United Nations ECE/CES/GE.41/2013/28/Rev.1

Distr.: General
23 September 2013

Original: English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Fifteenth Meeting
Geneva, 30 September – 3 October 2013
Item 6 of the provisional agenda
Operational aspects of censuses

Quality assessment and dealing with population over-coverage -Experiences and first results of the 2010 census round[1]

Note by the Vienna University of Economics and Business and Statistics Austria

Summary

Statistics Austria carried out its first register-based census in 2011 and thereby finalized the transition from a traditional census to statistics based on administrative data. This change is accompanied by various advantages, like diminishing costs, prompt availability and the removed burden for the respondents. However new challenges, like the evaluation of register-based data and the problem of population over-coverage, arise. Therefore a special framework for the quality-assessment of register-based data was developed. It delivers a quality indicator for each attribute of the census, which is derived from four hyperdimensions, each reflecting an important quality-related dimension. This enables to monitor changes in the quality of attributes in the processing of data as well as over time. Furthermore a method to deal with population over-coverage is described. For both topics preliminary results are presented.

[1] Prepared by Franz Astleithner, Predrag Ćetković, Stefan Humer, Mathias Moser, Vienna University of Economics and Business, Department of Economics, and Eva-Maria Asamer, Manuela Lenk, Henrik Rechta, Eliane Schwerer, Statistics Austria.

I. Introduction

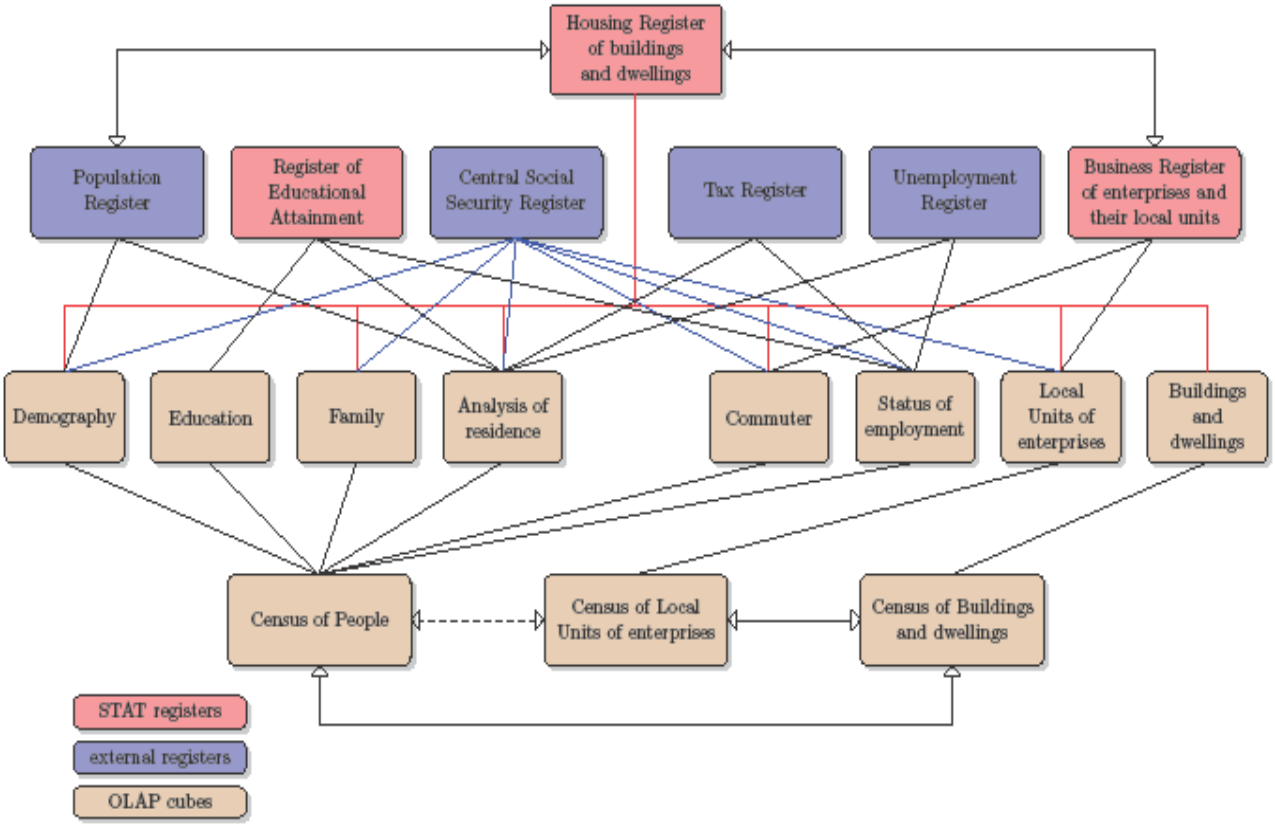
1. In Austria the last traditional census was carried out in May 2001. It was accompanied by a building and housing census as well as a census of local units of employment. Due to the challenging requirements of traditional censuses the importance of administrative data-sources for statistical purposes has been rising. The processing of data which has already been recorded by administrative authorities offers numerous advantages compared to survey data such as diminishing costs, removed burden for respondents and prompt availability of the data. Using data from existing registers ensures an optimal reflection of reality at reasonable expenditures. The combination of registers via unique linking variables enhances data quality and should help to harmonize definitions. Thus, an increasing number of National Statistics Institutes (NSI) promotes register–based censuses as a replacement for costly conventional censuses. In 2000 the Austrian council of ministers decided to establish the new method for the census of 2011 and in 2006 the regulatory framework came into force. Based on the population number of this census, the monetary amounts of fiscal equalization between municipalities and the federal financial authorities as well as the number of eligible voters for forthcoming elections are determined. Furthermore, information on commuters, education and employment offers important insights for economic and social policies. However, administrative data may apply definitions that differ from the needs of the NSI even though the data are of good quality (see United Nations, 2007, p. 3). Therefore, the NSI has to decide whether the data is adequate for the issue of interest.

2. In general, hardly any widely accepted procedures for the evaluation of register-based data exist. For this reason Statistics Austria developed a special quality-framework for this kind of data. Aim is to get a quality-indicator for each attribute in the census. These indicators are derived from four quality-related hyperdimensions. As information on attributes can be obtained from different registers, the quality of the outcome is improved, if all available information is combined. Another quality-related issue is the detection and reduction of population over-coverage. In this paper the development of the register-based census, the data privacy protection, a way to reduce population over-coverage and a general quality-framework for administrative data are presented. First selected results for the last two issues are given.

II. Development of the register based census and the principle of redundancy

3. A key question is to select the appropriate data source for the supply of the required information. The register–based census aims to cover all relevant variables that were formerly provided by traditional censuses. In this respect, the census in 2001 was the initial spark for the creation of some data sources, e.g. the Central Population Register (CPR), the Housing Register of Buildings and Dwellings (HR) or the Register of Educational Attainment (REDU). Prior to 2001 an interconnected network of population records was non-existent in Austria. By the introduction of the Housing Register in 2001 a Centralized Population Register evolved. The municipalities had to provide their records for the CPR. In 2004 the Housing Register of Buildings and Dwellings was synchronized with the Population Register for the first time. Moreover, the Register of Educational Attainment was founded during the census process of 2001.

Figure 1
The principle of redundancy in the Austrian register-based census



4. Register–based statistics depend on the quality of the information at the data holder. The data for the required variables are collected from as many sources as possible to check for consistency. In the Austrian register–based census this is accomplished by the combination of base and comparison registers. In this Figure the seven base registers are combined with the topics they supply.

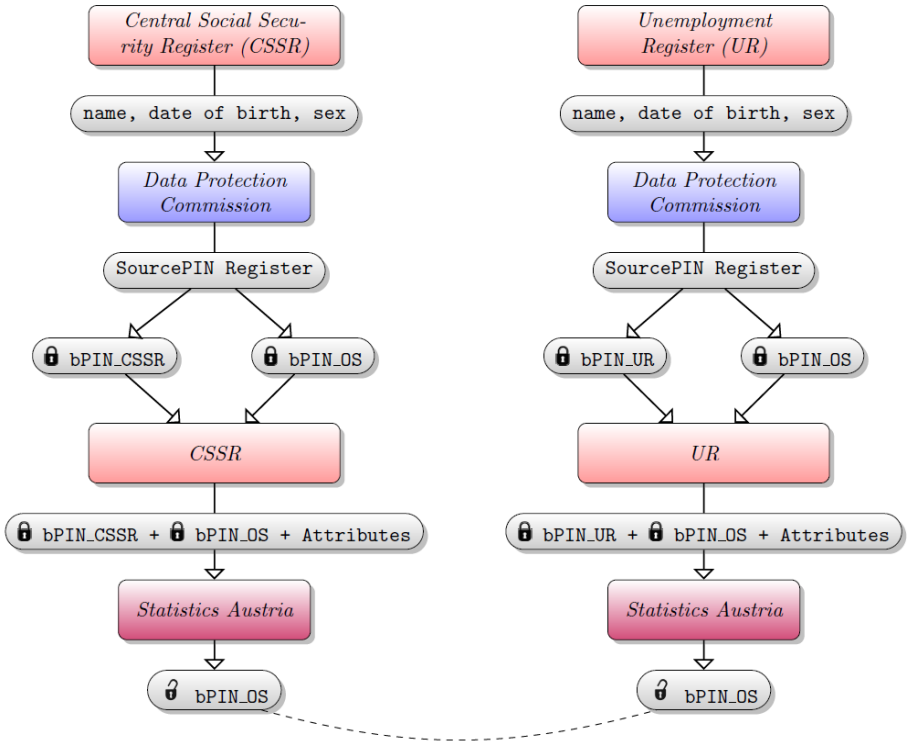
5. In Figure 1 all base registers of the census are illustrated and connected to the respective topics. The red-shaded data sources are maintained by Statistics Austria, the remaining information is provided by external data holders, like the Public Employment Service Austria (Unemployment Register: UR) or the Austrian Social Security (Central Social Security Register: CSSR). The Central Population Register forms the backbone of the census, since the units of analysis are individuals with main residence in Austria. To assure data quality, the base registers are backed up by seven comparison registers. These seven fields of administrative units are provided by 35 data holders and are mainly used for cross checks and the supply of information that is not or only partly available in the base registers (Berka et al., 2010, p. 300). Given the independence of the various registers as well as the autonomous process of data collection, the sources sometimes contain contradictory values for the same variable. Therefore, the principle of redundancy improves quality by acquiring information on sex, nationality, age etc. from as many registers as

possible. A particular method developed by Statistics Austria aims to identify one particular base register to provide the information for a certain variable, whereas the comparison registers are used to confirm the values in the base registers (Lenk, 2008, p. 3). While the registers provide sufficient information for most of the characteristics of the conventional census, some variables could not be covered by the register–based census. For instance the duration of the daily commute, colloquial language or religion are not captured by any of the registers at hand.

III. Data privacy protection

6. Since the names and the unique social security identification number of individuals are not part of the data delivery, a unique identification number is needed to merge the information from different data holders. To ensure data privacy, the introduction of a branch-specific personal identification number for official statistics (bPIN OS) is required before delivering the data to Statistics Austria. Each administrative branch in Austria, like the Austrian Social Security (CSSR) or the Public Employment Service Austria (UR) has its own bPIN. These 172-digit PINs, which should serve for privacy protected communication between public authorities via e-government, are derived by the Austrian Data Protection Commission (DPC) from the Source-PIN Register (Stammzahlregister). Figure 2 illustrates the achievement of data privacy protection through the derivation of branch-specific PINs as well as PINs for official statistics. The data holders demand these PINs from the DPC for each person by delivering the name, sex, date of birth, place of birth and address of an individual. The branch-specific PIN (bPIN) as well as the PIN for official statistics (bPIN OS) are derived from the Source-PIN Register, using a special and very complex algorithm developed by the DPC. The bPIN OS for register owners other than Statistics Austria is only provided in an encrypted form. The data holder has to send the data together with the encrypted bPIN OS to Statistics Austria. Moreover, the register owner delivers its own encrypted branch-specific PIN for each individual in order to identify the respective record in case of further inquiries by Statistics Austria (Lenk, 2008, p. 5). Each data holder is able to decrypt its own bPIN into a 28-digit number, hence only the NSI is able to decipher the bPIN OS and to use it as a common linking variable for the incoming register data. The described procedure of encrypting personal identification was first applied in a test census in 2006.

Figure 2
Achievement of data privacy protection in the register-based census



7. The individuals can be linked anonymously trough the branch-specific personal identification number for official statistics (bPIN OS).

IV. Dealing with Population over-coverage

8. Population over-coverage means that observations are included in a sample that should not be part of the sample. Since the results of the census are the basis for the financial equalisation between the municipalities and the financial authorities, the allocation of eligible voters for elections and for the sampling of surveys, good data quality is very important. To ensure the quality of the census Statistics Austria developed a process of residence analysis to detect and delete these supernumerary cases. In this process main residences in the CPR are questioned due to five reasons:

(a) A person has died before the reference date but still shows up in the CPR for the reference date (Died).

(b) An individual has multiple records for the reference date, for instance two or more main residences in the CPR (KIT-cases).

(c) A person was registered in Austria with a main residence for less than 90 days around the reference date (Minimum-stay).

(d) Persons who are registered in a municipality for not longer than 180 days but in another one before and after the reference period are considered as “census tourists”. These cases are counted to the latter municipality (Census tourist).

(e) A person has an entry in the CPR but no entry in any other register. In such a case it is reasonable to question the entry because of missing confirmatory "signs of life" (Single entry in the CPR).

9. If a case is questioned on the basis of reason one to four the corresponding municipality is asked to object to the exclusion of the doubtable case from the census. If the municipality isn't able to proof the existence of the case, it is excluded from the census, otherwise it is counted.

10. For reason five further investigations are carried out. However, some cases are included or excluded irrespectively of their doubtable statuses in the CPR according to a fixed rule-set for counting:

(a) If the place of main residence for the reference date was a detention centre or the individual declares to be in detention, the case is counted.

(b) If the place of main residence is a monastery or a convent, the case is counted.

(c) If the case refers to a child under the age of 15, the case is counted. Children are not questioned in written form and therefore counted anyway.

(d) If the entry in the CPR is changed by the individual within a certain period after the reference date, the case is counted. This is considered as a confirmation of the main residence for the reference date.

(e) If the gap of registration (no main residence) around the reference date in the CPR is less than 90 days, the case is counted (Registration-gap).

(f) If a child is born until the reference date and registered in the CPR within 90 days after the reference date, it is counted (Born before the reference day).

(g) If the entry of a person is deleted from the CPR by the municipality within a certain period after the reference date, the person is not counted (Entry deleted).

11. For the clarification of the other cases that had a single entry in the CPR a registered letter of clarification was sent to the place of main residence. In this letter[2] the individuals were asked to confirm the questionable status. The registration of the letters has the advantages that the return rates increase due to the official character and that information on the not-deliverable letters is available[3]. If these letters couldn't clarify the status, the municipalities were informed about the doubtable cases, whereupon they had the chance to confirm them.

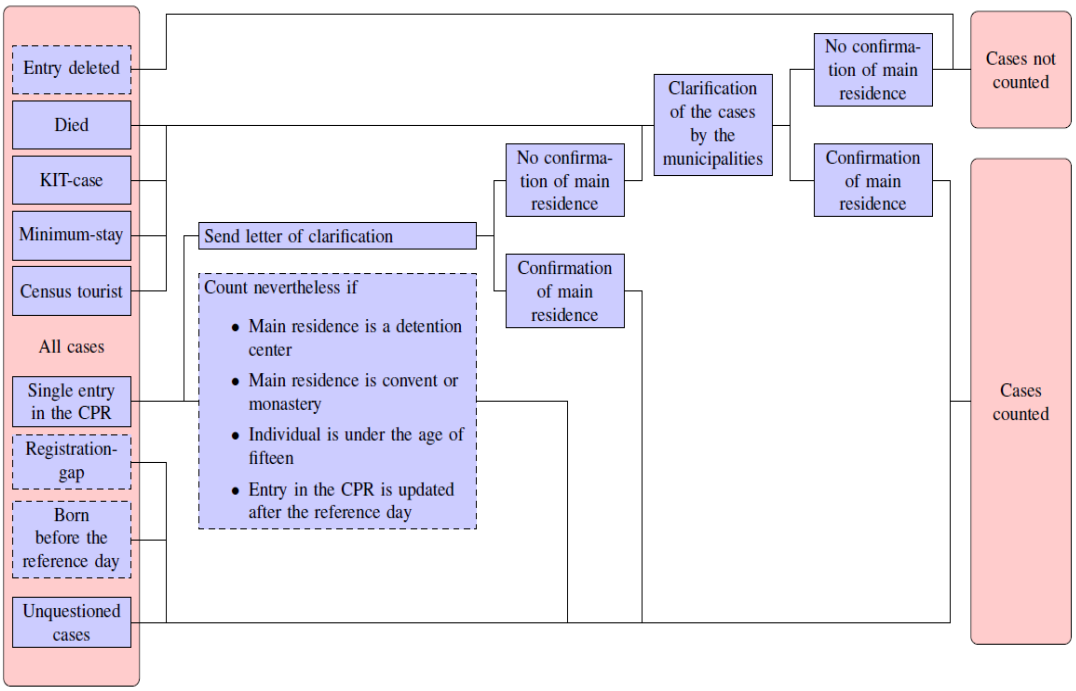
12. Figure 3 gives a graphical overview of the process of residence analysis.

[2] The letter, which was sent to all cases of clarification, contains only one single question: "Did you have your main residence in Austria at the reference date? -Yes or No."

[3] The reported categories are:

- Delivery not possible (RSB-A) (Addressee is unknown, Addressee moved, Drop-off point is deserted, Addressee died, Drop-off point irregularly used, problems to address addressee, Other Reasons)
- Letter wasn't accepted (RSB-B)
- Letter was accepted, but not answered (RSB-C)
- Main residence abroad (RSB-D)
- Other Reason (RSB-Sonst)

Figure 3
Process-flow for the residence analysis



13. Depending on the reason why a case is questioned, the different steps of the residence analysis are carried out. (Dashed frame signifies a decision according to the fixed rule-set for counting)

V. Results of the residence analysis for the census 2011

14. Technically the process of residence analysis was carried out in three rounds. In the first round cases which didn't show "signs of life" in the registers during the years 2008, 2009 and 2010 were questioned in November 2011, as it is reasonable to assume, that these cases are doubtable for the 2011 census as well. As the first survey was based on immediately available information, it could be close to the reference date of the 2011 census. This should improve the quality of the results of the first questionnaire. The survey for the second round was carried out in August 2012, where questionable cases that emerged in the 2011 census and haven't been covered in the first round were questioned. Round three consisted only of technical non-recognitions (Died, KIT-case, Census tourist and Minimum-stay) and was carried out in August 2012 as well. All the available information on each case is collected in a database. Thus, it is possible to reconstruct the confirmation or deletion of every single case.

15. For the census 2011 84,843 (about 1 per cent of all entries in the CPR) cases from the three rounds were sent to the municipalities for clarification. For the first round 54,378 cases with need for further clarification (Single entry in the CPR for 2008, 2009 and 2010) were detected. For 403 cases the address was deleted in the CPR after the reference date and for 120 cases the address was changed in the CPR in between the reference date and the day of the query from the CPR. Depending on the reason for the change of the status in the

CPR the cases were either counted for the census or not.[4] Respectively to 53,855 questionable cases a letter of clarification was sent. 255 cases could be deleted, because new information got available during the process. From the cases where the address stayed the same 13,934 (25.9%) were counted, 13,168 because the letter of clarification confirmed the main residence, 435 because they were confirmed by Statistics Austria according to the fixed rule-set for counting, 331 because there was no information on the delivery of the registered letter available. From the 523 entries that were updated after the reference date, 279 cases were counted, 46 deleted for technical reasons and 198 were still doubtable. Therefore, 39,894 cases were left from the first round and needed further clarification through the municipalities. For 5,557 cases the municipalities were able to clarify the status. In this stage 7,180 out of the 39,894 cases sent to the municipalities were counted; 229 because Statistics Austria could confirm the case according to the fixed rule-set for counting due to further information, 30 because the letters of clarification were delivered delayed, 3,054 because the municipalities were able to confirm the main residence and 3,687 cases were counted, because a sign of life showed up in a registers (see table 2). Respectively 4,013 cases were counted in the second round of clarification (see table 3). For the third round, consisting of 10,830 technical cases, the clarification request to the municipalities led to 244 counts, 7661 cases were deleted (see table 4). In total in round one 18 per cent, in round two 11.8 per cent, and in round three 2.3 per cent of the questionable cases were counted.

Table 1
Results of the letters of confirmation for the three rounds of clarification for the 2011 census

	Case with need for clarification	*Address in CPR stayed the same*				*Address in CPR is new or deleted*				*Need for clarification through the municipalities*
		Letters of clarification sent	*Doubtable main residence*	*Technical deletion*	*Cases counted*	*No letters of clarification sent*	*Doubtable main residence*	*Technical deletion*	*Cases counted*	
Round 1	54378	53855	39696	225	13934	523	198	46	279	39894
Round 2	47338	42019	29874		12145	5319	4245		1074	34119
Round 3										10830
Total	101716	95874	69570	225	29079	5842	4443	46	1353	84843

Table 2
Results from the first round of clarification in detail

Round 1				*Reason for confirmation*					
Reason why clarification letters didn't confirm cases		*Handed to the municipalities for clarification*	*Clarified cases from the municipalities*	*Confirmed by Statistics Austria*	*Main residence delayed confirmed by citizen*	*Main residence confirmed by municipality*	*Sign of life emerged*	*Cases counted*	*Share of counted cases %*
RSB-A	Delivery not possible	19232	1880	36	4	1415	1819	3271	17.0
RSB-B	Letter wasn't accepted	7974	1140	9	3	830	1085	1927	24.2
RSB-C	Lettter was accepted, but not answered	4102	1299	23	23	808	929	1783	43.5

[4] E.g.: If an individual changes her main residence the old one is counted for the census. If the municipality deletes the entry for technical reasons, the case is not counted for the census.

RSB_D	Main residence abroad	8304	1234	160	0	0	0	160	1.9
RSB_SONST	Other Reasons	282	4	1	0	1	37	39	13.8
RSB	All cases	39894	5557	229	30	3054	3867	7180	18.0

Table 3

Results from the second round of clarification in detail

Round 2		*Handed to the municipalities for clarification*	*Clarified cases from the municipalities*	*Reason for confirmation*				*Cases counted*	*Share of counted cases %*
Reason why clarification letters didn't confirm cases				*Confirmed by Statistics Austria*	*Main residence delayed confirmed by citizen*	*Main residence confirmed by municipality*	*Sign of life emerged*		
RSB-A	Delivery not possible	12592	1073	17	0	975	0	992	7.9
RSB-B	Letter wasn't accepted	8661	1210	7	2	1153	0	1162	13.4
RSB-C	Lettter was accepted, but not answered	4577	1679	22	294	1425	0	1741	38.0
RSB_D	Main residence abroad	3912	596	55	0	0	0	55	1.4
RSB_SONST	Other Reasons	4377	92	24	0	39	0	63	1.4
RSB	All cases	34119	4650	125	296	3592	0	4013	11.8

Table 4

Results for the third round of clarification in detail

Round 3 - Reason for the questioning of the cases	*Deleted cases*
KIT-cases	380
Individual died	3763
90-days-rule	3518
180-days-rule (only shifts assignment to the communities)	2925
Total	10586

Table 5
Results of the residence analysis for the 2011 census for Austria and the federal states

Federal State	*Population on October 31, 2011 according to the legal basis*	*Population in the CPR on October 31, 2011*[1]	*Balance of the 180 days rule (Census Tourism)* [2]	*Cases not counted*		*Share of the not-counted main residence in percent*[5]	*Number of citizens*[6]
				Technical[3]	Quality Assurance[4]		
Austria	8,401,940	8,472,421	-	-7,661	-62,820	0,83	7,461,953
Burgenland	285,685	286,778	-12	-237	-844	0,38	268,673
Carinthia	556,173	559,529	19	-429	-2,946	0,60	516,212
Lower Austria	1,614,693	1,622,053	-112	-1,472	-5,776	0,45	1,500,495
Upper Austria	1,413,762	1,419,780	37	-1,020	-5,035	0,42	1,295,694
Salzburg	529,066	535,715	9	-511	-6,147	1,24	462,162
Styria	1,208,575	1,217,506	-17	-934	-7,980	0,73	1,124,074
Tyrol	709,319	715,016	-6	-554	-5,137	0,80	631,323
Vorarlberg	370,440	372,094	18	-242	-1,430	0,44	321,465
Vienna	1,714,227	1,743,950	64	-2,262	-27,525	1,70	1,341,855

1) Data for October 31, 2011, 11:59^{pm}. Later changes are considered until June 30, 2012
2) Individuals are only counted in a municipality if they lived there for more than 180 days
3) a) Individuals died until the reference day b) individuals with multiple main places of residences c) individuals who lived less than 90 dies in Austria
4) Deleted after the letter of clarification didn't show signs of life
5) Is the sum of the balance of the 180 days rule and the cases not counted in proportion to Population in the CPR for the reference day
6) Persons with Austrian citizenship

16. Table 5 shows the outcome of the residence analysis and the resulting population for the 2011 census for Austria and the nine federal states. In Austria 7,661 cases were deleted because of technical reasons and 62,820 (cases sent to the municipalities minus counted cases) due to clarification process of quality assurance:

(a) 3,763 persons were deleted, because they died before November 1, 2011, but still showed up in the Central Population Register on the reference date.

(b) 3,518 persons were excluded from the census, because they haven't been registered in Austria for at least 90 days around the reference date. People who stayed shorter are – according to the census recommendations – regarded as visitors.

(c) For 2,925 persons the main residence was incorrect. With regard to the law for the register-based census persons who are census tourists shouldn't be counted in the census. People who moved from municipality A to municipality B before the census and moved back to municipality A after the census are only counted in B if they were registered there for at least 180 days. Hence, the rule for Census tourism only shifts individuals' assignment to a federal state, but doesn't change the overall population.

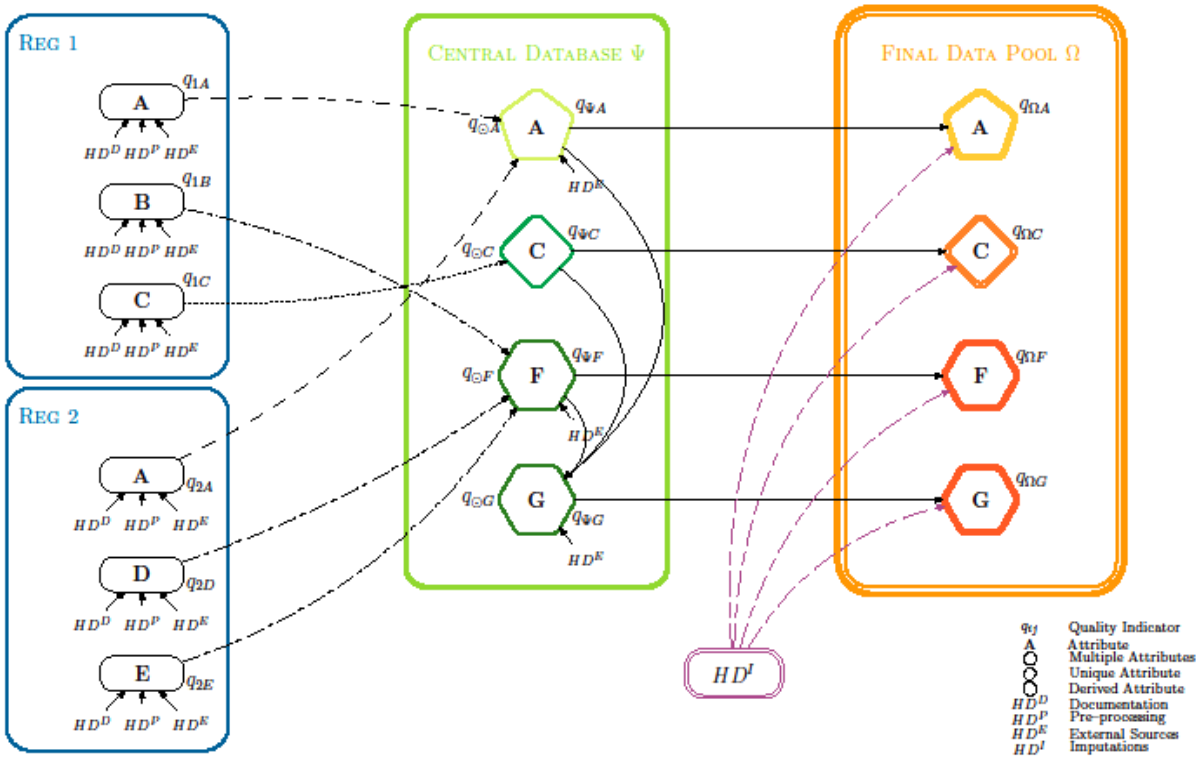
(d) 380 multiple records were detected and corrected. In such cases a person shows up more than one time in the Central Population Register. Therefore the resident's registration office has to decide which case to keep and which case to delete (KIT-Cases).

(e) 1,828 persons are counted in the census, even though they haven't been registered in Austria on the reference day. If a person was registered before and after the census, but not on the reference day and the register-gap was smaller than 90 days, the person is counted in the census nevertheless.

VI. A general framework for quality assessment of register based data

17. Especially the short transition time from a conventional census to a register-based census was challenging for Austria. In some European countries the interim period from gradually substituting survey data with administrative data lasted for about 20 years (see Ruotsalainen 2008). This allowed an intensive discussion on quality assessment by the NSIs and the data owners in these countries, whereas the transition schedule was very ambitious in Austria. Since the NSI is not responsible for the maintenance of the external data, the necessity of quality assessment in the process of register–based censuses arises. The quality analysis of register data has to satisfy several requirements such as transparency, accuracy and feasibility. This quality framework for the analysis of administrative data uses different hyperdimensions for the derivation of quality-indices (see Berka et al., 2012). The framework is closely tied to the data flow yet independent from data processing, which ensures that the processing is not influenced but evaluated. The data flow of the register–based census in Austria consists of three levels: the raw data (i.e. the registers), the combined dataset (Central Database, CDB) and the imputed dataset (Final Data Pool, FDP). Figure 4 illustrates the data processing, beginning with the receipt of raw data from the various administrative data holders. The information is connected via a unique key (bPIN OS) and merged to data cubes in the CDB. Further, the CDB data are enriched with imputations for item non-responses to create the FDP. Therefore, the FDP contains real and estimated information.

Figure 4
Quality Framework for register-based censuses



18. The quality framework consists of the three levels: raw data (blue), Central Database (green) and Final Data Pool (orange). Four hyperdimensions (HD^D, HD^P, HD^E, HD^I) aim to assess the quality for different types of attributes at all stages of the data processing. Finally, every attribute i in the statistics of administrative data obtains a quality indicator qi.

A. Quality assessment according to four hyperdimensions

19. The quality assessment is carried out via four hyperdimensions. Documentation (HD^D), Pre-Processing (HD^P), External Source (HD^E) are calculated for the base-registers and combined in the CDB, whereas the fourth hyperdimension, concerning the quality of imputations (HD^I), is applied on the FDP-level. The range of the hyperdimensions and the corresponding quality-indicators is from zero [worst] to one [best].

20. Independent from the actual values in the registers HD^D evaluates quality related processes at the data authority and the quality of documentation of data (for further details see: Berka et al. 2010). The degrees of confidence and reliability of the data holders are monitored by the use of a questionnaire containing 16 open-ended and nine scored questions (focusing on: administrative purpose, definitions, and data treatment). The NSI is therefore able to check for data collection methods or legal enforcements of data recording which may significantly influence the quality of the data. The questionnaires are answered by experts from the respective data holders and should thus deliver convincing results. The procedure is carried out for each attribute in each register. From the answers a quality-indicator concerning documentation is calculated for each attribute.

21. The second hyperdimension HD^P is concerned with formal errors in the raw data. Range errors in the data, item non-response and missing primary keys are detected in this step of the quality framework. The final result of this hyperdimension is given by the ratio of usable records to the total number of records. Again, this procedure is carried out for each attribute in each register. If the proportion of usable records for an attribute in a certain register is smaller than that of the same attribute within another register, the quality measure will accordingly be lower.

22. The third hyperdimension HD^E provides a comparison between the register–based data and an external source. In Austria the microcensus is a common benchmark for representative surveys and is assumed to be the best comparative dataset available. Checking for consistency with the external source offers the third quality measure which is the ratio between the number of congruent values and the total number of linked records. If the attribute of interest is not covered by the benchmark, local experts are asked for an evaluation of the data.

23. Now for every attribute in every register a weighted overall quality indicator (HD) can be calculated from the three hyperdimensions HD^D, HD^P and HD^E (for detailed information on the computation of the hyperdimensions see Berka et al. 2010)

24. Finally the fourth hyperdimension HDI evaluates the quality of the imputations on the FDP-level. According to the quality of the predictors and to the hit-ratio of the imputation-models a quality-indicator for the imputations can be derived (for further details see Ćetković et al. forthcoming).

25. Due to the quality indicators it is possible to evaluate the quality of each attribute at each level of the process-flow. Additionally the temporal changes in the quality can be monitored. This enables to compare the quality of the different data sources over time as well as the overall-quality of different census generations.

26. Quality assessment on the CDB and FDP level

27. The entire information from the registers is combined in the Central Database (CDB) which covers all attributes of interest for the census. Since there may be more than one data sources providing a certain attribute, a rule-set predefined by the NSI picks the most appropriate information from the underlying registers. Concerning the evaluation of quality for the CDB three types of attributes can be distinguished by their origin:

(a) Unique attributes exist in exactly one register, e.g. educational attainment (cf. attribute C in figure 2).

(b) Multiple attributes show up in several registers, e.g. sex (cf. attribute A in figure 2). The information from multiple sources is combined by a rule-set to derive the most appropriate value in the CDB attribute.

(c) Derived attributes are created based on different attributes, e.g. current activity status (cf. attributes F and G in figure 2). The registers do not contain any information for these attributes in the required specification.

28. For the quality of a unique attribute the quality-indicator is just the same as in the base-register. For multiple attributes the information on conflicting or matching estimations about the value of an attribute in the registers is combined to form a quality-indicator for each multiple attribute. To combine all available information the Dempster-Shafer theory of evidence – a special approach of fuzzy logic – is applied (see: Berka et al. 2012). For derived attributes the hyperdimensions HD^D and HD^P are applied to the registers. However the hyperdimension HD^E can be calculated on the register- and/or on the CDB-level. If there is more than one source for the derivation, the quality indicators of the relevant raw-

data attributes have to be weighted by their contribution to the derived attribute (for further information see: Berka et al. 2010)

29. The values from the CDB are complemented with imputations for item-non-responses (FDP). For every imputation a quality-indicator is computed. As the quality-indicator is set to zero for item-non-responses in the CDB and is set to the corresponding value of the hyperdimension HDI in the FDP, the overall-quality improves from the CDB- to the FDP-level through the process of imputation.

VII. Selected results from 2010

30. To give an example of the quality framework the results for a multiple attribute, the Legal Marital Status (LMS), are shown for the census test 2010. Information on the Legal Marital Status is contained in various registers. The spending registers are:

(a) Central Population Register (CPR)

(b) Unemployment Register (UR)

(c) Registers of Public Servants of the federation and the federal states (RPS)

(d) Central Social Security Register (CSSR: contains the sub-registers: HV,KA,KFA)

(e) Register of Social Welfare Recipients (RSWR)

(f) Register of Car Owners (RCO)

(g) Child Allowance Register (CAR)

(h) Central Foreigner Register (CFR is built by the sub-registers: AIS, FIS)

Table 6
Quality assessment for the spending registers for the attribute LMS

Register		HD^D	HD^P	HD^E	Q^{HD}
CAR		0.936495	0.96376	0.97143	0.95723
RPS		0.865071	0.98531	0.96087	0.93709
CPR		0.809524	1.00000	0.97296	0.92749
CSSR	HV	0.873003	0.95173	0.93918	0.921304
	KA	0.742927	0.98686	0.93698	0.888921
	KFA	0.682533	0.99058	0.98212	0.885077
UR		0.688876	0.96657	0.96319	0.872879
RSWR		0.834461	0.74422	0.89855	0.825745
CFR	FIS	0.396810	0.91791	0.87838	0.731035
	AIS	0.486762	0.455456	1.00000	0.647406

Table 7
Aggregated quality assessment on the CDB-level for the attribute LMS

	q_O	$CDB\text{-}HD^E$	q_Ψ	$q_{\Psi 0}$
CDB	0.95760	0.97382	0.96571	0.93512

Table 8
Aggregated quality assessment on the FDP-level for the attribute LMS

	$q_{\Psi 0}$	q_{Ψ}	HD^I	q_{Ω}	*Number of Imputations*
FDP	0.93512	0.96571	0.91016	0.96395	265,604

31. Table 6 shows the results of the quality assessment of the spending registers according to the hyperdimensions documentation, pre-processing and external source. The overall quality indicator for each register is calculated as the weighted average of HD^D, HD^P and HD^E. For example the Child Allowance Register (CAR) obtains 0.94 for the hyperdimension documentation, 0.96 for the hyperdimension pre-processing and 0.97 for the hyperdimension external source. The average from these three hyperdimensions is 0.96 and can be interpreted as an overall quality indicator of the Child Allowance Register for the attribute LMS. In a next step the Dempster-Shafer-Theory for the combination of evidence is applied and the quality indicator q_{Θ} for the CDB-Level is computed. Again, the results in the CDB are compared to an external source. The average of q_{Θ} and CDB-HD^E results in q_{Ψ}, the last quality indicator on the CDB-level. Table 7 shows the results of the quality assessment on the CDB-Level. For non-missing values the quality indicator q_{Ψ} equals 0.966. If we consider the missing values in the CDB, whit a quality-indicator set to zero, the last quality-indicator on the CDB-level for LMS $q_{\Psi 0}$ equals 0.935. On the FDP-level the quality of the imputations is evaluated (see Table 8). The hyperdimension is calculated from the quality of the predictors and the hit-ratio of the estimation model (HD^I=0.91). Therefore, the overall quality of the attribute LMS (q_{Ω} = 0.964) is equal to the weighted average of the quality indicators of imputed (HD^I) and non-imputed values (q_{Ψ}). Hence, the average of the final quality indicator in the FDP (q_{Ω}) is higher than the average of the final quality indicator in the CDB ($q_{\Psi 0}$) but lower than the average of the quality indicator for non-missing values (q_{Ψ}). Besides the assessment of the overall-quality as the average of all cases in the CDB (FDP), every single case obtains a quality-indicator. The more congruent values exist in the spending register, the higher is the quality-indicator for the multiple attribute.

VII. Conclusion

32. This paper presents the milestones in the transition from a conventional to a register-based census in Austria. It describes a structural approach for the quality assessment of administrative data. The quality of attributes can be assessed on three stages (raw data, combined data, imputed data). The evaluation is carried out along four hyperdimensions, each reflecting an important quality-related dimension. The quality framework results in a quality-indicator for each attribute in each register. Thus, the quality of attributes can be monitored in the processing of data. Furthermore it is possible to detect changes in the quality of single attributes or even census generations over time.

33. The second issue of this paper presents a way to deal with population over-coverage. A process of residence analysis developed and applied by Statistics Austria detects and reduces population over-coverage in the register-based census.

References

Berka, C., Humer, S., Lenk, M., Moser, M., Rechta, H., & Schwerer, E. (2010). A quality framework for statistics based on administrative data sources using the example of the austrian census 2011. Austrian Journal of Statistics, Volume 39, Number 4, 299-308.

Berka, C., Humer, S., Lenk, M., Moser, M., Rechta, H., & Schwerer, E. (2012). Combination of evidence from multiple administrative data sources: quality assessment of the austrian register-based census 2011. Statistica Neerlandica, Volume 66, Issue 1, 18-33.

Ćetković, P., Humer, S., Kausl, A., Lenk, M., Moser, M., Rechta, H., Schnetzer, M. (forthcoming). Quality measurement in administrative statistics and the assessment of imputations. Journal of Official Statistics (JOS).

Lenk, M. (2008). Methods of Register-based Census in Austria (Tech. Rep.). Statistik Austria, Wien.

Ruotsalainen, K. (2008). Finnish Register-based Census System (Tech. Rep.). Statistics Finnland.

United Nations. (2007). Register-based statistics in the Nordic countries. United Nations Economic Commission for Europe, New York and Geneva.