AGILE 2017 pre-conference workshop
9 May, Wageningen University and Research

http://www.agile-online.org/index.php/conference/conference-2017

Quality assessment of geospatial data: does it fit your needs?

How to deal with quality of geospatial closed, open and big data

Abstract

Due to the enormous growth of available spatial data, quality of spatial data is becoming a very important selection criterion to find the most adequate data for the intended use. Fitness for use is leading in determining quality of data. We developed a framework to assess the quality of a data set to meet a specified user requirements. In this workshop the framework will be discussed and validated against real world use cases.

Objective(s)

- To clarify and present the geospatial data quality framework on fitness for use
- Receive feedback and recommendations to improve the framework and support future research
- Integrate possible outcome(s) in the framework
- Create better data quality awareness among participants, feedback for presenters on their proposed approach
- Report on the results of the workshop as an official publication

Planned outcome(s)

- Better data quality awareness among participants,
- Discussion and feedback for presenters on their work and/or proposed approach.
- A report on the results of the workshop as a EuroSDR official publication

Subject

In this workshop, we present an overview of data quality and recent developments worldwide.

The keynote speaker Robert Jeansoulin (Université de Paris-Est-Marne-la-Vallée), co-author of “Fundamentals of spatial data quality" [Devillers et al; 2006], will elaborate on Essentials of Data Quality and Fitness for Use .

Our second invited speaker Karin Mertens (Quality Control Manager, National Geographic Institute, Belgium) will present the work of the Quality Knowledge Exchange Group (QKEN). Karin is a member of QKEN, a network established by EuroGeographics, the European National Mapping Agencies The purpose of QKEN is to discuss data quality and quality management issues. Over 43 active participants from 25 countries and established a network of data quality experts to support EuroGeographics policy towards European data interoperability, to share knowledge amongst members; and promote experiences on quality.

Next, we will present the current geospatial data quality framework developed for communication and assessing spatial data quality at the Expertise Center for Geospatial Data Quality at Wageningen University & Research [Vullings et al, 2015; Meijer et al, ISSDQ 2015]. It is based on the principle of ‘fitness for use’ [report geospatial data quality NCG workshop 26 June 2014 (in Dutch)] and is applicable to all kinds of geospatial data varying from closed to open data, big data and sensor data, to name a few.

Programme

9:30 - 12 AM	*Morning session* *Welcome, introduction*
	Keynote: Essentials of Data Quality and Fitness for Use. Presenter: **Robert Jeansoulin**, Université de Paris-Est - Marne-la-Vallée, co-author of “Fundamentals of spatial data quality"
	Work of the Quality Knowledge Exchange Network (QKEN)., a network established by EuroGeographics. Presenter: **Karin Mertens**, Quality Control Manager, National Geographic Institute, Belgium
	Data Quality assessment procedure as proposed by ECQSD, the Wageningen UR Expert Centre for Quality of Spatial Data. Presenter: **Jandirk Bulens**, Geo Information specialist at Wageningen Environmetal Research (Alterra), Netherlands
	Introduction of the cases of the participants
	Lunch
1:30 - 4 PM	*Afternoon session*
	The Data quality from NMA/academic perspective, Presenter: **Joep Crompvoets,** KU Leuven, Belgium
	Interactive session with use cases
	Discussion, outline of a research agenda

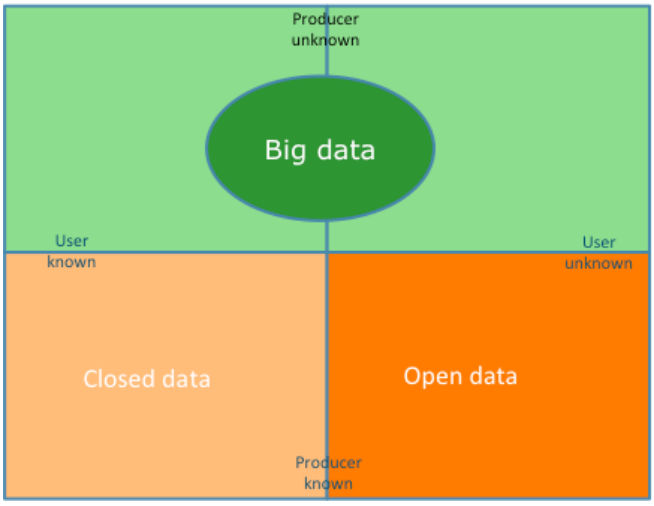
Call for cases

In the workshop we would like to use real world cases. We invite you, as a participant, to submit your case. This may be an existing case which you want to review, or a future case that you want to implement within your own organisation or environment.

A typical case could be:
Is the dataset provided by your mapping agency suited to calculate noise pollution in densely populated areas? In other words: what quality parameters are required for this use case and does the dataset comply?

Background

At the Expertise Centre for Geospatial Data Quality at Wageningen University & Research a geospatial data quality framework is developed for communication and assessing spatial data quality for specific use based on the principle “fitness for use”. This is ongoing work but the approach is presented in this workshop in real world cases. In our view, dealing with geospatial data quality depends on the interaction between the data producer and the data consumer. This is depicted in the following picture:



In case of closed data, the producer and the consumer are known, since the consumer has to apply for the needed data. Because of this the interaction can be complete. When dealing with open or big data the interaction between producer and consumer is much more difficult or directed only from one side since either the producer or the consumer is unknown.

Based on experience with geospatial quality projects (specifying criteria and auditing datasets) we defined this framework and used case studies [Meijer, ISDDQ 2015]to illustrate and specify the framework. The objective of the framework is to bridge the gap between producers and consumers. This refers to the geospatial data quality definition by improving communication at the consumer site (by specifying and elaborating the information needs) as well as on the producer site (by improving access to quality information and understanding of quality aspects of the data).

In the framework the user as a consumer plays a central role, since the consumer and the context of the usage determine the necessary quality (fitness for use). By describing the use case of the consumer we identify the relevant context to be the universe of discourse. The consumer often gives spatial data quality specifications within the identified Universe of Discourse of his or her application domain to his/her best knowledge. But many quality elements can be implicit and not known by the consumer. It is important to unravel the information need into criteria with the help of spatial data quality expertise. Based on this information we define the product that is wanted by the consumer. This can vary from 'plain' data provisioning to automated procedures like an App suitable for providing human services.

When dealing with open data we developed generic use cases to communicate quality elements with yet unknown users with unknown use. In case of big data additional challenges appear because we can state here that even communication in one direction does not exist. With regard to big data and quality we focus mainly on two of the 4 v's [*https://en.wikipedia.org/wiki/Big_data*] namely variability and veracity (uncertainty of data).

Additional publications:

- *Vullings, Wies et al; 2015; Spatial Data Quality: What do you mean?; AGILE 2015 – Lisbon, June 9-12, 2015*
- *Meijer, M et al.; 2015; Spatial data quality and a workflow tool; The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-3/W3, 2015 ISPRS Geospatial Week 2015, 28 Sep – 03 Oct 2015, La Grande Motte*