The Generations & Gender Programme Contextual Database: A Renewed Source of Contextual Data for Comparative Demographic Research

Arianna Caporali; French Institute for Demographic Studies (INED, Paris, France)

1. Introduction

To face demographic and societal challenges across Europe and beyond, it is important to improve the knowledge about the factors that affect family and life course dynamics across countries. For this, researchers need cross-country comparative data to study individuals' demographic behaviours, but also information on the political and socioeconomic contexts, in which individuals' behaviours are embedded. Next to international comparative survey data, it is therefore also important to provide cross-country comparative contextual data, as a service to the research community.

The Contextual Database (CDB; https://www.ggp-i.org/data/ggp-contextual-database/) of the Generations and Gender Programme (GGP; https://www.ggp-i.org/) supports researchers by providing cross-country comparable macro-level data on demographic, socio-economic, and policy contexts for about 60 countries. It was developed by the Max Planck Institute for Demographic Research (MPIDR) until 2018, when the French National Institute for Demographic Studies (INED) took over its coordination. The CDB is part of the GGP, which aims to improve the understanding of life course dynamics in Europe and beyond, and of the demographic, socio-economic, and political factors that influence them[1]. The CDB complements the individual level data of the Generations and Gender Survey (GGS), a cross-national longitudinal survey run in 20 countries on individual aged 18-79 years, about partnerships, fertility, work-life balance, gender relations, transition to adulthood, intergenerational exchanges, care and later life (Gauthier, Cabaço and Emery 2018).

GGP is currently preparing a new round of data collection, starting from 2020 (see the GGP website for further information). The aim is to update the previous surveys, which were run between 2002 and 2012, as well as to enlarge the geographic coverage to include new countries undergoing low fertility beyond Europe, such as in Latin America and in Asia. As part of this development, the survey questionnaire has been revised to include emerging factors that influence life course trajectories (e.g., the digitalization of life courses). Additionally, the fieldwork methodology has been improved, to ensure data comparability across a wide range of countries (GGP 2018). The data collection operations will be centrally managed by the GGP Central Coordination team at the Netherlands Institute for Demographic Studies (NIDI).

As part of the preparation of the new round of data collection, the CDB content and data collection methodology have also been revised. To face technical challenges, a new online tool was implemented. This paper presents what are these recent developments and aims at promoting the usage of the renewed CDB. After a brief history of the CDB, the paper describes its new content, the new data collection strategy, and the new online database tool. The renewed CDB aims to be a useful support tool for researchers interested in studying demographic trends especially in countries with low fertility, in a global perspective. The CDB data can be used as contextual information for analyses with data from the GGS, but also in conjunction with other data. The CDB also constitutes a data source for analyses of macro-level trends.

[1] The GGP was launched by United Nations Economic Commission of Europe (UNECE) in 2000 (United Nations 2000; Vikat *et al.* 2007; Gauthier, Cabaço and Emery 2018). Since 2009, the Netherlands Interdisciplinary Demographic Institute (NIDI) is in charge of the co-ordination of the project. INED contributes to the project taking care of the survey data documentation and online dissemination and, since 2018, of the Contextual Database.

2. History of the Contextual Database

In the aftermath of the launch of the GGP, an international working group and a group of national experts coordinated by the MPIDR developed the conceptual framework of the CDB[2]. The indicators included in the CDB were identified based on the topics, the scope, the theoretical and the methodological backgrounds of the GGP surveys (Caporali *et al.* 2013, 2016). The CDB contained more than 200 variables structured around sixteen topics[3]. Up to 2008, the institutions that were involved in the GGP conducted the data collection in a decentralized way. The decentralized data collection caused challenges to the purpose of compiling cross-county comparative data.

Between 2009 and 2012, the CDB coordination team based at the MPIDR received funding from the EU to extend the CDB into a broader comparative database (Caporali *et al.* 2013, 2016). The CDB coordination team complemented the data collections provided by GGP partner institutions with different sources, including international databases of supranational organizations or research consortiums. As a result, for the same indicator a variety of (national and international) sources were used, depending on the countries and on the time periods. The aim was to increase the number of indicators for which cross-country comparative data were available. In addition, an online dynamic tool, tailor made for the CDB, was developed, with useful functionalities for researchers (e.g., geo-coding and graphical functionalities). At the end of this phase, thirteen GGP partner institutions provided their country data collections. The CDB online tool contained 74 indicators for about 60 countries; data were available at the national level and, whenever possible, at the sub-national levels. However, after the end of the EU funding, the update of the data caused challenges for the CDB coordination team, as well as for GGP national partner institutions.

In 2018, INED has taken over the development of the database. This coincided with the preparation of a new round of survey data collection for GGP, called *GGP 2020*. Three main developments were required. First, it was necessary to renew the CDB content and to adapt it to the new global coverage and thematic scope of the GGP. Second, a new data collection strategy was desirable to ensure the regular update of the data and to relieve the burden on the GGP national partner institutions, which do not necessarily had the resources needed to collect the CDB. Third, we had to implement a new online software to allow for the transfer of the database to INED.

3. The new content

We identified a new "core" list of 36 indicators, in collaboration with the GGP Central Coordination team at NIDI and in consultation with a panel of demographic researchers. This list will be updated regularly. The following criteria were used: (1) Indicators should help explain or understand demographic trends; (2) indicators should be widely available and therefore be easily updated (ideally annually); (3) indicators should be available for a large number of countries as opposed to only European Union member states or only OECD countries. The new list of indicators draws from three main domains of macro-level indicators that, broadly speaking, help explain or understand demographic trends – especially in the context of developed countries with low fertility in a global perspective:

[2] See the CDB website for the more information on the history of the CDB and for the list of persons involved in its development at the MPIDR, in the international working group, or as national experts.

[3] The sixteen topics were: (1) General demographic indicators, (2) general economic & social indicators, (3) labour market and employment, (4) pension system, (5) parental leave institutions, (6) childcare policies and institutions, (7) military and alternative civilian service system, (8) unemployment, (9) tax/benefit system, (10) housing market and policies, (11) legal regulations of personal relations & family responsibilities, (12) education system, (13) health, (14) elderly care, (15) political system, (16) culture & values. The database consisted of three main data types: national-level time series, national-level policy histories (i.e., text descriptions of key policy changes over time), and sub-national levels (regions, provinces) time series. Whenever relevant, the data were by sex and age.

1) Economic domain: e.g., the influence of economic development on fertility, or the impact of unemployment on the life course trajectories of young adults;
2) Gender domain: e.g., the impact of gender equality on fertility;
3) Welfare and education domain: e.g., the role of social protection and education systems on fertility.

In addition, a forth domain includes summary demographic indicators. To help identify the indicators to be included in the "core" list, we consulted the Global Sustainable Development Goals Indicators Database (https://unstats.un.org/sdgs/indicators/database/), as well as other international databases. The data are now restricted to country level (sub-national data and descriptions of policies are no longer gathered).

The indicators from the pre-2018 platform (the one developed by the MPIDR) not included in the new list may still be found in "The Generations & Gender Contextual Database Archive". This group includes 61 indicators, mostly prepared between 2009 and 2011. Data are at the national level, and, wherever possible, at the sub-national level. For some GGP countries, data were compiled in collaboration with GGP partner institutions.

4. The new data collection strategy

The data included in the "core" list is compiled exclusively from international sources. This allows rapid update of the data with limited resources. The plan is to update these indicators annually. GGP national teams are no longer asked to provide data for the CDB, thus allowing them to focus all their resources on the collection of the survey data. For each indicator, we now consult only the data available in international databases and we identify the best source based on a set of criteria: comparability across countries, completeness of the data, great geographic coverage, long temporal coverage, and availability of metadata. We try to use only one source for each indicator[4]. At the time we write, it is not planned to update the indicators included in The Generations & Gender Contextual Database Archive".

5. The new online tool

In order to transfer the database to INED, we had to implement a new online platform. Instead of developing a tailor made product, we preferred a software already in use in other institutions. We believe that this helps software maintenance in the future. After having tested five tools[5], we implemented PX-Web (https://www.scb.se/en/services/statistical-programs-for-px-files/), the software developed by Statistics Sweden and used by other national and international statistical institutes (e.g., Statistics Norway, UNECE database). We chose this software for three reasons: (1) It is free of charges (also the updates); (2) we could get in contact with the users' community and learn from their experience; (3) it offers some of the main functionalities of the old platform. Among these functionalities, there are: selecting countries following geographic codes, visualizing detailed metadata by data entry, and plotting the data in a variety of graphical options.

6. Data availability as of October 2019

The CDB covers about 60 countries (Europe, North America, Asia and Oceania). In accessing the CDB, users can chose among 36 indicators structured in four groups: (1) Demography (e.g., TFR, mean age at childbearing, contraceptive prevalence, divorce rate, life expectancy, infant and

[4] For the demographic indicators we use the following sources: UN World Population Prospects, UN Estimates and Projections of Family Planning Indicators, UN World Marriage Data, UN Demographic Yearbook, UN Inter-agency Group for Child Mortality Estimation. For the economic indicators we use: World Bank, ILOSTAT, UN Human Development Report. For the gender indicators, the sources are: UN Human Development Report, World Economic Forum, ILOSTAT. For the welfare and education indicators, we use: ILO World Social Protection Report, UNESCO database.

[5] The software that we tested are: Beyond 2020, Nesstar, Eurostat Data Explorer, .Stat and PX-WEB.

maternal mortality, population aged 80+); (2) Economy (e.g., GDP per capita, labour force participation, unemployment rate, poverty rates, internet users); (3) Gender (e.g., gender development index, gender inequality index, global gender gap index); (4) Welfare and education (e.g., social protection expenditure, expenditure on educations, enrolment ratios). The time series are as far back as possible (for some indicators starting from 1950) and end with the most recent data available. Whenever applicable, the data are by sex. A fifth group, called "The Generations & Gender Contextual Database Archive", includes 61 indicators on demography, economy, labour and employment, child care, education, health, pension, tax and benefit. In this group, data are available up to 2010.

7. Conclusion and future perspectives

The renewed CDB has two main features. First, it provides easy access to key up-to-date contextual indicators about demographic, economic, gender, education and welfare-related issues in about 60 countries. These data can be used for macro-analyses and as contextual information for analyses linking the CDB data to GGS data or to other data (e.g., IPUMS, European Social Survey). Second, the CDB makes time series available in a software that offers essential functionalities for researchers, in terms of rich metadata, useful graphical options, and automatic geocoding. Through geo-codes, it is easy to match extracted CDB data with GGS data. Also NUTS and OECD coding schemes are supported, so to match CDB data with data from other surveys. These characteristics make the CDB a useful support tool for researchers interested in studying demographic trends, especially in countries with low fertility, in a global perspective.

In the future, the new "core" list of indicators will be updated annually. We are also open to enrich this list with new indicators, especially following the needs of researchers using GGS data. Similarly, if necessary, we will extend the geographic coverage of the CDB so to include new countries that will carry out the new round of GGP survey. In addition to quantitative indicators at national level, we plan to include a collection of complementary sources of national and sub-national contextual indicators and of "qualitative" descriptions of national policies.

References

Caporali, A., Klüsener, S., Neyer, G., Krapf, S., and Grigorieva, O. (2013). Providing easy access to cross-country comparative contextual data for demographic research: Concept and recent advances of the Generations & Gender Programme Contextual Database. Rostock: Max Planck Institute for Demographic Research (MPIDR working paper; WP-2013-001).

Caporali, A., Klüsener, S., Neyer, G., Krapf, S., Grigorieva, O., and Kostova, D. (2016). The Contextual Database of the Generations and Gender Programme: Concept, Content and Research Examples. *Demographic Research* 35 (9): 229-252. DOI: http://dx.doi.org/10.4054/DemRes.2016.35.9.

Gauthier, A.H., Cabaço, S.L.F., and Emery, T. (2018). Generations and Gender Survey Study Profile. *Longitudinal and Life Course Studies* 9(4): 456-465. DOI: http://dx.doi.org/10.14301/llcs.v9i4.500.

GGP (2018). GGP Technical Guidelines. Deliverable 2.2 of the GGP-EPI project funded under the European Union's Horizon 2020 research and innovation programme GA No: 739511. Available at (last accessed: 29/10/2019): https://www.ggp-i.org/data/methodology

United Nations (2000). *Generations and Gender Programme: Exploring Future Research and Data Collection Options*. New York/ Geneva: United Nations Publications.

Vikat, A., Spéder, Z., Beets, G., Billari, F., Bühler, C., Désesquelles, A., Fokkema, T., Hoem, J.M., MacDonald, A., Neyer, G., Pailhé, A., Pinnelli, A., and Solaz, A. (2007). Generations and Gender Survey (GGS): Towards a better understanding of relationships and processes in the life course. *Demographic Research* 17(14): 389–440. DOI: http://dx.doi.org/10.4054/DemRes.2007.17.14.