



Generations and  
Gender Programme



FamiliesAndSocieties

---

## Documenting cross-country comparability of survey data: the case of the Generations and Gender Programme

---

Arianna Caporali

Institut national d'études démographiques (Ined)

Survey Department

*Workshop on Comparative Survey Design and Implementation*

*City University, London, 27/03/2015*



UNITED NATIONS  
ECONOMIC COMMISSION FOR EUROPE

# Outline

---



- Main characteristics of GGP surveys.
- GGP methodology.
- GGP survey metadata and documentation.
- Challenges in documenting this type of surveys.
- Outlook on possible developments.



# Main characteristics of the GGP surveys

---



- Launched in 2000, as the continuation of the international surveys on fertility carried out since the 1960s (e.g. CFS, WFS, FFS).
- Aimed at studying how micro-level factors influence the relationships between generations and gender.
- Covering a variety of topics: fertility, partnership, health, care duty, etc.
- Designed as longitudinal panel studies (3 year intervals) on 18-80 year olds.
- Carried out in 19 countries (Europe and beyond).



# Methodology

---



- Based on a common questionnaire and guidelines for sample design and for data cleaning and harmonization, developed by the coordination team (i.e. NIDI, MPIDR, Ined).
- Adapted to the different national contexts by national teams or incorporated into existing surveys (e.g. Australian, Dutch and Italian cases).
- Leading to country differences in:
  - ✓ fieldwork methodologies,
  - ✓ “compliance” to the standard questionnaire (varying between 80% and 30% of the questions asked).



# Methodology

---



- Implementing *post hoc* harmonization procedures: e.g. label and routing checks, calculation of consolidated variables.
- Keeping country-specific values (i.e. different response categories) and country-specific variables (i.e. different question wordings).
- Requiring comprehensive metadata (i.e. descriptions of common questionnaire and guidelines for data collection and harmonization, as well as of country deviations).

# Groups of metadata items

---



1. Description of the harmonization process.
2. Information on country fieldwork methodologies provided by national teams.
3. Additional documents (e.g. methodology reports, presentations).
4. Detailed examination of each variable (e.g. question, universe, value labels, country deviations from the standard questionnaire, calculation method).
5. Description of variable availability across country datasets.

# Survey documentation

---



- Provided in compliance with the Data Documentation Initiative (DDI), the international standard for documenting social survey data expressed in XML.
- Implemented through the software package Nesstar, which allows to prepare metadata following the DDI-C specification, and to browse and analyse data and metadata online.

# GGP-Nesstar welcome page



The screenshot displays the GGP Online Data Analysis website. At the top left is the GGP logo (Generations & Gender Programme) with a silhouette of a diverse group of people. Below the logo is a navigation menu with buttons for 'Home', 'Data', 'Community', and 'About'. The main heading is 'Online Data Analysis' with a 'User Guide' button to its right. A search bar is present with a 'EN ?' language selector. On the left, there is a list of data categories: 'GGS Pooled datasets', 'GGS country data files', and 'GGS Variables Availability'. The central content area features a 'Welcome to the GGP Online Data Analysis' message, explaining that the page provides direct access to GGS surveys data and metadata through the Nesstar software system. It also states that users can search, browse, and analyze the data and corresponding documentation starting with the browse list on the left. Finally, it mentions that the main functionalities of the interface are explained in the GGP Nesstar Online Access [User Guide](#).



# GGP-Nesstar

---



<http://www.ggp-i.org/online-data-analysis.html>



# GGP-Nesstar content

The screenshot shows the 'Online Data Analysis' section of the GGP website. The header includes the GGP logo and the text 'Generations & Gender Programme'. Below the header is a navigation menu with 'Home', 'Data', 'Community', and 'About'. The main content area is titled 'Online Data Analysis' and includes a 'User Guide' button. A search bar is present, and a list of data options is shown: 'GGP Pooled datasets', 'GGP country data files', and 'GGP Variables Availability'. A red speech bubble points to the 'GGP Pooled datasets' option, containing three numbered points.

GGP Generations & Gender Programme

Home Data Community About

Online Data Analysis [User Guide](#)

- GGP Pooled datasets
- GGP country data files
- GGP Variables Availability

1. For users interested in cross-country analyses, there are **pooled data files** for each wave.
2. For users interested in specific countries, there are **country-specific data files** (17 wave 1 and 9 wave 2 datasets).
3. For both types of users, a **separate data file informs on variable availability across countries and waves.**

# Pooled data files

---



- Metadata common to all harmonized data files (e.g. keywords, abstract, description of the harmonization process and missing values).
- Possibility to understand for several countries at the same time, how country deviations from the standard questionnaire are dealt with in the harmonization process.

# Example of variable descriptions in the pooled data files



- Non-consolidated variable.

- Corresponding consolidated variable.

DESCRIPTION	TABULATION	ANALYSIS
Dataset: Generations and Gender Survey Wave 1 - Consolidated		
<b>Variable a148: Highest reached education level (country-spec. list)</b>		
<b>LITERAL QUESTION</b>		
What is the highest level of education you have successfully completed?		
<b>DESCRIPTIVE TEXT</b>		
Q 120 core questionnaire		
Country specific additional values:		
- BGR: "1101-has not studied in school, incl. illiterate".		
- DEU: "1407-still pupil", "1408-still in training", "1409-other education".		
- ITA: "1701-iscd 5A-6", "1702-iscd 5A", "1703-iscd 5A-5B".		
- AUS: "2401-still at school", "2402-(trade) certificate", "2403-post-graduate degree", "2404-not able to be determined".		
Only country specific values:		
- FRA: "1501-0 - isced97", "1502-1-2 - isced97", "1503-3A - isced97", "1504-3B - isced97", "1505-3C - isced97", "1506-5A-6 - isced97", "1507-5B - isced97".		
<b>Values</b>	<b>Categories</b>	<b>N</b>
0	iscd 0 - pre-primary education	1459   0.8%
1	iscd 1 - primary level	11574   6.5%
2	iscd 2 - lower secondary level	29400   16.5%
3	iscd 3 - upper secondary level	70286   39.4%
4	iscd 4 - post secondary non-tertiary	15745   8.8%
5	iscd 5 - first stage of tertiary	36059   20.2%
6	iscd 6 - second stage of tertiary	2042   1.1%
1101	has not studied in school, incl. illiterate	151   0.1%
1407	still pupil	134   0.1%
1408	still in training	276   0.2%
1409	other education	127   0.1%
1501	0 - isced97	85   0.0%
1502	1-2 - isced97	3205   1.8%
1503	3A - isced97	1087   0.6%
1504	3B - isced97	208   0.1%
1505	3C - isced97	2844   1.6%
1506	5A-6 - isced97	1789   1.0%
1507	5B - isced97	861   0.5%
1701	iscd 5A-6	57   0.0%
1702	iscd 5A	900   0.5%
1703	iscd 5A-5B	85   0.0%
2401	still at school	16   0.0%
2402	(trade) certificate	0   0.0%
2403	post-graduate degree	0   0.0%
2404	not able to be determined	50   0.0%
97	does not know	175
98	refusal	79
99	no response/not applicable	830

DESCRIPTION	TABULATION	ANALYSIS
Dataset: Generations and Gender Survey Wave 1 - Consolidated		
<b>Variable aeduc: Highest Education Level of Respondent</b>		
<b>DESCRIPTIVE TEXT</b>		
Consolidated variables.		
<b>Values</b>	<b>Categories</b>	<b>N</b>
0	iscd 0 - pre-primary education	1459   0.9%
1	iscd 1 - primary level	11574   6.9%
2	iscd 2 - lower secondary level	29400   17.7%
3	iscd 3 - upper secondary level	70286   42.2%
4	iscd 4 - post secondary non-tertiary	15745   9.5%
5	iscd 5 - first stage of tertiary	36059   21.6%
6	iscd 6 - second stage of tertiary	2042   1.2%
96	unknown	0
97	does not know	0
98	refusal	0
99	no response/not applicable	12959
<b>SUMMARY STATISTICS</b>		
Valid cases	166565	
Missing cases	12959	
This variable is numeric		
<b>NOTES</b>		
Copied from var. a148 (Q 120 core questionnaire).		

# Country data files

---



- Metadata common to all harmonized data files.
- For each country, possibility to merge datasets of different waves.
- Description of changes across GGP survey versions.
- Country-specific survey metadata and links to relevant documentation.



# Example of country-specific metadata in country data files



- Fieldwork metadata.
- Links to documents.

The screenshot shows the 'Online Data Analysis' interface for the GGP. The left sidebar lists 'GGP country data files' with a tree view expanded to 'France' and 'Generations and Gender Survey France Wave 1 & Wave 2'. The main content area is titled 'SAMPLING PROCEDURE' and contains detailed text about the sampling frame, coverage, and methodology for the 1999 census data used in the survey.

The screenshot shows the 'Online Data Analysis' interface for the GGP. The left sidebar is similar to the previous screenshot. The main content area is titled 'Dataset: Generations and Gender Survey France Wave 1 & Wave 2' and lists 'Related Materials'. It includes links to the 'Website of Étude des relations familiales et intergénérationnelles (ERFI)', 'Wave 1 French questionnaire (in French)', and 'Wave 1 French questionnaire - Show Cards (in French)'. It also lists 'Other References Note' with links to documents and presentations related to the survey.



# Example of country-specific metadata in country data files



- Fieldwork metadata.
- Links to documents.

The availability of these metadata varies across countries.

# Variable availability data file



The screenshot displays the 'Generations & Gender Programme' (GGP) Online Data Analysis interface. The main content area shows the variable 'x1111\_b: Opinion: Parents should provide financial help adult children when in difficulty'. The data is presented in a table with columns for 'Values', 'Categories', 'N', and a percentage. A horizontal bar chart is overlaid on the table, with blue bars representing the percentage for each category. The categories listed are country and wave combinations, such as 'Bulgaria W1', 'Russia W1', etc. The percentage for most categories is 4.3%, while some are 0.0%.

Values	Categories	N	Percentage
11.1	Bulgaria W1	1	4.3%
11.2	Bulgaria W2	1	4.3%
12.1	Russia W1	1	4.3%
13.1	Georgia W1	1	4.3%
13.2	Georgia W2	1	4.3%
14.1	Germany W1	1	4.3%
14.2	Germany W2	1	4.3%
14.51	Germany-Turkish subsample W1	1	4.3%
14.52	Germany-Turkish subsample W2	1	4.3%
15.1	France W1	1	4.3%
15.2	France W2	1	4.3%
16.1	Hungary W1	0	0.0%
16.2	Hungary W2	1	4.3%
17.1	Italy W1	0	0.0%
18.1	Netherlands W1	1	4.3%
18.2	Netherlands W2	1	4.3%
19.1	Romania W1	1	4.3%
20.1	Norway W1	1	4.3%
21.1	Austria W1	0	0.0%
22.1	Estonia W1	1	4.3%
23.1	Belgium W1	1	4.3%
24.1	Australia W1	0	0.0%
24.2	Australia W2	0	0.0%
25.1	Lithuania W1	1	4.3%
25.2	Lithuania W2	1	4.3%
26.1	Poland W1	1	4.3%
28.1	Czech Republic W1	1	4.3%
28.2	Czech Republic W2	1	4.3%
5	Systemiss	5	



# Conclusion & possible developments

---



- Surveys based on a decentralised model, like the GGP, mobilize a lot of resources of the coordination team for data documentation.
- In the future, we aim at continuing to enrich country-specific information.
- The consideration of upgraded DDI standard may help optimize the collection of country-specific metadata, as well as the management of the entire documentation.





Generations and  
Gender Programme



FamiliesAndSocieties

---

## Comments or questions?

---

Arianna Caporali  
Ined Survey Department  
[arianna.caporali@ined.fr](mailto:arianna.caporali@ined.fr)



UNITED NATIONS  
ECONOMIC COMMISSION FOR EUROPE