Towards a register-based census post 2021 in Germany: objectives, requirements and challenges

Note by the Federal Statistical Office of Germany*

Summary

While in the 2020 census round Germany continues to apply the combined model initially developed for the 2011 census, preparations towards a fully register-based census post 2021 have already started. This paper outlines the objectives, requirements and challenges of a change-over to a register-based census model covering the census areas of population, housing and dwellings, households and families, labour market, and educational attainment. Illustrated by selected examples, the paper discusses the current ideas in the fields of data provision (e.g. creating new registers or enhancing existing ones), quality assurance (e.g. to deal with issues of over- and under-coverage), and the establishment of the necessary record linkage procedures.

* Prepared by Thomas Körner (Thomas.koerner@destatis.de) and Eva Grimm (eva.grimm@destatis.de).
I. Introduction

1. Population and housing censuses are currently facing profound changes in terms of what is required of them. As shown by recent discussions at the European level (Eurostat 2017, 2018), which are also reflected by the recommendations of the German Statistical Council (2018) and the German Data Forum (2016), users are requesting census data more frequently, in a more timely manner, and with more detailed regional breakdowns. In 2021, a combined census will be conducted in Germany. Although the current methodology of a combined census was already successfully implemented in 2011, it is not going to fulfil the future requirements of population and housing censuses after 2021. For this reason, plans have been launched for a change-over to a fully register-based census model after the next census. This is in line with Eurostat’s plans for a fundamental redesign of European census-type statistics after the 2021 census round towards an increased use of data from administrative sources (Eurostat 2017).

2. By implementing a register-based census, the post-2021 censuses in Germany could provide grid-based results, results at higher frequencies and at a higher flexibility to respond to changing user needs. At the same time, cost as well as burden on the population could be reduced through an increased use of already existing administrative data compared to the combined approach. However, introducing a register-based census model in Germany is not straightforward for several reasons. The introduction of a register-based census requires access to relevant administrative data, high data quality, methods for quality assurance and the possibility to link data from different registers. Compared to countries that have already adopted register-based approaches, fewer suitable registers are available in Germany. Another challenge for the introduction of a register-based approach is the existence of strict data protection regulations that were established in the context of the last traditional German census in 1987 which make it challenging to link registers from different areas (Koerner/Dittrich 2017).

3. This contribution presents the current considerations of a change-over to a register-based census model (section II) and discusses basic requirements and challenges (section III). Regarding the development of solutions for some of these challenges, the paper discusses current ideas in the fields of quality assurance of register data (e.g. how to deal with issues of over- and under-coverage in the population register), the creation of supplementary registers for data provision (e.g. creation of a register on buildings and dwellings), and the establishment of the necessary record linkage procedures in the absence of a constant person identification number (section IV).

II. Requirements of a future proof census model

4. The implementation of a register-based census in Germany is driven by several goals which are presented in the following paragraphs (Körner/Krause/Ramsauer, 2019).

5. Grid-based results. The full potential of the census for local and regional decisions can only be tapped if data are provided in grid-based form as this also allows variable tabulations for non-administrative areas. This is reflected in current legal developments at the European level. As the data needs for the European Commission’s policy programmes are currently changing, the European Commission has initiated a new Framework Regulation for Census and Population Statistics, entitled European Statistics on Population (ESOP). This framework regulation is intended to establish the compulsory census data set of the European Union after 2024 and to replace the current regulations on census (Regulation No. 863/2008), population statistics (Article 3 of Regulation No. 862/2007) and migration statistics (Article 3 in Regulation No. 862/2007). The core requirement of the new framework regulation is the annual provision of grid-based results of the population figures from the reporting year 2024 onwards. Currently, census data on population and housing refer to administrative areas. However, the potential of the census can only be tapped for local and regional decisions regarding population and housing if data are provided in grid-based form, since this allows variable tabulations also for non-administrative areas.
6. **Higher frequency.** A census that is carried out under a combined census model constitutes a large-scale operation. For this reason, it can only be conducted at long intervals. However, in all topics of the census, there are variables for which the current frequency of ten years is considered too long by many users. Providing annual results for a limited subset of census variables could help to remedy this concern. For the time after the 2031 census, it is thus planned that results for several variables should be provided more frequently (monthly or yearly) which can only be accomplished by a register-based census model.

7. **Higher flexibility.** With data needs changing over time, data programmes will have to be more flexible in the future. In the current German system, intercensal population updates are used in order to determine the annual population figures in between the census years. As the method of intercensal population updates always requires a census as a baseline, it can only provide results for variables and breakdowns which were included in the census base. New variables or breakdowns cannot be introduced before the next census. In a register-based census approach, new variables can be added and analyzed more flexibly compared to intercensal updates that are relatively fixed between two census rounds.

8. **Shorter production times.** The size of the operations that are used for data collections in the combined census approach goes along with rather long production times. Due to this reason, results can only be produced with some delay. As the experiences of countries with register-based censuses illustrate, production times can be shortened by using a register-based approach.

9. **Reduction of cost and burden.** Interviewer-administered data collection is not popular due to its cost and due to the burden it imposes on respondents. Compared to traditional censuses, the combined census model in Germany has already led to significant cost reductions. However, it still includes components related to traditional data collection which are associated with rather high costs. The current combined census model includes a supplementary household sample survey and a complete enumeration of buildings and dwellings that go along with high investments of time and effort, as more than 8 million inhabitants and about 25 million owners of buildings and dwellings have to be surveyed. Several institutions have questioned whether the conduct of surveys is still justified for the purposes of the census. For example, the National Regulatory Control Council (2016, 2019) has advised to increase the use of existing administrative data sources in the census. The German Federal Constitutional Court has highlighted that collecting data from already available administrative sources is preferable to conducting surveys due to its lower intervention with fundamental constitutional rights of the population (Leischner/Bierschenk 2019). The basic idea of a register-based census is to replace the direct collection of data from the population with the use of administrative data held in existing administrative and statistical registers. It thus entails a reduction of burden on the population. Experiences from countries with purely register-based censuses indicate that significant cost reductions can be achieved (UNECE 2014).

III. **Basic prerequisites and challenges**

10. The introduction of a fully register-based census presupposes that a number of preconditions are met regarding register access, data quality and the possibility of linking the records from several registers. These requirements are discussed in the following paragraphs, referring to challenges that are imposed in Germany by tight legal restrictions, an incomplete register infrastructure and coverage issues in administrative data.

11. **Data quality.** To be used for statistical purposes, registers need to be of sufficient quality and approaches need to be developed to correct the data for errors. This is of particular importance in the case of administrative registers, which are usually not kept for statistical purposes and therefore often need data pre-processing. For example, as demonstrated by the results of the 2011 census, the population registers which constitute the main data source for the German population figures are subject to significant over- and under-coverage. In this paper (section IV.A), possible methods for the identification of over- and under-coverage in the population registers are discussed as an example.
12. **Data access in the context of an incomplete register infrastructure.** In order to move towards a register-based census, the necessary data in the fields of demography, buildings and dwellings, household and family types, and labour market and education must exist and be accessible to producers of official statistics. There are fewer suitable registers available in Germany than in countries which have already adopted a register-based approach. While a significant part of the information of interest is already available in registers, some still need to be created. For instance, neither nationwide registers on buildings and dwellings nor on the educational attainment of the population exist currently. This paper (section IV.B) discusses the possibility of creating a register on buildings and dwellings as an example.

13. **Procedures to link registers.** The development of suitable approaches to link data from different registers is a basic precondition for a register-based census. On the one hand, this is necessary as not all the variables required for a population and housing census are available in one single register. On the other hand, reliable and easily implementable linkage procedures are also required for quality assurance (e.g. in order to check for administrative signs of life) as well as to analyze the consistency of information included in more than one register. For a register-based census, two objectives of record linkage are of particular importance. First, at the level of persons, it must be possible to link the data of persons across several registers, ideally supported by a constant identifier. Second, for production of data on housing, household and family types, approaches need to be developed to link the persons to the dwellings they live in by the use of a building and dwelling identifier. Further units requiring record linkage procedures include addresses and institutions. In this paper (section IV.C) the possibility of linking register data without a unique identification (ID) number is discussed as an example.

### IV. Selected issues

#### A. Treatment of over- and under-coverage in the population register

14. Providing accurate data on the size and demographic structure of the population is one of the main tasks of a census. In Germany, the municipal population registers provide the main data source for the population figures as each person residing in the country is obliged to be registered in the population register of the municipality in which he/she lives. However, accurate population figures cannot be derived simply by counting persons that are registered, as the population registers are subject to significant over- and under-coverage issues. Since the population registers are updated by reported demographic events such as births or changes of residence, errors can arise, e.g. if people do not comply with their obligation to register and deregister when changing their place of residence. In the 2011 census, about 2.1 million cases of over-coverage and 1.3 million cases of under-coverage were detected by a sample survey and about 600,000 cases of duplicates by duplicate detection methodology (Bechtold 2013; Bund-Länder-Arbeitsgruppe “Einwohnerzahlen”, 2016: 8).

15. Different types of over- and under-coverage in the population registers can be distinguished (Table I). First, over-coverage can relate to persons who have died or persons who moved their main residence or centre of life to another country without reporting it to the authorities. Over-coverage can also be associated with duplicate records of persons who are registered more than once in several municipalities. Without correction, such cases would be erroneously counted in the population. Second, under-coverage can arise if persons who are born in or immigrate to Germany are not registered in the municipal population registers. Under-coverage can also be related to persons who have moved to another municipality and do not register in the new municipality but are deregistered “to unknown” (e.g. due to returned election documents) by the municipality of their former place of residence.

---

1 In the 2011 census, the total population figure of about 80.2 million inhabitants was derived after correction for cases of over- and under-coverage that were determined in the sample survey, duplicates that were determined with a duplicate detection methodology, persons living in special institutions who were determined by a complete enumeration survey and a clarification of discrepancies.
Moreover, under-coverage can also result if persons are registered only with secondary but no main residence in the population registers. Without correction, these persons would not be counted in the population although their actual place of residence is located in the country. Thirdly, there are cases of complementary over- and under-coverage that can arise if persons move their main residence or centre of life to another municipality within the country but do not report this change of residence to the authorities. Without correction, these persons would be counted in the population of a municipality in which they do not live any more (over-coverage) while they would not be counted in the population in the municipality they actually live in (under-coverage).

Table 1

Types of over- and under-coverage in the population registers

<table>
<thead>
<tr>
<th>Type of coverage error</th>
<th>Description</th>
<th>Consequence if not corrected</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Over-coverage in the population registers</strong></td>
<td>Omitted de-registration after death</td>
<td>Persons would be counted although they do not live in the country</td>
</tr>
<tr>
<td></td>
<td>Omitted de-registration when emigrating to another country</td>
<td>Persons would be counted in Germany as well as in another country</td>
</tr>
<tr>
<td></td>
<td>Omitted de-registration when moving the centre of life (main residence) to another country</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Registration at more than one sole or main place of residence.</td>
<td>Persons would be counted several times within the country (duplicates)</td>
</tr>
<tr>
<td><strong>Complementary over- and under-coverage</strong></td>
<td>No change of registration when moving within the country</td>
<td>Persons would be counted in the wrong municipality</td>
</tr>
<tr>
<td></td>
<td>No change of registration when moving the centre of life within the country (e.g. exchange of main residence with secondary residence)</td>
<td></td>
</tr>
<tr>
<td><strong>Under-coverage in the population registers</strong></td>
<td>Omitted registration after birth</td>
<td>Persons would not be counted although they live in the country</td>
</tr>
<tr>
<td></td>
<td>Omitted registration when immigrating from another country</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Omitted registration of persons who are deregistered “to unknown” although still living in the country</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Registration only with secondary residence</td>
<td></td>
</tr>
</tbody>
</table>

16. The population figures of the 2021 census will be based on a combination of data from the population registers and a survey comprising about 10 per cent of the population. In Germany, the maintenance of the local population registers is carried out by the respective municipalities. For the purpose of the census where cases of duplicates and persons that are registered only at secondary residences have to be identified, the register data from all municipalities will be merged into a single nation-wide statistical population database.

---

2 For example, if students move out of their parents’ home to another city and do not report their new residence to the authorities.
Based on this central database, duplicates can be determined by means of a duplicate detection methodology. Since this procedure does not necessarily require any surveys, it will be possible to adopt the principal approach and implement the automated duplicate detection methodology in the register-based census. However, the remaining types of over- and under-coverage will be identified via a sample survey in the 2021 census. Hence, for the transition to a fully register-based census after 2021 it will be necessary to develop a new methodology instead of the sample survey to correct for these errors.

Suitable ways to correct for these errors in a register-based approach are currently being explored. In order to identify over-coverage, the “signs of life” approach is considered, which is already used by other countries. The basic idea is to identify potential errors by linking personal records in the population registers with personal records in other registers that possibly contain administrative signs of life (comparison registers). Personal records in the population registers that do not relate to any sign of life in other administrative registers are considered potential cases of over-coverage. A statistical correction procedure can be used to decide whether a person who is registered in the population registers should be counted according to his or her presence in different administrative registers. For further clarification, persons who are considered potential cases of over-coverage can be contacted in order to verify their actual residence status (Austria, Spain).

When conceptualizing the signs of life approach for a German register census, the following fundamental questions need to be addressed:

- **Linkage of datasets.** A main challenge is the linkage of records of the central population dataset with records of the other registers without a personal identification number. An identifier for persons can facilitate the identification of matching records in administrative registers. However, as mentioned earlier, there is no universal personal identifier available in the German register landscape so far (see section IV.C).

- **Methods to identify under-coverage.** In addition to the detection of “pure” over-coverage, use of the signs of life approach is also being considered to detect cases of “pure” under-coverage as well as cases of complementary over- and under-coverage. By linking data from administrative registers with the population register data, potential cases of under-coverage might be identified if persons have left signs of life in administrative registers at a different place than in the population registers. A main challenge of this approach is to assign identified cases of under-coverage to their corresponding municipalities and, as geocoded population figures will be required after 2024, ideally to the correct grid cells. This requires the matching of local area codes or addresses. In order to achieve an efficient correction, local area codes (or addresses) of high quality and timeliness have to be available in the administrative registers which provide the information on signs of life. In order to determine the potential of the signs of life approach to detect under-coverage as well as over-coverage, a pilot study will be necessary. Alternatively, other methods for the identification of under-coverage will have to be found.

- **Identify suitable registers.** Another challenge of applying the signs of life approach in Germany is the identification of suitable registers which can be used for potential signs of life. In order to assess the eligibility of potential comparison registers, pilot studies will be necessary. The concept of such a study may include the linkage of records from population registers with those of potential comparison registers in order to assess potential cases of over- and under-coverage. Ideally, the cases thus determined might be compared to cases of over- and under-coverage that are found in the sample survey of the 2021 census.

- **Implementation within the administrative or the statistical domain.** There are two options currently being discussed as to where the signs of life approach could be implemented. The first and preferred option is to implement the signs of life approach within the administrative system. This idea refers to recent discussions of a modernization of the German register system, which includes the implementation of a basic administrative register of personal data. This register would be used as basic
data stock for the purpose of identity management in various administrative processes. If the signs of life approach were to be implemented as a regular method of quality assurance within this basic register, its results could be used for the purpose of statistics as well as for the purpose of directly correcting over-coverage and under-coverage in the original data sources, including the administrative population registers. However, both the time of its implementation and the explicit design of the basic register have yet to be determined. For this reason, it is not yet clear whether it will be available in time for the transition to a register-based census which is required before 2024. The second option is to implement the signs of life approach within the statistical domain as an interim solution. However, this option would not allow any feedback about identified cases of over- and under-coverage to the administrative population registers, due to statistical confidentiality requirements. However, it might be a transitional solution until the implementation of the administrative base register.

### B. Creation of a register on buildings and dwellings

19. Whereas in some areas, e.g. in the field of population data, existing administrative registers can be used as a basic data source for the construction of a register-based census, this is not the case in others. The issue of data on buildings and dwellings is discussed in this paper as a case in point. Besides the provision of data on the size and demographic structure of the population, providing data on buildings and dwellings and on the housing situation of the population is the second main task of a census. Currently, data on buildings and dwellings are collected in a complete enumeration among all 25 million building and dwelling owners. Collecting these data with a register-based approach would not only reduce the response burden considerably but may also create the potential to provide this type of information more frequently than every ten years.

20. However, to date there is no nation-wide register of buildings and dwellings in Germany that the census may use. Several options are currently being examined in order to nevertheless be in a position to collect information on buildings and dwellings using a register-based approach. The favourite option is the use of a land property database which is currently built by the German tax authorities for the purpose of collecting data for the ongoing reform of the property tax in Germany. As the variables needed for determining the property tax have not yet been defined, and as there is considerable variation regarding the different models under discussion, it cannot be determined today whether the land property database might be a suitable basis for collecting building and dwelling information. Even if one assumes that the model containing the most information relevant for the census is chosen, it still seems probable that a supplementary building and dwelling register containing the missing information will be set up. Another issue that would have to be solved is that the concepts used in the land property database are those required for tax purposes and only partly match the requirements of the census. This concerns the statistical units (e.g. buildings vs. economic units, the latter of which may comprise more than one building) as well as variables and items. It remains to be examined whether the definitions can accommodate the census needs and to what extent approximations of the variables of interest of the census might be possible.

21. If the land property database turns out not to be sufficiently appropriate to be used for census purposes, another option might be to create a (supplementary) building and dwelling register. Since the use in the census requires, for various reasons, the creation of an administrative register (and not a statistical register), such a register could equally be used as a data basis for various administrative purposes including city and regional planning processes, and the monitoring of land and dwelling market developments. Although the creation of a basic register of buildings and dwellings has been required on several occasions recently, the discussion about possible uses and users is still at an early stage. Such a register could also provide an opportunity to improve the data availability regarding commercial buildings.
22. When conceptualizing a building and dwelling register, the following three fundamental issues need to be addressed:

- **Collection of the information on the stock of buildings and dwellings.** A procedure needs to be defined for how the data on the initial stock of buildings and dwellings can be obtained.

- **Updating and maintenance of the register.** Once the initial stock is in place, a procedure needs to be established for an ongoing update of the register information. The update process needs to be in place directly following the collection of the stock data, including the definition of data transmission standards, in order to permit automatic updating of the register in the various authorities concerned.

- **Development of interfaces.** Technical interfaces need to be provided to establish a regular data exchange.

23. In addition to these aspects, a building and dwelling identification number is required for easy identification and updating of each of the units in the register. Such an ID number needs to follow a common logic and should ideally at least roughly indicate the position of the different dwellings inside a building. Furthermore, building and dwelling IDs need to be linked to the persons in the population register in order to allow statistical household generation as well as derivation of variables pertaining to the housing situation of the population (e.g. floor space per person).

C. **Linking register data without a unique ID number**

24. Establishing appropriate procedures to link data from different registers is a basic precondition of a register-based census. This requirement is not only due to the fact that not all of the variables of interest can be found in one single register. It is also necessary because quality assurance will have to rely upon comparisons between the registers, e.g. by the use of administrative signs of life.

25. Linkage of records from different registers can be achieved principally either by the use of a constant identifier available in all registers under consideration or by the use of identifying variables (e.g. name and date of birth) available in all registers to be linked. Constant identifiers are the most suitable, to ensure that different registers can be linked in a biunique way and that identifiers are unchangeable over time. Assuming that a central coordination is established to assign the identifiers, it is also highly likely that constant identifiers will achieve the most reliable linkage, minimizing false negatives as well as false positives. To meet the requirements of information technology (IT) security and data protection, constant identifiers can be encrypted (e.g. by using cryptographic hash functions) in order to limit the possibility of generating profiles of the units available in the registers.

26. In Germany, constant identifiers are at present available in single administrative spheres only and their use is legally limited to purposes within these spheres. For example, the tax identification number is a constant identifier used for various taxation purposes, but it must not be used for other purposes and is only kept in registers of the tax administration (except for the population registers, in which it is however limited to the use for taxation purposes, too). While there are ongoing discussions about introducing identity management based on a system including core personal variables and an identifier to be established across all registers, using a constant identifier to link registers for a register-based census is not possible in a short-term perspective.

27. Therefore, the Federal Statistical Office is currently exploring the possibilities of linking registers based on person variables such as the names, date of birth, place of birth etc., which might be used in a pseudonymized form in order to fulfill data protection requirements. Problems accompanying this alternative include the following: links might be equivocal if the data available do not allow a sufficient differentiation of the units included in the registers under consideration (e.g. if more than one person has the same names and date of birth). The reliability of the record linkage is limited by the fact that available person data might be subject to change over time. Even if names, dates of birth, and also sex are mostly stable person variables, their register entries can be changed over time. Also, the
degree harmonization of the shared variables in different registers may be limited, e.g. if different standards are applied to transcribe foreign scripts or different standards of character encoding are applied. Finally, the record linkage procedures have to be applied at every single data transmission required for the register-based census, leading to an increased effort at the statistical office or the unit in charge of linking the registers. Despite these obvious drawbacks, the Federal Statistical Office is currently exploring possible ways to establish record linkage on the basis of person variables for an interim period until constant identifiers can be made available in all registers required for conducting a register-based census.

28. A combination of constant identifiers and variables identifying person variables could also be envisaged, e.g. for an interim period. The precondition of such a combination would, however, be that at least one of the registers includes the person variables as well as the constant identifiers to be used for record linkage.

29. Besides the objective of providing population figures, censuses usually also aim to provide data on the housing stock, the housing situation of the population, and key variables related to household and family structures. In particular for the latter two objectives it is essential to have the possibility of linking the persons included in the population register with information on the dwellings in which they actually live and the other persons with whom they share their dwelling. Furthermore, persons in institutional households need be identified in the population count to be distinguished from persons living in private households. In order to achieve this requirement, a building and dwelling identification number needs to be introduced. While persons covered in several registers can be linked both by using a constant identifier and by using person characteristics (or a combination of both), in the case of buildings and dwellings only a constant identifier can be used. In the population register on the one hand and the building and dwelling register on the other, there is no overlap of information on which a record linkage procedure could be based. Therefore, the idea is currently being discussed of keeping the building and dwelling identifier in the building and dwelling register and subsequently introducing the identifier into the population register, so that all inhabitants can be assigned to one dwelling in the registration process. The details of such an approach, however, still need to be developed in further detail.

V. Conclusion

30. Changing user requirements lead to profound changes in census and population statistics. Providing results more frequently, in a more timely manner, more flexibly and with greater regional breakdown, will, at least in a long-term perspective, require the change-over to a register-based census. With further efforts to make progress in the digitization of public administration, the register infrastructure in Germany is under constant development. This development may lead to further harmonization and linking of registers, which could open up new opportunities for the use in official statistics and in particular the census.

31. Setting up a register-based census is model is a challenging undertaking, even in a country like Germany, where a combined census model has already been in place since 2011. The preparations, including access to existing registers and the creation of new registers, the creation of procedures to link register data in compliance with data protection requirements, and the development of appropriate quality assurance procedures, require an early start. First considerations for the census post-2021 therefore already started well in advance of the 2021 census.

32. This paper has presented examples of on which work is currently underway, i.e., the treatment of over- and under-coverage in the population register, the creation of a register of buildings and dwellings and the linkage of register data without a unique ID number. An exchange with other countries that are currently working on similar topics would be welcome.
References


National Regulatory Control Council (2019); Stellungnahme des Nationalen Normenkontrollrates gem. § 6 Absatz 1 NKRG. Entwurf eines Entwurf eines Gesetzes zur Durchführung des Zensus im Jahr 2021 (NKR-Nr. 4684, BMI).

