

kadaster



# Data Quality for Use: A Linked Data Approach

Erwin Folmer 25-01-2020  
Erwin.Folmer@kadaster.nl



# Dr.ir. Erwin Folmer

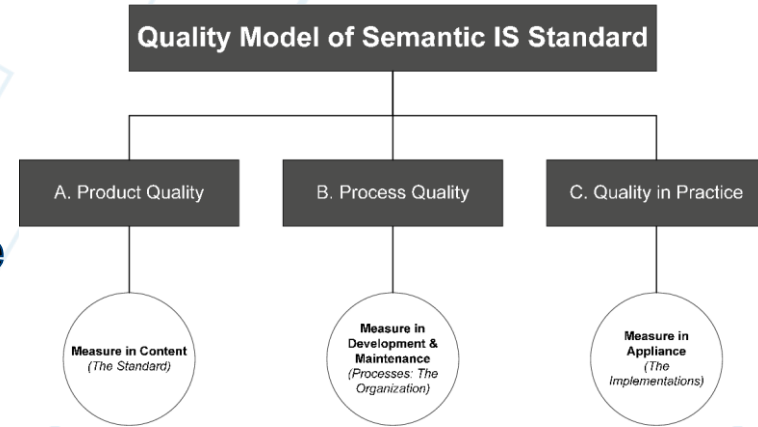
Data Science Team Lead @ Kadaster

Research Associate on Interoperability & Standards @ University of Twente

Board Member @ Platform Linked Data Netherlands

# Data Quality

- Juran – “Fitness for Use”
- Product – Process – Quality in Use
- Definitions



We see a lot of focus on “product quality”, and not on quality in use. (both in the standards world (my PhD) and in the spatial data world...)

# Two “Quality in Use” Problems for Spatial Data

1. Datasets can not be found (ISO 19115 metadata; example BAG, Universities)
  - a) Search in national spatial portal
  - b) Search in national/european portal
  - c) Search in google/google dataset search
  
2. Unable to determine quality (examples universities-BRT; 1005-BAG)
  - a) Quality statements are not in the metadata  
(implicit knowledge, or documents)

# Findings - User Research



	Data consultant (1)	Company director (1)	Software developers (2)	Mining engineer (1)	GIS officers/ technicians (3)	GIS analysts/ architect (5)	Heads of research (4)	Rank
Search using Google	✓	✓	✓	✓	✓	✓	✓	=1
Find Data.gov.uk search poor	✓	✓	✓	✓	✓	✓	✓	=1
Want data standardised for cross-referencing	✓	✓	✓		✓	✓	✓	2
Want to be able to find all geospatial data in one place	✓	✓	✓	✓		✓		3
Find licensing unclear	✓	✓	✓	✓				=4
Want a clear journey from catalogue to data	✓	✓	✓		✓			=4
Struggle with acronyms	✓	✓	✓					5
Want data hosted on a platform that does not move					✓	✓		=6
Want feature-level datasets					✓	✓		=6

# Solution: Focus & Attitude shift

## Attitude shift: from product quality to quality in use

- Not only improving the quality from 99.8% to 99.9%
- For some users 80% (intrinsic/product) quality will be more than enough
- More open data (data suppliers that think their data is not good enough for publication)

## Focus on Transparency!

- Dashboards
- **Transparency through Metadata**

# A. Dashboards

We have them...

- Custom build, many improvements, data supplier focussed (product quality), static.
- Move to standard products (BI), interactive and data user focussed (quality in use).



## BAG kwaliteitsdashboard voor afnemers

Over dit dashboard

**Dashboard**

Download

Feedback

Meer BAG

### Thema's

Nu getoond: Statusconflicten

### Verfijnen binnen thema

Nu getoond: Statusconflict nummeraanduiding - adresseerbaar object

Een nummeraanduiding met een huidige status mag niet gerelateerd zijn aan een adresseerbaar object met een beëindigde status.



Particulier

Kadaster.nl

Het Kadaster

Werkenbijhetkadaster.nl  
Over ons  
Nieuws  
Tarieven  
Nieuwsgelieven

Wat doen we

Registraties  
Producten  
Advies  
Woordenboek

Mijn Kadaster

Inloggen  
Abonneren  
Vraag en antwoord  
Wijziggen abonnement  
Onze abonnementen

English

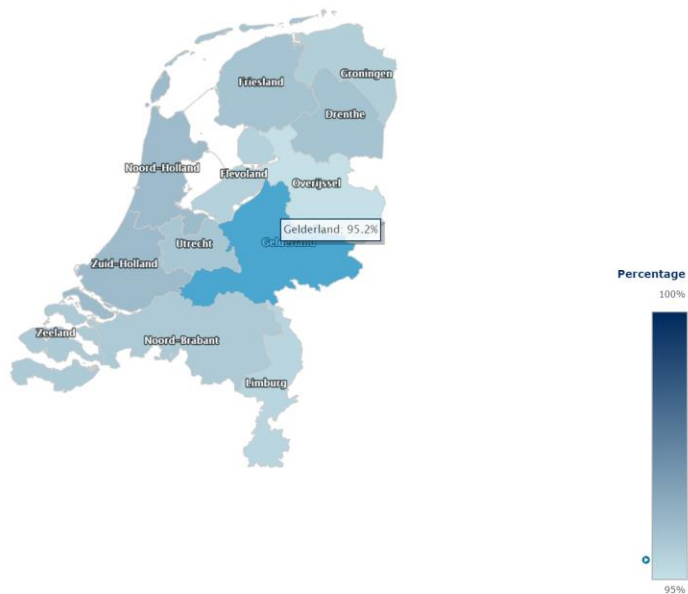
Kadaster.com  
About Kadaster  
International collaboration  
International relations

Contact

Klantenservice  
Kadasterkantoren  
Facturering  
Klachten en bezwaren  
Postadres algemeen



### BRT kwaliteit

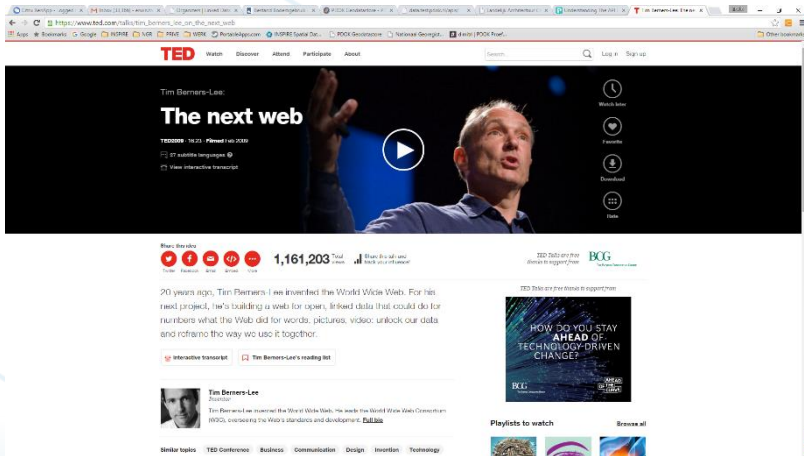


## B. Transparency through Metadata

- Metadata is the key to have more transparency for findability and usability
- The quality of the metadata has to be improved
- Metadata: Standards based (W3C-Linked Data).
- Data: Published as Linked Data offers opportunity

# Linked Data

**Linked Data: A way of publishing data (reusable, web standards, semantics, related to open and big data).  
Data is stored as triples (RDF standard) and can be (federated) queried with SPARQL.**



Everything has URIs: a webpage that can have metadata and be indexed

# Linked Data @ Kadaster

Example URI of Dataset, Model, Instance

e.g. <https://bag.basisregistraties.overheid.nl>



INDY/TECH

# THE MAN WHO'S MAKING GOOGLE MAPS SMARTER

My Blog    Events & Press

Rather than publishing online a database of railway station locations in the Netherlands and expecting a user to then query the database for “Amsterdam Centraal Station”, publish the database giving each record a URI so for example Amsterdam Centraal Station becomes;

<https://brt.basisregistraties.overheid.nl/top10nl/id/gebouw/102625209>

Now this is something I can paste into an email, tweet or even share on Facebook !

Kudos to the Dutch Kadaster for taking this approach and providing this example, Ordnance Survey you could do the same ?

This approach also results in such data becoming part of the “mainstream” web indexable and searchable, but I argue the key benefit is the “linkability”

The [Spatial Data on the Web best practice document](#), something of course I recommend you taking a longer look at provides many



♡    [Share icon]    Sep 12, 2017

Ed Parsons Retweeted

**Sam Zipper**  
@ZipperSam

Google Earth Engine + @NASA\_Landsat = annual 30 m resolution #irrigation maps!

Cool new study by @JillDeines:  
[onlinelibrary.wiley.com/doi/10.1002/20...](https://onlinelibrary.wiley.com/doi/10.1002/20...)



**Annual irrigation dynamics in t...**  
Sustainable management of agricultural water resources  
[onlinelibrary.wiley.com](https://onlinelibrary.wiley.com)

♡    [Share icon]    Sep 11, 2017

[Embed](#)    [View on Twitter](#)

## B. Transparency through Metadata

	A. Dataset	B. Model	C. Instanties
1. Vindbaarheid	1A. Dataset metadata	1B. Model metadata	1C. Instantie metadata
2. Kwaliteit (fitness for use)	2A. Dataset kwaliteit	2B. Model kwaliteit	2C. Instantie kwaliteit
3. Knowledge Graph			

## B. 1A – Standards & Results

<https://labs.kadaster.nl/dissemination/2019-11-26-Metadata-2-0/index.html#/18>



## B. 1B - Model

Disclaimer: Only for data sets that are published (the model) as linked data (for example BRT)

<https://labs.kadaster.nl/dissemination/2019-11-26-Metadata-2-0/index.html#/23>

## B. 1C – Instance - SKIP

Disclaimer: Only for linked datasets....URIs.

# B. 2A,B,C Standards & Results

To determine fitness for use

<https://labs.kadaster.nl/dissemination/2019-11-26-Metadata-2-0/index.html#/34>

## B. Results

On the test data:

- Improved metadata for publication of datasets, model, instances
- Improved index-ability of ...
- Improved quality annotation of ...

And later on...

- Improved findability in Google Dataset Search
  - (first datasets, then model and instances)
- Improved reusability through quality annotations

# And by that...

- (potential) users will be able to find the data
- (potential) users will be able to determine the fitness for use for their application.
- And that is fair to the users of spatial data.



Help us improve our products. [Sign up to take part.](#)

nature > scientific data > comment > article

a natureresearch journal

MENU ▾ SCIENTIFIC DATA

Search E-alert Submit Login

Download PDF

Comment | [Open Access](#) | Published: 15 March 2016

# The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, [...] Barend Mons

*Scientific Data* **3**, Article number: 160018 (2016) | [Cite this article](#)

84k Accesses | 1187 Citations | 1464 Altmetric | [Metrics](#)

An Addendum to this article was published on 19 March 2019

## Abstract

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measureable set of principles that we refer to as the FAIR Data Principles. The intent

Associated Content

Collection

[Scientific data](#)

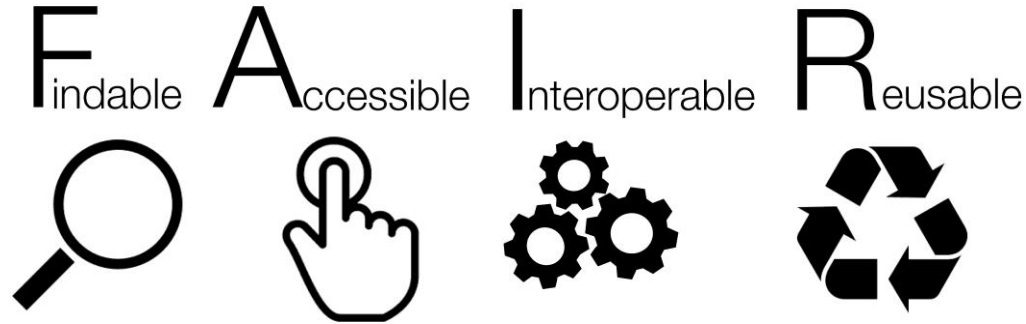
Collection

[Metadata Quality](#)

Sections [References](#)

- [Abstract](#)
- [Comment](#)
- [Additional Information](#)
- [References](#)
- [Acknowledgements](#)
- [Author information](#)
- [Ethics declarations](#)

# Conclusion



Attituded change by adopting the principles of  
FAIR Data.

# Final Conclusion

**Positive:** We know the “trick”, and it is not rocket science

**Negative:** Where is the “quality” metadata?  
(implicit knowledge)



kadaster



# Data Quality for Use: A Linked Data Approach

Erwin Folmer 25-01-2020  
Erwin.Folmer@kadaster.nl