

nccr —————>
on the move

Philippe Wanner

**Collection and Analysis of Quantitative
Data in the Field of Migration.
Past Trends, Current Status and
Future Prospects**

Working Paper #27
October, 2020

nccr on the move

National Center of Competence in Research –
The Migration-Mobility Nexus

nccr-onthemove.ch

Philippe Wanner (University of Geneva)

Collection and Analysis of Quantitative Data in the Field of Migration. Past Trends, Current Status and Future Prospects

The nccr – on the move is the National Center of Competence in Research (NCCR) for migration and mobility studies. It aims to enhance the understanding of contemporary phenomena related to migration and mobility in Switzerland and beyond. Connecting disciplines, the NCCR brings together research from the social sciences, economics and law. Managed from the University of Neuchâtel, the network comprises seventeen research projects at eleven universities in Switzerland: The Universities of Basel, Fribourg, Geneva, Lausanne, Lucerne, Neuchâtel, Zurich, ETH Zurich, the Graduate Institute Geneva, the University of Applied Sciences and Arts of Western Switzerland, and the University of Applied Sciences and Arts of Northwestern Switzerland.

The Working Papers Series is an online platform for academic debates by members and cooperating partners of the nccr – on the move. The authors are responsible for the analyses and arguments, which do not necessarily reflect those of the nccr – on the move.

nccr – on the move, University of Neuchâtel, Rue Abram-Louis-Breguet 2, 2000 Neuchâtel, Switzerland
Contact for the Working Paper Series: info@nccr-onthemove.ch

Abstract

To be understood, any social phenomenon must be documented, for example, using statistical indicators or figures. In this article, we show, on the one hand, that the complexity of migratory flows requires new ways of collecting information, fortunately made possible by advances in information technology and the development of data collection methods. On the other hand, using an example of Swiss databases, we indicate the new analytical potential offered by the development of comprehensive data systems. Finally, we discuss the potential offered by private data, in particular social networks and provider data, while also pointing out certain limitations.

Keywords

Population Registers, Demography, Data Sources, Big Data, Longitudinal Approaches

Acknowledgments

This work was supported by the nccr – on the move, which is financed by the Swiss National Science Foundation. The author acknowledges the nccr – on the move network office for the support during the finalization of this working paper.

Contact

philippe.wanner@unige.ch

Institute of Demography and Socioeconomics, University of Geneva, Uni Mail 40, 1211 Geneva

Contents

1	Introduction	5
2	Early Beginnings: Measuring Migration Flows in the Early 20th Century	5
3	Traditional Data Sources and Their Use	10
3.1	The Development of Comprehensive Registers	10
3.2	Opportunities for Analysis Related to the Development of Data Linkages	11
3.3	An Example: The nccr – on the move Longitudinal Database	12
3.4	Adding the Generational Dimension to the Data	14
3.5	Intergenerational Mobility of Migrants	16
3.6	Future: Internationalization of Registry Information	18
4	The Development of Analyses Based on Private Data	19
5	Conclusion	23
	Bibliography	25

1 Introduction

Since Ravenstein's pioneering work on the measurement of migration at the end of the 19th Century, the collection of data on international migration flows but also more broadly on contemporary social phenomena, has been influenced by the availability of collection tools. It has also been influenced by prevailing migration patterns – in other words, the way people move – and by the needs of immigration administrations and policymakers.

As early as 1970, developments in information technology encouraged the increasingly systematic collection of social data, and these data were gradually stored on computer supports. For their part, surveys multiplied, mainly in the countries of the Northern Hemisphere, to measure the level of integration of migrant populations, for example, on the labor market, or to capture certain information related to migration, such as remittances. These surveys were facilitated by better fixed telephone coverage of populations, allowing data collection at lower costs. This movement towards more systematic data collection was further accelerated in the last two decades of the 20th century when statistical software became more accessible and data processing became less cumbersome, and different tools or methods of analysis were developed.

This development gradually led to a situation where increasingly more individual information (about people or events, also known as microdata) became available. In the case of Switzerland, the 1970 census is available in the form of individual records harmonized with the subsequent decennial censuses. Prior to this date, only historical tables, which were sometimes – but not always – digitized, were available. The transformation of the systems on which the data are stored has led to a multiplication of analyses. This has come at the right time, as international migration has diversified, become more complex and changed. In fact, the complexification of migration and the hardening of the debate on migration have sometimes been at the origin of the development of statistics (Piguet 2005). The data have been produced at the right time to respond to different considerations related to migratory transformations.

After an introduction to the genesis of migration data collection in the second chapter, this paper addresses the current situation in the third chapter, considering the potential for using administrative data. The fourth chapter presents innovative approaches. The conclusion synthesizes the data available to researchers and discusses future opportunities.

2 Early Beginnings: Measuring Migration Flows in the Early 20th Century

In order to better understand the current potential of the data, it is useful to refer to the collection methods available before the emergence of information technology and to provide a short, non-exhaustive history of migration measurement.

In simple terms, migration prior to the Second World War was rather important (Hatton and Williamson 1994), even if it was slowed down due to the lack of the means of communication we know in the present. This migration took place in a climate marked by the gradual desire of states to better control their foreign or migrant populations, a situation that led to the introduction of various

control tools, such as legislation, the strengthening of immigration offices (such as the Federal Bureau of Immigration in the United States when the Immigration Act came into force in 1924) and the compilation of lists of people entering the countries¹. During this period before computerization, most registers, lists, and even surveys were organized in a manual and decentralized manner.

At that time, migratory movements, and even more so intercontinental movements, were not frequent. The cost of migration was so high and the journey so arduous² that a departure only made sense if it was made for a long period of time, even if the migration was not always definitive (Hoerder 2010, Hatton and Williamson 1994). Motivations for leaving one's country were linked to precarious living situations, so the desire to return was weak. This situation of permanence still exists in some migratory flows, particularly asylum flows. What is new in the present but was rare in the past is the search for short-term life experiences through training, internships, or jobs abroad.

From the point of view of measuring entry flows, migration accounted for a large share of international mobility: arrivals in the United States, well documented by the registers established, for example, at Ellis Island, were mostly candidates for immigration, i.e., for a permanent or at least long-term stay (especially for Italians, who were more inclined to return home). The number of short-term visitors was much lower. In contrast, in 2018, there were 2 million German visitors to the United States, 100 times more than new immigrants³.

The conditions of access at the time were ideal for recording migration flows. Passenger lists were also drawn up in the United States and are now available on the internet. They provide basic information on the arrivals who have been accepted at the end of the immigration procedure (identity of the arrivals, including gender and nationality, date of arrival and name of the ship). While no information is available on the socio-demographic characteristics of immigrants, since those who were admitted to the territory were staying there on a long-term basis, their characteristics can be traced back to the decennial censuses.

From the point of view of the quantity of information on flows, the data from immigrant registers of the time are impressive. They cover international immigration in a way that is certainly as accurate, if not more so, than current data, which suffer from the uncertainty associated with illegal migration. They are also still available on the internet.

Once in the host country, immigrants were, for the most part, left to their own devices and gradually integrated into the host population. Censuses, and possibly some university surveys, provided the necessary basis for monitoring their integration or even assimilation and their living conditions. In countries of immigration, such as Canada and the United States, census questions on the origin, ethnicity, year of arrival in the country, naturalization and citizenship, as well as on linguistic integration, were numerous – almost more numerous than they are currently – as shown in Table 1,

¹ <https://www.familysearch.org/blog/en/immigration-act-1924-ellis-island/> (consulted 06.03. 2020)

² See, in particular, among the testimonies of immigrants faced with a ten-day journey in precarious conditions, that of E. Steiner (1906). On the trail of the immigrant. Fleming: New York. Quoted by <https://www.geo.fr/histoire/histoire-dellis-island-la-porte-dentree-pour-le-reve-americain-192840> (consulted 06.03.2020)

³ https://www.tourmag.com/Etats-Unis-176-million-de-visiteurs-francais-en-2018_a98809.html (consulted 06.03.2020)

which lists the information contained in the questionnaires. As an illustration, the 1990 Swiss census only asked about the place of birth, nationality, residence permit and domicile five years ago, “classical” variables that do not allow for an in-depth analysis of integration processes. However, researchers in the field of migration have used census data for the analysis of migration flows (see for instance Piguet 2005).

Table 1: Information present in censuses conducted in two North American countries, 1900-1991

	United States of America			Canada		
	1900	1950	1990	1901	1951	1991
Place of birth	X	X	X	X	X	X
Place of birth – father	X	X				
Place of birth – mother	X	X				
Year of immigration	X		X	X	X	X
Number of years in the US	X					
Naturalization	X	X		X (year)		X
Color		X		X		
Citizenship		X	X	X	X	X
Origin					X	X
Race			X			
Hispanic origin			X			
Language spoken at home			X		X	X
Aptitudes in English			X	X (E+FR)		X
Place of residence 1y ago		X				
Place of residence 5y ago			X			

Source: Own compilation

The U.S. censuses of the time are digitized and publicly available (e.g., www.censusrecords.com). It is possible to consult individual records (see Figure 1). Researchers can also download some data from the U.S. censuses from various sites, for example, from the IPUMS website⁴, where the 1850 and 1880 censuses are listed exhaustively, while a 5% and 1% sample of the 1900 and 1910 censuses, respectively, are also available⁵.

⁴ www.international.ipums.org (consulted 06.03.2020)

⁵ https://international.ipums.org/international-action/sample_details#us (consulted 24.02.2020)

Figure 2: Example of a record that can be viewed on the internet, the 1900 Census

TWELFTH CENSUS OF THE UNITED STATES. 97 A

SCHEDULE No. 1.—POPULATION.

State New York County Brooklyn Supervisor's District No. 10 Sheet No. 3
 Enumeration District No. 44

Township or other division of county Rampart Town Name of Institution _____
 Name of incorporated city, town, or village, within the above-named division _____ Ward of city, _____
 Enumerated by me on the 11th day of June, 1900, Joseph H. King Enumerator.

LOCATION	NAME	RELATION	PERSONAL DESCRIPTION	NATIVITY			CITIZENSHIP	OCCUPATION, TRADE, OR PROFESSION	EDUCATION	SPEED OF REED					
				Place of birth of this person.	Place of birth of father of this person.	Place of birth of mother of this person.									
SEX	AGE	MARRIAGE	COLOR	SEX	AGE	MARRIAGE	NATIVITY	OCCUPATION, TRADE, OR PROFESSION	EDUCATION	SPEED OF REED					
											Male	Female	Male	Female	Male
1	44	Springstein, Estelle	Wife	W	20	Mar	1879	New York	New York	New York	11	Teacher	12	2	1
2		" " Eligam	Wife	W	15	Mar	1885	New York	New York	New York	11	Teacher	12	2	1
3		" " Malinda	Wife	W	12	Mar	1888	New York	New York	New York	11	Teacher	12	2	1
4		" " Catherine	Daughter	W	7	Mar	1893	New York	New York	New York	11	Teacher	12	2	1
5		" " Aaron	Daughter	W	5	Mar	1895	New York	New York	New York	11	Teacher	12	2	1
6		" " Abraham	Daughter	W	3	Mar	1897	New York	New York	New York	11	Teacher	12	2	1
7	45	Wilgram, Frederick	Head	M	20	Mar	1880	New York	New York	New York	11	Teacher	12	2	1
8		" " Mary Jane S.	Wife	W	17	Mar	1883	New York	New York	New York	11	Teacher	12	2	1
9		" " Selma	Daughter	W	12	Mar	1888	New York	New York	New York	11	Teacher	12	2	1
10		" " Sarah	Daughter	W	10	Mar	1890	New York	New York	New York	11	Teacher	12	2	1
11		" " Maggie	Daughter	W	7	Mar	1893	New York	New York	New York	11	Teacher	12	2	1
12		" " Raymond	Daughter	W	5	Mar	1895	New York	New York	New York	11	Teacher	12	2	1
13		" " Ethel	Daughter	W	3	Mar	1897	New York	New York	New York	11	Teacher	12	2	1
14		" " John	Daughter	W	2	Mar	1898	New York	New York	New York	11	Teacher	12	2	1
15	46	McDonnell, William	Head	M	20	Mar	1880	New York	New York	New York	11	Teacher	12	2	1
16		" " Mary	Wife	W	18	Mar	1882	New York	New York	New York	11	Teacher	12	2	1
17		" " John G.	Daughter	W	15	Mar	1885	New York	New York	New York	11	Teacher	12	2	1
18		" " Edward	Daughter	W	12	Mar	1888	New York	New York	New York	11	Teacher	12	2	1
19	47	Keeler, James	Head	M	20	Mar	1880	New York	New York	New York	11	Teacher	12	2	1
20		" " Thomas	Wife	W	18	Mar	1882	New York	New York	New York	11	Teacher	12	2	1
21		" " Tom	Daughter	W	15	Mar	1885	New York	New York	New York	11	Teacher	12	2	1
22	48	See, Benjamin	Head	M	20	Mar	1880	New York	New York	New York	11	Teacher	12	2	1
23		" " Sarah C.	Wife	W	18	Mar	1882	New York	New York	New York	11	Teacher	12	2	1
24		" " Elizabeth	Daughter	W	15	Mar	1885	New York	New York	New York	11	Teacher	12	2	1
25	49	Donahue, Elizabeth	Head	W	20	Mar	1880	New York	New York	New York	11	Teacher	12	2	1
26		" " Bernard	Wife	W	18	Mar	1882	New York	New York	New York	11	Teacher	12	2	1
27	50	Richards, Charles	Head	M	20	Mar	1880	New York	New York	New York	11	Teacher	12	2	1
28		" " Annie	Wife	W	18	Mar	1882	New York	New York	New York	11	Teacher	12	2	1
29		" " Robert	Daughter	W	15	Mar	1885	New York	New York	New York	11	Teacher	12	2	1
30		" " Alice	Daughter	W	12	Mar	1888	New York	New York	New York	11	Teacher	12	2	1
31	51	Campbell, James	Head	M	20	Mar	1880	New York	New York	New York	11	Teacher	12	2	1
32		" " Mary E.	Wife	W	18	Mar	1882	New York	New York	New York	11	Teacher	12	2	1
33		" " Mary M.	Daughter	W	15	Mar	1885	New York	New York	New York	11	Teacher	12	2	1
34		" " Florence S.	Daughter	W	12	Mar	1888	New York	New York	New York	11	Teacher	12	2	1
35		" " Nancy B.	Daughter	W	10	Mar	1890	New York	New York	New York	11	Teacher	12	2	1
36		" " William B. W.	Daughter	W	8	Mar	1892	New York	New York	New York	11	Teacher	12	2	1
37	52	Jamman, William	Head	M	20	Mar	1880	New York	New York	New York	11	Teacher	12	2	1
38		" " Emma	Wife	W	18	Mar	1882	New York	New York	New York	11	Teacher	12	2	1
39		" " Annie	Daughter	W	15	Mar	1885	New York	New York	New York	11	Teacher	12	2	1
40		" " William	Daughter	W	12	Mar	1888	New York	New York	New York	11	Teacher	12	2	1
41	53	Ward, William	Head	M	20	Mar	1880	New York	New York	New York	11	Teacher	12	2	1
42		" " William	Wife	W	18	Mar	1882	New York	New York	New York	11	Teacher	12	2	1
43		" " Annie	Daughter	W	15	Mar	1885	New York	New York	New York	11	Teacher	12	2	1
44		" " John S.	Daughter	W	12	Mar	1888	New York	New York	New York	11	Teacher	12	2	1
45		" " David M.	Daughter	W	10	Mar	1890	New York	New York	New York	11	Teacher	12	2	1
46	54	Ward, William	Head	M	20	Mar	1880	New York	New York	New York	11	Teacher	12	2	1
47		" " William	Wife	W	18	Mar	1882	New York	New York	New York	11	Teacher	12	2	1
48		" " Annie	Daughter	W	15	Mar	1885	New York	New York	New York	11	Teacher	12	2	1
49		" " John S.	Daughter	W	12	Mar	1888	New York	New York	New York	11	Teacher	12	2	1
50		" " David M.	Daughter	W	10	Mar	1890	New York	New York	New York	11	Teacher	12	2	1

Source: 12th Census of the United States (https://en.wikipedia.org/wiki/1900_United_States_Census)

Currently, the situation is not the same. The very nature of international migration has changed profoundly. Situations where people live between two or more countries are multiplying, with border work, detached workers or migrants going to sunny destinations in winter. The increasing mobility of the world's population, for its part, makes it difficult to distinguish at borders between migrants (who may end up staying longer than expected in a territory) and international visitors (who may leave the country quickly). The multiplication of arrival points in some countries and the increase in the number of visitors and tourists complicate the registration of people at points of entry. Certainly, computer tools make it possible to have statistics of visitors and migrants in the customs zones of airports or ports for countries that do not have customs agreements with their neighbors, as is the case, for example, in the United States. On the other hand, for countries that have opened their borders, for example, countries in the Schengen area, checks at crossing points are currently impossible.

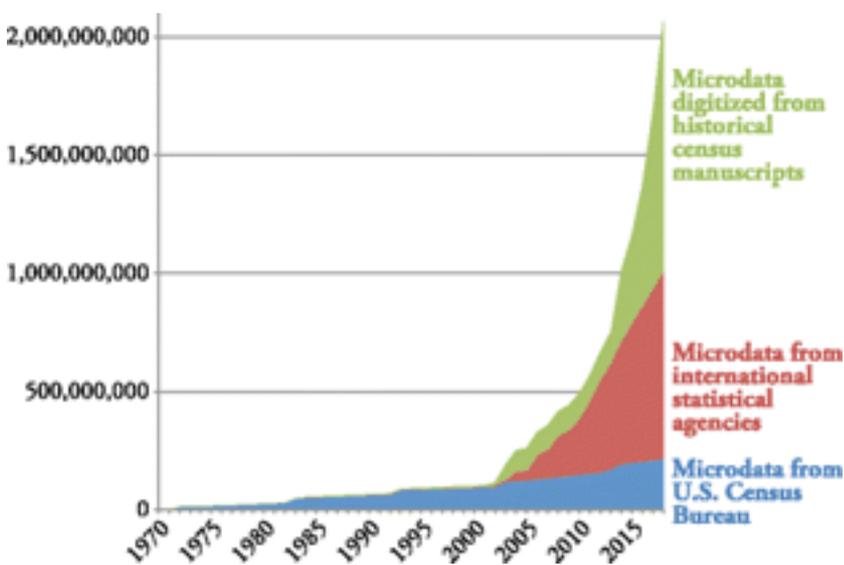
Given this shift, it is no longer “entry” that is measured, but “status”. The resident population includes persons with permits that give them certain rights, such as the right to work or to stay in the country temporarily or permanently. In Switzerland, holders of annual, permanent, international civil servant permits, as well as under certain conditions holders of short-term permits, asylum seekers and temporarily admitted persons, are included in the resident population. Undocumented migrants, on the other hand, are not included. They are sometimes estimated with varying degrees of precision. In Switzerland, Morlok et al. (2015) estimate between 58,000 and 105,000. These undocumented persons living in Switzerland are not part of the resident population and are not counted in any statistics.

Residence status allows a distinction to be made between migrants and visitors. The former corresponds to a definition proposed by the United Nations (“International migrants are defined as any person who changes his or her country of habitual residence”). For the United Nations, a person is a resident if he or she has lived in the country for most of the past year or intends to return to live there for 12 months (United Nations 1998). This definition is difficult to apply for alternative forms of mobility (nomadism, people sharing their lives between two countries).

Data on migrants, and alternatively on persons belonging to ethnic minorities or with foreign nationality, have multiplied, linked to the development of censuses and other surveys. Currently, the population is registered or enumerated many times, and microdata are multiplied. Moreover, at the initiative of statistical institutes and universities, such data are available to the research community. Ruggles (2014) estimates the number of digitized records (microdata) available in open access in 2018 (see Figure 2) at 2 billion and in limited access at 2.4 billion. The amount of these microdata has been increasing rapidly since the beginning of 2000.

In the following chapter, we focus on these “traditional” data, in the sense that they come from statistical offices, and analyze how they can meet the needs of migration research.

Figure 2: Number of available microdata for social scientists



Source: Ruggles 2014

3 Traditional Data Sources and Their Use

3.1 The Development of Comprehensive Registers

The abovementioned IT development has increased the number of administrative registers managed as individual records. Currently, every individual residing in an industrialized country, from a very young age, appears in many public registers (register of inhabitants, tax register, annuity register, civil status register, etc.) or private registers (insurance registers, bank registers, etc.). These registrations are essential for the management of many societal areas, such as national security, pensions or taxes or simply to determine who is domiciled in a commune or country.

The computerization of these registers and the speed of today's computers leads to a situation where the potential for using these registers for statistical analysis is high, provided that access to them is guaranteed. Such access is subject to privacy protection criteria, particularly for sensitive data. Statistical offices establish strict criteria in this respect. The solutions proposed by these offices are the removal of identifying data, the anonymization of individual data or the grouping of variables that are too precise and would allow identification (such as nationality). These solutions are essential for balancing privacy protection with society's interest in detailed analysis.

In Switzerland, official statistics have gradually included computerized registers to feed economic and social statistics. The first register, the civil status register, was introduced at the end of the 19th century. The statistics derived from this register make it possible to monitor phenomena such as births, marriages, divorces or deaths and to derive information such as life expectancy or the average number of children per woman. While in the past, these data were compiled in tabular form, since 1969, this statistic has been available in computerized form, making it easier to calculate indicators.

For a long time, the registers of inhabitants have been decentralized. Until 2010, each Swiss municipality kept a municipal register, the form and content of which varied⁶. At that time, the Federal Statistical Office did not have a population database for Switzerland. The harmonization of the municipal registers was the basis for the population statistics (STATPOP), which provides an individual record for each person living in Switzerland. Prior to 2010, registers provided researchers and the administration with individual records on the population of foreign nationality. The register of foreigners has existed since 1981 in computerized form and includes socio-demographic and residence status information for monitoring the foreign population. Until 2002, data on the professional activity were also included for all foreigners, before being suppressed for nationals of EU/EFTA member states.

A multitude of other registers are used by statisticians, sometimes in addition to the surveys that are carried out. For example, the register of individual accounts of the Central Compensation Fund for AHV (Old Age and Survivors' Insurance) gathers information on salaries. This centralized information, available in the form of individual records since 1982, provides the income of every worker in Switzerland. Thanks to this source, official surveys in Switzerland no longer need to ask questions about this sometimes personal dimension, which is often difficult to estimate precisely.

⁶ In some cases, such as Geneva, a cantonal register replaced the communal ones.

3.2 Opportunities for Analysis Related to the Development of Data Linkages

At the end of the first decade of 2010, Switzerland introduced a 13-digit personal identification number (social security number, NAVS13) that is both anonymous and universal. This number was implemented in public administrative registers but also in some private registers (e.g., the registers of the various health insurances). The introduction of this number, which is equivalent to the US social security number, enabled Switzerland to catch up with other countries in terms of its statistical infrastructure and to provide information approaching that of some Nordic countries, at least concerning economic and social statistics.

The creation of the harmonized and centralized register of inhabitants in 2010 and the implementation of this personal identification number were indeed two essential steps to develop social statistics and, for researchers, to extend certain analyses. Data matching or linkage means the pooling of different databases to assemble variables and information from several sources for the same individual.

Data matching dates back to the immediate post-war period (Dunn 1946) and increased with the formulation of probabilistic methods that allowed records from different sources to be pooled together, referring to the same individual (Newcombe et al. 1959, Fellegi and Sunter 1969). These methods aim to link data for the same individual in the absence of a personal identification number and using data common to two registers (such as surname or first name, date of birth). With the availability of a unique identifier, this approach is no longer necessary, and the basis for uncertainty and error is reduced.

Matching can be done within a single register or statistic (e.g., matching birth and death data in a civil register, matching information from different years). Matching may also involve different data sources. In both cases, data matching improves the understanding of social phenomena by adding information and/or creating longitudinal follow-ups of individuals by putting data from different years altogether. Data matching generally crosses disciplinary boundaries. For example, information on working life (unemployment episodes, changes in activity, etc.), on the family (births, marriages, etc.), on health (incidence of illness, hospital consultations, etc.), or on mobility behavior (changes in residence, etc.) can be linked. They, therefore, offer the possibility of describing the individual's life in a much more global context than can be done through the use of autonomous registers.

There are a few barriers to the use of these matched data for research. On the one hand, access to data and permission to perform matching is limited for obvious reasons of privacy protection. Different strategies are employed. In some countries, researchers do not directly access the data but may send analysis programs that a dedicated center runs, with the results then being produced in the form of tables or models. In other countries, matching takes place within a government agency before being passed on to researchers. In Switzerland, this is now the practice: researchers are invited to carry out their work on an offline computer at the Federal Statistical Office. Prior to the introduction of NAVS13, matching was carried out using probabilistic approaches outside the

Office. This was the case, for example, with the Swiss National Cohort⁷ or other projects on migration (Wanner 2012).

On the other hand, the data available in the registers are mainly those needed by administrations for management purposes. They are not always adequate for the specific needs of researchers. In the case of the foreign population, no register provides exhaustive information on the language skills of persons with a migration background, their employer, their activity rate (number of hours worked), their ethnicity, their social participation, or the failures encountered in the naturalization procedure. The lack of such information leads to frustration for researchers when carrying out analyses and interpreting the results and encourages them to consider other collections (e.g., surveys) to complete their analyses. It should be noted that surveys can also be matched to registers through the NAVS13.

In addition, some registers have varying levels of quality, especially for secondary variables, which are not necessarily essential for management purposes. For example, administrative tax registers have difficulties in providing reliable information on occupation, as it is not crucial for taxation. Inhabitants' registers show significant proportions of missing values for information such as the date of arrival in Switzerland or the country of origin or destination of quality migrants. These quality problems sometimes hamper the potential for analysis.

3.3 An Example: The nccr – on the move Longitudinal Database

In line with these developments, in 2015, we created a longitudinal database integrating different registers and surveys (Steiner and Wanner 2015). The idea behind this development was to provide migration researchers with access to exhaustive, detailed and original data and longitudinal indicators. The registers in this database have been partly linked with the NAVS13 number, partly through matching using variables common to the different files (date of birth, nationality). This database has been regularly updated since 2015 and extended by integrating the 2000 Census (by probabilistic methods since the NAVS13 is not available) and vital statistics data. The integration of the census is based on the dates of birth and nationality of persons, especially couples, observed in 2000 and found in 2010 in the population statistics or in 2000 in the register of foreigners.

The volume of data included in this system is presented in Table 2, there are nearly 300 million records with a number of variables ranging from 8 to over 100. Some registries with a NAVS13 number had to be cleaned up to ensure that passage keys (NAVS13) were correctly imputed to individuals and that only one key was assigned to an individual. This represented the bulk of the matching work, which was carried out on the premises of the Federal Statistical Office.

The individual records can be combined to reconstruct the partial life trajectory (between 1997 and 2018) of foreigners living in Switzerland, as well as of Swiss citizens (between 2010 and 2017). This database has been used for the preparation of longitudinal migration indicators⁸ and for specific, more targeted studies, which are currently being published (see for instance Lombard and Zufferey 2019, Zufferey and Wanner 2020, Zufferey et al. 2020).

⁷ <https://www.swissnationalcohort.ch> (consulted 06.03.2020)

⁸ <https://indicators.nccr-onthemove.ch/?lang=fr> (consulted 06.03.2020)

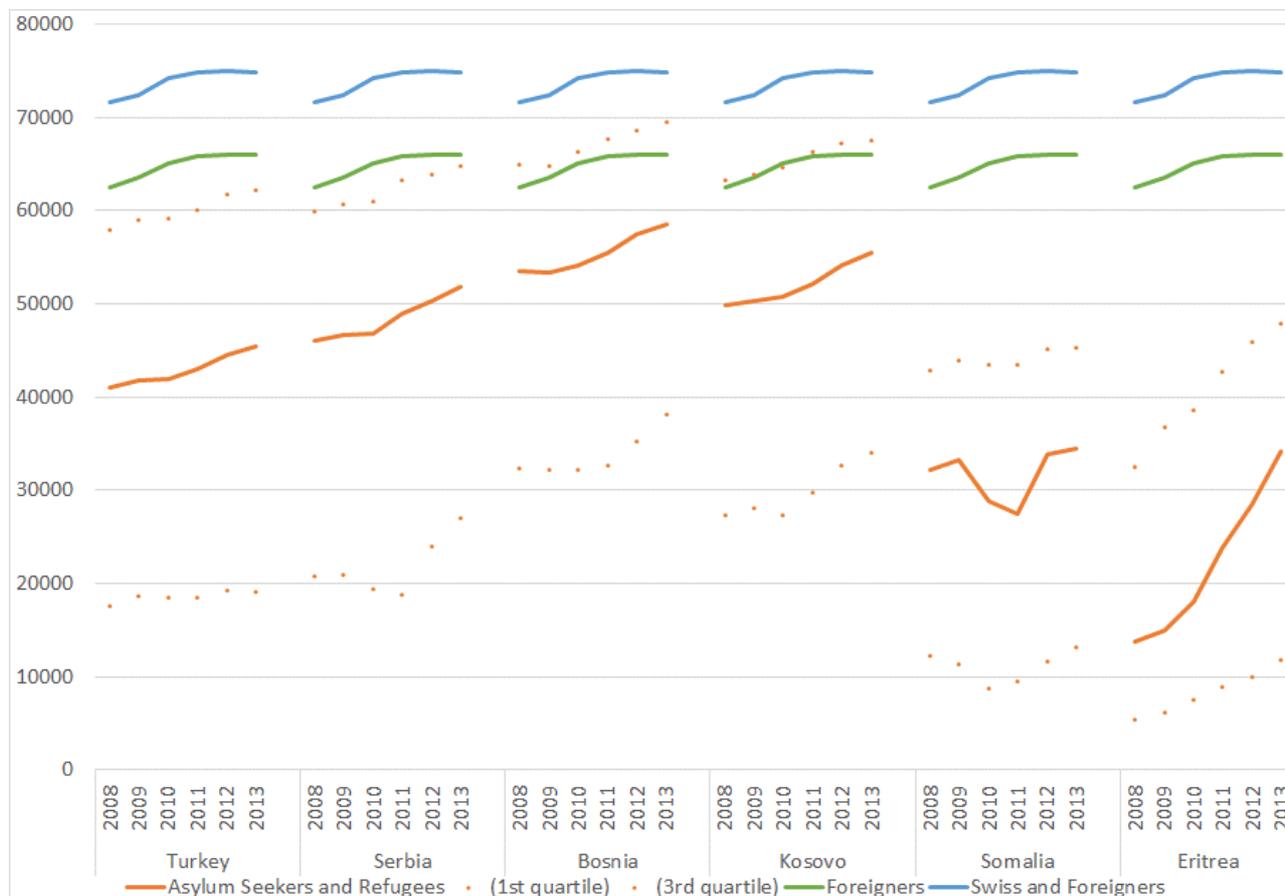
Table 2: Records, number of variables and number of records

Title	Type	Number of variables	Number of years	Records	Yearly average
Population Statistics	Status 31.12	59	9	75941000	8438000
	Movements	37	7	17157000	2451000
Structural Survey	Persons	103	8	2268000	284000
	Households	69	8	1929000	241000
	Household members	75	8	5049000	631000
Social Insurance Registers	Individual accounts	11	18	130695000	7261000
	Pensioners	22	2	4384000	2192000
Aliens Register	Status 31.12	28	14	24346000	1739000
	Movements	37	13	10861000	835000
Asylum Seekers Register	Status 31.12	25	14	2348000	168000
	Movements	36	13	1417000	109000
Census 2000		59	1	7288000	7288000
Vital Statistics	Adoptions	26	8	4000	1000
	Births	49	8	762000	95000
	Recognitions	23	8	169000	21000
	Weddings	68	8	369000	46000
	Divorces	64	8	147000	18000
	Partnerships	114	8	6000	1000
	Partnership breakdowns	68	8	1000	0
Unemployed Register		8	8	11020000	1378000
Total				296161000	33197000

Source: Own computation

Figure 3 from Wanner (2017) illustrates what data linkages allow. It has been constructed from different sources: population statistics, which provide a list of people living in Switzerland, their sex and age; the central register of foreigners, which gives the nationality of non-Swiss people living in the country before 2010; the register of asylum seekers, which makes it possible to verify the asylum status of the person; and the register of individual first-pillar accounts, which gives the professional income of each person paying contributions in Switzerland. Each of these registers provides specific information (number and characteristics of persons in the asylum field, median of salaries paid). However, combining them makes it possible to precisely describe phenomena that a single register does not cover: thus, as an example, the figure shows the evolution of income between 2008 and 2013 for different groups defined according to the reason for arrival (asylum or not), nationality and residence status for a group of persons defined according to labor market status.

Figure 3: First and third quartile median income of men participating in the labor market by nationality and year, 2008-2013



Source: Wanner 2017, SFSO and CdC Data

3.4 Adding the Generational Dimension to the Data

The potential for exploiting these linkages of registries can go further, as the following example will demonstrate. We use another database that we have developed in connection with the Federal Social Insurance Office, entitled WiSiEr (Wanner 2019). This database is even more ambitious: on the one hand, it includes a few other statistics from registers (social assistance statistics, housing, and building statistics) and many other variables not taken into account in the nccr – on the move database; on the other hand, it includes information from cantonal tax registers (11 cantons), which had to be taken over and harmonized as part of a rather voluminous work. Finally, it goes back to 1982 for data on assessed income, which increases the observation period.

A series of variables were included that refer to conjugal and parent-child relationships, taken from vital statistics. This is derived from the central database of civil status data INFOSTAR, introduced in 2001⁹. This database is fed by civil registry officers, who are in charge of recording the various civil status events occurring in Switzerland (birth, marriage, partnership, death, etc.). Events that concern the population residing in Switzerland but that occur abroad are also reported. This database has replaced the registers kept on paper and has made it possible to simplify the sometimes tedious procedures and, it seems, to avoid errors or duplicate registrations. INFOSTAR, like all

⁹ <https://www.bj.admin.ch/bj/fr/home/aktuell/news/2001/61.html> (consulted 18.03.2020)

administrative registers, includes the social security number of the persons concerned by civil status events. However, unlike most of the other registers, it links the different persons together by filiation or marital link. Thus, when a marriage is registered, INFOSTAR has the NAVS13 of both spouses. When a birth occurs, it has the number for the child and both parents. In the event of separation, the database can document the separation. Information on conjugal or filial relationships, including persons in partnership, is therefore available. These data have been communicated to the Swiss Federal Statistical Office, forwarded to population statistics, and integrated since 2010.

Analyses that we have had the opportunity to conduct from these data show certain limitations, which do not in any way affect the potential for using these data. These limitations are mainly because the civil register takes into account events occurring in Switzerland or among the Swiss population and therefore cannot take into account events occurring abroad before the migration. Thus, a couple married in a foreign country and arriving in Switzerland will not be formally considered as such until the information on their marriage can be traced back to the Swiss civil register (e.g., at the time of the birth of a child). There is therefore a fairly large amount of missing data (Table 3). While the information on the spouse is available for 99.3% of married persons of Swiss nationality, this proportion is only 41% for nationals of the European Union and EFTA, 56% for other Europeans, and 60% for those from other continents.

Table 3: Proportions of persons for whom information is available in STATPOP 2016 on spouses of married persons and parents (total population)

	Married Persons		Total Population		
	% with identified spouse	Sample	Father identified	Mother identified	Sample
Citizenship					
Switzerland	99.3	2623167	58.5	67.7	6318404
EU/EFTA	40.8	607653	9.9	10.3	1441232
Other Europe	55.7	217470	18.5	19.1	377692
Rest of the world	60.2	158550	10.3	11.9	359985
Total	85.1	3606840	46.4	53.4	8497313

Source: WiSiEr Statistics

The relatively low proportion of persons for whom information is available for EU/EFTA nationals is explained by the fact that many of them arrived after celebrating their marriage in a foreign country, and among a part of the married migrants, one spouse has remained in the original country. The length of stay in Switzerland is a factor since the rate of identified spouses increases with the length of time spent in the country. For the population as a whole, the proportion of persons for whom we have identification on the parent is rather low, approximately 50%, higher for Swiss than for foreigners and higher for mothers than for fathers. These figures must of course consider that identification of the father and/or mother is only possible when the parents are still alive and present in Switzerland¹⁰. They must be weighted by the analysis by age group (presented in Table 4).

¹⁰ More precisely, the NAVS13 number should have been attributed to the parent. It is possible that the parent is identified and the information still exists even after his or her emigration or death.

Among children under 10 years of age, almost all Swiss children have information on the father and mother. More than four out of five children of EU/EFTA nationality, more than 90% of those from the rest of Europe and approximately 70% of those from the rest of the world have information on their parents. For Swiss young adults (20-29 years), information is available for approximately 90%. The rates are lower for foreigners, partly because foreigners often arrive in Switzerland without their parents. On the other hand, information is often available for the second generation.

Table 4: Proportions of people for whom information is available in STATPOP 2016 on parents, according to age

Age	Switzerland		UE/EFTA		Other Europe		Other World	
	Father	Mother	Father	Mother	Father	Mother	Father	Mother
0-4	98.9	99.9	83.9	85.9	94.0	96.2	66.5	76.7
5-9	99.5	100.0	64.0	65.1	91.3	92.6	52.3	56.4
10-14	98.1	99.3	35.6	36.9	58.4	60.2	31.2	34.4
15-19	94.7	97.0	12.4	13.6	19.2	20.8	6.5	8.2
20-24	92.0	95.0	3.7	4.2	5.8	6.5	1.8	3.0
25-29	89.2	93.0	1.3	1.4	2.4	2.8	0.8	1.9

Source: WiSiEr Statistics

The information on filial and conjugal relationships (which also includes registered partnerships) has not yet been used or validated by the Swiss Federal Statistical Office. To our knowledge, no analysis has yet been carried out. In the following section, we show the potential of these data using a small example.

3.5 Intergenerational Mobility of Migrants

We are interested in the intergenerational mobility of second-generation immigrants (born in Switzerland), defined as those aged 25 to 34 in 2010. This age group refers to 597,500 people, and we have the paternal link for 85% of them and the maternal link for 89%. We have documentation of the level of education completed for 128,000 “child-father” couples. This information comes from the structural survey, which is matched with the other registers and is large-scale but not exhaustive. Table 5 compares the level of education of the father and his child. The green boxes refer to a positive intergenerational educational evolution, and the pink boxes refer to a negative evolution. It can be observed that when the father is educated at the secondary I level, in almost 92% of cases, the child does better. The social ascent is slightly lower for nationals from the rest of Europe than for other groups. When the father is tertiary educated, an intergenerational decrease in educational attainment is observed for 35% of the children. This decrease is less significant for young people from the rest of the world. When the father is of secondary II education, there is an ascent primarily for nationals from the rest of the world.

These data confirm that, on the one hand, the Swiss school system allows for certain social mobility, in particular, upward mobility. On the other hand, this mobility is more difficult for Balkan nationals, who make up the majority of the rest of Europe. This diagnosis deserves to be analyzed in more detail by specialists in the educational system.

Table 5: Comparison of father's and child's level of education by nationality (children aged 25-34 in 2010)

Children / Citizenship		Father		
		Secondary I	Secondary II	Tertiary
Swiss	Secondary I	8.0	3.8	2.0
	Secondary II	64.0	55.1	33.1
	Tertiary	28.0	41.1	64.9
EU/EFTA	Secondary I	7.9	4.7	2.7
	Secondary II	56.8	50.9	29.6
	Tertiary	35.3	44.3	67.6
Other Europe	Secondary I	11.4	5.9	3.3
	Secondary II	62.3	57.5	26.6
	Tertiary	26.3	36.6	70.1
Other World	Secondary I	8.4	5.0	3.0
	Secondary II	46.9	42.4	24.5
	Tertiary	44.7	52.6	72.5
Total	Secondary I	8.2	3.9	2.1
	Secondary II	62.0	54.7	32.5
	Tertiary	29.8	41.4	65.4

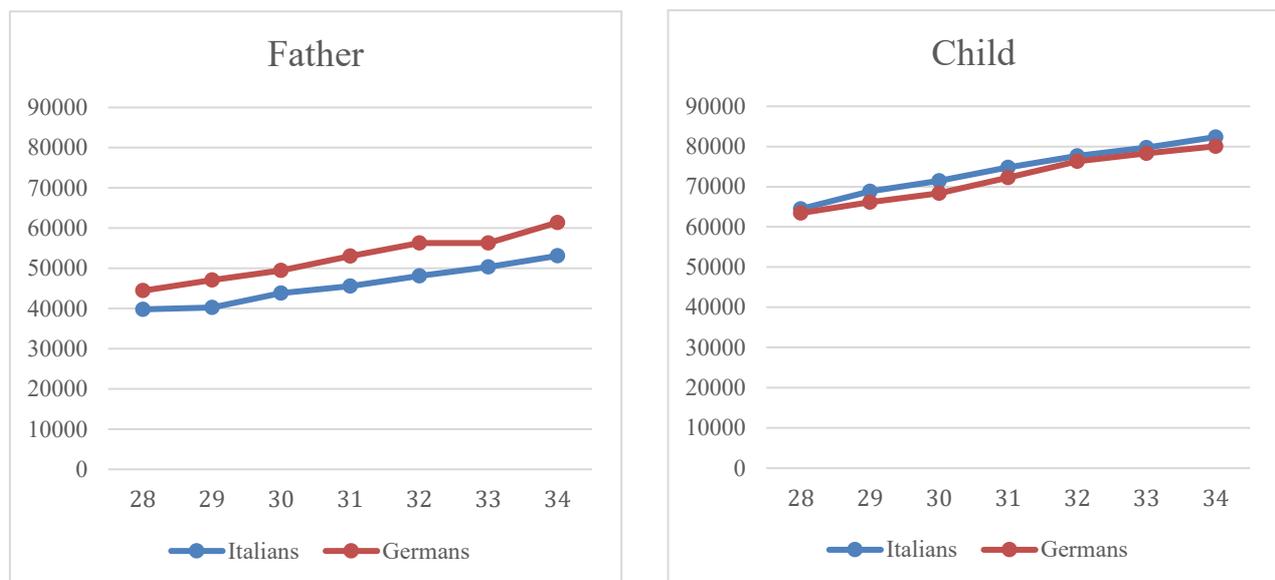
Source: WiSiEr Statistics

A second example based on family ties refers to the income of the parents and children around the age of 30. Figure 4 is based on the relationship between son/father couples followed between the ages of 28 and 34. Median incomes were analyzed for each member of these couples at a distance of approximately 30 years. Two origins, defined by the father's place of birth, are represented in the graph: Germans and Italians. In the first generation (fathers), German immigrants, when they were 28-34 years old (in the 1980s and 1990s), showed a clear advantage in median income compared to Italian immigrants of the same age. The right-hand side of the graph indicates that this advantage has now been erased, and second-generation Italians now have incomes equal to or even slightly higher than those of Germans. This second example seems to indicate that school has an integrative function: every child, whatever his or her national and/or social origin, has the same chances of success in the labor market (expressed by income at the beginning of his or her career). The results relate to Italians and Germans only and should, of course, be confirmed for other groups.

We do not wish to go further into the theme of intergenerational evolution, as these examples are intended to illustrate some of the possibilities offered by family ties. Children, parents and spouses can be easily identified as long as they are present and registered in the civil register. By extension, grandparents, brothers and sisters, step-parents, uncles and aunts, nephews and nieces, etc., can also be identified from filial and conjugal ties. The available data, therefore, make it possible to reconstitute families, even when the members of these families live in another household in Switzerland. Persons living abroad cannot be identified, which somewhat limits the analytical potential for migrant populations compared to the Swiss population.

Currently, family ties require that people are alive to be established. However, the data are updated annually. Thus, it is technically possible to analyze the situation at the same age of two generations, as we did in the previous example. However, in our example, we benefit from long-term data (available since 1982). For other phenomena that can be analyzed, such as family behavior, causes of death, employment status, and use of social assistance, this temporal coverage is no more than 10 years (the data generally go back to 2010). As time passes and this information is updated annually, however, the temporal coverage will increase.

Figure 4: Trends in median incomes of parents and children, aged 28-34, for the first and second generations of Italian and German migrants



Source: WiSiEr Statistics

All in all, these data are promising, provided one accepts the limitations in terms of the variables available and the quality of some of the variables. In the future, further extensions are possible in different disciplinary fields (e.g. health). In the field of migration, another development that is currently difficult to formalize may occur one day: the pooling of international data.

3.6 Future: Internationalization of Registry Information

The data and analyses presented here relate to Switzerland; however, some countries have already had this type of data for several decades; other countries have the same objectives of collecting and highlighting information and have or will have comparable data in the future. The comparison of migration indicators at the international level is therefore possible, subject to the different specificities of each country. It already provides a wealth of information that is essential for the formulation of migration policies.

However, register-based statistics could go even further. Researchers on migration, approaching issues of interest to them from a longitudinal perspective, have only data covering the period of the migrant's presence in the host country. Thus, in the case of Switzerland, it is possible to document the migrant's trajectory in Switzerland from his or her arrival to his or her eventual departure. If matches could be made between the registers of two countries, it would be possible to compare the

individual path before and after the migratory episode. Such comparisons would be useful for better documenting the factors behind migration and for assessing the benefits of migration for the migrant.

Technically, the matching of two registers from different countries can only be done by manual methods because there is no shared identifier between different countries. The current limitations refer to legal contexts, as the majority of countries do not allow the export of their data abroad. For this reason, access to foreign data for matching seems limited. However, in practice, it is quite possible to identify emigrants and immigrants by considering two countries (e.g., Canada and the United States) and to link them by individual characteristics. In the absence of access to records for each person, it is also possible to identify groups of migrants defined as precisely as possible (not only by nationality but also by age group, gender or education level) in both countries. Then, it becomes possible to analyze changes for these groups in, for example, family behavior or socio-economic status.

4 The Development of Analyses Based on Private Data

Alongside linked registers, a second promising avenue is that of massive private data, most often referred to as big data¹¹. These data have offered interesting breakthroughs in the field of public health and more particularly led to the development of digital epidemiology, which follows, for example, influenza or cholera epidemics, with an exponential increase in the number of articles using this support (Pastorino et al. 2019, Muin et al. 2014) and offering many advantages but also some disadvantages (Lazer et al. 2014). Currently, big data are considered by statistical offices a possible source of information (United Nations 2018).

In the social sciences, attempts at applications are multiplying, with for the moment a limited number of scientific publications in the major peer-reviewed journals. The majority of studies have been presented in specialized symposia. For the time being, applications based on big data should be classified more in the category of methodological trials and have not yet reached the stage of complementing (or even substituting) traditional data to provide answers to specific research questions, although the example that will be presented below shows the great potential of these data for decision-making.

The question of the possible substitution of traditional data by big data primarily concerns the countries of both the North and the South. In some areas, such as transport planning and, more specifically, traffic measurement, data from Google Traffic are gradually and advantageously replacing traditional collection methods based on surveys or vehicle counting loops. Although the vision is less accurate, the data are more comprehensive and less costly. In countries of the South or countries with weak statistical infrastructure, big data provides information that would not otherwise be accessible to different scientific disciplines. In this sense, private data represent enormous potential.

¹¹ The term big data is defined, e.g., by Bohon 2018 or Horrigan 2015, as comprehensive data whose creation was not motivated by research objectives. Thus, big data include the registers mentioned above (microdata). In this text, following common usage, we assimilate them to data produced by private companies, often based on information generated by the population itself through social networks or internet connections.

The very use of this data, to join Bohon's (2018) reflection, raises a fundamental question. Is this data a new tool for researchers active in different disciplines, e.g., public health, migration, or sociology? Does this data constitute a new scientific discipline? Specifically, for migration researchers, can this data be accessed and used for analysis? Or, on the contrary, is it necessary to develop collaborations with engineers and statisticians specializing in big data? To date, we realize that data accessibility is relative and that in most cases, a multidisciplinary approach is necessary. The new field of digital demography is still in its infancy but could develop.

Few related studies are published in major journals. One exception refers to the study by Zagheni et al. (2017), which analyzed the ability to measure the number of migrants in different countries of the world using the advertising platform of Facebook. This study was published by a leading journal, attesting to the interest of publishers in this type of methodological development. The authors extracted from this platform estimates of the audience size, as made available by Facebook when defining advertising campaigns. By indicating the expatriate population as a targeting criterion for a possible campaign, the authors obtained the number of corresponding Facebook accounts for different communities. Regression models comparing the data published by the states with the data from this platform show a relatively good ability to predict the national workforce.

The study by Ahas et al. (2017) focuses on transnationalism, defined as living in two different countries. The authors consider transnational migrants as people who have made between 5 and 52 trips to a foreign country, with a total stay of 92 to 273 days (i.e., between 25% and 75% of the annual number of days). They contrast them with commuters (who make more than 52 transnational trips), tourists (up to 4 trips and less than 182 days), and foreign workers (more than 182 days). Based on this typology and using roaming data from two access providers, the authors are able to quantify the number of people in each category, as well as the countries of origin or destination. As their data fill a gap, it is not possible for the authors to validate them by comparison with other sources.

These two studies show the potential of private data but also illustrate the difficulties inherent in using them. The authors are aware of some of these difficulties and the limitations that arise from them.

Access to data is sometimes problematic, as they are owned by companies that are not always willing to open their doors to researchers. For example, Zagheni et al. (2017) had to go through a public Facebook platform to establish their modeling and did not have access to precise information allowing them, for example, to validate Facebook accounts. They also did not have access to the algorithms that allow Facebook to estimate the size of the different expatriate populations and must therefore trust the information provided by the company. The reproducibility of study results, an important criterion for any scientific research, can also be questioned. Indeed, Facebook can at any time modify its algorithms for extracting the size of the populations under study, without the researchers being aware of it. Documentation of the algorithms is also missing.

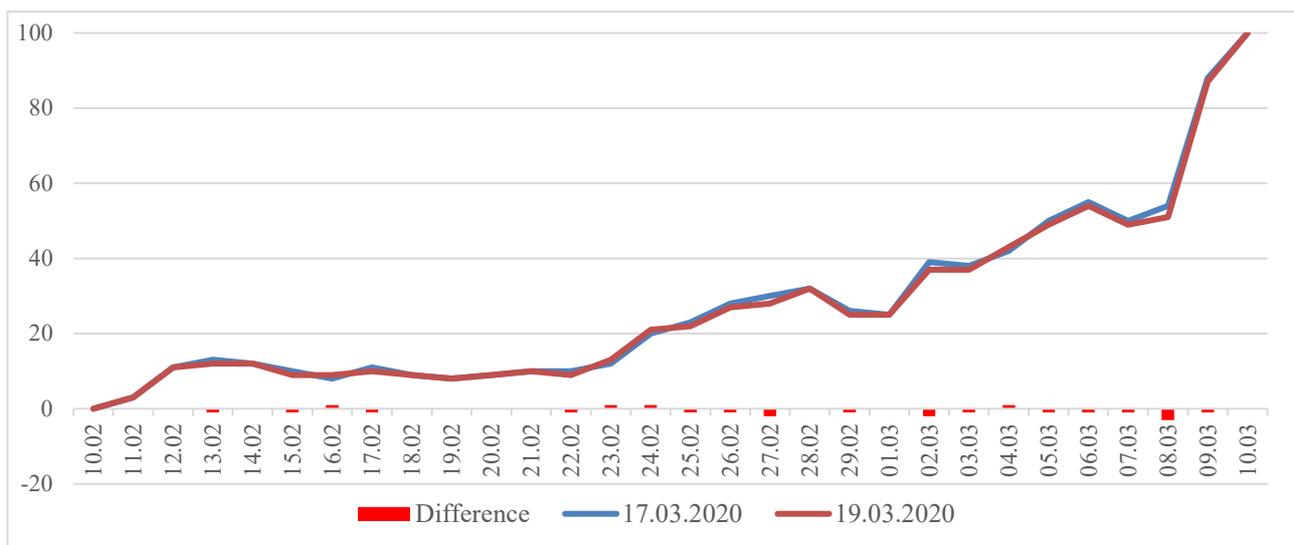
The lack of a reference population is also problematic. In their study, Ahas et al. (2018) use data from two access providers that may not be representative of all providers. Their analyses do not take into account individual behaviors, such as having more than one telephone, one for private use and one for business use or one for Estonia and one for another country, changing providers during

the year, etc. Therefore, there is much fuzziness in the data, which provides information that is probably useful but needs to be validated.

The rate of internet penetration and the use of social networks also poses certain problems. If the number of expatriates of a nationality identified by Facebook is increasing or decreasing, is it due to a change in the number of migrants or to a changing rate of use of this social network? The same question arises with studies using Google Trends to analyze migration flows or numbers. Thus, it is possible to demonstrate an association between the number of searches and migration flows (Wanner to be published). Moreover, the market share of Google is changing. First growing since 2004 and becoming quasi-monopolistic around 2015, Google is facing increased competition from other search engines, which are also used in different ways depending on the country or community. For this reason, an evolution of the index could be due to a change in practice. The control of these biases is illusory, as they are not documented.

With regard to Google Trends, which is widely used to monitor the evolution of certain social or medical phenomena, another problem is identified related to the updating of data. According to the instructions for use of the site, “Trends data is an unbiased sample of our Google search data. It’s anonymized (no one is personally identified), categorized (determining the topic for a search query) and aggregated (grouped together). This allows us to measure interest in a particular topic across search, from around the globe, right down to city-level geography.” According to Google, searches made several times by the same user over a short period of time are eliminated. The information that is transmitted refers to a scale from 0 to 100: “We index our data to 100, where 100 is the maximum search interest for the time and location selected”¹². In contrast, a value of zero means no search for the keyword studied in the sample.

Figure 5: Result of a Google Trends search on COVID-19, conducted two days apart for the period February 10 to March 10, 2020



Source: Own computation

The approach poses a major problem for its potential users, as it is based on a sample that appears to be updated on a regular basis. For example, Figure 5 shows the research indices for the term

¹² <https://medium.com/google-news-lab/what-is-google-trends-data-and-what-does-it-mean-b48f07342ee8> (consulted 19.03.2020)

“COVID-19” for all countries in the world and for the period between February 10, 2020, and March 10, 2020. The search was conducted on March 17, 2020, and then repeated two days later, with strictly the same parameters (period, keyword, geographical area). The curves are certainly close, but small differences are nevertheless observed for 19 of the 30 days under study, which can reach three points in the index ranging from 0 to 10 on March 8. These differences indicate that the reproducibility or duplication of a study based on these data is not guaranteed, although reproducibility is, of course, necessary for any scientific study.

Another problem that will someday have to be resolved is that access to private companies’ data is not systematically guaranteed. Access depends on the goodwill of these companies and sometimes involves a charge, which limits the use of these data, although they are produced by the users of the services that the companies offer.

It is also important to mention issues of individual data protection as a limitation of the use of private data. Generally, social science research is very comfortable with data anonymity. Researchers need to know the characteristics (age group, gender) of the persons analyzed in order to interpret the results, but they do not need information that can identify individuals. However, the data are never anonymous. From the moment a person activates the GPS from home to the office, he or she is at risk of being identified. The case of Strava¹³, a sports tracking application that unknowingly published intensity maps of sporting activities in military camps in Afghanistan¹⁴, is a perfect illustration of the risks associated with this type of application. These risks are increased when researchers have access to unpublished data.

These various elements suggest a cautious use of these data. However, their value is not questioned, especially in situations where data must be available urgently. In March 2020, less than a month after the start of the COVID-19 outbreak in Italy, an Italian study was published on the internet (Peppe et al. 2020) analyzing the impact of containment measures on internal mobility in Italy. The interest of this study is to show the evolution of mobility at the regional level, captured by a smartphone application (Cuebiq), to verify the impact of the measures taken during the COVID-19 pandemic on the mobility of the population and, by extension, the impact of the decrease in mobility on the spread of the virus.

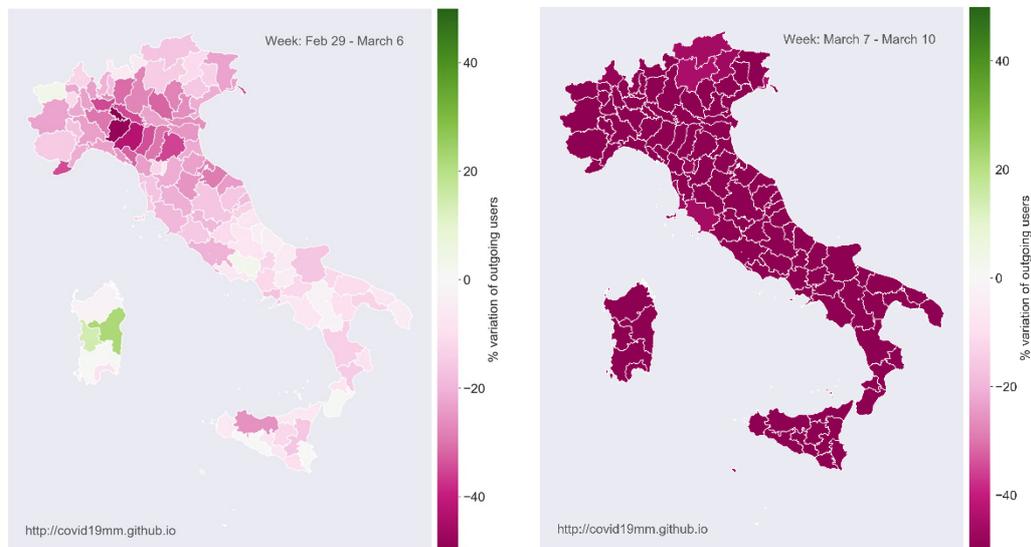
Figure 6, extracted from this study, shows the evolution, strongly decreasing, of this mobility between 29 February and 6 March, when the containment measures were limited to a few regions of Lombardy and Veneto, then between 7 and 10 March, just after the implementation of the national containment. The more detailed data, reduced to numbers of cases, will subsequently allow the analysis of the impact of reduced mobility on virus transmission and the number of people affected, which is essential for pandemic management and the evaluation of public health measures.

This study highlights the main advantage of big data, which is its immediate availability. In the past, the time required to obtain such information was several months, and the amount of information that could be collected was limited. Currently, the data are immediate and numerous and allow for real-time decision-making.

¹³ <https://www.strava.com/> (consulted 19.03.2020)

¹⁴ <https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases> (consulted 19.03.2020)

Figure 6: Evolution of mobility according to Cuebiq, from 29 February to 6 March and from 7 to 10 March



Source: Peppe et al. (2020)

5 Conclusion

With the development of statistical tools, there has been an important reflection on research needs regarding migration data (see, for example, Beauchemin 2014, Willekems et al. 2016). These different papers address the main challenges in terms of data collection and comparability, especially from an international perspective. Despite the existence of many limitations, innovative avenues are emerging, and this paper highlights some of them.

The purpose of this paper is then to show the main developments in data collection observed over the last few decades. These evolutions have made it possible to better document the migratory phenomenon and mobility in a context where flows have multiplied and become more complex. The rapid development of information technology has made it possible to adapt measurement instruments to the present reality.

Two directions were discussed. Administrative registers, which are usually exhaustive, provide the basis for analyses involving a large number of persons and offer interesting methodological extensions for the coming years. The interest of these registers is multiplied tenfold when they can be linked together and even more so when they cover a long period and allow longitudinal analyses.

The private data of social networks, telephone operators and the main players on the net represent a path to be investigated. While their interest in social science research has yet to be precisely established and while the biases of these data need to be taken into account and better documented, related analyses are multiplying. The first advantage of these data is their immediate availability, which in certain periods of crisis is an asset. An example is given with the monitoring of mobility in a period of an epidemic, where it is a question of having data rapidly available in order to verify the decrease in mobility and to understand how it is useful for health objectives, i.e., the reduction in

the number of cases. These data can also compensate for the lack of official data, as shown by Zagheni et al. (2017) for the numbers of migrants in different countries around the world.

The analyses presented here are not exhaustive, the aim, first of all, to show a few possible avenues, using examples that we have been able to propose based on Swiss data or international publications. We do not discuss a third way of collecting information, namely, population surveys (quantitative surveys). Surveys multiplied from the 1960s onwards and experienced a golden age with almost universal coverage of the population by fixed telephony. They have recently seen a new development with the availability of computer tools, which have democratized data collection. Surveys are an extremely useful means of collecting more personal or subjective information on the migration phenomenon, its experiences and consequences, in particular in a period where traditional censuses have been canceled in many countries, including Switzerland.

Other sources for collecting qualitative information could also be evoked and evolve according to the needs of users and the development of scientific methods. Their discussion, however, goes beyond the scope of this paper.

In the future, the next challenges for data collection will be the following: on the one hand, how to manage the important flows of information, sorting out the most telling data, and those that provide incorrect information or are difficult to comment on. How should data from diverse sources and collected in non-harmonized approaches be dealt with?

The risk is getting lost in this multiplication of sources and possibilities or having to sort out sometimes strongly divergent results depending on the source. The role of the researcher then changes. She no longer acts as just a producer and analyst of data but must also sort through information that is likely to lead to policy decisions. At the same time, she has to deal with a multitude of new data, which she will have to disseminate, value and comment on as correctly as possible.

On the other hand, how can data be accessed, especially when research budgets are becoming scarce? At what point should data become more freely accessible? Of course, it is not possible to give a precise answer to this question, but it can be seen that practices differ. In the case of private data, some researchers, for example, can access operators' data through collaborations, while others have no means of doing so. States, through statistical offices, could play a key role as a bridge between public producers and researchers. Otherwise, a risk of impoverishment of statistical systems could occur in some countries where operators are reluctant to share their data.

Bibliography

- Ahas R., Silm S., Tiru M. (2017). Tracking Transnationalism Originating in Estonia Through Mobile Roaming Data, *Estonian Human Development Report*. <https://inimareng.ee/en/open-to-the-world/tracking-trans-nationalism-with-mobile-telephone-data/>
- Beauchemin C. (2014). A Manifesto for Quantitative Multi-sited Approaches to International Migration, *International Migration Review* 48(4), 921–938.
- Bohon S.A. (2018). Demography in the Big Data Revolution: Changing the Culture to Forge New Frontiers. *Popul Res Policy Rev* 37, 323–341. <https://doi.org/10.1007/s11113-018-9464-6>
- Dunn H. L. (1946). Record Linkage, *American Journal of Public Health* 36, 1412–1416. <https://doi.org/10.2105/AJPH.36.12.1412>
- Fellegi I. P., Sunter A. B. (1969). A theory for Record linkage, *Journal of the American Statistical Association* 64(328), 1183–1210.
- Hatton T., Williamson G.J. (1994). *Migration and the international labor market, 1850–1939*. Routledge: London.
- Hoerder, D. (2010). *Cultures in Contact: World Migrations in the Second Millennium*. Duke University Press.
- Khoury M. J., Ioannidis J.P.A. (2014). Big Data meets Public Health, *Science* 28 Nov 2014, 1054–1055.
- Lazer D., Kennedy R., King G., Vespignani A. (2014). The Parable of Google Flu: Traps in Big Data Analysis, *Science* 14 Mar 2014, 1203–1205.
- Lombard A., Zufferey J. (2019). International Graduates in Switzerland: Transitioning into the Labor Market, *nccr – on the move Working Paper #21*. <https://nccr-onthemove.ch/publications/international-graduates-in-switzerland-transitioning-into-the-labor-market/>
- Morlok M. et al. (2015). *Les sans-papiers en Suisse 2015*. Basel: B.S.S. https://www.sem.admin.ch/dam/data/sem/internationales/illegale-migration/sans_papiers/ber-sanspapiers-2015-f.pdf
- Newcombe H. B., Kennedy J. M., Axford S. L., James A. P. (1959). Automatic linkage of vital records, *Science* 130, 954.
- Pastorino R., De Vito C., Migliara G., Glocker K., Binenbaum I., Ricciardi W., Boccia S. (2019). Benefits and Challenges of Big Data in Healthcare: an Overview of the European Initiatives, *European Journal of Public Health* 29(Supplement 3), 23–27. <https://doi.org/10.1093/eurpub/ckz168>

- Pepe E., Bajardi P., Gauvin L., Privitera F., Cattuto C., Tizzoni M. (2020). *COVID-19 Outbreak Response: First Assessment of Mobility changes in Italy Following Lockdown*.
<https://covid19mm.github.io/in-progress/2020/03/13/first-report-assessment.html>
- Piguet, E. (2005). *L'immigration en Suisse depuis 1948 - Une analyse des flux migratoires*. Zurich: Seismo.
- Ruggles S. (2014). Big Microdata for Population Research. *Demography* 51, 287–297.
<https://doi.org/10.1007/s13524-013-0240-2>
- Steiner I., Wanner P. (2015). Towards a New Data Set for the Analysis of Migration and Integration in Switzerland, *nccr – on the move Working Paper #1*. <https://nccr-onthemove.ch/publications/towards-a-new-data-set-for-the-analysis-of-migration-and-integration-in-switzerland/>
- United Nations (1998). *Recommendations in Statistics of International Migration*. Revision 1. New York, United Nations. https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles_and_Recommendations/International-Migration/SeriesM_58rev1-e.pdf
- United Nations (2018). *Report of the Global Working Group on Big Data for Official Statistics*. New York: United Nations Economic and Social Council, 17. <https://unstats.un.org/bigdata/bureau/documents/reports/GWG%20report%20-%202018-8-BigData-E.pdf>
- Wanner P. (2012). *La démographie des étrangers en Suisse*. Genève : Seismo.
- Wanner P. (2017). How Well Are Asylum Seekers and Refugees Integrating into the Swiss Labor Market?, *nccr – on the move Highlights #2*. <https://nccr-onthemove.ch/all-publications/highlights-2/highlights-2-3-2/>
- Wanner P. (2019). *Préparation d'une base de données sur la situation économique des personnes en âge d'activité et à l'âge de la retraite (WiSiER)*. Berne: Office fédéral des assurances sociales. <https://www.bsv.admin.ch/bsv/fr/home/publications-et-services/forschung/forschungsbereiche/WiSiER.html>
- Wanner P. (to be published). *How well can we estimate immigration trends using Google data?*
- Willekens, F., Massey D., Raymer J., & Beauchemin C. (2016). International migration under the microscope, Fragmented research and limited data must be addressed. *Science*.
<https://doi.org/10.1126/science.aaf6545>
- Zagheni E., Weber I., Gummedi K. (2017). Leveraging Facebook's Advertising Platform to Monitor Stocks of Migrants, *Population and Development Review*.
<https://doi.org/10.1111/padr.12102>

- Zufferey J, Steiner I., Ruedin D (2020). The Many Forms of Multiple Migrations: Evidence from a Sequence Analysis in Switzerland, 1998 to 2008. *International Migration Review*. <https://doi.org/10.1177/0197918320914239>
- Zufferey J., Wanner P. (2020). La distribution spatiale de la population étrangère en Suisse. *Social Change in Switzerland* 22. <https://doi.org/10.22019/SC-2020-00003>