

# Simplifying the Publishing Process for Open Geospatial Data

Andreas Divanis <sup>a,\*</sup>

<sup>a</sup> Chair of Cartography, Technical University of Munich, Munich, Germany – andreas.divanis@tum.de

\* Corresponding author

**Keywords:** open data, geospatial data, guidelines, accessibility

**Abstract:**

The publication of data as open data is driven by the idea that it will bring advantages to administrations, governance, public services, businesses, and citizens alike. The European Commission's "A European Strategy for Data" considers open data an essential resource. Today, actors worldwide participate, use, and promote open data. In many cases, data are now required by law to be published as open data, like in the case of the German "Act to promote electronic government" (EgovG), or making data open is a requirement for funding and partnerships when conducting research.

In 2023, Hahnel et al. published a report about the state of open data, an eight-year-long running survey and analysis of open data in the research community, the attitudes of the researchers towards sharing data, and their reported experiences. The results highlight some key issues. Only 23% of respondents reported receiving some form of support for making their data open, and the majority reported that they never received any support for sharing data. Even while attitudes toward opening data are positive, researchers need support in creating open data.

Furthermore, many open data guidelines, such as FAIR, formulate recommendations for data in general, but as researchers who work with Geospatial Data, we face some unique challenges. The data we produce comes in many dimensions, quantities, sizes, and different necessities for verifying quality dimensions, or we face restrictions due to business interests, security, and privacy concerns. Furthermore, spatial data pose difficulties with keeping data up to date, as we often physically have to move a human or a camera to a location to check the ground truth – a huge logistic challenge, the larger the dataset and the spatial spread. Furthermore, spatial data can cause privacy concerns, such as movement patterns making an individual's home or workplace visible.

As researchers in cartography, we have grappled with the issue of missing guidelines for creating good open spatial data. Therefore, we wanted to find out if we could devise a workflow that could guide us step by step while sharing spatial data in particular.

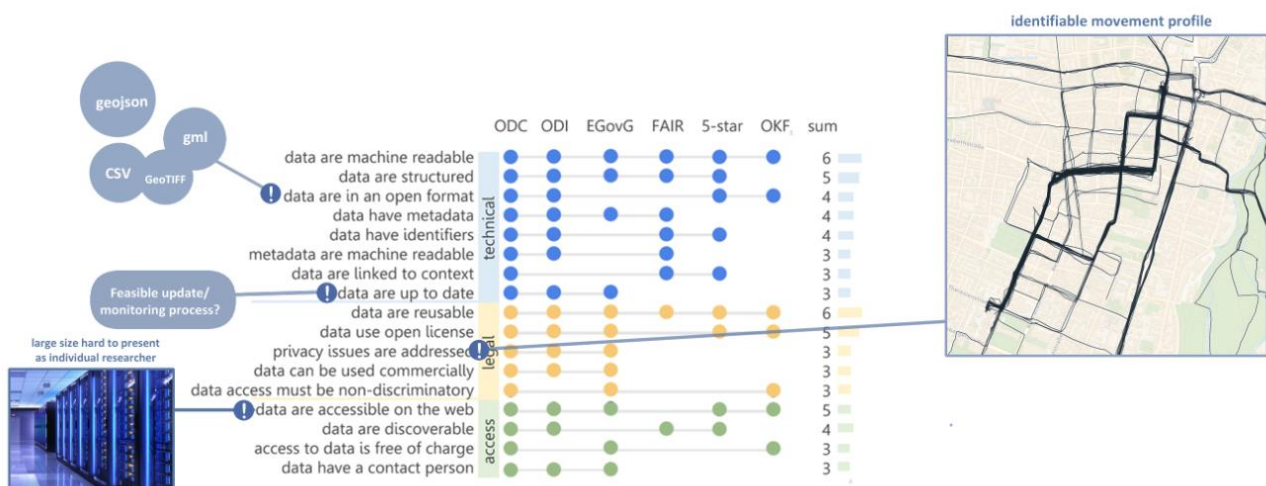


Figure 1. Subset out of generalized 42 requirements, filtered by "at least half agree" with some spatial data particularities

A significant aspect that makes publishing open data challenging for an individual researcher is the multitude of different technical, legal, and accessibility requirements that commonly used standards recommend. In other words, if a

researcher wants to publish open data in an accessible way, the researcher needs to be aware of many different heterogeneous requirements that each need different data preparation, processing, and legal steps.

Our work has compared requirements for publishing open data sets by six actively used institutions, organizations, and well-known practices. We compared the Open Data Institute's Data Certificate metrics, the Open Data Charter, the FAIR principles, the 5-Star Deployment Scheme for Open Linked Data, the Open Knowledge Foundation, and the German Act to promote electronic government. After comparing and generalizing the requirements defined there, we still found 42 unique requirements for the data. With an "at least half" approach, we defined a core set of 17 requirements (Fig. 1).

With two sets of our research data (a phenological point data set of a Bavarian citizen science project), we compared difficulties of trying to fulfil the requirements for multiple standards in a row versus preparing our data according to this core set of requirements and adjusting if we could match the individual standards from there. This saved us multiple days in preparing open data for different standards.

However, we noticed that the exact workflow still has to be adjusted, because in some cases, the early steps of the data collection process can influence whether we meet a specific standard (for example, how to assign rights to the data).

In this initial test, the spatial aspects of the data mostly came into play for the issues of monitoring and privacy protection. The dataset needed to be processed in more ways than expected, as we had to split the user information of the collector from the dataset to maintain privacy, as some (anonymized) users' data collection patterns were allowing to pinpoint their residence's location.

Furthermore, we still need to prioritize what recommendation we will give for data structure and types to retain the easiest way to adjust for individual variation in data preparation.

While our initial results look quite promising in simplifying the publishing process of open spatial data and benefiting researchers and data providers, additional tests with a larger pool of participants with more diverse spatial data should be our next step.