INSTITUT NATIONAL
DE L'INFORMATION
GÉOGRAPHIQUE
ET FORESTIÈRE

COMPTE RENDU DE LECTURE

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

Nicolas PY / Marie LAMBOIS

REDUCING CONSUMER UNCERTAINTY: TOWARDS AN ONTOLOGY FOR GEOSPATIAL USER-CENTRIC METADATA

https://**doi**.org/10.3390/ijgi9080488

L'utilisateur est confronté à un foisonnement de donnée, certes de plus en plus méta-documentées.

Ces métadonnées sont cependant 'orientées' producteur, et ne permettent pas à l'utilisateur de juger de l'adéquation à ses besoins.

Cette publication résume les travaux sur la communication à l'utilisateur d'informations d'adéquation a son besoin au travers d'une ontologie appuyée sur les standards,

ISPRS International Journal of Geo-Information

MDPI

Article

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

Hasti Ziaimatin *, 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: hasties.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

ISPRS INTERNATIONAL JOURNAL OF GEO INFORMATION (ISSN 2220-9964)

International

peer-reviewed

open access

journal on geo-information.

It is a journal of the ISPRS (International Society for Photogrammetry and Remote Sensing) and is published monthly online by MDPI.

ISPRS International Journal of Geo-Information — Open Access Journal

ISPRS International Journal of Geo-Information (ISSN 2220-9964) is an international **peer-reviewed** open access journal on geo-information. It is a journal of the **ISPRS** (International Society for Photogrammetry and Remote Sensing) and is published monthly online by MDPI.

- **Open Access** —free for readers, with article processing charges (APC) paid by authors or their institutions.
- **High Visibility:** The journal is covered by Science Citation Index Expanded (Web of Science), Scopus and INSPEC (IET).
- **CiteScore** (2019 Scopus data): 3.8, which equals rank 28/95 (Q1) in the category 'Earth and Planetary Sciences (miscellaneous)', rank 99/679 (Q1) in 'Geography, Planning and Development', and rank 13/40 (Q3) in 'Computers in Earth Sciences.'
- **Rapid Publication:** manuscripts are peer-reviewed and a first decision provided to authors approximately 15.2 days after submission; acceptance to publication is undertaken in 2.3 days (median values for papers published in this journal in the first half of 2020).
- **Recognition of Reviewers:** reviewers who provide timely, thorough peer-review reports receive vouchers entitling them to a discount on the APC of their next publication in any MDPI journal, in appreciation of the work done.

Impact Factor: 2.239 (2019) ; 5-Year Impact Factor: 2.402 (2019)

LA QUESTION

« Tandis qu'il y a des méthodes pour l'évaluation de la qualité interne des données géographiques, l'évaluation de la qualité externe reste une question ouverte.

Les utilisateurs, experts ou non, de données géographiques sont présumés en capacité de savoir de quel type de données ils ont besoin et dans quel entrepôt la trouver.

Même avec cette information disponible, les utilisateurs ne sont pas accompagnés dans la décision de l'adéquation de cette donnée à leur besoin, décision basée sur des métadonnées complexes dans les trop peu nombreux cas où ces métadonnées sont disponibles »

[…]

« Tandis que les efforts de standardisation ont permis l'amélioration significative de l'interopérabilité […] souvent les informations de qualité de la donnée ne sont pas transmises de manière standardisée à l'utilisateur »

LA PROPOSITION

L'étude apporte la conception d'une ontologie au travers de l'analyse de standards, vocabulaires, et l'élicitation de pratiques de producteurs et/ou utilisateurs.

Il est souhaité que cette ontologie permette de décrire la qualité et l'adéquation à l'usage :

- A différents niveaux de granularité : jeu de donnée, objet, attribut
- À l'aide d'un modèle de métadonnées unique, utilisable tant par un producteur qu'un utilisateur
- Intéropérable avec les modèles existants

Ontologie

Ensemble structuré de concepts, eux-mêmes organisés dans un graphe dont les relations peuvent être des relations sémantiques ou des relations de composition et d'héritage (au sens objet).

Source: wiktionnaire

L'EXISTANT

Le monde de la géo

Métadonnées ISO 19115
ISO 19157

Geospatial User Feedback

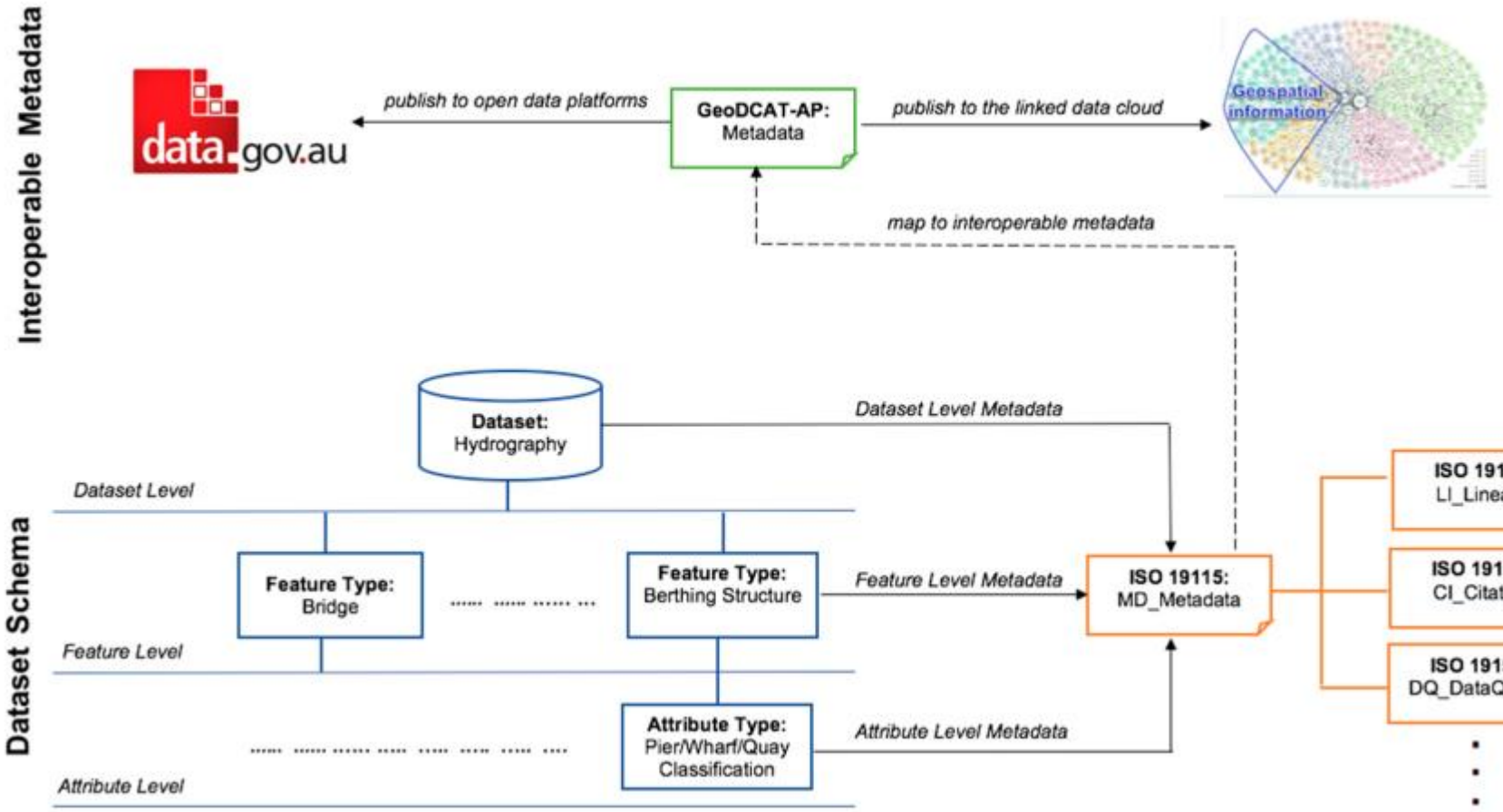
Geonetwork plugin :
https://github.com/metadata101/guf10

Le monde de l'OpenData

Métadonnées DCAT
Data Quality Vocabulary
(W3C)
Dataset Usage Vocabulary

Geonetwork plugin :
https://www.plan4all.eu/2018/03/team-8-exposing-guf-metadata-as-duv-in-the-geodcat-ap-output-of-geonetwork/

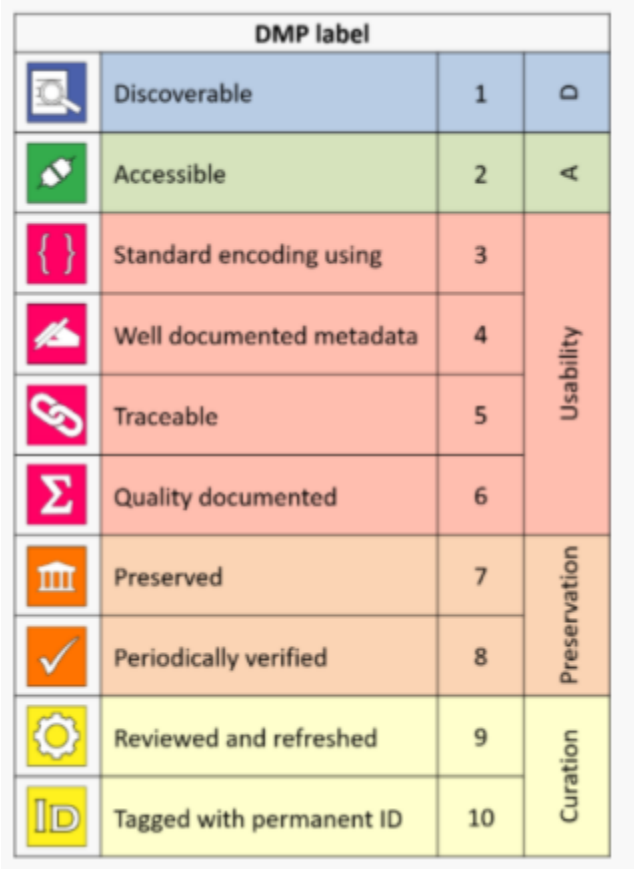
L'EXISTANT


Interoperable Metadata
data.gov.au
publish to open data platforms
GeoDCAT-AP:
Metadata
publish to the linked data cloud
Geospatial information
map to interoperable metadata
Dataset Schema
Dataset:
Hydrography
Dataset Level Metadata
Dataset Level
Feature Type:
Bridge
Feature Type:
Berthing Structure
Feature Level Metadata
ISO 19115:
MD_Metadata
ISO 19115:
LI_Lineage
ISO 19115:
CI_Citation
ISO 19157:
DQ_DataQuali
Feature Level
Attribute Type:
Pier/Wharf/Quay
Classification
Attribute Level Metadata
Attribute Level

L’EXISTANT : GEOVIQUA

GeoLabel

http://geolabel.info/DMP_generation.htm



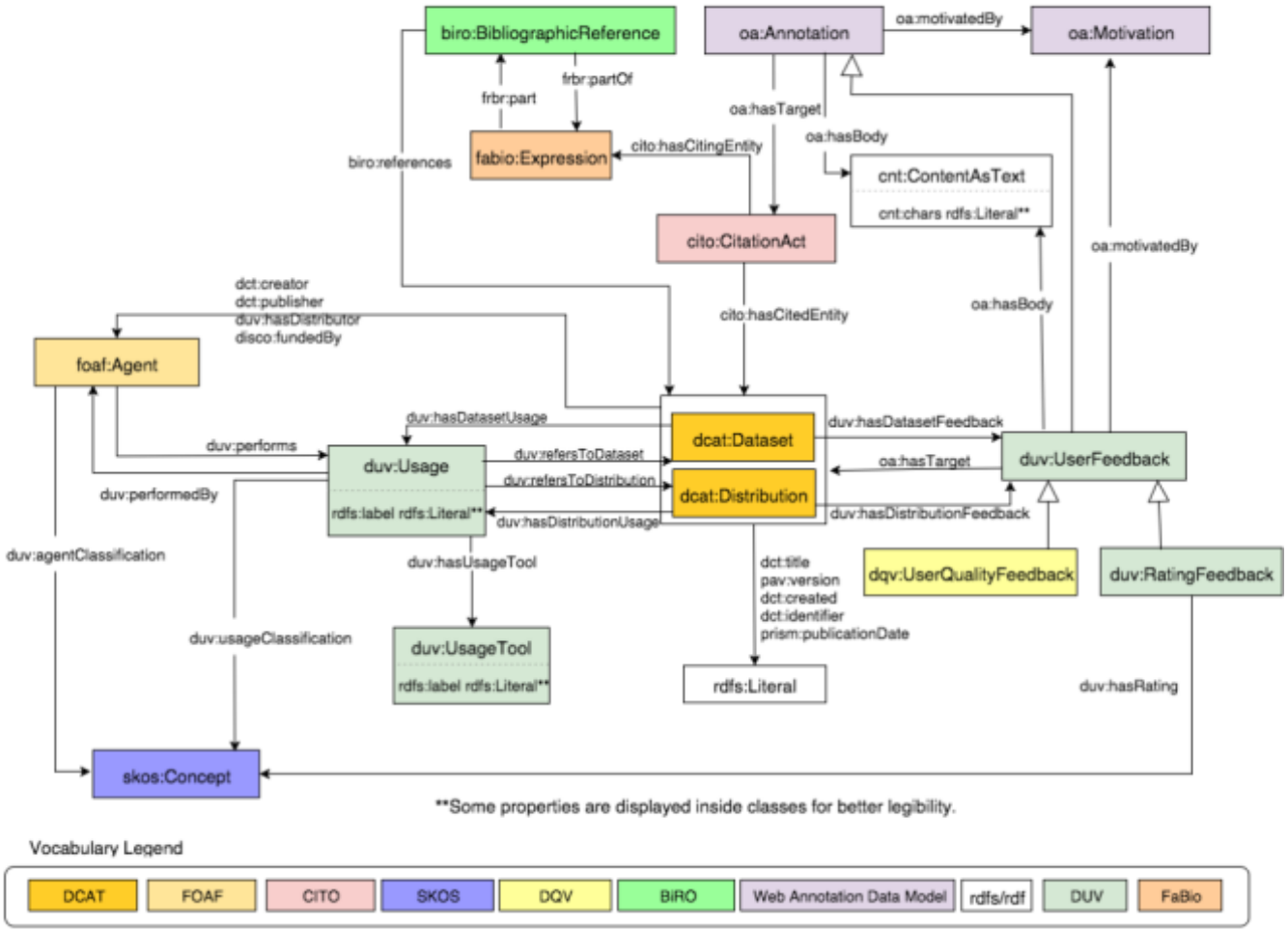
DMP label			
	Discoverable	1	D
	Accessible	2	A
	Standard encoding using	3	Usability
	Well documented metadata	4	
	Traceable	5	
	Quality documented	6	
	Preserved	7	Preservation
	Periodically verified	8	
	Reviewed and refreshed	9	Curation
	Tagged with permanent ID	10	

L ’EXISTANT : DUV

Dataset Usage Vocabulary

Recommandation du W3C

vocabulaire « glue »



LA MÉTHODE (CONTEXTE: AUSTRALIE)

i. Elicitation des points de vue producteur et utilisateurs sur la qualité des données et l'adéquation à l'usage

i. Par des entretiens semi-structurés (n=28)

ii. Par un questionnaire (n=15)

ii. Analyse qualitative et quantitative des pratiques

iii. Conception de l'ontologie GUCM (Geospatial User-Centric Metadata)

RÉSULTAT, LES ITEMS QUALITÉ

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

Data Quality Elements	Data Quality Sub-Elements	Direct Quote from Participant
Accuracy 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.	**Positional/Spatial Accuracy** 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."
	Attribute/Thematic Accuracy 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…"
	Temporal Accuracy 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 *accuracy of a time measurement*.	"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."
Completeness 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."
Logical Consistency 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"

RÉSULTAT, LES ITEMS QUALITÉ

Relevancy Relevancy (perceived relevancy) of a specific dataset to a user's intended uses and business purposes.	"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."
Currency 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.	"[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."
Reliability 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.	"…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."
Lineage (provenance information) historical information such as the source, producer, content, methods of collecting and processing, and geographic coverage of the data product.	"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."
Cost 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).	"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."

RÉSULTATS, LES ITEMS ADÉQUATION À L'USAGE

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
Producer profile: 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"
Dataset citation information: 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."
Data dictionary: 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."
Quantitative quality information: 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."
Soft knowledge: 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."
Compliance with standards: 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."
User ratings of the dataset (as a part of peer reviews and feedback): 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."
Community recommendations and advice (as a part of peer reviews and feedback): 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."
Independent expert reviews: 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..."

De nombreux champs à explorer…

RÉSULTATS, INTERVIEW SEMI-DIRIGÉS

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

RÉSULTATS, SONDAGE (N=15, AUSTRALIE)

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							

53% des sondés (n=15) avouent que leur organisation a au moins été une fois perturbée par un non/mauvaise compréhension de l'adéquation à l'usage d'une donnée

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

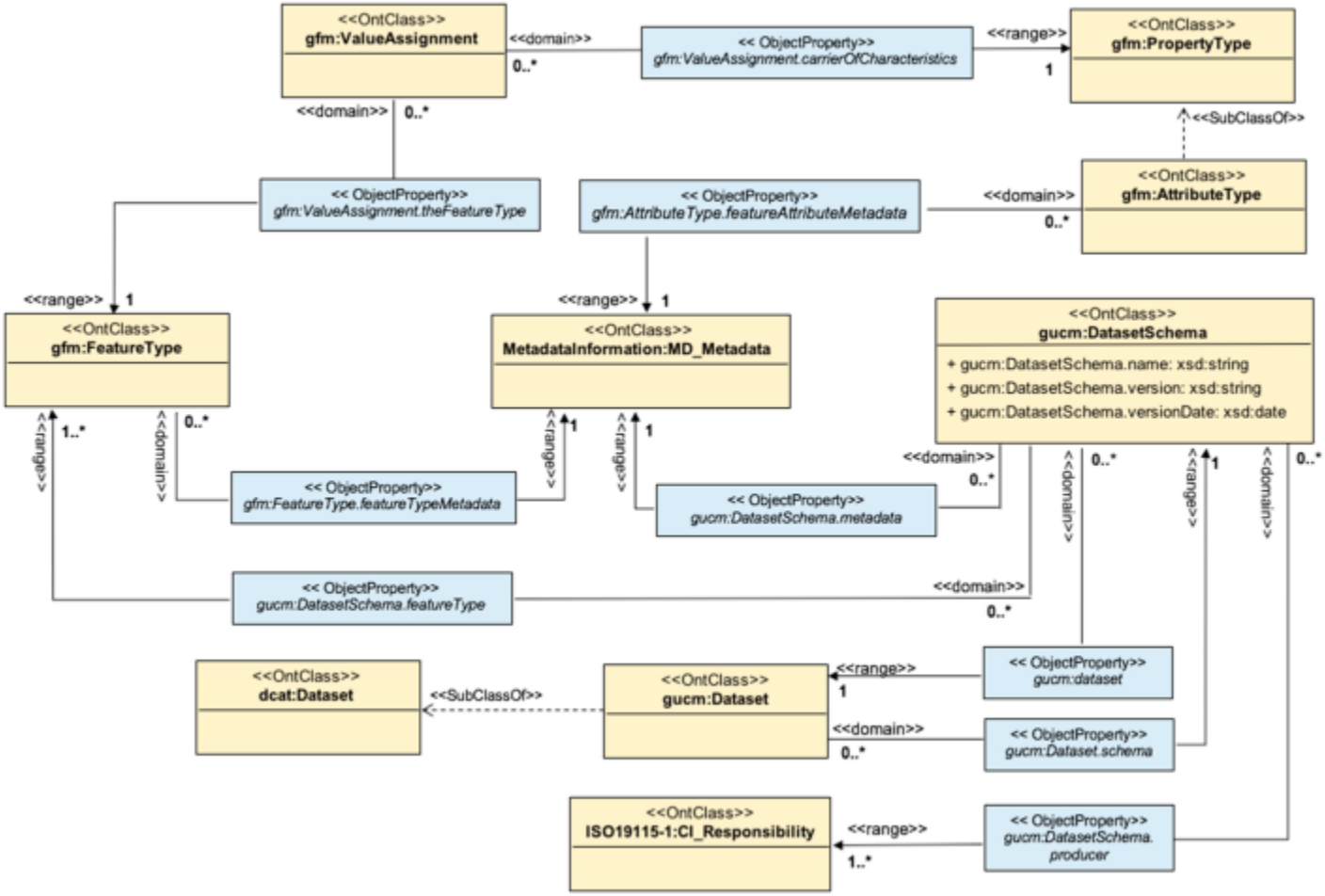
	N	Range	Minimum	Maximum	Mean	Std. Error	Std. Deviation	Variance
Experts' Review	15	5	2	7	4.67	0.361	1.397	1.952
Compliance with Standards	15	5	1	6	4.07	0.431	1.668	2.781
Community Advice and Recommendations (User Feedback)	15	4	2	6	4.27	0.300	1.163	1.352
Producer Profile (Reputation)	15	5	2	7	5.27	0.358	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							

RÉSULTATS, L'ONTOLOGIE GUCM

Trois composantes

La description du jeu de donnée

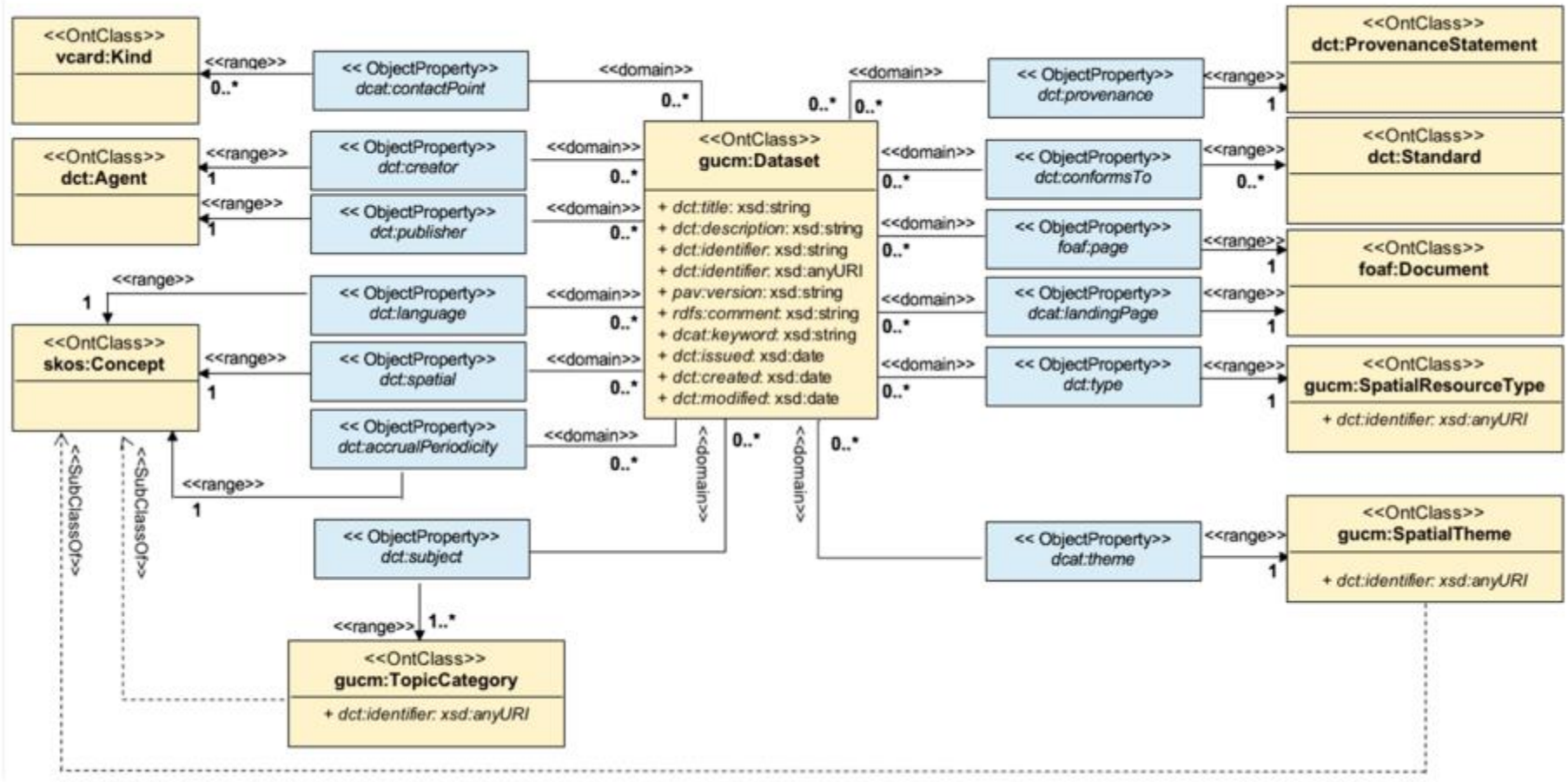
- à différents niveaux de granularité (jeu de donnée, objet, attribut)
- ISO 19109:2015, et son General Feature Model / Rules for application schema
- Métadonnées ISO 19115-1:2014 et ISO 19157:2013



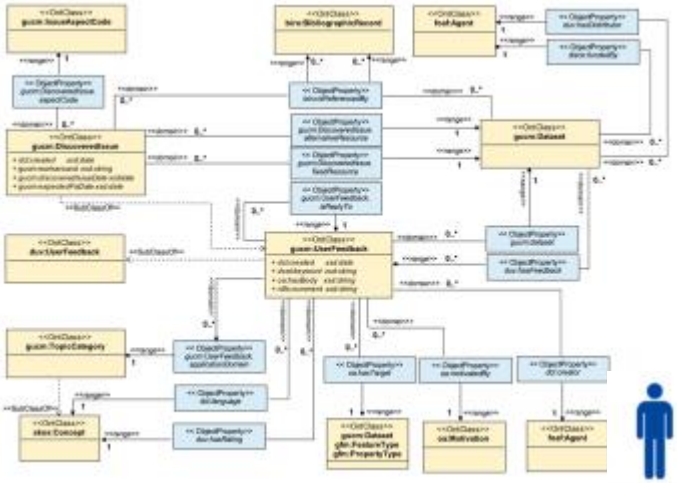
RÉSULTATS, L'ONTOLOGIE GUCM

Trois composantes

Métadonnées interopérables entre portails à composants spatiale ou non ; passerelle entre standards de métadonnées

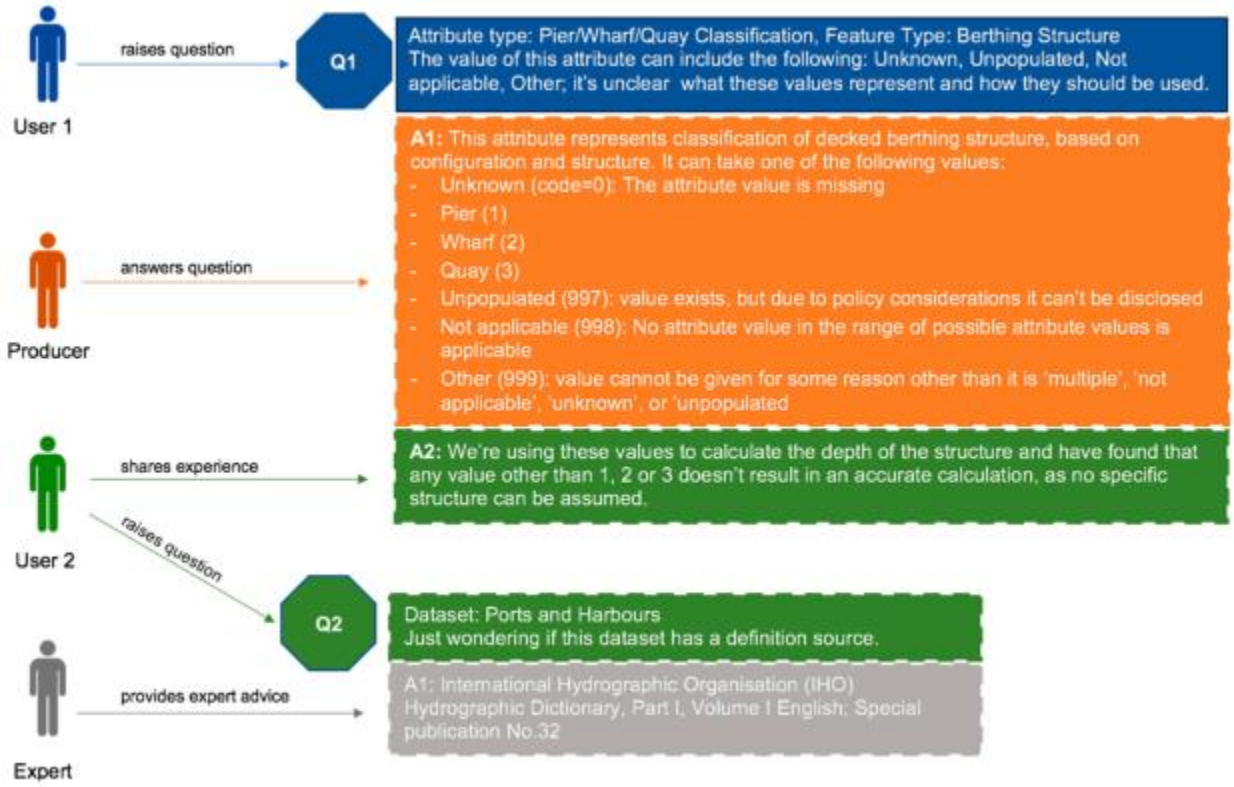


RÉSULTATS, L'ONTOLOGIE GUCM



Trois composantes

Feedback, utilise le Dataset Usage Vocabulary (DUV, recommandé par le W3C pour publier et utiliser des données sur le web), extension du vocabulaire du Data Catalog DCAT, incorpore le Geospatial User Feedback (GUF)



ET ENSUITE

Les auteurs envisagent

De poursuivre la collaboration avec le Quality Knowledge Exchange Network (QKEN (https://eurogeographics.org/knowledge-exchange/qken/)) d'EuroGeographics où l'IGN est représenté par Chantal Coulomb

De passer à une phase de validation et d'utilisation

- **Le GUCM sera implémenté par Western Australian Land Information Authority, i.e., Landgate (https://www0.landgate.wa.gov.au/)**
- **Le projet souhaite implémenter GUCM sur la plateforme open data du Gouvernement (https://data.gov.au)**
- **L'ontologie GUCM sera proposée au Australian Government Linked Data Working Group (http://linked.data.gov.au/)pour être inscrite en tant que registre hébergé sur le domaine data.gov.au**
- **Raffiner et tester l'ontologie via le Open Geospatial Consortium (OGC) Innovation Program (http://www.opengeospatial.org/ogc/programs/ip)**

APARTÉ: BEYOND ACCURACY: WHAT DATA QUALITY MEANS TO DATA CONSUMERS (1996)

http://dx.doi.org/10.1080/07421222.1996.11518099

Ability to be Joined With	Ability to Download	Ability to Identify Errors	Ability to Upload
Acceptability	Access by Competition	Accessibility	Accuracy
Adaptability	Adequate Detail	Adequate Volume	Aestheticism
Age	Aggregatability	Alterability	Amount of Data
Auditable	Authority	Availability	Believability
Breadth of Data	Brevity	Certified Data	Clarity
Clarity of Origin	Clear Data Responsibility	Compactness	Compatibility
Competitive Edge	Completeness	Comprehensiveness	Compressibility
Concise	Conciseness	Confidentiality	Conformity
Consistency	Content	Context	Continuity
Convenience	Correctness	Corruption	Cost
Cost of Accuracy	Cost of Collection	Creativity	Critical
Current	Customizability	Data Hierarchy	Data Improves Efficiency
Data Overload	Definability	Dependability	Depth of Data
Detail	Detailed Source	Dispersed	Distinguishable
Dynamic	Ease of Access	Ease of Comparison	Updated Files Ease of Correlation
Ease of Data Exchange	Ease of Maintenance	Ease of Retrieval	Ease of Understanding
Ease of Update	Ease of Use	Easy to Change	Easy to Question
Efficiency	Endurance	Enlightening	Ergonomic
Error-Free	Expandability	Expense	Extendibility
Extensibility	Extent	Finalization	Flawlessness
Flexibility	Form of Presentation	Format	Integrity
Friendliness	Generality	Habit	Historical Compatibility
Importance	Inconsistencies	Integration	Integrity
Interactive	Interesting	Level of Abstraction	Level of Standardization
Localized	Logically Connected	Manageability	Manipulable
Measurable	Medium	Meets Requirements	Minimality
Modularity	Narrowly Defined	No lost information	Normality
Novelty	Objectivity	Optimality	Orderliness
Origin	Parsimony	Partitionability	Past Experience
Pedigree	Personalized	Pertinent	Portability
Preciseness	Precision	Proprietary Nature	Purpose
Quantity	Rationality	Redundancy	Regularity of Format
Relevance	Reliability	Repetitive	Reproducibility
Reputation	Resolution of Graphics	Responsibility	Retrievability
Revealing	Reviewability	Rigidity	Robustness
Scope of Info	Secrecy	Security	Self-Correcting
Semantic Interpretation	Semantics	Size	Source
Specificity	Speed	Stability	Storage
Synchronization	Time-independence	Timeliness	Traceable
Translatable	Transportability	Unambiguity	Unbiased
Understandable	Uniqueness	Unorganized	Up-to-Date
Usable	Usefulness	User Friendly	Valid
Value	Variability	Variety	Verifiable
Volatility	Well-Documented	Well-Presented	

Figure 1. Data Quality Attributes Generated from the First Survey

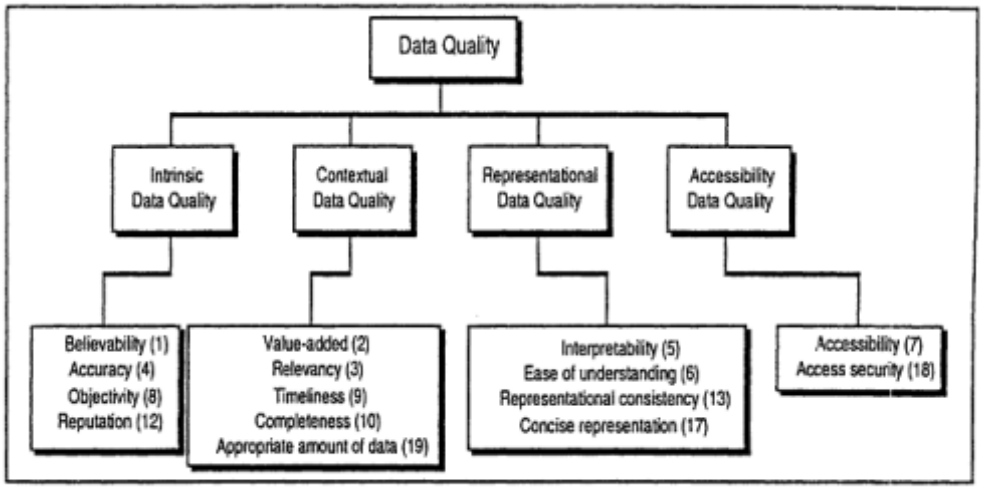


Figure 2. A Conceptual Framework of Data Quality