seriss SYNERGIES FOR EUROPE'S RESEARCH INFRASTRUCTURES IN THE SOCIAL SCIENCES

Deliverable Number: 2.1

Deliverable Title: Report on the use of sampling frames in European studies

Work Package: 2. Representing the population

Deliverable type: Report

Dissemination status: Public

Submitted by: SHARE ERIC

Authors: Authors: Annette Scherpenzeel (MEA), Angelica M. Maineri (TiU), Johanna Bristle (MEA), Senta-Melissa Pflüger (MEA), Ilziya Mindarova (TiU), Sarah Butt (ESS ERIC HQ/CITY), Stefan Zins (GESIS), Tom Emery (NIDI), Ruud Luijkx (TiU).

Date Submitted: January 2017

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 654221.

www.seriss.eu @SERISS_EU

SERISS (Synergies for Europe's Research Infrastructures in the Social Sciences) aims to exploit synergies, foster collaboration and develop shared standards between Europe's social science infrastructures in order to better equip these infrastructures to play a major role in addressing Europe's grand societal challenges and ensure that European policymaking is built on a solid base of the highest-quality socio-economic evidence.

The four year project (2015-19) is a collaboration between the three leading European Research Infrastructures in the social sciences – the European Social Survey (ESS ERIC), the Survey for Health Aging and Retirement in Europe (SHARE ERIC) and the Consortium of European Social Science Data Archives (CESSDA AS) – and organisations representing the Generations and Gender Programme (GGP), European Values Study (EVS) and the WageIndicator Survey.

Work focuses on three key areas: Addressing key challenges for cross-national data collection, breaking down barriers between social science infrastructures and embracing the future of the social sciences.

Please cite this deliverable as: Scherpenzeel et al. (2016). *Report on the use of sampling frames in European studies.* Deliverable 2.1 of the SERISS project funded under the *European Union's Horizon 2020 research and innovation programme* GA No: 654221. Available at: www.seriss.eu/resources/deliverables

The use of sampling frames in European studies

Content

Summary

Obtaining good probability samples is a key challenge for European cross-national studies in order to represent the population. This report gives an overview of the sampling frames which are used in countries participating in the four cross-European surveys cooperating in SERISS: the European Social Survey (ESS), the European Values Study (EVS), the Gender and Generations Program (GGP), and the Survey of Health, Ageing, and Retirement in Europe (SHARE). The overview will show where possibilities exist to jointly build and share sampling frames and where studies not using an existing population register can profit from the experience of other studies which do have access to such a register in the same country. It provides a valuable knowledge database of national sampling procedures and accessible population registers across Europe and in addition offers a way to improve harmonization of sampling frames and sample data across European surveys.

1. Introduction

SERISS Work Package 2 "*Representing the population"* is focused on ensuring that European surveys continue to remain state of the art when it comes to accurately describing phenomena in the population. The aim of most high quality surveys is to be able to draw inferences about a specific population by using probability-based sampling. This is a complex and expensive process in many European countries and the problems are compounded when one moves from national to cross-national surveys since the samples in each country must do justice to national specificity but at the same time be internationally comparative. This work package therefore aims to document and share the best of current practice in order to advance the state of the art and promote future harmonisation. Sampling experts and country teams from the four large cross-national face to face surveys involved in SERISS have put their efforts together to work on this aim: the Survey of Health, Ageing, and Retirement in Europe (SHARE), the European Social Survey (ESS), the Gender and Generations Program (GGP) and the European Values Study (EVS).

In the ideal case, all countries included in the surveys would have a probability-based sample from an official person register covering the population of interest. The availability of population registers that can be used as sampling frames varies a lot across countries, however, as do the regulations about who can or can't access the registers and what information can be obtained from them. In this first report of Work Package 2, we describe which sampling frames are used in the four large European cross- national surveys in practice and compare that with an overview of the central population registers which exist in European countries. The report also describes what other sampling frames or sampling methods are used in the countries that do not use a population register.

The objective of this report is to present:

a) An overview of the population registers which are known to exist in the European countries, based on existing publications.
b) An inventory of what is used in practice in the four major cross-European surveys of the moment, that is in the ESS, EVS, GGP and SHARE.
c) A comparison of the overlaps and differences between the surveys and within the countries.
d) A comparison of the list of sampling frames which are used in practice compared to the list of central population registers which exist in Europe, to determine which of the

central population registers are used for sampling in the surveys and which are not used.

e) A general conclusion about the status of probability-based sampling in cross-European surveys and the possibilities for synergy.

The aim of the SERISS work package "representing the population", of which this report is the first deliverable, is to attain synergy in sampling methods used across European surveys. The results presented in this first report provide the basis for this synergy by clarifying in which countries it would in principle be possible to use a common sampling frame for all studies, in which countries a joint effort to obtain access to the population registers for sampling purposes is needed, and in which countries the construction of an alternative common sampling frame may be considered. We will show that in many countries, the same central population register is already used: these countries would be the candidates for synergy in the sense of a common register based sampling frame. On the other hand, we will also list the European countries where registers are available but can apparently not readily be used. Finally, a third list consists of countries in which no population register is available at all: in a next stage of the work package planning we will explore how synergy can be achieved in quality control procedures on alternative methods of sampling. This is related to another goal of the work package: Improving the control over the quality of sampling frames used in European cross-national surveys, by comparing under- and over-coverage with respect to the target population and other sources of possible sampling biases. The topic of the quality of the registers and other sampling frames is not part of the present report but will be addresses in a subsequent report, SERISS Deliverable D2.2.

The present report starts with a short description of the four large European studies that co-operated in this work package project, with a focus on their sampling procedures and documentation. Next, we give an overview of relevant literature about available population registers in Europe and list the different registers found across publications. This is followed by a description of the expert survey which was set up across all countries participating in the four SERISS studies and the results of this survey concerning the use of sampling frames in practice, at present. We will compare the results with the known population registers across countries and identify the countries where the same registers are already used as sampling frames and those where available population registers are not used in practice. Finally, we will summarize the status of the use of sampling frames in the cross-European studies and provide an outlook on the possibilities of synergy in this domain. The report is supplemented by a comprehensive table contained in an excel file documenting the use of sampling frames across Europe. This table can be used by any survey researcher as an information source about sampling possibilities in different countries.

2. The sampling design of the ESS, EVS, GGP and SHARE

2.1 European Social Survey

European Social Survey (ESS): general description

The European Social Survey (ESS) is an academically driven cross-national survey of public attitudes and behaviour that has been carried out in 36 European countries since 2001 (also see http://www.europeansocialsurvey.org). Every two years face-to-face interviews are conducted with a fresh representative sample of the resident population aged 15 and over and living in private households. Consisting of a core questionnaire that remains the same in

every round alongside round-specific rotating modules, the survey covers a range of topics including: satisfaction with democracy, political engagement, health and wellbeing, social norms, attitudes to immigration, work and family, and attitudes to key areas of public policy including immigration, welfare state and energy and climate change. Seven rounds of the survey have been conducted to date with ESS Round 8 entering the field in September 2016. An overview of participating countries by ESS round can be found at: http://www.europeansocialsurvey.org/about/participating_countries.html .

The ESS has been a European Research Infrastructure Consortium (ESS ERIC) since November 2013, directed by Dr. Rory Fitzgerald. The Director is supported in the design and implementation of the survey by the ESS ERIC Core Scientific Team (CST). A National Coordinator in each participating country is responsible for liaising with the CST and ensuring that the survey agency appointed to conduct the ESS does so to the required specification.

European Social Survey (ESS): sampling

Target population: All persons aged 15 and over (no upper age limit) resident within private households in each country, regardless of their nationality, citizenship or language. The ESS target population definition includes people who are on holiday, away working or in hospital for less than 6 months as well as school-age children at boarding school and students sharing private accommodation but excludes people who have been away for 6 months or more, students away at university or college, temporary visitors and people living in institutions.

Sample size: The ESS Sampling Guidelines specify how the ESS approaches the goal of having comparable sampling errors of estimates, which is by aiming to have an effective sample size of 1500 in each participating country in each ESS round (800 for countries with an ESS population of less than 2 million). The calculation of the necessary net sample size to achieve the required effective sample size takes into account the estimated design effect, based on previous ESS rounds. The planned gross sample size is calculated by dividing the planned net sample size by the expected response rate multiplied by the expected rate of eligible persons in the sampling frame.

Sample design: The ESS has a set of rules to govern the sampling designs that are implemented to select the ESS samples in each participating country. These rules are described briefly in the Specification for Participating Countries and in more detail in the Sampling Guidelines: Principles and Implementation for the European Social Survey. The leading principal of ESS sampling is that the selection has to be based on a probabilistic rule. Thus convenience sampling and quota sampling are not permitted and the inclusion probabilities of the respondents should be known and calculable. Within this given set of rules and guiding principles the countries can design freely their sampling procedure. This is necessary, as countries have different endowments (including different access to population registers) which they can use to select a sample that adheres to the ESS sampling guidelines. Generally speaking there are three types of sampling frames, which are addresses of buildings, addresses of dwellings, and addresses of persons. For countries where no list of addresses is available at all, a random walk procedure is used to select buildings or dwellings.

Design approval and responsibility: In all countries that participate in an ESS round the National Coordination Team (NC Team) must prepare a so called Sampling Sign-Off Form

(SoF), which is a detailed (and technical) description of the sampling design that they are going to used.

Sample documentation: After they are finalized SoFs serve as a single point of reference for all sampling design related information. Although such information is also sorted in data collections and documentation systems its origin has to be the SoF. A summary of the sampling designs documented by the SoFs can be found in the Data Documentation Report that is published with every ESS data release. Each country has to submit a so called Sampling Design Data File (SDDF), which contains metadata on the sampling design. The SDDF is a dataset for the complete gross sample and it serves mainly to compute the design weights and to estimate the design effects.

2.2 European Values Study

European Values Study (EVS): general description

The European *Values* Study is a large-scale, cross-national, and longitudinal survey research program that investigates basic human values. The project originated in the 1970s, aiming at describing value differences, similarities, and changes within Europe. The study covers different value fields, such as life, family, work, religion, politics, and society (also see http://www.europeanvaluesstudy.eu/).

The study was fielded for the first time in 1981, and ever since it has been replicated every nine years (1990, 1999, 2008; the next wave will be fielded in 2017). In 1981, the study only covered 14 European countries, USA, and Canada; wave by wave, the survey research program has enlarged, reaching no less than 47 countries in wave 4 (2008). In each country, between 1000 and 1500 respondents are interviewed at each wave. An overview of participating countries by EVS round can be found at: http://www.europeanvaluesstudy.eu/page/participating-countries.html

As for the organizational structure of the EVS, the Council of Program Directors comprises Principal Investigators of each country, and it is in charge of managing the EVS project, discussing the general outlines and approving the final questionnaire and the survey method. The Theory Group develops the questionnaire, while the Methodology Group takes care of the quality of the project. The EVS Foundation aims at planning and promoting joint activities and offers assistance in fund raising. Finally, all daily responsibilities are delegated to the Executive Committee.

European Values Study (EVS): sampling

Target population: The population to be covered by data collection is broadly defined as individuals aged 18 or older, with no upper age limits, which have address of residence in the country at the date of the beginning of fieldwork, regardless of nationality and citizenship or language.

Sample size: The gross sample size is not planned in advance, because of differences in sampling design and response rates between countries. Only the effective sampling size is set, and it consists of 1200 individuals for countries with population over 2 million, and 1000 for countries with population less than 2 million. In 2008, an equal sample size of 1,500 respondents had been set everywhere. Yet, not all countries (mainly small ones) could reach that goal.

Sample design: According to the EVS rules, each country team defines the national sampling procedure and adjusts it to the country condition, even though remaining under the general rule of probability sampling. Single stage strategies are preferred, because of their clearer design and their lower risk of cluster effects. However, also multistage strategies are accepted, as long as national teams can calculate the loss of statistical power, and proportionally increase the number of individuals in the sample in order to obtain the projected effective sample size. When it becomes necessary to select the eligible respondent within a chosen household, the Kish Grid method is most frequently applied for identifying the person to interview; otherwise, further strategies employed for selection of respondents are the next or the last birthday method.

Design approval and responsibility: One of the executive branches of the EVS organization, the Methodology group, is in charge of providing general guidelines and recommendations about sampling design, and reviewing the national sampling plans. Each Program Director has to appoint a Sampling Director (or accept the function by himself/herself), who is responsible of the sample design and who operatively performs sampling tasks, such as providing the required documentation. Each country team has to obtain approval of the sampling design from the Methodology group before starting data collection, and the design chosen has consequences on the gross sample size. When on-field sampling is chosen, the Methodology Group proposes protocols for documentation of choice probabilities for each respondent.

Sample documentation: Sampling documentation consists of sampling plans, sampling frames and description of fieldwork procedures related to sampling, as well as data relevant for calculation of sampling weights. The complete sampling documentation (including data) has to be delivered by national teams to the central body of EVS.

2.3 Gender and Generations Programme

Gender and Generations Programme (GGP): general description

In a nutshell, the GGP is best defined as a "harmonized, large-scale, longitudinal, cross-national panel study of individuals & contextual database" (also see http://www.ggp-i.org). It is a longitudinal panel study covering the whole life course from 18 to 79 years of age. It collects retrospective information as well as prospective information collected through subsequent waves of the survey. The GGP covers a wide range of topics and collects data on: fertility and partnership histories, transition to adulthood, work-family balance, gender relations and gender division of housework, intergenerational exchange including informal and formal care, well-being and health, grandparenthood, and economic activity and retirement. The micro-level data are also complemented by a contextual database providing information about policies and the economic environment at the regional and country-level that may affect individuals.

The GGP is also a cross-national project currently covering 19 countries with data harmonized in a large database for cross-national comparisons. Moreover, 15 of these 19 countries have carried out subsequent waves of data collection (on the same individuals) allowing us to see changes over time in a variety of contexts. An overview of the participating countries by round in the GGP can be found at: http://www.ggp-i.org/. It is a large-scale project involving data collection from about 10000 individuals per country (including both men and women). Such large sample sizes are necessary to study specific population subgroups such as migrants or people at the extreme ends of the income

distribution, as well as to capture a sufficiently large number of life-events for statistical analyses.

The GGP Data collection is conducted by the national teams that involve either national statistical offices or national research institutes or often both. The GGP, including the survey and its instruments are developed centrally by a consortium of leading European centres of population research. GGP participating countries implement the GGS questionnaire, developed by GGP working group on Measurement. The GGP Data Harmonization team is tasked to make the submitted data as cross – country comparable as possible. As of 2016, the GGP is an emerging community of the ESFRI roadmap and aims to join the roadmap as a full member by 2020.

Gender and Generations Programme (GGP): sampling

Target population: The target longitudinal population is the resident non-institutionalized population aged between 18 and 79 selected at wave 1. Individual countries may limit the age ranges to specific periods within the life course but this should be done in consultation with the GGP coordination team. All respondents in the original sample will be interviewed in subsequent samples. At the second wave, the original sample is 3 years older and the age structure of the original sample hence is between 21 and 82 years old. If there is large attrition for a given group, supplementary sampling may be required but any refreshment respondents must be issued the wave 1 questionnaire for their first interview. In order to facilitate international comparisons, GGP recommends that each country minimizes as much as possible exclusions from the target population. Any country that excludes more than 5% of the target population must provide valid reasons for the proposed exclusions to the GGP coordination team.

Sample size: The minimum required number of respondents for GGP varies by country and is driven by the requirement to sustain robust analysis for a minimum number of events. It is recommended that interviews are conducted with, in priority: 1) At least 3,000 women in the reproductive age, i.e. 18-44 at wave 1 or 24-50 at wave 3; and 2) At least 3,000 men in the reproductive age (same age range).

Sample design: The sample should be designed for three waves: individuals selected for the longitudinal sample in year 1 at wave 1 are followed-up in year 4 (at wave 2) and in year 7 (at wave 3). GGP recommends that a probability sample is selected. There are many different types of probability sample designs, each useful in different situations. Each country can decide on which method they prefer, based on the availability and cost-effectiveness of different sampling frames, but quota sampling or any other form of non-probability sampling are not recommended. If a register sampling frame is used, GGP recommends stratifying the population into reproductive and non-reproductive ages (e.g., 18-50, 50-79) and as few regions as possible (e.g., aggregate regions wherever possible). If an area sampling frame is used, GGP recommends a multistage design in which the first stage sampling is geographical region. Detailed sampling guidelines for the GGP are given in Simard and Franklin (2005).

Sample documentation: The GGP provides detailed descriptions of the studies in each country including information on sampling at: http://www.ggp-i.org/data/methodology/data-documentation.

2.4 Survey of Health, Ageing and Retirement in Europe

Survey of health, ageing and retirement in Europe (SHARE): general description.

The Survey of Health, Ageing and Retirement in Europe (SHARE) is a multidisciplinary and cross-national panel database of micro data on health, socio-economic status and social and family networks of individuals aged 50 or older. SHARE's scientific power is based on its panel design that grasps the dynamic character of the ageing process.

The first SHARE data collection started in 2004 and has been repeated bi-annually. With the most recent extension SHARE now covers 26 countries of the European Union as well as Switzerland and Israel. Jointly with harmonized data from the English Longitudinal Study of Ageing (ELSA) and the Irish Longitudinal Study on Ageing (TILDA), pan-European research on effects of our ageing societies and their implications can be extended to all EU countries. The data collection with all new members will be wave 7 of SHARE and start in 2017. An overview of participating countries by SHARE wave can be found at: http://www.share-project.org/group-faq/faqs.html#1.2

In wave 3, the SHARELIFE questionnaire collected retrospective life histories. This will be repeated in wave 7 for respondents who were not yet part of the panel in wave 3. Context variables were collected as a data set of institutional information on the welfare state in Europe. They span a period from 1960 to 2009, over the participating countries and were compiled through national efforts.

In March 2011, SHARE became the first European Research Infrastructure Consortium (ERIC). Central coordination is located at MEA (Munich Center for the Economics of Aging), Max-Planck-Institute for Social Law and Social Policy, Germany. Managing director of SHARE-ERIC is Prof. Axel Börsch-Supan (Munich) and his deputy Prof. Guglielmo Weber (Padua). The country team in each participating country, consisting of a Country Team Leader (CTL) and a Country Team Operator (CTO), is responsible for the contact with the survey agency conducting the fieldwork and for the supervision of the data collection. The Management Board, consisting of internationally reputable researchers in the four scientific areas of SHARE (economics, health, health care, and social/family networks) is responsible for the strategic and scientific governance.

Survey of health, ageing and retirement in Europe (SHARE): sampling.

Target population: The target population of SHARE consists of persons of 50 years or older and persons who are a spouse/partner of a person of 50 years or older, who have their regular domicile in the respective SHARE country. A person is excluded if she or he is incarcerated, hospitalized or out of the country during the entire survey period, or is unable to speak the country's language(s). As the household level is important for most of the variables collected in SHARE, the spouses/partners of people aged 50 and older are included in the target population regardless of their own age. All SHARE respondents who were interviewed in any previous wave are part of the longitudinal sample. Additionally, refreshment samples are drawn regularly to maintain the representation of the younger age-cohorts and to compensate the effect of panel attrition on the sample size.

Sample size: Because funding and sampling resources vary across the participating countries, SHARE does not define a minimum net sample size but advises countries to maximize their net sample size with the available budget. An overview of the sample sizes of

all countries across waves can be found at http://www.share-project.org/data-access-documentation/sample.html .

Sample design: The sampling design is not restricted to be the same in all SHARE countries, but the basic principles of probability sampling with minimal coverage errors guides the choice of the national sampling designs. As a general rule, countries are allowed to use the best sampling frame available at each wave. For the target population of SHARE, a key feature any frame has to fulfil is the availability of reliable information on age. If this information is not available from a given sampling frame, a preliminary screening procedure to identify persons of 50 years or older has to be applied before starting the fieldwork. Most SHARE countries have access to population registers but in countries where this is not the case a random walk procedure is used in combination with a screening procedure. The most frequently used sampling design in the SHARE countries is a multistage stratified sampling design. Regional stratification schemes are recommended in order to ensure a good representation of different geographical areas within the country, improve efficiency of the survey estimates and reduce the costs of the interview process. If other relevant characteristics are available from the sampling frame – such as age and gender in the case of population registers – countries are advised to also use them for stratification. For more information see Börsch-Supan et al, 2013 and De Luca, Rossetti, and Malter, 2015.

Design approval and responsibility: In each countries which draws a baseline or a refreshment sample in a SHARE wave, the Country Team must complete a so called Sample Design Form, describing the proposed sampling design. On the basis of this form, the sampling design is evaluated and approved by the SHARE central coordination in Munich before the sample is drawn.

Sample documentation: The Sampling Design Forms are archived as a reference for the sampling information and the weighting design. In addition, each country that draws a baseline or refreshment sample has to submit a complete gross sample file, containing selection probabilities, geographical codes, strata and, if available, auxiliary variables for each sample unit. The gross sample file is used to compute the design weights.

3. Known central population registers in Europe: literature review

3.1 Definition and available registers

Population registers are, according to an acknowledged definition by the United Nations:

> "an individualized data system, that is, a mechanism of continuous recording, and/or of coordinated linkage, of selected information pertaining to each member of the resident population of a country in such a way to provide the possibility of determining up-to-date information concerning the size and characteristics of that population at selected time intervals" (United Nations, 1969, p. 1)

Therefore, population registers (PRs) identify lists of individuals living in a given country, provide varied information about each individual, and are updated regularly. They should cover the entire resident population. The purposes of such a list are manifold: indeed, even though their primary use is for administrative purposes, they can also be employed to produce demographic statistics about the population (Poulain and Herm, 2013).

To our knowledge, the most up-to-date overview of central population registers in Europe is the one by Poulain and Herm (2013). After drawing a brief history of population registers, they list the registers available in each country. Particular attention is devoted to the role of statistical offices and to the demographic information they have access to through the registers. Of the 30 countries under study (EU27-countries plus Switzerland, Iceland and Norway) 23 have population registers and most of them already have or are developing a central population register. Where there are no population registers, there are usually databases on individuals (it is the case, for instance, in Cyprus, Greece and Malta) in the form of local Civil Registers.

As for antecedents of Poulain and Herm (2013), in 1989 Redfern examined the main advantages and drawbacks of population registers from a statistical point of view (Redfern, 1989). The author focused on Western Europe, except for Austria and Switzerland, and discussed different types of population registers as well as the quality of registers existing at that time. Comparing Redfern's overview to Poulain and Herm's results, one may notice that the observed sample is smaller, and that, according to Redfern, France and Portugal have central population registers that co-ordinate administrative records (Redfern, 1989), while Poulain and Herm (2013) claimed no central population register exists in these countries. However, this contradiction is allegedly caused by heterogeneity in the definition of population registers: Redfern (1989) himself, for instance, pointed out that the French register is an unusual case since it holds personal reference numbers but not addresses. Moreover, Redfern's overview contains mainly information on local population registers and personal reference number usage. The author distinguished four clusters of countries according to the type of registration system. The first group comprises countries (Belgium, Denmark, Finland, Luxembourg, Norway, Sweden) with a full and effective system of population registration, as the registers are coordinated centrally, and individuals have a unique identifier. The second group is the "Intermediate" one (France, the Netherlands, Portugal, Spain), where the centralization of registers and/or the use of unique identifiers were not completed. The third group consists of countries with local population registers only (Germany, Greece, Italy), and is characterized by low quality of registration procedure. The fourth group comprises countries without any recognizable system of population registers (Ireland, the UK).

Eurostat's report "*Comparing data sources for measuring international migration in Central and Eastern Europe"* (Eurostat, 1997) referred to population registers as one of the sources of migration data. The report contains several comparative tables on potential sources of statistical data, types of population registers, their organization, and the population included in population registers. Overall, the information on national population registers is consistent with Poulain and Herm's (2013) results. As the authors in the Eurostat report claimed, "the centralization of population files at national level is now fully operational in the following countries of Central and Eastern Europe: Poland, Czech Republic, Slovak Republic, Hungary, Slovenia and Bulgaria" (Eurostat, 1997, p. 14). They also asserted that at that time only five EU members did not maintain "system of files for the management of the population at the level of the basic administrative unit" (Eurostat, 1997, p. 14) – Ireland, the UK, France, Portugal and Greece.

Figure 1: Structure of national civil registration. Source: Krabina, Prorok, and Tamm (2005, p. 3)

Krabina et al. (2005) summarized information on central and local population registers in Europe on a map (see Figure 1). At the time of report, the majority of Northern and Eastern European countries used central population registers. Portugal, Spain and Central European countries had local registers and Greece, France, Ireland and the UK did not use any population registers. Table 1 summarizes and allows comparing the availability of registers in each country according to different sources.

Table 1. Country's population registers across different sources.
Legend: CR: centralized register; LR: local registers; No: No population or civil registers are available; "-": country not considered in the study

LR	Redfern, 1989	Eurostat, 1997	Krabina et al., 2005	Poulain and Herm, 2013
Austria	-	-	CR	CR
Belgium	CR	-	LR	CR
Bulgaria	-	LR	CR	CR
Croatia	-	CR	-	-
Cyprus	-	-	-	Local civil registers
Czech Republic	-	CR	CR	CR
Denmark	CR	-	CR	CR
Estonia	-	CR(regional level)	CR	CR
Finland	CR	-	CR	CR
France	CR*	No	No	No
Germany	LR (West)	-	LR	LR, CR in some regions, CR of foreigners
Greece	LR	No	No	Local civil register
Hungary	-	CR	CR	CR
Iceland	-	-	-	CR
Ireland	No	No	No	No
Italy	LR	-	LR	LR, CR in preparation
Latvia	-	LR	CR	CR
Lithuania	-	LR	CR	CR
Luxembourg	CR	-	CR	CR
Macedonia	-	No	-	-
Malta	-	-	-	Local civic registers (Malta and Gozo)
Netherlands	LR	-	LR	LR linked online
Norway	CR	-	-	CR
Poland	-	CR	CR	CR
Portugal	CR*	No	LR	No
Romania	-	LR	CR	CR
Slovakia	-	CR	CR	CR
Slovenia	-	CR	CR	CR
Spain	CR	-	LR	CR
Sweden	CR	-	CR	CR
Switzerland	-	-	LR	LR; CR of foreigners
United Kingdom	No	No	No	No

* France has an unusual position, since according to Redfern it has a computerized central population register, but it does not comprise addresses. Also Portugal is marked as having a centralized register for the coordination of administrative records, although it cannot be fully considered a population register.
Note: Reporting on Redfern (1989) and Eurostat (2007), we put CR even though both local and central population registers exist. Krabina et al. (2005) divide centralization of population registers into national-based and regional-based. Moreover, please note that Poulain and Herm (2013) distinguish between PIS (Population Information System), PR (Population registers) and CR (Civil registers), although they do not provide a description of each of them: in our Table, we address mainly centralization. Therefore, in our table, "(Central) PIS" will be marked as CR (Centralized Register). Light green boxes indicate no contradiction across papers.

The situation of 17 out of 32 countries seems to be consistent across all articles, where the UK and Ireland have not had any population register, Austria, Czech Republic, Denmark, Estonia, Finland, Hungary, Luxembourg, Norway, Poland, Slovakia, Slovenia and Sweden have central registers and Italy, the Netherlands and Switzerland have local registers. Assuming similarity in authors' approaches, a development from local to centralized population registers can be seen in Bulgaria, Germany, Latvia, Lithuania and Romania. The opposite conclusion can be drawn for France, since only in one case the existence of a central population register for this country is testified, although it has been pointed out how this might be explained by a different definition of centralized population register and by the peculiarities of the French register (Redfern, 1989). Moreover, one may find different results for Belgium,

Greece, Portugal and Spain. Finally, information on population registers in Croatia, Cyprus, Iceland, Macedonia and Malta is only available in one article, so it is not possible to make comparisons.

In this report, we have chosen the most recent listing found in the literature: the publication of Poulain and Herm, as the basis for the comparison of known central population registers with the sampling frames used in the European cross-national surveys participating in the project.

3.2 Access to population registers

Access obstacles to population registers for researchers are one of the crucial problems addressed by this work. It can be noted that although many European countries manage individual level information into systems, population registers often are not even used for national censuses, though it may significantly decrease the costs. First of all there is a legal constraint: authorities should protect personal data. Moreover, there are practical issues, such as diversity in access procedures, search facilities, and fees. Another issue is the possibility of online access, since not every country in Europe provides remote access. The overall information on online access is reported in Figure 2.

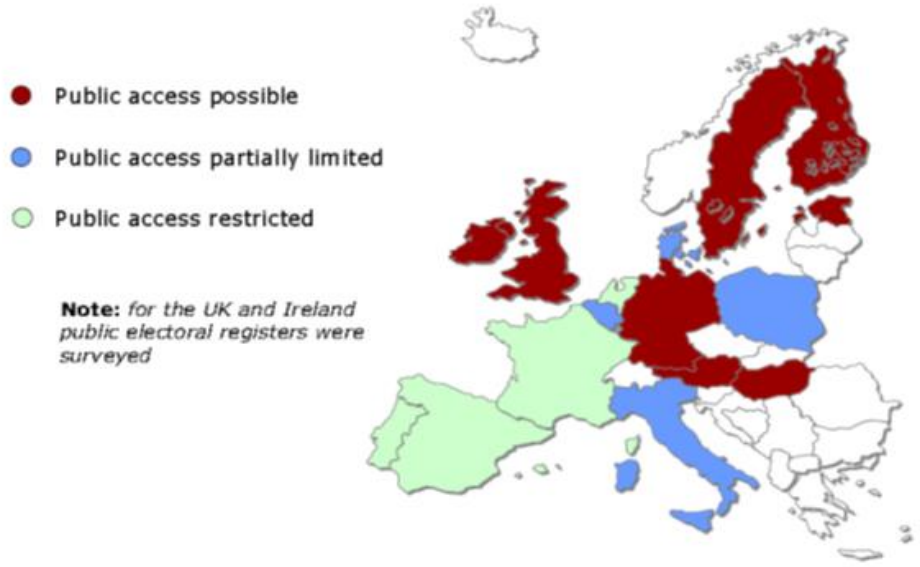
Figure 2: Online access to national population registers. Source: Krabina et al. (2005, p. 5)

Regarding the access issue, Poulain and Herm (2013) discuss the problem of equality and fairness among national population registers. While some countries provide advanced facilities and electronic datasets with micro data, others may not give access even within the country or simply make some restrictions on usage. In other words, a gap might occur between countries related to the availability and quality of registers; for this reasons, the authors called for a more equal access to data (Poulain and Herm, 2013).

The group of researchers working on the RISER project (Krabina et al., 2005) gave an overview on public access to population registers across Europe. Every person or company willing to collect data from a foreign population register may have not only language issues, but also juridical and technical problems. Overall, in the extreme case only public authorities have a right to ask for address details, while in the case of partial access companies that

require information from the population register have to claim the purpose of use or pass a legal check before the enquiry.

Figure 3: Public access to national population registers. Source: Krabina et al. (2005, p. 4)



As Nordic countries are the most experienced on population registers, they have established special services that provide micro data for researches (United Nations Economic Commission for Europe, 2007). Thus, in Finland researches may apply for samples and other use of data from the Longitudinal Census Data File and Longitudinal Employment File created in 1970 and 1987 respectively. In Sweden one may look for necessary information via the longitudinal integrated register *Louise* or use *MONA*, a special online service providing anonymized data by request. The more important finding is that register-based production is well established and the National Statistical Institutes in Nordic countries share anonymous micro data with research institutions by request. The main rule that they follow is providing researchers with only necessary data without any excessive information. For example, Statistics Denmark allows different organizations to obtain tailor-made sets of micro data online, since often companies were requesting data sets with the same demographic variables.

3.3 The use of population registers as sampling frames

In general, while there are several studies on register-based analyses and on the relationship between censuses and population registers, the literature focusing on the possible use of population registers as sampling frames is scarce. Redfern (1989), focusing on the UK, recognized that using population registers as a frame for sample surveys would be better than using the registers usually employed at those times. Yet, the author did not develop the argument further. In the UK, although no population register is available, Roberts et al. (1995) reported on using a register held by the Family Health Service Authority (FHSA) which contains all individuals registered with UK General Practitioners. According to the authors, this register has several advantages: it comprises almost 95% of the population, of all age; it is computerized; it comprises always information on gender, age and postcode; finally, data are updated regularly. Use of the register is possible after approval from FHSA's authorities. Major problems concerned the functioning of the software used by FHSA for sampling, and problems

with postcodes. Yet, in comparison to electoral rolls and telephone listing, which are overall more accurate, FHSA registers are updated regularly. In 1995, FHSA was disbanded and it is not clear what happened to the register and if it is still available. Leti, Cicchitelli, Cortese, and Montanari (2002) analysed the Italian situation, where municipal vital registers (*Liste anagrafiche*) are used as sample frames by the National statistical office (ISTAT) for some of its sample surveys.

In conclusion, the literature search resulted in a few studies about population registers' use for sampling purposes, each describing the situation for a single country only. Therefore, this report will contribute to the literature and to survey practice by investigating the employment of population registers as sampling frames from the point of view of experts working in large and cross-national survey research programs.

4. An expert survey about sampling frames in Europe: Method

To construct a comprehensive overview of the sampling frames which are used in the four SERISS studies (SHARE, ESS, GGP and EVS) and to create an inventory of the availability of auxiliary variables in these sampling frames, an expert survey was carried out. Between the end of April and beginning of May 2016, the researchers who are responsible for sampling and data collection in the countries included in the four large surveys received a questionnaire about the use of sampling frames and auxiliary data in their studies. The questionnaire was programmed as an electronic form and was sent by email to the country teams in each of the four studies, accompanied by an official invitation letter signed by the director of the respective study. Annex 1 shows the generic version of the questionnaire. Researchers of the country teams were asked in the email and letter to forward the questionnaire to the sampling expert who was responsible for their samples if they were not the experts themselves. This could also be a person at the survey agency to which the fieldwork is assigned. The questionnaire asked about the name and type of register actually used for the survey purpose, the responsible authority, the register's accessibility for different researchers and organisations, the amount of time it took to obtain a sample from it, the problems encountered, and the auxiliary variables obtainable from it. In addition, questions were included enquiring about other sources for auxiliary data that were used. The results of the questions concerning auxiliary data are presented in the SERISS deliverable 2.5, titled "Report on auxiliary data in available country registers" (Bristle et al, 2016)[1]. In the present report, we focus on the use of registers as sample frames in the latest rounds of the four SERISS studies.

The data from all questionnaires of the four studies are stored on the SHARE server at the Max Planck Institute for Social Law and Social Policy in Munich.

5. The use of sampling frames in European cross-national surveys: Results

5.1 Response to the expert survey

Table 2 shows the countries invited to participate in the expert survey in each of the four European studies involved and the countries which returned the completed questionnaire.

[1] The present report includes one more questionnaire, from the ESS Slovene country team, than the report of Bristle et al. for which this questionnaire was received too late.

No systematic pattern of the same nonresponding countries across the four studies was observed.

Table 2. Invitation and participation of countries in the expert survey across the four European studies.

Survey **Response**	ESS	EVS	GGP	SHARE
Invited for expert survey (N)	25	43	15	20
Completed expert survey (N)	21	33	10	19
Countries (abbreviations)	AT, BE, CH, CZ, DE, EE, ES, FI, FR, GB, HU, IE, IL, IS, LT, NL, NO, PL, RU, SE, SL	BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GE, GR, HR, HU, IE, IS, IT, LT, LU, ME, MK, MT, NL, PL, PT, RO, RS, RU, SE, SK, UA	AT, CZ, DE, EE, FR, HU, IT, NL, NO, SE	AT, BE, CH, DK, DE, EE, ES, ES-gi, FR, GR, HR, IL, IT, LU, NL, PL, PT, SE, SI

5.2 A comprehensive documentation file

From the data obtained from all country teams that participated, across the four surveys, a large overview and documentation table was constructed in excel. This table is the core part of the study and the work package deliverable, and can serve as a point of reference for any researcher who wants to draw a sample from the population in a European country. As the table is too large to include in this report text document, it is delivered as an excel file supplemental to the report (document "SERISS_WP2_D2.1_Annex2.xls", delivered with this report and publically available at: www.seriss.eu/resources/deliverables.
We shortly describe here the information that can be found in this file and how it is presented. In addition, we summarize some of the information that can be found in detail in this table file in the following paragraphs.

Figure 4 presents a screenshot of the excel documentation file as an illustration. The first column of the file contains all countries which participated in the expert survey. For each country, a separate row is included for the information obtained from each of the four European surveys: ESS, EVS, GGP and SHARE and an additional row for the information listed in the article of Poulain and Herm (2013). If a register was used as sampling frame, the name of this register is indicated as well as the responsible authority who owns/controls the register. In the next column is indicated what kind of sampling frame was used in each country in each of the four studies: A population register, another type of register, a geographical listing or database, or another sampling method. This information is summarized in Table 3 in this report. Furthermore, it is indicated what type of sampling unit

was obtained from the sampling frame (including each type of sampling frame mentioned): Individual address, building address, households/dwellings or other. The next column shows the wave and year in which the sampling frame was used in the study[2]. In the final 8 columns is shown to what degree each sampling frame is accessible, indicating: whether access is possible only nationally or also internationally, whether universities have access to it, non-commercial research institutes, statistical offices, commercial survey organisations, commercial marketers, and others.

Figure 4: Screenshot of sampling frame overview in Excel SERISS_WP2_D2.1_Annex2.xls.

	A	C	D	E	F	G	H	I	J	K	L	M	N
1	The use of sampling frames in four cross-European surveys												
2											Access by affiliation		
3	Country	Study	Name of register	Responsible authority	Type of sampling frame	Sampling unit	Used in Waves (Years)	Accessible for	Universities and affiliates	Non-commercial research institutes and affiliates	Statistical office(s)	Commercial survey organisations	Commercial marketers	Oth
4	Austria	ESS	Zentrales Melderec	Ministry of the Inter	population or civil r	Individual address	7 (2014)	international	No	No	Yes	No	No	Yes
5		EVS												
6		GGP	Zentrales Melderec	Federal Ministry of	population or civil r	Individual address	1 (2004), 2 (2012)	national	No	No	Yes	No	No	No
7		SHARE	Data Door	Österreichische Pc	register for specific	Building address	4 (2011)	international	Yes	Yes	Yes	Yes	Yes	No
8		Poulain & Her	Zentrales Melderec	Federal interior mir	Centralized population register									
9	Belgium	ESS	Rijksregister/Regis	Sectoraal Comité v	population or civil r	Individual address	1 (2002), 2 (2004)	national	Yes	Yes	Yes	No	No	Yes
10		EVS	Rijksregister	Commissie voor de	population or civil r	Individual address	1 (1981), 2 (1990)	national	Yes	No	Yes	No	No	Yes
11		GGP												
12		SHARE	Registre National /	Privacy Commissio	population or civil r	Individual address	4 (2011), 5 (2013)	don't know	Yes	Yes	Yes	No	No	No
13		Poulain & Her	Registre national d	Interior and equal (	Centralized population register									
14	Bulgaria	ESS												
15		EVS			random walk		5 (2017)	international	Yes	Yes	Yes	Yes	Yes	Yes
16		GGP												
17		SHARE												
18		Poulain & Her	Naselenie Esgraon	Regional developm	Centralized population register									
19	Croatia	ESS												
20		EVS			random walk		4 (2008)							
21		GGP												
22		SHARE	Registar osiguranil	Croatian Health Ins	register for specific	Individual address	6 (2015)	don't know	Yes	No	Yes	No	No	No
23		Poulain & Herm												

5.3 Type of sampling frames used across countries and studies

Figure 5 shows a map of Europe indicating where population registers have been used as sampling frames: The darker the shade of orange, the more studies in a country use a register. In addition, Table 3 shows which of the four studies in a country used a population register as the sampling frame, and which used a different register or a different sampling method. The graph as well as the table represent the current status of the sampling in the four studies: The use of sampling frames can vary within the same countries across different rounds of longitudinal studies.

Of the 83 returned questionnaires across the countries and studies included, 51 reported that some form of person register was used as the sampling frame (42 used a population register and 9 used a different type of person register such as an election register or health insurance register). No use of telephone registers was reported in any of the countries´ questionnaires. In total, 31 reported the use of alternative databases or procedures, such as geographical listings and random route procedures. It can be seen in Table 3 that the use of other methods of sampling than drawing from person registers is most common in the EVS. The EVS fieldwork started in the 1981, about 20 years earlier than the other three studies, when fewer possibilities to use population registers might have existed. Moreover, the EVS covers more countries than any of the other three studies (particularly compared to GGP and SHARE). The questionnaires completed by the EVS teams included 10 countries which were not available for the three other studies and which all indicated to have used another method of sampling than a person register: Bulgaria, Cyprus, Georgia, Malta, Montenegro, Macedonia, Romania, Serbia, Slovakia and Ukraine. In addition, other sampling methods

[2] Note that we asked the sampling experts in the country teams to report about the *current sampling frame* they were using for the study. The sampling frame used for one study in one country can vary over time, depending on changes in accessibility of frames and on operational options.

than drawing from a population register were reported in two or more studies in the following countries: Croatia, United Kingdom, Greece, Ireland, Lithuania, Portugal, and Russia. This list largely reflects the absence of population registers (central or local registers) in these countries as shown in the literature review in section 3.

Countries having a mix of reported sampling frames, hence a population register in one or more studies and other methods in other studies are: Austria, Estonia, Germany, Netherlands, Spain (all five having in majority population registers as sampling frames), Czech Republic, Finland, France, Italy, Luxembourg, Poland (all six truly mixed). France is a somewhat special case, as the literature overview indicated that the country has no population registers. Indeed, the EVS based the sampling in France on a random walk procedure and SHARE on the rolling population Census. However, the SHARE sampling expert completing the sampling questionnaire categorised this sampling frame as a population or civil register. In presenting the results, we adhere to the answers given by the sampling experts in the countries.

In 10 countries, the use of a population register is reported by each study's country team participating in our expert survey. The country teams of all four studies declared the use of a population register in Sweden. In nine countries: Belgium, Denmark, Hungary, Iceland (not shown in figure 5), Israel, Norway, Slovenia, Spain-Girona, and Switzerland, we do not have information about all four studies but the studies included all reported the use of a population register (see table 3). A special category is formed by the list of countries in which no use of population registers as sampling frames was reported at all, in any of the four studies. To determine whether registers are available at all in these countries, we will in the next section compare these countries with the lists of available population registers in Europe published by Poulain and Herm (2013). We will also show what other registers can be used for sampling in some countries. Furthermore, we compare in which countries the four European studies all use the same register and in which countries they use different registers as sampling frames.

Figure 5. Use of population registers as sampling frames for cross-national European studies. Source: SERISS expert survey about sampling frames in Europe, 2016.

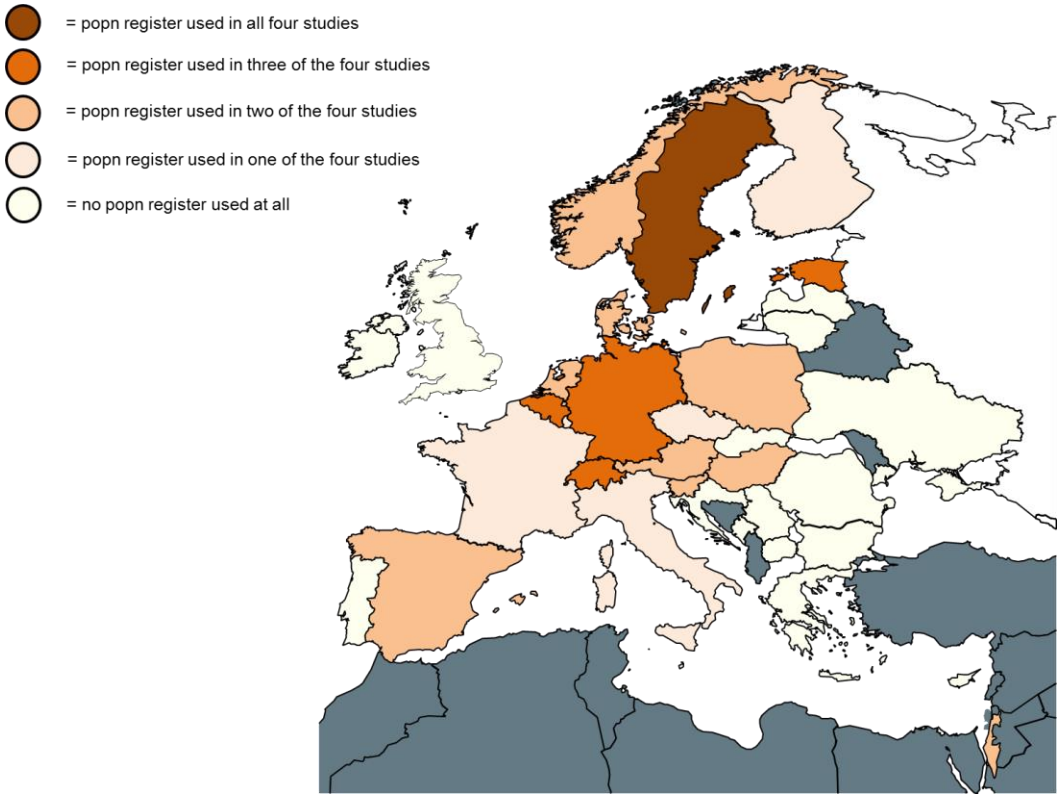


Table 3. Summary of sampling frames used in countries, across the four surveys. The frames were used for the data collection in different years, between 2004 and 2017[3].

	Study			
	ESS	**EVS**	**GGP**	**SHARE**
Totals				
Popn register	14	8	7	13
Other register	0	4	0	5
Other method	7	20	3	1
Country				
Austria	Popn register	-	Popn register	Other register
Belgium	Popn register	Popn register	-	Popn register
Bulgaria	-	Other method	-	-
Croatia	-	Other method	-	Other register
Cyprus	-	Other method	-	-
Czech Republic	Other method	Other method	Popn register	-
Denmark	-	Popn register	-	Popn register

[3] For the ESS, 11 countries´ questionnaires referred to the sampling frame used for all recent rounds up to the 2016 round, seven referred to all recent rounds up to the 2014 round, and one country (Iceland) referred to the 2012 round. For the EVS, 15 countries ´questionnaires referred to the sampling frame used for all recent rounds up to the 2017 round; another 15 referred to rounds up to the 2008 round; and one country (Hungary) referred to the 2014 round. For the GGP, five countries´ questionnaires referred to the sampling frame used in 2004; one to 2008 and one to 2012. For SHARE, five countries´ questionnaires referred to the sampling frame used for all recent waves up to the coming 2017 wave (for which they were preparing a sample already); six referred to all recent waves up to the 2015 round; five referred to either 2011 or 2013; and one referred to 2004. In each study, a two to three countries did not indicate the year of reference. The variation in years referred to within each of the studies reflects the fact that not all countries are always able to participate in each round or wave.

	Study			
	ESS	**EVS**	**GGP**	**SHARE**
Estonia	Popn register	Popn register	Other method	Popn register
Finland	Popn register	Other method	-	-
France	Other method	Other method	Other method	Popn register
Georgia	-	Other register	-	-
Germany	Popn register	Popn register	Other method	Popn register
United Kingdom	Other method	Other method	-	-
Greece	-	Other method	-	Other method
Hungary	Popn register	-	Popn register	-
Iceland	Popn register	Popn register	-	-
Ireland	Other method	Other method	-	-
Israel	Popn register	-	-	Popn register
Italy	-	Other register	Popn register	Other register
Latvia	-	-	-	-
Lithuania	Other method	Other method	-	-
Luxembourg	-	Popn register	-	Other register
Macedonia	-	Other register	-	-
Malta	-	Other register	-	-
Montenegro	-	Other method	-	-
Netherlands	Other method	Other method	Popn register	Popn register
Norway	Popn register		Popn register	
Poland	Popn register	Other method	-	Popn register
Portugal	-	Other method	-	Other register
Romania	-	Other method	-	-
Russia	Other method	Other method	-	-
Serbia	-	Other method	-	-
Slovakia	-	Other method	-	-
Slovenia	Popn register	-	-	Popn register
Spain	Popn register	Other method	-	Popn register
Spain-Girona	-	-	-	Popn register
Sweden	Popn register	Popn register	Popn register	Popn register
Switzerland	Popn register	Popn register	-	Popn register
Ukraine	-	Other method	-	-

5.4 Use of the same population registers for sampling

We have compared the population registers used as sampling frames in the four studies to the most recent list of known central population registers in the EU as published by Poulain and Herm (2013). Figure 6 as well as Table 4 show that in 16 of the countries in which a population register was used for one or more of the SERISS surveys, this was the central population register as listed in the publication of Poulain and Herm, and in 13 of these countries this central population register was used by at least two of the four studies. The darker the shade of orange in Figure 6, the more studies in a country used the register listed by Poulain and Herm. Table 4 shows in more detail which studies in each country used this register, and what the name of this register is, as given by the experts in the survey. In

addition, it shows that in four countries in which a different register than that listed by Poulain and Herm was used, this other register was used in at least two of the studies as well. The countries in which the same register was used in two or more of the SERISS surveys include all Scandinavian countries, most Western-European, some middle European countries and two southern-European countries (Italy, Spain) as well as Israel. In these countries, the possibility to jointly build sampling frames for the studies from the same register can be explored. For more details about the same registers that were used across studies in each country, we refer to the comprehensive excel results table that is delivered as documentation with this report (see at: www.seriss.eu/resources/deliverables).

Table 4 also indicated which countries were not covered by the listing of Poulain and Herm (2013) or in which they did not find a central register. These countries partly overlap with the countries in which only the EVS conducts surveys, using different sampling methods than register based samples (Georgia, Montenegro, Macedonia, Serbia, and Ukraine). However, Table 4 in addition shows that in 10 countries which do have a central population register according to Poulain and Herm this register is not used for sampling in any of the four studies. These countries are: Bulgaria, Cyprus, Czech Republic, Greece, Latvia, Lithuania, Luxembourg, Malta, Romania and Slovakia. Of these countries, Luxembourg uses a different register and the Czech Republic uses a Census as register, in one or two of the studies. The sampling experts from the different studies in the other eight countries all reported the use of other sampling methods which are not based on register sampling.

Figure 6. Use of the population register listed by Poulain and Herm (2013) as sampling frame for cross-national European studies. Source: SERISS expert survey about sampling frames in Europe, 2016.

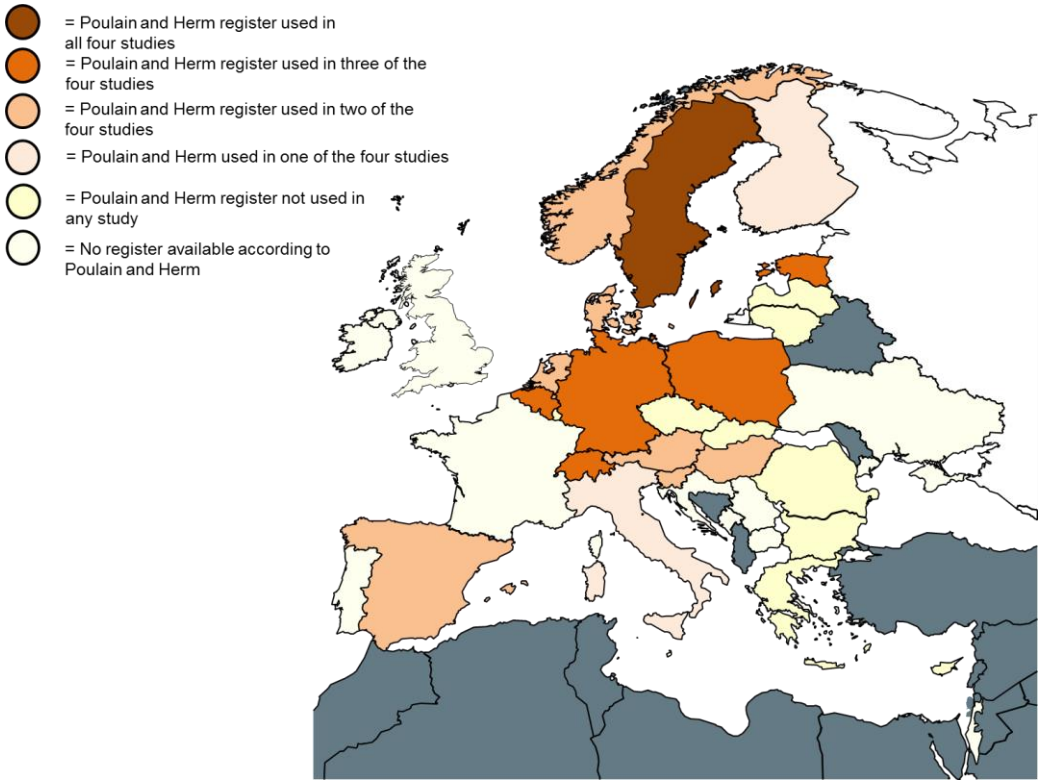


Table 4. Countries in which the known central register or the local civil register as listed by Poulain and Herm (2013) was used as sampling frame or in which the same register (other than the one listed by Poulain and Herm) was used across different studies.
Legend: None: The (central or local) population register listed by Poulain and Herm was not used in any of the four studies; "-": Country for which no central register was listed by Poulain and Herm; Underlined register name: The register reported by sampling expert is the same as listed by Poulain and Herm; Green coloured: At least two studies used the same register.

Country	**The register listed by Poulain and Herm was used in:**	**The same register[1] was used in:**	**Register name as reported in at least two studies or as listed by Poulain and Herm**
Austria	ESS, GGP		Zentrales Melderegister
Belgium	ESS, EVS, SHARE		Rijksregister/Registre national
Bulgaria	None		
Croatia	-		
Cyprus	None		
Czech Republic	None		
Denmark	EVS, SHARE		CPR
Estonia	ESS, EVS, SHARE		Population register
Finland	ESS		Population Information System
France	-	EVS, SHARE	Echantillon-maître (master sample based on Census)
Georgia	-		
Germany	ESS, EVS, SHARE		Einwohnermeldeamtregister
United Kingdom	-		
Greece	None		
Hungary	EVS, GGP		Népességnyilvántartó (Population Register)
Iceland	ESS, EVS		Þjóðskrá / National Register
Ireland	-		
Israel	-	ESS, SHARE	Population registry
Italy	GGP	EVS, SHARE	(1) Anagrafe; (2) Register of electoral lists
Latvia	None[2]		
Lithuania	None		
Luxembourg	None	EVS, SHARE	Social security registry
Macedonia	-		
Malta	None		
Montenegro	-		
Netherlands	GGP, SHARE		BRP (former GBA)
Norway	ESS, GGP		Central Population Register
Poland	ESS, EVS, SHARE		PESEL
Portugal	-		
Romania	None		
Russia	-		
Serbia	-		
Slovakia	None		

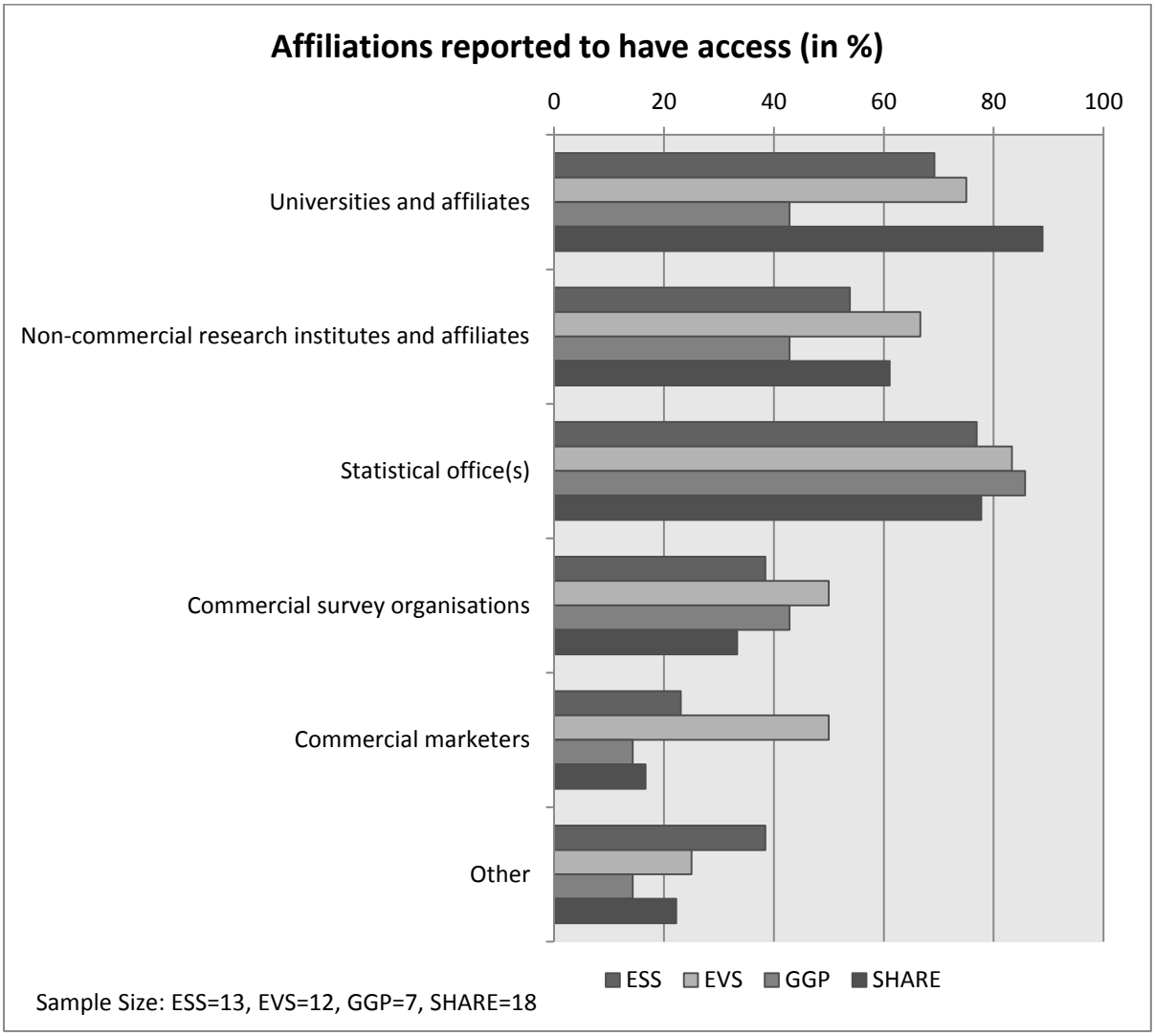
Country	The register listed by Poulain and Herm was used in:	The same register[1] was used in:	Register name as reported in at least two studies or as listed by Poulain and Herm
Slovenia	ESS, SHARE		Slovenian Central Population Register
Spain	ESS, SHARE		Municipal Population Register
Spain-Girona	-		
Sweden	ESS, EVS, GGP, SHARE		Navet/Total Population Register
Switzerland	ESS, EVS, SHARE		SRPH - Stichprobenrahmen für Personen- und Haushaltserhebungen
Ukraine	-		

[1] Another register than the register listed by Poulain and Herm (2013).
[2] Latvia was not included in the completed questionnaires of any of the four studies.

5.5 Access to population registers for sampling purposes

If the country sampling expert of a study reported that a register had been used as a sampling frame, the expert questionnaire asked them to specify to whom this register is accessible for sampling purposes: For universities, for non-commercial research institutes, for statistical offices, commercial survey agencies, commercial marketers, or also for others. Figure 7 lists the results across the four studies (detailed results per country can be found in the excel results table delivered with this report). Statistical offices seem to have the best access to population registers, closely followed by universities. Commercial survey agencies are granted access much less frequently, which might partly explain why several countries reported do not seem to use an available population register. As we could infer from the experts filling out our questionnaire as well as from our own experience, survey samples are often drawn or constructed by the commercial survey organisation that is hired to do the survey fieldwork.

Figure 7. Access to the reported register for sampling purposes. Includes only expert questionnaires in which the use of a register based sample was reported.



6. Access to population registers and data protection laws

In section 3 was described that not all countries in Europe have population registers and section 5 showed that even in countries where a population register is present, this is not always accessible or used in practice for sampling purposes. Universities and statistical offices are granted access in many countries but not all, and other research organisations more rarely have access. If we aim, in SERISS Work Package 2, at a joint effort to acquire more general access to population registers for sampling purposes, the logical first step would be to explore how the available registers that have not been used can be accessed in principle. First, countries have to comply with some general guidelines and principles established by supranational organs concerning data quality and protection in population registers. In this section, we therefore give a summary of the content of the major acts on data protection in Europe which are relevant for our aim of using population registers for survey sampling. We restrict the overview to the acts which are operative at this moment, preceding the forthcoming General Data Protection Regulation (GDPR) of the EU which will enter into application in May 2018.

The three major acts on data protection are (OSCE/ODIHR, 2009):

- The Organisation for Economic Co-operation and Development (OECD)'s "*Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*", firstly issued in 1980 and subsequently revised in 2013 (OECD, 2013);
- Council of Europes's (CoE) "Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data" (Council of Europe, 1981) ; and
- European Union's (EU) "*Directives on the protection of individuals to the processing of personal data and on the free movement of such data*" (European Union, 1995).

The OECD's guidelines (2013) established a version of Fair Information Principles (FIPs) which is now largely acknowledged around the world (Borgesius, Gray, & van Eechoud, 2015). These principles are:

- Collection Limitation Principle: there should be limits to the collection of personal data, which have to be obtained lawfully and, when appropriate, with consent of the data subject.
- Data Quality Principle: data should be relevant according to the purpose they are collected for, and as much accurate, complete and kept up-to-date as possible.
- Purpose Specification Principle: at the time of data collection, purposes have to be specified, and subsequent purposes should comply with the ones declared.
- Use Limitation Principle: data should not be disclosed for purposes that have not been specified, except under explicit consent of the subject or if decided by the authority of law.
- Security Safeguards Principle. Data should be protected. Nowadays, this principle also raises the issue of IT security: in case of electronic/digital registers, appropriate security measures should be adopted (OSCE/ODIHR, 2009).
- Openness Principle: there should be an overall policy of openness.
- Individual Participation Principle: Individuals should be able to access data and do have some rights on them (e.g. challenge to obtain rectification).
- Accountability Principle: the data controller should be accountable for complying with these principles.

The European Union (1995) and the Council of Europe (1981) directives on Data quality mostly overlap with the above principles, and can be summarized as follows:

Personal data undergoing automatic processing shall be:

- obtained and processed fairly and lawfully;
- stored for specified and legitimate purposes and not used in a way incompatible with those purposes;
- adequate, relevant and not excessive in relation to the purposes for which they are stored;
- accurate and, where necessary, kept up to date;
- preserved in a form which permits identification of the data subjects for no longer than is required for the purpose for which those data are stored.

(Council of Europe, 1981, p. 13)

Although the OECD´s "openness principle" could help to clear the road for the general use of population registers as survey sampling frames, it is likely that this aim is threatened by the "purpose specification principle", which is included in the OECD´s as well as in the European Union´s and the Council of Europe´s guidelines. In general, a population register is primarily an administrative tool and secondarily a tool for the production of demographic statistics (Poulain and Herm, 2013). The purposes specified at the time of data collection most likely do not include the use by survey researchers for sampling purposes and the transfer of person data (names and addresses) of sampled individuals to survey organisations.

Another obstacle is that many countries, in addition to adhering to the supranational guidelines, also have their own national data protection regulations regarding the use of population registers, which might be more restrictive than the supranational ones. However, it deserves further study whether the access restrictions experienced by the survey practitioners in our survey reflect national data protection regulations or whether these regulations would actually allow more general access than is now given in practice. With the forthcoming application of the GDPR, a single set of rules for data protection will apply to all EU countries. The harmonisation of data protection regulations throughout the EU will thus be improved and imaginably this could expand the possibilities to access registers in the countries which now have the most restrictive national protection regulations.

7. Discussion: Possibilities for synergy in the use of sampling frames in Europe

The objective of this report was to give an overview of sampling frames used in cross-national European surveys. The main output is the large overview and documentation file "SERISS_WP2_D2.1_Annex2.xls". This file is constructed in excel, delivered with this report and publically available at www.seriss.eu/resources/deliverables. It offers, firstly, an extensive documentation of the sampling frames used across the many countries included in the ESS, EVS, GGP and SHARE. Secondly, it can serve as a consultation source for survey practitioners and researchers in need of a sampling frame in a particular country or a set of sampling frames across different countries. Any new survey which is set up or existing surveys which are expanding can thus profit from the large experience gathered over years of surveying in Europe in the four large studies. As an example, SHARE extended to eight more countries in its seventh wave of data collection, in addition to the already participating countries. In the process of designing and preparing the samples in the new countries, the documentation excel file was already used to find appropriate sampling frames based on the experience of the GGP, EVS or ESS in these countries.

In addition to the delivery of the comprehensive excel documentation file, this report summarizes the findings across countries and across studies and compares the use of sampling frames in surveys to the availability of population registers in European countries as known from existing publications. From the overview given and the comparisons made, we can summarize the status of probability-based sampling in cross-European surveys in the following 8 points:

1. In ten countries, the use of a population register is reported by each study's country team participating in our expert survey. The country teams of all four studies declared

the use of a population register in Sweden. In nine countries: Belgium, Denmark, Hungary, Iceland, Israel, Norway, Slovenia, Spain-Girona, and Switzerland, we do not have information about all four studies but the studies included all reported the use of a population register.

2. In an additional 11 countries, the use of a population register is reported in one or more studies and other methods are given in other studies. These countries are: Austria, Estonia, Germany, Netherlands, Spain, Czech Republic, Finland, France, Italy, Luxembourg, and Poland.

3. In 18 countries, the same population register was used as a sampling frame for two or more studies. This was the case for: Austria, Belgium, Denmark, Estonia, France, Germany, Hungary, Iceland, Israel, Italy, Luxembourg, the Netherlands, Norway, Poland, Slovenia, Spain, Sweden and Switzerland.

4. In 16 of the total of 21 countries in which a population register was used in at least one study, this was the central or local population register as it is known from the most recent publication about available population registers across Europe, by Poulain and Herm (2013).

5. In ten countries the register which is available, according to the most recent publication about known central and local population registers in Europe, is not used as a sampling frame in any of the studies. These countries are: Bulgaria, Cyprus, Czech Republic, Greece, Latvia, Lithuania, Luxembourg, Malta, Romania and Slovakia. Of these countries, Luxembourg and the Czech Republic used a different register. In the other eight countries the sampling experts reported the use of other sampling methods, not based on register sampling.

6. Ten countries in which sampling experts participated in our survey do not have a known central or local population register according to the most recent published overview by Poulain and Herm, or are not considered in this publication: Croatia, Georgia, United Kingdom, Ireland, Montenegro, Portugal, Macedonia, Russia, Serbia, and Ukraine. Of these, Croatia and Georgia both indicated in at least one of the studies to have used an electoral register instead. Portugal and Macedonia indicated to have used other registers as well. The other six countries reported the use of other sampling methods which are not based on register sampling.

7. Statistical offices seem to have the best access to population registers, closely followed by universities. Commercial survey agencies have much less possibilities for accessing the registers, according to the country sampling experts participating in our survey.

8. The major acts on data protection in Europe, given by the OECD, the Council of Europe and the European Union, advocate some openness in use of the population registers but at the same time restrict the use to the purposes of the register as specified at the time of data collection. Such purposes are usually primarily administrative and statistical and might not encompass the use for social survey sampling.

The first result we found, that the four cross-European studies participating in the work package are all using one and the same population register for their samples in 10 different countries, is encouraging. It opens up possibilities for true synergy, by exploring ways to jointly build sampling frames from these same registers, from which each of the studies could then possibly draw its own sample. The countries in which we found mixed results, listed under (2) and (3) are especially interesting, since they offer the possibility to really profit from this expert survey and the exchange of experience between the four studies participating in the work package. With the information provided, we can, in the next stage, find out why some country teams do not or cannot use the population register that is accessible for other teams in the same country, and how we can improve that. For the countries listed under (5), having a population register which is not used as a sampling frame in any of the four studies, a major aim of the work package in the next years could be to found a working group or stakeholder group of European sampling experts which will support the country teams and help to put pressure on getting access to more of these registers. This effort has to look in more detail at the national and European data protection legislation to find ways to accommodate the importance of including good samples of the population in social surveys to the original purpose of population registers. Furthermore, it needs to be studied whether the access restrictions experienced by the survey practitioners in our survey reflect national data protection regulations or whether these regulations would actually allow more general access than is now given in practice.

A further question to possibly explore might be whether more registers could be used for sampling in our studies if the sampling request and sample drawing were done by university teams. As indicated in section 5.5, the samples for the four studies are often drawn or constructed by the commercial survey organisation that is hired to do the survey fieldwork, rather than by the university research team that has the scientific and financial responsibility for the study in their country. However, as the summarized result under (7) indicates, the commercial survey organisations have much less possibility to use the available population registers, so this might not be the optimal task procedure. Consequently, it also needs to be studied whether national data protection requirements allow that a sample obtained from the population register by the university team is then transferred by this team to a commercial survey agency which carries out the interviewing.

Most of the steps towards possible improvements in cross-European sampling which we propose here on the basis of our findings proceed from the assumption that the ideal sampling frame for all countries in the cross-European surveys is an official population register covering the population of interest. For the countries in which no central or local population register exists, as well as for countries where we do not succeed to get access to an existing register or countries where the register has known quality problems, we will nevertheless have to continue using alternative sampling methods. A secondary aim of the work package therefore is to jointly improve the quality control procedures for alternative sampling, for example by exploring the feasibility of building large common sampling frames using enumeration methods, from which different studies could draw their samples. In a subsequent report, SERISS Deliverable D2.2, we will evaluate the quality of the registers and other sampling frames used in European cross-national surveys.

References

Borgesius, F. Z., Gray, J., & van Eechoud, M. (2015). Open Data, Privacy, and Fair Information Principles: Towards a Balancing Framework. *Berkeley Tech. LJ, 30*, 2073-2073.

Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S., and Zuber, S. (2013). Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). International Journal of Epidemiology, 42, 992-1001.

Bristle, J., Butt, S., Emery, T., Luijx, R., Maineri, A.M., Pflüger, S-M., Scherpenzeel, A. and Zins, S. (2016) *Report on auxiliary data in available country registers*. Deliverable 2.5 of the SERISS project funded under the *European Union's Horizon 2020 research and innovation programme* GA No: 654221. Available at: www.seriss.eu/resources/deliverables

Council of Europe. (1981). Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Treaty No.108), ETS No.108 C.F.R.

De Luca, G., Rossetti, C., Malter, F. (2015). Sample design and weighting strategies in SHARE Wave 5. In: Malter, F. and Börsch-Supan, A. (Eds.). *SHARE Wave 5: Innovations & Methodology*. Munich: MEA, Max Planck Institute for Social Law and Social Policy.

Department of Statistics of the Republic of Lithuania. (1999). *The progress in the use of registers and administrative records*. Paper presented at the Conference of European Statisticians, Geneva, 1-3 March 1999.

European Union. (1995). Directive on the protection of individuals with regard to the processing of personal data and on the free movement of such data, No L 281/50 C.F.R. (1995).

Eurostat. (1997). Comparing data sources for measuring international migration in Central and Eastern Europe. In Eurostat (Ed.), *Eurostat Working Papers*.

Krabina, B., Prorok, T., & Tamm, H. (2005). *RISER: Trans-European eService in Civil Registration*.

Lange, A. (2014). The population and housing census in a register based statistical system. *Statistical Journal of the IAOS, 30*(1), 41-45.

Leti, G., Cicchitelli, G., Cortese, A., & Montanari, G. E. (2002). Il campionamento da liste anagrafiche: analisi degli effetti della qualità della base di campionamento sui risultati delle indagini. Rome: Presidenza del consiglio dei ministri. Commissione per la garanzia dell'informazione statistica.

Ludvigsson, J. F., Almqvist, C., Bonamy, A.-K. E., Ljung, R., Michaëlsson, K., Neovius, M., . . . Ye, W. (2016). Registers of the Swedish total population and their use in medical research. *European journal of epidemiology, 31*(2), 125-136.

OECD. (2013). The OECD Privacy Framework: Organisation for Economic Co-operation and Development (OECD).

OSCE/ODIHR. (2009). Guidelines on Population Registration. Warsaw: Organization for Security and Co-operation in Europe/Office for Democratic Institutions and Human Rights.

Poulain, M., & Herm, A. (2013). Central population registers as a source of demographic statistics in Europe. *Population, 68*(2), 183-212.

Redfern, P. (1989). Population registers: some administrative and statistical pros and cons. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 1-41.

Roberts, H., Rushton, L., Muir, K., Dengler, R., Coupland, C., Jenkinson, C., . . . Chilvers, C. (1995). The use of family health services authority registers as a sampling frame in the UK: a review of theory and practice. *Journal of epidemiology and community health, 49*(4), 344.

Simard, M. and Franklin, S. (2005). Sample design guidelines. In: United Nations Economic Commission for Europe (ed.). *Generations & Gender Programme*. *Survey Instruments*. New York and Geneva: United Nations: 3–14.

Thorvaldsen, G. (2011). Using NAPP census data to construct the historical population register for Norway. *Historical Methods, 44*(1), 37-47.

Thygesen, L. C., Daasnes, C., Thaulow, I., & Brønnum-Hansen, H. (2011). Introduction to Danish (nationwide) registers on health and social issues: structure, access, legislation, and archiving. *Scandinavian Journal of Public Health, 39*(7 suppl), 12-16.

United Nations. (1969). Methodology and Evaluation of Population Registers and Similar Systems In UN Department of Economic and Social Affairs (Ed.), (Vol. 15). New York: United Nations, .

United Nations Economic Commission for Europe. (2007). Register-based statistics in the Nordic countries. Review of best practices with focus on population and social statistics In United Nations (Ed.). New York and Geneva: United Nations

Annex

1. Questionnaire on National Registers - Generic version

2. Detailed sampling frame information “SERISS_WP2_D2.1_Annex2.xls”