Measuring Data Quality of Cadastral Data

Leikny GAMMELMO, Norway

Key words: Cadastre, Digital cadastre, Land management, Standards

SUMMARY

Data quality is a topic with many dimensions. One of these can be quality, for instance, seen from the perspective of one single user or for digitisation of a process. Or it might be the need for accurate data or data completeness. Other dimensions could be the need for available and relevant data. The cadastre is one source of data that can be of great importance in land management and development. The issue we want to address is that the cadastre is not of the calibre to fulfil its purpose as part of the land registration system in a digital world and play a key role in land management.

The Norwegian cadastre is a mixed dataset, collected over hundreds of years and under various legislation. Over time, the need has arisen for new data fields in the data matrix. This has led to varying quality among cadastral units and the buildings belonging to them. How should this be dealt with? One way is to accept that total control of data is impossible. It has been said that, when working with data, total control is impossible due to the extreme complexity of distribution and the pace of change. We must accept "data disorder", which means, among other things, variability of data quality. An analysis in Norway from 2019 concluded that collecting all the cadastral data that is missing, or "not good enough" given a specific standard, would take 20–50 years at the best estimate, and we cannot be sure that the quality we define today will be suitable for tomorrow's demands. We need a strategy for data quality, and it must involve all parties who use and providing cadastral data.

The aim of this paper is to answer and shed light on three questions:

- Why do we need to measure the data quality of cadastral data?
- What is a useful definition of data quality regarding cadastral data?
- How should the data quality of cadastral data be measured?

The methods used are literature study, workshops and hypothesis testing. Involving users and professionals has been important, connecting with them through workshops and involving them in the "Masterplan Cadastre" project.

Measuring Data Quality of Cadastral Data

Leikny GAMMELMO, Norway

1. INTRODUCTION

In our modern world, making processes digital is often a goal. Some of these processes need data from the cadastre. One example we can cite is the importance of the cadastre in the digital transformation of public administration – e-Government or Digital Government. In Norway, the cadastre is a mixed dataset, collected over hundreds of years and under various legislation. As a result, data has been used in this process which may or may not be in need of better quality. Since the quality differs, we need to improve it, and therefore we need to measure data quality. This leads to the answer to the first question in this paper: why do we need to measure the data quality of cadastral data? It is because we need to have some indications that we are heading in the right direction – towards better quality. It is also necessary to comment on the data in relation to digital processes: is the cadastre fit for purpose if that purpose is to play a key role in a digital planning process, the process of property formation or a building process?

The aim of this paper is to answer and shed light on the topic of measuring data quality and two other questions:

- What is a useful definition of data quality regarding cadastral data?
- How should the data quality of cadastral data be measured?

The research methodology is illustrated in Figure 1. It is a combination of literature search, hypothesis testing, and involving different people in workshops and discussions. Involving users and professionals has been important. It is all part of the "Masterplan Cadastre" project. Last autumn and winter, the project worked on a strategy for improving the quality of cadastral data. This work included seeking to elaborate a definition of data quality and the measurement parameters we need in our work of improving the quality of cadastral data.

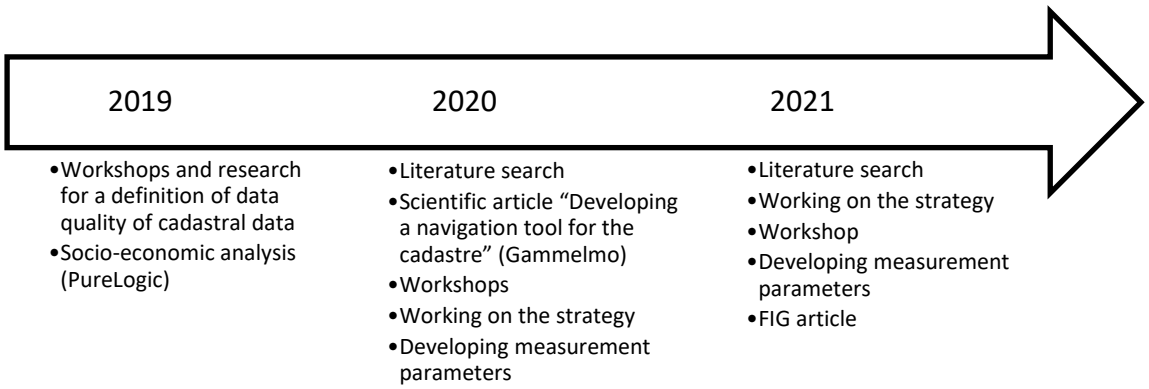

Figure 1: Research methodology

2. THE CADASTRE AND CADASTRAL DATA IN NORWAY

The organization and responsibility for the cadastre in Norway is divided. The Norwegian Mapping Authority is the central cadastral authority, and the municipalities are the local cadastral authorities. The local cadastral authorities perform the cadastral registration and are responsible for carrying out cadastral surveying (Cadastre Act, Section 5a). By agreement, the municipality may permit other parties to perform cadastral surveying on its behalf. Pursuant to the provisions of Chapter 5 of the Local Government Act, the municipality may by agreement permit another municipality or the central cadastral authority to maintain the cadastre on its behalf. The State, represented by the central cadastral authority, the Norwegian Mapping Authority, must organize, operate and administer the cadastral system, in accordance with Section 5a.

The register is open and available for everyone to use: Section 29 of the Cadastre Act states that everyone must have access to information about the cadastre and the data from the cadastre. The central cadastral authority and the municipalities must ensure that information about the cadastre is available on request.

The cadastre is a nationwide register that defines the existing location and dimensions of land parcels. The register should be a resource for both planning and administration, for example for procedures concerning real property and building applications. It can also provide important information during transaction processes and for mortgage applications.

Cadastral data does not necessarily mean the same thing to all parties. The simplest definition is to say that cadastral data is the data we find in the specific cadastre. ESRI states that "[c]adastral data defines the geographic extent of the past, current, and future rights and interests in real property and the spatial information needed to describe that geographic extent. These rights and interests are contained in land record documents." According to Williamson et al. (2010), cadastral data is "cadastral or legal parcels, properties, parcel identifiers, buildings, legal roads, etc".

In the Norwegian cadastre there are 3,357,773 cadastral units, 2,444,394 addresses (connected to a named road or area) and 4,305,568 buildings (data reported 9 February 2021). For each of these we can find a set of data that gives more information – for example, the accuracy of the boundaries of a cadastral unit, different areas and rooms in a building, and the coordinates of an address.

3. DATA QUALITY

In 2019, a socio-economic analysis was prepared which is part of the work to assess how we should proceed to improve the quality of the content of the cadastre. The analysis concluded that there was no unified definition or clear concept of quality for cadastral data in Norway. It was pointed out that the quality of the cadastre has different significance for different users.

As an example, it will be perceived as poor if the user finds that the relevant data field is incompletely filled in, has low accuracy, or the accuracy is incorrectly stated (PureLogic 2019:12). Therefore, we need to define what we mean by data quality.

In the search for a definition, we started with ISO/IEC 25012 and the six key dimensions recommended to be used when assessing or describing data quality from DAMA (2013). We then considered an article about dimensions of data quality concerning big data and using some of the same dimensions, Ramasamy and Chowdhury (2020). Cadastral data is not big data, but that does not mean this literature review is not useful. Another source was a specification from the Norwegian Digitalisation Agency (Difi 2017) and the Norwegian Mapping Authority (Kartverket 2015). It was also relevant to look at ISO 19157.

In a workshop with colleagues at the Norwegian Mapping Authority in 2020, further relevant literature was mentioned, and a risk was pointed out: the risk of lack of trust in the cadastre due to poor data quality. A relevant question was: what will happen if we do not manage to have a useful cadastre? What if we cannot provide the data needed and other registers are developed without a unified model?

Quality is “the degree to which a set of inherent characteristics fulfils requirements” (ISO 19157:2013). Data quality dimensions can be defined as “some thing (data item, record, dataset or database) that can either be measured or assessed in order to understand the quality of data” (DAMA 2013:3). According to Plotkin (2014) “[d]ata quality dimensions are ways to categorize types of data quality measurements. In general, different dimensions require different measurements. For example, measuring completeness requires analysing whether there is a value (any value) in a field, whereas measuring validity requires comparing existing format, patterns, data types, ranges, values, and more to the defined set of what is allowable.”

Table 1 provides some possible dimensions of data quality and some sources mentioning them. Some of them can be said to be different terms and slightly different definitions of the same concept. The quality dimensions can further be categorized into sub-dimensions, allowing more specialized quality dimensions with more specialized quality measures. For some dimensions it will be possible to use measurement parameters directly from the data in the cadastre, and for others will not. The dimensions we choose are those that can assist us, as “data quality practitioners” in accordance to DAMA (2013:3), when looking to assess and describe the quality of the data in our register.

Table 1: Different data quality dimensions and reference to sources.

Dimension	Source
Accuracy	ISO/IEC 25012, DAMA 2013, Difi 2017 and Pure Logic 2019
Completeness	ISO/IEC 25012, DAMA 2013, Difi 2017 and Pure Logic 2019
Consistency	ISO/IEC 25012, DAMA 2013 and Pure Logic 2019
Credibility	ISO/IEC 25012
Currentness	ISO/IEC 25012, Difi 2017 and Pure Logic 2019
Accessibility	ISO/IEC 25012
Compliance	ISO/IEC 25012 and Difi 2017
Confidentiality	ISO/IEC 25012
Efficiency	ISO/IEC 25012
Precision	ISO/IEC 25012
Traceability	ISO/IEC 25012
Understandability	ISO/IEC 25012
Availability	ISO/IEC 25012 and Difi 2017
Portability	ISO/IEC 25012
Recoverability	ISO/IEC 25012
Relevancy	Difi 2017
Uniqueness	DAMA 2013
Timeliness	DAMA 2013
Validity	DAMA 2013

Traceability is the capability to trace something and can be used as a means of identifying where the data comes from. It is probably a dimension on a higher level defining the source, like the Norwegian Mapping Authority. For cadastral data, this can be a sub-dimension and source to give information about how the data was collected – if the data was collected during a land survey, when handling a building application, or if the right holder submitted it.

Relevancy helps us to answer the question: does the dataset include an appropriate amount of data? Credibility, currentness and timeliness are dimensions concerning the user experience, and whether the data is fit for purpose. It can be measured, for example, by user feedback.

For obtaining relevant measurement parameters that can give us an indication of the quality of cadastral data, two dimensions are seen as especially relevant: accuracy and completeness. Compliance can also be part of measuring accuracy, because different data fields in the cadastre can be put together. Section 4 gives some examples of measurement parameters.

Accuracy is the degree to which data correctly describes the “real world” object. The scope is any “real world” object or objects that may be characterized or described by data, held as a data item, record, dataset or database. The unit of measurement is the percentage of data entries that pass the data accuracy rules (DAMA 2013). For cadastral data it will also be of interest to measure the number/quantity. Accuracy can also be said to be the “closeness of agreement between a test result or measurement result and the true value” (ISO 19157:2013).

Completeness is the proportion of stored data against the potential of "100% complete", a measure of the absence of blank (null or empty string) values or the presence of non-blank values. The range is 0–100% of critical data to be measured in any data item, record, dataset or database. The unit of measurement is the percentage (DAMA 2013). For cadastral data it will also be of interest to measure the number/quantity.

4. MEASURING THE DATA QUALITY OF CADASTRAL DATA

As seen in Section 3, more than one relevant data quality dimension can be part of a useful definition of data quality regarding cadastral data. But when it comes to actually measuring the quality and having the opportunity to see whether the project of improving the quality is moving in the right direction, there are two relevant dimensions: accuracy and completeness. The next step is to determine how the data quality of cadastral data should be measured. In this stage of the process, we need measurement parameters.

Table 2: Data quality dimensions, measurement parameters and the status in Norway.

Dimension	Measurement parameters	Status
Accuracy	Units with cadastral number as the only address	Nov. 2020: 79,976 Feb. 2021: 70,322
	Cadastral units with one or more fictitious boundaries	Mar. 2020: 117,427 Feb. 2021: 116,981
Completeness	Area lacking cadastral number	Nov. 2020: About 17%
	Cadastral units without a parcel in the cadastre map	Nov. 2020: 82,367 Feb. 2021: 82,005

Positional accuracy of boundary marks is a relevant measurement parameter of accuracy. In March 2020 there were over 20 million boundary marks registered in the cadastre. They have an accuracy with values ranging from 0 cm to 50 million cm. Just over 4.5 million of all boundary marks in the cadastre have an accuracy between 0 and 10 cm. An accuracy equal to 13 cm applies to approximately 3.5 million boundary marks. Just over 1.5 million boundary marks have an accuracy poorer than 200 cm. These data field can be used alone or together with data fields such as Year, because we have had different standards through the years.

For buildings, we register three different types of area; not all data fields are required for all buildings, and the data fields have come into use at different times; therefore, the parameter is only a starting point for now. The three different types of area are:

- BYA (bebygd areal) = built-up area
- BRA (bruksareal) = usable area
- BTA (brutto areal) = gross area.

The data field BYA was introduced in 1991. BRA was part of the register we had from about 1983 called the register for property, addresses and buildings (GAB). This register, together with the digital property map (DEK), became the Norwegian Cadastre in 2010 when the Cadastre Act came into force. The Cadastre Act introduced BTA, but the data field in the register was first implemented in 2011.

The numbers in Table 3 give an overview of the status from February 2021. There were 4.3 million existing buildings in total, and 4.9 million if we also include building alterations. The table shows the proportion of buildings without the different areas in total and for those registered from 2010 until today. The quality measurement parameters give us the status of completeness. The next step is to measure the accuracy of the data. If the data field is given the numeral 1, is this the real area of the building or is it just a way to avoid to showing a low level of completeness? Table 4 gives some examples of how this can appear in the cadastre.

Table 3: Data quality measurement parameters for buildings registered in the cadastre. The percentage shows the share without the data fields (status in February 2021).

Area data field	Percentage of all buildings and building alterations	Buildings and building alterations from 2010-2021
BYA	78%	22%
BRA	26%	10%
BTA	96%	85%

Table 4: Examples of how the data fields are registered in the Cadastre.

Type of building	BYA	BRA	BTA
Detached house with apartment	4	217	0
Detached house	126.6	241.5	254.8
Semi-detached house	78.3	65.7	72
Garage or outbuilding	2	1	0

5. CONCLUDING REMARKS

Through the Masterplan Cadastre project, the Norwegian Mapping Authority is working to complete a strategy to improve the data quality of cadastral data. The strategy is divided into 1) the vision and the goals for higher data quality and 2) the process we need to arrive there, together with an action plan. The ambition is for the cadastre to have the right data content and data quality necessary for processes that use cadastral data and for user needs. It must also be capable of supporting digital processes. Based on an analysis of the current situation and identified needs, three goals have been set for further quality work:

1. A common practice for cadastral registration in all municipalities.
2. Improving the quality of priority areas or data fields.
3. Streamlining the processes for collecting and updating cadastral data.

The process for attaining these goals needs to be agile and capable of acting “whenever a strategic opportunity offers itself”, instead of following a fixed long-term strategy. At the same time, the strategy, needs to provide clear priorities. Here we can use the dimension of relevancy. The Planning and Building Act and the Cadastre Act give the cadastre relevancy in land administration and management, but there can also be other processes and user needs which the cadastre should fulfil.

The Norwegian Mapping Authority will provide data and guidance, but most of the work will have to be done in the municipalities, due to their role as local cadastral authorities. They register the data in the cadastre, in most cases. The strategy will therefore provide an action plan with possible measures that can be implemented to improve the data quality of the cadastre. It is based on the work in the Masterplan Cadastre and proposals and requests received in an early consultation on the strategy document. In 2021 we will start some of these measures, and our goal is to have an action plan ready by the end of the year. When measures are to be prioritized, it is recommended that the improvement in quality will increase the use of the cadastre in the digital transformation of public administration. This can be the use of the cadastre in digital building processes, when calculating property tax or when reporting data to KOSTRA (Key figures on municipal activities by Statistics Norway).

The aim of the paper was to answer and shed light on three questions:

- Why do we need to measure the data quality of cadastral data?
- What is a useful definition of data quality regarding cadastral data?
- How should the data quality of cadastral data measured?

As stated above, we need a uniform definition of data quality suitable for our purpose. The foremost reason is to know if we are heading in the right direction, to see whether the measures we initiate actually improve the quality of cadastral data. Then we need to find a way of determining how to measure the data quality and which measurement parameters to use. This research is a step towards that goal.

The examples in Section 4 provide useful quality dimensions and measurement parameters for monitoring the development of the quality of cadastral data. They also provide documentation and a basis for further discussion. For example, when it comes to data fields for buildings, this leads to a discussion as to whether we should have three different types of areas or not. Other relevant questions are: who needs the data fields, and whether it would be more advantageous to obtain the data from other data sources. The Norwegian cadastre model is old, and the discussion of a new model will benefit from the results of measuring the data quality of the register today.

ACKNOWLEDGEMENTS

I would like to thank my colleagues, especially Morten Borrebæk, Beate Aase and Jan Häusler, for our discussions about data quality and for giving me valuable feedback on this paper. I am also grateful to the Masterplan Cadastre project and the "mythbusters" for providing valuable data.

REFERENCES

DAMA 2013 *The Six Primary Dimensions for Data Quality Assessment.* Produced by the DAMA UK Working Group on "Data Quality Dimensions". Accessible from https://silo.tips/download/the-six-primary-dimensions-for-data-quality-assessment

Difi 2017 *Spesifikasjon for beskrivelse av kvalitet på datasett. Versjon 1.0* Norwegian Digitalisation Agency. Accessible from: https://data.norge.no/specification/spesifikasjon-for-beskrivelse-av-kvalitet-pa-datasett/ [24 July 2020]

ESRI Web page article accessible from: https://www.esri.com/news/arcuser/1000/umbrella12.html [18 December 2020]

ISO 19157:2013 Geographic information – Data quality.

ISO/IEC 25012 Data Quality model. Accessible from: https://iso25000.com/index.php/en/iso-25000-standards/iso-25012?limit=5&limitstart=0 [23 July 2020]

Kartverket 2015. *Standarder Geografisk Informasjon. Geodatakvalitet, versjon 1.0 – 2015.* The Norwegian Mapping Authority. Accessible from: https://www.kartverket.no/globalassets/geodataarbeid/standardisering/standarder/standarder-geografisk-informasjon/geodatakvalitet-1.0-standarder-geografisk-informasjon.pdf [11 February 2021]

Plotkin, D. 2014 *Important Roles of Data Stewards*. Accessible from: https://doi.org/10.1016/B978-0-12-410389-4.00007-6

Ramasamy, A. and S. Chowdhury 2020. Big Data Quality Dimensions: A Systematic Literature Review. Accessible from: https://www.scielo.br/scielo.php?script=sci_arttext&pid=S1807-17752020000100303 [18 December 2020]

Williamson, I, S. Enemark, J. Wallace and A. Rajabifard 2010. *Land Administration for sustainable development*. FIG Congress 2010

BIOGRAPHICAL NOTES

Leikny Gammelmo holds a master's degree and a Ph.D. in land management and law from the Norwegian University of Life Sciences. She is the Chief Engineer working with the Cadastre and as Project Manager of "Masterplan Cadastre" at the Norwegian Mapping Authority.

CONTACTS

Chief Engineer and Project Manager, Leikny Gammelmo Ph.D.
The Norwegian Mapping Authority

NORWAY
Telephone: +47 32 11 80 00
Mobile: +47 976 63 668
E-mail: leikny.gammelmo@kartverket.no

www.kartverket.no