AGILE: GIScience Series, 3, 59, 2022. https://doi.org/10.5194/agile-giss-3-59-2022
Proceedings of the 25th AGILE Conference on Geographic Information Science, 2022.
Editors: E. Parseliunas, A. Mansourian, P. Partsinevelos, and J. Suziedelyte-Visockiene.
This contribution underwent peer review based on a full paper submission.
© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.

Modelling Inhomogeneous Geodata Quality in a Dataset's Metadata

Arne Rümmler [1], Christin Henzen [1], and Heiko Figgemeier [1]

[1]Chair of Geoinformatics, Technical University Dresden, Dresden, Germany

Correspondence: Arne Rümmler (arne.ruemmler@tu-dresden.de)

Abstract. Extensive data quality descriptions as a vital part of a dataset's metadata are widely accepted, albeit their provision in a formalized manner is often lacking. This is due to a number of problems that are frequently encountered by geodata producing scientists. As one of these problems, we identified missing, unknown or unused options to model inhomogeneity of data quality across space, time, and theme in a dataset's metadata. Detailed information of inhomogeneous geodata quality beyond dataset-wide statistical measures (variance, min, max, etc.) is often only described in dataset accompanying papers or quality reports. These text-based approaches prevent precise querying and hinder the development of advanced data discovery tools that could make valuable use of inhomogeneous data quality information. We propose a profile for the data quality vocabulary (DQV)[1] that allows to model inhomogeneous geodata quality. Considering established vocabularies typically used to describe geographic metadata, as well as ensuring compatibility with the default version of DQV, enhances the usability and thus, minimizes the effort for data producers to provide formalized descriptions of inhomogeneous data quality.

Keywords. geodata quality, metadata, linked data, reusability, DQV

1 Introduction

With the increasing usage of the **F**indable **A**ccessible **I**nteroperable **R**eusable principles (Wilkinson et al., 2016) in scientific cases, data producers become more and more aware of providing reusable data. Followingly, a rich description of the data quality becomes imperative to prevent misuse and to guide the process of data selection. Up to now, data quality is often described only in the dataset's accompanying papers or quality reports (Ariza-López et al., 2020). A reason for this are insufficient, unknown and/or unused structures to describe inhomogeneous geodata quality in a formalized, data-producer-friendly manner.

In geospatial data modelling, inhomogeneity of data quality is often caused by inhomogeneous input data, e.g., satellite imagery, governmental statistical data or in-situ measurements. Running a model with inhomogeneous inputs, results in inhomogeneous data quality of the model's outputs. For instance, for the dataset Global Forest Change (Hansen et al., 2013), data quality information is provided in the supplementary material in the form of text and tables; where it is disaggregated into different spatial, temporal and thematic scopes, e.g., error of forest loss from 2000 to 2012 in the tropical climate domain (p. 17).

If inhomogeneity of data quality is described in a non-textual manner, typically one of the following strategies is applied: 1) producing separate data quality datasets accompanying the original dataset (e.g., a raster of Root-Mean-Square-Errors), or 2) splitting the original dataset into smaller sub-datasets, so that the data quality is homogeneous for each resulting dataset. However, both are not suitable as a generic strategy: in the first option, the quality information is not part of the metadata and, e.g., cannot be used for discovery or evaluation. The second option is not feasible for every dataset, because measuring different quality indicators might require different splits, which results in spatially overlapping sub-dataset. With our approach, we contribute to the reduction of the required effort in reporting inhomogeneous geodata quality in a formalized way.

2 Requirements for a Metadata Profile to Model Inhomogeneity of Data Quality

This section frames requirements for a metadata profile that facilitates the modelling of inhomogeneous geodata quality. Our proposed concept serves as a means to enrich measurement information for any geodata quality indicator with a spatial and/or temporal extent, as well as with a thematic scope that constrains its validity. This allows to

[1]https://www.w3.org/TR/vocab-dqv/, April 20, 2022

measure a certain quality indicator several times, with different spatial, temporal, and/or thematic constraints, and to describe the results in one metadata set. The constraints have to be chosen so that the resulting geodata quality measurement is homogeneous, i.e., the constraints can be solely spatial, temporal or thematic, or a combination.

Since reuse of geodata happens across domains, interoperability of the constraints' descriptions is imperative. Moreover, as data producers assess the sufficiency of a measurement's homogeneity, the structure of the constraints' descriptions should be highly flexible, e.g., by permitting and encouraging the use of domain-specific terms that are properly defined. Consequently, requirement R1 addresses semantic aspects.

Lacking options to describe inhomogeneous data quality formally can lead to mixed forms of geodata quality descriptions, i.e., homogeneous data quality information is part of the metadata, while inhomogeneous aspects are covered somewhere else. Requirement R2 is split into two parts, which are both concerned with the completion of existing descriptions. The first part (R2.1) ensures that constrained geodata quality descriptions can be added to existing unconstrained descriptions. In addition, there are cases where existing data quality measurements are unconstrained, even if they are not valid dataset-wide. This can result from missing options to describe inhomogeneous data quality. Part 2 (R2.2) ensures that such errors are correctable.

R1 The constraints that are used to delimit a geodata quality measurement's validity should either be described by 1) proper terms[2] that are available in open ontologies/vocabularies[3] or by 2) defining them in established open formats.

R2.1 A measurement of a certain geodata quality indicator does not have to be constrained. A dataset's metadata can include constrained and unconstrained measurements of geodata quality measurements.

R2.2 Existing (unconstrained) measurements of geodata quality in a dataset's metadata can be enriched with (a) constraint(s).

The concept should be provided as profile of an existing data quality framework, which results in requirements on the framework (encoded using F.R). The framework has to be suitable to address quality indicators for spatial data (F.R1). Retroactive integration (see R2) requires that the profiled version of the framework is compatible with its default version (F.R2). Requirement R1 encourages using open ontologies/vocabularies, rather than defining the spatial, temporal and thematic constraints for each measurement of geodata quality. This can be facilitated by using technologies from the Semantic Web Stack (F.R3). Moreover, the usability of the framework is crucial. As indicated by Egli et al. (2021) and Ariza-López et al. (2020), a complex data quality model hinders data producers in (proper) application (F.R4).

F.R1 The used framework has to provide capabilities to implement any quality indicator relevant for spatial data.

F.R2 The profile has to be compatible to the default version of the framework.

F.R3 The framework should use technologies of the Semantic Web Stack, whenever possible.

F.R4 The framework's structure should be as simple as possible.

3 Review of Existing Quality Frameworks to Model Data Quality

In this section, we review quality frameworks against the developed requirements and provide a detailed review for the two most promising candidates – ISO 19157[4] and DQV. In the context of geodata quality, Barsi et al. (2019) present a data quality framework for remote sensing data that focuses on error propagation throughout the data lifecycle. Another framework in this domain is introduced with the Quality Assurance Framework for Earth Observation (QA4EO) (Hunt et al., 2021). Moreover, Senaratne et al. (2017) review literature to assemble quality indicators and according measurement methods to assess the quality of volunteered geographic information. Focusing on semantic aspects, the daQ Ontology facilitates describing the quality of linked datasets (Debattista et al., 2014), which can be considered influential in the development of DQV (Albertoni and Isaac (2021), Debattista et al. (2016)).

The international standard **ISO 19157** describes a framework to measure the quality of geographic data and is designed as a complement to ISO 19115-1[5], which defines general geographic metadata. Moreover, it comprises well-defined descriptions of quality indicators for geographical data that are referred to as data quality elements (DQ_Element). These data quality elements are grouped into five distinct categories of geographic data quality and an additional one to define user-specific data quality elements and thus, fulfill requirement F.R1 (F.R1 ✓). Furthermore, ISO 19157 comprises a register of evaluation methods for the different quality elements. However, ISO

[2]Proper denotes a complete description, i.e., a spatial term has to include a spatial extent, a temporal term has to include a temporal extent and thematic terms have to be defined comprehensively.

[3]In this publication, we use 'ontology/vocabulary' as a term that comprises vocabularies, taxonomies, gazetteers, thesauri, ontologies, and any other concept that is used to describe and/or link terms.

[4]https://www.iso.org/standard/32575.html, April 20, 2022

[5]https://www.iso.org/standard/53798.html, April 20, 2022

19157 is not based on concepts of the Semantic Web Stack (F.R3 ✗).

ISO 19157 already defines a concept to model inhomogeneous geodata quality by combining a quality element with a scope (ISO19115-1:MD_Scope). This combination is referred to as a data quality unit. A scope can be constrained spatiotemporally by an extent (ISO19115-1:EX_Extent), and thematically by a scope code (ISO19115-1:MD_ScopeCode) or a scope description (ISO19115-1:MD_ScopeDescription).

The spatial extent (ISO19115-5:EX_GeographicalExtent) can be described by using established formats (see ISO19115-1:GM_Objects) or by a description (ISO19115-1:EX_GeographicDescription), which has an identifier (ISO19115-1:MD_Identifier). This identifier refers to an object in a namespace, and the namespace can reference to an ontology/vocabulary by using the online resource element of the citation class (ISO19115-1:CI_Citation). The temporal extent (ISO19115-1:EX_TemporalExtent) can only be provided as value (ISO19108:TM_Primitive). However, the scope codes that are defined in ISO19115-1 do not allow to reference terms from an ontology/vocabulary - examples in ISO-19157 only constrain the scope thematically by pointing to feature classes in the dataset. The scope description class contains a free text field ("other") that could be used to link to a term, but that does not allow automated processing like type checking. In summary, the first requirement is only fulfilled partially (R1 ✗) and we do not consider ISO19157 to provide a simple and easy-to-use structure (F.R4 ✗).

The Data Quality Vocabulary (DQV) is a semantic data vocabulary (F.R3 ✓), designed along the Data Catalog Vocabulary (DCAT)[6], to describe a dataset's quality. Nonetheless, DQV can also be used independently to describe the quality of any resource on the Web (Albertoni and Isaac, 2021). DQV is built upon the three classes category, dimension, and metric proposed by Zaveri et al. (2015). A data quality measurement is performed against a certain quality metric. This metric is defined in a certain quality dimension, which belongs to a certain category. Thus, the category defines a broad concept of quality, the dimension further specifies this concept, and the metric describes options to measure a certain dimension.

The development of DQV is strongly guided by *reusing vocabularies* and *minimizing ontological commitment* (Best Practices on the Web[7] 15 and 16) (Albertoni and Isaac, 2021). DQV provides a generic and extensible concept regarding the definition of domain specific quality indicators, thus guaranteeing options to define any spatial quality indicator (F.R 1 ✓).

Albeit DQV cannot model inhomogeneous data quality yet, we choose DQV for the development of a profile. One reason for this is that it is a semantic data vocabulary (F.R3), which in turn simplifies the implementation of requirement R1. Further, we consider DQV easier to understand and to apply (F.R4). Moreover, tailoring ISO19157 to fulfill requirement R1 by changing the temporal extent class, as well as the items of the scope code/scope description class, would lead to a more complex structure. The required changes of the scope code/scope description class would include deleting items, which would lead to incompatibility with the default version (F.R2 ✗). In contrast, to constrain a geodata quality measurement in DQV, the RDF node that represents the measurement can simply be attached with the needed constraints (F.R4 ✓).

[6]https://www.w3.org/TR/vocab-dcat-3/, April 20, 2022
[7]https://www.w3.org/TR/dwbp/, April 20, 2022

4 Methods: Discussion of Design Choices and Implementation of the Profile

In this section, we discuss different possibilities to implement the constraints for a certain quality measurement in DQV. At first, we provide guidance to choose proper terms and possible formats (see R1). We then review the RDF properties that we used to model the constraints and discuss two implementation options.

Using terms from ontologies/vocabularies is in particular recommended for thematic constraints, e.g., to ensure a common understanding of used terms. An example of a well-known vocabulary for agricultural terms is AGROVOC (Caracciolo et al., 2013). If there is no proper term available in an existing ontology/vocabulary, comprehensive definitions can be implemented with the Simple Knowledge Organization System (SKOS)[8] or with the Web Ontology Language (OWL)[9]. Regarding spatial constraints, the Marine Regions Gazetteer (Lonneville et al., 2021) serves as an example to find proper and shared spatial terms. Commonly known serializations of geodata are GML[10] or WKT[11]. Finally, concerning temporal constraints, we assume that the required periods to temporally constrain a geodata quality measurement are often dataset specific, which results in rare occurrences of feasible proper terms. Encodings are available in ISO 8601[12] or OWL-Time[13].

The definition of the RDF properties is guided by the best practice on re-use of existing vocabulary (see Sec. 3). Consequently, we implement the spatial and temporal constraints with the Dublin Core (DC) Terms[14] *spa-*

[8]https://www.w3.org/TR/skos-reference/#, April 20, 2022
[9]https://de.wikipedia.org/wiki/Web_Ontology_Language, April 20, 2022
[10]Geographic Markup Language: https://www.ogc.org/standards/gml, April 20, 2022
[11]Well-known text: https://www.ogc.org/standards/wkt-crs, April 20, 2022
[12]https://www.iso.org/standard/70907.html, April 20, 2022
[13]https://www.w3.org/TR/owl-time/, April 20, 2022
[14]https://www.dublincore.org/specifications/dublin-core/dcmi-terms/, April 20, 2022

tial[15] and *temporal*[16]. Both properties fulfill Requirement R2. To model the thematic constraint, we defined the property *thematic*. Existing properties, like DCAT's *theme*[17] or SKOS's *inScheme*[18], target whole ontologies/vocabularies, instead of referencing specific terms that are defined in an ontology/vocabulary. We recommend encoding spatial constraints (DCT Locations) with the GeoSPARQL[19] ontology, temporal constraints (DCT Periods of Time) with the OWL-Time ontology and thematic constraints with the SKOS vocabulary.

To constrain a measurement, two implementations are feasible: defining a node that describes the constraints and that is subsequently attached to the measurement node (Figure 1 – version 1[20]) or adding the spatial, temporal and thematic constraints to the measurement node (Figure 1 – version 2).

```
@prefix dqv: <http://www.w3.org/ns/dqv#> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix profile: <https://ex.org/profile/#> .
@prefix : <https://ex.org/#> .

# version 1 - defining a separate constraint object
:qualityMeasurement a dqv:QualityMeasurement;
    dqv:isMeasurementOf :qualityMetric ;
    dqv:value 42 ;
    profile:hasConstraint :constraint .

:constraint a profile:Constraint ;
    dct:spatial <...> ;
    dct:temporal <...> ;
    profile:thematic <...> .

# version 2 - adding the constraints to the measurement
:qualityMeasurement a dqv:QualityMeasurement ;
    dqv:isMeasurementOf :qualityMetric ;
    dqv:value 42 ;
    dct:spatial <...> ;
    dct:temporal <...> ;
    profile:thematic <...> .
```

Figure 1. Two versions to apply spatial, temporal, and thematic constraints to a DQV quality measurement.

In both cases, the constraints are optionally - if the data quality is homogeneous, the constraints can simply be left out. Version 2 only allows reusing terms that are solely spatial (e.g., Europe), temporal (the 90s) or thematic (forest), but not a combination. Whereas Version 1 allows combinations, but is more complex. An example is the spatiotemporal term 'Little Ice Age' that can be considered a phenomenon of the Northern Hemisphere in the time span from about 1400 to 1850 (Matthews and Briffa, 2005). In version 2, we have to use two nodes to specify the example; one for the spatial extent and one for the temporal extent of Little Ice Age (see Fig. 3), whereas in version 1, we could simply use one node that holds the definition of the term 'Little Ice Age' as constraint (see Fig. 2). A common problem with these more complex terms is the co-existence of different domain-specific definitions. Approaches like 'deep' and 'broad' gazetteers[21] address this issue, e.g., by allowing co-existence of terms or fuzzy descriptions (Shaw, 2016). Although definition and reuse of such more complex terms has potential, currently proper definitions are missing in ESS. Consequently, we chose and applied version 2 in our projects, i.a., due to its simplicity. Additionally, the second version is coherent to the implementation of the spatial and temporal extents of datasets in (Geo)DCAT.

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dqv: <http://www.w3.org/ns/dqv#> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix xsd: <https://www.w3.org/TR/xmlschema-2/#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix profile: <https://ex.org/profile/#> .
@prefix : <https://ex.org/#> .

# this node could be available on the web
:northernHemisphere a dct:Location, gsp:Feature;
    gsp:hasDefaultGeometry [
        gsp:asWKT "POLYGON((-180 0,-180 90,180 90,180 0,-180 0))
        "^^gsp:wktLiteral
    ] .

# this node could be available on the web
:littleIceAge a profile:Constraint ;
    dct:spatial :northernHemisphere;
    dct:temporal [
        a dct:PeriodOfTime , time:ProperInterval ;
        time:hasBeginning [
            a time:Instant ;
            time:inXSDgYear "1570"^^xsd:gYear ;
        ];
        time:hasEnd [
            a time:Instant ;
            time:inXSDgYear "1900"^^xsd:gYear ;
        ]
    ] .

:measurementOfGeodataQuality a dqv:QualityMeasurement ;
    dqv:isMeasurementOf :aDataQualityMetric ;
    dqv:value "42"^^xsd:integer ;
    profile:isConstrainedBy :littleIceAge .
```

Figure 2. Little Ice Age - version 1: Little Ice Age is defined as separate node and includes the spatial and temporal constraint of the measurement. Its spatial extent is also defined in a separate node, since Northern Hemisphere serves as a reusable concept. The temporal constraint is defined inline by using a blank node.

4.1 Data and Software Availability

Both versions of the profile are published as open source RDF-TTL on GitHub[22] together with several examples for the recommended encodings of the constraints, and an im-

[15] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/#spatial, April 20, 2022

[16] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/#temporal, April 20, 2022

[17] https://www.w3.org/TR/vocab-dcat-3/#Property:resource_theme, April 20, 2022

[18] https://www.w3.org/2009/08/skos-reference/skos.html#inScheme, April 20, 2022

[19] https://www.ogc.org/standards/geosparql, April 20, 2022

[20] GitHub: https://github.com/rue-a/dqv_inhomogeneity/blob/master/listing1_implementation_versions.ttl, April 20, 2022

[21] A definition of the 'Little Ice Age' in the PeriodO gazetteer: http://n2t.net/ark:/99152/p0jk4xk2gz6, April 20, 2022

[22] https://github.com/rue-a/dqv_inhomogeneity, April 20, 2022

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dqv: <http://www.w3.org/ns/dqv#> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix xsd: <https://www.w3.org/TR/xmlschema-2/#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix gsp:  <http://www.opengis.net/ont/geosparql#>.
@prefix time: <http://www.w3.org/2006/time#> .
@prefix profile: <https://ex.org/profile/#> .
@prefix : <https://ex.org/#> .

# this node could be available on the web
:northernHemisphere a dct:Location, gsp:Feature;
    gsp:hasDefaultGeometry [
        gsp:asWKT "POLYGON((-180 0,-180 90,180 90,180 0,-180 0))
        "^^gsp:wktLiteral
    ] .

:measurementOfGeodataQuality a dqv:QualityMeasurement ;
    dqv:isMeasurementOf :aDataQualityMetric ;
    dqv:value "42"^^xsd:integer ;
    dct:spatial :northernHemisphere;
    dct:temporal [
        a dct:PeriodOfTime , time:ProperInterval ;
        time:hasBeginning [
            a time:Instant ;
            time:inXSDgYear "1570"^^xsd:gYear ;
        ];
        time:hasEnd [
            a time:Instant ;
            time:inXSDgYear "1900"^^xsd:gYear ;
        ]
    ] .
```

Figure 3. Little Ice Age - version 2: The measurement is provided with a spatial and temporal extent, which constrain it in space and time.

plementation for the Global Forest Change dataset (see Sec. 1).

The quality information of the Global Forest Change dataset is published in the supplementary material[23] of Hansen et al. (2013).

5 Conclusion and Future Work

Data quality information that shows an inhomogeneous distribution across a dataset is seldom part of the dataset's metadata. Our proposed DQV profile focuses on reducing the required effort for data producers to integrate this information in the metadata by proposing an implementation that (1) reuses known vocabularies, (2) only defines one additional RDF property, (3) aligns the profile's structure with DCAT and GeoDCAT, and (4) ensures the compatibility with the default version of DQV.

With our profile implementation and the developed examples, we show a proof-of-concept. Moreover, we identified several use cases for the profile as future work.

For instance, when using the profile for discovery, e.g. in metadata catalogs, it enables data users to formulate precise queries on a sub-dataset level of detail, thus facilitating a dataset's proper use. It furthermore fosters the development of advanced metadata/data quality analysis applications, like for accessing precise geodata quality information via a metadata catalog's API and directly utilizing it in data processing workflows.

Moreover, using a semantic data approach allows for seamless integration with other linked data concepts, thus encouraging novel developments. An example is the intersection with the provenance ontology PROV-O[24], which bears rich automation potential. Inhomogeneity in the data quality often results from inhomogeneous input data (see Sec. 1); by leveraging provenance information, the spatial, temporal, and/or thematic boundaries of input datasets can be traversed along the provenance graph to automatically derive constraints for data quality measurements in the output dataset, which in turn opens up possibilities to automatically apply meaningful data quality measurements.

Furthermore, the implementation of the example dataset 'Global Forest Change'[25] with our profile, exposed the benefits of using proper terms. Both in the publication and the supplementary material of the dataset, the authors refer to the used spatial extents as 'FAO climate domains'. With that, they mean the FAO Global Ecological Zones (GEZ) in the version of 2010. In the example, we implement the tropical, subtropical, temperate and boreal zones as reusable (and therefore referenceable) terms.

Generally, the application of the profile in conjunction with concepts like provenance, thematic vocabularies or essential (land use) variables[26] contributes to new prospects for knowledge modelling and gaining.

References

Albertoni, R. and Isaac, A.: Introducing the Data Quality Vocabulary (DQV), Semantic Web, 12, 81–97, https://doi.org/10.3233/SW-200382, 2021.

Ariza-López, F. J., González, P. B., Pau, J. M., Torres, A. Z., Pascual, A. F. R., Vergara, G. M., and Balboa, J. L. G.: Geospatial data quality (ISO 19157-1): evolve or perish, Revista Cartográfica, pp. 129–154, https://doi.org/10.35424/rcarto.i100.692, 2020.

Barsi, , Kugler, Z., Juhász, A., Szabó, G., Batini, C., Abdulmuttalib, H., Huang, G., and Shen, H.: Remote sensing data quality model: from data sources to lifecycle phases, International Journal of Image and Data Fusion, 10, 280–299, https://doi.org/10.1080/19479832.2019.1625977, 2019.

Caracciolo, C., Stellato, A., Morshed, A., Johannsen, G., Rajbhandari, S., Jaques, Y., and Keizer, J.: The AGROVOC Linked Dataset, Semantic Web, 4, 341–348, https://doi.org/10.3233/SW-130106, 2013.

Debattista, J., Lange, C., and Auer, S.: daQ, an Ontology for Dataset Quality Information, vol. 1184 of *CEUR Workshop Proceedings*, CEUR, Seoul, Korea, http://ceur-ws.org/Vol-1184/#paper9, 2014.

[23] https://www.science.org/doi/suppl/10.1126/science.1244693/suppl_file/hansen.sm.pdf, April 20, 2022

[24] https://www.w3.org/TR/prov-o/, April 20, 2022

[25] https://github.com/rue-a/dqv_inhomogeneity/blob/master/version_2/global_forest_change.ttl, April 20, 2022

[26] https://earthdata.nasa.gov/learn/backgrounders/essential-variables, April 20, 2022

Debattista, J., Auer, S., and Lange, C.: Luzzu—A Methodology and Framework for Linked Data Quality Assessment, J. Data and Information Quality, 8, https://doi.org/10.1145/2992786, 2016.

Egli, L., Fischer, J., and Henzen, C.: Reporting relevance, availability, and needs of quality information in Earth System Science Data, https://doi.org/10.5281/zenodo.5562326, 2021.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., and Townshend, J. R. G.: High-Resolution Global Maps of 21st-Century Forest Cover Change, Science, 342, 850–853, https://doi.org/10.1126/science.1244693, 2013.

Hunt, S. E., Albinet, C., Nickeson, J., Hall, A., Fox, N., Boccia, V., and Goryl, P.: A Quality Assurance Framework for Satellite Earth Observation Missions, in: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 608–611, https://doi.org/10.1109/IGARSS47720.2021.9553572, iSSN: 2153-7003, 2021.

Lonneville, B., Delva, H., Portier, M., Maldeghem, L. V., Scheppers, L., Bakeev, D., Vanhoorne, B., Tyberghein, L., and Colpaert, P.: Publishing the Marine Regions Gazetteer as a Linked Data Event Stream, vol. 2969 of *CEUR Workshop Proceedings*, CEUR, Bolzano, Italy, http://ceur-ws.org/Vol-2969/#paper8-s4biodiv, 2021.

Matthews, J. A. and Briffa, K. R.: The 'little ice age': re-evaluation of an evolving concept, Geografiska Annaler: Series A, Physical Geography, 87, 17–36, https://doi.org/10.1111/j.0435-3676.2005.00242.x, 2005.

Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., and Haklay, M. M.: A review of volunteered geographic information quality assessment methods, International Journal of Geographical Information Science, 31, 139–167, https://doi.org/10.1080/13658816.2016.1189556, 2017.

Shaw, R.: Gazetteers Enriched: A Conceptual Basis for Linking Gazetteers with Other Kinds of Information, pp. 51–64, Indiana University Press, http://www.jstor.org/stable/j.ctt2005zq7.9, 2016.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data, 3, 160 018, https://doi.org/10.1038/sdata.2016.18, 2016.

Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J., and Auer, S.: Quality assessment for Linked Data: A Survey, Semantic Web, 7, 63–93, https://doi.org/10.3233/sw-150175, 2015.