

Data Descriptor

Municipalities in the Czech Republic—Compilation of “a Universal” Dataset

Vít Pászto , Rostislav Néték , Alena Vondráková  and Vít Voženílek

Department of Geoinformatics, Palacký University Olomouc, 77146 Olomouc, Czech Republic; rostislav.netek@upol.cz (R.N.); alena.vondrakova@upol.cz (A.V.); vit.vozenilek@upol.cz (V.V.)

* Correspondence: vit.paszto@gmail.com

Received: 30 October 2020; Accepted: 19 November 2020; Published: 24 November 2020



Abstract: There have been many changes in the spatial composition and formal delimitation of administrative boundaries of Czech municipalities over the past 30 years. Many municipalities have changed their official status; they separated into ones that were more independent or were merged with existing ones, or formally redrew their boundaries due to advances in mapping technology. Such changes have made it almost impossible to analyze and visualize the temporal development of selected socioeconomic indicators, in order to deliver spatially coherent and time-comparable results. In this data description, we present an evolution of a unique (geo) dataset comprising of the administrative borders of the Czech municipalities. The uniqueness lies in time and topologically justified spatial data resulting in a common division of the administrative units at the LAU2 level, valid from 1995 to 2019. Besides the topologically correct spatial representations of municipalities in Czechia, we also provide correspondence tables for each year in the mentioned period, which allows joining tabular statistics to spatial data. The dataset is available as a base layer for further temporal and spatial analyses and visualization of various socioeconomic statistical data.

Dataset: Dataset is available at <https://gislib.upol.cz/portal/apps/sites/#/opendata/items/08a30f65288f40ccac2e31fc6ce6b908> or <https://tinyurl.com/superlayer>.

Dataset License: Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)

Keywords: municipalities; geodata; administrative boundaries; spatiotemporal analysis; open-data

1. Summary

The primary motivation for the compilation of the presented datasets comes from the need of the project (spatial differentiation and visualization of geodemographic processes, with a focus on households in an aging society in the Czech Republic) dealing with spatial differentiation and visualization of geodemographic processes in Czechia over the last 25 years. The project goal is to explore various geodemographic processes at a very detailed level, specifically at nomenclature of Territorial Units for Statistics (NUTS) local administrative units 2 (LAU2), commonly used in the European Union for statistical purposes [1]. While most of the national statistical data, especially those coming from censuses, are published for clearly defined administrative units [2], it is essential to have such units for spatial analyses and visualization in the form of spatial data. It is not that complicated to obtain the actual spatial data of administrative boundaries and corresponding statistical indicators for the selected year. However, when we want to explore trends in a more extended period (e.g., five, ten, twenty years), it is becoming a very complex task to prepare and harmonize spatial and statistical data. While we confronted our project vision with the actual state of the data quality, we encountered key issues we had to solve:

- “How can we display a given geodemographic indicator in a single large-format map considering all the changes in municipalities’ spatial composition?”
- “Is it even possible to explore trends of given geodemographic indicator in certain regions while we do not have unified municipal units?”
- “Will be geodemographic changes among municipalities and between years even comparable?”

The latter issue is also explicitly mentioned in [2] as one of the crucial limitations of statistical data, i.e., lack of mutual comparability due to frequent changes of administrative or statistical units. Although there are efforts to substitute administrative boundaries with regular grid structures (e.g., GEOSTAT population grid), and methods for transferring grids into administrative units (and vice versa) exist (e.g., [3,4]), and have been used in several scientific studies (e.g., [5,6]), there is, yet, no country in Europe (to the authors’ knowledge) deploying grids as an official approach for socioeconomic data curation—from collection, management, and analysis to visualization. Moreover, as mentioned by [7], the analysis and results should correspond with administrative boundaries as an essential requirement for real-world applications or practical applications of scientific conclusions (e.g., [8–10]). In other words, measures taken to tackle, e.g., the elderly population, changes will never be targeted to specific grid cells; instead, policy powers will be applied within the respective administrative boundaries. Therefore, our approach for data harmonization focused on keeping administrative boundaries for consequent analysis and visualization.

Our presented dataset unified municipal administrative units allowing analyses of data as they change over time. The added value of the dataset does not lie in the novel use of spatial methods—instead, we used a purposeful sequence of necessary individual tools—but in the final product in the form of a universal dataset. The data were curated with extremely high precision in order to meet the requirements of our project colleagues, demographers, working with data on individual citizens of Czechia (the non-public specialized database demographers obtained from the Czech Statistical Office containing detailed information from official state registers about every single citizen). This means, literally, ten million records that ultimately have to be in correspondence with the total sum of municipal statistics. Hence, our treatment and proper assignment of spatial representation of municipalities had to be 100% precise. Thus, individual and semi-manual data adjustment was time-consuming, heterogeneous (in terms of arisen complications), and necessary. Moreover, a number of diminutive, but significant, data adjustments had to be done. Finally, the article introduces a workflow from an Excel sheet to open data publishing. It is based on the cooperation between experts in cartography, geoinformatics, and demography.

We used an approach based on a principle of “common spatial denominator”, i.e., we used data aggregation into larger units with stable boundaries. Although we lost spatial detail to increase the time extent (more about these drawbacks in, e.g., [2,11]), the fact that we could use every single statistical data from 2019 backwards prevailed. In regards to the data itself, we obtained (a) a list of codes of municipalities for each year from 1995 to 2019, and (b) spatial data with administrative boundaries and their centroids for selected years only (1996, 2000, 2002, 2003, and from 2009 each year onwards).

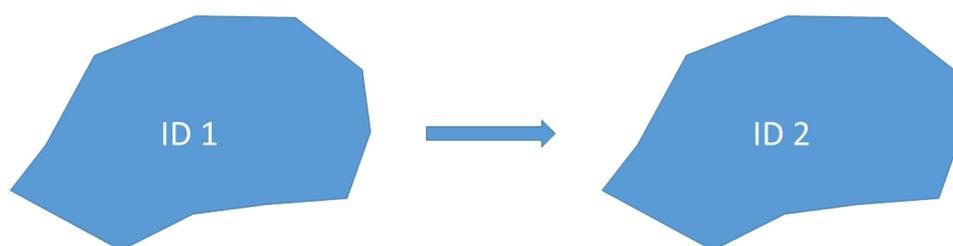
Additionally, spatial data were in two different generalization levels—older data (up to 2009) with coarser administrative boundaries, and newer data with high-precision boundaries. Most of the spatial data centroids also contained information about municipality IDs (compatible with the list of codes); however, in the missing years of spatial data, these were not available. More about spatial data curation is in the section Methods. In regards to non-spatial data, i.e., list of codes giving municipal ID and information about the absolute number of units, Table 1 provides an overview of their development in the observed period (1995 to 2019). Please note that Table 1 shows numbers of official (and existing) municipalities in particular years.

Table 1. The number (No.) of local administrative units (LAU2) in Czechia from 1995 to 2019.

Year	No. of Units	Year	No. of Units
1995	6232	2008	6249
1996	6233	2009	6249
1997	6234	2010	6250
1998	6242	2011	6251
1999	6244	2012	6251
2000	6251	2013	6253
2001	6258	2014	6253
2002	6254	2015	6253
2003	6249	2016	6258
2004	6249	2017	6258
2005	6248	2018	6258
2006	6248	2019	6258
2007	6249	-	-

However, we identified four broad problems during data processing, which many researchers share. These problems might present when dealing with administrative boundaries in general:

- (1) Splitting features: administrative units were divided over—it is clear from Table 1 that the number of administrative units increased from 6232 in 1995 to 6258 in 2019, which is the absolute difference of 26 units. In general, this number is not significant (given the total number of more than 6000 units); however, changes were detected in more than 70 cases over the period due to other discrepancies further described below. This first problem, which causes a need for identification and recoding of administrative units, is illustrated in Figure 3.
- (2) Merging features: administrative units were merged over time—although the changes in the absolute number of municipalities (Table 1) were not dramatic, we have to keep in mind that the division of administrative units (mentioned in point 1) was blurred by the merging of other ones. Nevertheless, the merged administrative units led us to find them and correct the spatial data in order to maintain a “common denominator” principle. Moreover, we had to select the ID code to be kept in data. The illustration of the second problem is in Figure 4.
- (3) Re-indexing features: administrative units changed their ID codes—in several cases, the municipalities changed their official status (due to changes in the systematic classification of settlement units) and had to be recoded (Figure 1).
- (4) Topology mismatch: spatial details of administrative boundaries changed—as mentioned before, the spatial data were provided in two different levels of detail in terms of administrative boundary precision (Figure 2), which complicated the spatial treatment of data. For instance, due to non-corresponding topology of boundaries, some of the formerly intended geographical information system (GIS) tools could not be applied (e.g., “select by location” tool), and “spatial join” remained as the only option. In general, any spatial-based calculations, such as areal measurements, choropleth map creation (when the area is needed), would be inaccurate. Therefore, this problem has to be solved as well.

**Figure 1.** Change in ID coding of administrative units.

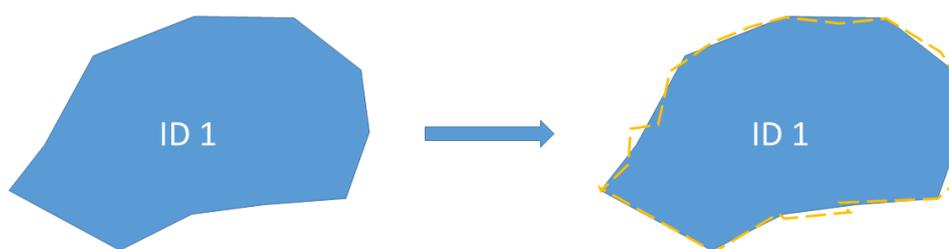


Figure 2. Change in spatial detail of administrative unit boundaries (orange dash line represents boundaries in more refined detail).

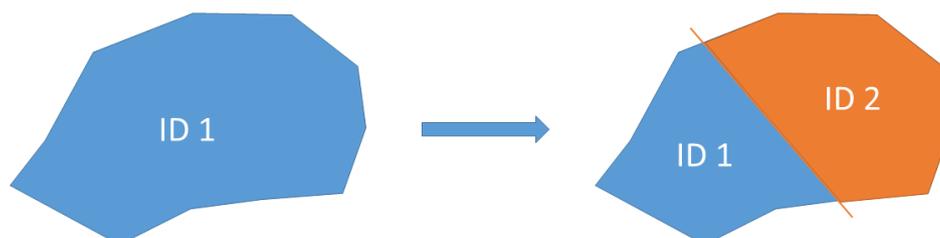


Figure 3. Principle of the division of administrative units.

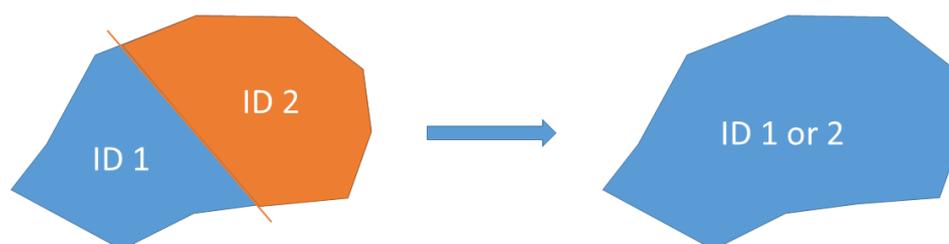


Figure 4. Principle of the merge of administrative units.

Although all of the problems mentioned above concerned approximately 70 municipalities, they could not be ignored. First, in several cases, they involved large towns and cities or military areas. Therefore, their exclusion from the dataset would diminish the quality of the data and consequent analysis and visualization. Second, the geodemographic dataset intended for joining with spatial data was obtained from the Czech Statistical Office, with extremely detailed characteristics about individual inhabitants. For that reason, all of the calculations and aggregations to administrative units had to be performed with 100% precision (in terms of total counts of inhabitants). Moreover, several specific “spatial” problems had to be treated individually (see details in Section 3.1). To summarize the final output, we prepared a dataset consisting of the spatial representation of aggregated administrative units valid for a period from 1995 to 2019, and the non-spatial part in the form of the correspondence table, where “old” (former) and “new” (based on aggregation) ID codes of municipalities are listed. We work-titled this dataset the “universal” or “superlayer” (we will use “universal” for the rest of the manuscript).

The dataset allows users to use statistical data from any year from the period and to link them with the spatial representation of administrative boundaries. The most significant advantage of the universal is that users can (1) compute derived indexes from any statistical data from 1995 onwards, and (2) analyze time-variability and trends of such data without further need of spatial data curation. An example of a combination of both benefits is in Figure 5 depicting an overall trend from 1995 to 2018 of the vital index [12] in Czech municipalities. This would not be possible without losing valuable information about some municipal units (by excluding them from the map) if the universal layer had not been created. Analogically, it is possible to perform year-to-year comparisons only by joining statistical data in given years with the correspondence table and consequently with spatial data (more in User Notes in Section 4).

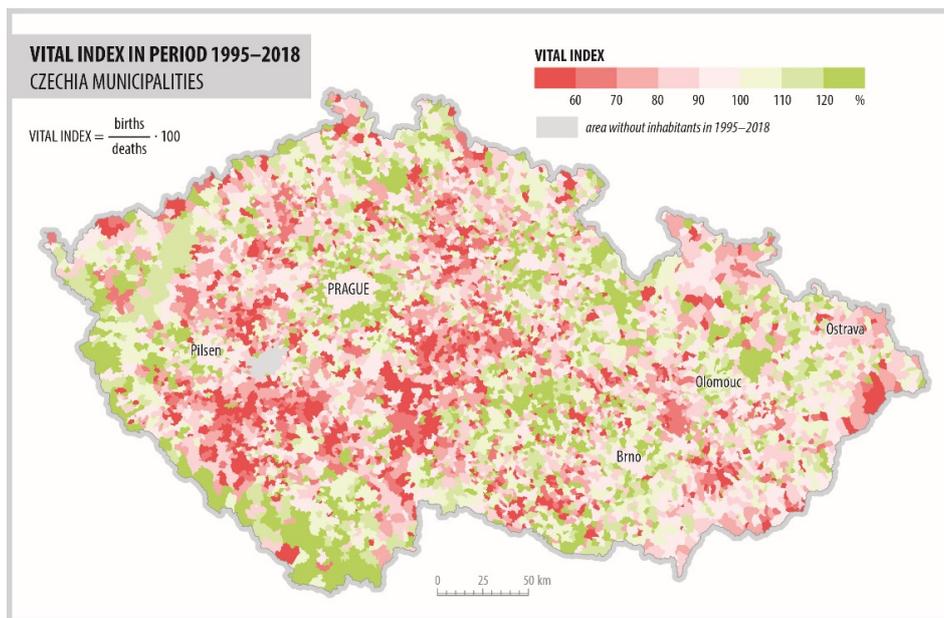


Figure 5. The overall trend in the vital index in Czechia during the period 1995–2018.

Similarly, other phenomena (data and layers) can be treated in the way we present in this paper, e.g., areas of geological regions, climatic types, and catchment areas. Moreover, we used a retrospective approach, which was challenging, because the current archiving and metadata instructions and tutorials are different than they were 15–20 years ago (if there were any at the time) [13]. Therefore, the presented universal dataset can save other researchers, regardless of the topic they deal with, a significant part of their time on data processing.

2. Data Description

As indicated in the previous chapter, we created a unique dataset (universal) that is ready to use for other researchers exploring Czech statistical data in the geographical or spatial context. The dataset consists of two parts—spatial and non-spatial data. Both parts are needed for proper linkage of universal with any other statistical data. We strive to keep the simplicity of this data description since we elaborate more on the most important methodological and technical details in the Methods section of this paper.

2.1. The Spatial Part of the Dataset

All spatial data treatment was performed in the ArcGIS Pro environment; therefore, the primary data format was the Esri geodatabase. To expand the target user group, we created an open data portal (see description in Section 2.3 and methodological background in the Methods chapter), where data can be downloaded in other formats (Esri shapefile, Geo JavaScript Object Notation (GeoJSON), TopoJSON, Keyhole Markup Language (KML)). All formats are compatible with standard GIS software tools. The resulting universal layer is in a vector format using polygons as the primary geometry representation of administrative boundaries. In total, 6215 records are equal to the number of “common denominators” representing administrative boundaries (Figure 6). The coordinate system set for the data is S-JTSK_Krovak_East_North (EPSG code 5514). Attributes are reduced to the most necessary ones:

- Shape geometry (Shape);
- Name of a municipality (Name);
- New ID code of a municipality (ID_code_n);
- Shape Length (Shape_Length);
- Shape Area (Shape_Area).

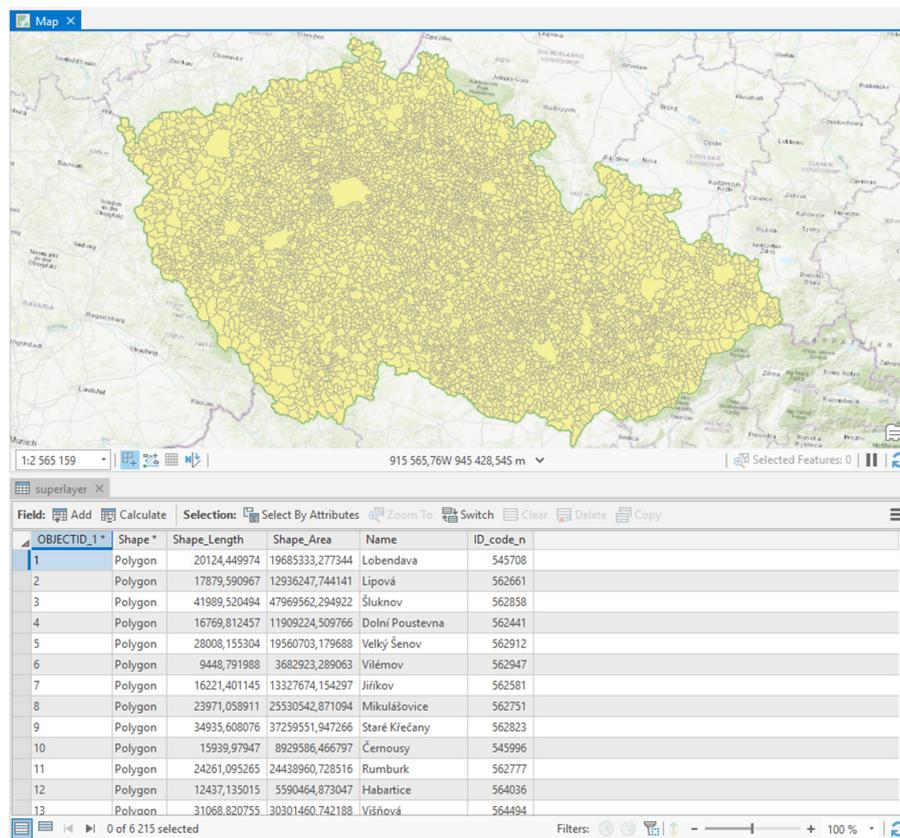


Figure 6. Snapshot from ArcGIS Pro environment showing data view on the spatial dataset with its attributes.

The spatial part of the dataset is a result of the geospatial analysis described in the Methods chapter. Initial provider of the administrative boundaries layer and their centroids is The Czech Office for Surveying, Mapping, and Cadastre (ČÚZK, <https://www.cuzk.cz/en>).

2.2. Non-Spatial Part of the Dataset

The non-spatial part of the dataset represents a correspondence table for a correct ID code assignment. The table serves as a converter between the original municipal codes and the newly assigned ones from the aggregation. The records in the table allow the user to transform any Czech statistical data at the municipal level (LAU2) from 1995 to 2019 into the resulting 6215 units that correspond with the spatial part of our universal dataset. It is to be noted that 6215 features are a result of data curation and represents common denominator units; therefore, the number differs from the numbers in Table 1 (showing official municipality counts). Although the creation of the non-spatial part is more demanding than its spatial part, the final output is straightforward to use. The non-spatial part is in the form of a table, primarily elaborated in MS Excel (XLSX format). Reasons for MS Excel usage are given in Section 3.1. Similar to the spatial part, various table formats-Comma-Separated Values (CSV), plain textfile (TXT)-are available through our open data portal.

The table itself is in a single Excel workbook containing 25 sheets, each representing one year from 1995 to 2019 (in the case of CSV and TXT formats, individual files are created for each sheet/year). The number of records varies in each sheet based on the total number of former/original municipalities (see Table 1). Thus, one row in the table relates to one municipality (LAU 2) unit. All sheets contain only two columns (Figure 7):

- Original code of a municipality as provided (ID_code_original);
- Newly assigned municipality code (ID_code_new).

	A	B	C	D	E	F
1	ID_code_original	ID_code_new				
6218	554111	554111				
6219	566292	566292				
6220	554006	554006				
6221	566250	566250				
6222	554057	554057				
6223	554138	554138				
6224	506320	506320				
6225	507113	507113				
6226	510891	510891				
6227	505927	505927				
6228	599247	599247				
6229	568252	568252				
6230	596124	596124				
6231	598984	555088				
6232	574716	555134				
6233	537012	537501				
6234	581411	581976				

Figure 7. Snapshot from MS Excel showing the data structure of non-spatial part of the dataset.

The non-spatial part of the dataset resulted from time-consuming individual edits and validation of municipal ID codes in each year. Original ID code refers to the actual situation in a given year; the newly assigned ID code brings the information of the aggregated unit (“common denominator”). Original ID codes were provided by The Czech Statistical Office (CZSO, <https://www.czso.cz/csu/czso/home>).

2.3. Open Data Portal

Since a requirement for providing data as available as possible, the authors decided to follow and “open data” concept. According to The Open Knowledge Foundation [14], open data is “data that can be freely used, shared and built-on by anyone, anywhere, for any purpose”.

Data are published on the Internet with no technical or legislative restrictions for users. Data are published from the primary/original source, but for users are available in several formats and open standards. Data should be published to the maximum extent as possible. There are several tools available for publishing open data as open data portals are the standardized solution for both publishing and sharing spatially based data. In general, open data portals are catalogues supporting spatial data formats from a technological point of view; therefore, it is the best solution for publishing unique layers, such as universal dataset, following the open data concept [15].

ArcGIS Hub (formerly the ArcGIS Open Data) is a community platform designed by Esri to share open data with the general public. Esri, a worldwide leader in GIS, provides a complex solution that is fully connected with the ArcGIS platform. It uses Esri Geospatial Cloud, which stores all created pages and all data (an advantage for organizations that do not want to manage data on their servers). ArcGIS Hub is designed for sharing both spatial and non-spatial data, which can be visualized directly on the platform using maps, tables, graphs, and the like. The general public can download datasets or their filtered parts in various data formats. Beside datasets, as a complex portal, it allows one to create maps or web applications, search inside ArcGIS platform content, and share content with other members. The portal includes a website for user-friendly access and a map browser. It can be used for creating a simple interactive map application—map overviews.

After data uploading to the portal, all vector data are automatically available for download in several formats: feature collection, Comma-Separated Values (CSV), Keyhole Markup Language (KML), Shapefile (SHP), Geo JavaScript Object Notation (GeoJSON), XLSX for Excel, and GeoServices Application Programming Interface (API). It is also possible to share data via Open Geospatial Consortium standards: Web Map Service (WMS) for rasters, Web Feature Service (WFS) for vectors,

and Web Coverage Services (WCS) for covers. ArcGIS Hub allows users to set permissions for workgroup members only or the public.

3. Methods

This chapter describes the main methodological steps taken during the creation of the universal dataset. The chapter is divided into three sections based on the nature of the data curation. General overview of data curation flow is depicted in Appendix A.

3.1. Treatment of the Data from a Spatial Perspective

Because the main goal was to compile a data set for analyzing and visualization of geodemographic indicators over time, the study started with obtaining the spatial data of Czech municipalities (LAU2). As mentioned above, we missed some years from an earlier time in our observed period (1995–2019). Therefore, the filling gaps started with the year 1996, i.e., one year (1995) backwards, and the following years onwards. By doing so, we had to check the changes between years with the use of the CZSO code list and manually change administrative boundaries accordingly if the municipalities merged. In the case of splitting the municipalities, we considered those and recorded their ID codes into the correspondence table. Consecutively, this procedure was repeated until the year 2019. This procedure ensured having administrative boundaries (polygons) spatially constant throughout the observed period (1995–2019).

Additionally, we received the official data from ČÚZK from 2009 onwards in finer spatial detail (see differences in Figure 8a), which forced us to decide what spatial data to use. We decided to keep the former (2008 and backwards) more generalized administrative boundaries because the final visualizations were intended to display the whole Czechia. Otherwise, the finer detail of administrative boundaries would cause additional cartographic problems in map-making (e.g., rendering errors of the final map due to the output map scale).

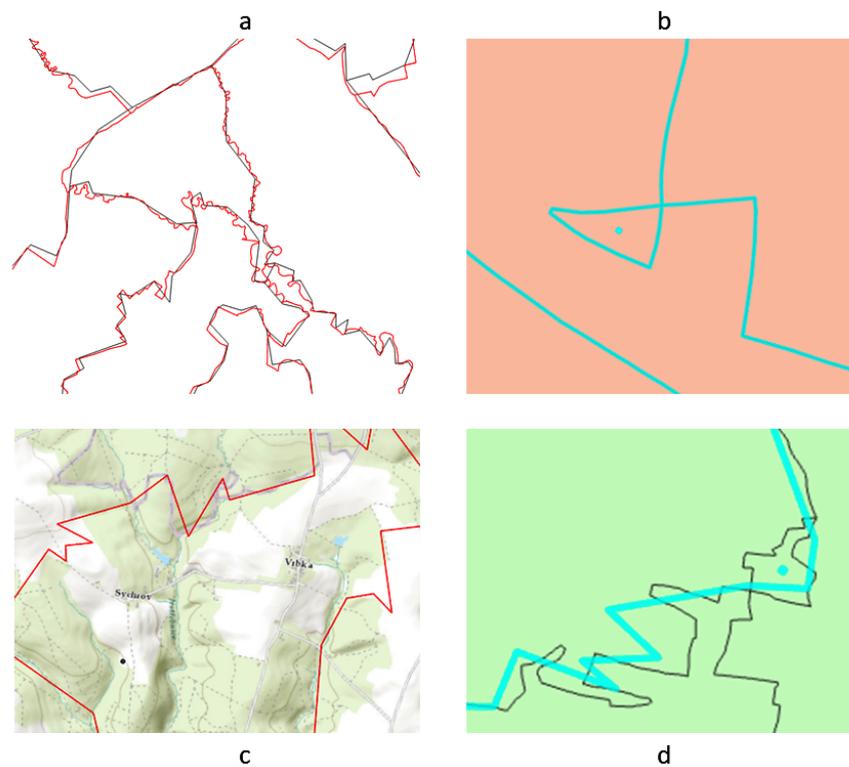


Figure 8. Spatial errors and issues raised during the spatial data processing (a—different spatial detail of administrative boundaries; b—incorrect location of centroid to municipality boundaries; c—missing centroid from other municipality; d—incorrect placement of other municipality’s centroid to different one).

As mentioned in the Summary chapter, spatial data obtained from ČÚZK were in a vector format as administrative boundaries (polygons) and their centroids (points with information on municipalities' ID codes). Therefore, it was necessary to link centroids with polygons to get ID codes within the polygons for future connection with other statistical data. From the analytical perspective, the “spatial join” tool served this purpose. In general, “spatial join” tool projects the attribute information from points to polygons based on their mutual geographical location. However, during “spatial join” application, errors emerged when point data carrying information with ID code did not fit within the right administrative boundaries—see Figure 8. Although these problems occurred in a small number of municipalities, they had to be corrected to fulfil the requirement of 100% correct ID assignment for future matching with tabular data.

In summary, the errors were first checked visually, corrected manually, and then cross-checked (and corrected) again after the data validation process. At this point, it is crucial to note that for such data adjustments, no automatic GIS tools could have been applied since automation had not been able to handle such issues in the data sufficiently. Automatic tools (or semi-automatic combination of GIS tools) is indeed beneficial for processing large datasets; however, often leaving some “outliers” unsolved (e.g., Figure 8b,c). In such cases, individual deviations in the data had to be treated manually, unfortunately. Since we did not work with other attribute data than municipality IDs (qualitative information), we could not even apply any of the methods commonly deployed in the modifiable areal unit problem (MAUP) [16], e.g., proportional redistribution of values.

Once we had all of the municipal ID codes contained in the polygon representation of administrative boundaries, we applied “spatial join” again for data validation to check the correct assignment of ID codes. In this procedure, the transformation of administrative boundaries (polygons) back into their centroids (points) occurred. As a result, 25 centroids (mostly laying on each other) for each municipality in Czechia were obtained. Consequently, we used “spatial join” again—target feature was a layer containing edited polygons of administrative boundaries based on the “common denominator” principle and all 25 centroids with ID code attributes. However, this time, we applied a different merge rule for attributes (coincidentally, also called “join”) in advanced settings of the tool, which allows listing all joined attributes within one record of the target layer. This helped us to:

- (1) Check the total number of ID codes—one municipality should contain 25 ID codes. If there were more/fewer ID codes, further data inspection and review was necessary. This validation step was in terms of quantity.
- (2) Check if there were two or more different ID codes—one municipality could have more ID codes, which indicated a split or merger of several municipalities. This validation step was in terms of quality.

All issues identified in spatial data processing were recorded and immediately taken into account in the non-spatial part of the universal dataset.

3.2. Treatment of the Data from a Non-Spatial Perspective

Since there were changes (mentioned in points 1 and 2 in the Summary section, and partially described in Section 3.1) emerging in different years, it was necessary to search for them individually in each year. Once the change was detected and identified, it has to be decided which ID code will be maintained. For change detection purposes, a combination of an official document on changes of administrative delimitation of Czechia from ČÚZK, official historical registers from CZSO, results from spatial data treatment, and other internet searches were used. Unfortunately, these sources were not mutually coherent, so individual changes had to be verified individually. In case of units division over time, the former ID code remained in one of those divided, and the newly established municipality received the ID code from the “common denominator” spatial unit with stable boundaries (fortunately, all the former ID code units existed throughout the whole period). Explained by the example—municipality A, divided into A1 and A2, while A1 kept its ID (based on its size or

importance within the settlement system) and A2 was assigned with an ID from the former bigger unit (usually the same ID as A1); instead of keeping the new one. If two or more administrative units merged over a given period, recoding was done backwards (all municipalities forming a new, bigger unit, were assigned with the identical ID). In other words, the newly established administrative unit (and its spatial representation) is projected back in time. Again, in the same logic, this newly created unit acted as a “common denominator”, therefore, kept in the final dataset.

Thus, every year after a detection of split or merger of the administrative unit, the former code replaces the new one, while the reference information about both codes was maintained. The correspondence table containing both codes represents a final product from non-spatial data preparation. This table allows linking any statistical table with the spatial part of the dataset (see more in the User Notes section).

Methodologically, we applied a combination of available tools in Microsoft Excel (e.g., look up function to search for differences, contingency table to cross-check overall counts, and so on), and the programming language R (functions `na.omit` and `setdiff`) in order to find differences between spatial and tabular data automatically. This combination of non-spatial tools was used rather for practical reasons. After initial preparation of the spatial dataset (with the use of geodatabase), demographers required us to deliver a list of municipalities in MS Excel as they commonly work in such tabular environments. The data was then cross-checked with the statistical tables demographers possessed and sent notes for corrections (highlighted in MS Excel) back. This iterative process was, therefore, easier to handle directly in Excel.

3.3. Open-Data Portal Design

Sharing universal layer via an open data portal is based on two steps: metadata and geometry preprocessing and publishing. Metadata is structured information that is used to characterize, identify, and interpret each dataset. Metadata is an essential parameter in the field of spatial data to be able to use the relevant data correctly [17]. Metadata should be part of every dataset and web service. In our case, the metadata are characterized as title, description, spatial extent (bounding box), author, date of publishing, license and permission, original source, number of features, and attributes specification (title, format, count statistics). All metadata for universal is available in a standardized format at URL: <https://gislib.upol.cz/portal/sharing/rest/content/items/08a30f65288f40ccac2e31fc6ce6b908/info/metadata/metadata.xml>.

If appropriate metadata is implemented, the publishing phase follows. Publishing is the formal and technical process of data publication. After publishing, data are visualized and available for download within the open data concept. Sharing with the public or only team-members is a crucial option. Within the ArcGIS platform, the user is asked for available data formats for download option, in our case, all possible options are available: native Esri spatial geodatabase as the original source, feature collection, CSV, KML, Shapefile, GeoJSON, XLSX, and standardized GeoServices API. All data, applications, and websites created through ArcGIS Open Data Portal are stored in the Esri Geospatial Cloud repository, which means that they are available for administrators for further updates. Open Data Portal of Department of Geoinformatics, Palacký University Olomouc, is available at Supplementary Materials. It contains dozens of datasets divided by categories. Each dataset is available via a specific URL. The case study of the universal layer is available via Supplementary Materials.

The interface of a specific dataset contains spatial, tabular, and attribute segments (see Figure 9). The interactive map provides a general overview of phenomena. It is not a fully-developed application with high-level of interactivity; it does not meet all cartographical rules. It allows the basic functionality of web maps—zoom in/out, pan, and attributes selection by click for data preview. The tabs allow the user to switch among spatial and tabular visualization of the whole dataset. Usually, the web-design is tested with real users before publishing, e.g., by eye-tracking methods [18,19]; however, since the Esri platform is devoted to fast, effective, and easy publication of data, it does not allow advanced options to change the user interface.

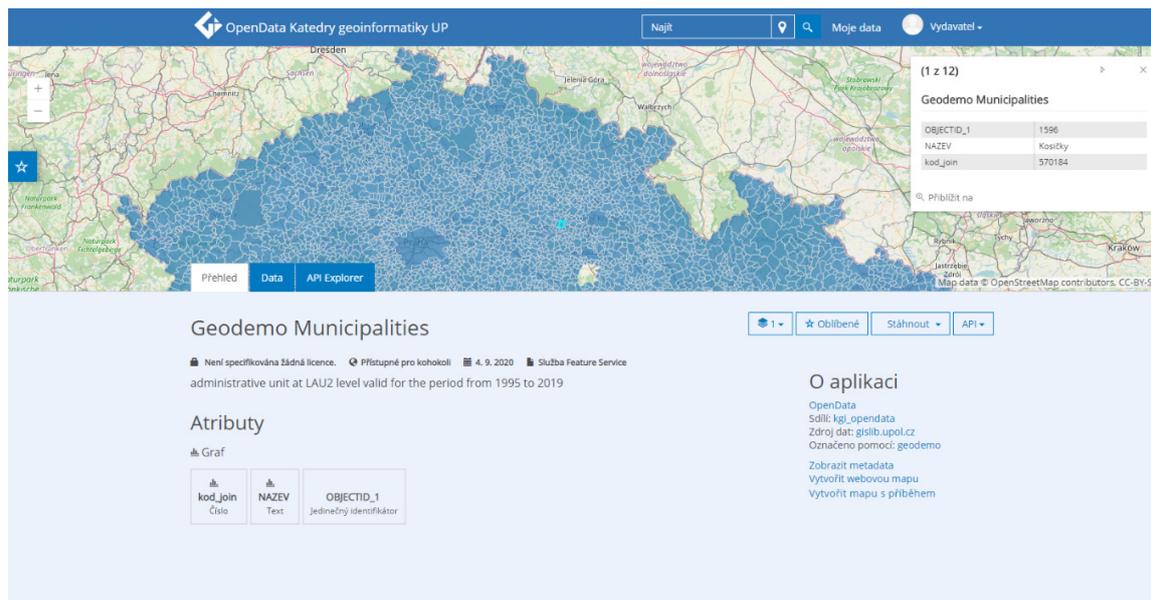


Figure 9. The user interface of the universal layer within Open Data Portal of Department of Geoinformatics, Palacký University Olomouc.

Moreover, it implements advanced filtering. The second part includes all metadata, including descriptions and attributes. Buttons allowing download in specified formats are fundamental from the users' point of view. If the Esri user is interested in the dataset, he could simply use it for a custom project within ArcGIS platform, directly upload this layer by "Make web map" button, and create highly interactive web map applications.

4. User Notes

In this section, we present a basic principle on how to connect the universal layer with any Czech statistical data. In order to join any Czech statistical data at the municipal level (LAU2) with the spatial part of our data for consequential spatial analysis or visualization, it is necessary to recode the municipality IDs of the statistical indicator with the use of our correspondence table. There are other ways to combine other Czech statistical data at the municipal level (LAU2) with our universal layer; however, we propose the following. First, it is crucial to recode the chosen Czech statistical data according to the correspondence table, i.e., find municipal ID codes that changed over time, and have the newly assigned value from the aggregation process. Once these records are identified, the calculation of selected indicator values should be performed (usually a simple count is satisfactory). The recoding process is illustrated in Figure 10. As the last step, recoded statistical data could be joined with the spatial part of the universal layer in GIS environment (in ArcGIS Pro by using "join" function). Obviously, the whole process is possible to perform vice versa, i.e., by joining the correspondence table with spatial data, first with duplicated records kept, and then joining the Czech statistical data with spatial data. Calculations of the indicator's values could be consequently performed directly in the GIS environment.

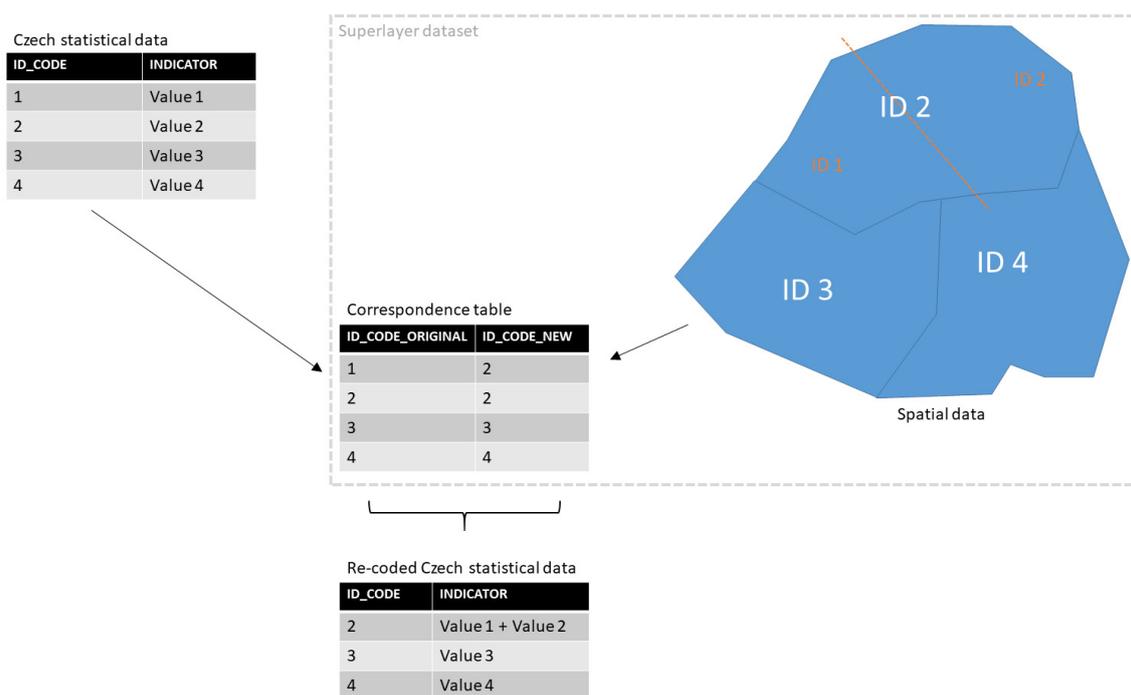


Figure 10. The Principle of connecting Czech statistical data with the universal dataset.

Finally, not to forget the availability of statistical indicators, they can be freely obtained on the Czech Statistical Office website: www.czso.cz.

Supplementary Materials: The described dataset is available online at <https://gislib.upol.cz/portal/apps/sites/#/opendata/items/08a30f65288f40ccac2e31fc6ce6b908> or <https://tinyurl.com/superlayer>.

Author Contributions: Conceptualization, V.P., V.V.; methodology, V.P., A.V., R.N.; data curation, V.P., A.V., R.N.; writing—original draft preparation, V.P., R.N., A.V.; writing—review and editing, V.V.; visualization, A.V., V.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by The Czech Science Foundation; grant number 18-12166S (Spatial differentiation and visualization of geodemographic processes with a focus on households in an ageing society in the Czech Republic).

Acknowledgments: The authors would like to thank Jitka Rychtaříková from Charles University (Prague) for her support in regards to thematic aspects of the data.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Schematic flow chart of the data curation

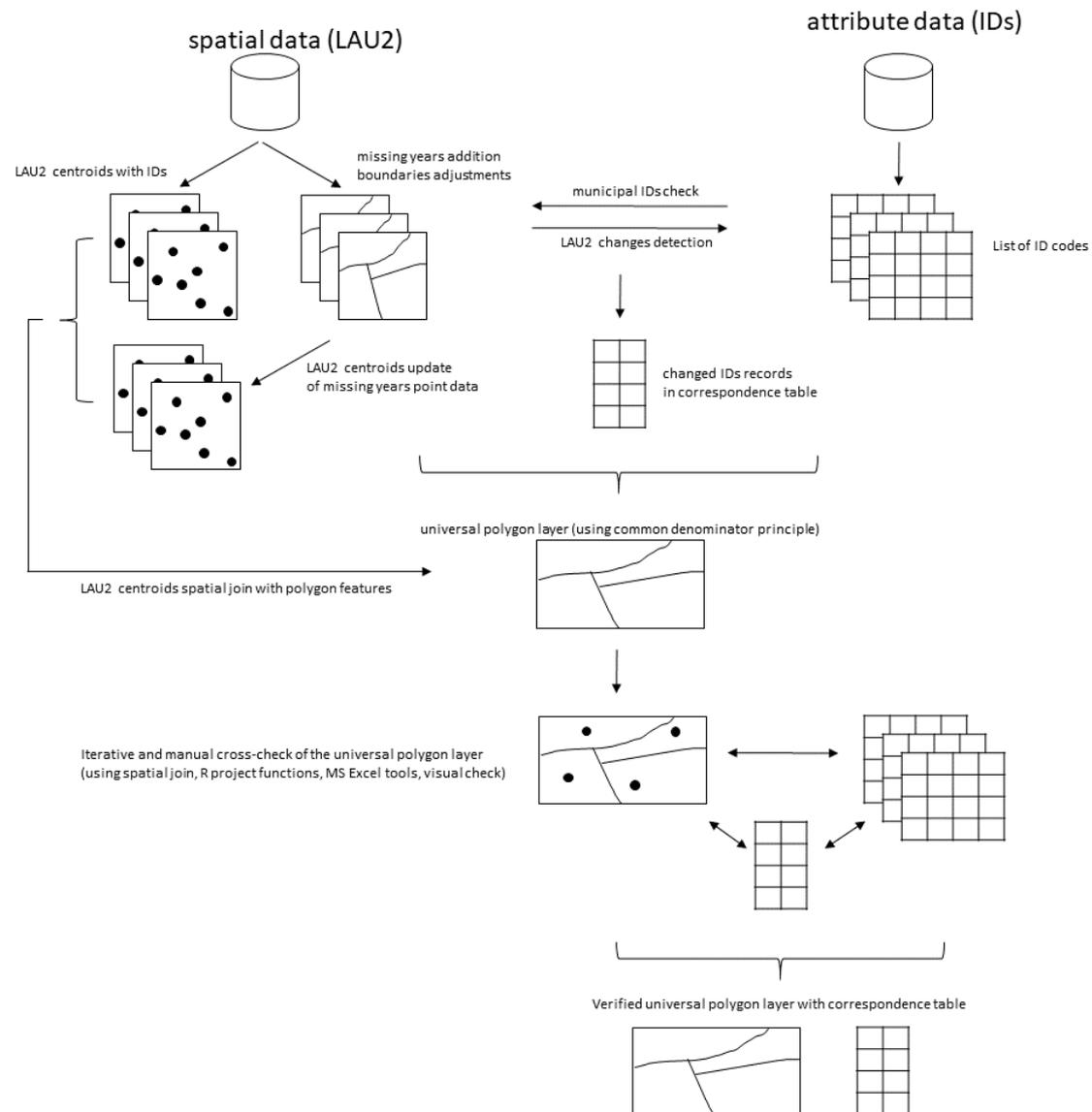


Figure A1. Graphical overview of the data curation work flow.

References

1. European Parliament. *Regulation (EC) No 1059/2003 of the European Parliament and of the Council of 26 May 2003 on the Establishment of a Common Classification of Territorial Units for Statistics (NUTS)*; European Commission: Brussels, Belgium, 2008; Volume 311118, pp. 1–411.
2. Gregory, I.N. Time-variant GIS databases of changing historical administrative boundaries: A European comparison. *Trans. GIS* **2002**, *6*, 161–178. [[CrossRef](#)]
3. Deichmann, U.; Balk, D.; Yetman, G. Transforming population data for interdisciplinary usages: From census to grid. *CIESIN Columbia Univ. Work. Pap.* **2001**, *200*, 1–19.
4. Monteiro, J.; Martins, B.; Murrieta-Flores, P.; Pires, J.M. Spatial Disaggregation of Historical Census Data Leveraging Multiple Sources of Ancillary Information. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 327. [[CrossRef](#)]

5. Quatrini, V.; Tomao, A.; Corona, P.; Ferrari, B.; Masini, E.; Agrimi, M. Is new always better than old? Accessibility and usability of the urban green areas of the municipality of Rome. *Urban For. Urban Green.* **2019**, *37*, 126–134. [CrossRef]
6. Calka, B.; Nowak Da Costa, J.; Bielecka, E. Fine scale population density data and its application in risk assessment. *Geomat. Nat. Hazards Risk* **2017**, *8*, 1440–1455. [CrossRef]
7. Alvioli, M. Administrative boundaries and urban areas in Italy: A perspective from scaling laws. *Landsc. Urban Plan.* **2020**, *204*, 103906. [CrossRef] [PubMed]
8. Hamilton, R.; Rae, A. Regions from the ground up: A network partitioning approach to regional delineation. *Environ. Plan. B Urban Anal. City Sci.* **2020**, *47*, 775–789. [CrossRef]
9. Pászto, V.; Brychtová, A.; Tuček, P.; Marek, L.; Burian, J. Using a fuzzy inference system to delimit rural and urban municipalities in the Czech republic in 2010. *J. Maps* **2015**, *11*, 231–239. [CrossRef]
10. Pászto, V.; Burian, J.; Marek, L.; Voženílek, V.; Tuček, P. Membership of Czech municipalities to rural and urban areas: A fuzzy-based approach. *Geogr. CGS* **2016**, *121*, 156–186. [CrossRef]
11. Kemp, K.; Charlton, M. Modifiable Areal Unit Problem (MAUP). In *Encyclopedia of Geographic Information Science*; Sage: Thousand Oaks, CA, USA, 2014; pp. 169–174.
12. Netušil, F.J.; Netušil, F. Vitální index obyvatels na Moravě a ve Slezsku. *Anthropologie* **1927**, *5*, 45–54.
13. Vondráková, A.; Voženílek, V. Concept of a Formalized Record of GIS Data Visualization for Map Creation. In Proceedings of the 15th International Multidisciplinary Scientific Geoconference SGEM, Albena, Bulgaria, 18–24 June 2015; pp. 765–770.
14. James, L. Defining Open Data. Available online: <https://blog.okfn.org/2013/10/03/defining-open-data/> (accessed on 9 September 2020).
15. Néték, R.; Loesch, B.; Christen, M. OpenWebGlobe-virtual globe in web browser. In Proceedings of the International Multidisciplinary Scientific GeoConference: SGEM, Sofia, Bulgaria, 16–22 June 2013; pp. 497–503.
16. Openshaw, S.; Taylor, P.J. The modifiable areal unit problem. *Quant. Geogr. A Br. View* **1981**, 60–69.
17. Néték, R.; Dostálová, Y.; Pechanec, V. Mobile Map Application for Passportisation of Sugar Beet Fields. *List. Cukrov. Reparske* **2015**, *131*, 137–140.
18. Brychtova, A.; Paszto, V.; Marek, L.; Panek, J. Web-design evaluation of the crisis map of the Czech Republic using eye-tracking. In Proceedings of the International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM, Albena, Bulgaria, 16–22 June 2013; Volume 1.
19. Popelka, S.; Herman, L.; Řezník, T.; Pařilová, M.; Jedlička, K.; Bouchal, J.; Kepka, M.; Charvát, K. User evaluation of map-based visual analytic tools. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 363. [CrossRef]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).