

Swiss Municipal Data Merger Tool: Open-Source Software for the Compilation of Longitudinal Municipal-Level Data*

Valentin Knechtl[†] and Alois Stutzer[‡]

September 13, 2021

Abstract

The Swiss Municipal Data Merger Tool (Swiss MDMT) offers a solution to a frequent data management problem encountered when compiling longitudinal datasets involving Swiss municipalities as the observational units. Due to municipal mergers, the number of municipalities in Switzerland declined from 3,095 in 1960 to 2,202 in 2020. As a consequence, manually securing the correct spatial reference when merging historical cross-sectional data is tedious and time-consuming. To facilitate this operation, the Swiss MDMT considers mutations at the municipal level and maps municipalities of a first point in time to municipalities in a second point in time based on information provided by the Swiss Federal Statistical Office's municipality inventory. The tool is distributed as an open-source R package and is freely available on CRAN.

Keywords: Data management, historical data, municipal research, open-source software, spatial reference

*We are grateful to Ulrich Matter, Kerstin Hansen, Kurt Schmidheiny, Andreas Ladner and Christian Sigg for helpful comments. We acknowledge financial support from the Swiss National Science Foundation and its National Center of Competence in Research - The Migration-Mobility Nexus.

[†]University of Basel, Faculty of Business and Economics, Peter Merian-Weg 6, 4002 Basel, Switzerland. Email: valentinknechtl@gmail.com.

[‡]University of Basel, Faculty of Business and Economics, Peter Merian-Weg 6, 4002 Basel, Switzerland. Phone: +41 (0)61 207 33 61, email: alois.stutzer@unibas.ch. (Corresponding author)

1 Introduction

Data at the municipal level constitute an important input for statistical analyses that investigate societal phenomena expected to differ across space. For applied statistical work, the change of geographical units over time thereby poses a major challenge if data are to be aggregated or compared longitudinally. In the case of Switzerland, mergers of jurisdictions at the municipal level are of first-order importance. The number of municipalities in Switzerland has decreased significantly in recent years; while there were still 2,899 municipalities in 2000, the corresponding number declined to 2,202 by 2020.

This research note introduces the Swiss Municipal Data Merger Tool (Swiss MDMT), an open-source R package that allows researchers to merge municipal level information from different points in time and to appropriately adjust the spatial references (Germ. Raumbezüge) so that a phenomenon is measured for the same spatial units over time. The software makes use of the Swiss Federal Statistical Office’s (SFSO) municipality inventory (Germ. Historisiertes Gemeindeverzeichnis), which is a database that contains standardized and up-to-date information about all mutations at the municipal level in Switzerland starting from January 1, 1960 (Verein eCH 2014; Bundesamt für Statistik 2017).¹ Longitudinal datasets are often provided with a unified spatial reference by the data owner.² In contrast, the issue of differing spatial references often occurs when an analysis involves data from different data sources. Especially cross-sectional data are normally provided with the spatial reference of the point in time of collection, as the primary objective is to capture a particular phenomenon within the spatial structure at that time. For instance, municipal statistics about citizens’ voting behavior or government finances are recorded for all the existing municipalities at a specific point in time. However, in secondary data analysis, researchers often aim to study developments over time, for example, the *evolution* of issue-specific voting behavior³, or local public finances⁴. In fact, there is huge potential for the analysis of local contexts based on the Swiss municipal structure (Horber-Papazian 2007, Ladner 1991, Steiner et al. 2014). This is further strengthened by the freely accessible information on municipal-level politics from Andreas Ladner’s Communal Secretary Survey that started in 1988 (see, e.g., Haus and Ladner 2020, Ladner 2003).

¹The association eCH is an entity that is closely related to the organization “e-government Schweiz”, which in turn consists of the Swiss Confederation, the Conference of Cantonal Governments (“Konferenz der Kantonsregierungen”) and the Swiss Association of Municipalities (“Schweizerische Gemeindeverband”).

²For example, the Swiss historical census data are linked to the spatial reference of December 5, 2000.

³Municipal-level data on voting behavior in decisions on national referendums and initiatives are, for example, analyzed in Baccini and Leemann (2020), Brunner and Kuhn (2018), Eugster and Parchet (2019), Meier et al. (2019) or Slotwinski and Stutzer (2019).

⁴There is a large literature in public finance that analyzes local income and corporate taxes as well as local government expenditures (see, e.g., Brülhart and Jametti 2006, Cappelletti and Soguel 2013, Galletta 2020 or Luthi and Schmidheiny 2014).

For analyses of developments at the municipal level over time, however, the appropriate and comparable spatial reference first has to be secured. One practice is to restrict the data sample to only those municipalities that were not involved in a mutation over the entire observation period. The spatial units are thus clearly comparable. However, with this approach, a possible selection effect is accepted because the subsample of municipalities that were not involved in a merger might be systematically different from the subsample of municipalities that underwent a transformation. In particular, the latter municipalities often tend to be smaller (and with regard to the aforementioned examples, might also be more rural and more likely to face fiscal difficulties). Another approach is to neglect the differing spatial references and merge municipalities in a “naive” way by using the SFSO municipality number available at any given point in time thereby accepting, first, structural breaks in the data series for the municipalities that emerged from a merger and, second, the exclusion of data from no longer existing entities. A more appropriate approach could be to combine those historical spatial units that refer to a given spatial unit in the last observational period. This approach works for municipal level datasets which fulfill the necessary property that the spatial aggregation is technically feasible. This holds, for example, for results of popular votes. Corresponding datasets contain information on the number of yes and no votes for all the municipalities that exist at the time of the vote. For example, let us assume that there are two municipalities A and B at time t_0 (i.e., the historical spatial units) and they merge to form municipality C at some later point of time t_1 (i.e., the spatial unit in the last observational period). Furthermore, let us suppose that we are interested in the yes-share in this new municipality C of a given popular vote which took place at t_0 . Thus, the yes-share in C is the sum of yes votes of A and B divided by the total sum of the yes and the no votes of A and B. In other words, the yes-share in C is the weighted average of the yes-shares of A and B. In sum, we suggest a programmatic solution and offer a corresponding tool that automates this latter approach, as it is tedious and time-consuming if performed manually. In the end, however, which of these approaches is best suited depends not only on the technical feasibility but also on the given research question.

The development of the Swiss MDMT relies on an official and well-documented data standard (called eCH-0071 Datenstandard Historisiertes Gemeindeverzeichnis der Schweiz) for the municipality inventory (see Verein eCH 2014). The existence of this e-government standard influences the tool development in two ways: On the one hand, it allows for a straightforward implementation of the mapping algorithm, and thus facilitates tool development. On the other, the mere existence of this data standard suggests a reasonable degree of continuity with regard to the data model utilized, which is a decisive factor in spurring additional initiatives.

Our tool fits and supports the more general enterprise in data science to make more and more data from an increasing range of sources publicly available. For quantitative analyses, the stylized data pipeline commonly consists of, first, getting hold of the raw data, second, accessing the

data sources and third, preprocessing data from potentially different data sources into the final dataset, ready for the actual analysis (for a discussion of this concept, see Matter 2019 or Sebei et al. 2018). Along this data pipeline, various cost-cutting developments can be observed: firstly, availability and straightforward accessibility of various data sources are enhanced by open data initiatives like the Swiss Open Government Data platform⁵ (for a general account, see Bundesblatt 2019 and Bürgi-Schmelz 2019). Secondly, accessing public databases is facilitated by open source interfaces (OSI) which provide high-level access to data from a web service by generating a flat dataset ready for statistical analyses from structured data available at that web service (Matter and Stutzer 2015; Matter 2018). One example of an existing OSI is the R package `pvsR` (Matter 2014), which is an OSI for accessing data made available by Vote Smart, a research organization that gathers data about various topics in U.S. politics. Thirdly, a tool like the Swiss MDMT helps to reduce effort during the preprocessing stage. To summarize, all these diverse developments along the data pipeline help facilitate empirical research.

In Section 2, we describe the Swiss MDMT, which is available as an R package via CRAN (see Knechtl 2020) and show a code example. The advantages of this approach compared to a naive merge are briefly illustrated in Section 3. Section 4 concludes.

2 Program functionality

In this section, the Swiss MDMT's core functionality is presented. The program automatically detects and maps municipalities that existed at a given point in time in the past to municipalities that exist at a later point in time. The program interface simply requires the dates of these two points in time. In response, the program returns a mapping table, i.e., a table which links the former set of municipalities to the latter set of municipalities. The underlying mapping algorithm exploits the information available in the municipality inventory.

Based on an example, we illustrate how the Swiss MDMT's mapping table is used in a complete merge.⁶ The example addresses the question of how the electoral support for the facilitated naturalization of third generation immigrants in Switzerland evolved over time at the municipal level. To empirically assess this phenomenon, we use data from two referendums on this topic that took place in 1983 and 2017. Naturally, the spatial reference of these datasets on the vote outcome at the municipal level differs, as each dataset is based on the municipal structure of the respective point in time. The first dataset (vote 1) has the spatial reference capturing the situation as of the 4th of December, 1983, and the second dataset (vote 2) the one as of the 12th of February, 2017.

⁵See <https://opendata.swiss>.

⁶The differences that emerge between a complete and a naive merge are presented and discussed in the next section.

To calculate the change in the yes-share over time and per municipality, the following three steps are necessary: firstly, the spatial reference of the earlier dataset needs to be adjusted to that of the later one using the Swiss MDMT's mapping table. Secondly, on this basis, the yes-share of vote 1 is calculated for the same spatial reference as vote 2. Thirdly, the two datasets with the yes-share are merged and the changes in the yes-share are calculated.

The following code examples in R are presented so that they can be easily understood by researchers who have some basic knowledge in R. If the tool is being used for the first time, the package needs to be installed from CRAN.

```
install.packages("SMMT")
```

Furthermore, in a new R session, the package needs to be loaded.

```
library(SMMT)
```

If the municipality inventory database is not yet available locally, it is obtained with the following command, which fetches the database and stores it in the current working directory of the R process. This command only needs to be executed once as it makes the municipality inventory locally accessible.

```
path_inventory <- download_municipality_inventory()
```

Once the R package is installed and the municipality inventory downloaded, the import function imports the database as an R object into the R environment.

```
mutations_object <- import_CH_municipality_inventory(path_inventory)
```

The mutations object is then used considering the two dates to obtain the mapping table.

```
old_state <- as.Date("1983-12-04")
new_state <- as.Date("2017-02-12")
mapping_object <- map_old_to_new_state(mutations_object$mutations,
                                       old_state, new_state)
mapping_table <- mapping_object$mapped
```

Table 1 displays an excerpt of the program output. The first record, i.e. Wiliberg, shows an unaltered municipality. Records two and three capture municipalities that were part of a merger, i.e., the municipality of Zofingen incorporated the municipality of Mühlethal in 2002, which, in turn, no longer exists.

Table 1: An excerpt of the mapping table which is the main output of the Swiss MDMT.

New state (February 12, 2017)		Old state (December 4, 1983)	
SFSO number	Municipality name	SFSO number	Municipality name
⋮	⋮	⋮	⋮
4288	Wiliberg	4288	Wiliberg
4289	Zofingen	4289	Zofingen
4289	Zofingen	4278	Mühlethal
⋮	⋮	⋮	⋮

This mapping table is the key element in performing a complete merge of the datasets in question. To exhibit the steps involved, let us assume that both popular vote datasets are available in the R environment and named as follows: `vote_1_1983` and `vote_2_2017`. The two tables contain the following columns: `yes_votes` and `no_votes`. Furthermore, the former dataset contains a column `bfs_nr_old`, whereas the latter contains a column `bfs_nr_new`. Note that for this code example, we use the functionality of further R packages which are thus loaded as well.

```
library(tidyverse)
```

To aggregate the data of the earlier vote, the following steps are required: first, municipalities that merged are grouped together. The mapping table contains the relevant information about which municipalities existing in December 4, 1983 belong together on February 12, 2017. Second, the grouped yes and no votes are added up. Third, the yes-share of vote 1 is calculated with the same spatial reference as vote 2.

```
vote_1_2017 <- left_join(mapping_table, vote_1_1983, by = "bfs_nr_old") %>%
  group_by(bfs_nr_new) %>%
  summarise(
    yes_votes = sum(yes_votes),
    no_votes = sum(no_votes), .groups = "drop"
  ) %>%
  mutate(v1_yes_share = yes_votes / (yes_votes + no_votes))
```

For the second dataset, which already has the proper spatial reference, the yes-share is calculated in a straightforward way.

```
vote_2_2017 <- mutate(vote_2_2017, v2_yes_share = yes_votes / (yes_votes + no_
  votes))
```

As the two datasets now share a common spatial reference, they can be merged by a join command, achieving a complete merge. And in a final step, the changes in the yes-share can be calculated.

```
complete_merge <- inner_join(vote_1_2017, vote_2_2017, by = "bfs_nr_new")
complete_merge <- mutate(complete_merge, difference = v2_yes_share - v1_yes_share)
```

The `complete_merge` object now contains the changes in electoral support for a facilitated naturalization of third generation immigrants across Swiss municipalities. This final dataset could now be used for further analyses of how citizens' preferences on naturalization changed over time.

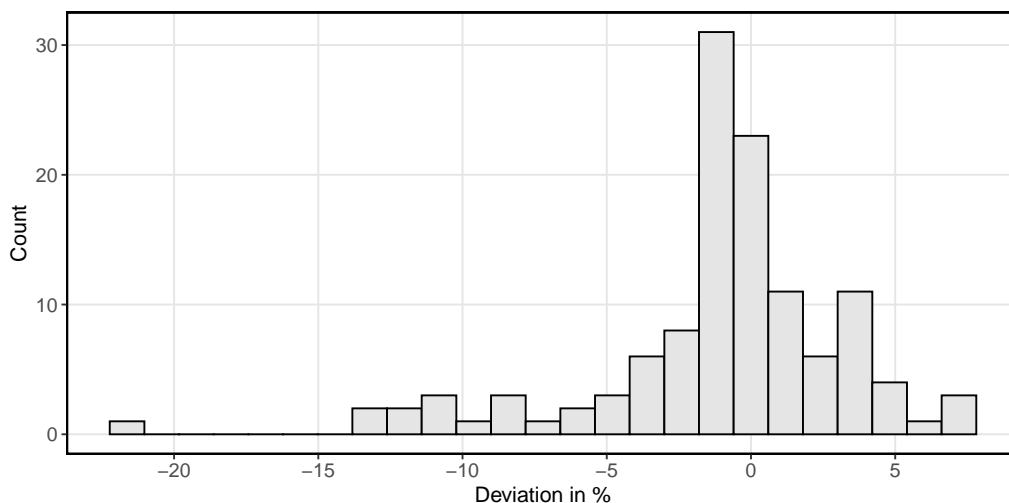
3 Comparison of the complete merge with the naive merge strategy

The advantage of using the complete merge strategy rather than a naive merge strategy in the example presented above is twofold. Firstly, observations are lost if the naive merge strategy is employed because the municipality numbers do not match (for example, if a new SFSO number was assigned after a merger of municipalities). In the example above, this is the case for 138 out of 2216 municipalities in the later period. Secondly, a bias is introduced in 122 cases because the same SFSO number no longer refers to the same spatial unit (and thus parts of the past municipality level information are ignored, for example, from Mühlethal, see Table 1). For our example, Figure 1 illustrates the resulting differences in the changes in the yes-shares between the two data merge strategies.⁷ In total, 11.7 % of the municipal data (regarding the stock in the later period) are either not considered or erroneous.

The deviation between the two strategies can be characterized along another dimension, i.e., the fraction of data points from original municipalities in the first period of time that is not considered when a naive merge is conducted. Interestingly, this fraction is not distributed uniformly across Swiss cantons because municipality mutations have occurred more often in certain regions of Switzerland than in others. Figure 2 depicts the fraction of municipalities as of January 1, 1960 that are considered in a naive merge with municipal data as of January 1, 2020 by canton. The canton of Glarus marks the extreme case: no observations would be included in a merged dataset. But also for the cantons of Neuchâtel, Fribourg, Ticino, Graubünden, and Thurgau, information

⁷Figure 1 includes the case of the incorporation of Mühlethal by Zofingen introduced above in Table 1. In the complete merge scenario, the yes-share in 1983 is the sum of the yes votes of Zofingen and Mühlethal divided by the sum of the yes and the no votes of Zofingen and Mühlethal. It equals 45.67 %. For the naive merge scenario, the yes-share in 1983 is simply the yes-share of Zofingen which is 46.14 %. The yes-share in 2017 is equal to 61.99 % (and it is the same for both strategies). Hence, for this example municipality, the change in the yes-share between 1983 and 2017 equals 16.31 % and 15.85 % for the complete and the naive merge, respectively.

Figure 1: Deviations from the actual changes in the yes-shares



Notes: This figure shows the deviations from the actual changes in yes-shares over time between the complete merge and the naive merge strategy for the 122 mutated cases for which the SFSO municipality number of one of the merging municipalities was carried over for the new entity. A positive / negative deviation signifies that the naive strategy overestimates / underestimates the change in the yes-share.

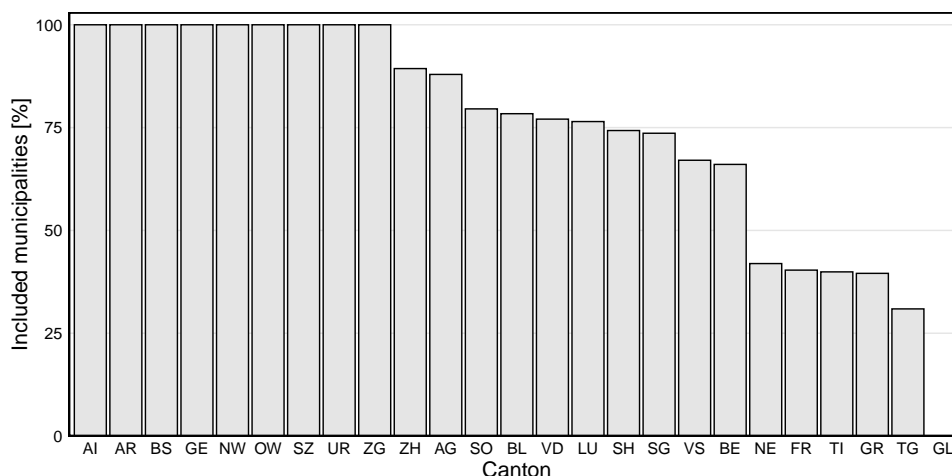
from less than 50 % of the original municipalities would be considered.

4 Concluding remarks

The Swiss MDMT automates the process of securing identical spatial references for different datasets as illustrated in the example above. The tool is meant to improve the quality of merged datasets compared to that of alternative procedures.⁸ Nevertheless, the appropriate strategy for the selection of municipalities that are considered in a longitudinal analysis depends on the research question. The same holds for how municipal level information is aggregated. If municipal mergers might affect the outcome variable per se, a separate selection and analysis of municipalities that were involved and not involved in mergers might be warranted. For example, consider the case of the municipal tax rate. The weighted mean tax rate might be a poor representation of a local tax situation in the past if the study is about tax competition. For tax competition, the variation in tax rates across spatial units is crucial and some weighted average tax rate might not easily serve as a counterfactual tax rate for the situation that would have existed if

⁸There are, of course, limitations: one is that not all types of municipal mutation can be mapped unambiguously. This refers in particular to secessions and exchanges of territory. However, this limitation is inherent to these types of mutations and not caused by the method used in the tool presented here.

Figure 2: Selection of municipalities included by canton when performing a naive merge between January 1, 1960 and January 1, 2020.



Notes: The percentage of original municipalities included if a naive merging strategy is adopted varies across Swiss cantons. For cantons without mutations, 100 % of the municipalities are considered in a naive merge, whereas the ratio of considered municipalities is below 100 % for cantons where mutations occurred between January 1, 1960 and January 1, 2020.

the municipalities had already been merged in the past. Regarding the limits to aggregation, there might be institutional municipality characteristics that cannot be meaningfully aggregated. For example, if the financial auditing in municipalities (Germ. Rechnungsprüfungskommission) that merged was organized differently, it is not readily possible to define and calculate a variable representative of these institutional conditions. Similar limitations might also arise even for characteristics that can be cardinally measured, like the size of the municipal council.

In the future, the Swiss MDMT could be extended with analytical functionality regarding the evolution of municipalities in Switzerland, such as one that lists all mutations for a given municipality within a given time period. Meanwhile, the tool will hopefully make life easier for researchers studying questions in the cosmos of Swiss local federalism.

References

- Baccini, L. and L. Leemann (2020). Do natural disasters help the environment? How voters respond and what that means. Forthcoming in *Political Science Research and Methods*.
- Brunner, B. and A. Kuhn (2018). Immigration, cultural distance and natives' attitudes towards immigrants: Evidence from Swiss voting results. *Kyklos* 71(1): 28-58.

Brühlhart, M. and M. Jametti (2006). Vertical versus horizontal tax externalities: An empirical test. *Journal of Public Economics* 90(10-11): 2027-2062.

Bundesamt für Statistik (2017). Historisiertes Gemeindeverzeichnis der Schweiz - Erläuterungen und Anwendungen.

Online: <https://www.bfs.admin.ch/bfs/de/home/grundlagen/agvch/historisiertes-gemeindeverzeichnis.assetdetail.4062821.html> [accessed: 05.11.2020]

Bundesblatt (2019). Strategie für offene Verwaltungsdaten in der Schweiz 2019-2023 (Open-Government-Data-Strategie, OGD-Strategie), 879-894.

Online: <https://www.admin.ch/opc/de/federal-gazette/2019/879.pdf> [accessed: 05.11.2020]

Bürgi-Schmelz, A. (2019). Wirtschaftliche Auswirkungen von Open Government Data.

Online: <https://www.bfs.admin.ch/bfs/de/home/dienstleistungen/ogd/dokumentation.assetdetail.11147089.html> [accessed: 05.11.2020]

Cappelletti, F., and N. Soguel (2013). Bureaucracy and the flypaper effect: Evidence from intercommunal fiscal equalisation in the canton of Vaud. *Urban Public Economics Review* 18: 38-67.

Eugster, B. and R. Parchet (2019). Culture and taxes. *Journal of Political Economy* 127(1): 296-337.

Galletta, S. (2020). Direct democracy, partial decentralization and voter information: evidence from Swiss municipalities. *International Tax and Public Finance* 27: 1174-1197.

Haus, A. and A. Ladner (2020). Wer hat die Macht in den Gemeinden? Eine Analyse über den Einfluss politischer Akteure auf die lokale Politik in der Schweiz. *Jahrbuch der Schweizerischen Verwaltungswissenschaften* 11(1): 66-80.

Horber-Papazian K. (2007). The Municipalities. In: Klöti U., P. Knoepfel, H. Kriesi, W. Linder, Y. Papadopoulos, P. Sciarini (eds.), *Handbook of Swiss Politics*. Zürich: Neue Zürcher Zeitung Publishing: (227-252).

Knechtel, V. (2020). SMMT: The Swiss Municipal Data Merger Tool Maps Municipalities Over Time. R package version 1.0.7. <https://CRAN.R-project.org/package=SMMT>

Ladner, A. (1991). *Politische Gemeinden, kommunale Parteien und lokale Politik. Eine empirische Untersuchung in den Gemeinden der Schweiz*. Zürich: Seismo.

Ladner, A. and R. Steiner (2003). Die Schweizer Gemeinden im Wandel: Konvergenz oder Divergenz? *Swiss Political Science Review* 9(1): 233-259.

Luthi, E. and K. Schmidheiny (2014). The effect of agglomeration size on local taxes. *Journal of Economic Geography* 14(2): 265-287.

Matter, U. (2014). pvsR: An R package to interact with the Project Vote Smart API for scientific research. R package version 0.3. <https://CRAN.R-project.org/package=pvsR>

Matter, U. (2018). RWebData: A high-level interface to the programmable web. *Journal of Open Research Software* 6(1): 1-12.

Matter, U. (2019). Big Public Data aus dem Programmable Web: Chancen und Herausforderungen. *HMD* 56: 1068-1081.

Matter, U. and A. Stutzer (2015). pvsR: An open source interface to big data on the American political sphere. *PLoS ONE* 10(7): e0130501.

Meier, A. N., L. Schmid, and A. Stutzer (2019). Rain, emotions and voting for the status quo. *European Economic Review* 119: 434-451.

Sebei, H., M. A. H. Taieb, and M. B. Aouicha (2018). Review of social media analytics process and big data pipeline. *Social Network Analysis and Mining* 8(1): 30.

Slotwinski, M. and A. Stutzer (2019). The deterrent effect of an anti-minaret vote on foreigners' location choices. *Journal of Population Economics* 32(3): 1043-1095.

Steiner, R., A. Ladner and P. Reist (eds.) (2014). *Reformen in Kantonen und Gemeinden*. Bern: Haupt.

Swiss Federal Archives (2017). Showcase Linked Data Service LINDAS Historisiertes Gemeindeverzeichnis.

Online: https://www.egovernment.ch/files/1115/0217/9053/Showcase_Gemeindeverzeichnis_de.pdf
[accessed: 05.11.2020]

Verein eCH (2014). Datenstandard Historisiertes Gemeindeverzeichnis der Schweiz.

Online: <https://ech.ch/de/standards/47999> [accessed: 05.11.2020]