



UNITED NATIONS  
ECONOMIC COMMISSION FOR EUROPE

---



*Report*

---

**6<sup>th</sup> Meeting of the GGP International  
Working Group**

---

Geneva, Switzerland  
13 May 2008

## Contents

GGP progress and current situation	2
Development of the GGP under the FP7 project	4
Standard tables	4
Data Cleaning and Harmonization Guidelines	5
Data Harmonization	5
Country presentations	8
Next IWG meeting	15

The present report reflects the floor discussions of all the agenda items. The presentations and prepared background documents are not summarized in this report; however, they are available on the GGP website.

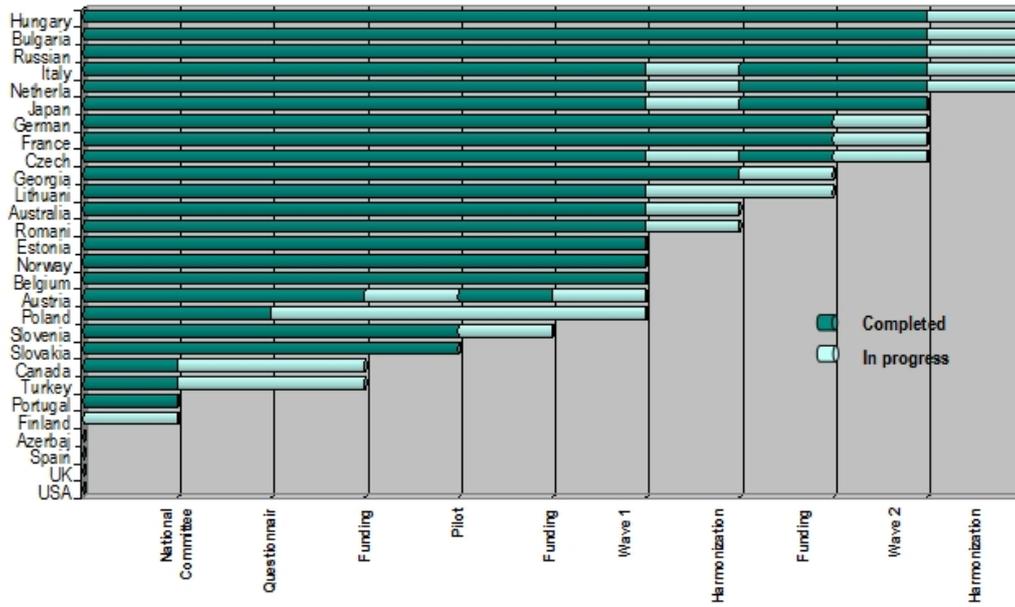
(<http://www.unece.org/pau/ggp/iwg/2008Geneva.htm>).

Representatives from 22 countries attended the meeting. Twelve were from national statistical offices, 34 from research institutes or universities and one from governmental institutions. Members of the Consortium Board (CB) and the coordinators of the GGP expert working groups also attended.

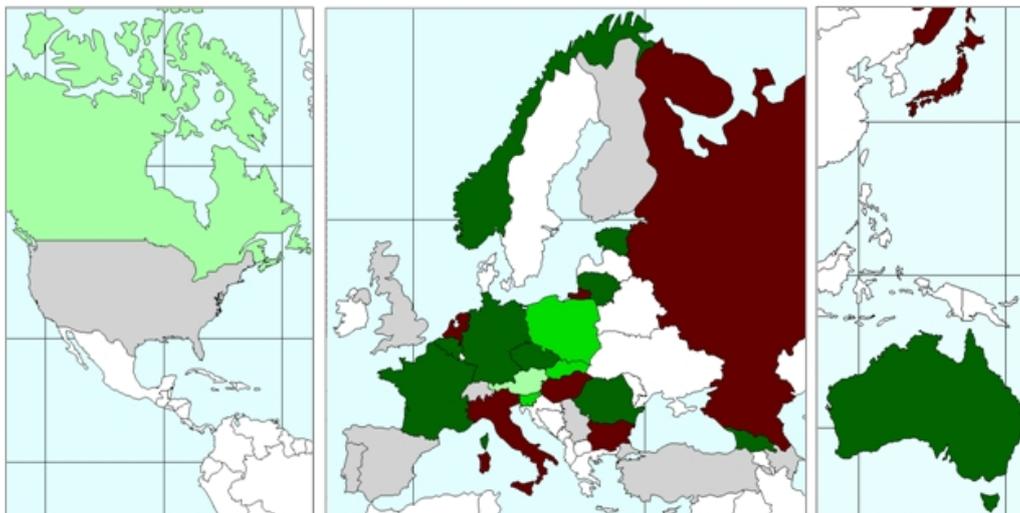
1. John Hobcraft, Chair of the CB, welcomed the participants to Geneva.
2. After a few opening remarks from Andres Vikat, Chief Population Activities Unit United Nations Economic Commission for Europe (UNECE, PAU), the agenda was approved without any comments.
3. The report from the IWG meeting in Ljubljana in 2007 was approved without any comments.

## GGP progress and current situation

4. The GGP Consortium announced two new members, Bocconi University (Italy) and Statistics Norway. Data harmonization, under the responsibility of Netherlands Interdisciplinary Demographic Institute (NIDI), is proceeding as planned. Six country files are harmonized and two are in progress. Data archive and dissemination was launched in January 2008 and information on data availability was not distributed beyond the IWG yet as the on-line data archive is not yet operational. The estimated launch of the on-line data archive was set for 1 July 2008. Progress in the implementation of GGP since the Ljubljana meeting is significant: Wave 1 is fielded for two more countries and is in the field or planned in 2008-2009 for further two; six countries have fielded Wave 2, and three more countries are preparing it. Six countries reporting on the Wave 2 launched in April are available on-line.



5. The discussion on analytical potential of the GGS datasets touched upon the issues relating to the Southern and Northern Europe divide. Southern Europe and the Mediterranean region are currently under-covered by the GGS but some promising developments have been reported from Spain.



## Development of the GGP under the FP7 project

6. Andres Vikat (UNECE, PAU) presented the project that has been selected for funding under the EU 7<sup>th</sup> Framework Programme for Research. The project aims at developing the GGP into a full fledged European Research Infrastructure.
7. The evolution of the GGP builds on enhancing the substantive and methodological dimensions of the Programme. The interdisciplinary approach, central to the GGP, should be expanded to include new areas, in particular psychological characteristics and expand some of the existing ones such as economic well-being, social support networks and life-course decision making. There is also a growing interest in collecting biomarkers and DNA in population surveys which should be taken into account. Further development of the GGP will consider methodological advances in order to improve the GGS and its potential.
8. The discussion about the future developments of the GGP addressed the question of funding of the implementation of the GGP in the participating countries. UNECE as the coordinator works according to the mandate given by the member countries and does not have any funds available. The countries seeking financial support should approach the country offices of the United Nations Population Fund (UNFPA), where they are available, and other potential national and international donors. However, a sign of commitment from the country governments is needed to show their interest in the programme.

## Standard tables

9. Francesco Billari, coordinator of the Analysis Working Group (AWG), presented the report on the development and the implementation of the GGS Standard tables. Currently there are still some outstanding issues that need to be resolved. The AWG is especially active on finalizing tables on gender as well as dealing with the treatment of incomplete income information.
10. The presentation of the imputation procedures for income variables incited a vivid discussion. Although doubts were expressed regarding the trustworthiness of the imputed data, the imputation overcomes difficulties with missing information and reduces certain biases in the data. The original non-imputed variables will be retained in the dataset

for the researcher's convenience. The imputation models have not yet been finalized as the activity is still in progress. Different variations of the imputation models will be considered. Secondary data sources that could be used for imputation have not been considered at this stage, but might be investigated in the future.

11. The tables on life course measures were also addressed. It was expressed that more attention should also be paid to the later stages of the life course.
12. In order to finalize the development of the standard tables they would need to be checked using all the currently available data. Final preparation of the standard tables is an interactive process between the AWG and the data harmonization team. Outstanding issues regarding the production of standard tables, such as treatment of income information and occupational coding, need to be resolved at the level of data harmonization.

## Data Cleaning and Harmonization Guidelines

13. Andrej Kveder (UNECE, PAU) presented the framework of data cleaning and harmonization. The issue of collecting the data on the survey implementation was raised, namely the requested re-translated questionnaires. The issues arising from the re-translation were recognized. However no feasible alternative could be suggested. The information on the national implementation and the degree of deviation from the core questionnaire need to be documented and presented to any potential data user.

## Data Harmonization

14. Niels Schenk (NIDI) presented the implementation of the harmonization of GGS Wave 1 data. The issue of missing survey design variables or design weights in the harmonized data file was raised. Currently this information is not available in any of the Harmonized Data Files (HDF), as the required information was not provided. The survey design information is important for estimating advanced statistical models, for the correct estimation of standard errors of estimates in particular. This type of information needs to be provided within the scope of the

HDF. The survey design variables need to provide the information to which sampling stages, clusters and strata each respondent belongs. Consortium will investigate the steps needed to make this information available to researchers.

15. Even though the information in the HDF is harmonized, some concerns were raised about the comparability of the harmonized information across countries and about the potential loss of information in harmonization. One of the main variables addressed was the income. Currently the income variable is left in a national denomination as reported. The labels for the bracketed information are also left in the national denominations with an approximate amount in Euros to provide some comparative information. In standard tables, quintiles are used in order to insure the comparability of the income information. Purchasing power parities were mentioned as one possible instrument for further harmonization of the income information.
16. Any loss of information as a result of the harmonization should be avoided or at least minimized. When the loss of information would be too great, country specific codes or even variables are retained and appropriately labelled. One of the variables addressed in this context was education. Due to very specific education systems some countries had to use a differential method of collecting the education information and hence producing non standard variables (e.g. Russian Federation and Georgia). The harmonization of these variables can lead to more significant loss of information. For the education variable, International Standard Classification of Education (ISCED) was used as a standard coding scheme.
17. Item nonresponse is not specifically addressed within the harmonization procedure. No imputation will be attempted apart from income information. The recommendation to data users is to use their preferential approach for dealing with item nonresponse, either omitting the missing information or using imputations. Some imputation of vital variables could be considered in the future.
18. Olivier Thévenon (Institut national d'études démographiques - INED) presented the data dissemination process and practical applications for GGS data access and use. The dissemination process of the GGP will be integrated into the INED tools, with some adaptations. The fact that information will be centralized and the secure web platform should facilitate access to the GGS data sets and the communication

between applicants and coordination. The purpose of the application procedure project is to allow users to request survey datasets on-line, on a Web User Space (WUS). The GGS documentation and data have to be conformed to the standards established by the Data Documentation Initiative. The development of the WUS represents an improvement to data access: the centralized information enable a better regulation and control into the process.

19. Data application requests are accepted or rejected on the basis of whether they are meant to be used for meaningful research. No evaluation of research proposals will be attempted. The applicant needs to be affiliated with a research institution as the data are officially released to the institution and not to the individual.
20. The issue a system alerting the data users of any new potential developments has also been raised. An automatic mailing system is already implemented in the current INED system and can be used for alerting the coordinators and data providers. Currently the data user needs to visit his/her WUS in order to receive new notifications. An alerting system using direct e-mail would be a major advantage and still needs to be developed.
21. The list of all accepted applications will be publicly available. Currently the list of research proposals is not yet accessible on-line. The publicly accessible list of projects should encourage cooperation research and push the competitive idea. It should be made explicit that overlap with other programmes is not grounds for rejecting a project proposal.
22. Each data user needs to honour the data dissemination agreement as well as data protection laws of the country where he/she is working. Any deviations from the signed data agreement are bound to those laws. Each institution is responsible for its researchers and also held responsible for any violations of the agreements.
23. Concerns were also raised about the fact that some of the GGS countries have not yet shared the data with the GGP community even the agreed upon maximum period of two years has been reached since the data collection. Very strict sanctions were advocated by some participants. It was agreed that the mechanisms of limiting the access to the international datasets are the most plausible sanctions the IWG can consider.

## Country presentations

**Austria** The funds for the GGS wave 1 have been finally secured. However the amount of funds received does not permit to implement the GGP fully. In order to comply with the funding restrictions, the age range of the respondents was restricted to the reproductive ages of 18-44. The target sample would include  $n = 3,000$  women and  $n = 2,000$  men. Preparations towards the Wave 1 are in progress. Austrian GGS wave 1 data should be provided to the international community by 2010.

**Belgium** Wave 1 pilot study was carried out in February-March 2007 by the Free University of Brussels and *Studiedienst Vlaamse Regering*, with the partnership of Statistics Belgium among others. The evaluation pilot study was completed in March-June 2007. Wave 1 questionnaire and sampling design were finalized. However, Statistics Belgium encountered some unexpected legal constraints in data collection. They were not allowed to collect certain types of data, in particular information on religion, political preference, sexual inclination, race and ethnicity. Consequently, pilot study was delayed for four months. It had implications on the questionnaire and on the sample design. The questionnaire was modified; it required substantive motivations. The sample design in the pilot study ( $n = 1,000$ ) was not incremental. People selected were contacted by reply cards and invitations. Among  $n = 107$  people who returned the invitation, only  $n = 72$  were interviewed. As a result, response rate of the pilot study was extremely low. There is a strong selectivity by age and educational level. A new contact procedure was required for Wave 1 fieldwork. The questionnaire was corrected and improved: translations and permission forms were added.

The targeted population is residential and non-institutionalized individuals aged 18-79. The population aged 79 and older will be followed up in Waves 2 and 3. In order to restore cross-sectional representativity, two age-groups 12-14 and 15-17 have also been drawn. The initial sample of  $n = 16,000$  was drawn from the National Register in January 2008. Regional supplement samples were added depending on observed response rates. There are is a disproportional stratification by region and a proportional one by age and sex. The initial sample has been fielded from March 2008 to June 2008. The fielding of regional supplement samples will be done in September- November 2008. GGS data would be available for commissioning authorities by middle of 2009. The availability of HDF depends on pre-harmonization activities.

The current situation is the following. Wave 1 fieldwork is being car-

ried out but 82.2% of selected people have not been attempted yet. They need to recruit some interviewers (221 interviewers were trained in February 2008). The refusal rate is 5.3%. The majority of respondents agree to be contacted for wave 2. In principle, resources for the three waves are secured. The preparation of GGS data files, of standard report and analyses and of the related documentation is in progress. Pre-harmonization activities started in July 2008. 70% of the variables of the contextual database are already available.

**Bulgaria** The Bulgarian Academy of Sciences and Coordination Research Centre for Social Development and Social Euro-integration is in charge of the organization and conducting of GGP in Bulgaria. The second wave of the GGS began in the mid-January, 2007. The organization and the fieldwork of the study went on until the end of June 2007. Wave 2 covers the same population that was sampled in 2004, for the first wave of GGS. The realized sample includes  $n = 9,364$  individuals, of whom 72.67% are from Wave 1.

The sample realization is varying across regions. The average response rate for the country is 72.7%. The lowest realization of the sample was registered in the capital of Sofia (59.34%) where the mobility of the population is highest. 27.3% of people from the original sample had not been contacted or interviewed. In general, interviewers didn't manage to contact the same person who participated to the first wave, or people refused to be interviewed. The first difficulty was the difficult access to respondents. Then, the main problems encountered are the understanding of the questions by respondents belonging to non-Bulgarian ethnic groups (Roma and Turks), the volume of the questionnaire and the duration of the interview.

**Czech Republic** The Department of Demography and Geodemography of Charles University of Prague completed the fieldwork of GGS Wave 1 in September 2007. The project received a research grant from the Ministry of Labour and Social Affairs.

The main study was carried out in 2005 on  $n = 10,006$  individuals. The sample is based on the 2001 census, restricted to the Czech nationality. They used two questionnaires: one for face-to-face interviews, one for self-completing. Thirty-seven per cent of the questionnaires were self-completed. The sample design is characterised by stratification by region, community size, sex, age, and education, based on demographic analysis. The response rate is 42% and the average length of interview is 77 minutes. The panel presents high nonresponse rates

for some questions; in particular, personal history questions, socio-economic questions and household income.

The Czech team had some problems with data harmonization and with the preparation of data files.

The second wave fieldwork is currently being carried out.

The micro-data are still restricted to the national use. No indication was given as to when the data will be submitted for international research.

**France** The first wave of the GGS was conducted by the INED and the INSEE (Institut National des Statistiques et Études Économiques) in 2005. The sample size is  $n = 10,079$  individuals aged 18-79 years who were interviewed with the CAPI questionnaire adapted to French specificities. The average duration of interviews was 63 minutes. In order to limit the attrition of respondents between the two waves, one mail was sent every six months to update the address book. 88% of respondents agreed to participate to the second wave. So for Wave 2, they expect around  $n = 8,350$  respondents.

The second wave has been in preparation since 2007. The pilot of the activity history carried out in 2007 met difficulties in obtaining precise dates for each period. The pilot survey of the second wave took place in May 2008. The average duration of the interviews was satisfactory (57 minutes). Some difficulties were encountered with the retrieval of information from Wave 1. The results of the pilot further indicate that the questionnaire module on values to be too long. Some sensitive questions caused negative reactions from the respondents, e.g. questions on fertility intentions to respondents with reported difficulties to conceive. The recording of periods of inactivity within the activity history module of the questionnaire also caused difficulties.

Wave 2 fieldwork is planned for October-November 2008. Data should be available in French format in September 2009; and the first results in December 2009. The third wave is expected in France in 2011. The contextual database is being updated. New information is being collected to complete it. The updated contextual database should be delivered by July 2008.

**Germany** The second wave of the GGS, conducted by the Federal Institute for Population Research in Germany, is currently being prepared. The field work will start in mid August 2008. The wave 1 sample is  $n = 10,000$  individuals. Only  $n = 5,200$  respondents agreed to be

re-contacted for the wave 2. The expected realized sample is thus only  $n = 5,000$  interviews. Wave 2 is also planned for a separate sample of Turkish migrants.

Very high attrition between wave 1 and 2 in Germany due to refusal to participate was identified as a major problem for comparative research. In France the re-actions of the respondents showed the opposite tendency and the majority agreed to be re-interviewed. The attrition in the German GGS is also very high even in comparison to the German social and economic panel.

**Georgia** The first wave of the GGS in Georgia was completed in 2006 with  $n = 10,000$  completed interviews. The harmonized data file is already available. The cooperation agreement was signed in July 2007 with the PAU, UNECE. The contextual database is available on-line since September 2007.

Four projects are supported by donors (international partners) to analyse the results of Wave 1 in Georgia on topics such as transformation of family and marital relations, intergenerational solidarity, and gender relationships. Georgian Centre of Population Research (GCPR) which is conducting the GGP in Georgia has five new partners for international collaboration on analyses and research: Centre D'études du monde Russe, Soviétique et post-Soviétique; Université Marc Bloch, Strasbourg; Moscow State University Higher School of Economics; Lithuanian Academy of Sciences Institute for Social Research; University of Texas, Austin.

Georgian Wave 2 is planned for 2009 and the preparatory works, supported by the INED, are processing.

**Hungary** In Hungary, the GGP, is coordinated by the Demographic Research Institute of the Hungarian Central Statistical Office (DRI-HCSO). The Hungarian Wave 1 was implemented in 2001-2002 on  $n = 16,363$  individuals. The Wave 2 carried out in 2004-2005 together with the Wave 1 corresponds to the GGS Wave 1. The sample size at Hungarian Wave 2 was  $n = 13,642$ . The harmonized Hungarian dataset contains information from the first and the second Hungarian waves.

The third Hungarian wave, which is scheduled for second half of 2008, and will correspond to the GGS wave 2. The results in term of percentage of response and successful interviews are quite satisfying: 83.2% of the interviews were successful. As the longitudinal sub-sample consists of 25-82 year-old respondents, two alternatives to top up the sample

have been identified: to interview 20-25 years old children of the respondents (study transmission of demographic behavior) or to create an independent sub-sample of the same age range. The Hungarian GGS was complemented with the sub-sample of Hungarian minority in Transylvania in 2007 ( $n = 2,600$ ).

**Italy** Italian GGS is based on the the 2003 Survey on households, known as Multipurpose on "Family and social subjects", where every family member regardless of age was interviewed for the first time: the sample size was of about  $n = 20,000$  households for a total amount of about  $n = 50,000$  people.

The follow-up survey on "Critical points in work paths in a gender perspective" funded by the Ministry of Labour and Social Affairs, was conducted in 2007 on a sub-sample of  $n = 10,000$  individuals aged 18-64 years in 2003. Originally the first survey was conducted using paper and pencil personal interviewing (PAPI), while the follow-up was conducted Computer Assisted Telephone Interviewing (CATI). Therefore, the sample was selected among the eligible respondents which were reachable by telephone. In order to reduced phone coverage limitations, when the telephone number of the target individuals were not found, the telephone numbers of other family members cohabiting with the respondent in 2003 were sought. The use of mobile numbers was also attempted.

Overall, there were around  $n = 35,000$  eligible individuals in Wave 1. Among those a random sample was drawn stratifying the individuals on gender, age and geographical location. The final realized sample has  $n = 9,997$  respondents.

Currently, the activities are devoted to writing a final report. At this stage no further follow-ups are envisaged due to the lack of funds available. There was no progress made towards the contextual database due to the lack of resources.

**Japan** The first and the second wave of the Japan GGS (JGGS-1 and JGGS-2) are included in the International Comparative Survey on Marriage and the Family, conducted by Aoyama Gakuin University.

JGGS-1 was conducted in the period 24 Feb-31 March 2004. The sample was built on a national two-stage stratified probability sample of  $n = 15,000$  men and women age 18-69. After JGGS-1, they made efforts to keep contact with the respondents for a following survey.

JGGS-2 was conducted in March-April 2007 by the Ministry of Health, Labour and Welfare. It focuses on male and female respondents aged between 18 and 49 at the time of JGGS-1.  $n = 4,568$  individuals were selected. A self-administrated questionnaire was distributed or mailed to selected people for data collection. The response rate for all Japan is quite satisfying (67.5% against 60.5% for JGGS-1). The contextual database is still in the process of data collection. Currently about 80% of numerical data and about 30% of text data are collected. The Japanese CDB should be ready by the end of the year.

In the future, many tasks have to be done to update and harmonize the Japanese surveys and datasets. They have to reshape the JGGS-1 dataset to fit a standard format. The JGGS-2 data has to be checked and cleaned. They should analyze the JGGS-1 and JGGS-2 data and work out policy implications for marriage and fertility. The Japanese contextual database has to be completed. Finally, the Japanese GGS needs to find a secure research fund for the JGGS-3.

**The Netherlands** The representative from the Netherlands Interdisciplinary Demographic Institute (NIDI) shared her experiences regarding sharing the data for international research use. Fast and easy access to the data of the Netherlands Kinship Panel Survey (NKPS) proved to be the main factors on extensive publication record using the NKPS.

**Norway** The Norwegian GGS results from the combination of two panel studies: the GGP panel study and the NorLAG survey. Two research units, Statistics Norway and the Norwegian Social Research, NOVA are in collaboration for the project entitled "Life course, Generations and Gender".

Wave 1 is being completed on  $n = 26,000$  people from a national sample (Statistics Norway) and  $n = 6,000$  more from NorLAG areas (NOVA). The fieldwork started in January 2007. Interviews were planned from January 2007 to June 2008. In order to follow-up respondents, letters were sent to all respondents before they were contacted by phone. Interviews were made by telephone (CATI) and self-completed questionnaires were sent and return by mail.

A GGP lottery was used to increase the response rates. The target was set to 60%, and the realized response rate is about 58% so far. Young people are more difficult to get in touch with than older people, but the later refuse to be interviewed more often. Among the respondents over 50, women refuse to answer more often than men.

Given that the GGS is built on two different surveys, data collection was made in three steps. Registry data including ID number of respondents and basic information such as date of birth, address, sex, status and household members, were added before the survey started. Other registry data were added after the survey, i.e. information on marital history, migration history, parental leave, income, educational activity and attainment, social insurance. Finally, about 84% of GGS questions are covered; about 14% of the questions are covered by administrative records and many questions were added.

Data files should be ready on October 2008 and available for foreign researchers at the end of 2009, on condition that they first apply to Statistics Norway, in accordance with Norwegian legislation.

**Poland** Wave 1 preparations are in progress.

**Russian Federation** The implementation of the GGS second wave in Russia was conducted by the Independent Institute Social Policy (IISP, Moscow). The design and standard survey instruments of the GGS were adjusted to the Russian context by the IISP and the Demoscope Independent Research Center (Moscow).

The sampling design is based on a multi-stage probability sample of households. Because of the lost respondents between the first and the second wave of the GGS, three samples were used to construct one: the panel of 21-82 years old people; the new respondents of 21 to 79 to replace those who left and new respondents of 18-20 years old. New respondents represent 30% of the new panel. Their insertion required the adaptation of the questionnaire. Two different questionnaires were used: the full version for the panel and an abridged version for newcomers. Some questions or modules were excluded from both or from the abridged one; additional questions and modules were also introduced. The resulting questionnaire was still very long and was at the limits of feasible interview time.

55.8% of the total wave 1 sample have agreed to be interviewed after 3 years and provided telephone numbers to be contacted. However, the provided information proved to be insufficient for a successful follow-up.

Some preliminary comparative results from the Wave 1 and Wave 2 were also presented. The sample distribution by the type of settlement reveals a relatively stable spatial distribution for the period. The partnership distribution by age and sex show that at any age, men

and women tend to cohabit more often than actually enter into a formal bond of marriage. The first cohabitation and marriage are being postponed in 2007 in comparison to the 2004 for both men and women.

## **Next IWG meeting**

24. The time and venue of the next IWG meeting will be decided at a later date.