

Metadata for an Ethical Data Practice in Cartography

Ester Scheck^{a,*}

^a TU Wien, ester.scheck@geo.tuwien.ac.at

* Corresponding author

Abstract: As with each map, which is the product of a design and decision-making process, including the selection of data to be used, each dataset is created in a specific context where conscious and unconscious choices are made. Reflection on this data context facilitates comprehension of the data, its assertions, limitations and potential biases and, consequently, enables more ethical and responsible decisions to be made during the map production process. In order to identify the relevant context information and its potential formats, various documentation and reflection frameworks are examined and related to the ISO standard on metadata for geographic information (ISO 19115:1-2014). The analysis demonstrates that the ISO standard provides a comprehensive metadata model encompassing numerous aspects of contextual information from the frameworks. However, the analysis also reveals some gaps and differences in the format. To illustrate the availability of context information in metadata, a small use case with three open datasets is analyzed. The results indicate that key aspects of the context of origin are missing, thereby limiting the datasets' evaluation. Finally, the paper outlines potentials and requirements for integrating and further developing context information frameworks and their use in cartography.

Keywords: metadata, open data, critical cartography, data context, ethics

1. Introduction

"The Numbers Don't Speak for Themselves", state Catherine D'Ignazio and Lauren F. Klein (2020, p. 149) in their book *Data Feminism*, referring in part to Donna Haraway's (1988) theory of situated knowledge. Data should not be seen as neutral or objective but as "products of unequal social relations, and this context is essential for conducting accurate, ethical analysis" (D'Ignazio & Klein, 2020, p. 149). As data, whether quantitative or qualitative, downloaded or self-collected, is the basis of any map, this assertion is just as relevant to cartography as it is to data science.

D'Ignazio and Klein's claim has many parallels with critical cartography, stating that every map has power and is not objective because of the choices made in the map-making process. I argue that the data used should already be seen as a product based on many conscious and unconscious choices and that the interaction with data at the beginning of any map-making process needs careful attention. Robinson et al. (1995) identify the transformation from the geographical environment to recognized geographical information as the first transformation in their model of fundamental information transformations in cartography. Each transformation, including data production, can affect the final map product. It is therefore important to promote an ethical and critical data practice in cartography. By this, I mean examining and using data in ways that consider the impact of power relations on data production, questioning widely accepted and normalized categories and

definitions, and striving for respectful and fair inclusion and representation of diverse groups of people.

All these aspects relate to the context of data production. Context is a broad and challenging concept to define. The Oxford online dictionary describes context as "the situation in which something happens, and that helps you to understand it" (Oxford Learner's Dictionary, n.d.). In terms of data context, it is possible to consider, for example, the methods of data collection, the definitions of classifications, the aims of data collection, the people involved and the limitations of the data (D'Ignazio, 2022). Looking at the production of data, the social, cultural, historical, institutional and material conditions and the related power relations can help to identify limitations of data, possible biases and a resulting ethical responsibility in the use of data (D'Ignazio & Klein, 2020). In line with the Oxford dictionary definition of context in general, it helps to understand the data.

This awareness is crucial, as a lack of context information can lead to misinterpretation and faulty analyses, confusion about the meaning of terms (e.g. 'forest'), or an overestimation of the level of detail if, for example, aggregation is not made transparent (Comber et al., 2007; Fischer et al., 2023; Krause, 2017; Whitfield, 2012). As data is increasingly used to inform political and economic decision-making, misinterpretation can have real consequences.

However, the context of data is often missing. While, thanks to the open data movement, more and more data is openly available online, documentation and metadata are often lacking (D'Ignazio & Klein, 2020). Moreover, domain-specific knowledge, in terms of geodata, for

example, the underlying conceptions of spatial referencing, cannot be taken for granted (Chmielinski et al., 2020; Devillers et al., 2007). Thus, while access to data is becoming easier, the lack of context and metadata, broadly defined as data about data (Comber et al., 2007), can hinder ethical data practice.

Following the introduction explaining the importance of context information of geodata for cartography, the scope of context information and its relation to metadata will be examined. First, the contextual information most frequently considered relevant in the literature will be analyzed. Documentation and reflection frameworks from critical data studies and ethical Artificial Intelligence (AI) and machine learning (ML) research will serve as a basis. These frameworks are then compared to the ISO metadata standard for geographical data (ISO 19115:1-2014). The investigation then continues with a small use case with three open datasets concerning the information provided in the metadata and its coverage of the previously identified contextual information elements. The work thus transfers a central approach of Critical Data Studies to cartography and provides starting points for further developing an ethical geodata practice.

Thematic maps are the primary focus of this study, given the recognition that geographical data can vary significantly in nature, with topological and thematic data exhibiting distinct characteristics. However, it is acknowledged that certain thoughts and approaches developed in this work may also apply to producing topographic maps.

2. Documentation and reflection frameworks for data context information

The scope of information encompassed by a data context is inherently extensive, and as such, a comprehensive definition cannot possibly be formulated. However, various documentation and reflection frameworks for datasets developed in Critical Data Studies or research on ethical AI and ML can be utilized as a point of departure. A framework, hereafter, is understood as a structured guideline for the analysis and documentation of specific processes, questions and topics. Such frameworks include the *Data Biography*, *Datasheets for Datasets* and *Data User Guides*, as referenced by D'Ignazio and Klein (2020). In addition, further frameworks have been identified from the literature and are listed in Table 1. Following a conceptual comparison of the various frameworks, an analysis will be conducted to identify overlaps in their contents, with aspects not mentioned in most of the frameworks being neglected.

Title	Author or publisher	Type	Comments
Data Biography	Krause (2017, 2023)	Reflection framework	
Data User Guides	Western Pennsylvania Regional Data Center (n.d.) (WPRDC)	Documentation framework	Template from 2015, developed based on needs by data provider
CIVIC Data Library of Context	CIVIC Software Foundation (2020)	Documentation framework	Developed in 2020 in a workshop, only example model available, limited background information
Datasheets for Datasets	Geburu et al. (2018), Microsoft Research	Documentation framework	Developed for AI and ML datasets

Aether Data Documentation	Microsoft's Aether Transparency Working Group (2022)	Documentation framework	Further development of Datasheets for Datasets
Data Nutrition Project	Holland et al. (2020), Chmielinski et al. (2020)	Documentation framework	Partly automated, developed for AI datasets
Data Cards	Pushkarna et al. (2022), Google Research	Documentation framework	Partly automated, developed for AI datasets

Table 1. Overview of the frameworks under comparison

The majority of the examined frameworks provide a template for their application, and these templates exhibit a variety of formats, ranging from a simple text document (Pushkarna et al., 2023; WPRDC, 2015) to a Google Sheets template (We All Count, 2023), a static website (Civic Software Foundation) and a PDF document (Geburu et al., n.d.; Microsoft's Aether Transparency Working Group, 2022; Pushkarna et al., 2023) to a web-based interactive user interface (Data Nutrition Project, n.d.). Within the template, all frameworks, except the *Data User Guide*, which consists of only eight aspects with occasional explanations, are structured in six to nine sections with up to 60 questions. For instance, the *Data Biography* follows five guiding questions: who, what, where, how, and why. The questions of *Datasheets for Datasets* are oriented on the life cycle of a dataset, and the sections comprise motivation, composition, collection process, preprocessing/cleaning/labelling, uses, distribution and maintenance. In certain frameworks, questions are open-ended, while others provide concrete examples or single or multiple-choice answers that offer a clear direction.

The provision of support materials, application examples and background information varies widely across the frameworks. The overarching objective of the frameworks can be categorized as either the critical reflection of a dataset on the part of the data user (*Data Biography*) or the documentation of a data production process by data providers (all documentation frameworks). However, for example, *Datasheets for Datasets* also emphasizes the reflection of data producers throughout the whole data production process. It is important to note that the target group of the frameworks is not always explicitly mentioned but can, in some cases, be derived from the wording and examples.

Four of the examined frameworks originate from scientific research about ethical AI and ML, aiming to promote a critical examination of training data. Of these four, three were developed by companies (Microsoft and Google). Despite this AI and ML background, most of the elements covered in these frameworks can also be applied to other types of datasets. However, none of the frameworks are explicitly designed for geodata.

The heterogeneity in the use of terms and the absence of examples or explanations present significant challenges when conducting a content comparison of the frameworks. It is assumed that, to a certain extent, differing terminology is employed to denote similar concepts or that the same terms are used with different understandings (e.g. data acquisition and data collection, data creator and data owner, lineage and provenance,

dataset type, data type and type of instances). This must be considered when interpreting the overlap in content described below.

Apart from a general description of the dataset, the frameworks frequently inquire about the motivation and purpose of the data collection. Additionally, the financing is often a point of examination. Typical metadata, such as the title of a dataset and its publication date, are often required. In comparison, it becomes evident that a multitude of temporal dates can be provided about a dataset, including the survey period, the time of creation, the covered timeframe, or the date of update. Also related to the 'who' aspect, the varying roles of the dataset creator, owner, publisher and maintainer are captured in different frameworks and are not easily comparable or clearly distinguishable. Notably, some frameworks request the provision of multiple roles, thereby illustrating the varying levels of detail and depth inherent to the frameworks.

Almost all frameworks require the type of data or features, including but not limited to textual information, images or tabular data. The number and content of the features or the data fields should also be noted in some frameworks. Furthermore, the maintenance and updating of the dataset are addressed in most frameworks.

Apart from the concise *Data User Guide*, all frameworks address sensitive data and human attributes that may be contained in the dataset and, as such, necessitate particular care. Specific questions include which demographic groups are described and how they were identified, how sensitive human attributes were captured (e.g. ethnicity, sexual orientation, socio-economic status) and whether individuals are identifiable or confidential data is included. The latter is negligible in open data, as data with these characteristics would not be published. For datasets representing people, four frameworks address the issue of consent to data collection. Four frameworks furthermore inquire about the procedures and considerations that underpin the ethical review of data collection and the potential implications thereof. In certain instances, privacy concerns and the consideration of systemic inequalities are also encompassed.

The data collection process is a recurrent theme in all frameworks, occasionally accompanied by detailed sub-questions concerning tools, mechanisms, technologies and software used, as well as the question of who collected the data. In certain frameworks, this furthermore includes the question of the representativeness of the data and possible limitations as a result, as well as the reference to or dependence on other data sets. The subsequent processing of data in preprocessing and data cleaning is also present in all frameworks, with varying degrees of elaboration. The four frameworks developed for ML processes additionally consider processes for labelling features and attributes.

All documentation frameworks (except for the *Data Biography* as a reflection framework) consider previous, intended or potential and/or unsuitable applications of the

datasets. This is related to terms of use and licenses, which occur in almost all documentation frameworks. The issue of access and publication is also addressed in all frameworks except the *Data User Guide*, which was developed to provide open data and where open access can, therefore, be assumed.

3. Context information and metadata

The examined frameworks aim to provide or collect data about a dataset, which can be described by the term 'metadata'. However, it is noteworthy that the frameworks seldom refer to metadata in terms of both name and concept. Some rare references include the definition of the *Data Nutrition Label* as a "standard format for metadata communication" (Holland et al., 2020, p. 8). The *Data Cards* and *Data User Guides* are described as concepts that extend metadata and the *CIVIC Data Library of Contexts* as an example of contextual metadata. The present chapter will examine the relationship between the frameworks and metadata in more detail after briefly introducing metadata and current related research topics.

Metadata is intended to facilitate three functions for data users: to locate data sources, to assess the usability of a dataset and to integrate multiple datasets (Comber et al., 2007; Henzen et al., 2013; Kalantari et al., 2021). Within the discipline of GIScience, there has been criticism regarding the usability and understandability of metadata, which affects the latter two goals (Devillers et al., 2007; Ziaimatin et al., 2020). Concurrently, there has been increased interest in research into the automated creation and maintenance of metadata to address the labour-intensive nature of metadata documentation and to improve machine readability. (Batcheller et al., 2009; Fischer et al., 2023; Kalantari et al., 2020; Wagner et al., 2021). Nevertheless, the need for metadata to comprehend and evaluate data contexts and dataset biases – and consequently, the limitations of automation, as emphasized by Gebu et al. (2018) – gets little attention.

The definition of metadata scope and content is typically established through specifications, guidelines and metadata profiles for data portals. In the field of GIS, the standards by the International Organization for Standardization (ISO), the guidelines provided by the Infrastructure for Spatial Information in Europe (INSPIRE) as well as the work from the Open Geospatial Consortium (OGC) should be given consideration (Ziaimatin et al., 2020). However, it is important to acknowledge numerous institutions' involvement in providing metadata templates and profiles. In this work, reference is made to the two ISO standards: ISO 19115-1:2014 (Geographic information – Metadata – Part 1: Fundamentals) (International Organization for Standardization [ISO], 2014) and ISO 19157-1:2023 (Geographic Information – Data quality – Part 1: General requirements) (ISO, 2023). OGC also applies ISO 19115-1:2014 (Open Geospatial Consortium [OGC], 2024), and the metadata elements defined in the INSPIRE regulation (1205/2008, 2008) can generally be seen as a subset of

the elements defined by ISO. Consequently, the ensuing discussion will focus exclusively on the aforementioned ISO standards.

ISO 19115-1:2014, the ISO standard for metadata of geographical data, can be applied to geo datasets and geodata services. The standard comprises twelve metadata classes, with data quality defined by its own standard (ISO 19157-1:2023). A few of the classes and attributes within the standard are mandatory to fulfill the standard, but most are optional or conditional. The implementation and encoding of the metadata model are defined in ISO 19139:2019 (Geographic information – XML schema implementation). To gain a comprehensive understanding of the information that can be documented based on the ISO standard, it is insufficient to merely examine the packages, classes, and attributes; code lists provided for certain attributes were also looked at.

A comparison of elements defined in the ISO standards with the context information mentioned in the examined frameworks demonstrates significant overlap. The ISO standard provides a comprehensive metadata model encompassing numerous aspects mentioned in the frameworks and additional geodata-specific components. Its format, however, differs significantly from the format of the frameworks: While the ISO standard offers a rather abstract model for documenting specific information in specific formats, it does not provide a clear guideline with concrete questions to answer, as most frameworks do. The ISO standard is also less accessible through a paid license and in terms of understandability.

In accordance with the reflection and documentation frameworks, the ISO standard permits the documentation of information pertaining to the sources utilized and the data production processes (LI_Lineage), albeit to a lesser degree of concreteness and detail. Participating parties, such as publishers, owners, maintainers or funding organizations, are not defined as individual ISO classes or attributes but can be provided through different role definitions specified in a code list (CI_Responsibility). Similarly, dates for dataset creation, publication, or updates, mentioned in many frameworks, are mentioned in a code list and can be provided (CI_Date).

The ISO standard provides coverage of additional context information elements that are frequently mentioned in the frameworks, like the motivation or intention of data production (MD_Identification.purpose), the maintenance and update policy (MD_Identification.resource-Maintenance), access and terms of use (more detailed than in most frameworks, MD_Identification.resource-Constraints), as well as the type of data and the data fields (MD_ContentInformation). It must be mentioned that the ISO metadata model can be applied at various levels of granularity. Consequently, metadata elements can also be provided for specific subsets or attributes to provide more details on the dataset content, as requested in some frameworks.

A marked distinction in the content becomes evident with regard to data concerning people and the ensuing challenges, limitations and potential biases, for instance,

with sensitive data, representativity or human attributes. Reflections about ethical implications and concerns are not incorporated in the ISO standard. This may be due to the prevailing perspective that geodata does not inherently pertain to the description of people. However, adopting a broad definition of geographic data as “data with implicit or explicit reference to a location relative to the Earth” (ISO, 2023, p. 3), geodata can also refer to people or people-related data can have a geographic reference; thereby aligning with the ISO standard. Therefore, this gap can be seen as relevant to the ISO standard, especially regarding ethical data practice for thematic cartography.

In addition to this discrepancy, the ISO standard provides a more limited range of options for the documentation of intended, unsuitable and actual uses of datasets. Only implemented uses can be contributed by dataset users (MD_Identification.resourceSpecificUsage).

Conversely, there are instances where information is not or less requested in the frameworks, yet it can be provided based on the ISO standard. Such information includes the data format, the language of the dataset, information about the metadata, related resources, the temporal resolution and the data quality. Furthermore, metadata elements specific to geographical data (spatial representation, spatial resolution, spatial extent, coordinate reference system and some data quality indicators) are unique to the ISO standard, as none of the frameworks have any geographical focus.

ISO 19157, which offers a model for the description of data quality of spatial data, defines five groups of data quality elements, namely completeness, logical consistency, positional accuracy, thematic quality and temporal quality, with subclasses appended to each. A number of these aspects are touched upon in four of the frameworks. However, they are not prominently featured there and are less differentiated.

A substantial congruence in the contents of the context frameworks and the ISO standard can be identified. The ISO standard incorporates most of the aspects referenced in many frameworks. Apparent gaps appear regarding ethical concerns, challenges and limits, and regarding data about humans. Beyond the frameworks, the ISO standard offers an extended definition of multiple geographic context elements and, in a highly differentiated manner, the aspects of data quality.

While a significant proportion of the identified data context elements can be documented within the ISO standard for metadata, this is rarely done in data publication. As previously stated, the context of data is frequently missing. A standard's existence and extensive scope do not guarantee the availability of the corresponding information (Quarati et al., 2021). Multiple studies have indicated that the quality of metadata in open data portals is often inadequate, with a significant proportion of metadata absent (Quarati et al., 2021; Ziaimatin et al., 2020). This lack of context information is especially problematic with regard to data quality and provenance, so the sources and production process of a

dataset (Fischer et al., 2023; Martino et al., 2019; Yang et al., 2013).

4. Use Case: Grätzlfarben

In order to test this hypothesis, a small use case was examined, which comprised three open datasets. The available metadata and the fulfilment of the requirements of the frameworks were compared. The selected use case is the web map ‘Grätzlfarben’ (Forschungsbereich Kartographie, TU Wien, 2024). This online application was developed by my research group with my participation. It displays the various land use types in Vienna, the capital of Austria, and enables users to print a postcard depicting the land use distribution in their respective neighbourhoods (Figure 1). The map is based on the open-source application ‘Kiezcolors’ by the Open Data Informationsstelle Berlin (ODIS, 2023).

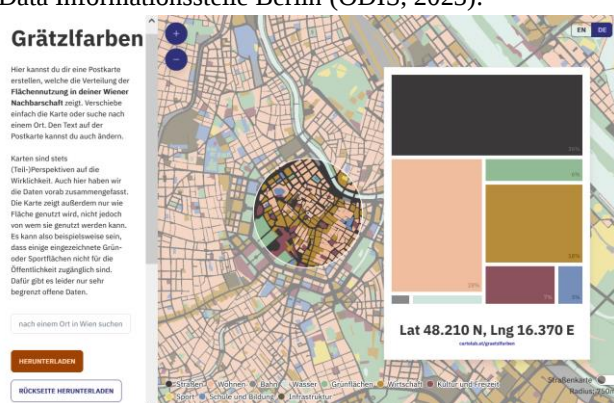


Figure 1. Screenshot of the web map “Grätzlfarben” (Forschungsbereich Kartographie, TU Wien, 2024)

The map requires area-wide polygon data for the various land use types in Vienna and, potentially, beyond. At present, the application utilizes Vienna’s open government dataset on land use mapping (Stadt Wien, 2024). However, the required information could also be extracted from alternative sources, such as the Austrian cadaster (Bundesamt für Eich- und Vermessungswesen, 2024b), which offers data on land use, or OpenStreetMap (OSM) (OpenStreetMap, 2024b). The ‘Grätzlfarben’ map was selected as it offers a practical and accessible example of a thematic map that can be used to examine various datasets regarding their fitness for use. However, it does not include any data about humans, which precludes testing an important aspect of the reflection and documentation frameworks. The examined datasets and references to the metadata and additional documentation that were reviewed are listed in Table 2.

Dataset	Metadata	Additional documentation
Landuse mapping	(Stadt Wien, 2024)	Description (Stadt Wien)
Cadaster	Various entries for different data formats, partly with varying metadata (Bundesamt für Eich- und Vermessungswesen, 2024c, 2024d, 2024e)	Product description (Bundesamt für Eich- und Vermessungswesen, 2024a) Interface descriptions (Bundesamt für Eich- und Vermessungswesen, 2023, 2024b)
OSM	Not available	OSM wiki (OpenStreetMap, 2024a)

Table 2. Examined datasets with references to metadata and additionally analyzed documentation

The OSM database, which is a volunteered geographic information database, does not offer any metadata. The heterogeneity of data contributors and data collection methods precludes the identification of many of the mentioned metadata elements to be identified for the database as a whole. The OSM wiki provides some general context information about the project, including motivation, financing, applications, access, license, data structure and data fields (tags). However, this meta information cannot be referred to individual data collections and features. Certain information, such as the date of the contribution, the most recent update, data contributor, or tools used, can partially be derived from features and changesets. Nonetheless, structural findings might be limited, and most context information aspects from the frameworks cannot be examined.

In the case of the land use mapping dataset by the city of Vienna and the cadaster dataset, several supporting documents were examined in addition to the metadata entries in the data portals. Basic content descriptions, licenses, and terms of use are available for both datasets. Information on updates and involved parties/contact points is limited or very general in both cases. The documentation examined did not contain any information regarding the motivation or purpose of the data collection. Furthermore, context information on the dataset's background and definitions and explanations for applied classifications for land use categories is limited. While a description of attributes and fields included in the datasets is available, the information in the interface description of the cadaster is more detailed than that for the land use mapping by the city of Vienna.

The spatial extent is provided in both dataset entries in the form of both text and geographic bounding boxes, thus allowing for both human and machine readability. The coordinate reference system is specified exclusively for the cadaster, contingent on the file format and the spatial extent. The land use mapping data portal entry does not provide a reference system. However, a web search shows that all open government datasets by the city of Vienna are published in EPSG 31256 (MGI / Austria, Gauß-Krüger East) (OGD Österreich, 2021).

The information provided on sources, collection, processing and validation is superficial for both datasets. The absence of detailed provenance information is notable despite its potential value, given the integration of diverse sources across both datasets. Data quality is addressed in both cases but not examined in detail. For the land use mapping dataset, furthermore, metadata on the date of publication, update, and creation are missing. Only the date of metadata updates is available, while the cadaster metadata offers more temporal information, allowing a better assessment of the currentness. The available information on the datasets' intended, actual and unsuitable use is also limited.

For background information, it is important to know that the cadaster, including its data on land use types, is regulated in ordinances (Bundesamt für Eich- und Vermessungswesen, 2014; Republik Österreich, 2010).

These documents could provide more information on the purpose, the data collection process and classifications. However, as this work focuses on the potential and status quo of metadata, this was not considered.

Concluding a fitness for use analysis for different datasets in the use case of 'Grätzlfarben' would require further analyses and tests for usability, in addition to the metadata and the context information. Consequently, at this point, no clear decision can be made regarding the most suitable dataset for this particular use case. However, the examination has revealed that the published metadata for both open government datasets is limited and covers only a few of the aspects mentioned in the frameworks and the ISO standards, with the cadaster offering more context information than the land use dataset metadata. Both datasets exhibit significant deficiencies, particularly regarding provenance, motivation, uses, and quality. While the use case cannot be regarded as representative, the results align with the findings from the extant literature. Furthermore, the comparison with OSM highlights the unique challenges of providing and extracting context information for crowd-sourced data.

5. Conclusions, discussion and outlook

Knowledge about a dataset's origin is vital for its interpretation, analysis, and utilization, also in cartography. The definition of context information can be initiated through a range of sources, including reflection and documentation frameworks, metadata standards, and guidelines. A comparative analysis of seven frameworks and the ISO metadata standard for geographical data reveals numerous similarities but also divergent focal points. Content-wise, the ISO standard encompasses most of the context aspects identified in the majority of the frameworks. However, ethical concerns, challenges and specific information for data about humans are not addressed. Conversely, the ISO standards ISO 19115:1-204 and 19157:1-2023 define multiple aspects that extend beyond the frameworks, particularly in terms of geographical aspects and data quality. The various formats can complement each other, especially regarding these differences. However, further research is required to determine the extent to which the examined context information elements are comprehensive for a reflective, critical and responsible data practice in cartography.

Despite the considerable differences in format and application between the frameworks (in the form of guidelines and templates) and the ISO standards (which define abstract models), there is considerable potential for the ISO metadata elements to serve as a foundation for a geodata-specific reflection- and documentation framework. With such a framework, data producers and data users in the GIS sphere could reflect on the context of the dataset they are producing or using by means of a series of standardized questions. The development and testing of such a guideline or checklist is a subject to be explored in future.

However, while the frameworks and the ISO standard offer directions on the types of context information, this does not necessarily mean that this information is available and published with open data. In practice, the metadata available is very limited, and in order to facilitate a deeper exploration of the data context by data users, it is necessary to identify the challenges and barriers to this limited provision, e.g. the effort and skills required. Ways to promote a more extensive sharing of context information along with the publication of any dataset have to be investigated. Adjustments to the technical workflow of providing metadata could be a starting point, for example, based on metadata editors, metadata profiles and the setup of geodata infrastructures. Such extensions' conceptual and technical needs and implementation are promising avenues for future research. Furthermore, there is a necessity to direct greater attention to the unique requirements and challenges posed by crowd-sourced data, like OSM, where alternative approaches to providing, extracting, and reflecting context information must be developed. While fostering the provision of context information through data providers is one approach, it would also be beneficial to explore opportunities for cartographers and data scientists working with datasets offering limited metadata and context information.

Whilst access to context information and a guideline for reflection are important for an ethical data practice, they represent merely the beginning of a cartographic workflow, ultimately resulting in a map. Consequently, it is important to consider how the results of contemplating the data context can influence and enhance the mapping workflow and the final product. The relationship between data reflection and cartographic decisions, such as generalization or classification, should be examined. Furthermore, opportunities to include information about the data context in the final product to increase transparency and trust should be examined.

Ultimately, it is imperative to consider not only the context of individual datasets but also to critically interrogate the structures of knowledge production as a whole (D'Ignazio & Klein, 2020, pp. 171–172): What kind of data is systematically absent due to specific power imbalances? Which datasets are not published due to specific interests? Whose knowledge is being suppressed? An ethical data practice in cartography goes beyond the thorough reflection of a specific dataset.

6. Acknowledgements

I want to thank my supervisor, Georg Gartner, and my colleague, Florian Ledermann, for their feedback. I acknowledge that some AI-supported tools were used for linguistic refinement in this paper (Grammarly and DeepL Write).

7. References

- Batcheller, J. K., Gittings, B. M., & Dunfey, R. I. (2009). A Method for Automating Geospatial Dataset Metadata. *Future Internet*, 1(1), 28–46. <https://doi.org/10.3390/fi1010028>
- Bundesamt für Eich- und Vermessungswesen (Ed.). (2014, April 1). *Information zur Neuordnung der BANU im Kataster*. https://www.bev.gv.at/dam/jcr:6040da95-68e0-472f-9112-173768422112/Information_zur_Neuordnung_der_BA_NU_im_Kataster.pdf
- Bundesamt für Eich- und Vermessungswesen (Ed.). (2023, July 26). *Digitale Katastralmappe und Grundstücksdaten Stichtagsdaten GPKG: Schnittstellenbeschreibung*. Version 1.0. https://www.bev.gv.at/dam/jcr:239729f2-83e1-400e-9e5f-9ff65cc02257/BEV_S_KA_Katastralmappe_Grundstuecksdaten_GPKG_V1.0.pdf
- Bundesamt für Eich- und Vermessungswesen. (2024a). *Kataster Stichtagsdaten: Informationen zum Produkt*. <https://www.bev.gv.at/Services/Produkte/Kataster-und-Verzeichnisse/Kataster-Stichtagsdaten.html>
- Bundesamt für Eich- und Vermessungswesen (Ed.). (2024b, December 4). *Katastralmappe SHP: Schnittstellenbeschreibung*. Version 2.9. https://www.bev.gv.at/dam/jcr:a6342749-e2c2-4525-9cee-3b7531474599/BEV_S_KA_Katastralmappe_SHP_V2.9.pdf
- Bundesamt für Eich- und Vermessungswesen (Ed.). (2024c, December 17). *Kataster Grafik Grundstücksverzeichnis GPKG Stichtag 01.10.2024: Datenkatalog BEV*. <https://data.bev.gv.at/geonetwork/srv/ger/catalog.search#/metadata/fae0e053-b22a-4bbd-a3d3-442fc632a816>
- Bundesamt für Eich- und Vermessungswesen (Ed.). (2024d, December 17). *Kataster Grafik INSPIRE Stichtag 01.10.2024: Datenkatalog BEV*. <https://data.bev.gv.at/geonetwork/srv/ger/catalog.search#/metadata/908ec148-91e9-4a36-b6c0-4ee52f6e5a50>
- Bundesamt für Eich- und Vermessungswesen. (2024e, December 17). *Kataster Grafik SHP Stichtag 01.10.2024: Datenkatalog BEV*. <https://data.bev.gv.at/geonetwork/srv/ger/catalog.search#/metadata/3d6a823f-08ba-41a8-b838-a841d5b82a59>
- Chmielinski, K. S., Newman, S., & Taylor, M. et al. (2020, December 11). The Dataset Nutrition Label (2nd Gen): Leveraging Context to Mitigate Harms in Artificial Intelligence. In NeurIPS (Chair), *Workshop on Dataset Curation and Security*. <http://arxiv.org/pdf/2201.03954>
- Civic Software Foundation (Ed.). *CIVIC Data Library of Context: Structured Context Demo*. <https://www.civicalibrary.org/context>
- Civic Software Foundation (Ed.). (2020). *CIVIC Structured Context Program*. <https://civicsoftwarefoundation.org/posts/structured-context-data-feminism>
- Comber, A., Fisher, P., & Wadsworth, R. (2007). User-focused metadata for spatial data, geographical information and data quality assessments. In M. Wachowicz & L. Bodum (Chairs), *AGILE*, Aalborg University.
- Data Nutrition Project. (n.d.). *The Dataset Nutrition Label*. Retrieved October 3, 2024, from <https://labelmaker.datanutrition.org/>
- Devillers, R., Bédard, Y., & Gervais, M. et al. (2007). How to Improve Geospatial Data Usability: From Metadata to Quality-Aware GIS Community. In *AGILE (Chair), AGILE Pre-Conference Workshop*.
- D'Ignazio, C. (2022). Creative data literacy. *Information Design Journal*, 23(1), 6–18. <https://doi.org/10.1075/idj.23.1.03dig>
- D'Ignazio, C., & Klein, L. F. (2020). *Data Feminism*. The MIT Press. <https://doi.org/10.7551/mitpress/11805.001.0001>
- Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata, 2008.
- Fischer, J., Egli, L., & Groth, J. et al. (2023). Approaches and tools for user-driven provenance and data quality information in spatial data infrastructures. *International Journal of Digital Earth*, 16(1), 1510–1529. <https://doi.org/10.1080/17538947.2023.2198778>
- Forschungsbereich Kartographie, TU Wien. (2024). *Grätzfarben*. <https://cartolab.at/graetzfarben/>
- Gebri, T., Morgenstern, J., & Vecchione, B., et al. (2018, March 24). *Datasheets for Datasets*. <http://arxiv.org/pdf/1803.09010>
- Gebri, T., Morgenstern, J., & Vecchione, B., et al. (n.d.). *Datasheets for Datasets Template*. Retrieved October 9, 2024, from <https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE4t8QB>
- Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, 14(3), 575. <https://doi.org/10.2307/3178066>
- Henzen, C., Mäs, S., & Bernard, L. (2013). Provenance Information in Geodata Infrastructures. In D. Vandenbroucke, B. Bucher, & J. Cromptvoets (Eds.), *Geographic Information Science at the Heart of Europe* (pp. 133–151). Springer International Publishing. https://doi.org/10.1007/978-3-319-00615-4_8
- Holland, S., Hosny, A., & Newman, S. et al. (2020). The Dataset Nutrition Label: A Framework to Drive Higher Data Quality Standards. In D. Hallinan, R. Leenes, S. Gutwirth, & P. de Hert (Eds.), *Data Protection and Privacy*. Hart Publishing. <https://doi.org/10.5040/9781509932771.ch-001>
- International Organization for Standardization (Ed.). (2014). *ISO 19115-1:2014: Geographic information — Metadata. Part 1: Fundamentals*. <https://www.iso.org/standard/53798.html>

- International Organization for Standardization (Ed.). (2023). *ISO 19157-1:2023: Geographic information — Data quality. Part 1: General requirements*. <https://www.iso.org/standard/53798.html>
- Kalantari, M., Syahrudin, S., & Rajabifard, A. et al. (2021). A Proposal for a User-Oriented Spatial Metadata Profile. *ISPRS International Journal of Geo-Information*, 10(6). <https://doi.org/10.3390/ijgi10060376>
- Kalantari, M., Syahrudin, S., & Rajabifard, A. et al. (2020). Spatial Metadata Usability Evaluation. *ISPRS International Journal of Geo-Information*, 9(7). <https://doi.org/10.3390/ijgi9070463>
- Krause, H. (2017). *Data Biographies: Getting to Know Your Data*. <https://gijn.org/stories/data-biographies-getting-to-know-your-data/>
- Krause, H. (2023). *An Introduction to the Data Biography*. <https://weallcount.com/2019/01/21/an-introduction-to-the-data-biography/>
- Martino, M. de, Rosim, S., & Quarati, A. (2019). Hydrographic Datasets in Open Government Data Portals: Mitigation of Reusability Issues Through Provenance Documentation. In E. Garoufallou, F. Fallucchi, & E. W. de Luca (Eds.), *Metadata and Semantic Research: 13th International Conference, MTSR* (pp. 307–319). Springer. https://doi.org/10.1007/978-3-030-36599-8_27
- Microsoft Research. (2022, August 25). *Data Documentation - Microsoft Research*. <https://www.microsoft.com/en-us/research/project/datasheets-for-datasets/>
- Microsoft's Aether Transparency Working Group (Ed.). (2022). *Aether Data Documentation Template (Draft 08/25/2022)*. <https://www.microsoft.com/en-us/research/uploads/prod/2022/07/aether-datadoc-082522.pdf>
- ODIS (Ed.). (2023). *Kiezcolors*. <https://kiezcolors.odis-berlin.de/>
- OGD Österreich (Ed.). (2021). *Metadaten data.gv.at*. <https://go.gv.at/ogdmetade>
- Open Geospatial Consortium (Ed.). (2024, January 22). *OGC e-Learning - Metadata - Introduction*. <http://opengeospatial.github.io/e-learning/metadata/text/main.html>
- OpenStreetMap (Ed.). (2024a, September 6). *OpenStreetMap Wiki*. <https://wiki.openstreetmap.org/>
- OpenStreetMap. (2024b, December 19). *OpenStreetMap*. <https://www.openstreetmap.org/>
- Oxford Learner's Dictionary. (n.d.). *context*. Retrieved January 13, 2025, from <https://www.oxfordlearnersdictionaries.com/definition/english/context?q=context>
- Pushkarna, M., Zaldivar, A., & Kjartansson, O. (2022). Data Cards: Purposeful and Transparent Dataset Documentation for Responsible AI. In ACM (Chair), *FACt '22: 2022 ACM Conference on Fairness, Accountability, and Transparency*, Seoul Republic of Korea.
- Pushkarna, M., Zaldivar, A., & Kjartansson, O. (2023, February 22). *Data Cards Playbook Templates*. <https://github.com/PAIR-code/datacardsplaybook/tree/main/templates>
- Quarati, A., Martino, M. de, & Rosim, S. (2021). Geospatial Open Data Usage and Metadata Quality. *ISPRS International Journal of Geo-Information*, 10(1). <https://doi.org/10.3390/ijgi10010030>
- Republik Österreich. (2010). *Verordnung des Bundesministers für Wirtschaft, Familie und Jugend über die Angabe und Definition der Benützungsarten und Nutzungen im Grenzkataster: Benützungsarten-Nutzungen-Verordnung – BANU – V, BGBl. II Nr. 116/2010*. https://www.ris.bka.gv.at/Dokumente/BgblAuth/BGBLA_A_2010_II_116/BGBLA_2010_II_116.html
- Robinson, A. H., Morrison, J. L., & Muehrcke, P. C. et al. (1995). *Elements of cartography* (6th ed.). Wiley.
- Stadt Wien (Ed.). *Realnutzungskartierung - Flächennutzung im Stadtgebiet*. <https://www.wien.gv.at/stadtentwicklung/grundlagen/stadtforschung/siedlungsentwicklung/realnutzungskartierung/>
- Stadt Wien (Ed.). (2024). *Realnutzungskartierung ab 2007 Wien - data.gv.at*. https://www.data.gv.at/katalog/de/dataset/stadt-wien_realnutzungskartierungab200708wien
- Wagner, M., Henzen, C., & Müller-Pfefferkorn, R. (2021). A Research Data Infrastructure Component for the Automated Metadata and Data Quality Extraction to Foster the Provision of FAIR Data in Earth System Sciences. *AGILE: GIScience Series*, 2, 1–7. <https://doi.org/10.5194/agile-giss-2-41-2021>
- We All Count (Ed.). (2023). *Data Biography 2023 Template Detailed*. https://docs.google.com/spreadsheets/d/1z5qlMasRLBaagd35nua5hVqH5VQZPTt2U_CQkspwlfM/edit?gid=0#gid=0
- Western Pennsylvania Regional Data Center. (n.d.). *Data User Guides*. Retrieved December 16, 2024, from <https://www.wprdc.org/en/data-user-guides>
- Western Pennsylvania Regional Data Center (Ed.). (2015). *Data-User-Guides: GitHub*. <https://github.com/rgradeck/Data-User-Guides/>
- Whitfield, P. H. (2012). Why the Provenance of Data Matters: Assessing Fitness for Purpose for Environmental Data. *Canadian Water Resources Journal / Revue Canadienne Des Ressources Hydriques*, 37(1), 23–36. <https://doi.org/10.4296/cwrj3701866>
- Yang, X., Blower, J. D., & Bastin, L. et al. (2013). An integrated view of data quality in Earth observation. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, 371. <https://doi.org/10.1098/rsta.2012.0072>
- Zaimatin, H., Nili, A., & Barros, A. (2020). Reducing Consumer Uncertainty: Towards an Ontology for Geospatial User-Centric Metadata. *ISPRS International Journal of Geo-Information*, 9(8). <https://doi.org/10.3390/ijgi9080488>