Quality indicators for the individual level – Potential for the assessment of subgroups

Session number 7
Date: 1st June 2016

Eva-Maria Asamer, Henrik Rechta, Christoph Waldner
Statistics Austria
Registers, Classifications and Geoinformation
eva-maria.asamer@statistik.gv.at
henrik.rechta@statistik.gv.at
christoph.waldner@statistik.gv.at

Madrid, May 31 - June 3

Quality assessment

Motivation

- Contemporary statistics based on administrative data take use of the principle of redundancy
- Assess the quality of all data sources + evaluate the process of combining the different sources

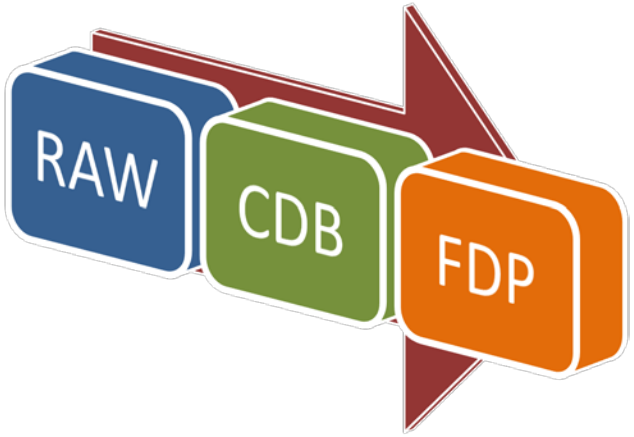
Application of a quality framework

- independent from data process →application to all statistics based on administrative data
- Data processes can be evaluated without influencing them

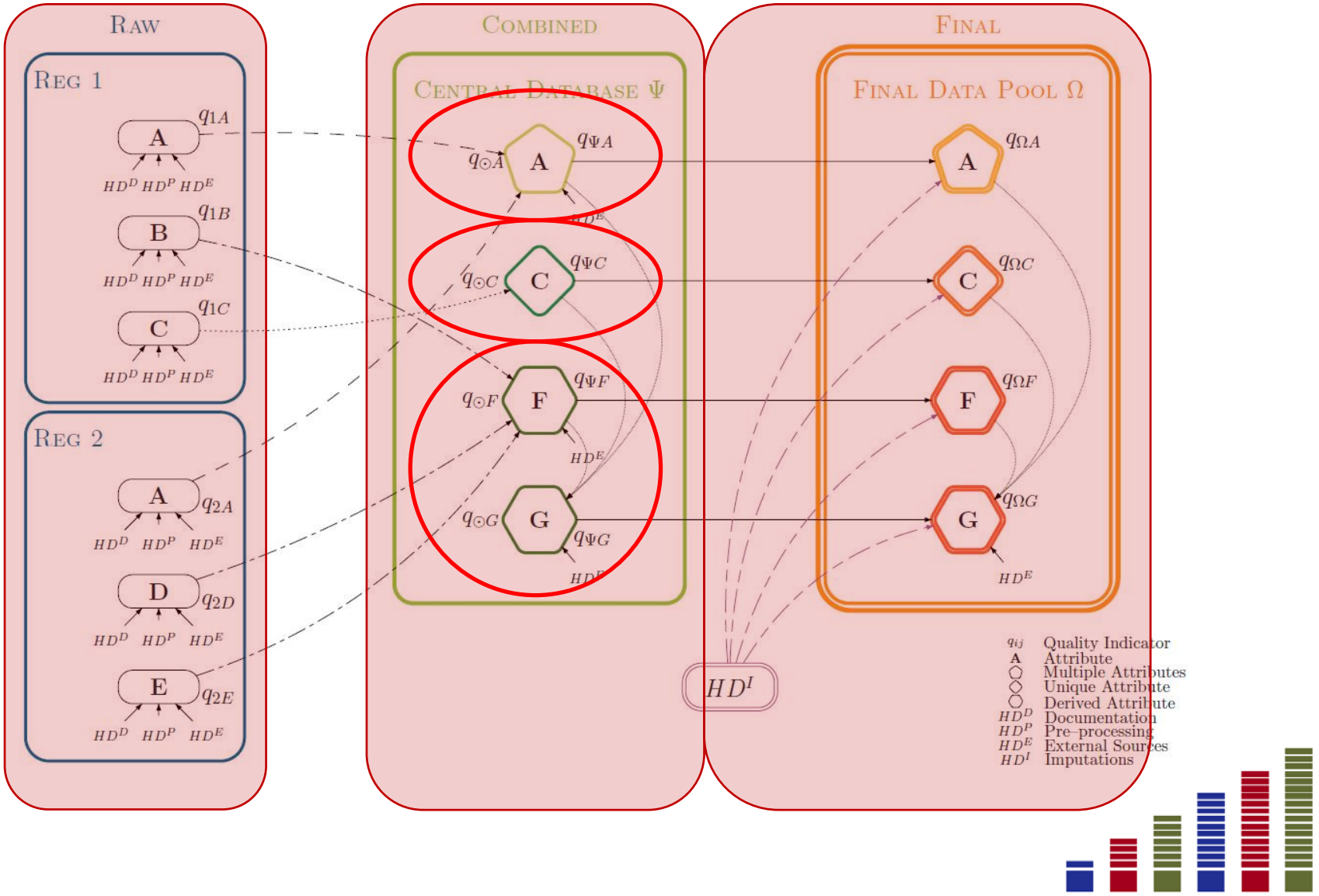
From Raw Data to the Final Data Pool

Three stages of quality evaluation

- Raw data
  - Registers provided by the data holders
- Central Database (CDB)
  - Combined information from the registers
  - Data is merged by a unique key
- Final Data Pool (FDP)
  - Final data including imputations



Quality framework



Usability of the results

Raw data

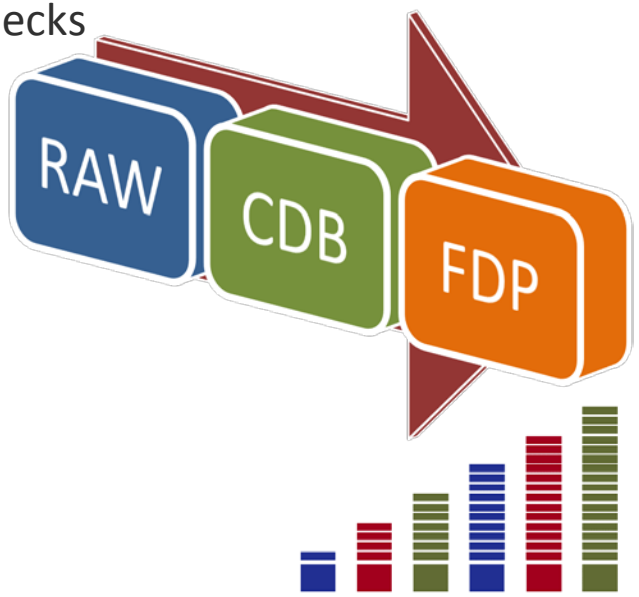
- Which register delivers a certain attribute with the highest quality indicator?
- Is there a register with a below-average quality for all delivered attributes?
- Annually monitoring of raw data quality (for each register and attribute)
- Assessing the usability of new data sources

Central Database

- Advancement of data quality through redundancy
- Comparison with prior data on this stage – plausibility checks

Final Data Pool

- Quality assessment for selected subgroups
- Comparison of attributes
- Comparison of results over time

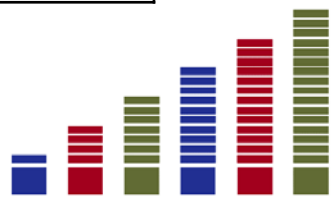


Results – a first look

Austrian register-based labour market statistics 2013

Average quality for multiple attributes per place of usual residence

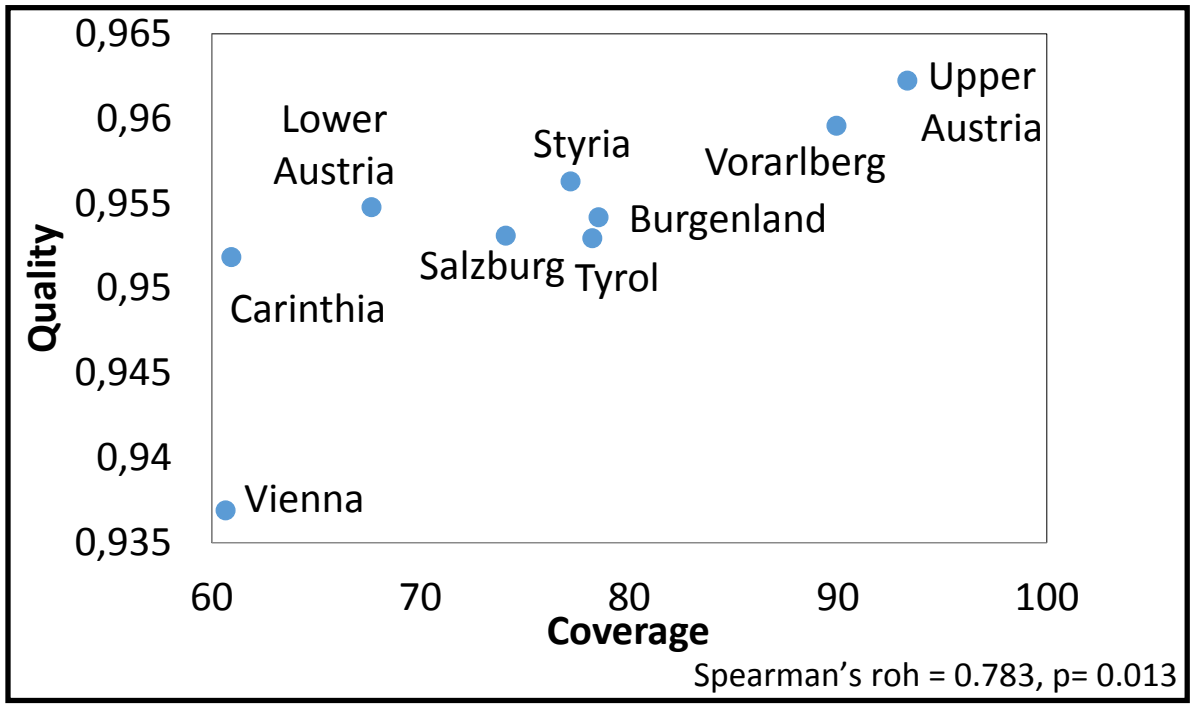
GEO	$\bar{q}_{\Omega\,\mathrm{AGE}}$	$\bar{q}_{\Omega\,\mathrm{SEX}}$	$\bar{q}_{\Omega\,\mathrm{LMS}}$	$\bar{q}_{\Omega\,\mathrm{COC}}$	$\bar{q}_{\Omega\,\mathrm{POB}}$
Austria	**0.999**	**1.000**	**0.952**	**0.991**	**0.991**
Burgenland	1.000	1.000	0.954	0.995	0.993
Carinthia	1.000	0.999	0.952	0.992	0.992
Lower Austria	1.000	1.000	0.955	0.993	0.989
Upper Austria	0.999	1.000	0.962	0.992	0.991
Salzburg	0.999	0.999	0.953	0.992	0.988
Styria	0.999	1.000	0.956	0.993	0.992
Tyrol	1.000	0.999	0.953	0.992	0.988
Vorarlberg	0.999	1.000	0.960	0.991	0.990
Vienna	0.999	1.000	0.937	0.986	0.993



LMS in Vienna I

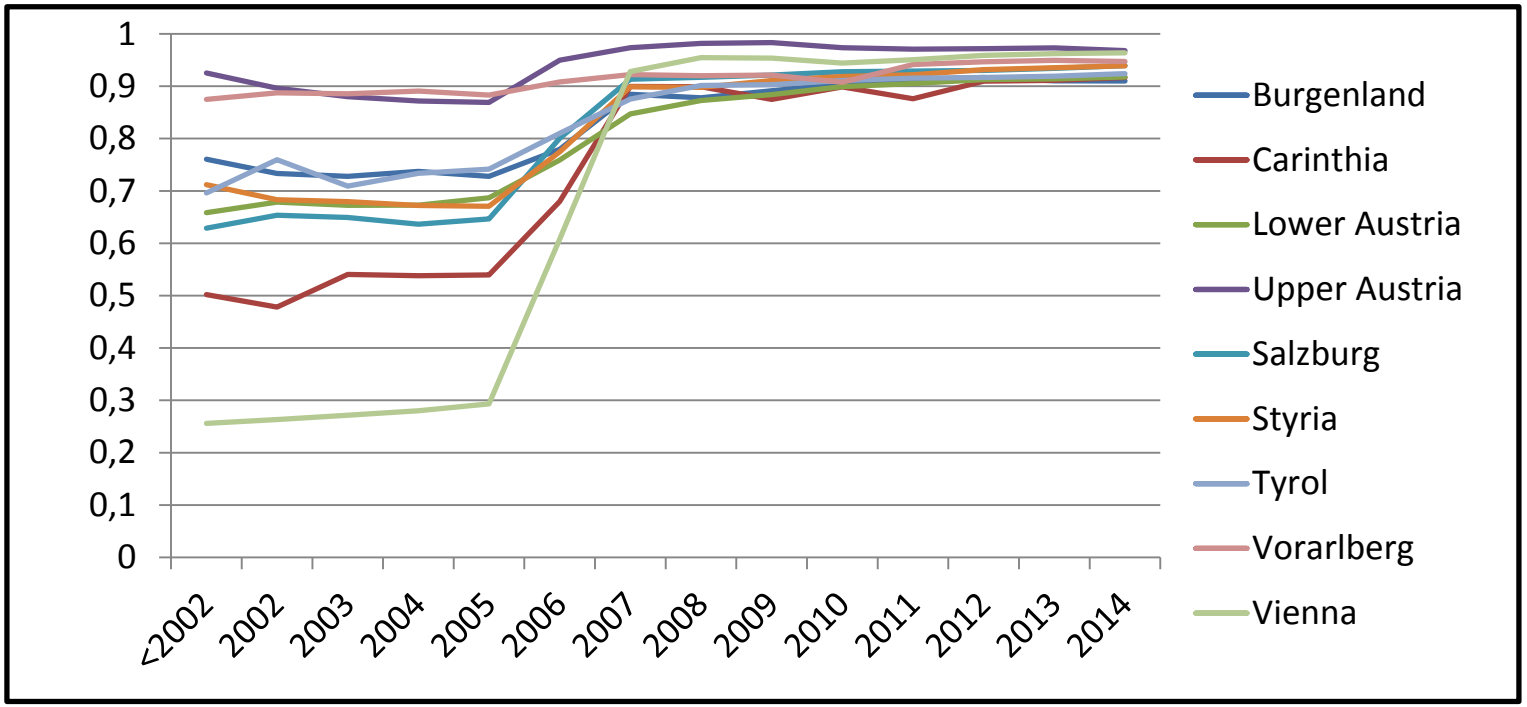
- Distribution: Vienna (10.27%), Austrian (7.77%) divorcees.
- Coverage: CPR main data source

Average quality and coverage rate for LMS of the CPR per laender



LMS in Vienna II

Coverage rate for LMS of the CPR per laender and date of registration



The principle of redundancy

- Same attribute in more than one source
- Comparison registers are used to confirm the values in the base registers
- Compute the quality via Dempster-Shafer theory on unit-level

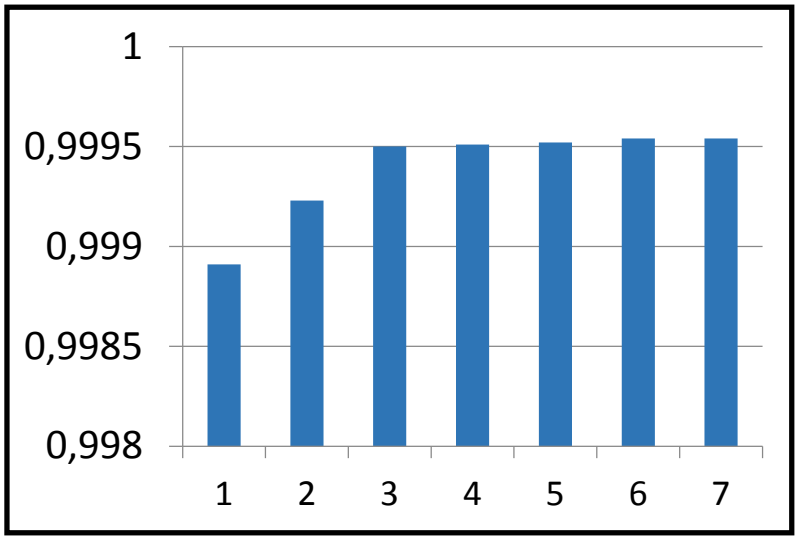
$q_{REG1} > q_{REG2} > q_{REG3}$

bPIN OS	SEX_REG 1	SEX_REG 2	SEX_REG3	SEX_VALID	Quality
⋮	⋮	⋮	⋮	⋮	⋮
ID3457	1	1	1	1	0,99
ID3458	1	1	2	1	0,90
ID3459	1	2	1	1	0,80
ID3460	2	1	-	1	0,30
⋮	⋮	⋮	⋮	⋮	⋮

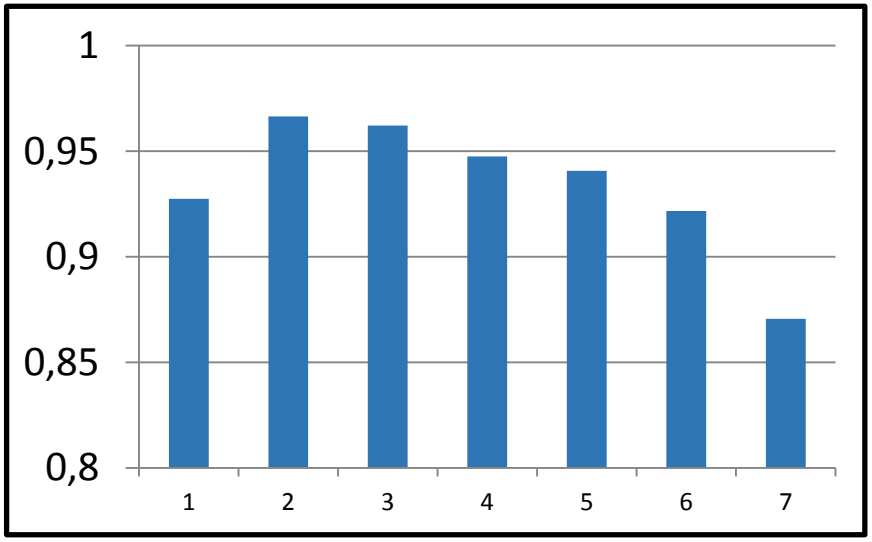
Redundancy

Average quality per the number of sources

AGE:



LMS:



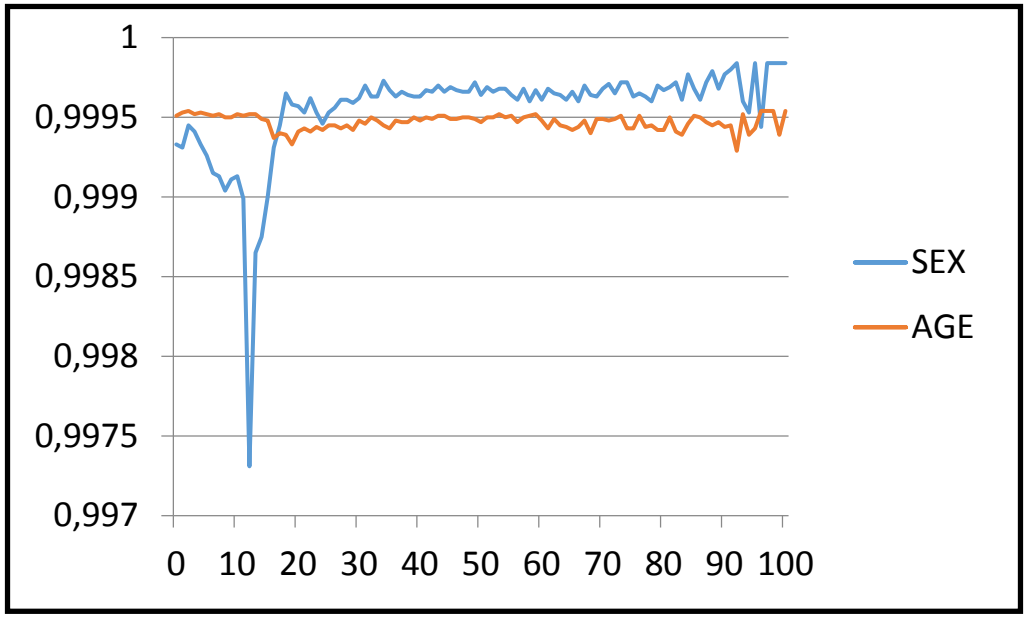
- Birthday is invariant from the date of excerpt
- LMS is not invariant from the date of excerpt

Selected subgroups

- Which are the values with lowest quality?
- How distributes the quality in relation to other attributes?
- What are the reasons for a worse quality ?

Selected subgroups: AGE

Average quality for AGE and SEX per age



- CPR vs. CSSR and FAR

Selected subgroups: POB

Average quality for POB and percentage of Austrians population

POB	$\bar{q}_{\Omega\ \mathrm{POB}}$	% of the Austrian population
Total	**0.991**	**100.00**
Republic of the Congo	0.991	<0.01
Democratic Republic of the Congo	0.871	0.01
People's Republic of China	0.988	0.17
Republic of China	0.902	0.02

- Similar names → confusions

Q 2016
European Conference on
Quality in Official Statistics

THANK YOU FOR YOUR ATTENTION