Stockholm University Linnaeus Center on
Social Policy and Family Dynamics in Europe, SPaDE

The Contextual Database of the Generations and Gender Programme: Concept, Content and Research Examples

Arianna Caporali, Sebastian Klüsener,
Gerda Neyer, Sandra Krapf, and Olga Grigorieva

Working Paper 2014: 4

The Contextual Database of the Generations and Gender Programme: Concept, Content and Research Examples*

Arianna Caporali

Service des enquêtes et sondages, Institut national d'études démographiques (Ined, France).

Sebastian Klüsener

Max Planck Institute for Demographic Research (MPIDR, Germany)

Gerda Neyer

Stockholm University, Demography Unit (SUDA, Sweden).

Sandra Krapf

Max Planck Institute for Demographic Research (MPIDR, Germany)

Olga Grigorieva

Max Planck Institute for Demographic Research (MPIDR, Germany)

Abstract: Differences in demographic behaviours across countries and sub-national regions have stimulated interest into the relationships between individual characteristics, and the context in which individuals are embedded. Analytical approaches that include contextual factors into statistical analyses of demographic behaviours require well-documented comparative data at the national, as well as the sub-national regional level. The Contextual Database (CDB) of the Generations and Gender Programme (GGP) has been set up to support such analyses by providing comparative data on demographic and socio-economic contexts, covering up to 60 countries in Europe, North America, Asia, and Oceania. This paper presents conceptual considerations and an overview over the content and the functionality of the GGP CDB. Research examples of studies applying multi-level models illustrate how data from this database can increase the analytical potentials of demographic analyses. The GGP CDB is a state-of-the-art research tool, offering well-documented comparative contextual data at the national and regional level. Although conceptually linked to the Generations and Gender Survey, the GGP CDB can also be used to analyse data from other surveys and to study macro-developments. It offers a number of advantages. This includes a high number of indicators specifically geared towards demographic analyses, which often provide extensive temporal and geographic coverage. Besides, the dynamic web environment provides high transparency on data sources as it offers metadata for each single entry. It also supports a number of geocoding schemes that are used by GGS and other surveys to denote region and country of residence.

Keywords: Database ; Contextual data ; Aggregate data; Cross-regional comparison; Cross-national comparison; Micro-macro links ; Multilevel analyses; Generations; Gender.

*This paper is also available as a *Stockholm Research Report in Demography*

1. Introduction

Over the recent decades the demographic landscape of Europe has become more diverse. Fertility rates have fallen below the replacement level in almost all European countries. Yet, some countries have managed to stabilize their fertility rates or even increase them close to replacement levels, while others have experienced long-term low and lowest-low levels. In some countries, the transitions to adulthood, to partnership, and to parenthood have become more heterogeneous. Concomitantly, family dynamics across the life-course as well as across gender and generations have turned to be more complex. In other countries, family formation has followed a rather stable pattern, and behavioural changes have been slow and minimal. At the other end of the life-span, life expectancy has risen particularly in Western Europe, while some groups in some parts of Eastern Europe have registered declines. Despite the gains in longevity in most European countries, there exist still huge and partly growing disparities in life expectancy across countries as well as within countries, for example, across educational and occupational groups, and among women and men (see contributions in Neyer et al. 2013a). These demographic developments and the persistent differences in demographic outcomes and demographic behaviour have spurred scientific and political interest in the role which economic, social, and political factors play in shaping demographic behaviour and demographic development. Central research questions raised are: Does the socio-economic context in which people live affect their demographic behaviour? Do contextual differences explain variations in demographic patterns across Europe? If context matters, which contextual factors matter? How much of the various factors which influence demographic behaviour and demographic outcome can be attributed to contextual factors?

These and similar questions on the relationship between contexts and demographic development are not new: Demographers have long recognized that the spatial and temporal environment in which people live affects their demographic behaviour and shapes

demographic outcomes. For example, in his seminal study on European marriage patterns in perspective, Hajnal links the changes and differences in marriage patterns across Europe and across time to cultural, social and economic factors, for example, to different paces in the development of European countries from agrarian to pre-industrial and industrial societies, to differences and shifts in the economic pre-requisites for household formation and household maintenance, and to changes in gender relationships (Hajnal 1965). Similarly, some of the "classic" fertility theories which relate variation in fertility decline and fertility levels across Europe to differences in economic development, modernization, secularization, value change, state formation, or gender equity attribute the demographic outcomes in question to contextual features (Notestein 1945; Coale and Watkins 1986; Lesthaeghe 2010; Watkins 1991; McDonald 2000). In most of these cases, contextual factors were rather used in a narrative manner, to describe the specificities of countries and their development or the economic, social, cultural, or political circumstances in which people have been living at a specific time. A narrative usage of contextual features offers valuable indications of the potential impact of macro-level aspects on micro-level behaviour. In addition, "thick descriptions" (Geertz)[1] might sometimes be the appropriate, if not the only, way to acknowledge the influence of contextual conditions on individual behaviour and to explain demographic patterns and demographic outcomes (Hoem 2008; Neyer 2013). However, even "thick" descriptions do not provide a sufficiently grounded, that is, statistically verified explanation for the effect of environmental aspects on demographic behaviour. Moreover, without the inclusion of macro-level variables in statistical models, it is not possible to assess the magnitude of the effects of contextual factors on individual behaviour. It is also not possible to distinguish between the strength of influence of various factors. Methodological,

[1] Geertz (1973) uses this term to point to the need to contextualize individual behaviour in order to understand it.

statistical and software developments over the past three decades have made it possible to move from a purely descriptive recognition of the context to a stringent inclusion of contextual indicators in statistical analyses of demographic behaviour. The prerequisites for such analytical approaches, that is, for hierarchical (multi-level) models are the availability of micro- and of macro-level data in a manner which allows researchers to link them in a methodologically sound way. The lack of such suitable data has long been an obstacle to the application of macro-micro models.

Statistical models which allow us to examine the effect of contextual factors on individual behaviour require a sufficiently large number of units at the macro-level (Bryan and Jenkins 2013; Stegmueller 2013). To study the effect of contextual factors on the behaviour of individuals in a single country, demographers need data on spatial units at a sub-national level, e.g., economic, social, cultural, political, demographic indicators for the region or the municipality in which a person lives.[2] For comparative research across countries, country-level indicators might also be used (provided there are sufficient countries in the analysis). In either case (single- or multi-country study), the macro-level indicators need to be comparable across the spatial units.

From its onset, the Generations and Gender Programme (GGP) has been dedicated to provide such data in order to facilitate rigorous, state-of-the-art research of the impact of contextual factors on individual demographic behaviour and demographic outcomes. To this end, the GGP has been set up to comprise both: national, comparative panel surveys with rich individual-level generations- and gender-specific information over time (the Generations and

[2] Which spatial unit of contextual data a researcher needs depends on her/his research question. Spatial units can be geographical (countries, regions, municipalities), economic (e.g., firms), educational (universities, schools), and so forth.

Gender Surveys, GGS)[3] as well as an accompanying GGP Contextual Database (CDB) (Macura 2002; Festy 2004; Vikat et al. 2007).

The aim of this paper is to provide an overview over the concept, content and functionality of the CDB, and to illustrate its potentials for research with examples. We start with the presentation of the concepts that guided the construction of the database and of the data collection process (Section 2). We then sketch the content of the database and its functionalities as well as its accompanying contextual data collection (Section 3). In order to illustrate the usefulness of the database, we present two examples of multi-level studies which made use of data from the CDB as well as from GGS-surveys (Lappegård et al. 2014; Neyer et al. 2011) (Section 4). We conclude with an outlook on the future of the CDB.

2. Conceptual framework and data collection procedure of the GGP Contextual Database

The purpose of the CDB has been two-fold:[4] On the one hand, the CDB should provide data that allow us to investigate the interrelationship between contextual factors and individual demographic behaviour. For this purpose, its data should correspond to the data collected in the GGS. On the other hand, it should allow us to test demographic theories, in particular those related to topics covered by the GGS. This refers, for example, to theories about the linkages between "female economic autonomy", "relative economic deprivation", "ideational change", "wealth flow", "gender equity/equality", etc. and transition to adulthood, marriage formation, family formation, union dissolution, relationship between grandparents-parents-children and so forth. To serve such a complex and broad purpose, the CDB should fulfil the following criteria:

[3] For more information on the GGS see the contributions in this Special Collection, as well as GGP (2014).
[4] For the history of the CDB, see, Caporali et al. (2013)

1. It should provide data in time-series to comply with the GGS's retrospective, prospective and panel structure.
2. It should provide data from a gender and a generation perspective as well as a life-course perspective to comply with the GGS's main dimensions regarding individuals and families. This implies to collect gender- and age-specific data over time.
3. It should provide data for comparative research across countries and within countries. It should therefore collect national as well as regional data.
4. It should provide data on economic, social, cultural, and political aspects to capture essential dimensions of the context in which people live and to allow testing demographic theories about the impact of contextual factors on demographic behaviour and demographic outcome (Neyer 2003).

These aims and criteria reach beyond the objectives of most other databases which existed at the time when the GGP started (at the beginning of this century). Most other research-oriented databases are built on the basis of a specific research question. Since the GGP is not limited to a focused research topic, but is designed for the demographic, sociological, and economic research community, the CDB has aspired to a broader perspective in content and in data coverage.

To structure the collection of data a four-way approach was used as a framework to guide the selection of indicators. First, the content of the GGS questionnaire served as a starting point for determining the relevant contextual domains (Festy 2002). Following a life course perspective, the focus was primarily on central transitions and careers over the life course, such as fertility career (becoming a parent, childbearing by parity, step-parenthood); activity career (comprising education, work, unemployment, retirement); partnership career (cohabitation, marriage, dissolution/divorce/widowhood). For each life course career, a

corresponding contextual domain for the CDB was identified by looking at which economic, social, cultural, and political factors influence the respective transition and life course event (Spielauer 2004). For instance, the transition to parenthood, which is captured in the GGS questionnaire via questions related to intentions to have a first child, becoming a parent, and childcare, was "contextualized" by trying to identify the socio-economic indicators that facilitate or constrain the intentions and the decision to have a child, and the possibility to care. Such macro indicators are, for example, the employment/unemployment rate in a region/country (since this influences income and the possibility to maintain work after childbirth) and the availability of childcare services in a region.

The second approach concerned demographic theories and hypotheses that can be related to the topics covered in the GGS questionnaire (Neyer 2003; Spielauer 2004, 2007). As above, the theoretical assumptions were linked to contextual indicators which could facilitate their testing. For example, indicators to capture "female economic autonomy" at the macro level are, e.g., female labour-force participation rate, share of female part-time workers, female unemployment rate, and women's wages (gender-gap in wages).

Since the GGP is specifically designed to investigate gender and inter-generational relations, attention was paid to maintain a gender and generational dimension in both collection guiding approaches mentioned above. On the one hand, this implied to look for gender- and age-specific indicators over time. On the other hand, grasping relationships goes beyond collecting gender and age-specific data. It comprises indicators which represent "qualitative" aspects of the relationship. Therefore, the potential indicators should reflect aspects which shape gender and generational relationships across the life-course: (1) equality, (2) agency, (3) social rights/social norms, and (4) risks and security (Neyer 2003). For example, levels of equality may be measured based on income distribution or the representation of different groups of the population in specific areas of public life (e.g.,

women's labour force participation). Agency may be evaluated based on the degree of access to social services (e.g., care services) or poverty rates. Social rights may be measured in terms of entitlements to the rights provided (e.g., parental leave). Risk and security may be captured in terms of the distribution of, e.g., health indicators, unemployment rate, social expenditure for vulnerable groups (e.g., families).

The third approach concerned methodological issues involved in the data analysis. As outlined above, to enable researchers to conduct multi-level comparative studies in combination with GGS micro-level data, the CDB had to match the retrospective, prospective, and geographical information collected in the survey (Racioppi and Rivellini 2002). In addition, it had to allow for the linkage over time between individuals and their geographical context, and between them and their membership groups. Furthermore, the data had to be comparative across countries and other units.

The fourth and final approach concerned the practical collection. It followed a two-step strategy. Since many of the indicators relevant for the GGP had never been collected for comparative research and/or over time, national GGP partners were asked to collect the data from national sources. This collection followed a template of collection guidelines for relevant contextual indicators structured around key topics identified via the approaches sketched above (see Spielauer 2004; Caporali et al. 2013 for a description of the development of the data collection procedure). The work by the national GGP teams produced a rich data collection of many indicators which were previously not available internationally. However, because many of these indicators were originally collected by national statistical or governmental offices following national criteria, comparability of these data across countries is limited, despite all attempts by the national GGP partners to derive as comparable data as possible from their existing national sources.

To enhance comparability of indicators across countries/units and over time, existing international databases, such as databases available at the European Union (Eurostat), World Bank, UNESCO, OECD, ILO, WHO, UN or research consortiums (e.g., Human Fertility Database, Human Mortality Database, Comparative Family Policy Database) were screened (Bisogno 2002; Caporali et al., 2013). Each indicator was checked with respect to cross-country comparability, completeness of the time-series, errors, deviation among definitions, notes, other documentations to understand the variables and possible irregularities/breaks in the time series, completeness of data sources; and so forth (see Caporali et al. 2013 for details). To complement internationally available data and to assess them, the internationally available data were compared to the nationally collected ones. Whenever possible, the national and comparative data were merged to assure completeness of time-series and of sub-national regional coverage for all countries. This strategy resulted in a large series of comparable indicators across countries, regions, and across time. To maintain the richness of the nationally collected (mostly non-comparable) data and to ease the usage of the internationally comparative data, it was decided to split the collected data into two parts: (1) the *Contextual Database* (CDB), which contains the internationally comparable data (see Section 3.1. and 3.2 below) and (2) the *Contextual Data Collection* (CDC), which contains the (mostly not comparable) data collected by the national GGP partners (see Section 3.3).

3. Content of the Contextual Database (CDB) and the Contextual Data Collection (CDC)

3.1. Content of the Contextual Database (CDB)

As we already lined out, the collection procedure for the CDB resulted in a large series of indicators that go partially beyond what is offered in other international databases. For example, the CDB provides rich data at the sub-national levels, long time series (as far back

as 1970 and, if possible, even earlier) for many indicators and extensive coverage of Central and Eastern European countries. To allow for theory testing data beyond the core European GGS-countries were included. For some indicators, the database therefore covers all countries in the UNECE region (Europe, Central and Western Asia, North America), as well as GGP countries in Asia and Oceania (Japan and Australia). The majority of indicators is harmonized on the country and/or the regional level over time.

The CDB contains data on a broad range of demographic and economic indicators as well as a selection of social and policy indicators. The indicators are organized in ten relevant domains: demography, economy and social aspects, labour and employment, unemployment, child care, education, health pension, culture and tax and benefits. In total, as of February 2014, 75 indicators are available[5], covering up to 60 countries (Europe, North America, Asia and Oceania). The CDB contains for example: indicators related to fertility (e.g., total fertility rate, mean age at birth, age-specific fertility rates, completed fertility, etc.) and to marriage and divorce (e.g., mean age at marriage, cohort ever married, total divorce rate), life expectancy indicators, indicators about education (e.g., school entry age and pupil-teacher ratio), different measures of gross domestic product, of poverty and of the Gini coefficient, indicators about labour-force participation, average wages, unemployment rates, indicators about pension (e.g., number of beneficiaries and exit age from the labour market). Most of the indicators are provided by age and by sex. Other indicators concern public expenditures, such as: spending on unemployment, childcare, education, health, pension, family allowances, and social protection.

[5] An overview of the available indicators for each country is provided in a table downloadable from the webpage (GGP 2014) in Excel format ("Overview - Available Indicators per Country"). The table provides a definition for each variable, and it tells the user whether the data are available at the national level only, or also at the sub-national regional level.

As mentioned, the data come from different sources, including international databases of supranational organizations or national institutions. Comparisons of these different sources allowed to create time series as complete as possible in terms of spatial and temporal availability of the data[6]. This cross-checking of the data guarantees the provision of high quality data from different sources.

The time series are described through detailed meta-information. Each indicator has a rich documentation which includes a definition of the indicator, a list of all of the national and international sources used to derive the data, and general comments about the sources used and the time series provided. Additionally, meta-.information is provided for each single data entry. This includes the following: information on the source; comments about possible breaks in the series due to revisions of data collection methods and/or changes in national and sub-national regional boundaries; deviation from the general indicator definition, and/or information on the calculation/estimation procedures to derive the given number, where applicable.

3.2. Special functions of the CDB

The database environment is set up as a dynamic system, based on a relational database. The web interface offers a dynamic choice of indicator values across countries, regions, and time (see Figure 1). This means that when picking an indicator, users have direct access to the data and to all related meta-information. This allows for making quick and informed choices as to

[6] In combining data from different sources there were two main strategies. First, for indicators about the core competencies of GGP national experts (e.g., demographic indicators) national sources provided by them are preferred. If the time series contain gaps, an effort is made to fill them with data from international sources that are comparable with the data provided by national collectors. The same international sources are used to derive data for missing countries. Second, for indicators that are harmonised across countries by supranational organisations (e.g., macro-economic indicators and labour market variables) these international sources are preferred. To ensure data consistency, an effort is made to avoid using different sources across countries for the same years. Examples on how the national and international sources were combined are available in Caporali et al. (2013).

extracting data according to the users' needs. Green or respective red flags signal whether an indicator contains solely cross-country comparative data or also non-comparative data. The latter are, however, the exception in the CDB. For those cases, where the provided data deviates for some countries or regions from the variable definition, this is documented in the meta-information. While the database also offers to immediately download all data available for an indicator, there is also the possibility to restrict the output based on certain selection criteria such as for specific years and geographical unites. Depending on the indicators, other selection features may be available (e.g., age and sex). In addition, users can choose the dimensions of the output (e.g., to organise the data columns by regions, by time, etc.). Data can be exported in different formats (e.g., CSV, XLS, and XML).

Figure 1: Choice of indicators about demography.

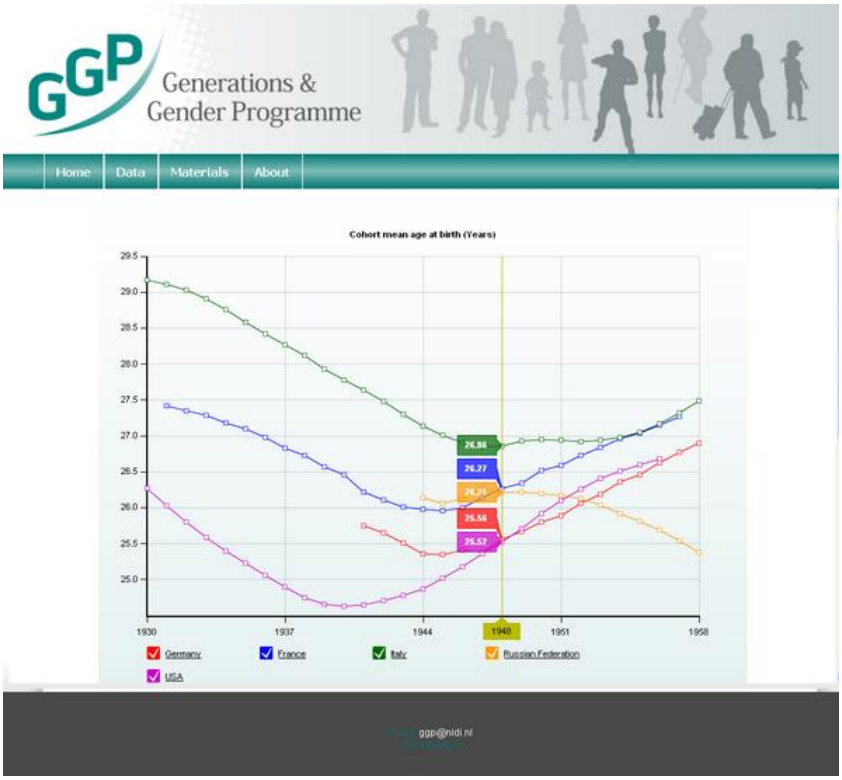
In Figure 2 we show an output example, with the small pop-up window below the centre providing meta-information for a single data entry. Users can access this meta-information either by clicking on the data cell in the output or in the process of defining the dimensions of the output. Here, the user can choose the "Single value column incl. meta data" output, which displays both the values and the meta-information in a single table.

Figure 2: Data output (with metadata for a single data entry and GGP geo-codes)

Furthermore, the CDB allows users to include an identity (ID) column in the output that provides the geocode used in the GGS survey to identify the place of residence of an interviewed person (see figure 2). With this code, the user should find it easy to match the extracted CDB data with the GGS data. In addition to the GGS codes, other regional coding schemes, such as NUTS and OECD, are also supported, which allows researchers to match the CDB data with data from other surveys (e.g., the European Social Survey).

The plot function gives an initial insight into the data (see Figure 3). Several dynamic options are available, including bar plot, line plot, and pie plot. These plots are interactive, allowing the user to zoom in on specific time periods, or to include or exclude countries and/or regions.

Figure 3: Dynamic plot function.



3.3. Content of the Contextual Data Collection (CDC)

The CDC provides detailed data for GGP member countries. As of February 2014 it covers twelve countries (Austria, Belgium, Bulgaria, Canada, France, Germany, Hungary, Lithuania, Norway, Romania, Russia). The data was collected by teams of national statistical offices, research institutes, or research departments within statistical offices that were involved in the GGP. For each country, there are data for approximately 220 indicators. Among these

indicators, there are around 95 national-level time series, 65 sub-national regional variables, and 60 policy histories which contain standardised descriptions of policy reforms[7]. Data go back to 1970, if possible. As it might be particularly difficult to obtain long time series for sub-national regional indicators, the focus of the data collection activities for these indicators is on the period after 2000. In contrast to the largely harmonized data provided in the CDB, the CDC contains a higher number of indicators that are not always comparable across countries. However, data in the CDC are very rich in terms of the national sources used, and comparability of indicators across regions within countries is given. The data are available and downloadable in Excel-format.

4 CDB in Practice

In order to demonstrate the potential of contextual data from the CDB for demographic research we present two studies by Lappegård et al. (2014) and Neyer et al. (2011). Lappegård et al. (2014) examine existing inconsistencies in explanations for the recent rise of childbearing in cohabitation. While some authors interpret it as a pattern of progress driven, among other aspects, by an increasing economic autonomy of women (e.g., Lesthaeghe 2010), others relate it to a pattern of disadvantage driven by economic uncertainties (Perelli-Harris et al. 2010a). In their paper, Lappegård et al. (2014) argue that the inconsistencies may be generated by the fact that these studies focus either on differences between countries or between sub-national regions or between individuals. Processes such as women's raising economic autonomy are often linked to welfare state developments (Sainsbury 1999; Esping-Andersen 2009). Thus, they might be particularly relevant to understand between-country variation, while processes of economic uncertainty might be more important to understand

[7] A comprehensive picture of the indicators included in the CDC is provided in Caporali et al. (2013), figure 1, pp. 7-9.

variation at the regional level, where a "general milieu of social disorganization" (Billy and Moore 1992) might emerge in regions with high structural unemployment. The CDB is very helpful to test such propositions, as it provides national-level data as well as regional level data.

For their study, the authors use individual-level data from the Harmonized Histories project (Perelli-Harris et al. 2010b) to study first birth among partners that are sharing the same household within cohabitation or marriage. The Harmonized Histories project comprises individual-level data from the GGS as well as from national surveys in countries not covered by the GGP. Lappegård et al. link their individual-level data with contextual data from the CDB and other sources. In total, 16 European countries are covered, which are subdivided in 116 regions. Contextual measures include social disapproval of cohabitation, the importance of religious norms, social norms related to the economic autonomy of women, and economic conditions. The first three are constructed from aggregated survey responses of the GGS, the European Social Survey (ESS), and the European Value Survey (EVS). The economic conditions data have been obtained from the CDB. The data are analyzed in a three-level multi-level logistic regression model. The first level constitutes the surveyed individuals, which are nested in their region of residence (second level) and in their country of residence (third level). This set-up allows the authors to make full use of the analytical potentials offered by the CDB data by simultaneously controlling for variation in contextual indicators at the regional and country levels, while being able to account for individual characteristics as well.

The outcomes of the analysis provide support for the propositions by Lappegård et al. (2014): The relevance of explanations varies by geographic scale. Considerations related to an increasing autonomy of women seem to be particularly relevant for understanding variation between countries. Arguments related to a pattern of disadvantage seem more important to

understand variation within countries where regions with higher unemployment show a significantly positive association with the likeliness that a birth occurs in cohabitation. Overall, the study provides an interesting example how the linking of survey data with contextual information at the regional and national level can contribute to improve our understanding of recent changes in family formation behavior in Europe.

The second example concerns the impact of economic opportunities and of economic uncertainties on childbearing intentions from a gender perspective (Neyer et al. 2011). The authors are interested in whether women's and men's childbearing intentions are differently affected by these economic contextual conditions. A plethora of studies has shown that economic uncertainty, measured as individual unemployment, lowers childbearing intentions and childbearing (see: Kreyenfeld et al. 2012). By contrast, economic opportunities, measured as being in employment, are found to have a varying effect on women's fertility intentions and childbearing (Matysiak and Vignoli 2008), while they seem to have an elevating effect on men's (Neyer et al. 2013b). Neyer et al. (2011) explore whether these patterns also hold if one considers the labor-market structures, that is, the economic opportunities and the economic constraints, in the region and in the country in which a person lives. The authors capture economic opportunities via female and male labor-force participation rates at the regional and at the national level, since these indicators can be regarded as indicators of a woman's and man's economic security and of the potential to have work in a region or country. Economic constraints are measured via regional and national unemployment rates, since these indicators reflect whether employment is scarce in a region or respective country.

Like Lappegård et al. (2014) Neyer et al. (2011) employ a three-level logistic regression model and find that including regional and national indicators of employment opportunities and employment uncertainty explains a substantial part of the unexplained variance of a model which only includes individual-level information. This is more so for

men than for women, for the childless more so than for parents. In general the national economic performance seems to have a stronger effect than the regional one. However, regional labor-market conditions matter for mothers' childbearing intentions. This may be an indication that due to work-care tensions mothers are more constrained by local labor market circumstances. The study nicely shows that economic opportunities and economic constraints in the region or the country affect women's and men's childbearing intentions differently. One can conclude from the study that to understand the relationship between economic factors and childbearing it is necessary to take a gender perspective and to consider the structure of economic opportunities as well as of economic uncertainties, both at the regional and the country level.

5. Conclusion

This paper has provided an overview of the conceptual considerations, content and functions of the GGP CDB. Although the main purpose of the database is to increase the analytical potentials of the GGS by providing contextual data for multi-level studies, it may also be useful for researchers who analyse individual-level data from other surveys as well as for researchers interested in studying macro-level trends. The main characteristics of the CDB are as follows: 1) it contains harmonised time-series comparable both across countries and years for a substantial number of demographic, socio-economic and some policy-related indicators; 2) it offers sub-national regional-level data for a large number of indicators; 3) it supports a dynamic linkage of the contextual data to IDs of established geocoding schemes for nations and regions (e.g. NUTS, GGS regional codes) to support the users in combining extracted CDB data with GGS data and data from other sources; and 4) it makes the data available in a dynamic, user-friendly web environment with comfortable and research-oriented functionalities such as detailed metadata documentation. The co-existence of all of these

features in the CDB makes it a unique support tool for researchers interested in the micro-macro linkages as well as for researchers interested in socio-economic structures and macro-level processes. We have demonstrated the potentials of the CDB using two examples of multi-level research. These show how a combined analysis of individual survey data with regional and national contextual level data can make important contributions to improve our understanding of demographic behavior in highly developed countries.

In order to enhance the analytical potentials of the CDB, future activities will particularly focus on increasing the number of harmonised policy indicators. In this endeavour the CDB-team aims to further intensify the collaboration with other database projects. This includes projects such as Anne Gauthier's Comparative Family Policy Databases (Max Planck Society 2014a), the Multilinks Database on Intergenerational Policy Indicators (Multilinks Project and Wissenschaftszentrum Berlin für Sozialforschung 2011), and the Population and Policy Database (Max Planck Society 2014b). The possibility of deriving aggregate national and regional-level data on values from individual-level survey data of the GGS, ESS and the EVS is also currently explored.

We conclude this paper with an outlook on databases of macro-level indicators and social science research: With the development of the internet an immense number of databases of contextual indicators have become available. They have often been collected and compiled by national or international administrative offices and organizations for other purposes than research. While in many cases they provide an excellent resource for researchers, they may in some cases not meet their needs. Research-oriented databases need to fulfil specific criteria: They should be theory-driven, and their content has to meet basic research principles. Data should be verifiable, reliable, and replicable. This requires a thorough documentation of the collection process, of the data sources, of the definition of indicators, and of the harmonisation procedure. There is also the almost insurmountable

tension between establishing a theory-driven, empirical-research oriented database and serving the multiple theoretical and methodological interests of the broader research community. We believe that this tension can only be mitigated by joint European research efforts, like the GGP, in establishing research-oriented databases in the long-run.

6. Acknowledgement

The work on the Contextual Database has been made possible by a design study grant from the European Union as part of the 7th Framework Research Programme (grant number 212749) and by financial support from the Max Planck Institute of Demographic Research, Rostock. Arianna Caporali acknowledges support from the Institut national d'études démographiques (INED), and from Stéphane Legleye, Olivier Thévenon, and Laurent Toulemon for having encouraged her work on the database. Gerda Neyer acknowledges financial support from the Swedish Research Council (Vetenskapsrådet) via the Linnaeus Center for Social Policy and Family Dynamics in Europe (SPaDE), grant registration number 349-2007-8701. We take the opportunity to thank the previous CDB working groups, the national GGP-teams who collected data in their countries and Anne Gauthier (NIDI) for granting allowance to reproduce the indicators in her Comparative Family Policy Database through the CDB. We also thank Fred Heiden and the IT-staff of the Max Planck Institute for programming the web environment and the student assistants at the Max Planck Institute for Demographic Research for support in data collection and data entry.

References

Bisogno, E. (2002). UNECE data of possible interest to GGP Contextual Database. In: *GGP Contextual database group: Activities and conclusions:* 23-26. (http://www.unece.org/fileadmin/DAM/pau/_docs/ggp/2003/GGP_2003_IWG002_BgDocCDBWG.pdf, last accessed: 2014-03-03).

Billy, J. O. G. and Moore, D. E. (1992). A multilevel analysis of marital and nonmarital fertility in the U.S. *Social Forces* 70(4): 977-1011. doi: 10.2307/2580198.

Bryan, M. L., and Jenkins, St. P. (2013). *Regression analysis of country effects using multilevel data: A cautionary tale.* Institute for Social and Economic Research. University of Essex.

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).

Coale, A., and Watkins, S. (eds.) (1986). *The Decline of Fertility in Europe*. Princeton: Princeton University Press.

Esping-Andersen, G. (2009). *The incomplete revolution. Adapting to women's new roles*. Cambridge: Polity Press.

Festy P. (2002). From GGS to GGP; from micro- to macro- via meso- levels. Variables in the survey [GGS] and in the contextual database [CDB]. In: *GGP Contextual database group: Activities and conclusions*: 18-21. (http://www.unece.org/fileadmin/DAM/pau/_docs/ggp/2003/GGP_2003_IWG002_BgDocCDBWG.pdf, last accessed: 2014-03-03).

Festy, P. (2004). GGP Contextual database group: first discussions, first conclusions. Available from the author: festy@ined.fr.

GGP (2014), Generations and Gender Programme, [electronic resource], Netherlands Interdisciplinary Demographic Institute. http://www.ggp-i.org/.

Geertz, C. (1973). *The Interpretation of Culture: Selected Essays*. New York: Basic Books.

Hajnal, J. (1965). European marriage pattern in perspective. In: Glass, D. V., and Eversley, D. E. C. (eds.). *Population in History*. Chicago: Aldine Publishing Company: 101-143.

Hoem, J.M., (2008). Overview Chapter 8: The impact of public policies on European fertility. *Demographic Research* 19(10): 249-260. doi: 10.4054/DemRes.2008.19.10.

Kreyenfeld, M., Andersson, G., and Pailhé, A., (2012). Economic uncertainty and family dynamics in Europe: Introduction. *Demographic Research* 27(28): 835-852. doi: 10.4054/DemRes.2012.27.28.

Lappegård, T., Klüsener, S., and Vignoli, D. (2014). Social norms, economic conditions and spatial variation of childbearing within cohabitation across Europe. Rostock: Max Planck Institute for Demographic Research (MPIDR working paper; WP-2014-002).

Lesthaeghe, R. (2010). The unfolding story of the Second Demographic Transition. *Population and Development Review* 36(2): 211-251. doi: 10.1111/j.1728-4457.2010.00328.x.

Macura, M. (2002). Executive Summary. The Generations and Gender Programme: A Study of the Dynamics of Families and family Relationships. New York/ Geneva: United Nations. (http://www.unece.org/fileadmin/DAM/pau/_docs/ggp/2003/GGP_2003_IWG002_BgDocProgrExecSumm.pdf, last accessed: 2014-03-03).

Matysiak A., Vignoli D. (2008). Fertility and women's employment: a meta-analysis. *European Journal of Population* 24:363–384.

Max Planck Society (2014a), Comparative Family policy Database [electronic resource]. Rostock: Max Planck Institute for Demographic Research. http://www.demogr.mpg.de/cgi-bin/databases/FamPolDB/index.plx.

Max Planck Society (2014b). Population and Policy Database [electronic resource]. Rostock: Max Planck Institute for Demographic Research. www.demogr.mpg.de/databases/PPD.

McDonald, P. (2000). Gender equity in theories of fertility transition. *Population and Development Review* 26(3): 427-439. doi: 10.1111/j.1728-4457.2000.00427.x.

Multilinks Project and Wissenschaftszentrum Berlin für Sozialforschung (WZB) (2011). Multilinks Database on Intergenerational Policy Indicators [electronic resource]. Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB). http://multilinks-database.wzb.eu/.

Neyer, G. (2003). Gender and Generations Dimensions in Welfare-State Policies. Rostock: Max Planck Institute for Demographic Research (MPIDR working paper; WP 2003-022).

Neyer, G., Vignoli, D., and Lappegård, T. (2011). *Gender, economic conditions and fertility. Does context matter for fertility intentions*. Paper presented at the Population Association of America Annual Meeting, Washington, DC, March 31-April 2.

Neyer, G. (2013). Welfare states, family policies, and fertility in Europe. In Neyer, G., Andersson, G., Kulu, H., Bernardi, L. and Bühler, C. (eds.) (2013). The Demography of Europe. Dordrecht: Springer Verlag: 29-53.

Neyer, G., Andersson, G., Kulu, H., Bernardi, L., Bühler, C. (eds.) (2013a). *The Demography of Europe*. Dordrecht: Springer.

Neyer, G., Lappegård, T., and Vignoli, D. (2013b). Gender equality and fertility: Which equality matters? *European Journal of Population* 29(3): 245-272. doi: 10.1007/s10680-013-9292-7.

Notestein, F. W. (1945): Population – the Long View. In: Notestein, F. W. and Schultz, T.W. (eds.). *Food for the World*. Chicago: Chicago University Press: 36-57.

Perelli-Harris, B., Sigle-Rushton, W., Kreyenfeld, M., Lappegård, T., Keizer, R., and Berghammer, C. (2010a). The educational gradient of childbearing within cohabitation in Europe. *Population and Development Review* 36(4): 775-801. doi: 10.1111/j.1728-4457.2010.00357.x.

Perelli-Harris, B., Kreyenfeld, M., and Kubisch, K. (2010b). Harmonized histories: Manual for the preparation of comparative fertility and union histories. Rostock: Max Planck Institute for Demographic Research (MPIDR working paper; WP-2010-011).

Racioppi, F., and Rivellini, G. (2002). The Contextual Dimension in GGP: Some Methodological Issues about Data Collection and Sampling Procedures. In: *GGP Contextual database group: Activities and conclusions:* 11-17. (http://www.unece.org/fileadmin/DAM/pau/_docs/ggp/2003/GGP_2003_IWG002_BgDocCDBWG.pdf, last accessed 2014-03-03).

Sainsbury, D. (ed.) (1999). *Gender and the welfare state*. Oxford: Oxford University Press.

Spielauer, M. (2004). The Contextual Database of the Generations and Gender Program: Overview, Conceptual Framework and the Link to the Generations and Gender Survey. Rostock: Max Planck Institute for Demographic Research (MPIDR working paper; WP 2004-014).

Spielauer, M. (2007). PART TWO - The Contextual Database of the Generations and Gender Programme. In: United Nations (ed.). *Generations and Gender Programme - Concepts and Guidelines*. New York/ Geneva: United Nations: 33-44.

Stegmueller, D. (2013). How many countries for multilevel modelling? A comparison of frequentist and Bayseian approaches. *American Journal of Political Science* 57(3):748-761. doi: 10.1111/ajps.12001.

Vikat, A., Spéder, Z., Beets, G., Billari, F., Bühler, C., Desesquelles, 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: 10.4054/DemRes.2007.17.14.

Watkins, S. C. (1991). *From provinces into nations: Demographic integration in western Europe, 1870-1960*. Princeton: Princeton University Press.