

Article

# Reducing Consumer Uncertainty: Towards an Ontology for Geospatial User-Centric Metadata

Hasti Ziaimatin <sup>\*</sup>, Alireza Nili  and Alistair Barros

School of Information Systems, Queensland University of Technology, Brisbane, QLD 4000, Australia;  
a.nili@qut.edu.au (A.N.); alistair.barros@qut.edu.au (A.B.)

\* Correspondence: hastim.matin@gmail.com; Tel.: +61-407-153-705

Received: 4 July 2020; Accepted: 9 August 2020; Published: 12 August 2020



**Abstract:** With the increased use of geospatial datasets across heterogeneous user groups and domains, assessing *fitness-for-use* is emerging as an essential task. Users are presented with an increasing choice of data from various portals, repositories, and clearinghouses. Consequently, comparing the quality and evaluating fitness-for-use of different datasets presents major challenges for spatial data users. While standardization efforts have significantly improved metadata interoperability, the increasing choice of metadata standards and their focus on data production rather than potential data use and application, renders typical metadata documents insufficient for effectively communicating fitness-for-use. Thus, research has focused on the challenge of *communicating fitness-for-use* of geospatial data, proposing a more “*user-centric*” approach to geospatial metadata. We present the Geospatial User-Centric Metadata ontology (GUCM) for communicating fitness-for-use of spatial datasets to users in the spatial and other domains, to enable them to make informed data source selection decisions. GUCM enables metadata description for various components of a dataset in the context of different application domains. It captures producer-supplied and user-described metadata in *structured* format using concepts from domain-independent ontologies. This facilitates *interoperability* between spatial and nonspatial metadata on open data platforms and provides the means for *searching/discovering* spatial data based on user-specified quality and fitness-for-use criteria.

**Keywords:** spatial metadata; fitness-for-use; uncertainty handling in spatial data; user-centric metadata; ontology; structured metadata; interoperability and open systems

## 1. Introduction

The quality of geospatial data has been the subject of extensive research in the GIS community for more than thirty years [1]. It has drawn considerable attention from the academic community and government agencies and, more recently, from industry. Geospatial data are subject to processes such as generalization, abstraction and aggregation; consequently, the transformed data can only provide an approximation of the real world, and often suffers from imperfect quality [2,3]. Thus, data consumers will always be exposed to some level of data uncertainty. Spatial data quality and uncertainty are two of the fundamental theoretical issues in geographic information science, where there is a keen interest to quantify, model and visualize the accuracy of spatial data in more sophisticated ways. Furthermore, spatial data supply chains today push data to the users via spatial Web portals or Web services. The value of this information depends on the ability to anticipate users’ needs and quality requirements. This approach is proving problematic given the unforeseeable and diverse nature of user requirements in the context of various application domains. The value of spatial data products is realized when the delivered knowledge enables users to achieve their intended purposes [4].

The quality of spatial data depends on the producer’s perception, i.e., internal quality, and the user’s perspective, i.e., external quality. Objective quality measures of geospatial data (internal

quality) relate to the “difference between the data and the reality that they represent” [5]. In the GIS domain, internal quality is often described in terms of the ‘famous five’ elements of geospatial data quality; i.e., *positional accuracy*, *attribute accuracy*, *temporal accuracy*, *logical consistency* and *completeness*. The internal quality of data can be improved during the course of data creation [3]. Subjective measures of quality (external quality) relate to a data source’s “fitness-for-use”; i.e., in order to assess the quality of data, we need to have information about the data in addition to the actual user needs [6]. It largely depends on user requirements and therefore the same product can be of different quality to different users. Despite the diversity in notions of internal (objective) and external (subjective) data quality, these two categories are closely linked together because, in order to evaluate external data quality, users will often require objective data quality descriptions. While there are methods for evaluation of internal quality of geospatial data, evaluation of external quality still remains an open issue. Experts and users of geospatial data are expected to know the type of spatial data resource they need, and the clearinghouse or geoportal that houses that data. Even with this information available, users are still left with the decision on fitness-for-use, based on complex metadata, for the few cases where such metadata exist [7].

To address issues of geospatial data quality, standardization bodies such as the International Organization for Standardization (ISO/TC211 (<https://committee.iso.org/home/tc211>)), Infrastructure for Spatial Information in Europe (INSPIRE (<https://inspire.ec.europa.eu/>)), Open Geospatial Consortium (OGC (<http://www.opengeospatial.org/>)), Dublin Core Metadata Initiative (DCMI (<http://dublincore.org/>)) and Federal Geographic Data Committee (FGDC (<https://www.fgdc.gov/>)), are actively working to establish, improve and extend geospatial data and metadata standards.

While standardization efforts have significantly improved metadata interoperability, the increasing choice of metadata standards poses a number of challenges: e.g., (i) Which standards should be used to describe quality? (ii) How much metadata should be provided to enable users to identify data sources that are fit for their intended uses? Finally, (iii) what quality information should be provided to make metadata “useful” and not just “usable” [8–12]? In addition, despite detailed recommendations of standardization bodies and the existence of formal metadata standards, often data quality information is not communicated to users in a consistent and standardized way [13].

There is currently a shortage of empirical research relating to how people interpret and use spatial data quality information for individual data sources in a real-world environment. A study conducted in the field of spatial data quality argues that fitness-for-use relates specifically to each individual use case; as a result, providing generic quality information is usually unhelpful to consumers [3]. Therefore, over the past decade, research has focused on the challenge of communicating fitness-for-use of geospatial data, proposing a more “user-centric” approach to geospatial metadata [11]. Researchers have proposed enriching metadata records with references to relevant literature (Citation information); less formal opinions from data producers; expert opinions of data quality; and user feedback regarding previous data use [10,14]. However, recent reviews suggest that these recommendations have not yet been put into practice, and there are no practical means for collating and searching user-focused metadata. Furthermore, many of the metadata records that are available, are in fact, incomplete [15,16].

The focus of our research is to communicate the quality of spatial data in terms of fitness to meet increasingly diverse applications, in order to enable spatial data users to make informed data source selection decisions. Towards this goal, our study aims to identify information that conveys fitness-for-use of geospatial data sources to users in spatial and nonspatial domains and to enable producers and users of geospatial data to describe fitness-for-use and metadata for datasets. More specifically, our study provides the design of an ontology through analysis of existing geospatial and domain-independent standards and vocabularies and requirements elicitation drawn from organizations across multiple sectors including councils, utility companies, government departments and social services, police forces, the health sector, insurance companies, engineering, consulting

and construction services, software and information technology companies, and the mining sector. Our study aims to achieve the following research objectives:

- O1. *Sufficiency of language*: the support of capturing and representing metadata and fitness-for-use descriptions for a dataset at various levels of *granularity*; i.e., the *dataset*, its *features* and *attributes*.
- O2. *Utility of language*: the support of communicating fitness-for-use of datasets, by enabling both producers and users of geospatial data to generate metadata and fitness-for-use descriptions, using a *single model*.
- O3. *Extensibility of language*: The support of domain-independent and widely adopted standards and vocabularies for describing metadata in order to facilitate *interoperability* between metadata descriptions from spatial and nonspatial domains, thereby making spatial data sources *searchable on open data portals*.

Our overall goal is to enable users to interpret spatial data quality in the context of their domain and for their intended purposes and applications, thereby providing users with the means to make informed data source selection decisions. Towards this goal, we designed an ontology to *communicate spatial metadata* and *implicit* knowledge of spatial data sources within various applications and domains. The continuous adoption of Semantic Web technologies [17], enabled us to transform the vocabulary into a more dynamic and well-grounded formalism; i.e., an ontology. In this paper, we introduce the *Geospatial User-Centric Metadata (GUCM) ontology*, which captures and represents *metadata* and *fitness-for-use* descriptions of geospatial data sources within various applications and domains, using machine-processable concepts defined in widely adopted ontologies. This in turn facilitates *interoperability* between spatial and nonspatial metadata on open data platforms. More concretely, this paper brings the following contributions:

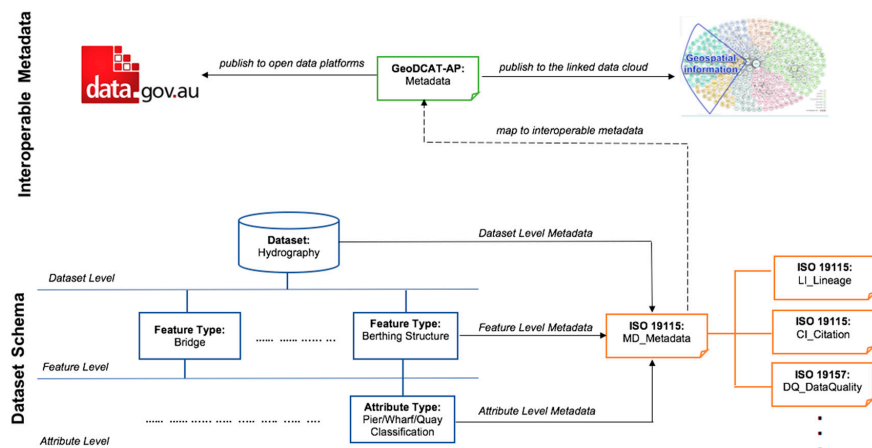
- (i) *Eliciting user and producer views on geospatial data quality and fitness-for-use*, by conducting (a) *semi-structured interviews* with spatial data producers and users from a variety of domains and applications; and (b) a *survey* to collect data from more diverse GIS communities. The survey aimed to examine findings from our industry engagements and to potentially identify additional quality themes (from the comments and suggestions) for assessing and evaluating spatial data quality and fitness-for-use;
- (ii) *Qualitative and quantitative analysis of our industry engagements* to identify the gaps that exist between *internal quality* (producer supplied) and *external quality* (consumer described). This analysis informed the design of the *Geospatial User-Centric Metadata ontology*;
- (iii) *Designed the Geospatial User-Centric Metadata (GUCM) ontology*, for communicating metadata and fitness-for-use of geospatial data sources to users, in order to enable them to make informed data source selection decisions.

### 1.1. Motivation and Significance

The GUCM ontology communicates spatial metadata and fitness-for-use descriptions of spatial datasets, using concepts from domain-independent and widely adopted vocabularies and ontologies. This structured information can be published to the Web of Data [18], and Open Data Portals such as the Australian government's public data platform (<https://data.gov.au>), a platform for discovering, accessing and reusing public data (Figure 1). This will in turn provide a means to *search* and *discover* spatial data based on metadata and dataset usage descriptions, in addition to facilitating *interoperability* between metadata for spatial datasets and metadata for datasets from nonspatial domains.

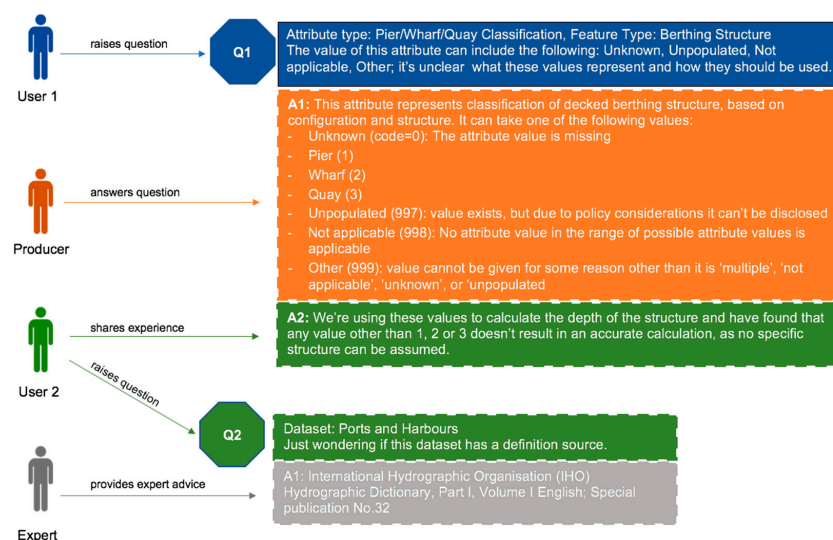
The ontology can be used not only to assess fitness-for-use of datasets in the context of specific applications and domains, but also to identify different use cases and users of datasets, as they describe fitness-for-use in the context of their application domains. In addition, producers can incorporate user-defined metadata into objective quality measures for their products, allowing providers to meet users' specific requirements. Furthermore, the vocabulary can be used to complement producer

metadata by presenting user-defined fitness-for-use descriptions for various applications of a dataset. Additionally, the hierarchical structure of ontologies enables metadata and fitness-for-use descriptions to target *various components* of a dataset; i.e., dataset, feature type, or attribute type. This in turn, facilitates dataset search and discovery based on metadata and usage descriptions for specific dataset components. Moreover, the ontology captures profiles of users that describe their experiences with spatial datasets in the context of specific use cases. These profiles can be used to assess the fitness-for-use of spatial datasets in light of attributes such as level of expertise, application domain and roles of users that provide these descriptions.



**Figure 1.** Interoperability of metadata from spatial and nonspatial domains on Linked Open Data and Open Data Portals.

The GUCM ontology enables both producers and users of geospatial data to create metadata and fitness-for-use descriptions using *one model*, rather than separate producer and user models. In other words, the ontology captures and represents producer-supplied metadata, in addition to users' experiences, feedback and fitness-for-use descriptions. Furthermore, GUCM enables producers and users to *communicate* and *discuss* data and its fitness-for-use in the context of specific use cases. Figure 2 illustrates examples of question-and-answer discussions between spatial data users, producers and experts.



**Figure 2.** Examples of user feedback metadata and communication among spatial data users, producers and experts.

## 1.2. Related Work

In this section, we present an overview of the related efforts and existing standards and specifications on data quality with particular emphasis on communicating fitness-for-use of geospatial data and “user-centric” approaches to geospatial metadata.

Describing data quality in terms of characteristics inherent to data means that we subscribe to a multidimensional conceptualization of data quality [19]. Briefly, these inherent characteristics, also called dimensions of data quality, include concepts such as accuracy, relevance, accessibility, currency, timeliness, and completeness. The initial work establishing the multidimensional conceptualization of data quality identified over 200 dimensions in use across surveyed organizations from different industries [20]. For most data uses, only a handful of dimensions are deemed important enough to formally measure and assess. The dimensions measured in data quality assessment should be those necessary to indicate fitness of the data for a particular use.

A recent study proposed a reference framework that supports measuring the overall quality of datasets [21]. The framework consists of eight quality items, three of which (accuracy, completeness and consistency) are particularly relevant to assessing quality of geospatial information. These include Accuracy, Completeness, Consistency, Redundancy (also considered as Conciseness), Readability, Accessibility and Availability, Usefulness and Trust (which covers the Reliability aspect).

Another empirical study produces semantic metadata to allow users to search, filter and rank datasets according to a number of quality criteria, thereby enabling users to discover relevant, fit-for-use datasets according to their requirements [22].

Ireland’s national mapping agency, Ordnance Survey Ireland (OSi), is responsible for the digitization of the island’s infrastructure with regards to mapping. OSi built its knowledge in a framework by generating data from various sensors (e.g., spatial sensors). A subset of the knowledge captured by this framework is transformed into geo-linked data. A recent initiative has set up a scalable linked-data quality assessment framework, in the OSi’s pipeline to continuously assess produced data in order to address any quality issues before publishing [23].

DaVe, a data value vocabulary that allows for the comprehensive representation of data value, enables users to extend it using data value dimensions as required in a given context. DaVe enables consensus on what characterizes data value and how to model it. This vocabulary enables users to monitor and assess data value during value creating or data exploitation efforts. It also enables the integration of different metrics that include many data value dimensions. This vocabulary is based on requirements from a number of value assessment use cases extracted from literature [24].

Assessing the quality and consistency of data requires data standards. Data standards are tools that enable interoperability and promote data quality. In other words, data quality is assessed by identifying important dimensions and measuring them. The key international organizations that develop standards for geospatial information are: (i) International Organization for Standardization (ISO) Technical Committee 211 for Geographic Information/Geomatics (<http://www.isotc211.org/>) and (ii) Open Geospatial Consortium (OGC) (<http://www.opengeospatial.org/>). The ISO, OGC and increasingly World Wide Web Consortium (W3C) (<https://www.w3.org/>) standards provide a set of constructs that enable data to be specified and delivered in a standardized interoperable manner. These standards provide a framework within which data products are developed. According to ISO 9000 [25], (Section 3.1.5 (formerly ISO 8402: 1994)), quality is defined as “the totality of characteristics of an entity that bears upon its ability to satisfy stated and implied needs.” And “The purpose of describing the quality of geographic data is to facilitate the comparison and selection of the dataset best suited to application needs or requirements” [26]. The ISO definition emphasizes that quality corresponds to the *intended use*; i.e., a dataset could be a perfect match to a particular use case, while being unsuitable to another user’s requirements. Therefore, validating the quality of geo-information is a key requirement for users and producers of spatial information. This in turn facilitates evaluation of fitness for purpose for particular applications, especially in the context of political reporting and decision-making [27].



Quality assured validation must adhere to the principles of transparency, traceability, independence, accessibility, and representativeness.

ISO Technical Committee ISO/TC 211, has developed a series of international standards that provide a conceptual modelling framework for geospatial information. This includes constructs that define how specific aspects of spatial information should be modelled regardless of application. These standards provide a framework within which information models can be developed for different application domains in a consistent manner. Both the ISO framework standards and the application schema developed from them are “information standards”. More specifically, ISO 19115-1:2014 Geographic info metadata fundamentals [28] defines the schema for describing geographic information and services through metadata. It represents information about the quality, spatial and temporal aspects, extent, spatial reference, distribution, and other properties of digital geographic data and services. ISO 19157:2013 Geographic information—Data quality [26], establishes the principles for describing the quality of geographic data and defines a set of data quality measures for use in evaluating and reporting data quality. It also defines and provides clarity to quality elements such as: completeness, logical consistency, positional accuracy, temporal quality, thematic accuracy and usability. ISO/TS 19158:2012 Geographic information—Quality assurance of data supply [29], represents a framework for quality assurance that is specific to geographic information. It is based on the quality principles and quality evaluation techniques of geographic information identified in ISO 19157:2013 [26], and the general quality management principles defined in ISO 9000 [25]. The framework defined in this standard provides customers with the assurance that internal and external suppliers are capable of delivering geographic information to the required quality.

While these standards provide an excellent framework for conveying quality to the user community, they are insufficient for assessing fitness-for-use of spatial data sources, as they are mostly focused on the processes used in producing data, rather than potential ways in which data are used in various applications and domains. In the majority of cases, metadata are not supplied or are incomplete and, where spatial metadata are supplied, information is complex and requires specialized expertise and knowledge to understand and interpret. In addition, end products are often derived from a variety of sources and datasets are used in various applications and domains. Therefore, a typical metadata document is not sufficient to effectively communicate “fitness-for-use” to consumers from a variety of domains and expertise levels [11]. In other words, it is difficult to issue simple statements about a dataset’s quality or to label a particular dataset as “the best”.

The Geospatial User Feedback (GUF (<https://www.opengeospatial.org/standards/guf>)) is an OGC standard, which defines a conceptual data model (OGC 15-097 (<http://docs.opengeospatial.org/is/15-097r1/15-097r1.html>)) and a practical XML encoding of the conceptual model (OGC 15-098 (<http://docs.opengeospatial.org/is/15-098r1/15-098r1.html>)). Geospatial User Feedback presents metadata that is mainly produced by the consumers of geospatial data products as they use and gain experience with those products. The standard enables users to document feedback items such as ratings, comments, quality reports, citations, significant events, etc. about the way in which data is used. Feedback items can be aggregated in collections and summaries of the collections can also be described. This standard, complements current metadata practices, where documents describing dataset characteristics and production workflows are generated by the producer of a data product. While this standard is a significant contribution to user-centric metadata, it does not support interoperability between spatial metadata and metadata from other domains.

The GeoViQua FP7 (<http://www.geoviqua.org/>) project significantly contributed to the Global Earth Observation System of Systems (GEOSS (<http://www.earthobservations.org/geoss.php>)) Common Infrastructure (GCI ([https://www.earthobservations.org/gci\\_gci.shtml](https://www.earthobservations.org/gci_gci.shtml))) by adding rigorous data quality representations to search and visualization GEO Portal (<http://www.geoportal.org/>) functionalities, prioritizing interoperability at all times. GeoViQua combines geospatial data with information on data quality and processing services within GEOSS catalogues. GeoViQua contributed to an enhanced, user-driven, and practical GEO Label (<http://www.geoviqua.org/GeoLabel.htm>) and thus

enabled increasing user trustworthiness over GEOSS data and services delivery. The GEO Label was proposed as a value indicator for geospatial data and datasets accessible through the GEOSS thereby assisting in searching activities by providing users with visual cues of dataset quality and possibly relevance. While graphical representation of metadata parameters enables users to easily screen the data, GeoViQua's quality model is based on its producer quality framework and its innovative user feedback model for geospatial data ([http://www.geoviqua.org/Docs/GeoViQua\\_book\\_v3.pdf](http://www.geoviqua.org/Docs/GeoViQua_book_v3.pdf)). This is in contrast to our proposed GUCM model, which aims to capture and represent producer-supplied and user-described metadata using a single model.

BIOME is a lightweight ontology designed for the management of datasets in the biodiversity domain [30]. Domain-specific metadata captured by the ontology conform to the INSPIRE directive [31]. In order to promote interoperability, the ontology is designed to establish relationships between INSPIRE concepts and concepts from widely used ontologies, such as Dublin Core (<http://dublincore.org/>), FOAF (<http://xmlns.com/foaf/spec/>), and CERIF (<https://www.eurocris.org/ontologies/semcerif/1.3/index.html>) ontologies. Consequently, biodiversity metadata records can be published as Linked Open Data [18]. While this initiative promotes interoperability, it is focused on dataset-level descriptors, in order to make biodiversity datasets searchable and understandable by domain experts. We contrast this with the Geospatial User-Centric Metadata ontology (GUCM) introduced in this paper, which facilitates metadata description at various levels of dataset granularity, i.e., dataset, feature types and attribute types for all datasets, irrespective of the domain. More importantly, the GUCM model facilitates and differentiates between metadata that is solely created and maintained by producers and metadata created by users. Furthermore, GUCM facilitates discussion and communication between producers and users of geospatial data.

The increasing ubiquity of spatial data has raised the need for seamless integration with other data on the Web. Efforts to clarify, formalize and harmonize spatial and Web standards have been recently completed by the Spatial Data on the Web Working Group (SDWWG (<https://www.opengeospatial.org/projects/groups/sdwwg>)), a collaboration between the OGC and the World Wide Web Consortium (W3C). In particular, the aim of the SDWWG has been to: (i) determine how spatial information can be best integrated with other data on the Web; (ii) discover how machines and people can determine that different facts in different datasets represent the same place, especially when "place" is expressed in different ways and at different levels of granularity; (iii) identify and evaluate current methods and tools and determine a set of best practices for their use; and (iv) complete the standardization of informal technologies already in widespread use.

The Data Catalog Vocabulary (DCAT (<https://www.w3.org/TR/vocab-dcat/>)) is a Resource Description Framework (RDF (<https://www.w3.org/RDF/>)) vocabulary developed to enable interoperability between data catalogs published on the Web. The DCAT Application profile for data portals in Europe (DCAT-AP ([https://ec.europa.eu/isa2/solutions/dcat-application-profile-data-portals-europe\\_en](https://ec.europa.eu/isa2/solutions/dcat-application-profile-data-portals-europe_en))) provides a specification, which is based on DCAT and is designed for describing public sector datasets in Europe. This specification enables discovery of datasets by searching across data portals, thereby making public sector data better searchable across borders and sectors. This can be achieved by the exchange of descriptions of datasets among data portals. GeoDCAT-AP (<https://joinup.ec.europa.eu/release/geodcat-ap/v101>) is an extension of DCAT-AP that describes geospatial datasets and services. It provides an RDF syntax binding for the union of metadata elements defined in the core profile of ISO 19115:2003 [32], and those defined in the framework of the INSPIRE directive [31]. It aims to facilitate the search and discovery of spatial datasets, data series, and services on general data portals, thereby making geospatial information better searchable across borders and sectors. This can be achieved by the exchange of descriptions of datasets among data portals.

The *Geospatial User-Centric Metadata ontology (GUCM)* presented in this paper aims to bridge the gap between producer and user views on geospatial data quality and fitness-for-use. Not only does the ontology capture *users' experiences* with spatial data sources in the context of different applications and domains, but it also captures information deemed necessary by users for assessing fitness-for-use,

such as dataset structure; i.e., details of feature types and attribute types, their relationships and links, allowed values, values appropriate for certain use cases, expert reviews of the data at different levels of granularity; e.g., review of the dataset, its features and attributes, producer profile, dataset citations and dataset distributions. Furthermore, GUCM facilitates metadata description for a dataset and all of its components, e.g., a metadata description, an expert's review, or a user feedback, can be associated with a dataset, a feature type or an attribute type. Finally, GUCM represents metadata using concepts from domain-independent and widely adopted ontologies, thereby facilitating interoperability with metadata from spatial and nonspatial domains on open data platforms.

## 2. Materials and Methods

In this section, we address a key objective of our study, i.e., eliciting user and producer views on geospatial data quality and fitness-for-use. In the requirements elicitation phase of our study, we conducted a total of 28 semi-structured interviews with spatial data producers and users from a variety of domains and applications. Following the semi-structured interviews, we conducted an online survey to collect data from larger and more diverse GIS communities. The survey aimed to examine findings from the interviews and to potentially identify additional quality themes and concepts (from the comments and suggestions) that are influential in assessing and evaluating spatial data quality and fitness-for-use.

### 2.1. Semi-Structured Interviews with Spatial Data Producers and Users

We selected participants from a broad spectrum of industries, in order to capture a wide variety of requirements for presenting and communicating geospatial metadata and determine the way in which users with different levels of expertise and competence in the geospatial domain, interpret spatial metadata. The objective of these interviews was to identify the ways in which geospatial data are produced, assessed and evaluated for various uses and purposes. Discussion was guided by a set of questions, used as high-level lines of inquiry. Interviews with geospatial data users aimed to elicit high-level user requirements for assessing and evaluating fitness-for-use of spatial data sources (see User Interviews Template). Interviews with geospatial data producers aimed to discover how producers of geospatial data describe the quality of their data sources (see Producer Interviews Template). We conducted interviews with local councils, utility companies (e.g., Melbourne Water), government departments and social services (e.g., Australian Bureau of Statistics, Department of Transport and Main Roads, Department of Premier and Cabinet), organizations providing architecture and planning services, the police service, organizations working for local governments (these organizations are not government departments or agencies), the health sector, the mining sector, software and information technology companies, organizations providing environmental services, insurance companies, engineering, consulting and construction services, and small businesses. Participants came from a diverse range of backgrounds, with 46% having worked with geospatial data for two to nine years. The majority of our interviews were conducted via a teleconference call, while a few were conducted as face-to-face interviews. In most interviews, one to two subject matter experts from the interviewed organization were present. The majority of the interviews took about an hour, with a few exceptions where interviews ran well over two hours.

Our semi-structured interviews were recorded, and audio recordings were transcribed. The transcripts were analyzed to identify high-level requirements for assessing fitness-for-use of spatial datasets. Analyses focused on identifying quality facets and deriving detailed user requirements that relate specifically to quality assessment of data sources for making data source selection decisions. As the collected data was qualitative, we analyzed the transcripts using *thematic data analysis*, "a method for identifying, analyzing, and reporting patterns (themes) within data" [33,34], to identify patterns of meaning across the study data.

After reviewing the transcripts several times, we generated a first cut code book. Over 45 codes were identified at this stage. Next, we did several rounds of consolidation and data reduction in our



coding by comparing text chunks and definitions for similarity. This iterative process of inductive coding resulted in identifying 20 codes, including 11 codes related to ‘quality dimensions’ and nine codes related to ‘fitness-for-use requirements.’ We then compared our inductive coding with data quality elements that emerged from our literature review in order to ensure that we name each code with a label that best describes the code. Where suitable labels were found, they were used for naming the codes in the final coding. Moreover, throughout the analysis, we used notetaking and memoing to refine our ideas [34], which supported categorization of the codes. The categorization of the codes was done by considering the similarity and differences of their underlying ideas. This resulted in grouping three codes including positional/spatial accuracy, attribute/thematic accuracy, and temporal accuracy as the three quality sub-dimensions of the accuracy quality dimension. However, this last step did not result in any changes to the codes representing fitness-for-use requirements.

## 2.2. Data Quality Survey

Following the semi-structured interviews with geospatial data producers and users, we conducted a survey to collect data from more diverse GIS communities (see Data Quality Survey). The survey was designed to examine findings from our interviews and to potentially identify additional quality themes that are influential in selecting data sources that are fit-for-use. The online questionnaire was designed based on the results of the analyses of our interviews with industry participants. The questionnaire comprised 32 questions:

- Questions 1–10 captured general information about the participants (e.g., identifying participants as spatial data users, producers or both; identifying participants’ level of expertise in geospatial data and metadata);
- Questions 11–17 and 26–28 aimed to gather participants’ views on geospatial data quality elements and sub-elements (outlined in Table 1);
- Questions 18–25 and 29 aimed to gather participants’ views on requirements for assessing fitness-for-use of datasets (outlined in Table 2);
- Questions 30 and 31 aimed to elicit participants’ perspectives on the usefulness of a geospatial user-centric metadata vocabulary for communicating dataset quality and fitness-for-use;
- Question 32 captured comments and suggestions on a geospatial user-centric metadata vocabulary, information that it should communicate and its potential role in enabling users to identify datasets that are fit for their intended purposes.

A seven-point Likert scale, ranging from “very unimportant” to “very important” was used to measure the importance of data quality elements and requirements for assessing fitness-for-use of data sources from the point of view of participants. The questionnaire was reviewed by several industry experts, resulting in minor updates and refinements. The project description and a link to the survey questionnaire were emailed to participants from a broad spectrum of industries, in order to ensure that survey results represented a variety of use cases and requirements from a wide range of users. Snowball sampling was also used to reach further potential participants, in addition to several reminders, ensuring that we collected as many responses as possible. In total, we received 15 responses; i.e., one response from each of the 15 distinct organizations that participated in the survey. Like the organizations we interviewed, these organizations predominantly use geospatial data in their area of work in Australia and New Zealand, hence the limited number of survey participants.

**Table 1.** Data quality elements, sub-elements and representative direct quotes from participants.

Data Quality Elements	Data Quality Sub-Elements	Direct Quote from Participant
<b>Accuracy</b> A measure of difference between the produced spatial data and the real world that it represents. It is a relative measure and often depends on some defined specification of a true value. Accuracy of data could be measured in terms of horizontal and vertical accuracy of captured data, correctness of object classifications (e.g., a road should not be misclassified as a river), and time stamp applied to the entities in the dataset.	<b>Positional/Spatial Accuracy</b> The difference between the recorded location of a feature in a spatial database or in a map and its actual location on the ground, or its location on a source of known higher accuracy. Positional accuracy can be refined to horizontal and vertical accuracy as it applies to horizontal and vertical positions of captured data.	“That level of quality and positional accuracy, can make a world of difference. So, like you were saying, when it comes to flood data, the positional accuracy of that flood study is more important than my household level geocode position. Which, if it’s give or take, one or two metres out of position, is less important, to me, than that, you know, to need a perfect flood study.”
	<b>Attribute/Thematic Accuracy</b> Denotes the correctness of object “classifications” and the level of precision of attribute “descriptions” in the produced data. For instance, a line in a dataset that denotes a river can be misclassified as a road; or a farm object can have the farmer or crops descriptions missing from it.	“As I said, one data set could be used for any, there might be fifty to a hundred columns of attributes within the data, but, when the individual components and the attributes of the data set are used by different people in different parts, it’s just not very good ... ”
	<b>Temporal Accuracy</b> Indicates the time stamp applied to the entities in the dataset. It is the difference between encoded dataset values and the true temporal values of the measured entities. It only applies when the dataset has a temporal (time) dimension in the form of [x, y, z, t]. This type of accuracy is identical with the concept <i>accuracy of a time measurement</i> .	“So, the quality, we have to have a good understanding as I said, of those input data sets and how they have changed and developed over time so that we can spot errors in those data sets, so they come through, so we are not building mesh block where we shouldn’t be building mesh blocks.”
<b>Completeness</b> Measures the omission error in the data and its compliance with data capture specification, dataset coverage, and at the level of currency required by the update policy. Highly generalised data can be accepted as complete if it complies with its specification of coverage, classification and verification.		“Yes, completeness, when I think of completeness I think about, well yes, I know that [name of the organisation removed] doesn’t have every address in the country that is actively used... There are many locations associated with the address and the data has to provide a type for each of the different locations associated with an individual address. So that it can be implemented appropriately for the business use.”
<b>Logical Consistency</b> Consistency as a general term is defined as the absence of conflicts or contradictions in a dataset. Logical Consistency relates to structures and attributes of geospatial data and defines compatibility between dataset objects – e.g., variables used adhere to the appropriate limits or types.		“... sometimes there’s no consistency between different producers on how that metadata is produced. That’s one thing, but then in terms of the attributes, the consistency that I was referring to was, in the example, was that, how it was actually, the definition that defined it, there may not necessarily be consistency there and that needs to be understood. For example, the first subclass that I was mentioning was ground water, surface water, or might be a meteorological station”

Table 1. Cont.

Data Quality Elements	Data Quality Sub-Elements	Direct Quote from Participant
<b>Relevancy</b> Relevancy (perceived relevancy) of a specific dataset to a user's intended uses and business purposes.		<p>"So, um, it's not that those points are wrong because they are correctly centroids of cadastral parcels, it's just, um, an alignment issue between our definition of the coast which is derived from GSIS Australia and the national cadastral. I think there is the succinct story, or you could provide, we could provide information on which points they were, and those points could be tagged, and people could deal with them in the appropriate way for their business use."</p>
<b>Currency</b> Currency is also known as timeliness (up-to-datedness) of data. Currency of data set differs from temporal accuracy, which relates to the time stamp applied to entities in a dataset.		<p>"[a dataset] that is updated weekly, so that we can have confidence that we have the most current representation of parcel information of title and ownership of parcels of land."  " So, you know, updating your addresses, updating data sets, it's not just the accuracy, it's also the currency."</p>
<b>Reliability</b> The extent to which a user perceives a data set to be trustworthy. Factors such as reputation and credibility of producer contribute to the user's perceived reliability of the dataset. Producer profile (if exists) can contribute to communicating reputation of producer and to the overall reliability of the dataset. The producer's identity alludes to how users perceive trustworthiness of a dataset.		<p>"... there is a degree of trust and knowledge that the data is fit for purpose and we sort of iterated through various different issues with the data and solved those issues as we have gone along. We don't, to be honest, we don't analyse the metadata that we get from [name of the organisation removed] because those questions are raised in a, in the quarterly meetings. I suppose, what is important is a knowledge that this is the best data that is available and then a good understanding of the actual limitations of that data."</p>
<b>Lineage (provenance information)</b> historical information such as the source, producer, content, methods of collecting and processing, and geographic coverage of the data product.		<p>"So, the metadata that is provided for those contains a lot of history, around where the data set originated from, what kind of sources were used to initially create it, and how it's updated, so that, I guess the history information is interesting to know how the data set came into being... The information around how the data set is maintained and updated is important, the frequency with which it's updated and knowing how the other authorities' information is fed into their process and then into their database."</p>
<b>Cost</b> Financial cost of a dataset for a user, considering their own financial circumstances (e.g., a user is able to and willing to pay more for a dataset, which better suits their intended purposes).		<p>"I guess there is more choice in say, imagery or Lidar, but that's more of a cost issue and a licencing issue and an ability to cost share with other authorities to obtain that."</p>

**Table 2.** Requirements for assessing fitness-for-use and representative direct quotes from participants.

Requirements for Assessing Fitness-For-Use of a Dataset	Direct Quote from Participants
<b>Producer profile:</b> the producer's profile can present information on reputation of dataset producer/provider. The information could contribute to the user's perceived reliability of dataset. Users tend to rely on spatial data from producers who they know.	"Yeah, so, definitely that information around the domain that they are working in and a small number of classifications of their abilities. So, intermediate or advanced or ... would be beneficial"
<b>Dataset citation information:</b> Some publications and journal articles report data quality checks, dataset use and evaluation which are useful for assessing quality of a dataset.	"... It definitely would be useful to know, that if it was actually used in publications and what not ... Because, if it's used and people say, oh that's accurate, well, how is that known? So, in some ways, if there is that, you know, validation by journals, that actually can become quite useful."
<b>Data dictionary:</b> information on every field, allowed values, types, formats, etc.	"So, we wouldn't even start looking at it (at a dataset) ... We'd probably dive straight into the data dictionary." "The data dictionary that is provided for the data set goes a long way towards enabling someone to understand how to use it for their purposes."
<b>Quantitative quality information:</b> providing a numeric quantification of some data quality aspects by creating a specification for the dataset or comparing it with other accepted reference sources (e.g., external vocabularies such as UncertML present statistical/quantitative definition of uncertainty). This quantitative quality information can cover information about spatial and temporal resolution; spatial and temporal scale; geometric correctness; horizontal, vertical and absolute accuracy; precision; error estimates; and uncertainty.	"Yes, and estimates around the, so, estimates of the accuracy in terms of percentages for new data that is added and estimates on the allowable errors for the historical information which is mostly digitised." "[We] will be driven by a process that will allow people to be able to, have some, sort of, standardisation in quantifying the quality and fitness for purpose and use of data."
<b>Soft knowledge:</b> Producer's comments (textual statements) that could help to evaluate fitness-for-use of a data product, such as comments on the overall quality of a data product, any known data errors, and potential use. This information could be updated periodically by the producer.	"The metadata statement is fairly complex to use, and I think trying to provide a more user-friendly description of those products and services, is exactly where producers need to go."
<b>Compliance with standards:</b> Dataset's compliance with national (if any) and international standards such as ISO 19157:2013, ISO 19115-1:2014, ISO 19115-2:2009, and Dublin Core.	"[Many data producers] conform with OGC and ISO standards. [However] it would be fair to say that for anybody who is trying to read an ISO compliant metadata statement, related to a data set, is not only just confusing, but really doesn't, you know, you can get lost in that."
<b>User ratings of the dataset (as a part of peer reviews and feedback):</b> quality ratings in the form of quality stars (e.g., four out of five quality stars) or any similar form of rating that conveys a quick visual feedback on overall quality of a dataset. Such rating is different from feedback and advice (from users and producer of a dataset) that is in the form of textual statements and can express more in-depth feedback on quality of the dataset.	"In a way, I think it (a rating system) would [be] beneficial. You may have a rating about quality." "... allowing data custodians access to a template of the processes, to be able to describe, and rate the quality and fitness for purpose of datasets that are being populated into that, and that's web services as well. So, it's an emerging space."
<b>Community recommendations and advice (as a part of peer reviews and feedback):</b> textual or verbal feedback from community of users on the quality of a dataset and advice on fitness-for-use of the dataset. It could also provide the underlying rationale for a rating (e.g., quality star rating) of a dataset. The interactions (e.g., brief Q&A and discussions) among the users could be via an online interactive tool (e.g., a discussion forum) that is specifically designed for this purpose or via other means of communication such as email and face-to-face meetings.	"We need to put some structure encoding around this so, as you say, it is queryable and people can make better use of people's understanding." "Allowing people to enter limitations that they have encountered would be, I could see the benefits of that, to kind of, generating a feedback to the supplier and capturing people's experience."
<b>Independent expert reviews:</b> expert value judgments from other organizations or businesses who are not the producer and user of a specific dataset, but have expert knowledge that could provide value judgments on the general quality, errors, domain of application of a specific dataset, etc.	"Most of the data sources that I get are from government agencies. So, there is already inherent, I guess, the assumption, that are of a certain credibility and alike. But that being said, I also use engineering drawings and get drawings from engineers..."

### 3. Results

#### 3.1. Qualitative Analysis of the Interviews

Our analysis identified spatial data quality elements and sub-elements used or required by the participants for assessing fitness-for-use of spatial data sources. Table 1 presents these data quality elements and sub-elements, their definitions and representative direct quotes. In addition, we identified quality themes and informational aspects of geospatial data sources, deemed useful and important for evaluating fitness-for-use; these quality themes and informational aspects are referred to as “fitness-for-use requirements” throughout this paper. Table 2 presents these requirements, their definitions and representative direct quotes from participants.

Tables 3 and 4 outline frequency counts of data quality elements and fitness-for-use requirements, identified from our interviews; i.e., the number of participants that identified each quality element and fitness-for-use requirement, as influential in assessing and evaluating fitness-for-use of spatial data. We note that these numbers could have been higher, had we employed a vote counting technique during the interviews. However, we avoided this technique, as the interviews were conducted as an exploratory, rather than confirmatory method of data collection.

**Table 3.** Frequency count of data quality elements from the interviews.

Data Quality Sub-Element or Sub-Element	Frequency
Positional/Spatial Accuracy	5
Attribute/Thematic Accuracy	4
Temporal Accuracy	4
Completeness	3
Logical Consistency	4
Relevancy	5
Currency	4
Reliability	5
Lineage (provenance information)	5
Cost	3

**Table 4.** Frequency count of fitness-for-use requirements from the interviews.

Requirement for Assessing Fitness-for-Use	Frequency
Producer profile (reputation of the producer)	4
Dataset citation information	4
Data dictionary	5
Quantitative quality information	5
Soft knowledge	6
Compliance with standards	3
User ratings of the dataset (as a part of peer reviews and feedback)	5
Community recommendations and advice (as a part of peer reviews and feedback)	5
Independent expert reviews	4

Furthermore, participants in almost all interviews indicated that while it is important to identify informational aspects of spatial data sources that are influential in assessing fitness-for-use, it is also important to enable users to discover spatial data on open data portals, using these informational



aspects as search criteria. In order to facilitate search and discovery of spatial data, spatial metadata should be released to the open data portals and be interoperable with metadata from nonspatial domains. Finally, we also note that participants commented on the importance of “ease of access”, “licensing”, “resolution” and “online forums”. These themes have not been included in our analysis, as they encompass quality elements and fitness-for-use requirements outlined in Tables 1 and 2; e.g., “user ratings” and “community recommendations and advice” outlined as fitness-for-use requirements in Table 2, are more specific requirements of the more broader theme, “online forums”.

### 3.2. Analysis of Survey Results

We analyzed data gathered from the survey using SPSS Statistics 23 software [35]. Analysis of responses to the first ten questions demonstrated that:

- 80% of participants identified themselves as both user and producer of spatial data;
- Most participants work in either “agriculture, forestry, and fishing” (26%) or “other services” (33%). We used the Australian and New Zealand Standard Industrial Classification (<https://www.abs.gov.au/ausstats/abs@.nsf/0/20C5B5A4F46DF95BCA25711F00146D75?opendocument>) to identify the industry represented by each participant;
- 93% of participants use data from external data providers;
- 86% of participants have a range of data sources to choose from;
- 46% of participants have worked with geospatial data for two to nine years;
- 93% of participants make data source selection decisions based on prior knowledge and experience;
- 53% of participants find selecting datasets that fit their needs a challenging task;
- 80% of participants consider metadata records or other supporting information when selecting data sources;
- 53% of participants believe that up to 25% manual effort is involved in understanding fitness-for-use of data sources;
- 6.7% of participants believe that metadata describing dataset quality do not follow any standards; 26.7% believe that such metadata are not provided; 33.3% believe that metadata that describe data quality are incomplete; and 33.3% believe that this metadata follow widely adopted standards;
- 53% of participants indicated that their organization has been impacted by not understanding (or misunderstanding) fitness-for-use of a dataset at least once.

Tables 5 and 6 represent the results of descriptive statistics for the importance of data quality elements and fitness-for-use requirements, respectively. As outlined in Table 5, the mean of responses to each of the ten data quality elements is higher than 5, indicating the importance of data quality elements in assessing fitness-for-use of data sources. More specifically, “Logical Consistency” and “Lineage/Provenance” have the lowest mean (5.20), while “Relevancy” has the highest mean (6.47), making it the most important quality element from participants’ perspective. Furthermore, as outlined in Table 6, “Quantitative Quality Information” has the highest mean (5.60), indicating its importance in assessing fitness-for-use of data sources, while “Dataset Citations” (mean: 3.13) and “User Ratings” (mean: 3.67) have a mean value below the neutral point, indicating their lesser importance compared with other requirements. In addition, analyses indicate that participants scored data quality elements higher than fitness-for-use requirements for identifying data sources that are best suited to their specific purposes.

Furthermore, descriptive analysis of questions 30 and 31 demonstrated that: (i) 86% of participants believe that a geospatial user-centric metadata vocabulary will be useful in making informed data source selection decisions and (ii) 93% of participants believe that describing metadata at various levels of granularity, i.e., describing metadata for a dataset, its feature types and attribute types, will be valuable in assessing fitness-for-use of the dataset.

**Table 5.** Descriptive statistics for data quality elements from the questionnaire.

	N	Range	Minimum	Maximum	Mean	Std. Error	Std. Deviation	Variance
Positional/Spatial Accuracy	15	4	3	7	5.73	0.345	1.335	1.781
Attribute/Thematic Accuracy	15	5	2	7	5.80	0.327	1.265	1.600
Temporal Accuracy	15	3	4	7	5.93	0.267	1.033	1.067
Logical Consistency	15	5	2	7	5.20	0.355	1.373	1.886
Completeness	15	5	2	7	5.40	0.412	1.595	2.543
Currency (timeliness)	15	3	4	7	6.13	0.256	0.990	0.981
Lineage/Provenance	15	5	2	7	5.20	0.380	1.474	2.171
Cost of Quality (Financial)	15	3	4	7	5.40	0.254	0.986	0.971
Overall Reliability of Data	15	3	4	7	6.07	0.228	0.884	0.781
Relevancy	15	3	4	7	6.47	0.236	0.915	0.838
Valid N (listwise)	15							

**Table 6.** Descriptive statistics for fitness-for-use requirements from the questionnaire.

	N	Range	Minimum	Maximum	Mean	Std. Error	Std. Deviation	Variance
Experts' Review	15	5	2	7	4.67	0.361	1.397	1.952
Compliance with Standards	15	5	1	6	4.07	0.431	1.668	2.781
Community Advice and Recommendations (User Feedback)	15	4	2	6	4.27	0.300	1.163	1.352
Producer Profile (Reputation)	15	5	2	7	5.27	0.358	1.387	1.924
Dataset Citations	15	6	1	7	3.13	0.533	2.066	4.267
Quantitative Quality Information	15	5	2	7	5.60	0.349	1.352	1.829
Soft Knowledge	15	5	2	7	4.80	0.312	1.207	1.457
User Ratings	15	5	1	6	3.67	0.347	1.345	1.810
Data Dictionary	15	5	2	7	5.53	0.350	1.356	1.838
Valid N (listwise)	15							

Finally, outliers were identified for data quality elements and fitness-for-use requirements. Outliers for data quality elements are:

- Attribute/thematic accuracy: only one participant in the financial and insurance services domain, scored this quality element low (2);
- Logical consistency: only one participant in the professional, scientific, and technical services domain, scored this quality element low (2);
- Completeness: only one participant in the professional, scientific, and technical services, scored this quality element low (2);
- Lineage/Provenance: only one participant from the “other services” domain, scored this quality element low (2);

Outliers for fitness-for-use requirements are as follows:

- Compliance with international standards: two participants from the “other services” domain, scored this requirement low (2);
- Community advice and recommendations (user feedback): only one participant from the “other services” domain, scored this requirement low (2);

- Reputation of dataset provider (producer profile): two participants (one from the “administrative and support services” domain, and the other from the “other services” domain) scored this requirement very low (1);
- Quantitative quality information: only one participant from the “other services” domain, scored this requirement low (2);
- Overall reliability: only one participant from the “other services” domain, scored this requirement 4. All other responses scored this requirement greater than 4;
- Relevancy: only one participant from the “professional, scientific, and technical services” domain, scored this requirement 4. All other responses scored this requirement greater than 4;
- Data dictionary (description of a dataset and its components; i.e., feature types and attribute types and their relationships): only one participant from the “administrative and support services domain”, scored this requirement low (2).

### 3.3. Geospatial User Centric Metadata Ontology

This section describes the *Geospatial User Centric Metadata* (GUCM) ontology. The requirements that underpin the design of the GUCM ontology have emerged from the analyses presented in Sections 3.1 and 3.2. Our industry engagements confirmed the requirement for a vocabulary that communicates geospatial data quality and fitness-for-use. In particular, analyses of our engagements revealed that 93% of the participants use data from external data providers and make data source selection decisions based on prior knowledge and experience. More than half of the participants (53%) indicated that selecting data sources that are fit-for-use is a challenging task and believe that up to 25% manual effort is involved in understanding fitness-for-use of data sources. In addition, 80% of the participants indicated that they consider metadata records or other supporting information for selecting datasets; however, metadata records are usually missing or incomplete. Furthermore, 86% of the participants stated that a vocabulary, which communicates spatial data quality and fitness-for-use will enable them to select data sources that are best suited to their application and domain.

The GUCM data model makes extensive reuse of existing standards and vocabularies, and where necessary extends these vocabularies in order to model information that emerged from our industry engagements for eliciting user views on geospatial data quality and fitness-for-use. GUCM is built using Protégé, a free, open-source ontology editor used to build knowledge-based solutions in areas as diverse as biomedicine, e-commerce, and organizational modelling [36]. More specifically, we used Protégé version 5.5.0, which provides full support for the OWL 2 Web Ontology Language (<https://www.w3.org/TR/owl-overview/>), and direct in-memory connections to description logic reasoners like HermiT and Pellet (<http://owl.cs.manchester.ac.uk/tools/list-of-reasoners/>), which were used to validate the ontology throughout the development phase.

From a conceptual point of view, the GUCM ontology (<https://csrcsprojres.s3-ap-southeast-2.amazonaws.com/gucmetadata-v1.owl>) comprises three main components: (i) Dataset Schema, (ii) Interoperable Metadata, and (iii) User Feedback.

The ontology can be accessed from the ESIP Community Ontology Repository (<http://cor.esipfed.org/ont?iri=http://reference.data.gov.au/def/ont/gucmetadata>). The following sections depict each component using the Ontology Unified Modeling Language (UML) profile; i.e., a formal specification of UML profile for RDF Schema (RDFS) and Web Ontology Language (OWL) [37]. OWL classes are depicted as UML classes, data properties as class attributes, object properties as association roles, individuals as objects and cardinality restrictions on association domain class as UML cardinalities.

Our analyses discovered that most participants (80%) are both users and producers of geospatial data; thus, GUCM enables both spatial data producers and users to describe and reference metadata. However, in order to ensure the integrity and trustworthiness of the described metadata, GUCM defines the producer and user of metadata represented by each of its components (please see “Metadata Producer” and “Metadata Consumer” information outlined for each of the components of GUCM in Sections 3.3.1–3.3.3).

The Dataset Schema component of the GUCM ontology captures geospatial metadata using vocabularies specific to the geospatial domain. Spatial metadata captured by this component can target a dataset or any of its components; e.g., a feature type or an attribute type. In other words, spatial metadata can be described for a dataset at various levels of granularity.

Public sector information has been increasingly released as open data in recent years. Governments across the world realize the potential benefits of releasing their information as open data. These benefits include more transparency, increase in governmental efficiency and effectiveness, and societal and economic benefits. The benefits of making datasets available as open data also apply to the private sector. Geographical data, such as topographical maps and the underlying Earth observation data, are top listed for release as open government data due to the high demand from users. In line with this initiative, the Interoperable Metadata component of the GUCM ontology represents spatial metadata using domain-independent and widely-adopted ontologies and vocabularies, in order to facilitate interoperability between spatial and nonspatial metadata on open data platforms.

The User Feedback component of the GUCM ontology aims to convey feedback, fitness-for-use and usage experiences of geospatial datasets to spatial data producers and users. As with the Dataset Schema component, the User Feedback component of GUCM enables feedback, fitness-for-use and dataset usage descriptions to be associated with a dataset or any of its constituents; e.g., a feature type or an attribute type. In addition, as is the case with the Interoperable Metadata component, user feedback is captured and represented using domain-independent and widely adopted vocabularies, thereby facilitating interoperability between metadata captured by this component and metadata on open data platforms.

### 3.3.1. Dataset Schema

Our analyses highlighted the significance of spatial data quality elements (Table 1) and importance of the description of a dataset and its components; i.e., feature types and attribute types and their relationships (Data Dictionary in Table 2). In addition, as outlined in Section 3.2, 93% of the participants indicated that describing metadata at various levels of granularity, i.e., describing metadata for a dataset, its feature types and attribute types, would be valuable in assessing fitness-for-use of the dataset. Therefore, the Dataset Schema component of the GUCM ontology aims to provide a full description of the contents and structure of a geographic dataset, using the application schema developed in compliance with the ISO 19109:2015 standard (<https://www.iso.org/standard/59193.html>). More specifically, this component uses the General Feature Model (GFM) from ISO 19109:2015 Geographic information—Rules for application schema. The General Feature Model is a metamodel for definition of features. It defines the meaning of the feature types and their associated feature attributes, feature operations, and feature associations contained in the application schema. The contents and structure of a dataset, defined in accordance with the General Feature Model, are used by the Dataset Schema component to describe and associate dataset metadata with specific components of a dataset, i.e., dataset, feature type or attribute type. Metadata is described in accordance with ISO 19115-1:2014 and ISO 19157:2013 standards, which are specific to the geospatial domain.

Figure 3 depicts the Dataset Schema component of the GUCM ontology. As depicted in Figure 3, the Dataset Schema component of GUCM can be conceptually divided into two main parts: (i) Part 1 that models the structure of a geographic dataset, and (ii) Part 2 that defines and associates metadata with the dataset, its feature types and attribute types. Both parts are discussed below.

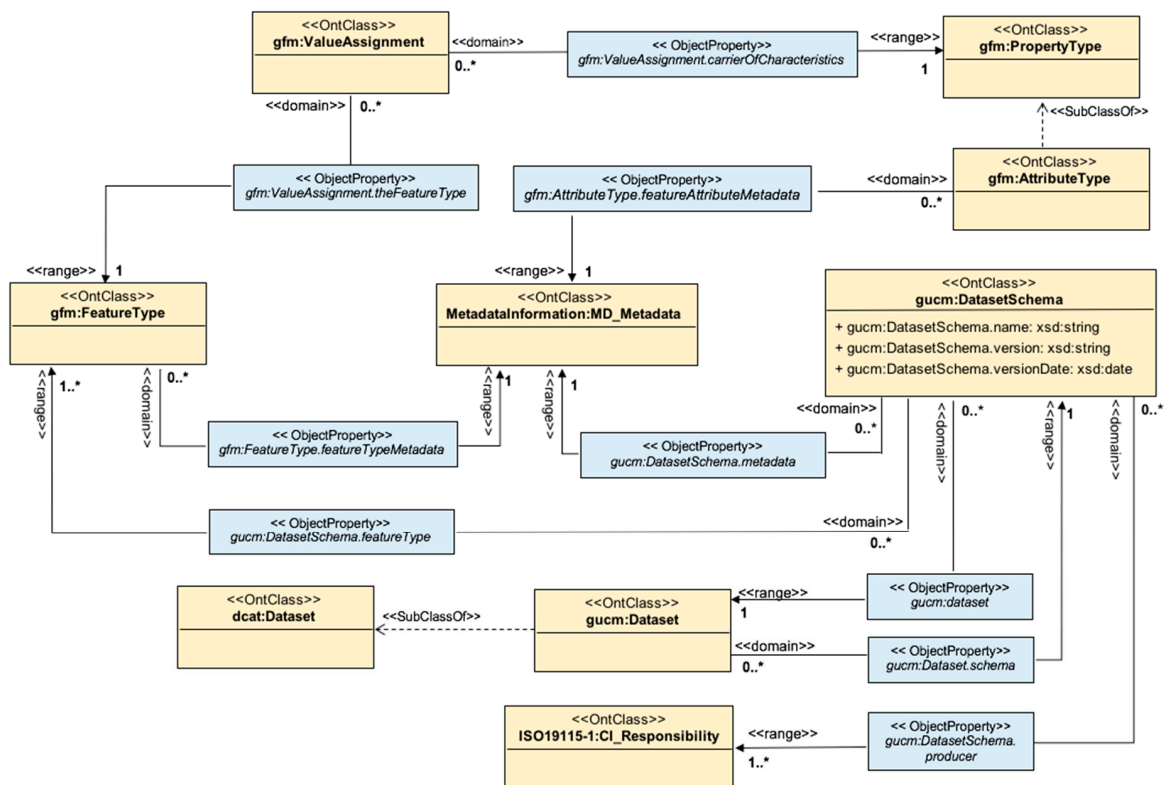


Figure 3. The Dataset Schema component of the GUCM ontology.

The central concept of Part 1 is **gucm:DatasetSchema**, which describes (**gucm: dataset**) the structure of a geospatial dataset (modelled by **gucm:Dataset**). A **gucm:DatasetSchema** defines (**gucm:DatasetSchema.featureType**) the dataset's feature types (modelled by **gfm:FeatureType**). An attribute type (modelled by **gfm:AttributeType**) is assigned (**gfm:ValueAssignment.carrierOfCharacteristics**) to a feature type (modelled by **gfm:FeatureType**) through a value assignment (modelled by **gfm:ValueAssignment**).

Part 2 of the Dataset Schema component of GUCM facilitates metadata description at various levels of granularity; i.e., the dataset schema (modelled by **gucm:DatasetSchema**) describes (**gucm:DatasetSchema.metadata**) metadata (modelled by **MetadataInformation:MD\_Metadata**) for the underlying dataset (modelled by **gucm:Dataset**). Metadata can also be described (**gfm:FeatureType.featureTypeMetadata**) for a feature type (modelled by **gfm:FeatureType**), or for (**gfm:AttributeType.featureAttributeMetadata**) an attribute type (modelled by **gfm:AttributeType**). The producer (**gucm:DatasetSchema.producer**) of a **gucm:DatasetSchema** is represented by **ISO19115-1:CI\_Responsibility**. See examples of metadata described by the Dataset Schema Component.

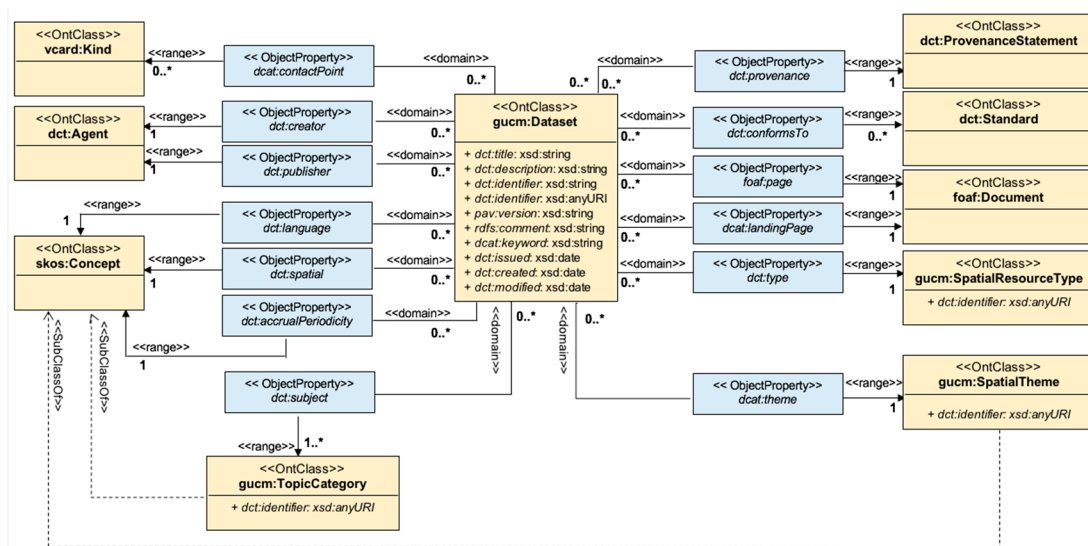
**Metadata Producer:** The producer-centric metadata represented by this component are described and maintained by spatial data producers for reference by spatial data users.

**Metadata Consumer:** Spatial data producers and users are consumers of metadata generated by this component. Users can reference metadata represented by this component to gain an understanding of a dataset and its structure; i.e., details of feature types and attribute types, their relationships and links. In addition, users can reference metadata defined for a dataset, its feature types and attribute types. Furthermore, metadata represented by this component can be used as reference for describing metadata through the User Feedback component of GUCM; i.e., users can associate their experiences, issues, limitations and suggestions, with a dataset or any of its constituents, as defined by this component; e.g., a limitation can be associated with a dataset, a feature type or an attribute type.

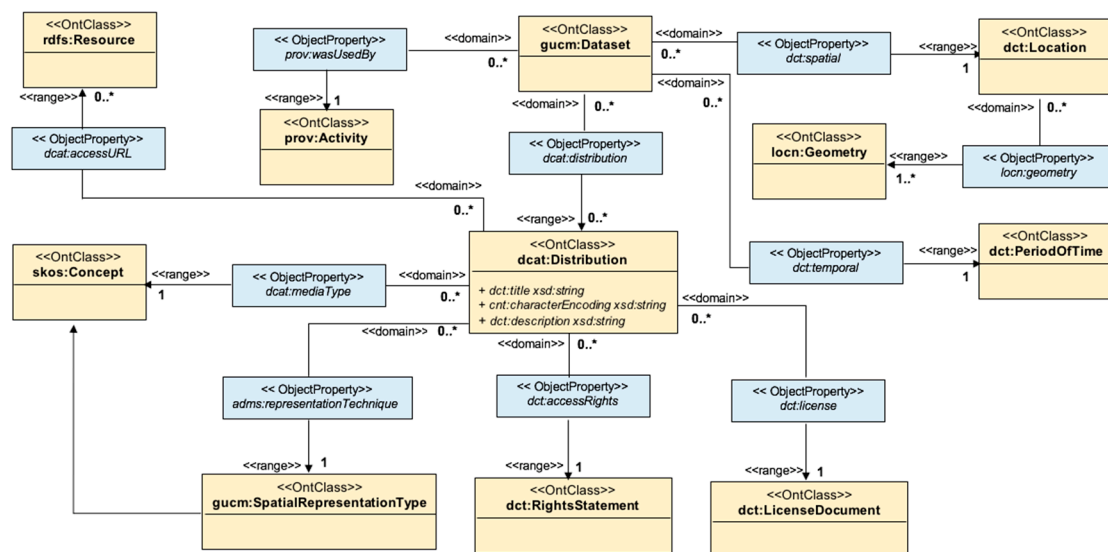


### 3.3.2. Interoperable Metadata

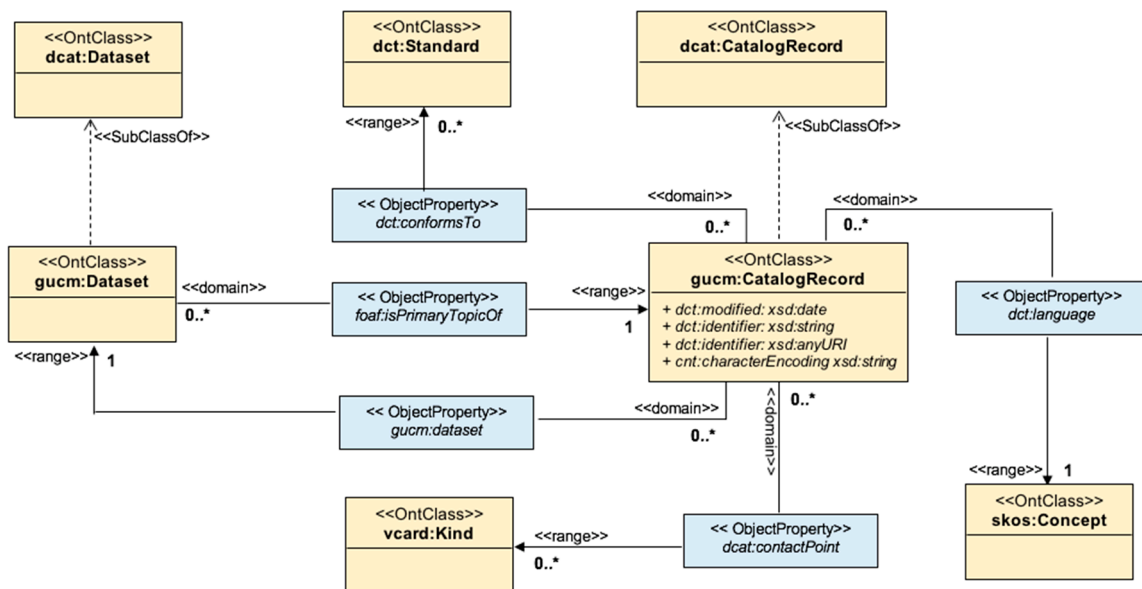
Analyses of our semi-structured interviews with producers and users of geospatial data highlighted the requirement for dataset search and discovery (please see Section 3.1). To this end, the Interoperable Metadata component of GUCM aims to facilitate interoperability between metadata descriptions from spatial and nonspatial domains, in order to make spatial datasets, dataset series, and services searchable on general data portals. The Dataset Schema component of GUCM facilitates metadata description using standards that are specific to the geographic domain; therefore, metadata descriptions generated by the Dataset Schema component cannot be exchanged with metadata descriptions from nonspatial domains. In order to facilitate interoperability between spatial metadata captured by the Dataset Schema component and metadata from other domains on open data platforms, the Interoperable Metadata component of the GUCM ontology presents metadata captured by the Dataset Schema component, using concepts from domain-independent and widely-adopted ontologies, based on the GeoDCAT-AP version 1.0.1 specification (<https://joinup.ec.europa.eu/release/geodcat-ap/v101>). GeoDCAT-AP is an extension of the DCAT application profile (DCAT-AP (<https://joinup.ec.europa.eu/solution/dcat-application-profile-data-portals-europe>)) for data portals in Europe for describing geospatial datasets, dataset series, and services. DCAT-AP provides a metadata profile that aims to provide an interchange format for data portals operated by EU Member States. DCAT-AP complies with and is based on the W3C Data Catalog (DCAT (<https://www.w3.org/TR/vocab-dcat/>)) vocabulary. GeoDCAT-AP also offers a syntax binding in RDF (<https://www.w3.org/RDF/>) for the union of metadata dimensions of the core profile of ISO 19115:2003 and metadata elements defined in the framework of the INSPIRE Directive [30]. Figures 4–6 depict metadata for describing datasets in accordance with the GeoDCAT-AP version 1.0.1 specification.



**Figure 4.** Interoperable Metadata—models responsible party and metadata point of contact, resource type, resource locator, resource language and metadata language, topic category, lineage, metadata standard name, metadata standard version, and maintenance information.



**Figure 5.** Interoperable Metadata—models temporal reference and metadata date, dataset access conditions and limitations, encoding and distribution format, spatial representation type, spatial coverage, conformity, and data quality.



**Figure 6.** Interoperable Metadata—models metadata on metadata (through dcat:CatalogRecord), metadata standard name, metadata standard version, metadata language, responsible party, and metadata point of contact.

The Interoperable Metadata component implements GeoDCAT-AP Core and GeoDCAT-AP Extended. GeoDCAT-AP Extended is a superset of GeoDCAT-AP Core. GeoDCAT-AP Core includes bindings for metadata elements of the INSPIRE metadata and metadata elements in the core profile of ISO 19115:2003 core for which DCAT-AP provides an RDF syntax binding. Those metadata elements for which DCAT-AP does not provide a binding are part of the GeoDCAT-AP Extended profile. GeoDCAT-AP Core is meant to enable the harvesting and re-use of spatial metadata records through DCAT-AP-conformant applications and services, including data portals and APIs. The alignments for INSPIRE and ISO 19115:2003 metadata elements that are not included in GeoDCAT-AP Core are defined in GeoDCAT-AP Extended. In addition to this, GeoDCAT-AP Core does not provide alignments from

metadata records concerning services, with the only exception of catalogue or discovery services, which are the only ones supported in DCAT-AP.

The GeoDCAT-AP specification is used to facilitate interoperability between metadata of spatial and nonspatial datasets in data portals in Europe. The bindings defined in GeoDCAT-AP for the RDF representation of INSPIRE metadata and the core profile of ISO 19115:2003 are based on widely adopted vocabularies, such as DCAT-AP. Therefore, this specification can also be used to facilitate the exchange of metadata descriptions for spatial datasets, dataset series, and services among data portals in other parts of the world. However, certain code lists recommended by the GeoDCAT-AP specification could be replaced to better suit the local context; for example, the INSPIRE spatial data themes (<http://inspire.ec.europa.eu/theme>) could be replaced with spatial data themes that may be more relevant to the local context; e.g., in the context of this study, we used themes defined by the Foundation Spatial Data Framework's Location Information Knowledge Platform (LINK (<https://link.fsdf.org.au/>)). See examples of metadata described by the Interoperable Metadata Component.

**Metadata Producer:** This component represents metadata described in the Dataset Schema component, using mappings identified by the GeoDCAT-AP version 1.0.1 specification. In order to prevent double data entry and data inconsistencies, an automated process should be used to add and update the value of elements for the Interoperable Metadata component, as corresponding elements in the Dataset Schema component are added and updated.

**Metadata Consumer:** Metadata represented by domain-independent and widely-adopted vocabularies, in accordance with the GeoDCAT-AP version 1.0.1 specification, can be published to the Linked Data Cloud and Open Data Portals [18], such as the Australian government's public data platform (<https://data.gov.au>) in the context of our research study. This will in turn facilitate interoperability between metadata from spatial and nonspatial domains (Figure 1). Therefore, spatial data users and producers, including users and producers of spatial data on Linked Data Cloud and open data portals, are consumers of metadata represented by this component.

### 3.3.3. User Feedback

Our analyses identified other quality themes and informational aspects of geospatial data sources, which users deemed relevant and important for evaluating fitness-for-use (Table 2). The User Feedback component of the GUCM ontology captures these fitness-for-use requirements, including metadata that represent users' *implicit knowledge* of spatial datasets; i.e., knowledge representing users' experiences with datasets in the context of specific application domains. These metadata encompass user feedback, experiences, comments, questions and answers, description of encountered problems, proposed solutions and publications describing those problems, dataset ratings and fitness-for-use descriptions of geospatial datasets. Metadata captured by the Dataset Schema component represent dataset characteristics and production workflows generated by the producer of a spatial dataset. Metadata captured by the User Feedback component complement metadata captured by the Dataset Schema component by presenting user feedback, comments and fitness-for-use descriptions for various applications of a dataset. In addition, the User Feedback component facilitates communication and discussion between producers, experts and users of geospatial data. Figure 2 illustrates an example of Q&A between producers and users of a dataset.

The User Feedback component of the GUCM ontology uses the Dataset Usage Vocabulary (DUV (<https://www.w3.org/TR/vocab-duv/>)) to describe user experiences, citations, and feedback about a dataset. The Data on the Web Best Practices Working Group (<https://www.w3.org/TR/dwbp/>), which is a W3C Recommendation for publishing and using data on the Web, recommends the DUV for citing published data, conveying feedback between users and producers, and defining descriptive metadata that provide insights to users about how published datasets can be used. The DUV is an extension of the Data Catalog (DCAT (<https://www.w3.org/TR/vocab-dcat/>)) vocabulary version 1.0, which provides metadata for citing, describing usage, and conveying feedback on published datasets and distributions. As a domain-independent and open vocabulary, the DUV encourages producers to add descriptive



Figure 7, a **gucm:UserFeedback** targets (oa:hasTarget) a dataset (modelled by **gucm:Dataset**), a Feature Type (modelled by **gfm:FeatureType**) or an Attribute Type (modelled by **gfm:AttributeType**, subclass of **gfm:PropertyType**).

A **gucm:UserFeedback** can be a reply to (**gucm:UserFeedback.isReplyTo**) another user feedback (modelled by **gucm:UserFeedback**). An issue discovered with a dataset (modelled by **gucm:DiscoveredIssue**), a type of **gucm:UserFeedback**, can be cited in (**biro:isReferencedBy**) a publication (modelled by **biro:BibliographicRecord**), categorised by (**gucm:DiscoveredIssue.aspectCode**) a code (modelled by **gucm:IssueAspectCode**). A **gucm:DiscoveredIssue** can identify (**gucm:DiscoveredIssue.alternativeResource**) an alternative resource to the dataset (modelled by **gucm:Dataset**). It can also identify (**gucm:DiscoveredIssue.fixedResource**) the fixed resource (modelled by **gucm:Dataset**); i.e., the resource in which the discovered issue is resolved. See examples of metadata described by the User Feedback component.

**Metadata Producer:** Metadata represented by this component are predominantly described by spatial data users; however, spatial data producers and experts can also contribute to the generation of metadata. As mentioned above, the DUV encourages producers to add descriptive metadata tailored to meet users' domain-specific needs. In the context of the User Feedback component of GUCM, a producer can provide an answer to a question raised by a user or describe metadata about a dataset or any of its components; e.g., description of possible values for an attribute of a feature type and their appropriate uses in specific contexts and applications. Figure 2 illustrates examples of metadata captured by the User Feedback component, while illustrating the contribution of producers and users to the generation of metadata.

**Metadata Consumer:** Spatial data producers and users are consumers of metadata generated by this component; users use these metadata to determine if a dataset is fit for their specific purposes, and to seek guidance from data producers by participating in Q&A discussions (Figure 2). Producers can use feedback and users' experiences with datasets, to improve their data products and meet users' domain-specific needs and requirements. In other words, user feedback generated by this component can be analyzed over time and used to refine and enhance metadata represented by the Dataset Schema component. This will in turn render the producer-supplied metadata more relevant to users' specific needs and requirements. Figure 2 illustrates an example of a user requesting information about possible values of an attribute. If this information is consistently demanded by the users, the Dataset Schema component can be updated to include possible values for attributes of feature types.

## 4. Discussion

The Geospatial User-Centric Metadata ontology (GUCM) presented in this paper aims to communicate quality and fitness-for-use of spatial data sources to users, in order to enable them to make informed data source selection decisions. More specifically, GUCM aims to bridge the gap between producer and user views on geospatial data quality and fitness-for-use. Our industry engagements aimed to identify this gap and elicit users' requirements for identifying datasets that are fit for their intended uses and purposes. The GUCM ontology is designed based on the requirements that emerged from analyses presented in Sections 3.1 and 3.2.

### 4.1. Contributions to Knowledge

As outlined in the Related Work section, previous studies and initiatives have tried to communicate spatial data quality by capturing users' feedback about a data source or providing users with visual cues of dataset quality and possibly relevance. The main features that clearly distinguish the GUCM ontology from previous efforts are as follows:

- *Capturing metadata in structured form*—as outlined in the introduction section, one of the objectives of this study is to make spatial data sources searchable on open data portals (O3). The GUCM ontology captures and represents metadata and fitness-for-use descriptions of spatial datasets, using concepts from domain-independent and widely adopted vocabularies and ontologies.



The structured metadata described and captured by the GUCM ontology can be published to Open Data Portals and the Web of Data [18], providing a means to search and discover spatial data based on metadata and fitness-for-use criteria, in addition to facilitating interoperability between spatial and nonspatial metadata on open data platforms.

- *Representing producer-supplied and user-described metadata using a single model*—as outlined in the introduction section, one of the objectives of this study (O2) is to enable both producers and users of geospatial data to describe metadata and fitness-for-use descriptions of datasets using a single model. Internal quality, modelled by the *Dataset Schema* component, and external quality, modelled by the User Feedback component, are captured and represented using the *same model*, rather than separate producer and user models. As mentioned in Section 3.3, in order to ensure the integrity and trustworthiness of metadata descriptions, the model differentiates between metadata that is solely created and maintained by producers (Dataset Schema), and metadata created by users, producers and experts (User Feedback).
- *Enabling metadata description at various levels of granularity*—one of the objectives of this study (O1) is to facilitate metadata and fitness-for-use descriptions for datasets and their components. The hierarchical structure of the ontology enables metadata and fitness-for-use descriptions to target various components of a dataset; i.e., dataset, feature type or attribute type. This in turn, facilitates dataset search and discovery based on metadata and usage descriptions for specific components of a dataset.
- *Facilitating communication and discussion between geospatial data producers and users*—the GUCM ontology enables producers and users of geospatial data to generate metadata and fitness-for-use descriptions using the same model (O2). This in turn facilitates communication and discussion between geospatial data producers and users. The User Feedback component of GUCM facilitates communication and discussion between spatial data users, producers and experts (Figure 2). In addition, metadata captured by the User Feedback component can be used to improve producer-supplied metadata (Dataset Schema) over time. This will render the producer-supplied metadata more relevant to users' specific needs and requirements.
- *Providing contextual information for metadata*—the GUCM ontology captures profiles of users that describe their experiences with spatial data sources and contribute to user-centric metadata by sharing their insights and implicit knowledge of data sources. In addition, the ontology captures the applications and domains within which metadata are described. The user profiles and application domain information associated with metadata can be used to put metadata and fitness-for-use descriptions in context when assessing the suitability of data sources for specific uses and purposes.

#### 4.2. Study Limitations

This study has also identified a number of limitations, as discussed below:

- We took great care to ensure that our industry engagements for eliciting user requirements represented a broad spectrum of industries (please see Section 2.1 for a complete list of participating industries); however, we were unable to arrange interviews with some industries such as the military. Future work will aim to include industries that were underrepresented or missing in the requirements gathering phase of this study.
- We engaged geospatial users and producers from diverse GIS communities in our local context, i.e., Australia and New Zealand. However, in order to create an all-encompassing solution for assessing fitness-for-use of geospatial data, requirements elicitation should include a wider group of geospatial users and producers from around the globe. For example, the Spatial Data Quality Working Group of the Open Geospatial Consortium Technical Committee (<http://www.opengeospatial.org/projects/groups/dqdwg>), which conducted the geospatial data quality online survey in 2008 ([http://portal.opengeospatial.org/files/?artifact\\_id=30415](http://portal.opengeospatial.org/files/?artifact_id=30415)), used randomized

sample technique to reach a large number of GIS users and vendors, where respondents came from seven continents. Our future work will also focus on extending the collaboration that was initiated with our European partners during this research initiative. More specifically, we will continue to collaborate with the Quality Knowledge Exchange Network (QKEN (<https://eurogeographics.org/knowledge-exchange/qken/>)) of EuroGeographics, in order to share insights and experiences and uncover additional informational aspects of spatial data that are influential for assessing fitness-for-use of geospatial data sources. This information will be used to refine the GUCM model, which will in turn lead to a more inclusive vocabulary to enable spatial data users to assess fitness-for-use of geospatial data.

## 5. Conclusions

In this article, we addressed the problem of communicating spatial data quality and fitness-for-use, in order to enable users to identify spatial data sources that are best suited to their intended uses in the context of specific applications and domains. We introduced the Geospatial User Centric Metadata (GUCM) Ontology for communicating fitness-for-use of spatial data sources to users in the spatial and other domains. The requirements that emerged from the analyses of our industry engagements provided the foundation upon which the GUCM ontology was designed. Our analyses aimed to identify the gaps that exist between internal quality (producer supplied) and external quality (consumer described). In addition to the main features that distinguish GUCM from previous efforts for communicating spatial data quality and fitness-for-use (please see Section 4.1), the ontology is designed to capture and represent metadata, as use cases emerge or evolve over time; e.g., due to changes to user requirements or physical objects or natural phenomena. In other words, metadata description is not restricted by use cases, as contextual information, such as application, domain and user profiles are captured to put metadata into context.

Future work will focus on continuing this initiative into a validation and utilization phase, in order to demonstrate the practical value of the GUCM ontology and its potential impact. The GUCM ontology is in the process of being implemented by Western Australian Land Information Authority, i.e., Landgate (<https://www0.landgate.wa.gov.au/>). Landgate will implement the ontology within their Web portal, which is powered by CKAN (<https://ckan.org/>), in order to test and validate the design in a real-world environment. In addition, the project aims to implement the vocabulary within the Australian Government public data platform (<https://data.gov.au>), a platform for discovering, accessing and reusing public data. Towards this goal, the project will liaise with the Australian Government Linked Data Working Group (<http://linked.data.gov.au/>) to publish the GUCM ontology using some part of the data.gov.au domain. Implementing the GUCM ontology in the data.gov.au platform will enable the project to assess its usefulness for communicating metadata to spatial data users and facilitating interoperability between spatial and nonspatial metadata. Furthermore, the implementation will assess the extent to which the GUCM ontology facilitates collaboration between producers and users of spatial data for communicating spatial metadata and fitness-for-use of spatial datasets in the context of various applications and domains. Lessons learned from implementing GUCM within the Landgate and Australian Government public data platforms will be used to refine the design of the ontology. The outcome of this iterative design process will be communicated to spatial data users, through Landgate and the Australian Government public data platforms and feedback will be sought from a wider range of participants through an online questionnaire. In addition, the project aims to gather insights from an expert community by participating in the Open Geospatial Consortium (OGC) Innovation Program (<http://www.opengeospatial.org/ogc/programs/ip>) to prototype and test the ontology.

It is important to note that the GUCM ontology reuses many widely adopted ontologies and vocabularies. At the time of designing the ontology, some of these vocabularies, were only available as an RDF or OWL file and not published in any domain. Examples of these vocabularies are ISO 19109:2015 and the GUF standard. Therefore, the GUCM ontology imports all the ontologies that

it reuses; consequently, despite the fact that GUCM is a very light-weight ontology, it contains a large number of imported vocabularies. Future work will also focus on de-coupling these ontologies from GUCM, e.g., publishing them or working with other standardization bodies to publish them to the Web of Data, where they can be reliably accessed using their published URIs. This will in turn reduce the size and complexity of GUCM, rendering it into a very lightweight ontology that can be easily and effectively updated and maintained. This has been done as part of exposing G-NAF (<https://www.pisma.com.au/products/g-naf>), a dataset containing all physical addresses in Australia, to the Linked Open Data and for G-NAF referencing standard definitions for address related concepts that apply to the geospatial domain (<http://pid.data.gov.au/websrv/reference/def/ont/iso19160-1-address>). As a short-term exercise, the GUCM ontology can be decoupled from the ontologies that it reuses, and all ontologies can be published on a server. GUCM can be modified to access these ontologies using their existing URIs. However, as a long-term initiative, e.g., in order to publish the ontology to the data.gov.au domain, an official “Persistent ID” could be sought on data.gov.au through the Australian Government Linked Data Working Group (<http://linked.data.gov.au/>) and all URIs can then be redirected to their persistent space.

Finally, it is worth noting that implementation of GUCM is quite an important task, as data, metadata, applications, and use cases are constantly changing. Hence, GUCM would have to evolve to accommodate such changes. We aim to provide application programming interfaces (APIs) to streamline the process of capturing and representing metadata, as data and user requirements change over time. APIs will also facilitate integration of GUCM metadata with metadata from spatial and nonspatial domains in other platforms, such as open data platforms. We expect this to be an iterative process, through which these APIs will be refined. Our aim is to minimize the effort required in updating and customizing the APIs, as data, user requirements, use cases and applications evolve.

**Author Contributions:** Conceptualization, Alistair Barros and Hasti Ziainatin; Methodology, Hasti Ziainatin; Software, Hasti Ziainatin; Validation, Hasti Ziainatin; Formal Analysis, Hasti Ziainatin and Alireza Nili; Investigation, Hasti Ziainatin, Alireza Nili, Alistair Barros; Resources, Alistair Barros; Data Curation, Hasti Ziainatin; Writing—Original Draft Preparation, Hasti Ziainatin and Alireza Nili; Writing—Review & Editing, Hasti Ziainatin, Alireza Nili and Alistair Barros; Visualization, Hasti Ziainatin; Supervision, Alistair Barros; Project Administration, Alistair Barros, Hasti Ziainatin and Alireza Nili; Funding Acquisition, Alistair Barros All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Australia and New Zealand Cooperative Research Centre for Spatial Information (CRCSI), Project Agreement 3.16. This project agreement was made between Spatial Information Systems Research Limited, PSMA Australia Limited, Queensland University of Technology, The State of New South Wales through its Department of Finance, Services & Innovation Spatial Services Division, Land Information New Zealand, Western Australian Land Information Authority (Landgate) and The State of Victoria through its Department Of Environment, Land, Water & Planning (DELWP).

**Acknowledgments:** We are very grateful to our industry participants, who generously donated their time to assist in our studies by engaging in our interviews and completing the online questionnaire.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Devillers, R.; Stein, A.; Bédard, Y.; Chrisman, N.; Fisher, P.; Shi, W. Thirty years of research on spatial data quality: Achievements, failures, and opportunities. *Trans. GIS* **2010**, *14*, 387–400. [CrossRef]
2. Goodchild, M.F. Sharing imperfect data. In *Sharing Geographic Information*; Onsrud, H.J., Rushton, G., Eds.; Centre for Urban Policy Research: New Brunswick, NJ, USA, 1995; pp. 413–425.
3. Devillers, R.; Jeansoulin, R. Spatial data quality: Concepts. In *Fundamentals of Spatial Data Quality*; Devillers, R., Jeansoulin, R., Eds.; ISTE Ltd.: London, UK, 2006; pp. 31–42.
4. Arnold, L. Spatial Data Supply Chain and End User Frameworks: Towards an Ontology for Value Creation. In *GeoValue Workshop*; Curtin University: Perth, Australia, 2016.
5. Goodchild, M.F. Foreword. In *Fundamentals of Spatial Data Quality*; Devillers, R., Jeansoulin, R., Eds.; ISTE Ltd.: London, UK, 2006; pp. 13–16.

6. Chrisman, N.R. The error component in spatial data. In *Geographical Information Systems: Overview Principles and Applications*; Maguire, D.A., Goodchild, M.F., Rhind, D.W., Eds.; Longman: White Plains, NY, USA, 1991; pp. 165–174.
7. Ivánová, I.; Morales, J.; de By, R.A.; Beshe, T.S.; Gebresilassie, M.A. Searching for spatial data resources by fitness-for-use. *J. Spat. Sci.* **2013**, *58*, 15–28. [[CrossRef](#)]
8. Gahegan, M. The Grid. Bringing Data Producers and Consumers Closer? In *NIEeS Workshop on Activating Metadata*; Cambridge University Press: Cambridge, UK, 2005.
9. Longhorn, R.A. Geospatial standards, interoperability, metadata semantics and spatial data infrastructure. In *NIEeS Workshop on Activating Metadata*; Cambridge University Press: Cambridge, UK, 2005.
10. Comber, A.J.; Fisher, P.F.; Wadsworth, R.A. User-focused metadata for spatial data, geographical information and data quality assessments. In Proceedings of the 10th AGILE International Conference on Geographic Information Science, Aalborg University, Aalborg, Denmark, 8–11 May 2007; pp. 1–13.
11. Goodchild, M.F. Putting research into practice. In *Quality Aspects of Spatial Data Mining*; Stein, A., Shi, W., Bijker, W., Eds.; CRC Press: Boca Raton, FL, USA, 2009; pp. 345–356.
12. Brown, M.; Sharples, S.; Harding, J.; Parker, C.; Bearman, N.; Maguire, M.; Forrest, D.; Haklay, M.; Jackson, M. Usability of geographic information: Current challenges and future directions. *Appl. Ergon.* **2013**, *44*, 855–865. [[CrossRef](#)] [[PubMed](#)]
13. Boin, A.T.; Hunter, G.J. Do spatial data consumers really understand data quality information? In Proceedings of the 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Lisbon, Portugal, 5–7 July 2006; pp. 215–224.
14. Comber, A.J.; Fisher, P.F.; Wadsworth, R.A. Approaches for Providing User Relevant Metadata and Data Quality Assessments. In *Geographical Information Science Research UK Conference (GISRUK)*; National Centre for Geocomputation, National University of Ireland: Maynooth, Ireland, 2007; pp. 79–82.
15. Goodchild, M.F. The future of digital earth. *Ann. GIS* **2012**, *18*, 93–98. [[CrossRef](#)]
16. Ellul, C.; Foord, J.; Mooney, J. Making metadata usable in a multi-national research setting. *Appl. Ergon.* **2013**, *44*, 909–918. [[CrossRef](#)] [[PubMed](#)]
17. Berners-Lee, T.; Hendler, J.; Lassila, O. The semantic web. *Sci. Am.* **2001**, *284*, 34–43. [[CrossRef](#)]
18. Bizer, C.; Heath, T.; Berners-Lee, T. Linked data: The story so far. In *Semantic Services, Interoperability and Web Applications: Emerging Concepts*; IGI Global: Hershey, PA, USA, 2011; pp. 205–227.
19. Lee, Y.W.; Pipino, L.L.; Funk, J.D.; Wang, R.Y. *Journey to Data Quality*; Massachusetts Institute of Technology: Cambridge, MA, USA, 2006.
20. Wang, R.Y.; Strong, D.M. Beyond accuracy: What data quality means to data consumers. *J. Manag. Inf. Syst.* **1996**, *12*, 5–33. [[CrossRef](#)]
21. Batini, C.; Scannapieco, M. *Data and Information Quality*; Springer International Publishing: Cham, Switzerland, 2016; Volume 43.
22. Debattista, J.; Lange, C.; Auer, S.; Cortis, D. Evaluating the quality of the LOD cloud: An empirical investigation. *SWJ* **2018**, *9*, 859–901. [[CrossRef](#)]
23. Debattista, J.; Clinton, E.; Brennan, R. Assessing the quality of geospatial linked data—experiences from Ordnance Survey Ireland (OSi). In Proceedings of the SEMANTiCS Conference, Vienna, Austria, 11–13 September 2018.
24. Attard, J.; Brennan, R. A semantic data value vocabulary supporting data value assessment and measurement integration. In Proceedings of the 20th International Conference on Enterprise Information Systems, Madeira, Portugal, 21–24 March 2018.
25. ISO. *ISO 9000:2015 Quality Management Systems—Fundamentals and Vocabulary*; ISO: Geneva, Switzerland, 2015.
26. ISO. *ISO 19157:2013 Geographic Information—Data Quality*; ISO-Standard and Swedish SIS Standard; ISO: Geneva, Switzerland, 2013.
27. Congalton, R.G. Accuracy assessment and validation of remotely sensed and other spatial information. *Int. J. Wildland Fire* **2001**, *10*, 321–328. [[CrossRef](#)]
28. ISO. *ISO 19115-1:2014 Geographic Information-Metadata-Part 1: Fundamentals*; International Standards Organization: Geneva, Switzerland, 2014.
29. ISO. *ISO 19158:2012 Geographic Information—Quality Assurance of Data Supply*; ISO: Geneva, Switzerland, 2012.

30. da Silva, J.R.; Castro, J.A.; Ribeiro, C.; Honrado, J.; Lomba, Â.; Gonçalves, J. Beyond INSPIRE: An ontology for biodiversity metadata records. In Proceedings of the OTM Confederated International Conferences “On the Move to Meaningful Internet Systems”, Amantea, Italy, 27–31 October 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 597–607.
31. European Commission. Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 Establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). *Off. J. Eur. Union* **2007**, *50*, 1–14.
32. ISO. *ISO 19115:2003 Geographic Information-Metadata*; International Organization for Standardization (ISO): Geneva, Switzerland, 2003.
33. Braun, V.; Clarke, V. Using thematic analysis in psychology. *Qual. Res. Psychol.* **2006**, *3*, 77–101. [[CrossRef](#)]
34. Miles, M.B.; Huberman, A.M.; Saldana, J. *Qualitative Data Analysis: A Methods Sourcebook*; Sage Publications: Thousand Oaks, CA, USA, 2018.
35. George, D.; Mallery, P. *IBM SPSS Statistics 23 Step by Step: A Simple Guide and Reference*; Routledge: London, UK, 2016.
36. Gennari, J.H.; Musen, M.A.; Fergerson, R.W.; Grosso, W.E.; Crubézy, M.; Eriksson, H.; Noyb, N.F.; Tu, S.W. The evolution of Protégé: An environment for knowledge-based systems development. *Int. J. Hum. Comput. Stud.* **2003**, *58*, 89–123. [[CrossRef](#)]
37. Gašević, D.; Djuric, D.; Devedžić, V. *Model Driven Engineering and Ontology Development*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2009.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).