25th Meeting of the Wiesbaden Group on Business Registers
- International Roundtable on Business Survey Frames

Tokyo, 8 – 11 November 2016

Nadine Bachholz, Christian Schakerl

Statistics Austria

Session No. 5

Technology

Monthly bitemporal historicized data management within the new business register application at Statistics Austria

Introduction

When the first Austrian Business Register went live in the year 1995 it was based on a not very flexible environment. One of the major drawbacks was that within this business register only an annual historiography was feasible – so each maintenance operation within a unit would seem to be valid for the whole year, and if several changes occurred during one year, only the last one would show as all the others would be overwritten. When, for example, a unit was active from February to November it would show in the final dataset as inactive for the whole year. Therefore a wrong picture of the Austrian economy was presented.

When Statistics Austria decided to design a new maintenance application, one of the major objectives – next to a higher flexibility – was to implement a new data saving approach which should constitute a monthly bitemporal historiography. Those improvements allow Statistics Austria to convey a highly enhanced picture of the economy.

This paper describes shortly the design of a bitemporal database, the advantages of having a monthly bitemporal historiography and how this system is implemented within Statistics Austria.

Monthly bitemporal historiography

Bitemporal combines

- "business time" which is the time period during which a value/characteristic is correct in the real world
- and "system time" which is the actual moment during which a value/characteristic is stored in the database

Both – business and system time – have a beginning and an ending.

Monthly bitemporal historiography

There are two types of characteristics in the Austrian Business Register. Most are monthly historicized, some are not. For the monthly historicized characteristics business time as well as system time is saved, whereas for the others only the system time is saved whenever a maintenance operation takes place.

In which way a characteristic is saved is decided each time, a new characteristic has to be put into the system. When switching from the old application to the new one, the leaders of the two maintaining crews sat together and decided for each characteristic of the old system, whether it had to be saved monthly bitemporal or if a system time historiography would be sufficient.

RECHT_ID	KZ_R	HIST_V...	HIST_...	GUELTIG_VON	GUELTIG_BIS	R_F...
183521814	R969D681J	201601	999999	2016-09-02 06:18:08.724	9999-12-31 00:00:00.0	SA

Picture 1 – example of a new database entry

Picture 1[1] shows a new database entry – the unit R969D681J was entered into the system. It was entered on 2nd of September 2016, 6:18am (GUELTIG_VON) as system time; however, as it is relevant for the business register from January 2016 onwards the business time (HIST_VON) is 201601. The unit was flagged with "SA", which is the standard value of a legal unit.

Both ending times HIST_BIS and GUELTIG_BIS are "open".

RECHT_ID	KZ_R	HIST_VON	HIST_...	GUELTIG_VON	GUELTIG_BIS	R_F...	TAET_ID
183521814	R969D681J	201601	999999	2016-09-02 06:18:08.724	2016-09-02 06:52:27.425	SA	183521810
183521855	R969D681J	201601	201606	2016-09-02 06:52:27.425	9999-12-31 00:00:00.0	SA	183521810
183521860	R969D681J	201607	999999	2016-09-02 06:52:27.425	9999-12-31 00:00:00.0	LF	183521856

Picture 2 – example of a database entry for a monthly bitemporal saved characteristic

Picture 2 shows what happens in the database, when a monthly bitemporal characteristic is altered. Looking at the first row shows the same database entry ID (RECHT_ID). The value "SA" from 06:18 is now invalid, as a maintenance operation took place at 06:52. Now "SA" is only valid for the period 201601 to 201606. For the period from 201607 onwards the new value "LF" (agriculture and forestry) is valid. The NACE was changed as well (TAET_ID changed), in order to find out, which NACE has been

[1] RECHT_ID: the ID of the database entry
KZ_R: the ID of the legal unit which was maintained
HIST_VON/HIST_BIS: from which month on and until which month the value is valid
GUELTIG_VON/GUELTIG_BIS: from which moment on is the value set in the database, and when was it changed
R_FLAG: is a flag which marks, why a unit is in the business register –SA being the standard value, LF being the marker for agriculture and forestry units)
TAET_ID: ID of the alphabetical index entry

valid, one has to check the respective table with the respective TAET – ID. Therefore it is possible to see, that this unit was categorized in NACE 95.29 (picture 2a) and changed to NACE 01.43 (picture 2b).

TAET_ID	TAET_LIST_A...	T_OEN...	T_OENACE_FMT	T_OENAC...	T_OENACE_AV...	T_OENACE...	T_SPE...
183521810	AN	95290	S 95.29-0	527402	GA 52.74-02	2016-09-02	J

Picture 2a – NACE entry before maintaining the unit

TAET_ID	TAET_LIST_A...	T_OEN...	T_OENACE_FMT	T_OENAC...	T_OENACE_AV...	T_OENACE...	T_SPE...
183521856	AN	01430	A 01.43-0	012200	AA 01.22-00	2016-09-02	J

Picture 2b – NACE entry after maintaining the unit

Within the business register of Statistics Austria the business time is a monthly historiography:

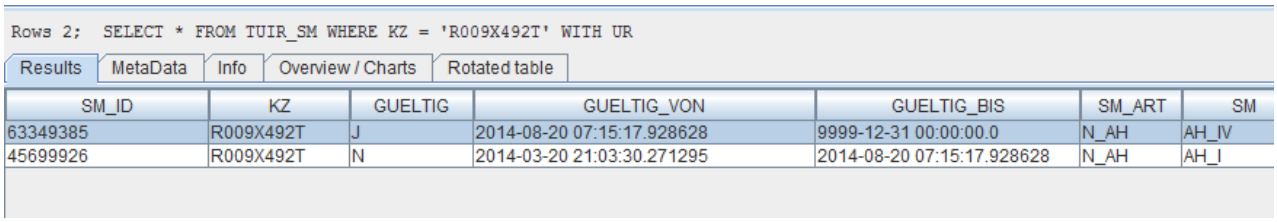
Picture 3 – example of the monthly historiography in Statistics Austria's BR

Picture 3 shows the monthly historiography in the business register. In principle all units may be altered as long as the structural business statistics for the respective year is not completed. The picture shows the above mentioned unit which is only relevant for the register since January 2016. Any time a characteristic is modified, the colors of the month buttons change. This makes the daily maintenance tasks easier for the responsible persons.

As not all changes are valid from that month on, when the maintenance operation takes place, each time a user alters a unit, he/she has to set the month (for characteristics which are saved monthly bitemporal), from which onward the change was effective. Maintaining for future time periods is not possible in the business register of Statistics Austria.

Whenever a maintenance operation takes place, that exact moment is timestamped. Hence it is possible to find out at any given time how any unit and its respective characteristics did look like before AND after the dataset was changed.

System time



Rows 2; SELECT * FROM TUIR_SM WHERE KZ = 'R009X492T' WITH UR

Results | MetaData | Info | Overview / Charts | Rotated table

SM_ID	KZ	GUELTIG	GUELTIG_VON	GUELTIG_BIS	SM_ART	SM
63349385	R009X492T	J	2014-08-20 07:15:17.928628	9999-12-31 00:00:00.0	N_AH	AH_IV
45699926	R009X492T	N	2014-03-20 21:03:30.271295	2014-08-20 07:15:17.928628	N_AH	AH_I

Picture 4 – a characteristic which is not saved on a monthly basis

Picture 4 shows an example for a characteristic which is not covered monthly. Here the image within the database is a little different compared to the monthly bitemporal historiography. The characteristic "SM" with parameter value "AH_I" (unit is currently in AH_I survey) was valid for unit R009X492T from 2014-03-20 until 2014-08-20. Since the exact moment when "AH_I" became invalid, "AH_IV" (unit is currently part of AH_IV survey) is now the new parameter value. To be able to easily differentiate between currently valid and invalid values, an additional characteristic is implemented: GUELTIG – which gives no more information than "this value is currently valid in the database, YES/NO" – which makes queries a lot easier than having to input long timestamps in order to get back the correct figures.
Other examples for characteristics which are saved with system time-stamps only are the current values of VAT and employees – which are overwritten each month. However, tax and social security send their data in a way, in which we can easily derive which value was valid in each month.

Excursus: self-programmed vs. automatic historiography

The development of the new maintenance application started in the year 2008. The database system on which the application was based did not support an automatic bitemporal historiography. As this was one of the most favoured desires in the development, the functionality had to be programmed by Statistics Austria IT specialists.
In the self-programmed version each maintenance operation has its own database entry ID. Whenever a characteristic is changed, the "old" datasets system ending time is set to the moment of the operation. At the same time a new dataset with a new beginning date and open ending date is set.

In 2012 Statistics Austria switched to DB2-10. This system allows in principle a bitemporal historiography by its own design – it is setting time-stamps for each maintenance operation. In contrast to the Statistics Austria self-programmed version of bitemporal historiography, DB2-10 uses a so called "system historiography table" where all invalid values of any characteristic are stored. However, as DB2-10 uses a different approach of logging than the self-programmed version, and as the implementation of the new maintenance system was already nearly finished, switching to the auto-bitemporal historiography

of DB2-10 was not possible without major drawbacks. Furthermore DB2-10 has a business historiography which is pre-set to one day; as a monthly historiography is completely sufficient for Statistics Austria requirements and as the current solution is working fine, we will not change our approach for now.

Conclusion

Having switched to a monthly historiography allows Statistics Austria to present a more correct picture of the Austrian economy. Units, which are only active for a few months will now show as being active for the respective period, and will not show as being inactive, just because they were not active at the last month of the year – this is a huge improvement in comparison to the old Austrian Business Register.

Units, which are only active for a few days within one month will show as inactive for the whole month, however, due to the bitemporal historiography it is possible to derive which units were active at any given date in one month. Therefore one can theoretically query at any moment in time how many units where active or inactive or had a specific characteristic.