Strategie Geoinformation Schweiz
Stratégie suisse pour la géoinformation

Assessment of Linked Geodata Needs in the Linked Data Community

Final Report

18.03.2025

Bern University of Applied Sciences
Business School
Institute Public Sector Transformation (IPST)

Imprint

This report was commissioned by the Swiss Federal Office of Topography (swisstopo) and prepared by the Institute Public Sector Transformation at the Business School of Bern University of Applied Sciences.

It was commissioned as part of the Action Plan 2024 (Action 2-24-06) of the Swiss Geoinformation Strategy (SGS) and the Federal Office of Topography (swisstopo). This report contains the observations, opinions, results, and conclusions of researchers and experts in the field of Linked Data.

Author: Dr. Jurek Müller, Institute Public Sector Transformation, Bern University of Applied Sciences

Contact person at swisstopo: Pasquale Di Donato

Table of Contents

1 Project Background

To date, Linked Geodata in Switzerland is mostly provided by the Federal Spatial Data Infrastructure (FSDI) through their GeoSPARQL triplestore[1]. GeoSPARQL[2] is a standard and a query language that enables geospatial queries on RDF triplestores and thus combines the analytic strengths of knowledge graphs and geoinformation systems. However, as the transformation to RDF including a rigorous metadata annotation is time and resource-intensive, to date only a small amount of the Swiss Open Geodata is available as Linked Data. At the time of this report, the FSDI provides 16 Linked Geodata sets which is dwarfed by the ca. 1'000 datasets published on map.geo.admin.

The goal of this project was to identify geodata categories and geodata sets that should be prioritized for transformation and publication as Linked Geodata based on the actual needs of the Linked Data community in Switzerland. With this, the impact of future resources spent for the transformation to Linked Data can be maximized.

To achieve this goal a survey among the Linked Data community was conducted followed by a quantitative and qualitative analysis of the results. Based on this analysis a ranked list of geodata categories and geodata sets was compiled.

Section 2 describes the methodology of the survey and its analysis. The survey results are then presented and discussed in Section 3. The limitations of this study are discussed in Section 4. Finally, the results and insights are summarized in the conclusion.

2 Approach

The data collection consisted of an online survey among the Linked Data Community.

2.1 Survey

The online survey was created using the survey tool Qualtrics[3] and was distributed as described in Section 2.2. The survey was conducted from 13.11.24 to 16.12.24 and was available in English, French, and German.

It was divided into three main sections. In the first section, the respondents were asked about their professional context as well as their familiarity with Linked Data, Geodata, and Linked Geodata.

The second section asked the respondents to list any Liked Geodata sets that they were missing during their previous professional work or that they perceive as especially valuable when published. At this point, the goal was to get answers without prior priming with example categories or datasets.

Finally, in the third section of the survey, respondents were asked to select geodata categories that they perceived to provide the most value when published as Linked Geodata from a predefined list. Respondents were then asked to rank their selected categories from most to least important and for each provide example datasets and their potential use cases.

The list comprised a total of 26 geodata categories from three domains: "Base Information and Planning" (8), "Nature and Environment" (8), and "Population and Economy" (10). These correspond to the eCH-0166[4] Standard also used by the Swisstopo geodata catalog on maps.geo.admin.ch. To clarify the scope

[1] https://www.geo.admin.ch/en/linked-data-service-linking-geodata

[2] https://www.ogc.org/publications/standard/geosparql/

[3] https://www.qualtrics.com/

[4] http://www.ech.ch/sites/default/files/dosvers/hauptdokument/STAN_d_DEF_2013-09-23_eCH-0166_V1.1_Geokategorien.pdf

of each category for the respondents, the category names were supplemented with short descriptions crafted based on the datasets available on maps.geo.admin.ch under the respective categories.
At the end of the survey, respondents were asked to leave their email addresses for potential follow-up questions.
A copy of the survey questions is attached at the end of this report.

2.2 Reaching the Linked Data Community

To reach the still relatively small Linked Data community in Switzerland with the survey, different outreach approaches were combined.

1. The Institute Public Sector Transformation (IPST) is an active member in the Linked Data community, regularly organizing community events such as the Linked Data Day or the Linked Data Meetup. It also maintains a Linked Data newsletter with 212 members at the time of the survey. This newsletter mailing list served as the primary distribution channel for the survey.
2. Additionally, through manual search on Google and LinkedIn 10 organizations and 4 additional individuals were identified that work with Linked Data professionally. They were contacted individually and asked to forward the survey to any interested parties.
3. To reach any Geodata experts with affinity to Linked Data, the survey was posted in relevant Geodata Forums.
4. Finally, to reach any persons or organizations outside this network, the survey was promoted on LinkedIn via the IPST account.

These combined efforts resulted in 96 respondents of which 29 fully completed the survey. Roughly estimated based on the number of survey invitations sent by mail, this corresponds to an estimated response rate of about 12.8 %, which, considering the specificity of the topic, is an expected rate.

2.3 Analysis

To allow consistent interpretation of the results across and between all survey questions, partial responses (67 of 96 responses) were excluded from the analysis. Additionally, for the main part of the analysis, respondents who indicated to possess no familiarity with the topic of Linked Data (3 respondents) were also excluded (see Section 3.1). The quantitative analysis of the survey as well as the data visualizations included in this report were done using Python.

For the analysis of question N.Q6, where respondents were asked to rank previously selected geodata categories, the average rank for each category was calculated, by assigning the last possible rank (26), in cases where the category was not previously selected (see Section 3.3). For the calculation of the weighted average rank the assigned rank was multiplied by the results of question F.Q2 (“How familiar are you with the topic of Linked Data?”) for which values from 0 (“Not familiar at all”) to 1 (“Extremely familiar”) were assigned.

For the analysis of the free-text questions, answers were coded and categorized as follows:

- Answers to the question F.Q1a (“What industry do you work in?”) were grouped into four main categories: Administration, IT / Data, Research, and Other (see Figure 2).
- Datasets and topics mentioned in the answers to the questions N.Q1 (“In your previous contact with Linked Data, were there certain geodata that you would have liked to be available as Linked Geodata?”) and N.Q2 (“What types of geodata do you think would have the greatest use value for you or the greater public if they would be made available as Linked Geodata?”) were assigned to the 26 geodata categories described in Section 2.1. Additionally, the explicit data sets mentioned were harmonized and grouped where appropriate to allow further analysis (see also Section 3.2).
- Text answers for N.Q6 (“Rank the previously selected categories of geodata from most use value (top) to least use value (bottom) as Linked Geodata.”) describing example datasets were similarly harmonized.

Excel tables listing the original answers and coded values accompany this report.

3 Results

3.1 Professional Background of Respondents

To shed light on the professional background of the respondents, questions at the start of the survey asked about the respondent's industry (F.Q1a), their job title (F.Q1b), and what topics and themes their work focuses on (F.Q1c). The responses to the first question are visualized in the form of a word cloud in Figure 1, showing a large representation among the administration, research, and the IT / Data domain. For this analysis, answers were categorized accordingly.

Figure 1: Word cloud of answers to question F.Q1a (industry)

Figure 2 confirms that the most represented group is administration employees followed by respondents from IT / data industries and research. Answers categorized as "Other" include health, archives, education, construction, and more.

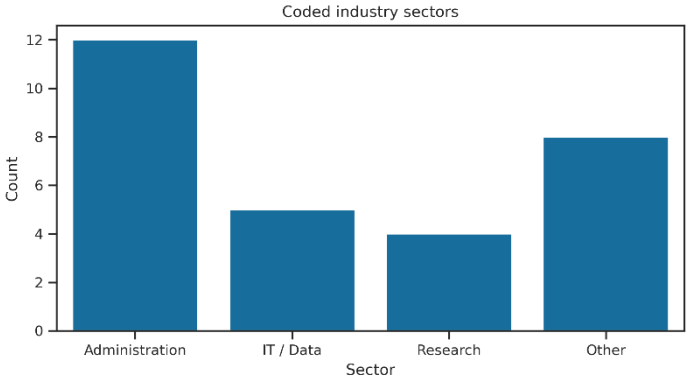

Figure 2: Answers to question F.Q1a (industry) grouped into four categories.

Indicated job titles from Question F.Q1b reveal that most respondents from administration and research also work on the topic of "data", either in a conceptual, organizational, or technical capacity.

This is further highlighted by the results from question F.Q1c about the thematic focus of the respondents. The word cloud visualization in Figure 3 shows a strong focus on data topics including Open Data, Geodata, Linked Data, Data Governance and Data Infrastructure.

Taken together, these results show that the survey managed to reach data experts and data users both from within and outside of the public administration.

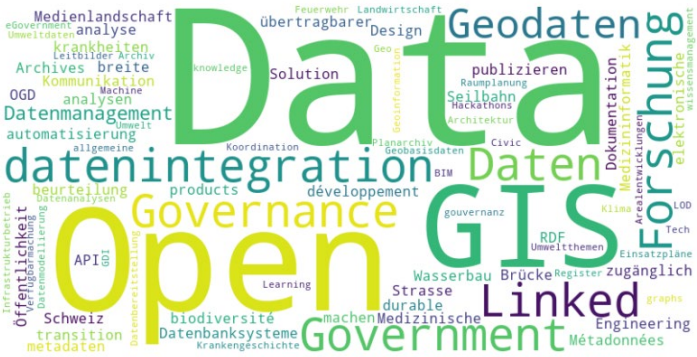

Figure 3: Word cloud of answers to question F.Q1c (thematic focus).

3.2 Familiarity with Linked (Geo)data

To assess if respondents belong to the target audience of this study, that is the Linked Data Community, the survey included a question that asked for the level of familiarity with the topic of Linked Data (F.Q2). Figure 4 shows that the level of familiarity is diverse, but that the majority (72 %) of respondents are at least moderately familiar with the topic with 19 % even indicating extreme familiarity.

Three respondents indicated having no knowledge of Linked Data, which is why they were excluded from the subsequent analysis presented here and in Section 3.3.

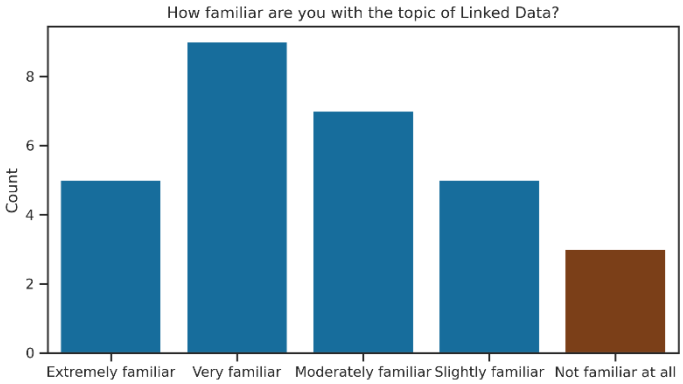

Figure 4: Self-reported familiarity with the topic of Linked Data. Respondents reporting no familiarity were excluded from the subsequent analysis.

With respect to the investigated topic of Linked Geodata the survey posed an additional question to assess the current prevalence of geodata within the Linked Data community. Respondents were asked to indicate how often they work with Geodata, Linked Geodata, and GeoSPARQL.

Figure 5 shows that while all considered respondents work with geodata at least rarely, their touchpoints with Linked Geodata are infrequently and less than half (46 %) of the respondents considered have worked with GeoSPARQL before. One reason might be the still limited supply of Linked Open Geodata in Switzerland.

Taken together, the results show that the survey reached the target audience of Linked Data experts, who all indicated experience with geodata. However, familiarity and practical experience with the specific subset of Linked Geodata is less common.

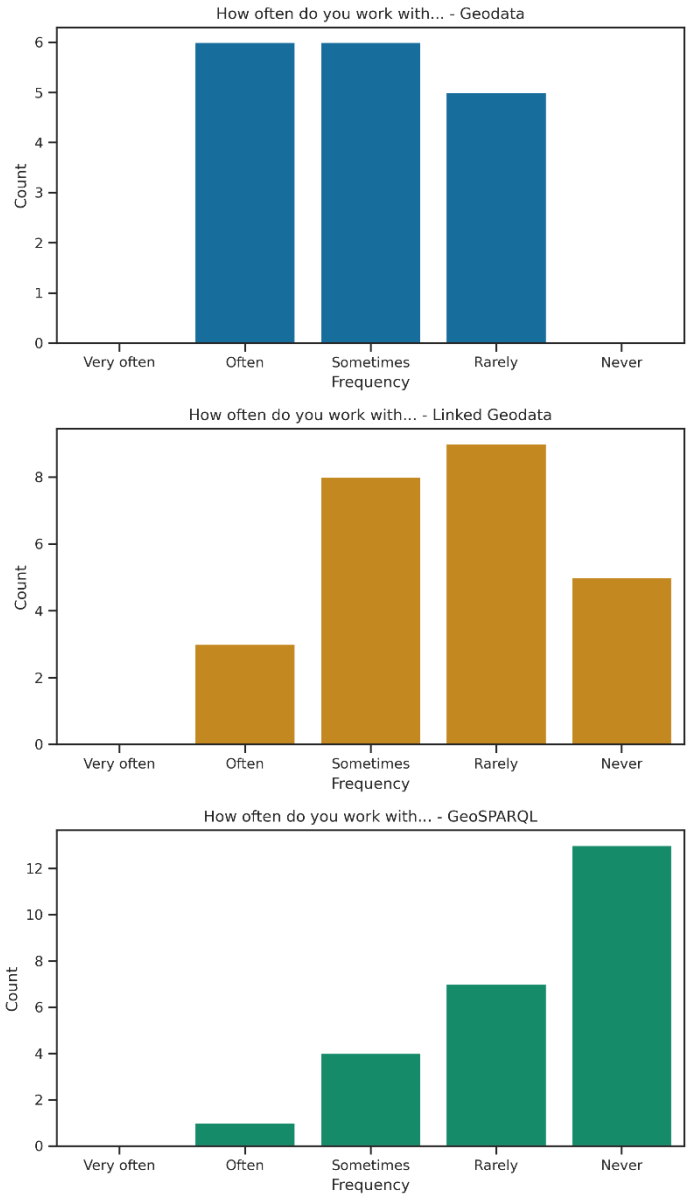

Figure 5: Self-reported frequency of how often the respondents work with geodata, Linked Geodata, and GeoSPARQL.

3.3 Mentioned and Selected Categories

Before priming respondents with lists and descriptions of geodata categories, the survey posed two free text questions that asked about Linked Geodata sets that the respondents missed during their previous work with Linked Data (N.Q1) and geodata which, in their eyes, would deliver the most use value when transformed to Linked Geodata (N.Q2).

The mentioned types of geodata or specific datasets were grouped into the 26 geodata categories described in Section 2.3. Figure 6 shows how often individual categories were mentioned at this stage of the survey. The top three categories mentioned are “Boundaries”, “Building, Infrastructure”, and “Transportation”.

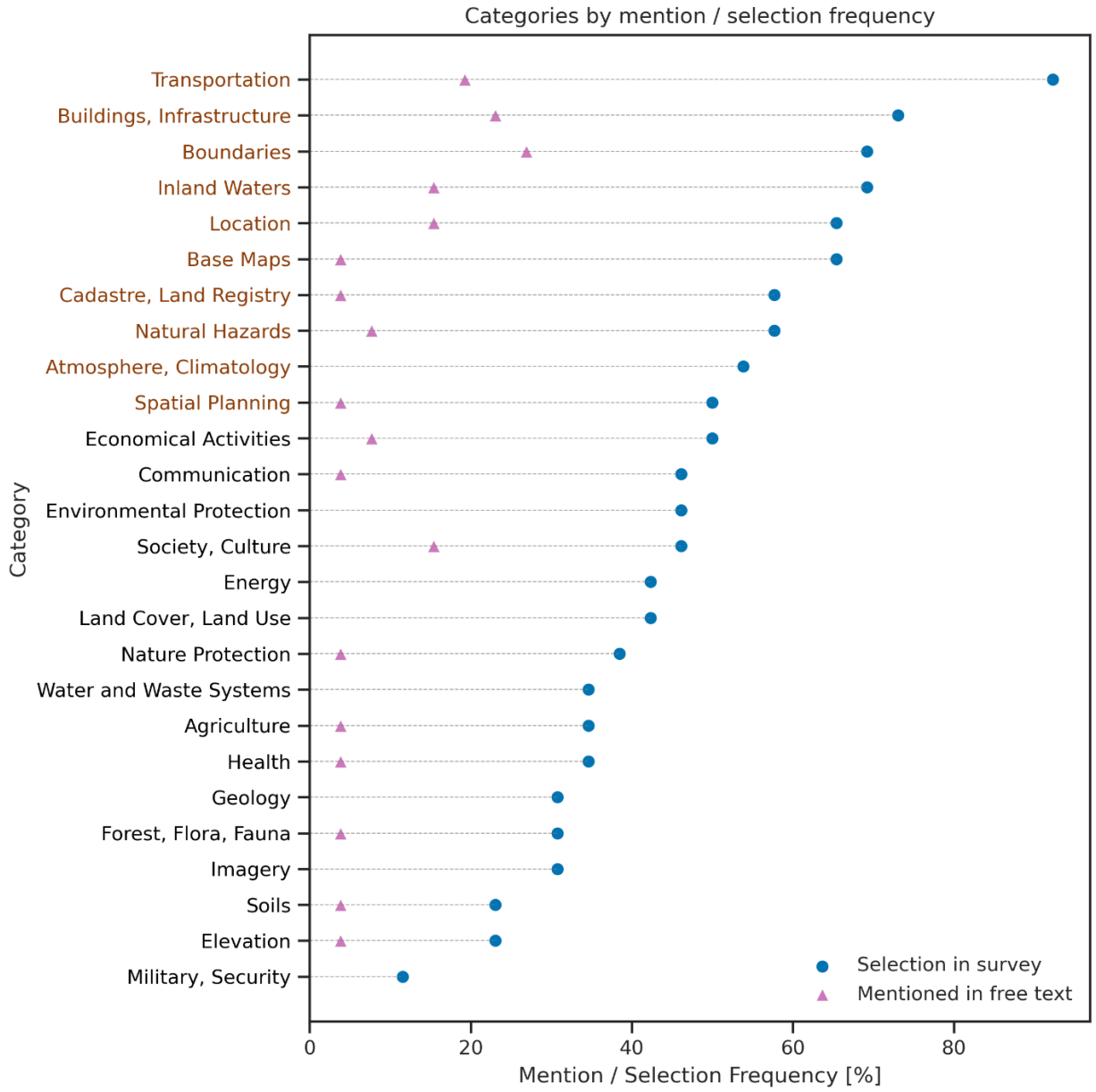

Figure 6: Frequency of categories mentioned in the free text questions N.Q1/2 (triangles) and selected in the questions N.Q3-5 (circles). The top 10 categories are colored orange.

Next, respondents were presented with a list of geodata categories and descriptions per one of the three domains (N.Q3-5) and asked to select the categories that would have the most use value as Linked Geodata. Figure 6 shows how often each category was selected. Compared to the free text mentions, we find that the top three categories remain the same, however with reversed order and with “Inland Waters” joining “Boundaries” in a shared third place. In first place, “Transportation” was selected by 92 % of the analyzed respondents.

Taken together these results show that the top selected categories are also the categories most often perceived as valuable with respect to Linked Geodata even before any priming with the category list and descriptions.

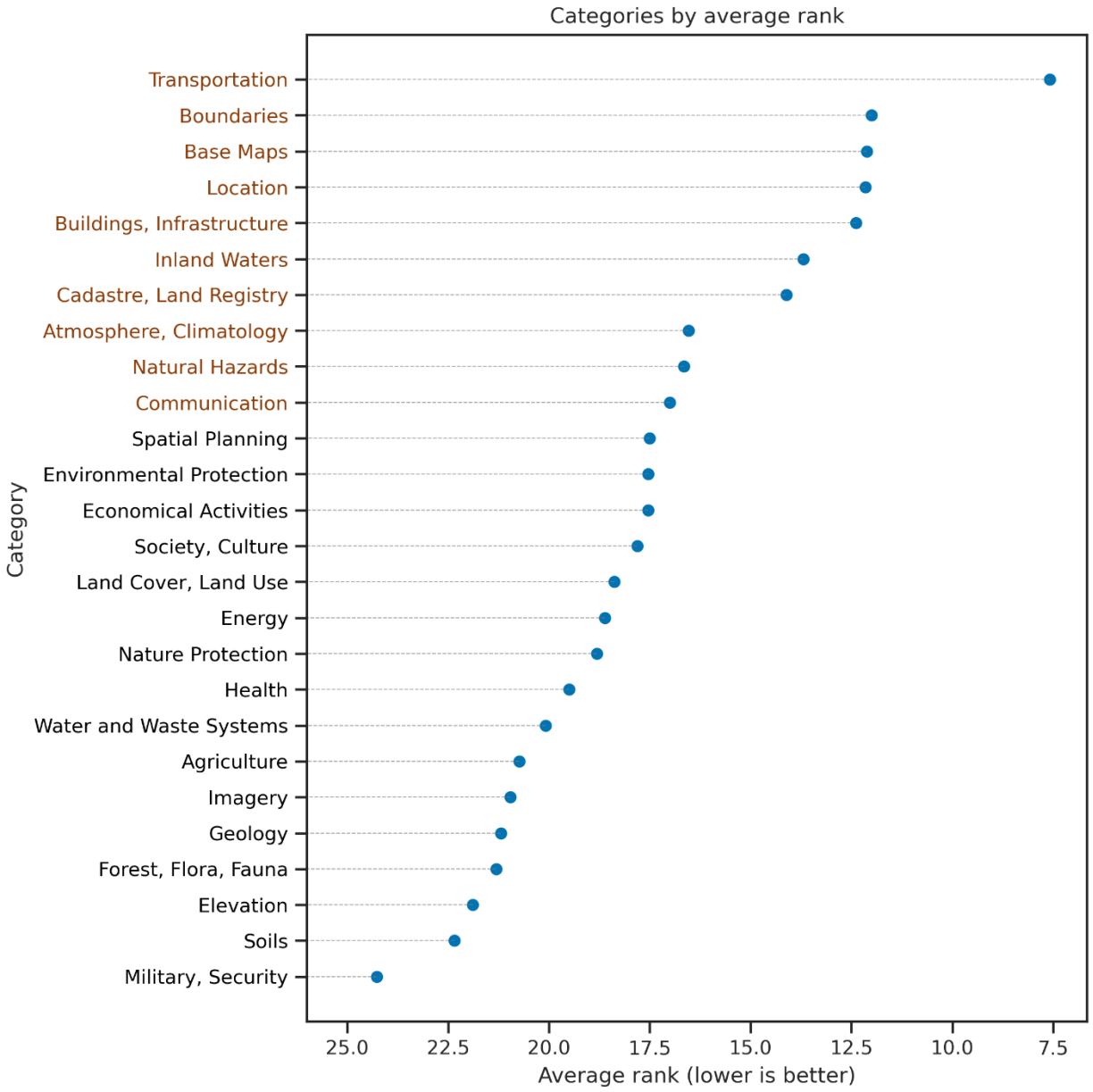

Figure 7: Geodata categories by calculated average rank (see Section 2.3). The top 10 categories are colored orange.

3.4 Ranked Categories

After selecting the most valuable geodata categories, respondents were asked to rank their selection using drag-and-drop interactions (N.Q6). Based on the individual rankings, the average rank for each category was calculated as described in Section 2.3. The results are shown in Figure 7.

It shows a similar ranking of the top 10 categories compared to when considering selection frequency alone (Figure 6), although with some categories rising (e.g. “Base Maps”) or dropping (e.g. “Buildings, Infrastructure”) within that group. Sorted after average rank the top three categories are “Transportation”, “Boundaries”, and “Base Maps”.

As this study aims to highlight the needs and opinions of Linked Data Experts and Practitioners, in a final step, a weighted average rank for each category was calculated by using the self-reported familiarity of the respondents with the topic of Linked Date (Figure 4) as weights to the respondents ranking (Figure 8).

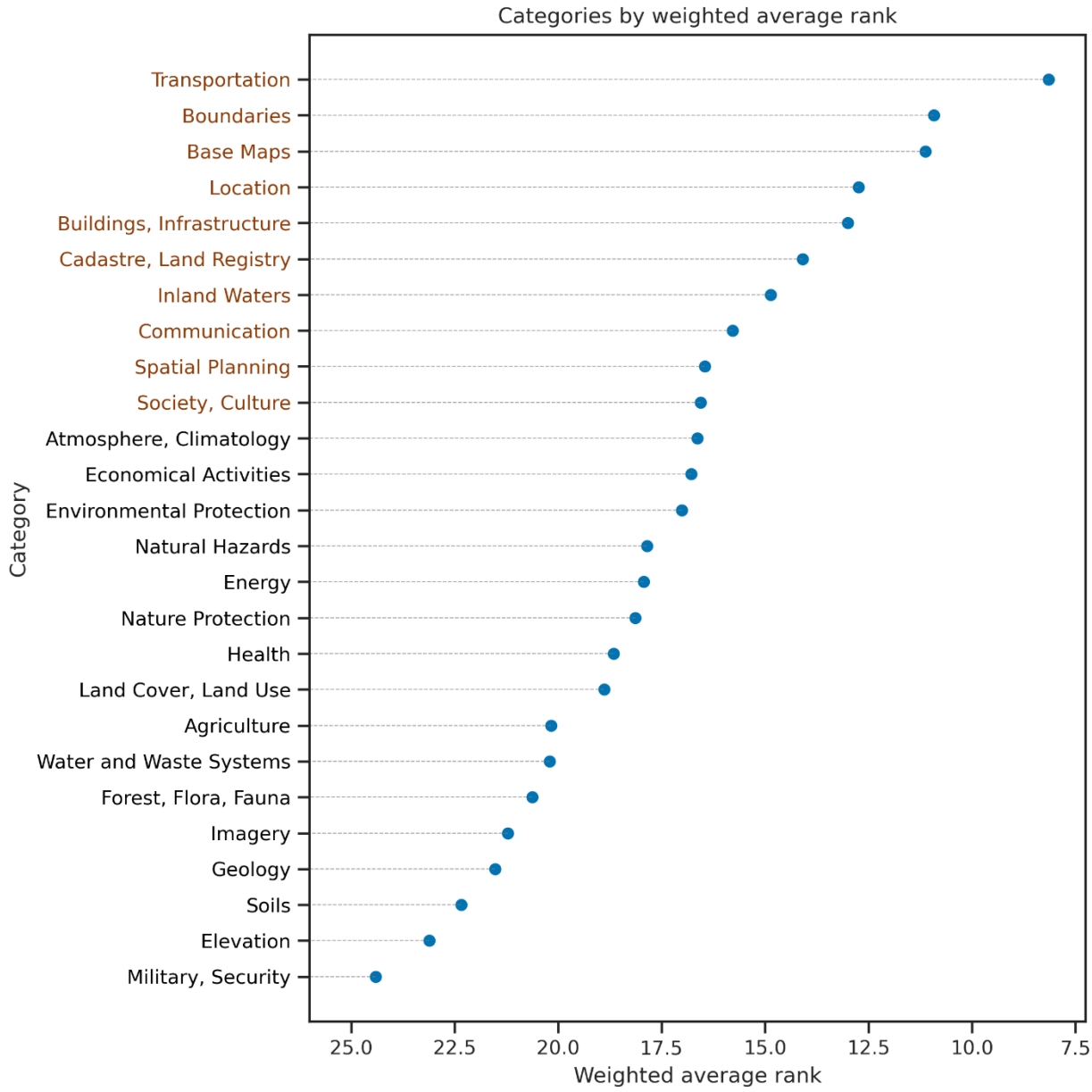

Figure 8: Geodata categories by calculated average rank weighted by the familiarity of respondents with the topic of Linked Data (see Section 2.3). The top 10 categories are colored orange.

Figures 7 and 8 show that the top 5 categories are identical independent of the use of the average or the weighted average rank, lending further confidence in the robustness of the results. Places 5-6 are also similar although in a different ordering. The largest difference between the two calculation methods is the category of "Society, Culture" which climbs from place 14 using the average to place 10 using the weighted average.

Across methods, the geodata categories perceived as the least valuable as Linked Geodata are consistently “Soils”, “Elevation”, and “Military, Security”.

Categories	Weighted Average Rank	Average Rank	Harmonized Datasets (N.Q6)	Harmonized Datasets (N.Q1 & N.Q2)
Transportation	5.173076923	7.576923077	Traffic data (2x); Street network (2x); Transportation network CH (2x); Transprotation Infrastructure; SBB plan; Airspaces; Transportation hubs; Public transport stops; Traffic Infrastructure	swissTNE; Accidents; Transportation network CH
Boundaries	6.932692308	12	Administrative borders (Municipalities, Districts, Cantons, Postal Code, etc.) (6x); Historisized administrative borders (2x); Language Regions	Administrative borders (Municipalities, Districts, Cantons, Postal Code, etc.) (4x); Historisized administrative borders (2x)
Base Maps	7.057692308	12.11538462	Vektormaps (2x); OpenStreetMap; swissTLM	OpenStreetMap

Figure 9: Excerpt of the excel file accompanying this report that lists the datasets mentioned by respondents per category.

3.5 Dataset Identification

As part of the question to rank their selection (N.Q6) respondents were asked to provide example datasets for each category that hold the most potential use value as Linked Geodata. These free-text answers were harmonized and grouped to allow an analysis of how often certain datasets or groups of data were mentioned. Mentions of municipal, cantonal, or postal code boundaries for example were harmonized and grouped into “Administrative borders (Municipalities, Districts, Cantons, Postal Code, etc.)”.

The same methodology was applied separately for the specific datasets mentioned in the free text questions N.Q1 and N.Q2 (see also section 3.3). The resulting harmonized datasets and the number of their mentions were consolidated into an Excel file that accompanies this report. Figure 9 shows the respective data for the top 3 categories by weighted average rank.

Noticeably, this list of mentioned datasets includes datasets that are already available as Linked Geodata through the FDSI Linked Geodata service. These include some of the most mentioned datasets, such as addresses or administrative borders of municipalities, districts, and cantons, as well as less mentioned datasets, such as traffic accidents, radio/TV coverage areas, or public transport stops. On the one hand, this indicates that the FDSI is already covering part of the demand from the Linked Data community, but on the other hand, the Linked Data community seems to be not fully aware of the already existing supply of Linked Geodata.

The free text answers also include references to OpenStreetMap (OSM) data that some respondents would like to be able to link with the data offered through the FSDI. Two respondents refer to the QLever project as an example, which, among others, offers the full OSM data as Linked Geodata[5]. Enabling geospatial queries that link OSM data with Linked Geodata from geo.ld.admin could potentially greatly increase the use value of the published datasets.

Taken together, the results as discussed above and as shown in Figure 9, will allow to tailor the future publication of Linked Geodata sets to the needs of its primary user base, the Linked Data community.

[5] https://qlever.cs.uni-freiburg.de/osm-planet

4 Limitations

The largest limitation of this study is the small sample size of the survey, which results from the still limited size of the Linked Data Community in Switzerland and the even more specific topic of Linked Geodata. However, although the small sample size does not allow for rigorous statistical analysis, the results of the survey provide the FSDI with an empirical basis for identifying and prioritizing datasets for future Linked Geodata publishing.

5 Conclusion & Recommendations

This study investigated the needs and opinions of the Linked Data community with respect to the topic of Linked Geodata. The goal of this study was to identify geodata categories and datasets that match the needs of Linked Data users, allowing to maximize impact with future Linked Geodata publications. This goal was achieved by conducting and analyzing an online survey among the Linked Data community. The analysis included investigating the professional background of respondents, creating ranked lists of geodata categories based on respondents' mention and selection frequency, average rank, and weighted average rank as well as identifying requested datasets per category based on free-text answers.

The analysis of the professional background showed that, even in the Linked Data community, touchpoints with Linked Geodata and technical tools like GeoSPARQL are still rare. This may partly be attributed to the still relatively small amount of data that is available as Linked Geodata but might also be grounded in a still generally low knowledge about the specificities of Linked Geodata and its advantages. The offering of additional tutorials, courses, or showcases might thus increase the use of already published datasets and thus their impact.

The analysis of the geodata categories showed that there are categories that are widely recognized as especially valuable in the context of Linked Geodata, such as “Transportation” or “Boundaries”, and that this result is relatively robust with regard to the applied metric. Category rankings as well as identified datasets are consolidated in an Excel file accompanying this report. This data will form an empirical basis for the prioritization of the publishing of new Linked Geodata sets through the FSDI.

The category with the largest perceived use value as Linked Geodata across all calculation methods is “Transportation”. The specific datasets mentioned for this category reveal that the desired data mostly relates to the multimodal network of transportation including transportation infrastructure. The networked nature of this data makes it an ideal candidate for publication as Linked Geodata. The FDSI, as of the time of this report, offers only a small part of this information as Linked Data, such as the public transport stops or the official street directory. Transforming and Publishing datasets emerging from the undergoing initiative “Transportation Network CH”[6] such as the “swissTNE Base”[7] dataset which provides a reference network of the Swiss multimodal transportation system, could thus be a valuable first priority.

The analysis also showed that some datasets that were mentioned are already published as Linked Geodata through the FSDI, above all current and historic administrative borders. The expressed demand for these datasets underlines their importance but also suggests that knowledge about their availability as Linked Geodata is not as widespread within the Linked Data community. Additional measures to disseminate and advertise the Linked Geodata offer of the FDSI thus might be needed to maximize the impact of past and future Linked Geodata Publications.

[6] https://www.swisstopo.admin.ch/en/project-transportation-network-ch

[7] https://www.swisstopo.admin.ch/en/landscape-model-swisstne-base