré par l'équipe "Mise à disposition" du service des enquêtes et sondages de l'INED.
Contact: enquetes@ined.fr

<http://nesstar.ined.fr/webview>

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

GGP Generations & Gender Programme [Subscribe to Mailing List](#) [Login to GGP User Space](#)

[Home](#) | [Data](#) | [Community](#) | [About](#)

Data / Browse the GGS Data

[Online Codebook & Analysis](#) [User Guide](#)

- GGS Pooled datasets
- GGS country data files
- GGS variable availability
- GGP 2020 - Wave 1 - Version 1.0

Welcome to the Online Codebook & Analysis

This page provides direct access to GGS survey data and metadata through the Nesstar software system.

You can search, browse and analyze the data and the corresponding documentation starting with the browse list on the left.

The main functionalities of the interface are explained in the GGP Nesstar Online Access [User Guide](#).

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

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

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*

<i>Features to maintain</i>	<i>Features to improve</i>
<ul style="list-style-type: none">● Compatible with DDI● Visualization of metadata about fieldworks● Variable tabulations and graphs● Data analysis functionalities (e.g., linear regression and cross-tabulation)● <i>Nesstar Publisher</i>, a user-friendly DDI files editor	<ul style="list-style-type: none">● 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...

<i>Pros</i>	<i>Cons</i>
<ul style="list-style-type: none">• DDI compliant• Open source tool• Supported by Harvard University and CESSDA• User-friendly interface and search functionality• Considered for implementation by French archive (PROGEDO Quetelet Diffusion)• Visualization of variable distributions with an external tool	<ul style="list-style-type: none">• Needs to be customized• Needs a well established IT infrastructure• No data analysis• No DDI files editor (Variable metadata can only be edited through the API or an external tool)

3. Evaluate the selected tools

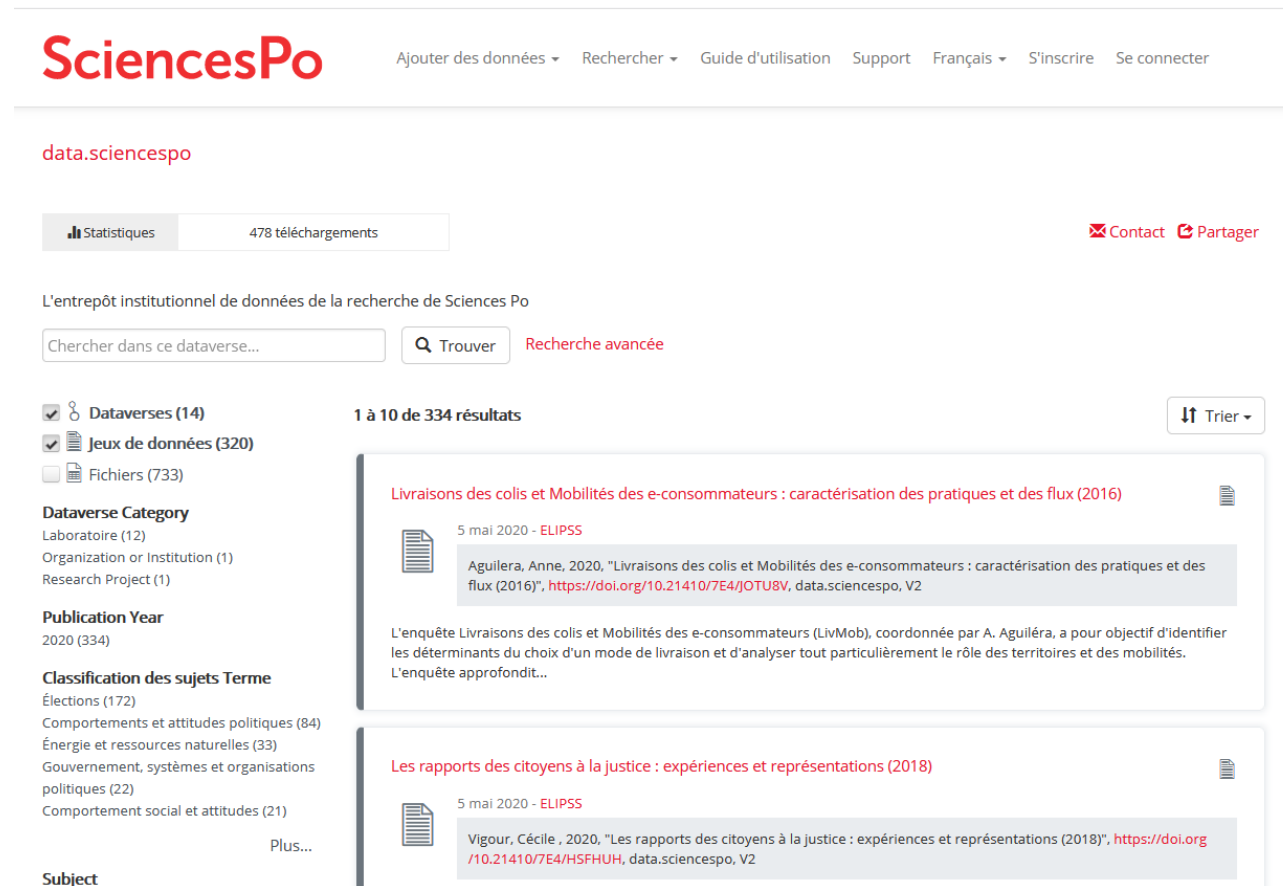


...3.1 Dataverse...

<i>Looked-for characteristics</i>	<i>Evaluation</i>
<ul style="list-style-type: none">• Be compatible with DDI	+
<ul style="list-style-type: none">• Be harvestable by CESSDA catalogue	+
<ul style="list-style-type: none">• Be compatible with Quetelet PROGEDO Diffusion data portal	?
<ul style="list-style-type: none">• Target the largest research community	+
<ul style="list-style-type: none">• Installable on INED servers	?
<ul style="list-style-type: none">• Maintain the strengths of Nesstar (e.g. variable tabulation & data analysis)	+/-
<ul style="list-style-type: none">• Improve the weaknesses of Nesstar (e.g. search functionalities)	+

3. Evaluate the selected tools

...3.1 Dataverse - example



The screenshot shows the SciencesPo Dataverse website. At the top, the SciencesPo logo is on the left, and navigation links for 'Ajouter des données', 'Rechercher', 'Guide d'utilisation', 'Support', 'Français', 'S'inscrire', and 'Se connecter' are on the right. Below the header, the URL 'data.sciencespo' is displayed. A navigation bar includes 'Statistiques' and '478 téléchargements'. On the right, there are 'Contact' and 'Partager' buttons. The main content area is titled 'L'entrepôt institutionnel de données de la recherche de Sciences Po' and features a search bar with the text 'Chercher dans ce dataverse...', a 'Trouver' button, and a link to 'Recherche avancée'. On the left, there are filters for 'Dataverses (14)', 'Jeux de données (320)', and 'Fichiers (733)'. Below these are sections for 'Dataverse Category' (Laboratoire, Organization or Institution, Research Project), 'Publication Year' (2020), and 'Classification des sujets Terme' (Élections, Comportements et attitudes politiques, Énergie et ressources naturelles, Gouvernement, systèmes et organisations politiques, Comportement social et attitudes). A 'Subject' filter is also visible. The main results area shows '1 à 10 de 334 résultats' and a 'Trier' button. Two search results are displayed: 1. 'Livraisons des colis et Mobilités des e-consommateurs : caractérisation des pratiques et des flux (2016)' by Aguilera, Anne, 2020, with a DOI link. 2. 'Les rapports des citoyens à la justice : expériences et représentations (2018)' by Vigour, Cécile, 2020, with a DOI link.

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

3. Evaluate the selected tools

3.2 NADA...

<i>Pros</i>	<i>Cons</i>
<ul style="list-style-type: none">• DDI compliant• Free of charges• Already installed at INED• Visualization of variable distributions• User-friendly interface and search functionality• Compatible with Nesstar DDI version	<ul style="list-style-type: none">• No data analysis possible• No DDI files editor

3. Evaluate the selected tools

...3.2 NADA...

<i>Looked-for characteristics</i>	<i>Evaluation</i>
<ul style="list-style-type: none">• Be compatible with DDI	+
<ul style="list-style-type: none">• Be harvestable by CESSDA catalogue	?
<ul style="list-style-type: none">• Be compatible with Quetelet PROGEDO Diffusion data portal	?
<ul style="list-style-type: none">• Target the largest research community	+
<ul style="list-style-type: none">• Installable on INED servers	+
<ul style="list-style-type: none">• Maintain the strengths of Nesstar (e.g. variable tabulation & data analysis)	+/-
<ul style="list-style-type: none">• Improve the weaknesses of Nesstar (e.g. search functionalities)	+

3. Evaluate the selected tools

...3.2 NADA – examples



Inventaire d'enquêtes Demostaf

ACCUEIL > INVENTAIRE D'ENQUÊTES DEMOSTAF

RECHERCHE PAR MOT-CLÉ
dans la description de l'étude
dans la description des variables

FILTRE PAR PAYS 7
 Toutes
Affichage / Augmenter la sélection

FILTRE PAR ANNÉE
Afficher les études menées entre 1870 et 2020

FILTRE PAR ACCÈS AU DONNÉ ?
 Toutes
 Données disponibles dans une archive externe
 Données non disponibles

FILTRE PAR COLLECTIONS 3
 Toutes

Inventaire d'enquêtes Demostaf

COLLECTIONS CATALOGUE GÉNÉRAL BIBLIOGRAPHIE

Trouvé 305 études parmi 305

Trier les résultats par: Pays | Année | Titre | Popularité

Afficher 1-15 études parmi 305

1 2 3 4 5 Suivant >

Enquête sur les indicateurs du paludisme (2017-2018)
Burkina Faso, 2017-2018
Par : Institut national de la statistique et de la démographie (INSD) - Ministère de l'économie, des finances et du développement, Burkina Faso
Collection: Enquêtes de santé
Créé le: Jul 19, 2019 Dernière modification: Jul 19, 2019 Vues: 905

Enquêtes PMA2020 (Performance Monitoring and Accountability 2020) (2014 - en cours), Vagues 1 - 6
Burkina Faso, 2014-2018
Par : Institut Supérieur des Sciences de la Population (ISSP) - Université Ouaga I Pr Joseph Ki-Zerbo
Collection: Enquêtes sur la fécondité, la contraception, la planification familiale
Créé le: Aug 13, 2019 Dernière modification: Aug 13, 2019 Vues: 2433 Citations: 1

Observatoire de santé et de population de Ouagadougou (2008 - en cours)
Burkina Faso, 2008-2017
Par : Abdramane Soura - Responsable de l'Observatoire de santé et de population Ouagadougou, ISSP, Université de Ouagadougou
Collection: Enquêtes de santé
Créé le: Jul 28, 2017 Dernière modification: Jul 28, 2017 Vues: 2254 Citations: 37

NOM	LIBELLÉ	QUESTION
a401a_a	Who does hh tasks: Preparing daily meals	Preparing daily meals

PDF | Print
Who does hh tasks: Preparing daily meals (a401a_a)
Fichier: GGS_Wave1_France_V.4.3.

Aperçu

Type: Discrète Enregistrements valides: 6085
Format: numeric Invalide: 3994
Largeur: 2
Décimales: 0
Intervalle: 1-2147483624
Invalide: 97, 98, 99

DÉFINITION
Q 401.a.a core questionnaire

Country specific additional values:
- AUS: "2401-shared equally between all household members".
- NOR: "2001-other person from the list of providers and receivers".

Questions et instructions

PRÉ-QUESTION
Now I would like to ask you some questions about who does what in your household. Please tell me who does the following tasks in your household, choosing your answers from the card: card 401-1 if R lives alone; card 401-2 if there is a partner in the household; card 401-3 if there is no partner in the household.

QUESTION LITTÉRALE
Preparing daily meals

CATÉGORIES

Valeur	Catégorie	Enregistrements	
1	always R	1909	31.4%
2	usually R	899	14.8%
3	R and partner equally	953	15.7%
4	usually partner	990	16.3%
5	always partner	1309	21.5%
6	always or usually other persons in the household	13	0.2%
7	always or usually someone not living in the household	12	0.2%
2001	other person from the list of providers and receivers	0	0.0%
2401	shared equally between all household members	0	0.0%
97	does not know	0	
98	refusal	0	
99	no response/not applicable	3994	

Avertissement: ces statistiques indiquent le nombre d'enregistrements trouvés dans les fichiers de données, et non des nombres pondérés. Ils ne peuvent pas être interprétés comme étant représentatifs de la population concernée.

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...

<i>Pros</i>	<i>Cons</i>
<ul style="list-style-type: none">• Compatible with DDI version 3.2<ul style="list-style-type: none">→ Comparisons between variables over time for longitudinal surveys such as GGP• Comes with a DDI files editor, <i>Colectica Designer</i>, with interesting features for GGP<ul style="list-style-type: none">→ Conversion of Blaise files to DDI• Installed on local servers or on a cloud• Allows to browse variable distributions• Training workshops for users	<ul style="list-style-type: none">• 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

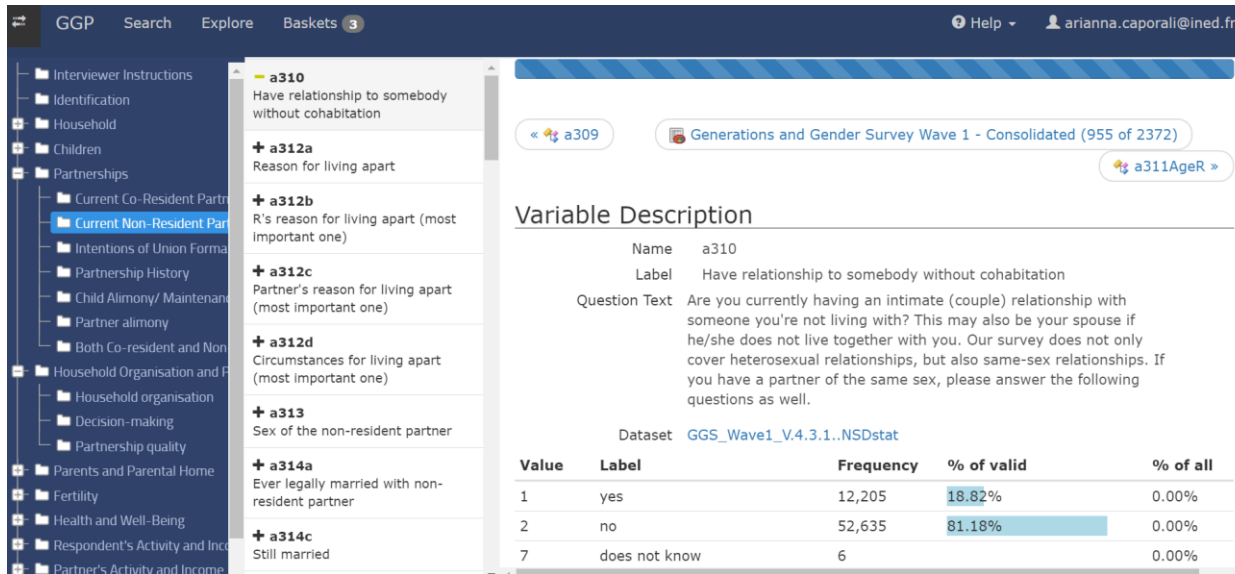
3. Evaluate the selected tools

...3.2 Colectica Portal...

<i>Looked-for characteristics</i>	<i>Evaluation</i>
<ul style="list-style-type: none">• Be compatible with DDI	+
<ul style="list-style-type: none">• Be harvestable by CESSDA catalogue	+
<ul style="list-style-type: none">• Be compatible with Quetelet PROGEDO Diffusion data portal	?
<ul style="list-style-type: none">• Target the largest research community	-
<ul style="list-style-type: none">• Installable on INED servers	?
<ul style="list-style-type: none">• Maintain the strengths of Nesstar (e.g. variable tabulation & data analysis)	+/-
<ul style="list-style-type: none">• Improve the weaknesses of Nesstar (e.g. search functionalities)	+

3. Evaluate the selected tools

...3.2 Colectica Portal – examples



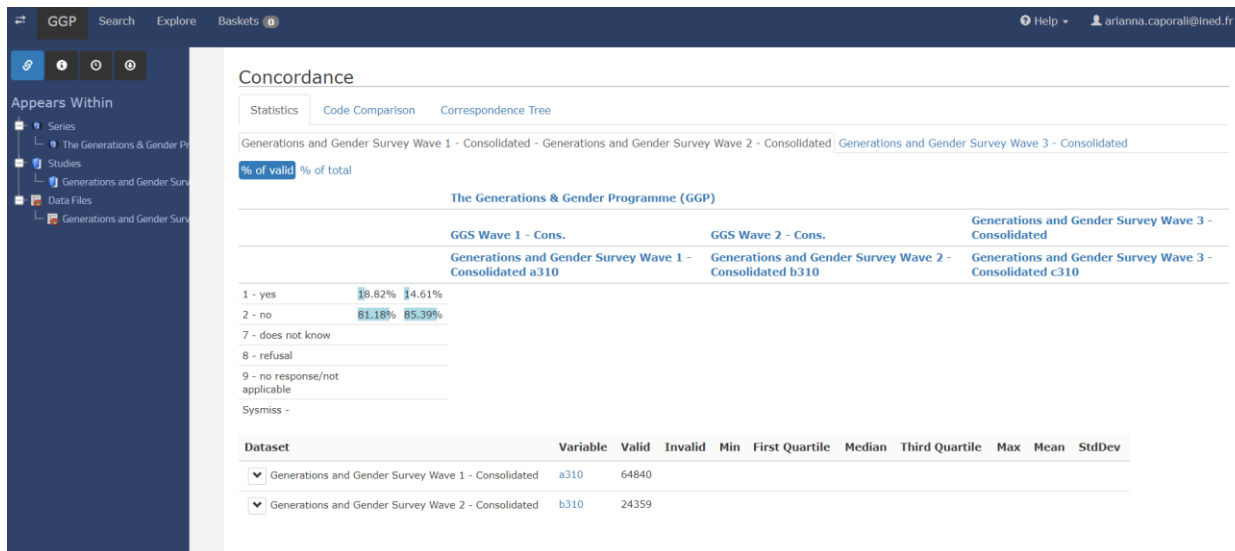
Variable Description

Name: a310
 Label: Have relationship to somebody without cohabitation
 Question Text: Are you currently having an intimate (couple) relationship with someone you're not living with? This may also be your spouse if he/she does not live together with you. Our survey does not only cover heterosexual relationships, but also same-sex relationships. If you have a partner of the same sex, please answer the following questions as well.

Dataset: GGS_Wave1_V4.3.1..NSDstat

Value	Label	Frequency	% of valid	% of all
1	yes	12,205	18.82%	0.00%
2	no	52,635	81.18%	0.00%
7	does not know	6		0.00%

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



Concordance

Generations and Gender Survey Wave 1 - Consolidated - Generations and Gender Survey Wave 2 - Consolidated - Generations and Gender Survey Wave 3 - Consolidated

	% of valid	% of total
1 - yes	18.82%	14.61%
2 - no	81.18%	85.39%
7 - does not know		
8 - refusal		
9 - no response/not applicable		
Sysmiss -		

Dataset	Variable	Valid	Invalid	Min	First Quartile	Median	Third Quartile	Max	Mean	StdDev
Generations and Gender Survey Wave 1 - Consolidated	a310	64840								
Generations and Gender Survey Wave 2 - Consolidated	b310	24359								

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