

Analysis of existing migratory data production systems and major data sources in eight South-East European countries

Synthesis Report

Éva GÁRDOS

Hungarian Central Statistical Office

Irén GÖDRI

Hungarian Demographic Research Institute



2013

This research paper was developed in the framework of SEEMIG Managing Migration and its Effects in SEE – Transnational Actions Towards Evidence-based Strategies.

SEEMIG is a strategic project funded by the European Union's South-East Europe Programme. Project code: SEEMIG - SEE/C/0006/4.1/X. Lead Partner: Hungarian Central Statistical Office.

The research paper was prepared within the SEEMIG activity *Enhancing data production systems of migration and human capital in the South-East European area* coordinated by the Hungarian Demographic Research Institute at the Hungarian Central Statistical Office.

The information published here reflects the authors' views and the Managing Authority is not liable for any use that may be made of the information concerned.

© Gárdos, Éva – Gödri, Irén
All Rights Reserved.

Information for reproducing excerpts from this research paper can be found at www.seemig.eu. Inquiries can also be directed to: Hungarian Demographic Research Institute H-1024 Budapest, Buday László u. 1-3. or by contacting dri-seemig@demografia.hu.

Suggested citation: Gárdos, Éva – Gödri, Irén (2013): *Analysis of existing migratory data production systems and major data sources in eight South-East European countries*. Synthesis Report developed within the project 'SEEMIG Managing Migration and Its Effects in SEE – Transnational Actions Towards Evidence Based Strategies'.

<http://www.seemig.eu/downloads/outputs/SEEMIGDataSystemsSynthesisReport.pdf>

Acknowledgements

This paper has been elaborated by the SEEMIG experts at the Hungarian Demographic Research Institute and the Hungarian Central Statistical Office, in close cooperation with other contributing partners. The results are based on the SEEMIG country reports on existing migratory data production systems and data sources. The authors would like to express their thanks to all partners for their much appreciated inputs, to all reviewers, to the Transnational Working Group Members and Panel of Experts, as well as to discussants at the SEEMIG meeting in Belgrade: *Nada Stropnik* (Scientific Research Centre of the Slovenian Academy of Sciences and Arts) and *Tamás Kiss* (Romanian Institute for Research on National Minorities). Special thanks go to *Dragana Avramov*, independent appraisal expert of the project, for her useful comments on an earlier version of the report and valuable recommendations.

CONTENT

List of abbreviations and acronyms	6
1. INTRODUCTION	7
2. ADMINISTRATIVE DATA SOURCES.....	11
2.1. Conditions of the use of administrative sources in international migration statistics in SEEMIG countries	11
2.2. Types of administrative data sources	13
2.2.1. Population register	14
2.2.2. Register of Foreigners.....	16
2.2.3. Register of Asylum Seekers	17
2.2.4. Social Security Database.....	18
2.2.5. Other administrative data sources	19
3. STATISTICAL DATA SOURCES, SAMPLE SURVEYS.....	22
3.1. General overview	22
3.2. Full-coverage statistical data collections of the target population	22
3.2.1. Statistical register	23
3.2.2. Statistical data collection connected to administrative procedures	24
3.3. Survey-type data sources.....	26
3.3.1. Internationally standardized surveys.....	27
3.3.1.1. <i>European Union Labour Force Survey (EU-LFS)</i>	27
3.3.1.2. <i>European Union Statistics on Income and Living Conditions (EU-SILC)</i>	33
3.3.1.3. <i>OECD's Programme on International Student Assessment (PISA)</i>	33
3.3.2. National surveys	34
3.3.2.1. <i>Micro censuses</i>	34
3.3.2.2. <i>Other national representative surveys</i>	35
3.3.2.3. <i>Migration-targeted surveys</i>	35
4. CENSUS DATABASES.....	39
4.1. The types of the last censuses and the reference date	39
4.2. The population covered by censuses	40
4.3. Migration-related data	42
4.4. The availability and quality of census data	45
5. CONCLUSIONS, CHALLENGES AND RECOMMENDATIONS.....	49
6. REFERENCES.....	56
APPENDIX	59
General overview and evaluation of using administrative data for statistical purposes.....	59

List of Tables

Table 1: Administrative data sources in SEEMIG countries	13
Table 2: Characteristics of full-coverage statistical data sources in SEEMIG countries	23
Table 3: Characteristics of LFS survey in SEEMIG countries.....	31
Table 4: Examples of migration specific surveys in SEEMIG countries	36
Table 5: Type of the last census, the reference date and the population covered by census in SEEMIG countries.....	47
Table 6: Main migration-related data included in the censuses in SEEMIG countries	48

List of abbreviations and acronyms

EC	European Commission
EEA	European Economic Area
EEC	European Economic Community
EFTA	European Free Trade Association
EU	European Union
EU2	Bulgaria, Romania (which became members of the EU on 1 January 2007)
EU8	Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovakia, Slovenia (which became members of the EU on 1 May 2004)
EU15	Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Malta, the Netherlands, Portugal, Spain, Sweden, United Kingdom
EU-LFS	European Union Labour Force Survey
EU-SILC	European Union Statistics on Income and Living Conditions
HCSO	Hungarian Central Statistical Office
HDRI	Hungarian Demographic Research Institute
ILO	International Labour Organization
ISTAT	Italian National Institute of Statistics
OECD	Organisation for Economic Cooperation and Development
PIN	Personal Identification Number
PISA	Programme on International Student Assessment
SEE	South-East Europe
SEEMIG countries	Austria, Bulgaria, Hungary, Italy, Romania, Serbia, Slovakia, Slovenia
UN	United Nations

1. INTRODUCTION

Efforts to improve migration statistics have been made over the past 100 years or so, and many steps have been taken towards the standardisation of concepts and measurement.¹ Nevertheless, there is still lack of accurate and reliable migration data, and inconsistencies in data collection and measurement are still prevailing. Despite improvements in the harmonisation of migration-related data-collection, there are differences in the concepts used by national statistics, which are related to the history of the respective countries, as well as to their historical migration patterns (Fassmann, Reeger and Sievers 2009). Inconsistent migration data, the differences in concepts and measurement impede the international comparison of migration flows and stocks.

Issues related to international migration have been given increasing attention over the past decade, as it was recognized that international migration could play an important role in national development. The need for comprehensive and comparable migration statistics became more and more imperative. In 2007 the European Parliament adopted a new regulation on migration statistics², which provides clear definitions of immigration and emigration, and lists the migration indicators that must be transferred to Eurostat. This new regulation preserved the concept of *usual residence* and the *duration limit of one year* included in United Nation's recommendations adopted in 1997 – which defined an international migrant as a person who changes the country of his or her usual residence, and distinguished between long-term and short-term migrant on the basis of duration of stay (UN 1998). However, this concept and the range of migrants covered by this definition is considered increasingly vague, due to the diversification of migration types and increasing complexity of the phenomenon in the recent decades (Herm 2006, Fassmann 2009, Sik 2012).

Since the 2000s several projects have aimed at improving the reliability and validity of migration statistics, and promoting the comparability of migration data mainly in the European Union. Examples are COMPSTAT – Comparing National Data Sources in the Field of Migration and Integration (2001-2002)³, THESIM – Towards Harmonised European Statistics on International Migration (2004-2005)⁴, MIMOSA – Migration Modelling for Statistical Analyses (2007-2009)⁵, PROMINSTAT – Promoting Comparative Quantitative Research in the Field of Migration and Integration in Europe (2007-2010)⁶. These projects have also revealed that comparative research in migration in Europe is hindered by differences in definitions and sources and in the coverage of migration statistics, as well as by lack of relevant data and low reliability of existing data.

Most of the above mentioned projects dealt with EU member states or Western European countries where migration and compilation of migration statistics have a long tradition, going back to decades. Little has been done up to date to address migration issues and migration data in a broader European context. The SEEMIG project (2012-2014) – under which the present paper was developed

¹ The first set of recommendations on international migration statistics proposed by the International Emigration Commission (created within the structure of ILO) date back to 1922, and recommended that agreement should be reached on uniform definition and methodology to record migration (Herm 2006). Since then several set of recommendations regarding concepts, definitions and techniques to measure migration were proposed, but these recommendations have not been widely implemented in the course of data collections.

² *Regulation (EC) No 862/2007* of the European Parliament and of the Council of 11 July 2007 on Community statistics on migration and international protection and repealing Council Regulation (EEC) No 311/76 on the compilation of statistics on foreign workers.

³ <http://research.icmpd.org/1243.html?&F=zhoragodox#c2296>

⁴ Poulain, M., N. Perrin and A. Singleton (eds.) (2006), <http://www.uclouvain.be/en-14230.html>

⁵ Bijak, J. and A. Wiśniowski (2011), <http://2011.isiproceedings.org/papers/650247.pdf>

⁶ Kraler, A. and D. Reichel (2010), http://www.prominstat.eu/drupal/?q=system/files/PROMINSTAT_FINAL_REPORT.pdf

–, focuses on the South-East European (SEE) region, where countries started to face challenges of migration and migration statistics in the late 1980s and in the 1990s. Building on the results and experience of previous projects, SEEMIG intends to find ways to improve the system of migration statistics, increase the reliability and accuracy of data and decrease data gaps by taking into account special features of South-East European migratory patterns and of the data collection practices in the countries under review.

This paper analyses the major data sources on international migration in the context of human capital and labour market processes and gives a cross-national evaluation of statistical data production processes in the SEE region. This activity – as part of SEEMIG Work Package *Enhancing data production systems of migration and human capital in the South-East European area* – aims to highlight specificities of data sources and data production processes related to availability, reliability and comparability of migration-related data in different SEE countries and look beyond all of these. Analysis of data production systems in SEEMIG countries provides insights into processes of migration-related data production, and reveals the strengths and weaknesses of different data sources. This exploration and comprehensive description of data production systems is also necessary for the building of a SEE-level comparative statistical dataset containing all the major longer-term demographic, migration, human capital and labour market indicators relevant for developmental models and patterns (see: Fassmann and Musil 2012).

Although the available administrative data sources and their contents are closely connected with the migration policies of various countries, this document is not intended for policy evaluation and examination of the immediate impact these policies have on the scope of available data. Nevertheless, the SEEMIG project through various activities will also address the broader policy aspects of managing migration.

This report is a cross-national synthesis based on eight country reports of SEEMIG partner countries: Austria, Bulgaria, Hungary, Italy, Romania, Serbia, Slovakia and Slovenia. Country reports were elaborated by SEEMIG partners on a national level, based on a common methodology and a uniform system of criteria. The specifications for the country reports were prepared by the Hungarian Demographic Research Institute (HDRI) based on comments received from the Lead Partner and project partners. HDRI was responsible for the overall co-ordination of the analysis of data production systems. The *Conceptual Framework* (Fassmann and Musil 2013) and *Data Requirement Paper* (Fassmann and Musil 2012) prepared in the frame of Work Package *Conceptual framework for modelling longer term migratory, labour market and human capital processes*, were considered a comprehensive basis for designing the methodology of the country reports. In addition, a range of completed projects (e.g. THESIM, MIMOSA, PROMINSTAT) and publications also facilitated national-level analyses.

After the mapping of register-based databases, conducted censuses, international- and national-level sociological surveys, various data production systems were analysed in each partner country in order to identify positive features and inadequacies that hamper comparative analyses from a longer-term developmental perspective. The concepts and the categories of migrants and migration, including immigration/emigration flows, immigrant/emigrant stock, asylum seekers, return migration, irregular migration, remittances, and problems related to the definitions have all been taken into consideration. Project partners were asked to include all administrative and statistical data sources which contain data on the above-mentioned domains in their analyses, to provide critical description of current data production systems, and to highlight the changes that have occurred since 2001, corresponding to the 'newer period' of development. According to the common methodology, the country reports had to identify the breaks in the time series and their causes. It should also be mentioned that there might be more data sources with relevance to migration in a particular country than what the country report covers, but for various reasons the researchers or statisticians have no information about them. This mostly relates to the administrative data sources whose usability for

migration statistical purposes has remained hidden till recent past. In order to have a complex picture of available data sources in SEEMIG countries we additionally asked the participants to provide information on data sources that are likely to exist in most countries, but were not mentioned in the country report.

The national level analyses that served as key inputs for this report describe and evaluate the different types of above-mentioned migration-related data sources, data collection activities and data production processes in the given countries. They also include the institutional frame of data collection and data production, as well as the availability and quality of data. Individual country reports can be found on the project's website at www.seemig.eu and the methodological paper used as a harmonised template for the country reports can also be downloaded from there.

The aim of this comparative synthesis report is to bring together the main findings of the country reports highlighting the general, common features and differences as regards, for example, methods of data collection and data production, definitions used, availability of data, as well as the institutional and legal backgrounds concerning these differences. Cross-national comparison and evaluation of existing data production systems reveals – besides shortcomings – best practices of migration-related data collection in South-East Europe. This will be a platform for – and contribute to – recommendations for data enhancement, and strategies which aim to enhance data production. Based on the analysis of data sources and data collections, action plans will be developed and trainings will be designed for national, regional and local level administrations in order to enhance data supply and collection. SEEMIG is also a policy and development project, seeking to build the capacities of national, regional and local authorities to collect and utilise statistical data better in their planning, and to introduce evidence-based policy making and implementation. The national and cross-national evaluation of statistical data production systems in the SEE region aims to contribute to this.

The analysis concerning data production systems and major data sources is intended to be comprehensive and valuable for national/regional/local authorities or other entities responsible for data production, stakeholders (decision makers and civil servants) who apply or use the data, as well as international organisations and statistical bodies.

This analysis sheds light on similarities and differences in data collection and data production systems of the SEEMIG countries. As a range of entities directly concerned with data production, the SEEMIG partnership is directly involved in the compilation and delivery of data, and the analysis should therefore serve as a basis for supporting efforts to harmonise data in participating countries. Furthermore, the analysis also provides recommendations for administrative bodies who are in charge of keeping registers and providing data for statistical purposes on the one hand, and for those who rely on data when drafting policy frameworks at different territorial levels on the other. A wider audience, i.e. decision-makers and experts, will be directly involved when formulating recommendations and/or strategies in the respective national contexts – but further stakeholders should also be considered.

Countries within the scope of the project have different migration data production systems. Data relevance and data coverage regarding emigration and immigration also differ across and within the countries. Considering EU membership and the time of accession there are four categories represented within this project, though this categorisation does not necessarily correspond to the typical historical patterns of migration that characterise these countries. Nevertheless, to clarify and understand data sources and data quality we used the following grouping of the countries:

- EU15: Austria, Italy;
- EU8: Slovenia, Slovakia, Hungary;
- EU2: Romania, Bulgaria;
- EU Candidate: Serbia.

Old European Union member states (EU15) are represented by Italy and Austria. Both countries may now be regarded as immigration countries, although Italy has long experiences of emigration, and Austria has also been an emigration country for long. From the EU8 countries, which became members of the EU on 1 May 2004, this project includes Slovenia, Slovakia and Hungary, which have different experiences of immigration and emigration. The EU2 Romania and Bulgaria became EU members with the second wave of enlargement on 1 January 2007, and are largely characterised by emigration. Finally there is one country in the project, Serbia, which is not an EU member but a candidate country, and also has strong history of emigration.

The summary report is divided into three main parts. In the first one (chapter 2), we describe administrative data sources that contain data with the potential to be used for producing migration statistics. The second part (chapter 3) provides details of full-coverage statistical data collections that are closely linked to administrative procedures, as well as general sample surveys – which collect variables with migration relevance, such as citizenship, country of birth, place of foreign residence, etc. –, and specific migration-targeted surveys. The Labour Force Survey (LFS) is emphasised among the sample surveys because it is an internationally standardised instrument having – despite all its deficiencies – migration-related potential that has not been used to its full extent. Finally, the third part (chapter 4) describes censuses with high relevance to migration statistics even though they are conducted either every five or ten years and have different methods of implementation: exclusively register-based, exclusively enumeration-based (traditional), or a combination of these (such as register-supported census).

2. ADMINISTRATIVE DATA SOURCES

2.1. Conditions of the use of administrative sources in international migration statistics in SEEMIG countries

“Enlargement of the European Union has brought an added geographical and political dimension to the scale of the phenomena associated with migration. It has also brought a further impetus to the demand for accurate, timely and harmonised statistical information. There is also an increasing need for statistical information regarding the profession, education, qualifications and type of activity of migrants” (*Regulation (EC) No 862/2007*). The overall demographic problems of developed countries, resulting from declining fertility rates and rates of mortality, accompanied by increases in individual life expectancy and the increased permeability of the world hold increasing socio-economic relevance of migration. As the above-mentioned regulation emphasises, despite calls for harmonised data throughout the European Union these are still lacking. The main reason for this is that certain aspects of migration processes can only partially be assessed in a costly manner (if at all) by using sample surveys.

The SEEMIG countries have diverse experiences with using administrative data for statistical purposes due to the different legal background, different statuses of registers and different official attitude to secondary utilization of non-statistic data.

Considering the accessibility of personal-level data, the statistical authority of Slovenia is in an almost optimal situation because it has overall entitlement to access register data at the necessary level of aggregation. Moreover, the Personal Identification Number (PIN) is used in Slovenia as a key identifier for data access and for data integration in official and other administrative records. This usually enables organisations involved in data collection to collect it only once. Where this is not the case, the data source with the most complete data coverage and up-to-date information is used in the preparation of statistics. In Austria another solution was found. A special, anonymised PIN (pBK code) for statistical use was developed, which allows Statistik Austria to use anonymised administrative data at a personal level and to link data. The statistical institute gathers data from appropriate data sources and edits and links them into the population register (POPREG), which is the pivotal statistical database. PINs are also widely used in Bulgaria, not only by governmental authorities and institutions, but also by private companies. PINs are used, amongst other purposes, for the social security system, tax administration, for health insurance purposes, at employment agencies, for election lists, at banks, and by mobile operators. Nevertheless, the statistical office is not authorised to have direct access to the information systems of other institutions that contain data identified by the PIN, but is supplied microdata files based on PINs and containing information needed for the production of statistics. The system of PINs was also introduced in Hungary in the second half of the 1970s, but following the political and economic changes of 1989 the constitutional court drastically limited the use of PINs, although everybody is provided with one at birth, or in the cases of foreigners when any appropriate permit or certificate is issued. Other identification codes are used and are planned to be introduced in order to substitute the use of PINs in several domains of life.

The utilisation of administrative data is most effective if the data sources can be linked. The legal regulation of data linkage differs greatly between countries. In countries where a general PIN is used in administrative records data linkage is much easier and more effective than in the other countries where this is not possible.

The official connection between the statistical office and other data owners regarding the use of administrative data varies among countries. Accessibility of data and the relationship between the

Austrian and Slovenian statistical institutes and other data owners is fairly good. In cases where data are directly used for statistical purposes, Austrian and Slovenian statistical offices are usually involved in the design of the data collection or/and in the latest modification of the data source. In Hungary the right of the statistical office to use non-statistical data has also been increasingly appreciated and it is increasingly involved in administrative data collection issues by the data owners – although in many cases cooperation is not satisfactory basically due to the resistance of many ministries to providing administrative data for the HCSO at individual level.

Romanian migration statistics are a result of full-coverage statistical data collections rather than administrative registers. Immigrants in the registers are defined as foreign citizens, who come to Romania with the agreement of Romanian authorities to settle in the country. Emigrants are defined as Romanian citizens who choose their residence abroad in agreement with Romanian authorities. Consequently, the target populations of registers can not cover the migrants as defined in statistical requirements as there is no information on immigrant Romanian citizens and out-migrant foreign citizens.

In Serbia the roles and the activities of every relevant body for migration management are defined by a law specifying regulation and management of migration issues. The greatest challenge in Serbia is coordinating actions, which includes monitoring and managing migration. The need for coordinating competence between various institutions that monitor migration and migrants and for an improved cooperation between the relevant national institutions as the first precondition being harmonisation of definitions and updating of databases was recognised with the latest change in regulation of migration management, which anticipated formation of a coherent system for the collection, organisation, and exchange of data on migration.

Summarising the experiences of the SEEMIG countries, administrative data sources used in most countries for the purpose of migration statistics are the population register, the register of foreigners, the register of asylum seekers and the social security database. The former supports the counting of the officially defined foreign population residing in a country as regards immigrant stocks, emigrant stocks, immigration flows, emigration flows, re-migration flows, asylum seekers and acquisition of citizenship. The latter provides details about their lives, including economic activity, occupation, whether they commute to work, etc. Centralisation and computerisation engender significant potential advantages to these registers and their use.

The ministry responsible for interior affairs is usually involved in the storage of migration-related records in all countries.

Although statisticians and researchers increasingly use administrative data sources for migration statistical purposes, in many cases users' knowledge on registers is deficient.

On the basis of international experience, the use of administrative data is much more effective in small countries where keeping central registers is much easier than in countries with large territories and populations such as Italy. Problems that statistical experts or researcher have to face in using administrative data are compounded by further difficulties when multi-central data collections are available.

2.2. Types of administrative data sources

We describe below the administrative registers that are potentially usable for statistical purposes, irrespective of their actual (and current) use (for an overview of these see *Table 1*). Moreover, we characterise and contrast the different related practices of the countries regarding migration statistics.⁷

Table 1: Administrative data sources in SEEMIG countries

Country		General data sources	Migrant-specific data sources
EU15	Austria	<ul style="list-style-type: none"> - ZMR: Central Register of Residents, Federal Ministry of the Interior - BKA: Federal Criminal Office Database, Austrian Criminal Intelligence Service - HV: Central Social Security Register, Main Association of Social Insurance Carriers - Unemployment Database, Public Employment Service 	<ul style="list-style-type: none"> - BFIS: Federal Alien Information System, Federal Ministry of the Interior <ul style="list-style-type: none"> o FIS: Alien Information System (EEA nationals and third-country nationals), Federal Ministry of the Interior o AIS: Information System on Asylum Seekers, Federal Asylum Office o BIS: Information System on Federal Care of Asylum Seekers, Federal Asylum Office o Smuggler Database, Federal Criminal Intelligence Service - Work permits, Public Employment Service
	Italy	<ul style="list-style-type: none"> - Municipal Population Registers <ul style="list-style-type: none"> o Resident change of population o Live birth data by citizenship 	<ul style="list-style-type: none"> - Entry visas, Ministry of Foreign Affairs - Register of emigrants (Aire), Ministry of the Interior - Register of residence permits (third-country nationals), Ministry of the Interior and Italian National Institute of Statistics - Acquisition of Italian citizenship, Ministry of the Interior - Foreign workers' remittances, Bank of Italy - Sub-registers of Municipal Population Registers <ul style="list-style-type: none"> o Foreign citizens (stock)
EU8	Slovenia	<ul style="list-style-type: none"> - Central Population Register, Ministry of the Interior - M Forms: Reports of data regarding pension, disability and health insurance, parental insurance and unemployment insurance; the Health Insurance Institute - Balance of payments, Bank of Slovenia - Employment register based on social security insurance data 	<ul style="list-style-type: none"> - Register of Aliens (EEA nationals and third-country nationals), Ministry of the Interior - Asylum Register, Ministry of the Interior - Data on illegal migrants; Ministry of the Interior, Police - Work Permits, Employment Service
	Slovakia	<ul style="list-style-type: none"> - Population register, Ministry of the Interior 	<ul style="list-style-type: none"> - Evidence of Foreigners (ECU), Ministry of the Interior <ul style="list-style-type: none"> o EEA nationals o Third-country nationals - Information system Migration and International Protection, Ministry of the Interior - Information system of employed foreigners, Central Office of Labour, Social Affairs and Family - Information System MIGRA (of refugees), Migration Office of the Ministry of the Interior - Central Register of the Acquisition and Loss of the Citizenship of the SR, Ministry of the Interior

⁷ Although country reports are available only for the SEEMIG countries, we try to provide general conclusions on the types of registers that may provide usable information in any country of the world. This summary therefore does not list each individual practice, just typical ones. Full details are available in the country reports.

Country		General data sources	Migrant-specific data sources
	Hungary	<ul style="list-style-type: none"> - Central register of personal data and addresses, Ministry of the Interior - Educational register, Office of Education - Register of social security PINs, National Health Insurance Fund Administration - Register of pension insurance, Central Administration of National Pension Insurance - Database of National Tax and Customs Administration 	<ul style="list-style-type: none"> - Central Immigration Register, Office of Immigration and Nationality (OIN) <ul style="list-style-type: none"> - EEA nationals (included irregular migrants) - Third-country nationals (included irregular migrants) - Refugee Affairs Information System, OIN - Register of foreign workers, National Employment Service
EU2	Romania	<ul style="list-style-type: none"> - Population register, Ministry of Internal Affairs - Register of the education, Ministry of Education 	<ul style="list-style-type: none"> - Register of short-stay visa applicants, Ministry of Internal Affairs - Registers of EU/EEA/Swiss Confederation citizens and third-country nationals residing in Romania, Ministry of Internal Affairs - Register of asylum seekers and persons granted protection in Romania, Ministry of Internal Affairs - Register of applications for reacquiring Romanian citizenship, National Agency for Citizenship
	Bulgaria	<ul style="list-style-type: none"> - Unified System for Civil Registration and Administrative Service of Population (Population Register), Ministry of Regional Development and Public Works - Health insurance contributions fund register, National Revenue Agency - Register of the education, Ministry of Education and Science 	<ul style="list-style-type: none"> - Registers on Bulgarian citizenship, Ministry of Justice <ul style="list-style-type: none"> o Register of persons who have acquired Bulgarian citizenship by naturalisation o Register of persons who have lost their Bulgarian citizenship o Register of persons who have had their Bulgarian citizenship reinstated - Information on refugees, asylum and humanitarian status, State Agency for Refugees - Register of Foreigners (EEA nationals and third-country nationals), Ministry of the Interior - Register of EU citizens residing in the Republic of Bulgaria, Ministry of the Interior
EU Candidate	Serbia	<p>Population register does not exist, Social Security Database is available but not used for migration statistics</p>	<ul style="list-style-type: none"> - Migration database, Ministry of the Interior <ul style="list-style-type: none"> - Immigrants - Asylum seekers - Acquisition of citizenship - Irregular migration - Return of citizens of RS – based on Readmission Agreement - Register of refugees and internally displaced persons (IDPs) from Kosovo and Metohija, Commissariat for Refugees and Migration

2.2.1. Population register

The population register is a regulated administrative database. It is centrally maintained and generally contains information about the entire population residing in a country. It primarily aims to record residence-related data of persons registered in a country. The population register exists in every SEEMIG country except Serbia, where it is only just being established following new legislation which came into force at the end of 2012. Coverage of the foreign population is slightly different by country depending on the legislation. In Hungary only foreign people considered as permanent residents are included, and consequently people from third countries with residence permit (which entitles the owner for temporary staying in the country) are excluded. This differs from most countries, where the population register covers all people resident in a country for more than three months.

In most countries the population register also provides information for statistical or scientific purposes on naturalisations – persons acquiring the citizenship of a country through application. In some countries (such as Bulgaria, Italy and Romania) the administrative register of acquisition and loss of the citizenship is used for statistical purposes separately from the population register, but those granted citizenship are also reported in the population registers.

Although Romania has a population register it only covers Romanian citizens irrespectively of their country of residence and it is not used for statistical purposes, except the electoral list. The Slovakian population register is not used for statistical purposes either, except for in the preparation of censuses where it is applied as the statistical frame of the data collection.

The centralised electronic registers developed at different times in SEEMIG countries, ranging from the beginning of the 1990s to the beginning of the 2000s.

Information for updating register data (birth, death, change of the name, change of marital status) is generally directly entered into the population register through the official system of administration. A change of addresses is statutorily declared by concerned people. This obligation is complied with unevenly in different countries. For example, in Austria failure to declare the main place of residence is considered an administrative offence and may result in a fine; registration is frequently used in official documents so data users need to have confidence that the content is accurate and reliable. However, people in Austria can have an unlimited number of secondary residences, which can mean that the register does not provide reliable information on the actual place of residence. Nevertheless, according to expert opinion the register provides reliable information in statistical terms. In Hungary, where certain advantages (selection of a school or the price of car insurance) are connected to the address of the usual residence, the registered data frequently do not reflect the location of the true residence.

In Italy there is a special registry (Aire) in every municipality, which gathers the registrations of all Italians who have transferred residence to another country for more than a year. A living person with Italian citizenship can appear either in the population register or in the Aire. Despite this, the number of Italian emigrants and Italian returned migrants is underestimated.

A population register is a collection of data originating from many other administrative data sources. Migration-related personal pieces of information in the population register are usually citizenship and the country of birth. Further data are name, sex, date of birth, marital status and address. Austrian records additionally include country of previous residence for immigrants, and country of destination for emigrants. The Slovenian and Bulgarian population registers comprise information on family relationships, with each person's record including the mother's, father's, children's and spouses' PINs. Although the Slovakian register cannot use PINs, the family relations of those permanently residing in the country are still recorded. The population register in Romania also includes data on educational attainment, occupation and military status. The Italian register contains specific information concerning education, working condition, working position, sector of economic activity and citizenship.

In order to ensure interoperability among administrative data collections the population register provides data to other administrative datasets. Consequently, the population register is connected to many others and this guarantees a relatively high quality of data. The population register, even if it can be considered reliable, reflects the *de jure* rather than the *de facto* population. Moreover it does not contain the date of entry of the foreign population.

Population registers are not usually set up to meet the demands of population or migration statistics. Consequently, they have the problems that generally characterise administrative data (e.g.: relating to the de-jure status rather than de-facto one; including only the indicators necessary for the administrative process). Furthermore, the usage for statistical purposes differs by country. In Austria, for example, the so-called *POPREG* ("POPulation REGister") has been set up as a "statistical twin" of

the Central Register of Residents); it serves statistical purposes rather than the original, administrative register, and is the basis for the newly developed register-based demographic system at *Statistik Austria*, which has been fully operational since 2002. This contrasts with the case of Italy where the population register is held by the statistical office and where data is collected from the municipalities.

In most countries the difference between the administrative concept of a migrant (actual or intended duration of stay for one year or more) poses the problem of how to come up with accurate estimations of the actual migrant population. Indeed, there is no information in the register that relates to the intention (or not) to stay for more than three months, the time limit of obligation to apply for residence permit, not taking into consideration the statistical concept of the migrant that uses one year of time limit.

Data on emigration is likely to be under-estimated in all countries, due to a lack in incentives for self-deregistration for persons moving abroad, particularly in the case of temporary emigration. For this reason Austrian registration offices additionally try to gather information on emigration retrospectively, in order to preserve the quality of migration statistics. However, the Austrian authorities have insufficient resources and tools for further improving coverage. Recently introduced legislation in Hungary accommodated to the practice of the population that can be characterized in this respect as avoiding legal rules, and overturned the obligation of the population to declare temporary emigration. Considering the emergence of international migration this shortcoming of the reporting system creates additional administrative problems, as well as non-reliable statistical data. The underestimation of out-migrants, especially in the case of emigration countries, yields the lack of accurate figures on the population size which in itself is a problem, but in addition biases basic demographic as well as certain economic and social indicators. Characterizing the different practices followed by countries, in Italy information about the emigrant Italian population is gathered in a separate register.

The population register tends to have the most potential among administrative registers from the point of view of the actual size of the migrant population and its distribution by basic characteristics. This data collection can cover the widest part of the population including migrants. The population register is generally fed by several other data sources, making it possible to cross-check data before final inclusion in the register. Population register data is regularly and methodically updated.

2.2.2. Register of Foreigners

The register of foreigners is generally operated as an independent administrative register according to legislation, containing information on applications for residence, issued and rejected residence permits, issued and rejected visas and data on law enforcement measures regarding foreigners. In the SEEMIG countries that are EU members, foreigners who have the right to free movement and reside in the country are also recorded in this register. Italy is an exception among SEEMIG countries because it only registers so-called third country nationals. Registers in SEEMIG countries usually work according to regulations passed between 2005 and 2010. From the statistical point of view it is very important to have harmonised data from the different data sources and this can only be achieved if they are integrated. The Hungarian register of foreigners announces foreign citizens and changes of their data into the population register. The Slovenian Register of Aliens has also been integrated with the Central Population Register.

Foreigners with the right to free movement in the EU and residence in a Member State are in most countries handled in a different database of the foreigners' register than citizens of third countries. The coverage of the former (part of the) register is moderate, as in some countries EU citizens are neither obliged to enroll in the register of foreigners nor to the population register (Bulgaria, Italy, Slovakia) while in others they frequently fail to record themselves (Austria).

Residence permits are classified according to the type of residence (permanent or temporary) and according to the reason for stay in the country including the possibility of pursuing work. In each country residence permits of several durations are available, and these differ by country.

A person's record is entered when the first application is submitted or the first residence permit is issued. Updates are connected to changes in the legal or personal status of a person, or as a result of an administrative procedure (e.g. new application, termination of residence). The recorded person-related data are suitable for providing an overview of the distribution of the foreign population that have entered the country in a certain period by the most important demographic characteristics, such as age, sex, citizenship as well as the type of the stay. The country of birth is usually not included (Austria, Hungary, Italy, Romania, Serbia, Slovenia). Different countries collect different additional pieces of information. In Austria, Slovakia and Hungary information on educational attainment, occupation or economic activity is included in the records and experiences with it are similar: statistical usability is very limited either because the data are not captured exhaustively, or because they only refer to first registration and are not subsequently updated. The Slovenian register comprises not only occupational status but marital status too. The latter is also registered in Slovakian records, where data on family members (parents, spouse, children and siblings) of the third country's citizens are also stored.

In Romania, the register of foreigners embraces the same groups of the population as it does in other countries. However, it is not yet used for statistical purposes, because the responsible ministry does not transfer data directly from the registers to the statistical office, and it has its own annual publication on the stock and flow of immigrants. In Bulgaria, promising cooperation has started between the administrative and statistical authorities to improve the data quality, and the statistical office is already authorised to use data with personal identification, thus making it possible to link individual records – crucial for production of reliable and correct statistics on the number of immigrants and the usually resident population. The Slovakian statistical office is allowed access only to aggregated data, though the office works with the ministry to improve the quality of data.

In Serbia, as in the most countries, the Ministry of the Interior is responsible for data collections on the migrant population. However, only the statistical office can access data processed and published by the ministry in the so-called Migration Profile. The database comprises information on immigration flow, immigrant stock, asylum seekers, acquisition of citizenship, as well as on return of Serbian citizens according to the Readmission Agreement.

Recording irregular migration in the broad system of the register of foreigners is mentioned by the Hungarian and Serbian country reports.

The register of foreigners is not a full-coverage source on stocks of foreigners because only asylum seekers who are 'accepted for processing' and 'recognised' are recorded, and the group of EEA or Swiss citizens is insufficiently covered because persons concerned often fail to apply for this documentation. A general problem of the migration data in all countries is that emigration is hard to capture. Even if emigration is estimated on the basis of expired settlements or residence permits the problem of permits without expiration date remains. Consequently, the registers overestimate the stock of foreigners. Moreover, actual duration of stay frequently cannot be stated due to inability to link consecutive records of the same person.

2.2.3. Register of Asylum Seekers

This register is generally maintained to record asylum procedures from application to the decision of the authority. Data refer to all persons whose stay is regulated by the national asylum law. Regulations in most countries were updated between 2005 and 2010.

The register files data for processing asylum claims, mainly personal data on asylum seekers and the state of the procedure. Single records are compiled per asylum claim, and separate records are hitherto not consecutively interlinked. Records generally comprise the following variables: name, date of birth, sex, citizenship, residence and state of the asylum procedure. In the Austrian dataset information on ethnicity is also included in textual form, but only for procedural reasons relating to asylum and this is not used for statistical purposes. The Slovenian register comprises many other pieces of information: the place of birth, marital status, education and profession, data on employment and income, religion, and data on medical conditions. The information under data protection is collected with written consent. Like the Slovenian register, the Serbian register is also very rich in information on refugees. Cooperation with the data owner is satisfactory and the data are accessible to the statistical office. Although Serbia already has separate records for asylum seekers and refugees from the former Yugoslav republics, the data cover only those who were forced to leave their places of residence in the former Yugoslav republics on account of the events of the disintegration of Yugoslavia and the civil war that took place in the period of the 1990s. The Romanian register, on the top of the basic information, includes data regarding the family of the applicant, the route travelled from the country of origin to Romania, and data regarding previous asylum applications in EU member states or non-EU countries. The Slovakian register includes information on religion as well.

In Bulgaria the responsible authority produces statistics as well, and the statistical office is informed on a monthly basis about foreigners who are granted refugee or humanitarian status through the population register.

As the authority generally pays little attention at the time to the sequential number of asylum claims submitted by a person, and registers maintain data from a certain point of time only, there is usually no information on first-time asylum seekers. However, the Slovenian register, as an exemption, is able to differentiate between the first and subsequent applications.

2.2.4. Social Security Database

This register generally embraces all persons under the social security system, employed and self-employed persons as well as dependent family members, having either compulsory, voluntary or other types of state insurance. It is compulsory for persons who are gainfully employed in Europe to provide social insurance contributions. Private or semi private schemes were not mentioned in any of the country reports as the actual or possible data source of migration statistics.

Social security registers are perhaps the richest sources of data in terms of migration statistics. This is because they cover a wide range of socio-demographic factors. However, different amounts of data are used for the purpose of migration statistics in the different countries. This database usually makes it possible to select different types of immigrants. However, as pointed out by Serbian experts, social security data in Serbia are relevant and reliable for requirements of compulsory social insurance, but not for an analysis of categories of migrants and migration events. Namely, the Serbian database includes data from all employed persons in the country and citizens of Serbia who work abroad but do not have compulsory social insurance in the foreign country, as well as all foreigners who are studying or undergoing specialised training.

The foreigners' register generally provides reliable information on immigration, but those leaving the country frequently do not declare this fact to register. The social insurance database comprises regularly updated data that contribute to having more exact information on the usually resident population. The Bulgarian statistical office searches additional information channels to improve coverage of migrants. Improving the under-estimated number of emigrants is considered an important task. As such, they use health insurance data on persons who have declared refusal of payment of health insurance contributions. In Hungary there is a similar desire to have more precise

information and to make use of health insurance data on the size of the immigrant population living in the country and on the Hungarian and foreign citizens who leave the country.

The following migration-relevant variables are usually included in databases of the social security: age, sex, citizenship and legal information on insurance. In Hungary, the country of birth, place of residence and the date of commencement of insurance are also available. The Austrian register collects additionally occupation type, occupational status, place of employment, (declared) income, number of employment relationships and access to long-term benefits (e.g. in case of unemployment). Beyond basic personal data, Slovenian records include data on education, occupation, type of work contract, type of work, hours worked per week and the work permit number. In Hungary and Slovenia, personal data are directly updated from the population register. Only Slovenia uses social security data sources for employment statistics with the citizenship variable to separate migrants and nationals. Overall, data may be considered rich for statistical purposes in this country.

2.2.5. Other administrative data sources

The **register of visas** is also mentioned as a means of measuring a certain immigration segment in the Italian and Romanian country reports. The granting of visas to foreign citizens varies depending on the citizenship, the country of usual residence, and the length of and reason for the stay in the country. EU citizens can enter the countries within the Schengen system without needing a visa. Besides selecting foreigners by the above-mentioned factors, the most important limitation of this type of administrative data source in relation to migration statistics is that the duration of stay authorised by issuing a visa does not reach the standards of the statistical definition of an immigrant.

Illegal or irregular migration is defined as a situation where a foreigners' entry, stay or work is contrary to the existing laws which regulate their entry, residence and economic activities. There are only three SEEMIG countries that mentioned data sources other than the register of foreigners concerning irregular migration. Slovakia and Slovenia maintain databases specified for migrants, while Austria uses the general criminal database and the so called Smuggler Database. Administrative **registers** on irregular migration maintained by **police authorities** are mostly available, but irregular migration remains very much a hidden phenomenon and is therefore poorly covered. The amount of statistical detail and usability varies greatly by country. However, data on refused entries at the border, persons who were caught due to illegal border crossing and persons who illegally resided in the country are available in criminal databases. Citizenship is usually contained within the collected data, but in certain countries the exact country of citizenship is not available or accessible. However, in Slovenia all statistics are differentiated by citizenship. In the Slovak Republic a database of the flows of irregular migrants and their characteristics is maintained in the framework of the "Information System on Migration and International Protection". In most of the SEEMIG countries the statistical institute does not use the data of criminal database for migration statistics certainly in order to avoid stigmatization of the migrant population, and the data owner directly provides indicators for national and international bodies.

The available statistical information on irregular migration is in general poor and varies greatly between countries. The Austrian authorities issue several publications on irregular migration: on organised smuggling according to citizenship and other variables, quarterly crime statistics broken down by citizenship, criminal cases committed by foreigners from selected countries, and security reports on organised crime, human trafficking, prostitution, smuggling and illegal migration. In Slovenia yearly reports on irregular migration are produced and various indicators of irregular migration are included within it.

The Serbian and Hungarian foreigners' registers include information on irregular migration. Serbian data relate to denied entry, return of persons illegally residing in the country as well as victims and

perpetrators of human trafficking. For denied and returned people their citizenship is among the recorded data. The scope of the Hungarian registers is different. Third-country nationals are included if they have been ordered to leave the territory of the country, are subject to compulsory confinement, expulsion, exclusion or detention, are prohibited from leaving the country, if they have been detained, arrested or taken into custody in Hungary, or if they have been affected by some extraordinary event (i.e. death, accident resulting in serious injury, etc.). The reasons are only recorded in cases where criminal proceedings are initiated. EEA citizens and family members of those who are registered are recorded if they are subject to expulsion or entry ban or restriction of personal liberty. The registered data comprise the place of birth and citizenship. It can be supposed that the police and the immigration authority maintain registers on irregular migration in other countries, too, but this part of migration statistics is under-developed; there is currently no common methodology and therefore the relevant data sources are not considered as components of the migration data assets. Even in Hungary, where these data are accessible for statistical purposes in personally not identifiable form, the statistical office has never required them because these data provide very little additional information and international standards are missing in this statistical domain.

Another administrative data source which is related to the labour market is available in six countries: Austria, Hungary, Italy, Romania, Slovakia and Slovenia. The content again varies greatly by country, particularly regarding coverage, content, accessibility and the methods for maintaining the databases. These registers are kept by the offices responsible for employment issues. In Slovenia, only those who need work permits are registered, while in Hungary the content of registered information is more detailed for the third-country nationals than for those of European Economic Area (EEA) countries, and registered data of all employed foreigners are the same in Slovakia irrespective of the citizenship. The Italian register overestimates the foreigners as it uses the country of birth to identify foreigners, while the Romanian one underestimates them.

In EU countries, data gained via the register on work permits will decrease in significance in the future because the share of immigrants EU nationals who do not require work permits is increasing steadily and significantly. It seems that other data sources (e.g. personal tax data combining with the compulsory social insurance data) will be more informative than the register of work permits. This notwithstanding, in some countries this data source provides an important contribution to migration statistics/information (e.g. Slovenia). It may be supposed that the significance of the register of work permits depends on the share of third countries' citizens among all foreigners who are employed. In most cases data on employment of foreigners are not only collected and used for administrative purpose by the competent office, but are also used and published as statistical indicators too.

In Austria we see an exceptional situation, as they have a database specified for statistical and research purposes, maintained by the Labour Market Service. It is built on data stemming from the administrative registers of the Labour Market Service on the one hand, and from the Central Social Security Register on the other. These databases include citizenship among the variables.

Registers of the education system such as the Public Education Database and the Higher Education Database in Hungary can also provide migration related information. These have been developed since the beginning of 2000s and kept for recording data of maintainers, institutions, employees, pupils and students. Personal data are reported and updated in relation to events. Educational identification numbers are attached to all employees as well as pupils and students who participate in the education. The personal subregisters cover the following data for employees: personal data, addresses, identification number, education, professional qualifications, and details of the employer. Collected data for pupils and students includes personal data, addresses, an identification number, and data of the educational institution attended. The migration-related pieces of information are citizenship, country of birth and country of permanent residence. Citizenship is not included in the

employees' subregister of the public education register. For foreign students information on the grounds of stay in Hungary is also included.

Similar data source is also available in Bulgaria and Romania, but it is used for migration statistics purposes only in Italy. The Hungarian system is still under development, the Bulgarian and Romanian databases suffer from quality problems.

Transferred **remittances** can be considered one of the indicators describing benefits of emigration. In certain countries it is included among regularly produced statistics, while in others information is less often used from a migration perspective. Data are provided by the central bank of the country. Nevertheless, even in the countries where data on remittances would be useful for statisticians and researchers in the domain of migration this indicator has substantial shortcomings if it is produced according to the methodology defined by relevant EU regulation, because remittances under a certain amount are not included. Moreover, the banks are often unable to identify the country of residence of the final transferee of the transaction and this results in poor quality of the international territorial break down of the bank system statistics. On the other hand, remittances are the transfers running through the regular intermediation channels, whereas the informal channels remain excluded.

In Italy, the National Observatory on Migrants' Financial Inclusion was set up in 2012. This operational body serves to analyse and monitor migrants' inclusion within financial processes. Every year it presents research which studies in depth the main characteristics of the bank–migrant relationship as regards savings, credit, money transfer services and micro finance. The research agenda also covers the channelling of remittances through official channels as well as the development of migrant savings.

3. STATISTICAL DATA SOURCES, SAMPLE SURVEYS

3.1. General overview

This chapter maps various statistical data sources and sample surveys conducted in the SEEMIG countries. The main aim is to provide an overview of key specialised surveys which *focus specifically on migration* and surveys with only *some migration-related questions*, as well as *micro censuses* and other sociological surveys that have *migration relevance* in the countries examined.

Comparison across countries' data sources requires clarification of the general and the particular character of the given statistical data sources or surveys, since data collection and data coverage might vary in the eight countries. Furthermore, countries differ in terms of their migration experience, which influences the purpose and target of country-specific data as well as coverage of migrants. Country reports cover various data sources, and it may be assumed that the presented data sources correspond to countries' interests and the relevance of different sources. In addition, they may address emigration, immigration or both emigration and immigration. Consequently, the mapping of statistical data sources and sample surveys aims to clarify the above-mentioned similarities and differences of data sources, and to present best practices and attempts that countries have made (or are planning to make) to overcome data source deficiencies. These best practices may serve as examples to other countries.

Some of the surveys are comparable, as they are part of the European statistical data collection and are conducted in all countries according to international standards using the same concepts and methods (e.g. *Labour Force Survey*), while others are different and follow national concepts which are hardly comparable.

When mapping migration-related data sources, the dimensions of both emigration and immigration have to be considered. By definition, **immigration data** – as stock statistics – are easier to collect since resident population consists of nationals, foreign citizens and/or foreign-born persons. Consequently, data sources more often provide data on immigrants. Nevertheless, range and relevance of immigrant data is low in countries where immigration is a marginal issue. At the same time, gathering and clarifying **emigration data** is problematic and requires special effort. Although administrative data on deregistering of those leaving the country would provide data on emigration, this data is only partly relevant and is often misleading. More generally, emigration data are usually under-represented as a result of various factors concerning data collection or the nature of emigration.

In the following sections we group the data sources presented into two relevant groups: (1) full-coverage statistical data sources and (2) survey type data sources. Further grouping – by different aspects – is possible as follows: general data sources and migrant-specific data sources (respectively migrant-targeted surveys), international surveys and national surveys, periodic (or repeated, respectively panel) surveys and one-time surveys, as well as surveys with a focus on immigration or emigration. Types of full-coverage statistical data sources and sample surveys will be discussed in different sections since relevance and comparability of these sources are largely different.

3.2. Full-coverage statistical data collections of the target population

There are many reasons why a statistical institute relies on full-coverage statistical data collections. The statistical register is one example of such a data collection; it is a continuously updated set of objects for a given population, containing information on identification and other attributes. A

statistical register can be used to support the statistical surveying process of the register population, but it can also serve as a data source for the production of statistical indicators. The register contains the current and historical statuses of the population and the causes, effects and sources of alterations in the population. Register data of population units are stored in a structured database. A register is compiled using several data sources, including administrative and statistical ones. The overlapping data of data sources used for feeding the statistical register are used for cross-checking, and the appropriate one is selected by predefined regulations or rules.

The other major group of full-coverage data collections consists of those that are attached to administrative procedures, which substitute and supplement the administrative data. This kind of survey is used when good quality administrative data are not accessible for statistical use or the information provided by the administrative data source does not meet statistical needs. The connection to an administrative process makes such data collections cheaper than a sample survey with specially trained interviewers.

Depending on the scope of administrative data sources available for statistical purposes, and on the whole statistical system, full-coverage statistical data collections differ greatly between countries (see: *Table 2*).

Table 2: Characteristics of full-coverage statistical data sources in SEEMIG countries

Country		General data sources available	Migrant-specific data sources
EU15	Austria	-Statistical population register (POPREG), Statistical Office -Educational Attainment Register, Statistical Office -AMDB: Labour Market Database, Labour Market Service	
	Italy	Education database, ISTAT	
EU8	Slovenia	- Statistical Survey on Student Enrolment in Tertiary Education, Statistical Office	
	Slovakia	-Statistical demographic register (marriage, divorce, birth, death, change of permanent resident), Statistical Office	
	Hungary	- Demographic database (marriage, divorce, birth, death, change of permanent resident), Statistical Office - Educational statistical databases, Ministry of Human Resources	- Statistical survey on people acquired Hungarian citizenship, HCSO - Statistical survey on foreign citizens with settlement document, HCSO
EU2	Romania	-Demographic data collection of (deaths, births, immigrants and emigrants), Institute of Statistics	- Exhaustive survey on immigration and emigration flow, Ministry for Internal Affairs and Institute of Statistics
	Bulgaria	-Information System Demography (marriage, divorce, birth, death, change of present address, change of citizenship), National Statistical Institute	
EU Candidate	Serbia	<i>No full-coverage statistical data sources are used for migration statistics</i>	

3.2.1. Statistical register

In Austria the statistical office compiles the so-called POPREG⁸ database, which should be considered a statistical mirror register of the population register, discussed in the second chapter of this

⁸ POPREG = POPulation REGISTER. Since 2011, the census has also been register-based, relying on the POPREG as a central component.

summarising report, and thus the major statistical data source on population stock, population change and demographic events (births, deaths, internal migration, immigration, emigration, and other events). POPREG was prepared in parallel to the last traditional census in 2001, and became fully operational in 2002. Due to direct linking to authoritative administrative registers, POPREG provides exhaustive population statistics for all territorial levels at any point in time. The Austrian statistical office renders the administrative population register data statistically usable by applying statistical concepts. The following variables on personal characteristics are included in POPREG: sex, date of birth, date of death (if applicable), marital status, current citizenship and country of birth.

In Austria the Educational Attainment Register also provides information for the Register-based Population Census. This data system has been in place since the 2003/2004 academic year and it is maintained by the Austrian statistical office and contains personal data on successfully accomplished educational attainment. Data on educational attainment includes all institutionalised forms of education at public and private schools, universities or similar educational establishments. Units of this statistical register are schools, classes, pupils and staff, including teachers and non-pedagogic staff. Core data on pupils: sex, age, citizenship, mother tongue, special educational needs, start of compulsory school attendance, status of educational attainment at reference date, current education, school success in precedent reference period, attendance of foreign language lessons and final exam. Major annual educational data are published on the basis of this register. The following variables are available in the register of higher education: age, sex, citizenship, social security number or equivalent surrogate identification code, place of residence, appellation and type of studies, date of enrolment and of graduation, occupation, occupational status at enrolment, study-related stays abroad (country, purpose, funding and duration). The variable “country of birth” is considered neither in public nor in the higher education system.

3.2.2. Statistical data collection connected to administrative procedures

Demographic databases

The most common type of full-coverage statistical data collection is the system of vital statistical surveys; this was mentioned by the Bulgarian, Hungarian, Romanian and Slovakian country reports. (The instrument for gathering demographic data can be an administrative database as in Italy.)

The Bulgarian Information System Demography comprises data on population stocks, all vital events, country of birth and changes of citizenship. Although it is held by the statistical office, information is based on official documents. Development efforts are mainly directed to searching new data sources and establishing cooperation with other national institutions that collect personal data. The statistical office does not produce data on emigrant and immigrant flows. Data on emigrant and immigrant stocks have only been produced since 2008. The statistical office has since 2010 used the information held by the National Revenue Agency on persons who have declared refusal to pay health insurance contributions due to leaving the country for more than 183 days during one calendar year in order to improve the coverage of the information system, especially concerning emigration data. Individual data is received based on a person’s PIN.

Hungarian vital statistics are based on a data collection commonly used for administrative and statistical purposes. When a vital event occurs, the data provider has to declare the necessary information to the authorities only once. Reported data relate to administrative and statistical needs, and every authority has access only to the data it is allowed to by law. According to the present legal rules, all vital events (birth, death, marriage, divorce) that occur in the territory of Hungary are

registered while those that occur to Hungarian citizens outside the country are not⁹. Demographic data hold great potential from the point of view of migration, but they are hardly ever utilised because they cannot provide valuable information on the migrant population due to its small proportion. Collected data are as follows: sex, date of birth, marital status, permanent residence, temporary residence and citizenship, number of live births and living children, educational attainment, economic activity, occupation and occupational status. Place of birth has been included among the collected data since 31 May 2013. Data are accurate and are generally complete and reliable with some exceptions (education and occupation).

In Romania, the registration of demographic events is closely linked to administrative procedures, yet the National Institute of Statistics uses its own paper-based questionnaires in a statistical review of deaths, births, immigrants and emigrants using the same concepts as the administrative registers. Some survey data concerning both emigration and immigration are not reliable. In terms of immigration, a considerable ratio of immigrants is ethnic Romanians with Romanian citizenship who decided to move to Romania and they did so without notification. In terms of emigration, the majority of out-migrants have little interest in declaring emigration to local authorities. According to statistical estimation, emigration statistics capture less than ten per cent of the legal outflows from Romania. Consequently, the majority of effective in-migrants do not appear in statistical office statistics as immigrants while emigrants are not registered. Difference, according to population stock data is sizeable and presents a real problem in a strong emigration country.

The Slovakian statistical office has a similar full-coverage data collection system of vital statistics. The country report of the data production system and the break-down of the published data on vital events and the population suggest that citizenship and country of birth are included in the demographic questionnaires, but published data are not available broken by these variables. However, international migration data are published by country of origin and country of destination, respectively, which is rather unusual.

Other full-coverage statistical data collections

Bulgaria and Romania use administrative data sources, as it was shown above, to have education related indicators of the migrant population. Austria compiles a statistical register as it was presented earlier, while Italy and Slovenia operate statistical databases which can be used for the same purpose, and Hungary has both.

Statistical Survey of Student Enrolment in Tertiary Education is a mandatory, annual statistical survey in Slovenia, collecting individual data on enrolment of students at every year of study and at application for graduation in higher vocational and higher undergraduate and postgraduate studies. The data are transferred to the Statistical Office by the education institutions that collect the data. The data on foreign citizens who study in the country can be derived from the survey. All students who do not have Slovenian citizenship (after matching with the Central Population Register data) are considered foreign students. Students, for whom citizenship is not found, even after matching with the Central Population Register, are shown under the category “Not classified”. Socio-demographic, labour market and human capital characteristics are collected on students. Data are also available by country of residence and citizenship. The data are only used for statistical purposes by the Statistical Office of the Republic of Slovenia.

There is a compulsory educational statistical database in Hungary similar to the Slovenian one described above. Besides the gradually improving administrative register of education, there is also a

⁹ From 31 May 2013 the target population of demographic data collection is broader: among the population movement events it is necessary to observe deaths occurred abroad of those who have permanent address in Hungary, and births delivered abroad by mothers who have permanent address in Hungary, if these events have been registered by the authority in Hungary.

compulsory full-coverage statistical database related to public and higher education. The ministry engaged in educational issues is responsible for data collection and publication. The statistical system was developed in cooperation with the Hungarian Central Statistical Office (HCSO) and data collections are in line with EU methodological requirements of education statistics, covering information of maintainers, institutions, employment, children, pupils, students, teachers and masters (teachers in colleges and universities). Migration-related data are citizenship and the country of residence. Citizenship is not collected for employees in public education.

Pupils by citizenship (Hungarian and non-Hungarian) by completed secondary level attainment, by classes and individual citizenship, by the country of residence, those graduated (from secondary school) by sex and citizenship, successful professional examination by profession and citizenship (Hungarian and non-Hungarian) are covered by the public education data collection. Numbers of foreign students at several points of higher education and those in student hostels are gathered in the higher education data collection. The response rate is 100 per cent, as the ministry closes the data set only and when it is complete. From the next year on (2014) statistical data on education will be produced from administrative data sources.

In Hungary there are two other statistical data collections of full coverage, which are the responsibility of the statistical office. One covers *people who have acquired Hungarian citizenship*, and the other relates to *foreign citizens with settlement documents*. The purpose of these surveys is to know the sociological characteristics of those who want to live in the country. The data collection on new citizens covers everybody who granted Hungarian citizenship whether or not having residence in the country. Data with relevance from a migration point of view in these data collections are as follows: country of birth, (previous) citizenship and previous foreign residence. The statistical data collection is bound with the administrative procedure and data provision is voluntary in both cases. The response rate is quite high (more than 80 per cent) for the data collection on citizenship. However, the response rate of the other survey is much lower.

Although education database is also available in Italy, its usability in migration statistics is unremarkable.

3.3. Survey-type data sources

Sample-based surveys could be important sources for migration-related research because they provide richer information than administrative data sources. In many cases – as we have outlined before – even basic data about migrants (e.g. employment, level of education) are incomplete or even missing in registers. Surveys, however, cover a wide range of topics and therefore – depending on the objective of the survey – enable detailed analysis of various characteristics, and integration indicators of the migrant population. For instance many indicators of migrant integration proposed by the Zaragoza Declaration¹⁰ could also be calculated only based on survey data.

The SEEMIG country reports include internationally standardized and significant national surveys which contain data about migrants or migration-related events. Among these, there are both *general surveys* representative of the total population, including migration modules or within which migrants can be identified by particular characteristics (usually citizenship or country of birth), and *specialized surveys* representative of the migrant population (or some special sub-group of migrants). In the case of surveys representative of the total population, the main restrictions of the analysis stem from *under-representation of migrants* and their *low number* in the sample. In addition, even when

¹⁰ The Zaragoza Declaration, adopted in April 2010 by EU Ministers responsible for immigrant integration issues, and approved at the Justice and Home Affairs Council on 3-4 June 2010, has identified a limited number of policy areas of relevance for integration: employment, education, social inclusion and active citizenship (see more detailed: http://ec.europa.eu/ewsi/UDRW/images/items/doc/13055_519941744.pdf).

migrants can be identified in the sample, important migration-related questions are often missing (e.g. date of migration, reason for migration, legal status of the migrant, etc.), which could be important background variables for analysis. Finally, surveys targeted directly at migrants often lack a proper sampling frame (it is difficult to produce a representative sample). Reaching migrants in the course of the survey and their low response rate (due to language difficulties and lower willingness to provide information) is also a problem.

3.3.1. Internationally standardized surveys

A great advantage of internationally standardized surveys is that due to harmonised questions and methodology results are comparable across countries. However, if the sample size is not sufficiently high, the aforementioned problem of low numbers of migrants included in the sample is nonetheless present.

The main harmonized European surveys which could provide important data on immigrant integration in different areas are: the *European Union Labour Force Survey* (EU-LFS), the *European Union Statistics on Income and Living Conditions* (EU-SILC) and the OECD's *Programme on International Student Assessment* (PISA). As stated also in the Zaragoza Declaration (2010), these could be used as data sources for the proposed common indicators. They allow comparison between migrant (foreign national as well as foreign-born) and non-migrant population on the one hand, and across different countries on the other.

The Labour Force Survey is the international survey that was considered more or less important by all SEEMIG country experts from the point of view of migration data (a description was included in all the country reports). Therefore, we provide a more detailed overview on LFS in the countries examined, focusing mostly on the availability of migration-related data.

The EU-SILC and the PISA surveys, although with migrants identified in their samples, were assessed by SEEMIG partners in general as less important or not relevant regarding migration data. This is mainly due to the considerably smaller sample sizes of these surveys, which means that migrants are more likely to be under-covered. However, these surveys – besides their limitations – serve as important data sources for comparative analyses about the social and economic situation of migrants, as well as about the performance of students with immigrant background (Eurostat 2011, OECD 2010, 2012).

3.3.1.1. European Union Labour Force Survey (EU-LFS)¹¹

The most comprehensive type of general sample survey covering the resident population aged 15 and over in private households is the Labour Force Survey (LFS), which has a wide range of questions regarding labour market indicators. In addition to its comprehensive character, the essential advantage of the LFS is comparability, since data collection by countries uses the same concepts, definitions and classifications and covers the same set of characteristics in each country. Nevertheless, there are differences in the immigrant and emigrant groups covered in different countries.

According to Eurostat, LFS is conducted in the 27 Member States (and EFTA countries, except Liechtenstein) and three Candidate Countries in accordance with *Council Regulation (EC) No 577/98* of 9 March 1998. Participation in the LFS is compulsory in three SEEMIG countries: Austria, Italy and Slovakia, and voluntary in the others.

¹¹ This part was prepared with the contribution of Ágnes Hárs and Zsuzsa Blaskó, for which contribution the authors are grateful.

The LFS provides quarterly results on labour market participation of people aged 15-74, as well as on persons outside the labour market. Persons carrying out obligatory military or community service are not included in the target group of the survey, as is also the case for persons in institutions or collective households. The national statistical institutes are responsible for selecting the sample, preparing the questionnaires, conducting direct interviews among households, and forwarding the results to Eurostat in accordance with a common coding scheme. The data collection mostly covers the years from 1993 onwards (in Austria and Serbia from 1995). In general, data for each country are available depending on its EU accession date¹². Although LFS has a comparative character, being part of the EU international statistical programme, data collection is a national responsibility and data gathering and data coverage, in spite of EU regulations, differs in some degree across countries¹³.

Applying LFS as a data source for migration research first requires specification of the possibilities and limits of it both for immigration and emigration data production. The whole population of LFS is interviewed for *citizenship* and *country of birth*. This allows some data production on **immigration** (diplomatic personnel, foreign representatives and members of their families are excluded). Nevertheless applicability depends on sample size, frequency of data collections, target group coverage and some national differences in data collection by country.

In addition to standard statistical data production based on citizenship and/or place of birth, discrete migration-specific data collection has focused on immigrants as well. In **2008 an ad hoc module** of the EU-LFS was conducted on the '*Labour market situation of migrants and their immediate descendants*'¹⁴. These data were collected in the EU27 Member States, Switzerland and Norway, and differed in terms of countries with high versus low shares of immigration: a detailed questionnaire module ('full module') was applied in the former group, while a simple module ('light module') was applied in the latter¹⁵. The full module was applied in only two SEEMIG countries: Austria and Italy (collecting also information on the main reason for migration, legal barriers on the labour market as well as qualification and language issues), while the light module - with questions about year of citizenship acquisition, country of birth of mother, country of birth of father, and length of residence in the host country - in countries with low numbers of immigrants: Romania, Bulgaria, Hungary, Slovenia and Slovakia. In Serbia this ad hoc module was not conducted at all.

Beyond immigration-related data, some **emigration data** are also collected in LFS. Since temporarily absent household members are included in the survey, some limited aspects of emigration can also be measured. Such an analysis can be based on the standard question of the LFS, asking the employer's country of residence. Data refer to those employees who are currently employed abroad (or to those persons whose last employer was abroad prior to returning to the present resident country and that it was the person's last job). These data refer only to a special subgroup of labour emigrants. A targeted ad hoc module of the LFS focusing on emigration has not been conducted yet.

The overview above has outlined the possibility of the LFS regarding both directions of migration. Applicability of migration data differs considerably by country depending on the size of emigration versus immigration and, correspondingly, the coverage of the target population in the LFS. Most of the countries have LFS data for a long period, even covering migration data. However, due to national conditions, data collections and coverage of target population (both emigrants and immigrants) differ by country.

¹² Based on <http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/lfs>.

¹³ General comparison across countries regarding sample size, reliability, etc., is not provided in this short overview. More comprehensive discussion on the comparative character of LFS is available on the EUROSTAT homepage, or in: *Labour Force Survey in the EU, candidate and EFTA countries. Main characteristics of national surveys, 2009*; Collection: Methodologies and working papers; Eurostat, 2011.

¹⁴ Commission Regulation (EC) No 102/2007 of 2 February 2007 adopting the specifications of the 2008 ad hoc module.

¹⁵ More detailed information on the ad hoc module available: http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/EU_labour_force_survey_-_ad_hoc_modules.

Most of the SEEMIG country reports referred to the LFS as the most comprehensive survey with migration relevance, but also as the one with serious deficiencies. In all cases LFS *under-represents the number of migrants* and sampling allows only moderately detailed analysis. Most countries consider the LFS as a source of immigration data, while at the same time some countries with sizeable emigration, particularly Hungary and Romania, consider the possibility of measuring labour emigration using the particular LFS question. This question, aimed at those working abroad and identified by employers' country of residence, is a relevant though incomplete indicator of labour emigration.

A pilot study in the frame of the SEEMIG project aims to collect emigrant data and the contact details of emigrants linked to the households in the home country, so as to develop more comprehensive emigration statistics based on the LFS¹⁶. Due to its standard methodology and relatively large sample size, the LFS is an obvious choice for such a methodological experiment (for details on the lessons learnt from pilot study see: Blaskó – Jamalia 2014).

Table 3 provides a comparative overview of SEEMIG countries' characteristics of the LFS regarding migration-related data, mainly based on descriptions of country reports. **Immigrant population** is under-represented in each country in the project. As the most characteristic interpretation of the Italian report states, *immigration is less substantial in the Labour Force Survey than appears in the civil registry*. There are various reasons for this: the hidden nature of certain groups of the immigrant population, language barriers, the particular spatial distribution of immigrants, which is often not in correspondence with the sampling methodology, and potentially high non-response rate due to irregular work or irregular status. According to an EC Report, non-response rates are higher among recent migrants and non-EU migrants (EC 2008). In addition, only private households are included in the LFS, and that may also add to the under-representative character of immigrant data since many migrants live in collective households (e. g.: in institutions for migrant workers or asylum seekers). Even Austria, with a high share of immigrants and serious efforts to cover non-nationals in the LFS, reports that *non-nationals are generally under-recorded due to language barriers despite multi-lingual interviewers and questionnaires*. For these reasons results are only valuable at the aggregated territorial level¹⁷. Language barriers are of a different nature in Slovenia, where most immigrants originate from the successor states of ex-Yugoslavia. Here *the questionnaire is in Slovenian but, if necessary, most of the interviewers are able to conduct interviews in Serbian or Croatian*. Still, due to the lack of a foreign language questionnaire, a small but important part of immigrants who speak no or poor Slovenian, Serbian or Croatian are excluded from the LFS survey in Slovenia.

To overcome the obvious shortcomings of capturing immigrants through the LFS there is a currently on-going project in Hungary at the HCSO financed by European Integration Fund (Európai Integrációs Alap) (EIA/2012/2.6.2). This project aims to increase the availability of foreigners in the LFS survey by mitigating language-barriers during the fieldwork. Activities include training interviewers to contact people who do not speak Hungarian and designing self-administered questionnaires in foreign languages. At the same time another pilot project is being prepared to boost the immigrant subsample in order to achieve a sufficient number of immigrants in the LFS. In Italy in order to improve the quality of LFS data for immigrants by decreasing language barriers ISTAT conducted a cognitive test related to LFS 2008 ad hoc module.¹⁸

¹⁶ Moving somewhat further from the standard LFS approach, but still building on LFS's main features, the SEEMIG pilot study on emigration – implemented in Hungary and Serbia – used the LFS to build up an indirect sample of emigrants from the country of origin and tried to reach them directly in a second phase of the research.

¹⁷ As mentioned in the Austrian country report, *Register-Based Labour Market Statistics* may become a valuable data source in the future, presumably due to the possibility of better coverage of immigrants.

¹⁸ The cognitive interview identifies those aspects of the question that lead to errors in the answers. The aim is to stimulate the respondents to realise the difficulties encountered in answering questions. As a result of the cognitive test related to the Italian LFS, the average comprehension level of the wording of the questions rose from 71% to 87% (Albisinni, Marzilli and Pintaldi 2008).

Despite all the above-mentioned deficiencies, the LFS is an important source of data in countries where the size of immigration is considerable, though it is hardly applicable for any comparison in countries where size of immigration is moderate (as seen in the Romanian and Bulgarian reports).

As for **emigration**, the LFS is applicable to identify some – mostly short-term – labour emigration but has no capacity to cover the total emigrant population. The particular question aimed at those working abroad identified by employers' country of residence is a relevant though incomplete indicator of short-term labour emigration. The best discussion of this possibility is in the Romanian country report, a country with considerable emigration experience. Hungary also applies the LFS for exploring labour out-migration - the relevant questions have been included in the LFS since 1999¹⁹. It is recognised, however, that because the indicator applied defines a special subsample and the sample is small, a very cautious interpretation of this data is needed. It should be noted that the way household membership is defined in the LFS also has implication on the definition (and hereby on the size) of labour out-migration measured by the LFS. Since the definition of household membership is not fully harmonised across countries, and refers to different time periods, data on labour out-migration is not internationally comparable²⁰. Moreover, poor specification of the notion of 'household member' can also lead to misunderstandings and therefore false and biased results on domains based on this concept.

Besides the 2008 ad hoc module (included in almost all countries conforming to EC regulation), individually initiated, national supplements on migration are rarely included in the LFS. An exception is Hungary, where questions on *intentions of being employed abroad* were included in LFS surveys from 2002 to 2009, and questions on *ever worked abroad* and circumstances of employment were included nearly every year between 2002 and 2009.

Furthermore, as it was also mentioned in the Romanian report, for countries characterised by considerable emigration and therefore by large emigrant stock (like Romania, Bulgaria and Serbia) the LFS surveys of the main destination countries could also provide substantial data on labour-force out-migration.

All in all, the essential advantage of the LFS is the comparability across countries and between foreign and national population, as well as the large number of variables, which makes detailed analysis possible. Data on migration are mostly under-represented and not really applicable for assessing the actual size of migration. Still, the analytical character of the survey is crucial. Basic data are insufficient for analysing labour emigration in general terms, but additional elements of data collection would allow more precise analysis of labour out-migration. The next ad hoc module on migration, which will be included in the LFS in 2014, will be a useful tool to collect data about migrants in a more systematic way, at least about the most relevant labour market indicators. Two sub-modules will be included: one about the background of migrants and their descendants and another one on the obstacles to labour market participation. The former (besides the country of birth of parents and level of education of parents) contains question on the last country of work abroad and the reason for migrating into the host country (European Statistical System 2013).

¹⁹ In case the household member is working abroad at the time of the survey, *country of employment abroad* is recorded. In case of a special group - those who are not currently working but who had their last employment abroad - the country of the last employment is also registered.

²⁰ In Hungary, for example, people living abroad for no more than one year and maintaining an economic relationship with the household should be included among members of the household. In Romania persons who live temporarily (for less than six months) elsewhere (in Romania or abroad) and persons who moved elsewhere for a longer period than six months, but maintain close relationships with the household (children studying elsewhere, students, household members working elsewhere, convicts and prisoners, military personnel, etc.) are considered members of the household.

Table 3: Characteristics of LFS survey in SEEMIG countries

Country		LFS data available	Migrant data available		Comments
			Immigrant data	Emigrant data	
EU15	Austria	Since 1995 yearly based, since 2004 quarterly based – connected to the Microcensus (MZ) installed in 1968.	Foreign citizenship (since 2003), and foreign country of birth (since 1995). All persons whose regular residence is in Austria. 2008 ad hoc module: detailed focus on labour market situation of migrants and their descendants; country of birth of each parent maintained since then.	Not mentioned.	Results are only valuable on aggregated territorial level due to restricted representativeness. Only survey-type source mentioned. Since 2004 critical groups (i.e. non-nationals, foreign adolescents) are better depicted.
	Italy	Since 1959 (several reshaping), since 2004 (new sample) quarterly based.	Foreign citizenship (since 2005), and foreign country of birth (since 2005). Does not include illegal immigrants or immigrants who are legal but do not reside in Italy (covers population that is duly listed in the municipality civil registry). 2008 ad hoc module: detailed focus on labour market situation of migrants and their descendants; country of birth of each parent maintained since then.	Those temporarily abroad are included.	Not really appropriate for application for migration purposes (less substantial than the civil registry). The innovations in 2004 broke the continuity of economic indicators. The main data available online allow a distinct overview of nationals and foreigners.
EU8	Slovenia	Since 1993 yearly based, since 1997 quarterly based.	Foreign citizenship (since 2002), and foreign country of birth (since 2002); all persons with usual place of residence in Slovenia. + the year of immigration & the country of origin. 2008 ad hoc module: short version; included foreigners with permanent and temporary residence.	Not mentioned.	Questionnaire only in Slovenian, but interviews in Serbian and Croatian language too (most immigrants in Slovenia originate from other successor states of ex-Yugoslavia). Not really applicable for migration purposes due to under-representation.
	Slovakia	Since 1993 quarterly based.	Foreign citizenship (since 2003), and foreign country of birth (since 2003) + year of residence in SR, country one year ago. Foreigners who intend to stay for more than one year are included. Foreigners with temporary residence are not subject to the LFS. 2008 ad hoc module: short version (focus on migrants and their descendants).	Possible to obtain labour emigration. Outputs for selected years were published only in some research works.	Low reliability, applicable to the total population only. Data about migrant categories are not published.

	Hungary	Since 1992 quarterly based.	Foreign citizenship (since 1997), and foreign country of birth (since 1997); country of residence one year before (since 1999). + year of immigration (since 2001), country of residence one year ago (since 1999). 2008 ad hoc module: short version (focus on migrants and their descendants); questions have remained since then.	Household members living abroad less than one year are included if they have common income/consumption with the surveyed household. Short-term labour force out-migration: in case the household member is working abroad or last employment was abroad country of employment is recorded (since 1999). Intention of being employed abroad (from 2002 to 2009).	Low number and under-representation of immigrants in the sample; special sub-sample and low sample size in case of labour force emigration. There is only access to anonymised data. Data published (in HCSO publications) do not contain data by citizenship or country of birth. Migration-related data are published in some research work.
EU2	Romania	Pilot survey in 1993, yearly based in 1994 and 1995, since 1996 quarterly based.	Foreign citizenship (since 2004, with low reliability), country of origin. No data on country of birth. Foreign citizens with permanent residence are included. 2008 ad hoc module: short version, information on migrants and their descendants and reason for immigration.	Nationals who live less than six months elsewhere (in country or abroad), or longer than six months but maintaining close contact with the household are included. Short-term labour force out-migration (since the country of workplace is registered); data on this have not been published yet. Does not contain any specific questions regarding emigration.	Low share of immigrants and low reliability of immigration data, but considered an important source of information on short-term labour emigration. Data are not separately published for Romanian citizens and foreign nationals, or by country of origin.
	Bulgaria	First in 1993, between 1994–1999 three or four times per year, 2000–2002 four times per year, since 2003 quarterly based.	Foreign citizenship (since 2001, with low reliability), foreign country of birth (since 2006). Foreign citizens living in the country more than one year or intending to stay more than one year. 2008 ad hoc module: short version (reduced information).	No emigration has been measured using LFS. Persons who left the country one year or more and persons permanently living abroad are not considered household members.	Small sample, small proportion of immigrants in the country, therefore immigrants not well represented and the LFS does not provide reliable data. Data are not published.
EU Candidate	Serbia	Since 1995 (yearly), since 2008 (two times a year); in 2013 three rounds are planned.	Foreign citizenship and foreign country of birth. Foreigners who reside or intend to reside in the territory of Serbia for one year or longer. 2008 ad hoc module was not conducted. Data on citizenship are not published, but may be made available upon request.	Definition: persons who have been absent for less than one year and do not intend to be absent for more than a year are included in regular data collection (in households).	For migrants who have been in Serbia for less than one year and intend to stay no more than one year, only the basic demographic data are collected – data have not been processed. Only survey-type source mentioned. Low sample size, less frequent (two times a year).

3.3.1.2. *European Union Statistics on Income and Living Conditions (EU-SILC)*

EU-SILC is a source for comparative statistics on income distribution, poverty, social exclusion and other living conditions at the European level. It was launched in 2003 in seven countries, and in 2012 it was implemented in 27 European Union countries, Croatia, Iceland, Norway, Switzerland and Turkey – so in all SEEMIG countries except Serbia (which joined this survey only in 2013). It provides two types of annual data: cross-sectional data pertaining to fixed time periods, and longitudinal data pertaining to individual-level changes over time, observed periodically (usually over a four year period). The common framework (i.e.: common procedures, concepts and classifications), as well as the harmonised list of target variables aim at ensuring comparability.²¹

Although EU-SILC does not specifically target migrant population, immigrants could be identified - similarly to the EU-LFS – on the basis of *citizenship* and *country of birth* (but information on how long they have been in the country is not gathered). Both indicators are collected only for persons aged 16 and over. Since EU-SILC only covers people living in private households, migrants living in collective accommodation for migrant workers or asylum seekers are not included.

As it was already mentioned in the case of LFS, the coverage of migrants is problematic in this case as well, and more particularly the recently arrived and poorly integrated migrants are likely to be under-covered. The non-response rate of some migrant groups is also higher than average due to their language difficulties and lower willingness to provide information. The illegal (or irregular) and temporary migrants in particular are likely to be underrepresented. Moreover, the measurement of migrants is more limited on the basis of EU-SILC, due to its considerably smaller sample size than that of the EU-LFS.²²

As a result of the above listed deficiencies of the EU-SILC regarding migration-related data, the majority of SEEMIG partners considered this data source unimportant from the point of view of migration research, except Austria and Italy, the two main immigration countries where this survey was launched in 2003 and 2004, respectively. In other SEEMIG countries the most common reason of considering this data sources useless were the small sample size and the under-representation of immigrants (compared to their actual share in the population), but the low share of immigrants in the country (Bulgaria, Romania and Slovakia), lack of data on citizenship (Slovenia), and restricted classification of country of birth (Slovenia) were also mentioned.

In the Italian report one national survey which is connected with EU-SILC was also described in the country report (see later).

3.3.1.3. *OECD's Programme on International Student Assessment (PISA)*

PISA is an international survey launched by the OECD in 1997 on the reading, mathematics and scientific competencies of 15-year-olds students who attended at least the seventh grade of elementary school. It is conducted every three years with the first data collection in 2000, then in 2003, in 2006, in 2009 and in 2012. It aims to assess the extent to which students near the end of compulsory education can apply their knowledge to real-life situations.²³ Data are comparable over time and among countries.

²¹ Based on: http://epp.eurostat.ec.europa.eu/portal/page/portal/income_social_inclusion_living_conditions/introduction.

²² The sample size of EU-SILC in SEEMIG countries varied between 4,344 and 20,928 households in 2008, which was in most of the countries (except Slovenia and Slovakia) just one third of the EU-LFS sample size in the same year.

²³ For more information see: <http://www.oecd.org/pisa/aboutpisa/>.

PISA offers a possibility to analyse the performance of 15-years-old students with immigrant background as well. This is important because the education is a key factor for the integration process of immigrants. Students with an immigrant background include students who are first- or second-generation immigrants: *first-generation immigrant students* are foreign-born students whose parents are also foreign-born, *second-generation immigrant students* are those who were born in the country of assessment, but whose both parents are foreign-born (OECD 2010).²⁴

Moreover, PISA distinguishes between *international students* – students who have moved from their country of origin with the purpose of studying, and *foreign students* – those who are not citizens of country where they are enrolled. These latter may be long-term residents, or indeed, have been born in the country. Therefore, international students are a sub-group of foreign students. (OECD 2012)

Although all SEEMIG countries participated in the last PISA survey, only Slovenian, Austrian and Italian SEEMIG partners regarded it as really relevant data source from the point of view of migration (and it was included only in the Slovenian country report). Students in the survey are selected from a random sample of schools according to their age. Between 4,500 and 10,000 students are tested in each cycle in each country. Despite the fact that foreign and/or foreign-born students could be identified in the sample (as well as those with foreign-born parents), in most of the SEEMIG countries – particularly in those with low share of immigrant population – they constitute a very small, and at the same time a very heterogeneous group (as regards their country of origin and language) within the sample. Consequently, in general there are serious coverage problems of students with immigrant background. Their small number in the sample and their under-representation implies that the results about their performance need a cautious interpretation.

Nevertheless, in Slovenia (where the country of birth is used as a criterion for defining an immigrant student) the sample is considered large enough for aggregate comparisons of the performance of immigrant students with the performance of national students, though too small for more detailed analysis.²⁵

3.3.2. National surveys

Among the national surveys described in the country reports, there are ones which represent the whole population, and migrants (or a certain group of them) may be identified therein, and many others are ‘migrant-targeted’ which aim specially to reach immigrants or emigrants. An overview of these surveys will be provided below, pointing out best practices that could serve as examples. One particular example of a sample-based survey aimed at the whole population is the *micro census*, which provides information in the inter-census periods and due to its relatively large sample size may include higher number of immigrants than other national surveys, so it could be considered as data source for migration research. However, a micro census has been carried out in only two SEEMIG countries: Austria and Hungary.

3.3.2.1. Micro censuses

The Austrian Microcensus (MZ) has a quarterly character and embraces the LFS and the microcensus on dwellings. In the Austrian country report the MZ is regarded as the most comprehensive source of information on the labour market and human capital attributes of foreign nationals and foreign-born persons, and the only source of information about self-employment of these groups. *MZ data makes*

²⁴ Students who have at least one parent born in the country of assessment are considered native students.

²⁵ In the PISA 2009 data collection, there were 333 schools/educational programmes and 7,810 students included in the sample in Slovenia.

it possible to combine several variables referring to educational attainment and labour market-related characteristics and consider place of birth and duration of stay.

The Hungarian country report referred to the 2005 micro census as possible survey data of immigration. The importance of the micro census for migration-related issues lies in its relatively large sample size (two per cent of the population), which could cover sufficient numbers of immigrants (considering the relatively low number of foreigners in Hungary it is nearly impossible to get relevant information on immigrants through general population surveys with a smaller sample size). However, the micro census includes only a group of immigrants who reside legally in Hungary (their proportion was 1.4 per cent of the total population in 2005), and they were also slightly under-represented in the sample (1.27 per cent). The Hungarian micro census does not include information about emigrants.

Although the micro censuses are important data sources in the field of migration research, principally regarding immigrant stocks, they include only officially registered foreign citizens and collect insufficient data on illegal immigrants or those who have arrived recently (Fassmann 2009).

3.3.2.2. Other national representative surveys

In Italy an annual survey has been conducted since 1993 on “*Aspects of daily life*”. This sample survey (based on a sample of households selected from the municipal registers) is part of an integrated system of social surveys (*The Multipurpose Surveys on Households*) and collects fundamental information on individual and household daily life. It is a national representative survey, but it also includes data on citizenship (distinctions made between Italian, foreign and stateless persons) and in some cases on the country of birth of members of the selected families. However, as mentioned in the country report, this information has not been systematically used and validated since the selection criteria of the analysis units (the actual families) do not always guarantee any statistical representativeness of this specific population segment.

3.3.2.3. Migration-targeted surveys

Finally, we present a selection of targeted migration surveys which are important for understanding the characteristics of migration in SEEMIG countries and for describing particular problems.²⁶ However, according to SEEMIG partners, in some of the countries migration-specific surveys have not been implemented, while in others these surveys are diverse in character: the research focus reflects the main migration tendencies of the respective country.

Most important are the immigration surveys in Italy and emigration surveys regarding Romanians’ emigration experiences. However, immigration surveys in Italy are based on a relatively large number of immigrants and correspondingly, emigration surveys on Romanian emigration supposed there to be a considerably large number of emigrants. Bulgarian surveys also address emigration; nevertheless data sources are less precise about the number of emigrants and they mainly focus on their characteristics (see details later). As for Hungary, the peculiarities of immigration from neighbouring countries and a policy focus on newly naturalized citizens is reflected in migration-specific surveys. The most important migration-targeted surveys in SEEMIG countries are presented in *Table 4*.

²⁶ Other surveys may certainly exist, including those of a scientific nature, which cannot be examined in the context of the SEEMIG project.

Table 4: Examples of migration specific surveys in SEEMIG countries

Country		Migration-targeted surveys	
		Immigration focus	Emigration focus
EU15	Austria ²⁷	- Gfk Survey (2010, 2011 and 2012) - Survey on the Social and Living Conditions of Students in Higher Education (1998, 2002, 2006 and 2011)	
	Italy	- Income and living conditions (ISTAT), since 2009 - Social conditions and integration of foreign citizens (ISTAT), since 2011 - Special survey on irregular migration (ISMU)	
EU8	Slovenia ²⁸		- Emigration of researchers in the years 1995–2004 (2004) - Emigration of researchers in the years 2004–2009 (2009–2010) - Potential brain drain of the Slovenian scientists (2005) - Emigration of former Erasmus students, generations 2005–2010 (2011–2012) - Potential emigration of the Slovenian students and young graduates (2004)
	Slovakia	<i>Migrant-targeted surveys have not been implemented according to SEEMIG partners.</i>	
	Hungary	- Immigrants 2002, 2006 (HDRI) - Immigrants in Hungary, 2009 (IMS) - Immigrant Citizens Survey, 2011–2012 (IMS)	- Hungarians on the Austrian labour market, 2008–2009 (OFA)
EU2	Romania		- Community census – local authority survey on emigration, 2001 (IOM) - Living abroad temporary, 2006 (Soros) - Romanian migrants in Italy, 2011 (ISMU), Italian project
	Bulgaria		- Territorial mobility of population (2001), NSI - Migration and migration behaviour of the population (2011), NSI - Sample survey of Bulgarian and foreign citizens departing from Bulgaria (2012), NSI
EU Candidate	Serbia	<i>Migrant-targeted surveys have not been implemented according to SEEMIG partners.</i>	

Some of the country reports listed special (usually) small-scale surveys, which are rather different in scope, focus and quality. It is still important to present them to prove that migration research has various sources in addition to the basic data sources. Moreover, some research addresses specific groups of migrants.

In Italy the *Income and living conditions* survey (funded by the Ministry of Labour and Social Policies) was carried out by ISTAT for the first time in 2009, joining EU-SILC by using the same methodological

²⁷ In Austria the *Gfk Survey* (2010, 2011 and 2012) focuses on the process of integration (perceptions, contacts and discrimination); *Survey on the Social and Living Conditions of Students in Higher Education* (1998, 2002, 2006 and 2011) includes variables like citizenship, place of birth, place of acquisition of highest education apart from information on educational attainment, occupation and occupational status. But these surveys were only mentioned and not discussed in much detail in the Austrian country report, due to the fact that it was dedicated to illustrate the main data-sources.

²⁸ In Slovenia some emigration-targeted surveys have been implemented in the last decade, but due to their narrow focus they were not included in the country report. (In the surveys about emigration of researchers in 2004 and 2009-2010 research organisation were surveyed, while in 2005 only researchers who held a degree.) Furthermore, neither opinion surveys nor non-representative surveys have been included in country report.

tools for the survey (questionnaires, survey techniques, methods of correction, imputation techniques, etc.), but on a sample of 6,000 households resident in Italy *with at least one foreign member*. To facilitate fieldwork, the survey questionnaires were translated into the ten languages most commonly spoken by foreign citizens who are resident in Italy. Thus the picture of households with foreigners – including data about living conditions, housing deprivation, and material deprivation, etc. – provided by this survey may be compared with that of households made up exclusively of Italians (obtained from the EU-SILC survey, which was also carried out in 2009). This survey is a good example how an in-depth analysis can be carried out among immigrants in a way that ensures comparability with the host population, including sufficient number of foreigners, ensures also representativity, and utilising the methodology of an international survey.

Another migrant-targeted survey in Italy, on *Social condition and integration of foreign citizens*, started in 2011 and aimed to provide information on the living conditions of foreign citizens, including naturalised citizens who acquired Italian citizenship after birth. The sample consists of foreign (or naturalised) persons of the selected families in selected municipalities (12,000 families across 800 Italian municipalities). The sample statistically represents the foreign citizens residing in the country, but selection is made from the civil registry, so only registered foreigners are included. However, the gathered data are very rich, and as described in the Italian country report, it covers different aspects of life, so it enables very detailed analysis, though no publications or online data related to this research have been available up to date.

A special sample survey, *Estimates of irregular migration*, is also conducted in Italy on the basis of regular data collection of “aggregation centres” visited by undocumented migrants (for social contacts, health care, religion, leisure or simply for everyday needs). A random sample of “centres” is chosen and a weighting procedure ensures the representativeness of the sample. The reports are routinely produced by the ISMU Foundation for Integration and Multi-ethnicity (Milan) and estimates have been available since 1991.

In Romania a *Community census* was carried out in 2001 at local authority level (sponsored by the IOM). The methodology aimed to gather information about different groups of emigrants by interviewing authorities (e.g. persons who left the country after 1989 and were staying abroad at the time of the survey; persons who lived abroad after 1989, but were present in the locality at the time of the survey). Though the non-response rate was extremely low, the validity of the investigation concerning aspects of information quality and representativity, as well as the fact that the questionnaires were filled by “key-informers” is problematic. However, the investigation is considered more informative than a simple personal survey.

The *Living abroad on a temporary basis* survey conducted in 2006 (sponsored by the Soros Foundation) was a well-prepared and well-designed research project to reveal emigration size and structure. After the national-level representative survey, households with work-related migration experience were contacted on two regional samples based on the snowball method. Qualitative fieldwork was then carried out in the main destination countries (the contacts of migrants were collected during the quantitative surveys carried out in the two micro-regions). Besides these, in the Serbian–Romanian border area the phenomenon of cross-border migration was also studied. This research is a good example of how necessary it is for migration surveys to be complex and to explore various aspects of migration, a phenomenon which is complex and diverse in its own right.

Romanian Migrants in Italy, a survey in 2011 (founded by the Italian source ISMO) addressed the migration intentions of Romanian migrants in Italy. The main aim was to reveal the impact of the free visa regime and on Romania’s EU accession. The research sample was primarily based on the statistics of the Italian National Statistics Office on Romanian migrants, then in a second phase on a randomly selected sample of Romanian migrants who visited the aggregation centres, and third on a snowball sample. This survey is important because it provides empirical data on the temporary or circular character of migration. At the same time, it clearly shows that studying out-migration is

much more effective if it is conducted in the host country with the involvement of local research groups.

In Bulgaria the *Territorial mobility of the population (2001)* and *Migration and migration behaviour of the population (2011)* surveys were included in the Census Programme and conducted parallel to the census. The aim of these surveys was not only to study migration processes in detail, but also to discover the perspectives and reasons for a given migratory attitude. The latter covered the “usual resident population” aged 15 or more. Different groups were distinguished, dependent on their predisposition and readiness for migration: potential migrants, labour emigrants, short-term migrants, tourists and visitors, not travelling abroad. Linking survey data with the census data was possible based on the person’s PIN.

The *Sample survey on Bulgarian and Foreign Citizens Departing from Bulgaria*, conducted by the National Statistical Institute (NSI), is a specialised sample survey carried out on a monthly basis among Bulgarian and foreign citizens passing border checkpoints (three main airports and five land border checkpoints are included). The survey ensures statistical data on the trips of Bulgarian citizens travelling abroad and visits of foreigners to Bulgaria. The survey data is used for production of more precise emigration estimates and especially estimates on Bulgarian citizens leaving the country (according to the purpose of the trip and length of intended stay). Survey results may be used as a weighting factor in the distribution of the emigration data available (on persons who declare interruption of payment of health insurance contribution) by destination countries or some other characteristics of emigrants.

In Hungary there were various national research programmes on immigration in the last decade, but few were of a representative nature. *Immigrants 2002–2006*, an exception, was a two-wave representative panel survey (conducted by the Hungarian Demographic Research Institute), but focused only on immigrants from neighbouring countries (as they are the largest group). It covers very rich thematic topics, thus enabling detailed and – considering its panel design – longitudinal analysis of the integration of immigrants. *Immigrants in Hungary, 2009* (Institute for Minority Studies, HAS) examined the situation and strategies of six immigrant groups in Hungary in 2009 (identifying them by citizenship) using a questionnaire survey and snowball sampling. *Immigrant Citizens Survey, 2011–2012* (Institute for Minority Studies, HAS) was part of the survey co-ordinated by the King Baudouin Foundation (KBF) and the Migration Policy Group (MPG), with the objective of describing the effects of integration policy in the EU. It focused on foreign-born immigrants, who had been resident for more than a year.

Besides these immigration surveys, a special emigration survey, *Hungarians on the Austrian labour market (2008–2009)* was carried out in Hungary. The research focused on Hungarian employees in Austria, their status, migration strategies, and perspectives. Snowball and ‘respondent driven sampling’ were carried out. Despite its low sample size the survey was an important – and effective – attempt to reveal increasing labour force out-migration in Hungary (or at least a segment of it). It also serves as a possible example of the fact that a premise for the success of such surveys relies on close *cooperation between the country of origin and the country of destination* of migrants.

4. CENSUS DATABASES

Taking into account the shortcomings and limited reliability and comparability of administrative data sources on migration, as well as the problems related to the representative nature and small sample size of surveys, population censuses have advantages which render them important for analysing migration, especially regarding immigrant stocks. Compared with other data sources, they contain data about the totality of the population and thus about various groups of migrants²⁹ which are more accurate, reliable and more detailed. Some countries have detailed data about the foreign and foreign-born population living in their territories only at the time of the census; in between census periods only estimates are available.

Since census questions cover a number of different topics they offer detailed data about areas which are important in terms of the integration of migrants (e.g. educational attainment, economic activity and occupation, living arrangements and living conditions). In most countries, a full range of data about such characteristics of immigrants is not (or only partially) available from other sources. Another major advantage of census data is that they enable us to compare the characteristics of immigrants with the characteristics of nationals acquired through the same methods, and also that different migrant populations (which are originally included in different registers) also become comparable.

A drawback of census data, however, is that being updated every ten years they are unsuited to identifying short-term changes in the number and composition or position of the foreign or foreign-born population and only allow us to grasp changes that are on a larger scale. Due to the long time lag between two consecutive censuses some changes in the structure of population could remain hidden. Furthermore, if we compare the census data of different countries, we must also take into account the differences which exist in terms of the population covered and the definitions used, etc.

4.1. The types of the last censuses and the reference date

Of the eight countries participating in the SEEMIG project, Austria and Slovenia carried out register-based censuses in 2011. In both countries this was the first time that the traditional census was replaced by a register-based census³⁰, which is why it may be claimed that these are countries just after transition from the traditional to a register-based census. This was preceded in both countries by a long process of preparation, and the transition may be seen as a major achievement by the respective statistical systems³¹.

The most important advantage of a register-based census is that it entails lower costs and human resources, and does not burden the population by answering long questionnaires. In addition, data processing is faster and data become available more quickly than in the case of a traditional questionnaire-based data collection. In this way the censuses may be carried out more often than the usual ten-year cycle. For example, in Austria and Slovenia it is expected that censuses will take place every five years from 2011 onwards³².

In cases of register-based censuses, data stem from a range of basic registers containing the required information on the population to be enumerated. The heart of the census in Austria is the Central

²⁹ In most EU countries the census includes all international immigrants, regardless of their legal status. The only difference between countries is the minimum required stay to be counted as resident (Fassmann 2009).

³⁰ Countries with entirely or largely register-based censuses in Europe are: Denmark, Finland, Iceland, Norway and Sweden.

³¹ In Slovenia the register-based orientation of statistics began in the 1970s following Nordic countries' experiences.

³² In Austria in 2006 a test census for reference date 31 October 2006 has already been conducted.

Register of Residents (ZMR), and the other base registers are the Housing Register of Buildings and Dwellings (HR), the Business Register of Enterprises and their Local Units (BR), and the Register of Educational Attainment (EAR), all of which are maintained by Statistics Austria, as well as the Central Social Security Register (CSSR), the Unemployment Register (UR) and the Tax Register (TR, though not including data about the income). Several comparison registers assure quality assurance.

In Slovenia the three basic sources used are the Central Population Register (for data on the population), the Household Register (for data on households), and the Real Estate Register (for data on dwellings). Besides these, a range of different sources were utilised, such as Register of Spatial Units, the Business Register, the Statistical Register of Employment, regular statistical surveys on births, migration, student enrolment in tertiary education, tertiary education and on recipients of scholarships (students), Population Census 2002, and databases on unemployed persons, graduates, national examinations, recipients of pensions, insured persons, recipients of social transfers and income tax.

In the cases of register-based censuses, data from different registers are linked by the so-called anonymised pBK code in Austria and by the PIN (personal identification number) in Slovenia.

In Italy the last census was conducted in an innovative way: traditional census (full field enumeration) was combined with data from registers and sample surveys. Sampling techniques for the collection of socio-economic data were applied (though only to the largest municipalities that are capital of province or with at least 20,000 inhabitants). Two different enumeration forms were used: a short form, to collect data on the demographic and housing characteristics of the whole resident population, and a long form, designed for a sample of households, including demographic topics as well as socio-economic variables, such as educational level, occupational status and commuting. The latter enables estimations concerning the socio-economic set of variables included in EU regulations.

In other SEEMIG countries (Bulgaria, Hungary, Romania, Slovakia, Serbia) there were traditional censuses, based on questionnaire enumerations, carried out with enumerators or – in Bulgaria, Hungary and Slovakia – the self-enumeration method (with paper-based or online questionnaires). The method of data provision was selected by the respondent. Online fill-in was an option for the first time in the census history of Hungary and Bulgaria. Moreover, in Bulgaria, for the first time, administrative registers data were used for checking the coverage and accuracy of census results.

The 2011 censuses were performed – for the first time in the majority of countries – according to European standards, namely to *European Union Regulation (EC) No 763/2008* on population and housing censuses. According to this, in order to ensure the comparability of the data provided by member states the data of the censuses should refer – contrary to the previous round – to the same reference year (see *Table 5*). Besides Regulation 763/2008, one of the most important documents is CES Recommendations for the 2010 Censuses of Population and Housing (jointly prepared by UNECE and EUROSTAT), which defines core topics, i.e. topics that are obligatory for including in the contents of basic census forms (United Nations 2006).

4.2. The population covered by censuses

According to *EU Regulation (EC) No 763/2008*, the population covered by censuses means the population at its 'usual residence' at the reference date. The usual residence is defined as 'the place where a person normally spends daily periods of rest, regardless of temporary absences for purposes of recreation, holidays, visits to friends and relatives, business, medical treatment or religious pilgrimage'. This way, the resident population also includes the foreign population as long as the person has been living in the country for a longer period, at least 12 months according to the regulation, or plans to spend at least such a period in the country. It also includes nationals who are

abroad for a period not exceeding 12 months. This way, censuses also cover migrants, regardless of their actual legal status, the only condition being the duration of their actual or planned stay.

There is, however, a difference among the examined countries in the way they handle the duration limit for *foreign citizens living in the country* (see Table 5): some countries consider a shorter period of stay than that of the harmonised definition (e.g. three months are required in Austria, Hungary and Slovakia). In case of Bulgaria, Romania and Serbia, a period of 12 months is required. In Slovenia, where the register data are collected, the only persons included in resident population are those who have a registered permanent or temporary residence. Those who have been abroad for one year or more (and gave notice of their departure) are excluded from the former group, while in the latter those with the total duration of stay at least one year are included. In Italy no time limit was included as residence criterion, they counted all foreign citizens and stateless persons with usual place of residence in Italy who possessed a valid permit to stay (even if physically absent on the date of the Census). Foreign citizens temporarily present were also enumerated, but they were not counted in usually resident population (only in 'present population'). Non-EU foreign citizens without possession of valid permit of stay in Italy were counted as persons temporarily present.

In Austria, the definition of the '*place of usual residence*', applied for defining the resident population in 2001 (in the course of the last traditional census), referred to the main residence, i.e. 'the centre of a person's life irrespective of the *actual duration of stay*', rendering the date of enumeration decisive. The 2001 census therefore contains non-nationals without Austrian citizenship, who were registered with their main residence in Austria, as well as Austrian nationals residing abroad. In 2011 in the course of the register-based census the three-month criterion is generally applied (for nationals and non nationals as well)³³. In Slovakia, the census contains foreigners who were present in the territory of the Slovak Republic at the time of the census, but those living in Slovakia less than 90 days had to fill in only gender, date of birth and citizenship.

The practices of the examined countries regarding *nationals living abroad* are mostly uniform: those temporarily staying abroad, meaning shorter than 12 months (and not expected to exceed that term), are treated as belonging to the resident population (except Austria where the three-month criterion is considered). However, in Serbia persons who had been living abroad for more than a year but re-enter the country at least once a week (i.e. commuters) are also considered members of the resident population. A further special case is Romania, where a special questionnaire was completed for those living abroad for a long period of time (more than 12 months), and also for those living abroad temporarily, but only the latter group was considered part of resident population (so in their case this special questionnaire was a complement to the standard census questionnaire).

In Hungary in the 2011 Census emigrants who were temporarily (less than 12 months) abroad at the reference time were also included for the first time: the whole personal questionnaire was completed for them. As regards persons staying permanently abroad ('the length of staying abroad has reached or is expected to reach 12 months') only their number was recorded on the dwelling questionnaire, no other data were collected about them. (Whole households that had moved abroad could of course not be included in this data collection, due to the lack of data provider.)

There is also a difference among the SEEMIG countries in the way various special groups are regarded in terms of residency. *Asylum seekers* are usually excluded from the resident population (except in Austria), but in Italy, Hungary and Serbia those who otherwise met the criteria of the usual resident population are included; in Slovenia, Romania and Bulgaria those who have granted refugee status are also included. *Foreign diplomats* and *national diplomats* are treated in a more uniform fashion: the former are excluded, while the latter are included (except in Austria where both are

³³ A person who resides abroad for at least three months (after having resided in Austria for at least three months) is counted in emigration statistics, while those residing in Austria for at least three months, are counted in immigration statistics.

included if they have the main residence in the country, and in Italy where foreign diplomats are also included if they do not hold a diplomatic passport, even if live in diplomatic or consular premises). Undocumented (unregistered), irregular migrants are excluded in all countries.

4.3. Migration-related data

The census topics were divided into ‘core’ and ‘non-core’ topics. It was highly recommended that countries collect information with respect to ‘core topics’, while ‘non-core topics’ were optional – recommendations were included on these topics for guidance for those countries that decided to include them in their census (United Nations 2006).

The following migration-related core topics were proposed to be covered in the Population and Housing Censuses by the Regulation (EC) No 763/2008: *country/place of birth, country of citizenship, previous place of usual residence and date of arrival in the current place; or place of usual residence one year prior to the census*. These were obligatory topics for the national level and for the geographical levels NUTS 1, NUTS 2 and NUTS 3. Besides these, questions referring to *ever resided abroad and year of arrival in the country (from 1980)* were also obligatory but only on the national level and for NUTS 1 and NUTS 2. Though statistical institutes in Europe have agreed on the collection of information on the year of arrival of international migrants, as well as the place of residence one year beforehand, the global census recommendations prepared by the United Nations Statistical Division consider a question about residence five years beforehand to be ‘more appropriate for collecting data for the analysis of international migration’ (United Nations 2008).

The Commission on International Migration Data for Development Research and Policy recommends that all countries ask three questions in their national census: *place/country of birth, country of citizenship, and place/country of residence either one year ago or five years prior to the census* for each person enumerated (Center for Global Development 2009). Their suggestion was that in the 2010 round of censuses all these relevant questions should be included, allowing a separate response for each individual country of birth, citizenship or previous residence, and that this information should be publicly disseminated, tabulated by sex, by age and by level of education³⁴. Furthermore, the Commission recommends that all countries collect data on previous residence in the manner most appropriate to needs and circumstances.

The two key questions that serve to identify the foreign population, respectively immigrants (*country of citizenship, country of birth*) were part of the census in all SEEMIG partner countries (see Table 6)³⁵. While the former serves to identify the actual foreign population, the latter also enables identification of those immigrants who have already been naturalised. As regards data relating to migration history (e.g. previous residences, date of arrival), the picture is no longer uniform: *the previous place of usual residence* (the actual country if abroad) is included in all cases, but *date of arrival in the current place* was missing for Austria, Italy and Slovenia. However, ‘*year of arrival in the country*’ was included in all countries. Moreover, in Slovenia they also asked for the data about first and the last arrival. The question regarding ‘*country of residence one year ago*’ only appears in Italy, and Bulgaria (and in Serbia it could be derived indirectly), while ‘*country of residence five year ago*’ was enquired only in Italy.

³⁴ This is only possible if the census form is designed to record all possible countries of citizenship, birth and previous residence. For this reason, countries that allow partial responses (such as check boxes with only five countries, plus a catch-all ‘other’ category) are not considered as complying with the recommendations (Center for Global Development 2009).

³⁵ It is not completely applicable to the former Yugoslav Republics. According to international recommendations, it is only necessary to take into account time of splitting the former country (SFRY). For example, persons born before splitting Yugoslavia are not considered to be foreign born.

Naming the country of residence one or five years ago captures the migration thus outlined, while naming the previous place of residence, if this was abroad, might record movements over an even longer period of time. At the same time, a drawback of the latter question is that data about international migration gets lost for people who moved residences within a destination country after their arrival there. Thus, it would be better to apply the two approaches (inquiring about previous place of residence and place of residence one year before) jointly, but this only happens in Italy and Bulgaria³⁶. For a similar reason, the time of moving to the present place of residence does not necessarily correspond to the time of migration. Thus it is worth (at least in the case of people born abroad) completing a question about the time of arrival in the country.

In order to identify the group of ‘ever-international migrants’ (persons who have ever changed their country of usual residence), a new topic was included in 2011 censuses: individuals were asked whether they had ever had a usual residence abroad, and for those who ever resided abroad, information on the year of last arrival in country³⁷. This identified not only immigrants (foreign-born population), but also those native-born who have ever resided abroad and then returned (after at least one year abroad). This is a core topic (*ever resided abroad* and *year of arrival in the country*) and the information is of more relevance in a former emigration country (like Italy). Data on this were included in the census in all countries except Austria³⁸ and Romania.

A further important, although non-core topic is related to the *ethno-cultural characteristics* of the population. Questions on the *ethnic, linguistic and religious affiliation* of migrants are relevant since, particularly in the SEE region, citizenship and country of birth do not necessarily reveal what minority group the individual might belong to. The ethnic and linguistic plurality of the region makes it reasonable to be acquainted with the ethnic identity of migrants (immigrants and emigrants alike), though this can also be seen in an ambivalent light. Administrative data sources do not record these characteristics and therefore relevant information (apart from possible rare survey data) only comes from censuses. These data are especially important from the point of view of examining the integration of migrants, since in any relevant case these will be able to show that a certain group of immigrants is of the same ethnicity as the receiving population and that they do not constitute an ethnic minority in the destination country. It is also important to collect data besides on mother tongue about command of language, and linguistic and cultural affiliations for the purpose of integration analysis (e.g. the Hungarian census asks about the language the individual most frequently uses in the family or with friends). It is also an interesting question for those leaving the country (and staying abroad temporarily or over the long term) whether minorities are over-represented.

Of the countries examined, Austria’s census did not contain (neither in registers nor in previous census rounds) data on ethnicity, mother tongue or religion. The same holds true for Slovenia, because these data are not available in the registers. Questions on ethnicity, mother tongue or religion are not included in the questionnaire of the 2011 Census in Italy either. In all the remaining countries, however, the census contains all three indicators (*Table 6*). Nonetheless, Hungary followed a different practice when asking these questions: it was possible to give a second ethnicity, while only one option was provided in other countries. Another difference was that in Romania and Slovakia, the question regarding ethnicity was an open question and responses were categorised later on. At the same time, since answering these questions was not mandatory in any of the countries the response rate is questionable. Another problem related to foreign citizens’ ethnic

³⁶ In Bulgaria, however, the question regarding previous place of residence only covers the changes in place of residence between 2001 and 2011.

³⁷ This question does not provide information on interrupted stays.

³⁸ The date of arrival, which is captured in the ZMR (and therefore also in the POPREG), is the date of registration of main residence in Austria. In principle this information could be generated for the census, but this is not carried out as standard.

affiliations emerged in Romania: in the case of foreign citizens instead of their real ethnicity the country of origin was registered as ethnic affiliation³⁹.

Besides those listed, in certain countries there were further questions related to migration in the census. Thus, for instance, in Romania and Serbia the question about the *reason for migration* was included with regard to both those staying abroad and those arriving in the country (in Romania the latter only applies to persons temporarily present). At the same time, in Romania, as we have mentioned, there was a special questionnaire for those living abroad (even though they are not all included in the resident population – only those living abroad temporarily), which, besides basic demographic characteristics, also recorded the *destination country*, the *time of their departure* (grouped into categories), and *reason for departure*. In case of employment abroad, the *field of work* and the occurrence and frequency of *remittances* were also recorded. Applying such a special questionnaire was clearly made necessary by the large-scale emigration from Romania that has been characteristic of the past decades. The experience they yielded, however, is worth considering by all other countries which have long-term or recent experience of emigration. At the same time we also have to bear in mind that the validity of data collected about the household members staying abroad is not always sufficient: apart from basic data (destination country, time of departure), more detailed information (i.e. reason for migration, length of time intended to spend abroad) is not always exactly known by the family member.

By contrast, in Italy, a country that receives a considerable number of immigrants and has a massive foreign population, there was a separate sheet (called an ‘Individual sheet for foreign citizens not residing in Italy’) for people who were present on a temporary or occasional basis at the time of the census (e.g. visiting students, weekly commuters, as well as visiting tourists, short vacations, short-term medical treatments, visiting friends or relatives, etc.), but they were not counted in the resident population.

For countries with a significant number of immigrants it is also important to know the *country of birth of parents*⁴⁰, in order to identify the group of second-generation immigrants (descendants of immigrants). Nevertheless, a question referring to this was only included in the Italian census, where there has been a major increase in the number of foreign citizens since the 2001 census and the population with a foreign background (including second-generation immigrants and naturalized citizens, too) has increased even more. In Austria, this question has only been included in the micro census since 2008. Although the group of so-called second-generation migrants, who are born in the given country and have its citizenship but born from immigrant parents, also deserves attention in other countries, censuses of other countries do not extend to this question. Slovenia is the only country where data on first residence of a person and his/her parents and grandparents were derived in 2011 for the first time, despite (or because) of the register-based census, so data on first, second and third-generation migrants are available.

Another important data item about migration which censuses of different countries handle in divergent ways is the question of *second citizenship* (in the case of dual citizens)⁴¹. While in certain countries foreigners who preserve their own citizenship and also acquire that of the receiving country do not appear as foreigners (in terms of citizenship) because they only record the citizenship of the given country (e.g. in Italy), in other countries (e.g. in Hungary, Romania) both citizenships are recorded even if neither of them is that of the receiving country. Both citizenships were also recorded in Slovakia (e.g. the Slovak and other), but in cases where a person did not have Slovak citizenship only one citizenship was recorded. In Austria, earlier traditional censuses recorded

³⁹ For instance, a Kurdish immigrant from Turkey was registered as Turkish. Or an ethnic Romanian immigrant from Ukraine (if he/she did not possess Romanian citizenship) was registered as Ukrainian.

⁴⁰ This was a non-core topic in the CES Recommendations.

⁴¹ According to the EU Regulation a person with two or more citizenships has to be allocated to only one country of citizenship, in a determined order of precedence.

multiple citizenships up until 2001, but from 2011 onwards this was not included in the census, as the registers which serve as a basis do not contain relevant data. Thus if one of the citizenships is Austrian, this is most likely to be recorded (just like Italian citizenship in Italy).

As regards foreigners who have been naturalised (i.e. when the respondent has not carried the citizenship of the country since birth), an important question is *the year when the new citizenship was acquired*. Acquiring citizenship is an important milestone in the integration process, since this brings a whole line of rights and entitlements to the immigrant and thus makes integration easier. In Italy, the census contains data as to whether the respondent has held Italian citizenship since birth or acquired it by naturalisation, but there is no information about the year when citizenship was acquired. In the Austrian census (which relies on information derived from the Central Register of Residents ZMR) the date when a *Locally Competent Registration Authority* is informed about a person's change in citizenship is recorded (by quarter of a year). Statistics on naturalisations are currently not directly linked to the ZMR, hence data linkage based on bPK codes is not feasible (though micro data are available)⁴². In other countries the question on the year of acquisition of citizenship was not touched upon at all.

The data listed in *Regulation (EC) No 763/2008* as core topics concerning socio-demographic, economic, labour market, and human capital characteristics of the foreign or foreign-born population were included in the 2011 census of all countries.

4.4. The availability and quality of census data

The EU regulation also defines criteria for assessment of the quality of census data, such as their *relevance, accuracy, timeliness and punctuality, accessibility, clarity and comparability*. The relevance of these data is beyond debate and also their accuracy usually exceeds that of data from other data sources.

Timeliness (which refers to the delay between the reference time of the census and the availability of results) is undermined in certain countries by the fact that although the data were recorded in 2011, only preliminary figures have been published to date (July 2013). Furthermore, even in countries where the final data have been published, migration data have not been processed and the majority of data relevant from a migration point of view are not accessible (for example in Hungary even the final data published in March 2013 only included the number of foreign citizens resident in the country by sex and by main countries of citizenship, but no other break-down was provided⁴³). So this information could be out-of-date by the time of publication.

Nevertheless, census data often highlight important differences between the population counted by the census and the number of population resulting from the population estimation, particularly in countries where there has been significant immigration or emigration between the last two censuses. In Italy, one of the main problems of the census is that the foreign population is not fully covered. A survey on the coverage claims that even after the previous census, in 2001, a coverage rate of some 90 per cent was estimated for the entire country and nearly 75 per cent for metropolitan areas. The Post Enumeration Survey after the census of 2011 proved again this discrepancy between the number of people enumerated and those included in the population registers, even though the discrepancy is clearly shrinking compared to the previous period. The reason for this discrepancy is, as mentioned in the SEEMIG country report, partly the imperfect measurement of migration in the civil registry (especially because of failed cancellations due to emigration resulting in overestimation) and partly the already mentioned coverage error of the

⁴² From 2014 on the Central Citizenship Register (ZSR) in Austria will deliver more detailed information concerning this.

⁴³ As regards the number of persons resident abroad, the addition is made whereby 'this may be viewed as the minimum number of persons living abroad'.

census, which resulted in fewer foreign citizens included compared with the real number resident in the country.

In other countries (e.g. Romania, Hungary), the problems concerning registering emigration cause the discrepancy between the population enumerated and the population estimated on the basis of vital statistics. The under-registration in the number of emigrants causes over-estimation in the number of both the stock of foreigners and the whole population. In Romania's case the latter showed such a significant surplus (12.7 per cent) in 2011 that - as pointed out in the country report - questions were raised about the reliability of other demographic indicators delivered by the Statistical Office. In Hungary, census data concerning the stock of foreigners remained significantly behind the figure included in registers (by about 60,000 persons). This partly revealed the inaccuracy of registers data, but at the same time, also raised the question of the coverage of foreigners by the census (just as in the Italian example).

Although we may assume that inclusion of foreign nationals in the census is helped by having the questionnaire translated into several languages, the Italian example shows that even though the questionnaire has been available in eleven languages since 2011, accessing foreigners remains a problem.

Coverage error was a result of other factors in Serbia: ethnic Albanians living in southern parts of Serbia refused to provide information⁴⁴. According to the Serbian SEEMIG team, in the 2011 census in two municipalities in the southern part of Serbia (Bujanovac and Preševo) there was under-coverage of the census units owing to the boycott by most of the members of the Albanian ethnic community. The call for a boycott of the 2011 census in the Raška area had no significant impact on the responsiveness of the citizens. However, "according to Statistics Bureau data, the ethnic Albanian minority boycotted the census in southern Serbia, while the Bosniaks in the Sandzak region⁴⁵ mostly ignored the call for a boycott"⁴⁶.

An important requirement regarding the quality of census data is the *comparability* both in time and across countries. The former may be compromised by changes in the definition of the census population, changes in the definition and measurement of the various characteristics, and transition to the register-based census. The first of these cases exemplifies Romania, where the definition of the resident population has changed (in line with EU norms) compared to the previous, 1992 census, but the concepts used in the population registry did not follow this change. Comparison over time in South-East European countries is also hindered by the fact that migration-related census data in this region before 2001 was rather scarce. Comparison across countries improved considerably compared to earlier censuses after adopting EU norms (standard definitions and methods and standardization of certain questions), though at the same time, differences remain among different countries as regards the population covered and these need to be kept in mind when making comparisons.

⁴⁴ They were covered in all censuses since 1948, except for 1991 and 2011.

⁴⁵ However, the Sandzak region does not exist according to the law on territorial organisation.

⁴⁶ See http://newsdemo.atlasproject.eu/asset_demo/news/details/en/details.html?ci=06a14a1a-d504-4af3-8913-98463021904b.

Table 5: Type of the last census, the reference date and the population covered by census in SEEMIG countries

Country	Type of the last census	Reference date	Residence criterion				Special groups included			
			For foreign citizens (or stateless)		For nationals		Asylum seekers	Foreign diplomats	National diplomats	
			Duration of stay in country	Intended/actual	Duration of stay abroad	Intended/actual				
EU15	Austria	Register-based	31 October 2011	>3 months (90 days)	Actual	<3 months (90 days)	Actual	Yes	Yes, if main residence in AT	Yes, if main residence in AT
	Italy	Combined (enumeration, register, survey)	9 October 2011	Usual place of residence in Italy + possession of regular documentation for staying in Italy (no time criterion)		Less than 12 months	Actual & intended	Yes (who met the criteria of usual res. pop.)	Yes, but only if they do NOT hold a diplomatic passport	Yes
EU8	Hungary	Traditional	1 October 2011	>3 months	Actual	Less than 12 months ¹	Actual & intended	Yes (who met the criteria of usual res. pop.)	No	Yes
	Slovakia	Traditional	21 May 2011	>3 months ²	Actual	Less than 12 months	Actual & intended	No	No	Yes
	Slovenia	Register-based	1 January 2011	>12 months on the basis of registered permanent or temporary resident	Actual & intended	Less than 12 months	Actual & intended	Granted refugee or humanitarian status or asylum	No	Yes
EU2	Romania	Traditional	20 October 2011	>12 months	Actual	Less than 12 months ³	Actual & intended	No, those with refugee status yes	No	Yes
	Bulgaria	Traditional	1 February 2011	>12 months	Actual & intended	Less than 12 months	Intended	Granted refugee or humanitarian status or asylum	No	Yes
EU Candidate	Serbia	Traditional	30 September 2011	>12 months	Actual & intended	Less than 12 months, or for a year or longer, but kept re-entering the country at least once a week	Actual & intended	Yes (who met the criteria of usual resident population)	No	Yes

¹ In Hungary those who live abroad for more than one year were also counted (without any other data on them).

² Foreigners living in Slovakia for less than 90 days only had to fill in gender, date of birth and citizenship.

³ In Romania a special questionnaire was completed by those *living abroad for a long period of time* (more than 12 months) or *temporarily*, and another special questionnaire in the case of persons *temporarily present*.

Table 6: Main migration-related data included in the censuses in SEEMIG countries

Country		Country of citizenship	Country of birth	Previous place (country) of usual residence	Date (year) of arrival in the current place	Country of residence one year ago	Country of residence five years ago	Ever resided abroad (at least one year) & country	Year of arrival (or return) in the country	Ethnicity	Mother tongue	Religion
EU15	Austria	Yes	Yes	Yes	No	No	No	No ¹	Yes	No	No	No
	Italy ²	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No
EU8	Hungary	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes
	Slovakia ³	Yes	Yes	Yes	Yes	No	No	Yes	Yes (latest)	Yes	Yes	Yes
	Slovenia	Yes	Yes	Yes	No	No	No	Yes	Yes (first & latest)	No	No	No
EU2	Romania ⁴	Yes	Yes	Yes	Yes	No	No	No	Yes ⁵	Yes ⁶	Yes	Yes
	Bulgaria	Yes	Yes	Yes (change of usual place of residence between 2001 and 2011)	Yes	Yes	No	Yes (since 1980)	Yes	Yes	Yes	Yes
EU Candidate	Serbia ⁷	Yes	Yes	Yes	Yes	Could be derived indirectly	No	Yes	Yes	Yes	Yes	Yes

¹ Could generally be generated for persons born since 2002 (based on ZMR, due to bPK code).

² In Italy additional migration-related questions refer to place of birth of the parents (see in the text).

³ In Slovakia the reason for immigration to current permanent residence was also asked.

⁴ In Romania additional migration-related questions were asked about persons away for a long period of time or temporarily, and from those temporarily present (see body text).

⁵ More exactly: date of settling in Romania.

⁶ However, in Romania the country of origin was registered as ethnic affiliation for foreign citizens.

⁷ In Serbia the reason for moving to the country, as well as the reason for leaving the country and the length of absence was also asked.

5. CONCLUSIONS, CHALLENGES AND RECOMMENDATIONS

The current general tendency in statistical methodology is to use administrative data sources as widely as possible in order to decrease costs and the response burden on data providers. Administrative data make up a peculiarly big proportion of data sources of migration statistics as well. Data on immigrants gathered in administrative data sources are relatively rich, though in certain cases many important data are incomplete, as it can be seen in the detailed descriptions of SEEMIG country reports. Various situations of migrants are covered by administrative data, such as official stay and residence in a country of destination, labour force participation, enrolment in education, etc. Information on their socio-demographic, economic, labour market and human capital characteristics is also collected.

Almost all EU member states (and other European countries) have been moving (or are intending to move) towards an increased use of administrative data sources for statistical purposes, either as a substitution or as a complement to information previously collected by surveys. This shift is taking place in all areas of statistics. Although certain countries are at different stages of this process, steps have already been taken in most of them to reinforce legal bases for more effective use of administrative data for statistical purposes and to better regulate the interplay of actors in charge of data production. However, there are still many obstacles and challenges to the utilisation of non-statistical data, especially in terms of access to data at the necessary level of aggregation, and in terms of the compatibility of different data sources and their accommodation to statistical needs.

There are many administrative registers usable for migration statistics in all SEEMIG countries and some of them have not yet been exploited. The reasons for this are multifarious, ranging from lack of governmental will to use administrative data for statistical purposes, to inadequate legal regulation, and a failure of cooperation between the administrative and statistical parties. Moreover, concepts and classification used by authorities often do not meet statistical requirements, and register data are of low quality. For migration statistics it would be essential to access administrative data at the personal level but this is not possible in some SEEMIG countries, since these data are usually aggregated. In certain countries migration-related data sources do not fulfil the necessary minimum quality criteria for statistical use. Regarding migration aspects one of the most frequent problems is the under-coverage of the migrant population.

In Serbia, the only non-EU, but candidate country in the project, the use of administrative data for migration statistics is rather limited and statistical aspects in registers are vindicated less than in EU countries. Furthermore, the registers related to migration need to be established or improved. Hampered access to administrative data, the lack of cooperation between administrative data collectors and the statistical office, as well as the lack of published definitions used in registers contribute to the rather poor estimation procedures on the basis of available data sources. However, new legislation (Law on Migration Management, November 2012) offers the potential for positive change.

Besides all deficiencies, the use of administrative data for statistical purposes is inevitable, especially in the field of migration. Data collected in electronic formats offer such budgetary advantages for statistical offices that eventually all states will take them into account. Nevertheless, statisticians and researchers have to be familiar with the *problems of re-using data collected for non-statistical purposes* (see the details in the *Appendix*):

- in most countries data owners have an opposing attitude regarding the access of data at the necessary level due to the lack of legal conditions for the utilization of administrative data for statistical purposes;

- the lack of close cooperation between data owner public authorities and the statistical institute, which could be necessary for the persistent good quality of data collected for non-statistical purposes;
- the lack of public knowledge on and agreement with the use of administrative data for statistical purposes;
- the definitions used in administrative data sources are different from the statistical standards;
- the quality is frequently not the same for all data in an administrative register;
- the use of administrative data necessitates the application of statistical methods that are more sophisticated than traditional ones (data integration and inventive estimations complement the palette of statistical procedures).

Specific issues of reliability have to be taken into account when making use of migration-related register-based data. Migrants (particularly from third countries) frequently have a strong interest in acquiring permission to stay and then settle or work in the host country, which can lead to manoeuvring among different types of applications and which may bias the data (Gárdos et al. 2009). These non-sampling errors cannot be measured exactly.

Specific *requirements concerning the use of administrative data* include:

- deep insight into the legal regulation, including practical conditions of the data collection as well as the concepts used,
- deep insight into all of the data sources, both administrative and statistical, which comprise data in the scope of the research interest,
- integration of the data sources being more effective than the use of single data sources,
- the need to define a key variable, a certain type of PIN, which integrates and enables personally identifiable records,
- omitting any piece of information from the administrative data source, if there is a strong and reasonable doubt about its quality (similarly in the case of full-coverage statistical data collections).

In order to improve the quality of migration-related data resulted from **administrative sources** the following *recommendations* should be considered in all SEEMIG countries:

Integration of the foreigners' registers and population registers

In the European and global context only a few countries, such as Austria and Slovenia among the SEEMIG countries, can say that the migration-related data currently available are able to meet international requirements and can consequently be evaluated as being of good quality. This is particularly true of the population registers. The population register should cover the whole population residing in the country including foreigners. However, foreign citizens with a residence permit are not necessarily registered in the population register on the one hand, and the population register may also include foreign citizens whose residence permit has expired on the other. Therefore, the population register and the register of foreigners may provide different estimates. This situation argues at least from the perspective of migration statistics for integrating the foreigners' and population registers.

In most EU countries the coverage of EEA citizens in the foreigners' register is not precise as they are not obliged to report themselves or a change of their usual place of residence within the EEA. However, similarly to citizens of the country (nationals), they do have to declare their residence in the population register. Regular cooperation between data handling institutes and harmonization of data in the two registers can contribute to having more precise information on the foreign population living in a country.

Completion of administrative data sources with new migration-related variables

Harmonising different data-sets is essential, as it would be desirable to include *country of birth* in various fields of statistics, e.g. on labour market or educational attainment. As a next future step, it would constitute a major enrichment of the data to include the variable *country of birth of parents*. This would allow more multi-faceted analyses of migratory patterns and processes, although there are valid reasons to be concerned about the systematic collection of such data. There may be a risk of their abuse for political purposes. The term ‘migrant’ can be considered a strong form of categorisation, particularly to those who may not consider such a label important or relevant to their everyday lives. The terms ‘second-generation migrant’, ‘third-generation migrant’, etc., can also be problematic in this respect. However, if there are no sufficient data about the migrant population (or those with migrant-background) for statistical and scientific investigation it is not possible to implement effective governmental and non-governmental integration programmes. For example, Scandinavian countries have information available for such purposes. Nevertheless, the right balance between political risk and the need for further data must be sought in each individual country.

Furthermore, the collection of data on *previous and next residence* of migrants in registers would improve the quality of the national migrant statistics. This would contribute to learning more about the paths of foreign migrants. The more mobile a person is, the higher the probability that the countries of origin and citizenship are different. The question to be included would simply ask about the country from where the migrant has arrived or the country of destination if the migrant is leaving.

The *previous country of residence* of immigrants would be especially important to be known – and not only from census data – in countries which have separate citizenship law for ethnic kin minorities from neighbouring countries (e.g.: Hungary, Romania etc.). There might be persons among immigrants from these neighbouring countries who already have the citizenship of the receiving country at the time of their arrival, despite the fact that they were not born in the destination country nor have they lived there before (so they are not return migrants). The identification of these immigrants could only be accurate by country of birth and previous country of residence together.

The use of a unified system of PINs

Administrative data, in many cases, can be effectively utilized for statistical purposes through data integration, which requires common identification system in the different data files. This highlights the need to use a unified system of PINs. Moreover, the idea of an EU-level personal identifier can also come up with respect to migration issues, especially considering the shortcomings of emigration data. Nevertheless, it should be kept in mind that the PIN clearly serves the purposes of governmental interoperability initiatives, mostly at the national level. Using an EU-PIN could be proposed to policy makers because it contributes not only to better migration data but to having clearer and more effective national programs of social insurances and benefits.

Steps towards improving emigration statistics on national level

The collection of some specific data will remain a challenge in the future, for instance reliably tracing emigrant flows and stocks, because these data are still based on estimates or on de-registrations as recorded in the population register. In general, in all SEEMIG countries, emigration is likely to be underestimated in administrative data sources, as they only show an administrative reality. There are many reasons why people do not declare their leaving the home

country for a longer period, such as trying their luck without final decision, keeping social benefits, subsidies and other entitlements etc.

In order to improve emigration statistics on national-level, *incentives* (e.g. a continuation of receiving social benefits after leaving the country) should be introduced into the (de)registration systems *both for persons involved in migration and administrative staff taking part in the data collection process*. However, the underestimation of emigration results in an overestimation of the net migration, and therefore inaccurate population estimations. This is a problem particularly in the case of emigration countries, where therefore Statistical Offices do not have appropriate figures on the population size and basic demographic indicators. Consequently, national-level administrative data sources alone are insufficient for describing the migration profile of a country, and therefore other tools should be used.

Steps towards improving emigration statistics on international level

International efforts should be made to harmonise the foreigners' registers to a certain extent in order to enable them to be used for mirror statistics, thus improving statistical data on emigration for countries of origin. This is inevitably important for emigration countries. The problem of enumerating emigrants cannot be solved on the national-level. While it is difficult to count emigrants by administrative means, censuses and sample surveys also have difficulties in detecting (or covering) absent persons, especially when the whole household has left the country.

Cooperation of statistical offices, especially between those of emigration countries with the ones of destination countries would contribute to improving population estimations and the related statistical indicators. The need for closer *collaboration between countries* is not a new idea: it was already highlighted by the United Nations in 1953, and *improving the coordination of migration statistics at the international level* has also been claimed (Herm 2006). Recently, in order to compensate the weaknesses of emigration data in sending countries, the use of immigration data of the destination countries was proposed by UN Economic Commission for Europe (2010). Moreover, methods for estimating missing migration flow data (Raymer 2008) and for harmonising data available from migrant sending and migrant receiving countries (De Beer et al. 2010) has already been developed. Although international cooperation in order to enhance the comparability of migration data is very important, it is also difficult to implement.

Besides administrative data sources **sample surveys** are another important source for migration-related data. They could cover a wide range of topics, enable a more problem-oriented data collection, and therefore a detailed analysis of various characteristics and integration indicators of migrants. Surveys are especially important tools for migration-related data collection, as the growing diversity and complexity of migration movements make it increasingly difficult for administrative data sources to cover the entire migrant population.

Deficiencies of migration statistics, difficulties of migration-related data collection and the scarcity of internationally comparable data are evident in most European countries. As has been pointed out by experts before (e.g. Fassmann, Reeger and Sievers 2009), collection and aggregation of data from national statistics (e.g. by OECD or Eurostat) can only result in international-level databases of better quality – and thus a breakthrough in migration analysis – if considerable improvements are made at the national level in terms of data collection practices, or if a standardised, survey-based, regularly conducted migration-related data collection is introduced in Europe (similarly to already existing international surveys on other topics). So besides the improvement of administrative data sources, either adequate regular statistical data collection or *systematic, internationally standardised surveys on migration* could offer a solution for migration-related data shortage.

Since this latter option is not likely to become a reality in the near future, it would be advisable to *design surveys at the national level, which are linked to regular international surveys* (and thus

ensure comparability and consistency), but conducted on a separate representative sample of immigrants (as in the Italian example). The problem of under-representation and the small sample size of immigrants – which are the main limits of surveys representative of the whole population – could be mitigated this way. However, precise registers which ensure a proper sampling frame are necessary for this (and these could be made available for sending countries too).

Also *the usability of existent large international surveys should be improved* (i.e. the EU-LFS, EU-SILC). In countries with large-scale immigration, increasing the coverage of foreigners in the surveys could be a solution (by *mitigating language barriers during fieldwork, boosting the immigrant sub-sample or adjusting this to the real territorial distribution of immigrants*). In countries with large-scale emigration the *inclusion of supplementary questions in already existing international surveys regarding emigration or labour force out-migration* should be proposed: e.g. out-migration of current or former household members or other well-defined relatives. The experiences of and lessons learnt from the SEEMIG pilot study should be considered in this regard (see: Blaskó – Jamalia 2014). Regarding *return migration*: additional questions on previous (work) experience abroad (on the reason for leaving the country, the date of first departure/last return, the duration of residence abroad etc.) should also be included in these surveys.

At the same time, *harmonisation of the definition of ‘household member’* in international surveys is recommended for improving the comparability of these surveys’ data. Furthermore, the *harmonisation and wider selection of non-response codes* is also recommended, in order to be able to identify when the whole household had gone abroad.

The importance of surveys and statistical data sources differs markedly by countries, depending on the relevance, size and intensity of migration. Countries covered by the SEEMIG project are partly immigration countries with high versus low immigration, and partly emigration countries with intensive or moderate emigration; some countries have experienced both emigration and immigration at various periods and sometimes on a large scale (see more detail in: Fassmann – Musil – Gruber 2014). Consequently, detailed national level surveys have been conducted in countries where emigration and/or immigration has been important and has attracted more policy attention. The focus of the surveys reflects the main migration tendencies of the respective country. Among the national surveys overviewed, there were some (highlighted in sub-chapter 3.3.2.) that could serve as good examples worthy of following and/or improving upon in other countries too.

The advantages offered by surveys with richer data content should definitely be exploited, since in the majority of countries – as this overview has shown – administrative data sources (to differing degrees) lack important characteristics of immigrant population, like some of the already mentioned Zaragoza indicators. Their exploration can only be achieved by targeted surveys. At the same time, the possible *introduction of longitudinal surveys* should also be considered, as they would provide essential information about the integration process of immigrants (relevant in countries with considerable immigrant populations), as well as about the changing patterns of emigration.

As regards surveys on emigration or labour out-migration, an important lesson learnt here is that the opportunities for sending countries are limited – while an innovative pilot methodology, as mentioned above, was developed within SEEMIG to reach out-migrants (people staying abroad for a defined period). Also, we have seen that *cooperation between research groups of the country of origin and those of the destination country* greatly increases the success of such surveys.

The last type of data sources discussed in this report is the **census**. Censuses are especially important for migration analyses since they collect data on various migrant groups (foreign citizens, foreign-born persons and ever-international migrants) using an EU-level harmonised methodology. Last censuses were conducted complying with EU standards in all SEEMIG countries (even in EU candidate state Serbia). Data identifying migrants (country of citizenship and country of birth) were included in all the censuses, as well as data identifying ethnic minorities (ethnicity, mother tongue and religion)

in many cases. But further *harmonisation of questions referring to migration history* (former places of residence and date of arrival in the current place) is recommended, which would make comparison across countries easier. Although these questions were asked in some form in all countries, inclusion was not uniform.

There are also differences across countries in terms of the population covered in the census – regarding both immigrants and emigrants (e.g. 3 months vs. 12 months resident criterion, those living abroad included vs. excluded, asylum seekers included vs. excluded). The further *harmonisation of population covered by censuses* is also recommended.

Although the advantages of register-based censuses have been stressed for a long time, and developments seem to be moving in this direction, only two SEEMIG countries (Austria and Slovenia) had such censuses in 2011. The reason for this is that such censuses need long-term planning and preparation, as well as substantial financial resources to set up adequate data systems, since data which are not included in any kind of register cannot be covered by a census. Besides *suitable legislation, source registers with unique identifiers that enable data sources to be linked* are also needed.

Since Member States have a choice whether to include further questions in the census besides the mandatory ones, countries considerably affected by some form of migration (immigration or out-migration) might consider including more questions in this topic. Regarding immigrant population, *inclusion of the following topics is recommended*, based on the comparative overview (see more details in chapter 4.4.):

- *Legal status of foreign citizens*: it would be particularly important in case of third-country nationals (immediately after arrival and at the time of the census);
- *Date of acquisition of the citizenship and the type of naturalization*: in case of individuals who acquired the citizenship of the given country later than their date of birth; in this regard the case of new citizens who acquired their citizenship when new states were created (after dissolution of former States) needs special attention;
- *Recording all citizenships*: in the case of individuals with dual citizenship (in countries where this has not been a practice yet, and the dual citizenship is allowed by law), and *harmonization of the registration mode of multiple citizenship*;
- *Country of birth or (former) citizenship of parents*: in order to identify second-generation migrants.

Some of the above (like country of birth of parents and citizenship acquisition) were already recommended by the UN Economic Commission for Europe as non-core topics for the 2010 Censuses, along with reason for migration (UNECE 2006).

While censuses provide a more or less sufficient overview of the population with immigrant background, and enable comparison of various immigrant groups and at lower geographical units, they are less satisfactory in terms of *out-migration* and *return migration*. Part of return migration (in the case of migrants staying abroad for a period of more than a year) is recorded by censuses in most SEEMIG countries, while out-migration is usually not or less well recorded. Based on the census, the size of the population decrease in comparison with population estimation based on registers can be observed, i.e. people who left the country between two consecutive censuses and possibly their significant share in some age groups. But the date of emigration, the destination country and the reason for migration is only discovered if there is a separate questionnaire (or emigration module) about this – completed by former household members living in the country of origin or by former neighbours. Obviously, this only records some part of out-migration, and reliability of data received in this manner is rather questionable.

Although registering out-migration is a problem to be solved in most countries, it is also a problem that seems to exceed the opportunities provided by the census. Similarly to surveys, censuses have limitations in counting absent persons, especially in case of the emigration of the entire households. Despite this, and taking into account the previous attempts made for estimating emigration through an emigration module in the census (Camilleri 2006), *inclusion of some supplementary census questions regarding emigration and labour out-migration, as well as return migration* is recommended. Although national-level development of censuses could also be an important step, the real improvement would be if all suggested topics would be regulated at the EU-level.

In addition, the cooperation and data exchange between countries is important in the case of census data, too. Immigration data from destination countries are important for estimating out-migration from countries of origin. In this regard the new dissemination system of census data, the Census Hub developed by Eurostat – which is expected to be officially announced in the summer of 2014⁴⁷ – is an important and expectedly very useful development.

On the whole we can conclude that the enhancement of migration-related data sources should be initiated in all major fields (administrative, survey, census), because only a harmonised, complex structure of all these data sources can provide an accurate and comprehensive picture on the phenomenon of migration.

⁴⁷ See more: http://www.unec.org/fileadmin/DAM/stats/documents/ece/ces/ge.50/2014/Topic_2_Eurostat.pdf.

6. REFERENCES

SEEMIG Country Reports

Bleha, B. – Jurčová, D. – Mészáros, J. – Šprocha, B. – Vaňo, B. (2013): *Analysis of existing migratory data production systems and major data sources in Slovakia*. Country report developed within the project 'SEEMIG Managing Migration and Its Effects – Transnational Actions Towards Evidence Based Strategies'.

<http://www.seemig.eu/downloads/outputs/SEEMIGDataSystemsCountryReportSlovakia.pdf>

Fassmann, H. – Gruber, K. – Musil, E. (2013): *Analysis of existing migratory data production systems and major data sources in Austria*. Country report developed within the project 'SEEMIG Managing Migration and Its Effects – Transnational Actions Towards Evidence Based Strategies'.

<http://www.seemig.eu/downloads/outputs/SEEMIGDataSystemsCountryReportAustria.pdf>

Gárdos, É. – Gödri, I. (2013): *Analysis of existing migratory data production systems and major data sources in Hungary*. Country report developed within the project 'SEEMIG Managing Migration and Its Effects – Transnational Actions Towards Evidence Based Strategies'.

<http://www.seemig.eu/downloads/outputs/SEEMIGDataSystemsCountryReportHungary.pdf>

ISS, SORS (2013): *Analysis of existing migratory data production systems and major data sources in Serbia*. Country report developed within the project 'SEEMIG Managing Migration and Its Effects – Transnational Actions Towards Evidence Based Strategies'.

<http://www.seemig.eu/downloads/outputs/SEEMIGDataSystemsCountryReportSerbia.pdf>

Kiss, T. (2013): *Analysis of existing migratory data production systems and major data sources in Romania*. Country report developed within the project 'SEEMIG Managing Migration and Its Effects – Transnational Actions Towards Evidence Based Strategies'.

<http://www.seemig.eu/downloads/outputs/SEEMIGDataSystemsCountryReportRomania.pdf>

Kostova, M. – Yakimova, E. (2013): *Analysis of existing migratory data production systems and major data sources in Bulgaria*. Country report developed within the project 'SEEMIG Managing Migration and Its Effects – Transnational Actions Towards Evidence Based Strategies'.

<http://www.seemig.eu/downloads/outputs/SEEMIGDataSystemsCountryReportBulgaria.pdf>

Piovesan, S. (2013): *Analysis of existing migratory data production systems and major data sources in Italy*. Country report developed within the project 'SEEMIG Managing Migration and Its Effects – Transnational Actions Towards Evidence Based Strategies'.

<http://www.seemig.eu/downloads/outputs/SEEMIGDataSystemsCountryReportItaly.pdf>

Stropnik, N. – Kump, N. – Povhe, J. – Jurišić, B. (2013): *Analysis of existing migratory data production systems and major data sources in Slovenia*. Country report developed within the project 'SEEMIG Managing Migration and Its Effects – Transnational Actions Towards Evidence Based Strategies'.

<http://www.seemig.eu/downloads/outputs/SEEMIGDataSystemsCountryReportSlovenia.pdf>

Literature

Albisinni, M., Marzilli, E. and Pintaldi, F. (2008): Cognitive Test related to Labour Force Survey ad-hoc Module on Migrants and their Immediate Descendants. Presentation at the European Conference on Quality in Official Statistics. Rome, 9-11 July 2008.

Bijak, J. and Wiśniowski, A. (2011): *Statistical Modelling of International Migration Flows*, Int. Statistical Inst.: Proc. 58th World Statistical Congress, 2011, Dublin, <http://2011.isiproceedings.org/papers/650247.pdf>

Blaskó, Zs. and Jamalia, N. (2014): *Surveying emigration I. Report on the first stage of the SEEMIG pilot study in Hungary and Serbia*. Research report developed within the project 'SEEMIG Managing Migration and Its Effects – Transnational Actions Towards Evidence Based Strategies'.

<http://www.seemig.eu/downloads/outputs/SEEMIGPilotReport1.pdf>

Camilleri, R. (2006): *Estimating emigration through an emigration module in the census. A proposal for next censuses*. MEDSTAT Programme. European Communities: Luxembourg.

Center for Global Development (2009): *Migrants Count. Five steps Toward Better Migration Data*. Report of the Commission on International Migration Data for Development Research and Policy. Washington.

De Beer, J., Raymer, J., Van der Erf, R. and Van Wissen, L. (2010): Overcoming the Problems of Inconsistent International Migration data: A New Method Applied to Flows in Europe. *European Journal of Population* (2010): 26: 459-481.

European Communities (1999): *Use of Administrative Sources for Business Statistics Purposes – Handbook on Good Practices*, Eurostat. http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/CA-24-99-897/EN/CA-24-99-897-EN.PDF

European Statistical System (2011): *European Statistics Code of Practice*. For the national and community statistical authorities, Eurostat. http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-32-11-955/EN/KS-32-11-955-EN.PDF

European Statistical System (2013): ESS agreement – Labour Force Survey (LFS) ad-hoc module 2014 on the labour market situation of migrants and their immediate descendants. Published on 5th March 2013. http://epp.eurostat.ec.europa.eu/portal/page/portal/pgp_ess/0_DOCS/estat/ESS%20agreement%20LFS%202014%20module%20with%20annex%20EN.pdf

Eurostat (2011): *Migrants in Europe. A statistical portrait of the first and second generation*. Luxembourg: Publications Office of the European Union. http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-31-10-539/EN/KS-31-10-539-EN.PDF

Fassmann, H. (2009): European migration: Historical overview and statistical problems. In: Fassmann, H., U. Reeger and W. Sievers (eds.) *Statistics and Reality. Concepts and Measurements of Migration in Europe*. IMISCOE Reports, Amsterdam, University Press: 21-44.

Fassmann, H., Reeger, U. and Sievers, U. (eds.) (2009): *Statistics and Reality. Concepts and Measurements of Migration in Europe*. IMISCOE Reports, Amsterdam, University Press.

Fassmann, H. and Musil, E. (2012): *Data requirement paper for measuring longer term migratory, labour market and human capital processes*. Paper developed within the project 'SEEMIG Managing Migration and Its Effects – Transnational Actions Towards Evidence Based Strategies'. <http://www.seemig.eu/downloads/outputs/SEEMIGDataRequirementPaper.pdf>

Fassmann, H. and Musil, E. (2013): *Conceptual framework for modelling longer term migratory, labour market and human capital processes*. Working Paper developed within the project 'SEEMIG Managing Migration and Its Effects – Transnational Actions Towards Evidence Based Strategies', <http://www.seemig.eu/downloads/outputs/SEEMIGConceptualPaper.pdf>

Fassmann, H., Musil, E. and Gruber, K. (2014): *Dynamic Historical Analysis of Longer Term Migratory, Labour Market and Human Capital Processes in the SEEMIG Region*. Synthesis report developed within the project 'SEEMIG Managing Migration and Its Effects – Transnational Actions Towards Evidence Based Strategies'. <http://www.seemig.eu/downloads/outputs/SEEMIGHistoricalAnalysisSEERegion.pdf>

Gárdos, É., Gödri, I., Hárs, Á., Hablicsek, L., Illés, S., Molodikova, I., Melegh, A., Sárosi, A., Tóth, P. P. and Tóth, J. (2009): *Fejlesztési tanulmány. A bevándorlási és integrációs statisztikai rendszer fejlesztése. (Developmental study. Development of the statistical system of immigration and integration)*, EIA/2007/3.2.3.1, KSH Népeségstudományi Kutató Intézet (HDRI), Budapest.

Herm, A. (2006): Recommendations on International Migration Statistics and Development of Data Collection at an International Level In: Poulain, M., N. Perrin and A. Singleton (eds.) (2006): *THESIM Towards Harmonised European Statistics on International Migration*. Presses Universitaires de Louvain: 77-106.

Kraler, A. and Reichel, D. (2010): *Statistics on Migration, Integration and Discrimination in Europe*. PROMINSTAT Final Report, ICMPD, June 2010. http://www.prominstat.eu/drupal/?q=system/files/PROMINSTAT_FINAL_REPORT.pdf

OECD (2010): *PISA 2009 Results: Overcoming Social Background – Equity in Learning Opportunities and Outcomes* (Volume II), <http://dx.doi.org/10.1787/9789264091504-en>

OECD (2012): *Education at a Glance 2012: OECD Indicators*, OECD Publishing. <http://dx.doi.org/10.1787/eag-2012-en>

Poulain, M., Perrin, N. and Singleton A. (editors) (2006): *THESIM Towards Harmonised European Statistics on International Migration*. Presses Universitaires de Louvain, 744 pages, <http://www.uclouvain.be/en-14230.html>

Raymer, J. (2008): Obtaining an overall picture of population movement in the European Union, In: J. Raymer and F. Willekens (eds): *International migration in Europe: Data, models and estimates*, John Wiley & Sons Inc, Chichester: 209-234.

Regulation (EC) No 763/2008 of the European Parliament and of the Council of 9 July 2008 on population and housing censuses, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:218:0014:0020:EN:PDF>

Regulation (EC) No 862/2007 of the European Parliament and of the Council of 11 July 2007 on Community statistics on migration and international protection and repealing Council Regulation (EC) No 311/76 on the compilation of statistics on foreign workers, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32007R0862:EN:NOT>

Sik, E. (2012): Doomed to failure – with some chance to success: Migration statistics in the 21st century. Presentation at the SEEMIG Launching Conference: *'Migration, demography and labour markets. Long term processes and linkages'*, 19-20 September 2012, Bratislava, Slovakia.

Statistics Finland (2004): *Use of Registers and Administrative Data Sources for Statistical Purposes Best Practices of Statistics Finland*, http://www.stat.fi/tup/julkaisut/kasikirjoja_45_en.pdf

UNECE (UN Economic Commission for Europe) (2006): *Conference of European Statisticians Recommendations for the 2010 Censuses of Population and Housing*. Prepared in cooperation with the Statistical Office of the European Communities (EUROSTAT), United Nations: New York and Geneva.

UNECE (UN Economic Commission for Europe) (2007): *Register-based Statistics in the Nordic Countries – Review of Best Practices with Focus on Population and Social Statistics*. http://www.unece.org/fileadmin/DAM/stats/publications/Register_based_statistics_in_Nordic_countries.pdf

UNECE (UN Economic Commission for Europe) (2010): *Guidelines for Exchanging Data to Improve Emigration Statistics*. Prepared by the Task Force on Measuring Emigration Using Data Collected by the Receiving Country, United Nations: Geneva.

UNECE (UN Economic Commission for Europe) (2011): *Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices*. United Nations: New York and Geneva. <http://unstats.un.org/unsd/EconStatKB/Attachment442.aspx>

United Nations: *Fundamental Principles of Official Statistics*, <http://unstats.un.org/unsd/methods/statorg/FP-English.htm>

United Nations (1998): *Recommendations on Statistics of International Migration*. Statistical Papers Series M, No. 58, Rev. 1. Department of Economic and Social Affairs, Statistics Division, New York.

United Nations (2008): *Principles and Recommendations for Population and Housing Censuses*. Statistical Papers Series M, No. 67, Rev. 2. Department of Economic and Social Affairs, Statistics Division, New York.

Zaragoza Declaration (2010): http://ec.europa.eu/ewsi/UDRW/images/items/docl_13055_519941744.pdf

APPENDIX

General overview and evaluation of using administrative data for statistical purposes⁴⁸

What are administrative data?

Administrative data are registers established for administrative purposes.

Administrative records contain data concerning enforcement of rights or doing so as required by law which are often related to budgetary revenues and expenditures. The statistical use administrative data requires comprehensive and reliable data sources that cover the country as a whole, and can be individually considered consistent from the point of view of methodology and quality. The fact that more and more administrative data sets are available for statistical purposes is partially a result of the rapid development of information technology over the past few decades. The most important records are kept in electronic formats, which makes it possible to take advantage of all the benefits that come with the statistical use.

It must be emphasised that knowing individual data is the aim of an administrative register, while this is only expedient in statistical data production. Hereinafter we discuss administrative data as potential data sources for statistical data production that, according to the Code of Practice, must be carried out by the statistical authority independent of political and other external interference.

The advantages of using administrative data sources

Cost savings

A statistical survey is a costly data-collection process. Although information technology exempts human resources from a significant part of the processing burden, statistical data collection is still quite labour intensive, especially in terms of finding respondents and arranging interviews with them, phases of work that will probably never be fully automated. Moreover, most of the administrative data have now been made accessible in electronic format. Consequently, there is no reason to collect information more than once.

The resource needs of a census carried out in the traditional manner are much higher than those of sample surveys, because the tasks are to be carried out on a much larger scale. Statistical offices that hold traditional census data require extra funding for this work, because it is too costly to be funded from their own regular budgets. This explains the fact that censuses can be carried out in a traditional manner very rarely, usually every ten years or so, and thus certain statistical indicators can be produced only over such a long time period, while the general public demand more and more recent data. Traditional censuses are exposed to political priorities due to their large cost requirements.

Although the cost of establishing a statistical system based on administrative sources can reach the amount that is necessary to set up a system using statistical data collection, the operational costs of the latter are usually considerably lower. Censuses conducted in European countries in 2000-2001 show the differences of the two methods in terms of costs. Most countries use administrative data

⁴⁸ This section was prepared based on following documents: European Communities 1999, Statistics Finland 2004, European Statistical System 2011, UNECE 2007, 2011, UN: Fundamental Principles of Official Statistics.

sources combined with traditional data collections in censuses, and consequently the unit cost varies widely. In countries where the census was basically carried out by statistical data collection, the per capita cost amounts to more ten times than that in Finland, where the census was based entirely on administrative sources. In Finland, the cost of compiling census data was 0.2 EUR per person while it was close to 14 EUR in Switzerland⁴⁹. We can say that this is the strongest argument in favour of using administrative data more widely in statistics.

Access to administrative sources is usually free, especially if the data come from the public sector. However, even if the data owners charge costs for the production of data files needed by a statistical authority, the total amount will be significantly lower than if the same data had been collected via a statistical questionnaire.

Reducing the burden on respondents and other resources

Use of any external data source for statistical purposes means that the primary data provider is exempted from fulfilling requests for new data, which in turn reduces the burden on data providers. In many countries, such as Hungary, reducing the burden on data providers, particularly in the sphere of business, is a strategic goal of the government.

Reducing the burden on data providers also figures among EU regulations. The duties of statistical institutes including the utilisation of administrative data for statistical purposes have been laid out in the Code of Practice of the European Statistics.

Frequency of statistical data production

It is one of the other advantages of the use of administrative sources that in some cases it is possible to produce statistics more frequently, without increasing the burden on respondents while the cost only increases slightly. In countries where administrative data sources are used, annual census indicators can be calculated, while in others where traditional methods are followed, it may only be done once every five or ten years.

Coverage

Administrative sources are mostly kept for implementing national public programmes, so from this perspective they are full coverage, often ensuring complete or nearly complete coverage of the target group concerned. Consequently, in most cases there are no sampling errors and non-responses. As a result, more accurate and detailed estimates can be provided for sub-populations, such as small areas, or other respondents of unique characteristics.

Timeliness

The use of administrative resources increases the timeliness of the statistical data by providing access to more up-to-date information on variables. Statistical surveys, on the other hand, usually take a long time from conception to producing results, and this is especially true of annual or *ad hoc* data collections.

Utilisation of these three advantageous properties of use of administrative data for statistical purposes is particularly possible when the occurrence of the investigated event is relatively rare, and its distribution in space, society or time, etc., is uneven. In such cases statistical indicators can be produced with a statistical survey of a very large and complex sample, which is very time-consuming and costly, and therefore it can be prepared infrequently. Moreover, it is not clear whether they indeed characterise the analysed phenomenon. In these cases it is better to use administrative data sources (if any). It is much cheaper and the required statistics can therefore be published more frequently.

⁴⁹ Source: Documentation of the 2000 round of population and Housing censuses in the EU, EFTA and Candidate Countries (Eurostat), http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-CC-04-002/EN/KS-CC-04-002-EN.PDF.

Administrative data is mainly useful for deriving objective indicators, and certainly cannot take over the role of sociological surveys or that of subjective data; however, they can serve as a framework for sample surveys based on a statistical questionnaire.

Public image of statistical activities

Public opinion about sharing of information varies considerably from country to country, and this is especially true in cases where exchange of data between different government agencies is involved. If the public generally accepts and supports the sharing of data, the statistical office's prestige can be increased by making better use of existing data sources, as its activity is more efficient and cost effective.

Challenges in the use of administrative data sources

A number of advantages of using administrative data sources are listed above; however, there are also challenges in using these data sources for statistical purposes. Some of these are related to specific data sets only, but most of them are general.

Access to administrative data

One of the main obstacles to widespread use of administrative data sources for statistical purposes is lack of access to the data. The conditions necessary for ensuring access for statistical use are multi-faceted: legal, political, organisational, and technical. The most important is the legal criteria making it possible to use administrative data for statistical purposes. The other conditions are less important. Nevertheless, being familiar with them is useful in order to smooth the flow of data as well as to minimise the risk of any problems or misunderstandings between data owners and statistical associations.

Legal conditions

The Statistics Act is typically the legal rule that defines the role and tasks of statistical agencies. It usually stipulates the conditions under which the institutions producing official statistics may have access to external data and use them. If the statistical organisation has no general authorisation to access any data, except those of a 'secret' nature in order to fulfil its legal duties, then sector-specific laws and regulations specify the terms that define the use of non-statistical data sources for statistical purposes. In addition, the statistical regulations of the European Union provide for the accessibility of administrative data, which is binding on the member states. So it may be possible that two or more laws regulate access to administrative data.

Although the 'statistical law' of the European Union (European Statistics 223/2009 Regulation) declares that member states' national statistical agencies 'shall have access to administrative data sources, from within their respective public administrative system, to the extent that these data are necessary for the development, production and dissemination of European statistics', it also states that 'practical arrangements and the conditions for achieving effective access shall be determined where necessary by each member state ... within their respective spheres of competence'. European Union regulations on certain specific statistical areas, however, go further and eliminate this dependence on the local legal rules. An example of this is the European regulation on the statistical business register, which provides unlimited access to all administrative sources whose data are needed to compile it.

While national regulations define the conditions for accessing non-statistical data, they also determine the boundaries of use. Constraints, such as those stating that administrative data can be used only for specific statistical purposes and that personal data have to be protected are very

common. This last clause even prevents access to individual business information when personal and business data overlap (e.g. the self-employed).

Cooperation with public authorities

A unified effort on the part of public authorities, in order to build a statistical system that use data sources collected for non-statistical purposes to the greatest possible extent, can only be truly effective if it is supported by strong and clear commitment at governmental level and there is close cooperation between the authorities concerned. It is absolutely necessary for producing good quality statistical data that the data owner ensures so-called descriptive information, i.e. information that is necessary for the accurate interpretation of data of the received file. If possible, data reception and data quality should be formally agreed with the data owner in order to reduce risks of corrupted data quality due to scant data collection, insufficient data entry or delayed data transfer.

Setting up large, comprehensive statistical data systems (e.g. a census) on the basis of administrative data makes it particularly necessary for the statistical agency to have strong bargaining power in relation to other authorities.

Public agreement

To ensure that administrative data be considered predictable and reliable data sources of statistics, it is necessary for the public to agree with the sharing of data among governmental agencies. In some countries, depending on the political atmosphere and historical experience, it is relatively easy to achieve the agreement compared with others. In some countries people are wary of losing control of their own data, while in others the population fears surveillance. It is very difficult to overcome such uneasiness, but a persistent, well-designed educational policy can reduce it significantly. Clearly and unambiguously worded notices on the limits and rules of data use must be issued so that the public may understand sensitive data obtained for statistical purposes can never be accessed by other government authorities.

The fifth principle in “The Principles of Official Statistics” published by the United Nations states that any data source can be used for statistical purposes, whether statistical survey or administrative record. Statistical agencies may take into account the quality, timeliness, cost and the burden on respondents when choosing the appropriate data source. The sixth principle states that individual data obtained for statistical purposes by statistical organisations should be kept strictly confidential and only used for statistical purposes; administrative data can therefore only flow one way. The two principles comprise the main messages of the awareness-raising activities.

Different definitions

If the necessary administrative data of appropriate quality is already at the disposal of statistics, we may still face many problems, arising from the fact that these data were originally collected for non-statistical purposes, and because the two goals are often not the same. Even if a register is considered accurate, its data can differ to some extent from those that a statistical survey would collect. Differences may manifest in terms of:

- units,
- concepts,
- nomenclatures used,
- reference times between the two types of data.

Administrative data reflect the *de jure* situation, while the statistical data reflect the *de facto* situation; if the legislation, means of maintenance of the register, or lack of the public discipline permits then the results obtained from the two types of sources may differ significantly from each other. Thorough investigation should ascertain what various indicators are suitable for producing and what their limitations are before using administrative data for statistical purposes. Finally, it is

necessary to clarify what the statistical expectations are of the data, and if the concepts used by the administrative data source meet statistical requirements.

Limited availability of information

Register data are usually exact and robust because they are collected for administrative purposes. Sometimes a few pieces of information (e.g. education or occupation) are also included but regular maintenance of them is vague. Generally, there is no way to update them and the authority is not interested in whether these data have changed. If the authority does not pay the same attention to all data in a record the less important data will be less precise. That is why it is rarely worth having further data collected just for statistical purposes in an administrative register. Only the variables can be used that have been collected independently from the statistical or research concepts. Administrative data will never be eligible for investigating opinions, emotions or attitudes; this needs specified sample surveys.

Use of more sophisticated statistical methods

Often a single administrative register is not sufficient to achieve a statistical goal, but the integration of more will provide the necessary data set. In certain cases, data integration is the result of a complex statistical procedure rather than a simple connection of units or copying of sub-registers into one dataset.

Given the conceptual differences, complex estimation procedures play a greater role when using administrative data alone or integrated compared to statistical ones. Register-based statistics need thorough knowledge of the investigated event, process and its background as well as the characteristics of the primary data collection system, its shortcomings and reasons for this. Bad and unimprovable data must not be used.

Conclusions

Overall, it can be stated that in spite of all the problems associated with the use of administrative data, the advantages far outweigh the disadvantages. Most of the problems that arise with the use of administrative data can be solved, or at least reduced in similar ways – as in numerous areas of statistics –, with effective planning and organisation, a thorough knowledge of data sources, creative thinking and the active cooperation of the partners involved.

However, the specific problems concerning use of administrative data do not undermine their inclusion in statistical data production. Indicators produced either from administrative or statistical data sources give only estimates about a reality that we do not know completely. An indicator can be biased in different ways, depending on the analysed phenomenon and the nature of the data source included.

If there is a definite intention of the government to use secondary data as widely as it is possible for statistical purposes, most of the challenges can be overcome. However, there are some real obstacles like the lack of good will and the rejection of the additional workload by the top management of responsible institutions, as well as low priority and low quality of register data. The negative public opinion with regard to the use of personal data is also often referred to, but highly overestimated.