



**nccr —→  
on the move**

**Ilka Steiner and  
Philippe Wanner**

---

**Towards a New Data Set for  
The Analysis of Migration and  
Integration in Switzerland**

**Working Paper #1  
May, 2015**

**Ilka Steiner and  
Philippe Wanner  
(University of Geneva)**

---

# **Towards a New Data Set for The Analysis of Migration and Integration in Switzerland**

The nccr – on the move is the National Center of Competence in Research (NCCR) for migration and mobility studies, which has been operational since 2014. The overall scientific aim of the nccr – on the move is to offer a comprehensive and interdisciplinary framework for understanding the character of contemporary migration and mobility to Switzerland. It was established by the Swiss National Science Foundation with the objective of fostering an innovative and competitive field of research on migration and mobility in Switzerland. Managed from the University of Neuchâtel, the network consists of some 60 researchers from social sciences, economics and law, who collaborate in 19 projects that are based at the universities of Neuchâtel, Basel, Bern, Fribourg, Geneva, Lausanne, Lucerne, and Zurich.

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, Faubourg de l'Hôpital 106, 2000 Neuchâtel, Switzerland  
Contact for the Working Paper Series: [info@nccr-onthemove.ch](mailto:info@nccr-onthemove.ch)

## Abstract

To fully understand migratory flows, their determinants and their consequences for the host countries, researchers increasingly use methods based on life course – or longitudinal – approaches. Such approaches aim at following migrants from their arrival until their departure and at analyzing, among other factors, their integration and assimilation processes into the host society. These kinds of analyses inform to which extent (and after which duration of stay) migrants benefit from the same opportunities as native citizens. The introduction of a unique Personal Identification Number (PIN – corresponding to the social security number) in 2010 in the Swiss administrative registers and a new decree, which came into force in 2013, regulating data linkage for statistical purpose, allow today the development of longitudinal statistics based on population registers and official surveys. One objective of the NCCR On the Move was to create such a database in order to develop longitudinal analyses on migration issues. In this context, the aim of this working paper is to review the process leading to the creation of a new longitudinal database (called Swiss Longitudinal Demographic Database – SLDD). Therefore, we first describe its conceptual framework, by presenting the different available registers, the aims and the target population. Secondly, the article discusses the data linkage and the validation procedures. The paper ends with two examples of possible applications.

## Keywords

Longitudinal data, migratory trajectories, foreign resident population, Germans, Switzerland

## Acknowledgments

We thank the SNSF for the funding of the project. We also thank Jonathan Zufferey for his help when checking the quality of the data set as well as the SFSO staff for their expertise and availability: Marco Buscher, Christoph Freymond, Christel Gigandet, Claude Joye and François Rais.

## Contacts

[ilka.steiner@unige.ch](mailto:ilka.steiner@unige.ch), [philippe.wanner@unige.ch](mailto:philippe.wanner@unige.ch)

Institut d'études démographiques et du parcours de vie, Faculté des Sciences de la Société  
Uni Mail, 40, bd Pont d'Arve, 1211 Genève 4

## 1. Introduction

The increase in international mobility was one of the most challenging social transformations of the 19th and 20th centuries, leading today to a wide range of questions related to the number of migrants and their social and cultural integration. With the progressive disappearance of borders within the EU countries at the end of the 20th century, migration patterns progressively diversified in terms of either the motivations or the socio-demographic and economic characteristics of migrants. This diversification and the rising number of migrants take place in a cautious and even suspicious political and social context. To avoid mistrust or erroneous interpretations of migratory flows and their consequences for the host countries, it is essential that researchers provide robust analyses. To reach this objective, statistical data are required that not only measure and describe the migratory flows but also contribute to the discussion concerning the advantages and disadvantages of international migration.

In Western countries, visa or residence permit statistics available until the mid-twentieth century gave an estimate of the number of new immigrants and sometimes of their socio-demographic characteristics. Once inside the country, the immigrant was “forgotten”, from a statistical point of view, until his/her departure was recorded. Thus, migratory statistics generally focused on flows and only in rare cases on stocks of migrants (at the time of a census or if a population register was available). Nowadays, this descriptive approach is insufficient.

At the individual level, migration is one component of the life course involving different strategies: some individuals choose to stay in their host countries and others re-emigrate or migrate back and forth between two or more countries. A life course – or longitudinal – approach is therefore required to fully understand migration and its determinants. Such an approach aims at following migrants from their arrival until their departure and at analyzing, among other elements/factors, their integration and assimilation processes into the host society, especially whether they benefit from the same opportunities as native citizens.

Record linkage approaches were initially developed after World War II for the welfare state and the taxation system, introducing large files on individual citizen (Herzog et al. 2007: 82). According to Henry, the way was opened by geneticists: « Plus récemment est apparu le souci d'utiliser au maximum pour des études génétiques de longue haleine les divers renseignements sur une même personne collectés à diverses occasions, sous différentes formes, par divers services – recensement, état civil, service d'immigration, services sanitaires – etc. » (1963: 248).

Data linkage became widespread with the development of new statistical methods (Fellegi and Sunter 1969). Nowadays, the soaring rise of information technology significantly contributes to the development of scientific research in social sciences in its use of administrative registers (population registers, registers of foreigners, etc.) and in its adoption of a longitudinal approach instead of a transversal one. This new approach also allowed the transition from description to explanation of social facts (Henry 1963).

Switzerland followed the trend towards longitudinal approaches, however with a delay due to the absence until 2010 of a unique Personal Identification Number (PIN – corresponding to the social security number: AHV number) in the administrative registers and to the inaccessibility for

researchers of identification variables such as names or surnames. Today, this PIN allows the rapid and deterministic record linkage of registers and the identification of information for an individual in different data sets. For example, the Parliamentary Control of the Administration (PCA) published an evaluation of the residence of foreign nationals under the Agreement on the Free Movement of Persons (PVK 2013). In order to analyze foreigners' work and residence trajectories, the authors linked several administrative data sources, amongst others data from the Central System of Migration (ZEMIS), the Unemployment Register and the Social Security Register. Before the introduction of the AHV number, probabilistic linkage approaches had to be used to provide a longitudinal follow-up of residents with a resulting lower quality. In recent years, two demographics projects in Switzerland have constructed longitudinal data bases, by applying a probabilistic data matching approach. The first one, the *Swiss National Cohort*<sup>1</sup>, is still ongoing, and reconstitutes from the 1990 Census the life course of residents up to their death (Bopp and Gutzwiller 2007). Therefore, the 1990 and 2000 Censuses were linked together using probabilistic data linkage methods. The Central Register of Foreign Nationals was also requisitioned in order to assess the impact of migratory flows on mortality rates. However, the absence of a PIN complicated the linkage procedure and led to a reduced quality of the matched trajectories (Zufferey 2014). The second application was undertaken in the framework of a project addressing the question of the demographic behavior of foreign citizens (Wanner 2012b), including the analysis of re-emigration and naturalization behavior (Pecoraro 2012a, 2012b). This project was based only on the linkage of the Central Register of Foreign Nationals with the 1990 and 2000 Censuses, and was therefore limited in terms of available information.

Another explanation for the late development of longitudinal systems in Switzerland was the unclear legal situation of data linkages, statistically speaking, since the use of these linkages was finally regulated by the Federal Law of Statistics only at the end of the first decade of this century. In 2013, the Federal Department of Home Affairs (FDHA) launched a partial revision of the Federal Statistics Act (FSA), including on the issue of data linkages (Art 14a)<sup>2</sup>, in order to legislate on the statistical use of administrative registers. The new decree<sup>3</sup>, which came into force on 15th January, 2014, allows for the linkage of different data sources by the Swiss Federal Statistical Office (SFSO) staff as well as for the involvement of external researchers: "In order to fulfil its statistical duties, the Federal Office may link data, provided the data is rendered anonymous. In the event that data links involve data considered especially sensitive or that data links generate personal profiles, the linked data must be deleted on completion of the statistical analysis. The Federal Council regulates the details." (Art. 14. A.I.). Because of the restriction specified by this article, the data linkage process needs to take place at the SFSO headquarters and in collaboration with its staff.

In this context, Switzerland is today at the beginning of a new era of quantitative research on migration. The experience of other countries has led to great expectations regarding the opportunity to develop studies on migration and integration processes. The Nordic countries for instance base the description and explanation of migratory issues on administrative registers. For example Statistics Norway has access to the Central Population Register (CPR) since 1964, a centralized register that provides information on education, labor force participation, income and other topics

---

<sup>1</sup> Swiss National Cohort, <http://www.swissnationalcohort.ch/>, consulted on 23.05.2014.

<sup>2</sup> Federal Statistics Act (FSA) – <https://www.admin.ch/opc/en/classified-compilation/19920252/index.html>, consulted on 07.08.2014.

<sup>3</sup> Ordonnance du DFI concernant l'appariement de données statistiques, 17.12.2013 – <https://www.admin.ch/opc/fr/official-compilation/2013/5475.pdf>, consulted on 07.08.2014.

(Aalandslid 2007; Vassenden 2008). English-speaking countries also developed longitudinal studies on migrants, including longitudinal surveys (for instance the Longitudinal Survey of Immigrants to Canada – LSIC), which question on a regular basis a cohort of migrants after their arrival in the LSIC case, 6 months, 2 years and 4 years after their arrival, see for instance Zhao et al. (2010). In 2014, the creation of a new data set for the analysis of migration and integration in Switzerland was initiated in the framework of a new National centre of competences in research (NCCR) on migration called “On the Move” in cooperation with the SFSO. The aim of this paper is to review the process of its construction. Therefore, we first describe its conceptual framework, by presenting the different available data sets, the aims and the target population. Secondly, we focus on the data linkage and the validation procedure of the demographic longitudinal data base. The paper ends with two examples of possible applications.

## 2. Conceptual Framework

The creation of the longitudinal data base (called Swiss Longitudinal Demographic Database – SLDD) proceeds in two different stages. First, a demographic base is constructed; in which the Central Register of Foreign Nationals (ZAR) and the Population and Households Statistics (STATPOP) are linked. In a second step, other data sets, such as the Structural Survey (RS) or the Unemployment Statistics are incorporated to this demographic base. The statistics that were used or are going to be used are presented in chapter 2.1 and the aims and the target population are described in chapter 2.2.

### 2.1 Data Sets

With the rise of information technology, our societies have multiplied the number of public and commercial registers which include a wide range of information on the resident population. For the social sciences, a number of these registers are informative, especially after pooling them together. Table 1 contains a non-exhaustive list of registers, selected according to their relevance for the study of migration and structural integration, as determined by the following criteria:

- Exhaustivity: the registers contain information on the whole of the target population. The Structural Survey, which replaces the traditional Census since 2010 and which is based on a sample of approximately 200,000 permanent residents, presents the only exception.
- Availability of a PIN, the so-called “AHV number” (Social Security Number), at least since 2010. Before 2010, the registers contained internal identification numbers.
- Information on the migratory status of the population (place of birth, citizenship) and/or structural integration (labour market participation, education, etc.).
- Except for the Central Migration Information System, they also include Swiss citizens.

Table 1: A Selection of Population Registers in Switzerland

Description	Available since	Legal Basis	Statistical Unit	Main Variables
Population and Households Statistics (STATPOP)				
The Population and Households Statistics (STATPOP), which are based on the population and household register, are part of the surveys conducted within the framework of the new population Census, initiated in 2010	2010	Swiss Federal Law of 22 June 2007 on the Federal Population Census (CC 431.112 Swiss Federal Ordinance of 19 December 2008 on the Federal Population Census (CC 431.112.1)	Individuals and households within the permanent and non-permanent resident population	Individuals: date of birth, gender, marital status, citizenship, place of residence, place of birth, place of previous residence, duration of stay since last immigration, residence permit (for foreign citizens). Movements: births, deaths, arrivals, departures, acquisitions of the Swiss citizenship Households: persons living in the same dwelling, household's composition
Central Migration Information System (ZEMIS)				
Register managed by the Swiss Federal Office for Migration (FOM). It contains several data sources, including the Central Register of Foreign Nationals (ZAR) and the Automatic Registration System of Persons (AUPER). Since 2010, this register is included, from a statistical point of view, in the Population and Households Statistics	ZAR since 1982, AUPER since 1985	Swiss Federal Ordinance (142.513) April 12, 2006	Foreign population	Date of birth, gender, marital status, citizenship, place of residence, permit, duration of stay
Swiss Social Assistance Statistics				
The Swiss Social Assistance Statistics, collated and published by the Federal Statistical Office, collects information that makes it possible, inter alia, to determine the number of social assistance recipients, the type of means-tested benefits that are granted, the duration of the benefits and the family structure of the recipients	2004 at the national level	Federal Statistics Act, 9 October 1992 and its Ordinance on the Conduct of Federal Statistical Surveys (SR 431.012.1), 30 June 1993. Revised Asylum Act of 16th December 2005 (SR 142.31): Art. 102a of the Swiss Statistics on Social Assistance Recipients concerning the transfer of anonymized and aggregated data by the SFSO to the Swiss Federal Office for Migration (FOM)	Individuals and households	

Structural Survey (RS)				
The main focus of this survey is the observation of the socio-economic and socio-cultural structures of the population in Switzerland.	2010	Swiss Federal Law of 22 June 2007 on the Federal Population Census (CC 431.112) Swiss Federal Ordinance of 19 December 2008 on the Federal Population Census (CC 431.112.1)	Persons aged 15 or more of the permanent resident population living in private households	Language, religion, education, work and employment, mobility of commuters, types of households and families, status of occupation of the dwelling and rent to be paid
Buildings and Dwellings Statistic (BDS)				
The BDS provides information on the structure of the whole building and dwelling stock and on the population's living conditions (e.g. occupation density per room or dwelling, floor space per person).	2009	Federal Act of 22 June 2007 on the Federal Population Census (SR 431.112) Ordinance of 19.12.08 on the Federal Population Census (SR 431.112.1) Ordinance of 30.06.1993 on the Conduct of Federal Statistical Surveys (SR 431.012.1) Ordinance of 31.05.2000 on the Federal Register of Buildings and Dwellings (GWR-VO, SR 431.841).	Residential buildings and dwellings	Geo-coordinates, building category, period of construction, type of heating, power source for heating and warm water, number of floors, number of rooms, floor space

Source: SFSO online, amongst others: <http://www.bfs.admin.ch/bfs/portal/en/index/themen/00/05.html> and <http://www.admin.ch/opc/de/classified-compilation/20050566/index.html>, consulted in January 2015.

Additional registers could certainly be added to this list, such as the fiscal registers (which are managed at the cantonal level) and the register of private enterprises. Further, registers in the field of health (registers of tumours, hospitalizations, professional accidents, etc.) can be used to describe the health condition of individuals. Moreover, surveys can also provide added value when they are linked with registers. The aim of an integrated system is to put together all available information in order to better understand the migrants' living conditions.

Finally, several data sets were not selected as they are based on samples that do not allow precise analyses at the regional level or for specific groups of foreigners. This is the case of the TREE survey<sup>4</sup> or the Swiss Household Panel.<sup>5</sup>

Importantly, a specificity of this SLDD is its flexibility. The selection of registers that are used followed a bottom-up approach, through which the researchers' anticipated needs defined the registers that were integrated. The further extensions of the database will be defined according to the needs and requirements of both researchers and policy makers.

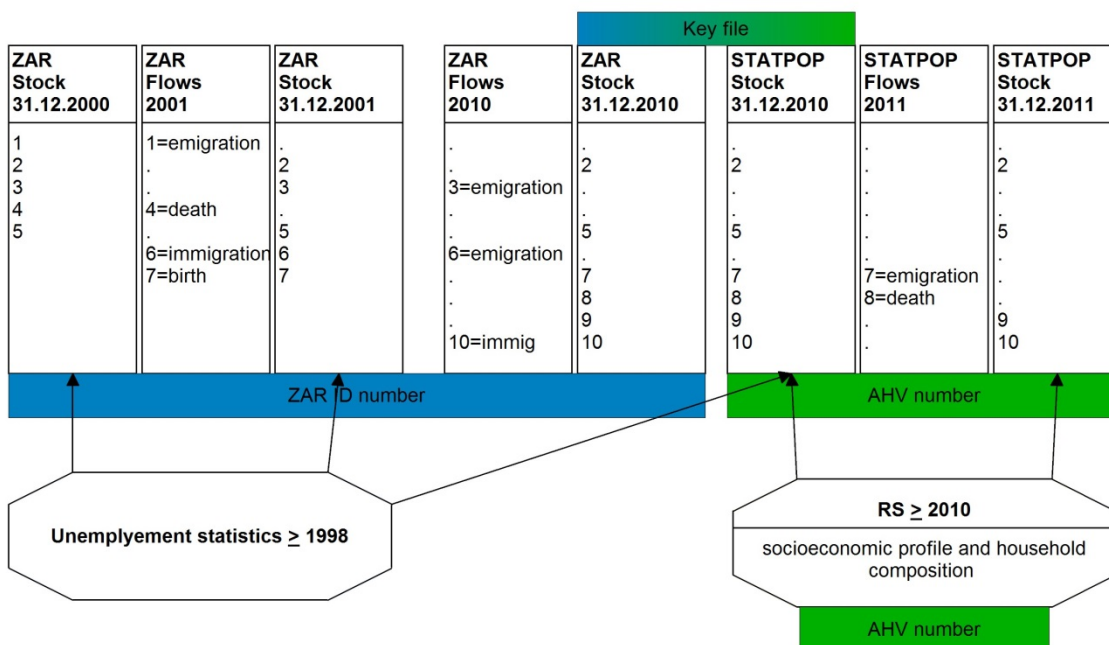
<sup>4</sup> TREE – Transition from Education to Employment, <https://tree.unibas.ch/fr/home/>, consulted on 23.05.2014.

<sup>5</sup> SHP – Swiss Household Panel, <http://forscenter.ch/en/our-surveys/swiss-household-panel/>, consulted on 23.05.2014.

## 2.2. Aims and target population

As exemplified in Figure 1, the aim is to construct a database that allows the follow-up of every foreigner since his entry into the target population, for instance by immigration or birth, until his/her dropping out through naturalization, death or emigration. The re-immigration of formerly emigrated foreigners is also documented. Furthermore, the database allows the tracing and monitoring of successive internal movements (changes in residence), and other changes such as in civil status or permit category. Other indicators referring to the socio-demographic structure of the population and the migratory status (such as the reason for immigration and the date of arrival in Switzerland) are also documented. The period under study so far ranges from 1997 to 2013.

Figure 1 : Demographic Base and Extensions



Source: The numbers stand for individuals. For instance person number 7 was born in 2001 and present in Switzerland until the end of 2010, before he or she emigrated in 2011.

Until 2010, only the foreign resident population was integrated to the SLDD since no population register containing information on Swiss nationals, was available. Every type of resident detaining a foreign nationality was considered, except foreigners in the asylum process (asylum seekers and provisionally-admitted refugees) who will be integrated in a further step: short-term residents (permit L), resident foreign nationals (permit B), settled foreign nationals (permit C) and resident foreign nationals with gainful employment (permit Ci). From 2010, the entire resident population, of both Swiss and foreign nationality, is considered. Residents are classified according to their primary legal place of residence. Secondary communes of residence are not considered. In order to take into account the frequent change in the commune of residence observed in Switzerland during the period, we converted the information on the commune(s) of residence each year in order to refer to the situation at the end of 2013.

### 3. Methods

#### 3.1. Demographic data

The SLDD is constructed by linking the Central Register of Foreign Nationals (ZAR) and the Population and Households Statistics (STATPOP). What follows is a description of the two data sets and their quality.

##### 3.1.1. ZAR – Central Register of Foreign Nationals

Foreign nationals living in Switzerland and detaining a residence permit are registered in the Central Migration Information System (ZEMIS), which is managed by the Swiss Federal Office for Migration (FOM, see Table 1). It exists in its current form since 2008 and resulted from the consolidation of the Central Register of Foreign Nationals (ZAR) and the Automatic Registration System of Persons (AUPER). The ZAR "serves the economic supervision of migration flows and the attribution of permits. It was introduced in the early 1970-ies in a political context characterized by an agitated debate on the presence of the foreign resident population and by xenophobic popular votes, which pushed the government to be better informed regarding foreigners living in Switzerland." (Wanner 2012: 24).

Until 2010, the ZAR data were regularly transferred from the FOM to the SFSO, which validated them in particular to guarantee coherence with other population statistics. The corrections notably concerned birth and deaths (extracted from vital statistics BEVNAT), which were missing or communicated belatedly to the FOM.<sup>6</sup> Since the end of 2010, these statistics were integrated into the STATPOP statistics.

The ZAR is composed of two standardized files: the "stocks" data file contains a list of all foreign residents (as well as cross-border workers) with a legal permit on the 31 December of each year. This file contains information on the demographic and socioeconomic characteristics of every foreigner. Since the residence permit is no longer linked to a work permit for EU/EFTA citizens, there is sometimes a time gap between a change in the socioeconomic status and its recording. In particular, the quality of socioeconomic variables is questionable.

The "flow" data file records all modifications concerning the foreigner's presence or status during the year. Four different types can be distinguished: *entries* into the target population, namely births, internal and international migration, *exits*, such as death, naturalization<sup>7</sup> and emigration, *modifications* in the foreigner's civil or legal status (permit) as well as *administrative corrections*. Both files, stocks and flows, also contain information on the migratory context, for instance: arrival date<sup>8</sup>, residence permit, nationality, country of birth, reason for immigration and nationality of the spouse (Swiss or foreigner).

<sup>6</sup> SFSO's corrections are one explanation for differing results regarding the foreign resident population between SFSO and FOM.

<sup>7</sup> Following naturalization, the foreigner is considered as a Swiss citizen and does not figure anymore in ZAR, even though he/she kept his/her original nationality and therefore detains dual-citizenship.

<sup>8</sup> According to Wanner 2012, the arrival date is inferred from the issuing date of the first residence permit. In case this information is not available, the last arrival date is registered.

### 3.1.2 STATPOP – Population and Household Statistics

Since 2010, the demographic data collection system in Switzerland underwent an important reorganization. As other European countries, Switzerland abandoned the traditional census and replaced it with surveys and a harmonized set of Population and Household Statistics called STATPOP. These statistics are based on the communal population registers as well as on the ZEMIS statistics and include every resident (Swiss and foreigners) of Switzerland. Since 2010, the presence in Switzerland of a foreigner is therefore attested by both communal population registers and the ZEMIS statistics. Information from both sources is compared by the SFSO and records are selected using a standardized prioritization process.

STATPOP is also composed of two files, stocks and flows. Available information available includes demographic characteristics and the identification of the household. Its creation brought a harmonization of demographic data. It allows for comparative analyses between the Swiss and foreign residents. Also, new variables have been introduced, such as the municipality of residence one to five years ago or the date and type of the acquisition of the Swiss citizenship (in any), as well as the tracking of emigrants, owing to the registration of their country of destination. Moreover, other registers were adapted in order to match the structure of STATPOP and the new social security number was introduced into other registers and surveys (e.g. the Swiss Labor Force Survey).

ZAR stock and flow records were produced until 2010; however, results based on stock records have only been published by the SFSO until 2009, because STATPOP stocks were available since 2010. Regarding the flow data, ZAR is still used in 2010, due to the fact that technical problems linked to the new demographic data production system prevented the production of STATPOP flows 2010. This transition has some incidence on the reconstitution of the longitudinal database.

### 3.1.3. Quality of the data

The quality of administrative data is often questionable (Roos et al. 1987: 46), as the data are constructed for administrative reasons and not for research purposes. In this context, the different actors, namely the administration and the population implicated in the registration process have a crucial influence on the quality of the data. Migrants for instance are obliged to announce all changes in their status to the authorities. Delays between the event occurrence and its registration are therefore often observed. Also, the management of ZEMIS and STATPOP is decentralized and the data are entered at a communal level. Before 2010, this led to discrepancies at regional level with regards to the way corrections were done (Wanner 2012: 20). Moreover, gaps or duplicate records were observed, namely regarding non-systematic corrections (Wanner 2012: 19). In the ZAR for instance, international migration flows were often used to correct other information: There were cases where a characteristic was recorded wrongly. Instead of entering a correction record when identifying the error, the employee in charge of the register in a municipality sometimes recorded one emigration with the wrong information and a new immigration with the correct one. Even though the stock at the end of the year was correct, such corrections overestimated international migration flows. Furthermore, the information contained in the two data sets “flows” and “stocks” were not always synchronized and coherent, leading to missing stock records, even

though a person immigrated at the end of the year. Finally, the change in the data production system in 2010 led to a certain discontinuity in the data.

Despite all the administrative errors that are observed, the overall quality of both ZAR and STATPOP statistics is considered as sufficiently good to enable the longitudinal follow-up and statistical description of residents in Switzerland (Wanner 2012: 20).

### 3.2. Record Linkage

“The principle of record linkage is simply to compare and bring together records from two (or more) sources that are believed to relate to the same real-world entity and whose (presumably overlapping) description can be matched in such a way that they may then be treated as a simple record” (Berti-Equille 2007: 112). According to Herzog et al. (2007: 81), the easiest way to represent personal, verified identification numbers (PINs). Contrary to most projects where record linkages are based on the comparison of non-modifiable characteristics (such as name, sex, date of birth, etc.), we have such PINs at our disposition.

The Central Register of Foreign Nationals (foreigners – ZAR) and the Automatic Registration System of Persons (asylum seekers and refugees – AUPER) contain each an independent PIN at least since the late 1990s (Wanner 2012: 29). Until 2007, they each had their own PIN (respectively ZAR PIN and AUPER PIN) which were replaced with the introduction of ZEMIS by a common PIN (the ZEMIS PIN) in 2008. Before 2008, when obtaining refugee status and therefore a residence permit, an asylum seeker was deactivated in the AUPER register and assigned with a new PIN number in the ZAR register, leading to difficulties to follow the trajectory of such foreigners before and since the recognition of his/her refugee status. The introduction of ZEMIS and the ZEMIS PIN improved this situation. Since 2010, the ZEMIS further includes the new social security number (AHV number<sup>9</sup>). STATPOP also contains this AHV number.

The introduction of the new AHV number in 2010 was one of the elements improving the capacity to build the Swiss Longitudinal Demographic Database, but more generally to improve Swiss statistics. This identification number was introduced in almost all administrative registers as well as surveys, and hence allows for an easy linkage of data.

As mentioned, STATPOP is a compilation of different data sources. Regarding foreigners, records are extracted on the one hand from the new communal population registers, on the other hand from ZEMIS. A foreigner can therefore have several records, which are compared and chosen by a standardized prioritization process. The data set therefore contains simultaneously the ZAR, AUPER, ZEMIS and AHV number. For this project, SFSO provided the extraction of these four PINs, if available, from STATPOP 2010-2013.

In a first step, the extractions were reorganized in order to obtain a file (called “key file”) with one record per individual containing the different PINs at the close of different years (2010-2013). For the foreign population, this file contains 2,749,788 records. Normally the PIN numbers should stay

---

<sup>9</sup> Federal Social Insurance Office, The new Social Security Number, <http://www.bsv.admin.ch/themen/ahv/00011/02185/index.html?lang=de>, consulted on 11.02.2014.

the same whatever the year, but it was not systematically the case. A change in one of the four PIN numbers is observed for 0.5% of the persons considered, when they left for example Switzerland for some years and got a new number attributed at their return or due to an update in the AHV number. For instance, 11,321 foreigners received two different AHV numbers, and 837 other foreigners received two different ZEMIS numbers between 2010 and 2013. This situation led to different data cleaning procedures in order to guarantee a correct linkage. In particular, in case of multiple AHV numbers, the last number attributed is considered as the reference. It is anticipated that in the future this reference number will be kept even if AHV number updates occur.

This key file establishes a link between both ZEMIS and STATPOP databases. However, it has some gaps: in particular the bridge between the ZAR/AUPER PINs and the ZEMIS PIN is guaranteed only for foreigners who were present in Switzerland in 2008 or thereafter. For those who left the country before 2008, the ZEMIS PIN is not available, thereby limiting the possibility to link the AUPER and ZAR registers: in such cases we have to link both sources using probabilistic linkages procedures based on non-modifiable variables such as date of birth or gender.

Another limit is due to the file being based on STATPOP data from the years 2010 to 2013, which limits the ability to correct possible mistakes during the period preceding 2010. This leads to the necessity to organize validation procedures in order to control the quality of the data and the coherence of the results.

Finally, as previously mentioned, the AHV number can be updated. Such updates can be easily identified for foreign citizens, as we can compare this information with the ZEMIS number. For Swiss citizens, such updates are not traceable.

*Table 2: Number of Observations and PINs in the Different Data Sets, by Origin*

Source	Year	Observations			PINs		
		Foreigners	Swiss	Total	Foreigners	Swiss	Total
<b>Flows</b>							
ZAR	1997-2010	10'861'392	-	10'861'392	3'079'723	-	3'079'723
STATPOP	2011/2012	2'652'713	2'408'556	5'061'269	1'052'445	1'124'673	2'177'118
Total	1997-2012	13'514'105	2'408'556	15'922'661	3'540'500	1'124'673	4'665'173
<b>Stocks</b>							
ZAR	1997-2010	24'345'519	-	24'345'519	3'437'758	-	3'437'758
STATPOP	2010-2013	7'770'778	25'094'611	32'865'389	2'443'559	6'350'448	8'794'007
Total	1997-2013	32'116'297	25'094'611	57'210'908	4'073'518	6'350'448	10'423'966
Total		45'630'402	27'503'167	73'133'569	7'614'018	7'475'121	15'089'139

Source: authors' calculation

Table 2 shows the numbers of observations as well as the number of the different PINs contained in every data set. For the whole period under study, the number of individuals included in the SLDD reaches 15 million (representing more than 73 million records, as each individual can be observed up to 15 times) and there are 16 million events (such as migration, change of status) concerning 4.7 million inhabitants.

In order to guarantee anonymity in accordance with the FSA, the SFSO generates random PINs that have no resemblance with the AHV number. This PIN is an alphanumeric variable with 46 positions. The algorithm allowing the creation of this PIN is deposited at the SFSO, allowing for a standardized anonymisation of further data sets (i.e. STAPTOP 2014, 2015, etc.).

*Table 3: Fictive Example of a Trajectory*

Movement	Date	Permit	Municipality	Gender	Civil status	Birthday	Arrival year	Source
immigration	06.11.2010	7	261	1	2	03.01.1979	2010	
	31.12.2010	7	261	1	2	03.01.1979	2010	ZAR
	31.12.2010	7	261	1	2	03.01.1979	2010	STATPOP
emigration	05.12.2010	7	261	1	2	03.01.1979	2010	

Source: authors' own presentation.

The result of the linkage procedure is grouped in a mega file containing for every individual a record for every event (for instance, immigration or emigration) as well as a record at the end of each year (see Table 3 for an example).

### 3.3. Validation Procedures

As mentioned by Berti-Equille, “the quality of data mining results and the validity of results’ interpretation essentially rely on the data preparation process and on the quality of the analysed dataset” (2007: 101). Therefore the validation of the data is an important step of data production. Roos and Nicol (1999) discuss several sources of error within administrative registers. Firstly, administrative errors, such as duplicates, gaps, recordings of different dates and contradictory events need to be considered. Due to the nature of administrative data, events such as relocation and change in marital status are captured at reporting by the registrant and can therefore be registered with a delay. Moreover, the date reported may further conflict with existing information. All these limitations should be taken into account when analysing the data.

However, as mentioned before, the quality of the population registers is considered to be good, especially for the variables directly related to the attribution of a residence or work permit (for instance the type of permit or the reason of immigration). The quality can be further improved through different correction procedures, for instance in order to avoid (the limited number of) double registrations.

Secondly, design restrictions, such as discrepancies between information available in various sources are common. However, this does not apply to our project, since SFSO introduced standardization practices during the procedure of data production.

Thirdly, record linkage errors, such as incomplete data, can lead to failed linkages and to false positives (records that link two different individuals).

In order to verify the data, different tests were undertaken, in particular the control of erroneous linkages, the plausibility of sequences of events and their completeness. Once these tests were completed, a second phase involved the identification of the sources of the errors, by consulting the SFSO or by running plausibility tests with control variables. The third phase then consisted in deciding how to correct the data.

### 3.3.1. Duplicate Records

As previously mentioned, the different statistics (ZEMIS, population register, etc.) are taken into account to compute the final version of the STATPOP statistics. This step also includes the deletion of duplicate records. However, in a limited number of cases, this deletion was not carried out, for instance when a discrepancy was observed for one or more variables. In such cases, the linkages allow the identification of duplicate records. Table 4 shows a fictive example, where the two records were not matched since the gender of the person was wrongly reported in one record. In such cases we deleted the duplicate record.

*Table 4: Mistaken Duplicates (Fictive Example)*

Movement	Date	Permit	Municipality	Gender	Civil status	Birthday	Arrival year	Source
immigration	06.11.2010	7	261	1	2	03.01.1979	2010	
	31.12.2010	7	261	1	2	03.01.1979	2010	ZAR
	<b>31.12.2010</b>	7	261	<b>1</b>	<b>2</b>	<b>03.01.1979</b>	<b>2010</b>	<b>STATPOP</b>
	<b>31.12.2010</b>	7	261	<b>2</b>	<b>2</b>	<b>03.01.1979</b>	<b>2010</b>	<b>STATPOP</b>
emigration	05.12.2010	7	261	1	2	03.01.1979	2010	

Source: authors' own presentation.

### 3.3.2. Erroneous Linkages

In some cases we suspect that records concerning two different persons were linked together. This is probably due to confusions concerning the AHV numbers, leading the SFSO to attribute in 2010 a temporary AHV number for one of the two individuals, but not the following year. The difficulty in such cases is to determine whether records are duplicated (meaning that there is in fact one person recorded twice) or if the AHV number is incorrect (meaning that two persons have the same AHV number). The occurrence of such cases is very small (probably less than 20). An example is provided in Table 5: the AHV number of the first trajectory (person 1) was changed twice (numbers with 222 indicate temporary numbers). In 2011, the AHV number was the same as that of Person 2, probably another person. Such cases were forwarded to SFSO which has access to complementary information (for instance the person's name) and was able to correct the data.

*Table 5: Erroneous Linkage (Fictive Example)*

ID Number	Date	Permit	Municipality	Gender	Civil status	Birthday	Arrival year	Source
<b>Person 1</b>								
222.111.333	31.12.2010	7	261	1	2	03.01.1979	2010	STATPOP
<b>756.224.123</b>	<b>31.12.2011</b>	<b>7</b>	<b>261</b>	<b>1</b>	<b>2</b>	<b>03.01.1979</b>	<b>2010</b>	<b>STATPOP</b>
222.111.334	31.12.2012	7	261	1	2	03.01.1979	2010	STATPOP
<b>Person 2</b>								
756.224.123	31.12.2010	7	261	1	2	03.01.1979	2010	STATPOP
<b>756.224.123</b>	<b>31.12.2011</b>	<b>7</b>	<b>261</b>	<b>1</b>	<b>2</b>	<b>03.01.1979</b>	<b>2010</b>	<b>STATPOP</b>
756.224.123	31.12.2012	7	261	1	2	03.01.1979	2010	STATPOP

Source: authors' own presentation.

### 3.3.3. Incoherence between Information from Different Sources

Validation procedures also showed that for some variables, such as civil status, the value may not be coherent.

*Table 6* presents a fictive example of a man who got married in April 2004, according to the ZAR 'flow' file. All subsequent records however identify the person as being single.

*Table 6: "Married Singles" (Fictive Example)*

Movement	Date	Permit	Municipality	Gender	Civil status	Birthday	Source
marriage	05.04.2004	7	2831	1	<b>M</b>	16.09.1974	ZAR flows
emigration	12.09.2004	7	2831	1	<b>S</b>	16.09.1974	ZAR flows
immigration	30.11.2004	7	2831	1	<b>S</b>	16.09.1974	ZAR flows
	31.12.2004	7	2831	1	<b>S</b>	16.09.1974	ZAR stocks

Source: authors' own presentation. Civil status: M = Married S = Single.

The information recorded for marriages, permit renewals or status modifications is often questionable in terms of its quality, but there is no way to validate or not such information. The choice is to prioritize the information available in the 'stock' file. In cases such as presented in

*Table 6*, we suggest to consider the civil status as recorded on December 31, 2004 (i.e. single).

### 3.3.4. Multiple International Movements

As already mentioned, migration codes were often used in ZAR in order to correct or update status records such as the civil status or the commune of residence. In the example in *Table 7*, the sex of the foreigner was wrongly codified. In order to correct it, an immigration and an emigration were recorded in October 2010. In order not to overestimate migratory flows, such movements need to be corrected. Therefore, when both immigration and emigration take place the same day, both events are corrected.

Table 7: *Erroneous Migration Records (Fictive Example)*

Movement	Date	Permit	Municipality	Gender	Civil status	Birthday	Arrival year	Source
immigration	06.01.2010	7	261	1	2	03.01.1979	2010	ZAR
<b>Emigration</b>	<b>31.10.2010</b>	7	<b>261</b>	<b>1</b>	<b>2</b>	<b>03.01.1979</b>	<b>2010</b>	<b>ZAR</b>
<b>Immigration</b>	<b>31.10.2010</b>	7	261	<b>2</b>	<b>2</b>	<b>03.01.1979</b>	<b>2010</b>	ZAR
	31.12.2010	7	261	2	2	03.01.1979	2010	ZAR

Source: authors' own presentation.

### 3.3.5. Duplicated Events

In some cases, records for the same person are not coherent. In the example in Table 8, two entries in the population register were recorded: an immigration and a birth. The source of such an error is probably the delay occurring before the registration of some events such as a birth. In our example, the birth was recorded with a two months delay, but the administration had, in the meantime, recorded an immigration in order to register the child.

In such a case, the correction depends on the place of birth. If the birth occurs in Switzerland, the birth record is retained and the immigration movement is deleted. When the birth is declared to have occurred abroad, the immigration record is retained.

Table 8: *“Birth after Immigration” (Fictive Example)*

Movement	Date	Permit	Municipality	Gender	Civil status	Birthday	Arrival year	Source
Immigration	06.02.2010	7	261	1	2	03.01.2010	2010	ZAR
<b>Birth</b>	<b>11.03.2010</b>	7	<b>261</b>	<b>1</b>	<b>2</b>	<b>03.01.2010</b>	<b>2010</b>	<b>ZAR</b>
	31.12.2010	7	261	2	2	03.01.2010	2010	ZAR

Source: authors' own presentation.

Other corrections were also applied. For instance, in case of duplicate records (two births, two deaths, two naturalizations), one of the records was systematically deleted.

### 3.3.6. Complete and Incomplete Trajectories

The evaluation of the completeness of trajectories is based on different rules regarding the beginning and the end of a trajectory.

The beginning of the trajectory is defined by an immigration movement or a birth. In some cases, corrections (for example the recognition of an immigration that had not been documented before) can also provoke the beginning of the trajectory. For foreigners already living in Switzerland before the beginning of observations (1998), the “stock” record as of 31.12.1997 represents the beginning of the trajectory.

The end of the trajectories is assumed in case of departure, death, and naturalisation or (in some cases) corrections (recognition of an emigration that was not documented before). For people residing in Switzerland at the end of the observation period (2013), the “stock” records as of 31.12.2013 represent the end of the trajectory.

A presence at only one time (for instance at the end 2011) without any information regarding the entry or the departure is not considered as a trajectory and is therefore deleted.

Some trajectories abruptly end without information on emigration, death or naturalisation. This was specially the case in 2010 when the transition to STATPOP occurred. The contrary is also true, as some foreigners suddenly appear in STATPOP without previous evidence of their presence in Switzerland. This can be partly explained by the differences in the definitions of registers and by the way in which information is recorded. For instance, STATPOP includes some people with cross-border permits or waiting for a permit, but already living in Switzerland. Some foreigners can also lose their permit but stay in Switzerland. Such discrepancies are difficult to take into account, but can be partially resolved by specifying the population included in the analyses. It is particularly important to consider the same inclusion criteria for each database (for instance by excluding all diplomats or persons without a permit in each database).

Even after corrections, approximately 15,000 individuals in ZAR 2010 are not in STATPOP at the same date. Conversely, 30,000 persons are in STATPOP but not in ZAR. Hypotheses regarding such cases include the omission of a departure, a mistake during the registration or transitional situations (for instance people already emigrated but for whom the departure has not been announced at the communal level – STATPOP).

### **3.4. Final Content**

As mentioned before, the SLDD contains information referring to all Swiss citizens living in Switzerland from 2010 to 2013 as well as foreigners living in Switzerland from 1997 to 2013 with a regular permit. This represents approximately 15 million individuals followed over a period of up to 15 years.

The SLDD is organized in a way which allows its regular extension, year after year. Such an update will increase the period under observation and therefore the capacity of the database to provide useful information on migration and integration.

Moreover, the database is also organized to include other statistics, for instance those mentioned in Table 1. Up to now, the Structural Surveys 2010, 2011 and 2012 were linked with the SLDD providing useful information on the socioeconomic characteristics of residents in Switzerland. The database is already in use in different research projects. To provide some examples, the next sections present two on-going analyses.

## 4. Possible Applications

To illustrate the SLDD's interest, two short examples of possible applications are given.

### 4.1. The Temporality and Recurrence of Current German Migration patterns

With the principle of free movement of persons which offers the right to move, travel, study, work, settle and retire anywhere within the EU/EFTA, temporary and circular forms of migration have become more common (Faist 2008). Using the SLDD, we can not only measure the duration of stay in Switzerland to document this new form of temporary migration, but also precisely define the kind of migration as well as the probability of emigrating and returning to Switzerland.

A first descriptive analysis of the SLDD concerns the recurrence of immigration. Table 9 provides for instance the distribution of the immigration characteristics among Germans aged 18 years or more, arrived in Switzerland between 2002 and 2007 and observed during the five years following their immigration. In the majority of the cases the migrants arrived and stayed for at least 5 years (39%) or emigrated again without re-entering the country during the observation period (39%). However, some of the German immigrants can be considered as recurrent migrants (23%), since they immigrated twice or more to Switzerland (probably Germany).

Moreover, the data shows that German migrants emigrating once are either leaving the country up to one year after their immigration (72%) or at the end of the observation period (13%). On the other hand, recurrent migration patterns do in most of the cases not exceed duration of residence of more than one year (91%). This is probably due to their migratory intention, coming for a short term-migratory project and with a short-term permit (permit B or L).

*Table 9: Types of Migration Patterns among Germans Arriving in Switzerland between 2002 and 2007*

	Number	Percentage
Immigration ->	80'424	38.61
Immigration -> Emigration	80'716	38.75
Immigration -> Emigration -> Immigration	15'283	7.34
Immigration -> Emigration -> Immigration -> Emigration	13'993	6.72
Immigration>2 / Emigration>2, Immigration>Emigration	9740	4.68
Immigration>2 / Emigration>2, Immigration=Emigration	8133	3.90
Total	208'289	100%

Source: Steiner (2015)

Factors influencing the mobility patterns can also be investigated using the SLDD: they include for instance gender, age, and matrimonial status. In particular, mobility is higher among younger migrants than among older ones.

## 4.2. Longitudinal Measure of Integration

Another application of the database concerns the integration processes of migrants to Switzerland, in particular in the labour market. This question has received increasing attention during the last decade, especially after the diversification of migratory flows during the 1990s. The request for integration indicators also increased and the SFSO introduced some years ago a first attempt to monitor integration, using different indicators regularly computed from surveys (Kristensen 2014). One of the limits of the traditional measure of integration is related to the transversal approach which is systematically adopted. The indicators are regularly computed (for instance at the end of every year) and the values are compared with each other in order to discuss the level of integration. This approach does not take into account the characteristics of the migrant population which can change from one year to another, in particular when migration levels are high. Therefore, indicators are measured among populations that are not comparable in terms of age, country of origin or level of education.

The SLDD allows the measure and description of the integration processes according to the date of arrival, the country of origin as well as other criteria such as the level of education, socioeconomic characteristics, and the purpose of migration. This will not only enable the measure of integration of one specific group, but also the estimation of the rhythm of integration, elements which can vary according to socioeconomic and sociocultural attributes.

This approach is based on the assumption that the starting point from which migrants evolve to converge to the normative standard of native Swiss is different for each individual and depends on individual and group characteristics. While some migrants can be described as integrated from the outset, due to cultural proximity and a higher level of professional qualifications, others need to catch up on different dimensions in order to be considered similar to native Swiss. Thus, the rate at which new migrants converge to the average level of native Swiss, according to different parameters measured at different points in time, will be computed with the help of the SLDD.

## References

- Aalandslid, Vebjørn (2007). Using register data to monitor the immigration and emigration of immigrants. United Nations Expert Group Meeting on Measuring international migration: Concepts and methods. New York: United Nations.
- Berti-Equille, Laure (2007). “Measuring and Modelling Data Quality for Quality-Awareness in Data Mining”, in Guillet, Fabrice and Howard J Hamilton (eds.) *Quality measures in data mining*. Berlin: Springer. 43, p. 101-126.
- Bopp, Mathias and Felix Gutzwiller (2007). *Die Swiss National Cohort 1990-2000 – eine soziodemografische Datenbasis für longitudinale Gesundheitsanalysen*. Institut für Sozial- und Präventivmedizin der Universität Zürich: Swiss National Cohort Report Nr. 5.
- Faist, Thomas (2008). “Migrants as transnational development agents: an inquiry into the newest round of the migration–development nexus.” *Population, Space and Place*, 14(1): 21-42.
- Fellegi, Ivan P. and Alan P. Sunter (1969). “A Theory for Record Linkage.” *Journal of the American Statistical Association*, 64(328): 1183-1210.
- Henry, Louis (1963). “Réflexions sur l’observation en démographie.” *Population*, 18(2): 233-262.
- Herzog, Thomas N., Fritz J. Scheuren and William E. Winkler (2007). “Record Linkage - Methodology”, in Herzog, Thomas N., Fritz J. Scheuren and William E. Winkler (eds.) *Data Quality and Record Linkage Techniques*. New York: Springer.
- Kristensen, Ewa (2014). *Methodenbericht zum Indikatorensystem der Integration der Bevölkerung mit Migrationshintergrund. Konzepte, Methoden, Auswahlverfahren und Quellen*. Neuchâtel: Bundesamt für Statistik (BFS).
- Pecoraro, Marco (2012a). “Devenir Suisse. Les facteurs intervenant dans le choix de se naturaliser”, in Wanner, Philippe (ed.) *Dynamisme démographique de la population étrangère*. Seismo.
- Pecoraro, Marco (2012b). “Rester ou partir : les déterminants de l’émigration hors de Suisse”, in Wanner, Philippe (ed.) *La démographie des étrangers en Suisse*. Zurich: Seismo.
- PVK, Parlamentarische Verwaltungskontrolle (2013). *Evaluation zum Aufenthalt von Ausländerinnen und Ausländern unter dem Personenfreizügigkeitsabkommen*. Bern: Parlamentarische Verwaltungskontrolle PVK.
- Roos, Leslie and Patrick Nicol (1999). “A research registry: uses, development, and accuracy.” *Journal of clinical epidemiology*, 52(1): 39-47.
- Roos, Leslie, Patrick Nicol and Sandra Cageorge (1987). “Using administrative data for longitudinal research: comparisons with primary data collection.” *Journal of Chronic Diseases*, 40(1): 41-49.

Steiner, Ilka (2015). “Towards a new longitudinal data set for the analysis of migration and integration in Switzerland.” International Workshop on Methodological Challenges for the Study of Return and Circular Migration. TEMPER Project, Madrid, 22-23 January.

Vassenden, Kåre (2008). “Data basis for Statistics Norway’s migration statistics”, in Daugstad, Gunnlaug (ed.) Immigration and immigrants 2008. Oslo–Kongsvinger: Statistics Norway, p. 163-166.

Wanner, Philippe (2012). “Données et méthodes”, in Wanner, Philippe (ed.) La démographie des étrangers en Suisse. Zurich: Seismo.

Zhao, Jun, Li Xue and Tara Gilkinson (2010). “Health status and social capital of recent immigrants to Canada.” Health Policy Research Bulletin, 17: 41-44.

Zufferey, Jonathan (2014). Pourquoi les migrants vivent-ils plus longtemps ? Analyse des différentiels de mortalité des populations suisse et étrangères 1990-2008. Université de Genève: PhD thesis.