

Data Quality in Remote Sensing

C. Batini ^{a*}, T. Blaschke ^b, S. Lang ^b, F. Albrecht ^b, H. M. Abdulmutalib ^c, Á. Barsi ^d, G. Szabó ^d, Zs. Kugler ^d

^a University of Milano-Bicocca, Italy - batini@disco.unimib.it

^b Dept. of Geoinformatics, University of Salzburg, Austria – (thomas.blaschke, stefan.lang, florian.albrecht)@sbg.ac.at

^c Dubai Municipality, Dubai, UAE - HUSSEINMA@dm.gov.ae

^d Department of Photogrammetry and Geoinformatics Budapest University of Technology and Economics,
Budapest, Hungary – (barsi, gyszabo, kugler.zsofia)@eik.bme.hu

ISPRS ICWG III/IVb

KEY WORDS: data dimensions, Data quality dimensions, big Earth data

ABSTRACT:

The issue of data quality (DQ) is of growing importance in Remote Sensing (RS), due to the widespread use of digital services (incl. apps) that exploit remote sensing data. In this position paper a body of experts from the ISPRS Intercommission working group III/IVb “DQ” identifies, categorises and reasons about issues that are considered as crucial for a RS research and application agenda. This ISPRS initiative ensures to build on earlier work by other organisations such as IEEE, CEOS or GEO, in particular on the meritorious work of the Quality Assurance Framework for Earth Observation (QA4EO) which was established and endorsed by the Committee on Earth Observation Satellites (CEOS) but aims to broaden the view by including experts from computer science and particularly database science. The main activities and outcomes include: providing a taxonomy of DQ dimensions in the RS domain, achieving a global approach to DQ for heterogeneous-format RS data sets, investigate DQ dimensions in use, conceive a methodology for managing cost effective solutions on DQ in RS initiatives, and to address future challenges on RS DQ dimensions arising in the new era of the big Earth data.

1. INTRODUCTION

1.1 Background

Data Quality (DQ) is of growing importance in Remote Sensing (RS), due to the growing relevance that RS data have in planning and operational decision of public bodies and private firms, and the huge amount of digital services (or apps) that exploit RS data. One may postulate that for many years RS data were difficult to access, expensive and required a lot of knowhow. The ease of collecting, storing, and processing RS data went along with the advent of what some are calling the fourth paradigm of science. Recently, DQ issues have evolved from the area of databases, where relational tables are considered, to the wider set of perceptual and linguistic data representations, such as (perceptual) drawings, maps, images, videos, and (linguistic) structured texts and loosely structured texts. For all such different types of data representations, a comprehensive analysis is provided in (Batini & Scannapieco, 2016) of quality dimensions and metrics that are relevant for the representation; furthermore, techniques and methodologies are provided for the life cycle of DQ management, in the two phases of DQ assessment and improvement. The evolution of dimensions and methodologies from the area of “small” data to the rapidly emerging area of “big data” is discussed as an open problem.

1.2 Aims and scope

In this paper we address several open issues that are currently investigated on how to migrate/adapt/enrich such topics in the area of remote sensing data. First of all, we notice that remote sensing data had since the beginning some of the characteristics

of big data, due to volume and velocity of data acquisition and processing involved in many applications, while RS data are typically structured data that come along with metadata. We will therefore cautiously discuss the questions whether RS are in fact “big data” and will therefore build on earlier work. We particularly refer to the Quality Assurance Framework for Earth Observation (QA4EO; <http://QA4EO.org/>) which was established and endorsed by the Committee on Earth Observation Satellites (CEOS; <http://ceos.org/>) as a direct response to a call from the Group on Earth Observations (GEO; <http://earthobservations.org/>).

At first, (Section 2), we address the issue of data dimensions classification. The efforts of standardization bodies and scholars concern DQ in general in GIS, while there is no consensus in the literature on a stable taxonomy for DQ dimensions and metrics in RS data. We propose a taxonomy to organize remote sensing quality dimensions.

In Section 3 we consider RS applications in which datasets are characterized by heterogeneous representations, including raster images, vector images, georeferenced maps, relational tables. Emphasis is put on outlining the information resources used in organizations towards a common conceptual representation. Second, we assess the quality of information considering such the underlying basic (geometrical, temporal etc.), yet data quality dimensions identified in section 2.

Section 4 analyses DQ dimensions currently in use and their relationships under different use scenarios under the aspect of

* Corresponding author

DQ. Here we discuss models proposed for utility and for the utility assessment process in RS applications.

In Section 5 the economic perspective in the area of RS projects related to DQ improvement is investigated, sketching a methodology for the net benefit optimization of the information resource; here, the optimization refers to economic benefit as compared to the required investment on data quality management. Section 6 discusses the evolution of RS DQ dimensions and methodologies in the new era of big data. We discuss different terms for “big” remote sensing data, including datasets that originate from social networks and media, from the Internet of Things or from mobile computing.

2. A TAXONOMY OF DATA QUALITY DIMENSIONS IN THE REMOTE SENSING DOMAIN

2.1 Data dimensions

As a first issue, we address different topics of data dimensions classification. The efforts of standardization bodies and scholars concern DQ in general in Geographic Information Systems (GIS), while there is no consensus in the literature on a stable taxonomy for DQ dimensions and metrics in RS data. Since we cannot investigate this issue comprehensively due to limited space, we show how remote sensing quality dimensions can be organized in a multilevel taxonomy by enriching a one-level taxonomy proposed in (Batini et al. 2015) and (Batini & Scannapieco 2016). In order to obtain a sound classification, we extend the concept of “structural characteristics” from GIS data in general, where it corresponds to topological, geometrical, metrics, thematic characteristics, to RS data; for example the characteristics of images in remote sensing can be complemented with RS specific data acquisition information e.g. on orbits, flightpaths or atmospheric conditions which provide a ‘second level’ DQ information.

2.2 Data quality dimensions in remote sensing

As a merit of a long-term initiative the GEO QA4EO guidelines define Quality Indicators as “a means of providing a user of data or derived product (which is the result of a process) with sufficient information to assess its suitability for a particular application”. In addition, this information should be based on a quantitative assessment of its traceability to an agreed reference or measurement standard (ideally SI), but can be presented as numeric or a text descriptor, providing the quantitative linkage is defined.

Prior to the systematic discussion of the data quality dimensions, remote sensing itself has to be clarified. Our understanding is, that photogrammetry (both aerial and close-range) as well as laser scanning (both airborne, terrestrial and mobile) are taken as part of remote sensing. Remote sensing is furthermore meant with the used algorithms, techniques to process the acquired data sets. Let us have a look on the quality dimension clusters and go into the details.

The **Resolution** is an outstanding important cluster in remote sensing. *Spatial resolution* describes the pixel size, but it is a base to categorize the sensors, like e.g. very high resolution (VHR) imaging sensors. Speaking also about laser scanning, the point density belongs to this quality description. *Radiometric resolution* gives information about the image quantization: how many different intensity values are represented in the pixels of the captured image. *Spectral resolution* is also a substantial measure in remote sensing, because typically the used images

have more than three channels. Multispectral, hyperspectral or ultraspectral images belong to remote sensing – all are based on this quality measure. *Temporal resolution* controls the usability of a monitoring system; it informs about the frequency of the captured images.

Precision in the cluster **Accuracy** has fourfold meaning in remote sensing: (1) *Geometric or spatial precision* expresses how homogeneous is the image rasterization. It has strong connection to the geometric resolution and depends from the camera or sensor model. (2) *Radiometric precision* of a remotely sensed image informs about the goodness of the intensity values, containing the effects of the atmosphere, the vertical angle of the Sun (its position in the sky), but e.g. in radar imagery the surface permeability, too. (3) *Spectral precision* describes the accuracy of the boundaries of the spectral bands; which can be critical in hyperspectral imaging. (4) *Temporal precision* depends also from the temporal resolution; it gives how the capturing date and time was accurate.

Positional accuracy refers to the connection between image (coordinate system) and map (coordinate system), depending on the amount, distribution and accuracy of the ground control points (GCPs). Subtypes of positional accuracy can be absolute and relative. These positional accuracies can be understood in horizontal and vertical meaning.

Thematic precision can also be defined in the accuracy cluster, but it is dependent from the data processing, e.g. image classification. It speaks about the recognition or classification goodness. *Semantic accuracy* similarly to its GIS representation is a very interesting quality measure; it is based on the thematic precision but is a hierarchical term. (In case of a road for example, the thematic precision covers the features about the road axes, boundaries, drainages etc. Semantic measure would then be informative about any paved road as aggregates of the elementary classified components.) *Attribute accuracy* has strong correlation with the thematic accuracy (in some sense, it is the same: mapping of road types from remotely sensed images, e.g. asphalt pavement vs. block pavement). In stereophotogrammetry a height of a building is an attribute, which depends on the measurement accuracy, image scale, point identification accuracy, orientation parameters – it is a derived accuracy measure. There are further thematic accuracy related measures in remote sensing. The *classification accuracy* (sometimes *correctness of classification*) is a group of measures, where overall accuracy, average accuracy, producers’ accuracy, users’ accuracy and Cohen-kappa are the most widely used terms. These are derived from error (confusion) matrix containing amounts for classified and reference pixels of all analyzed classes.

Temporal validity can be used for a map instead of an image having the meaning how long e.g. a land cover map like CORINE is valid.

In the cluster **Completeness** the *Data completeness* is dealing with the completeness of an image, handling for example the effect of shadowing objects, sun flares on water surfaces or masking out by an object (e.g. propeller of a UAV). *Spatial completeness* is a feature on the area coverage. In photogrammetry (especially in stereophotogrammetry) its 3D version, the *stereo completeness* has extreme importance. In monitoring systems and applications the *Temporal completeness* term features how the taken images represent a complete time series. The *thematic completeness* measure describes the image interpretation quality how the expected and defined classes are evaluated. This feature is important with the use of e.g. multiple classifiers.

The *spatial redundancy* in the **Redundancy** cluster is a value being proportional to the number of overlapping images or to the

overlapping area. By the use of stereo image evaluation or Structure-from-Motion (SfM) algorithm it has great importance. The *temporal redundancy* has correlation to the controllability in time series: how many timely repetitions of the images are in the data set.

The **Readability** cluster has *spatial readability* and *Radiometric readability*; which measure in photointerpretation, how the objects can be identified/separated in spatial and radiometric manner. Both measures have strong connection to the relevant resolution features.

The **Accessibility** cluster has a focus in remote sensing on the sources of the images; the data providers controls how the acquired images can be obtained, speaking about satellite or airborne images. Some “accelerations” are established for remote sensing imagery: the International Charter on "Space and Major Disasters" for example. In monitoring systems, the fast accessibility is critical, but has temporal, technical and legal aspects.

In remote sensing we can speak about *spatial consistency* in the **Consistency** cluster. It represents the quality of image interpretation/understanding: how are the different objects or classes recognized/evaluated integrally. A bridge above a water surface, like river can be detected in pixel-wised manner, but the question is how coherent they are in the output map. This phenomenon has very close to the *thematic consistency*, where the recognition integrity is represented in this way. The *topological consistency* is defined mainly for network-type surface objects, like roads or rivers, where the connection of all atomic segments are rated by this measure. Urban mapping focuses on the built environment objects, where e.g. house-parcel inclusions are described by this feature. The *temporal consistency* is for monitoring again, representing for example the possibility or impossibility of land cover changes in time. Having multiple data sources (even airborne or terrestrial), their integral usage can be qualified by this measure.

2.3 Additional aspects for satellite remote sensing data

Atmospheric conditions particularly influence the usability of optical remote sensing data. As a consequence a major quality dimension of RS data usability is cloud cover. Data quality dimension related to cloud cover can be described as data completeness, *radiometric completeness*. In addition to this, the distribution of cloud cover highly influences the usability of the images in spatial terms. Therefore cloud cover distribution on images can be defined in terms of *spatial completeness*.

Radiometric accuracy of image data is not only influenced by the effects of the atmosphere and the path length but topographic effects of shadows. Regarding optical systems shadows account for radiometric distortions in data. In contrast, topographic effects on active microwave data accounts for a higher geometric than radiometric distortion due to its oblique nature. The latter effect is related to *spatial completeness* of the data.

Another phenomenon related to spatial completeness of remote sensing data is performance related. Sensors often face hardware

problems like sensor failure or malfunctions. It can be both understood as *spectral* and *spatial incompleteness* in data.

Radiometric accuracy is enhanced by in field and on-board sensor calibration procedures in many cases. Accuracy becomes essential especially in sensors where radiometric measurement is assisted by on-board calibration of measured values in each acquired data scan line. Thermal IR remote sensing is a good example for that. Another example is the on-board calibration of optical sensors by deep-space view. In this case radiometric noise reduction of images is the main objective of the procedure.

Further to this satellite sensors have to be geometrically calibrated not only in laboratory but also in-flight after launching enhancing image geometry periodically. Lens distortion, boresight misalignment and time synchronization can be improved from time to time (Jacobsen, 2006). The resulting outcomes influence *spatial/geometrical accuracy* of the images. Moreover, *spatial accuracy* is further influenced by the processing of images. Satellite data can be acquired in different preprocessing contexts. Depending on the processing policy of different provides, data quality can strongly be at variance. Spatial correctness can be measured and enhanced by reference locations, both by end-users or by the data provider controlling spatial accuracy.

Redundancy measure is an essential objective in processing of remote sensing data. The information content of images is significantly higher than the useful information obtained as a result of image processing. Reduction of information is mainly achieved by aggregation of similar element of an image. It is mainly carried out by spectral classification, segmentation or other clustering strategies reducing information content to achieve usefulness of information for end-users. The scale of both *spatial* and *radiometric redundancy* in image processing is highly depending on the objectives of the end-user. Therefore no general metrics can be assigned to measure redundancy.

The processing of hyperspectral images accounts for a great challenge in image information reduction. Several processing chains were developed in the past specifically to assist information extraction and information aggregation in hyperspectral image cubes. Not only the processing but also the visualization of image cubes introduces a type of spectral information reduction itself.

Parallel to this, image layers/spectral channels contain a great amount of information duplication. Different channels can be affected by high spectral correlation depending on the spatial location in the image. Several strategies exist to reduce high correlating information during image processes like principal component analysis. It assists the reduction of information content and aggregates high spectral variation in a reduced amount of spatial layers. This can be defined as *spectral redundancy* of images and can be measured by quantitative means.

	Relational data	GIS	Remote sensing specifics (on top)
Accuracy	Syntactic a. Semantic a. Currency Timeliness Schema a. wrt the model Schema a. wrt requirements	Precision (Spatial, Geometric, Temporal, Thematic) Positional accuracy Absolute position accuracy Relative position accuracy Gridded data pos. accuracy Horizontal accuracy Vertical accuracy Geometric precision Semantic Thematic Quantitative attributes accuracy Non quantitative attributes accuracy Temporal validity Correctness of classification	<i>Here, several categories are currently identified at a 'meta-level' according to data capture (orbit/flightpaths, atmospheric conditions, 3D information in relation to the IFOV)</i>
Completeness	Value completeness Tuple completeness Attribute complet. Relation completeness Schema completeness	Data Completeness Schema Completeness Spatial Completeness Temporal Completeness Thematic Completeness Spatial Completeness Temporal Completeness Thematic Completeness	<i>Very difficult topic: Can RS data ever be "complete"?</i>
Redundancy	Schema minimality Schema normalization	Map Schema minimality Map Instance minimality Thematic minimality	<i>Very difficult, yet specific topic: redundancy is partially necessary for RS data, e.g. for stereoscopic images</i>
Readability	Schema readability Instance readability	Map Schema readability Map Instance readability Thematic readability	
Accessibility	Web site accessibility Accessibility to disabled persons	Web Site accessibility Accessibility to disabled persons	
Consistency	C.. through integrity constraints C through edits	Spatial Logical Conceptual Domain Format Topological Temporal Thematic	<i>Very important topic which includes some RS specifics in respect to orbits/flightpaths</i>
Trust of Sources	Provenance Lineage	Veracity (Objectivity, Truthfulness, and Credibility), Trustworthiness Provenance Lineage	<i>Maybe the aspect with the highest degree of specificity of RS data</i>

Table 1. Data Quality Dimension in GIS and RS: This table is mainly based on Batini & Scannapieco 2016. The last column is only a first and indicative version of ongoing work of this ISPRS working group and will be replaced for the final print version).

3. DATA QUALITY FOR REMOTE SENSING: SPECIFICS

In most real world applications that use RS data, datasets are characterized by heterogeneous representations, such as raster images, vector images, georeferenced maps, relational tables, etc. Batini & Scannapieco (2016) discuss how to extend methodologies for DQ assessment and improvement (for a detailed comparison of such methodologies see (Batini 2009)) to take into account sources made of heterogeneous information types. In (Batini 2016) semi structured data are considered, but the approach can be followed also for other data types. The main idea underpinning the extended methodology is to determine the information resources used in an organization or community to a common conceptual representation and then to assess the quality of information considering such homogeneous conceptual representation.

When revisiting the general concepts in section 2, it may not be obvious why remote sensing (RS) data quality needs to be discussed separately. We argue that it is not about the data dimensions as seen in computer science (and as depicted in table 1). Rather, the physical dimensions of the data derived through remote sensing platforms bear specifics which are typically classified in regard to a) platforms (orbits, flightpaths) and b) sensors/cameras used. The remotely sensed data thus, in the majority of cases are stored in regular or irregular raster format types or points (LiDAR) that are to be processed further. Again, while the data are not different from a Computer Science point of view, information about the data generation process can make a difference!

We may refer to the quality as a measure of the difference between the data and the reality that they represent. Under such a view it may become clearer that the data and the corresponding reality diverge (Goodchild 2006) and can we relate the quality of remote sensing data to the following major factors:

1. The relations of the produced raster or primitive vector / raster data sets to the reality presented
2. If the data set doesn't present the reality in a direct manner, then its relation to the product / products which consequently present the reality is considered as an important quality factor.
3. The fitness of use; that is the level of details the model should comprise to present the level of details from the reality
4. The level of details from the reality shall be decided according to the purpose / purposes, or anticipated future goals
5. The imbedded data richness that is comprised by an image, including the metadata.

Further, a quality model replicating the reality should encompass all aspects of quality, covering a large spectrum and affects the entire process of the acquisition, management, communication, and use of geographic data (Devillers & Jeansoulin 2006). Herein two major ('super') levels are realized, namely the data as such and the process of deriving the data and meta-information associated to such a process. Quality Assurance thus shall insure the anticipated quality of the product or the data, whereas Quality Control shall provide the measure of closeness of the model elements presented by the data to the element of the sought reality, which should fall within the range set in relation to the fitness of use, and the quality of details reflected by how accurately the elements of the reality model can be presented with minimum uncertainty. Furthermore, the uncertainty provides information regarding the robustness and usefulness of the model quality. Whereas, the quality affects decisions based on the model or its

analysed data or perhaps information, decisions maybe regarded as the conclusion of an informed and logical process, in which the treatment of uncertainty must be present (Devillers & Jeansoulin 2006).

The described elements of (RS) data quality now can be used to describe the strength of both the operations to collect and build the data, and the model or the data itself. Examples include: different types of accuracies (see also table 1); completeness; consistency; precision and many more. For those RS specific DQ elements, basic geometrical; attribute and other model descriptions can be utilized. The model representing a specific reality related to a certain use, can be assessed in a similar way like assessing maps using standards, but to firmly assess the model It would be advisable and important to recognize that there is a tacit correspondence between the technically possible, the economically viable, and the utility for the intended uses (Chrisman 2006).

4. DATA QUALITY DIMENSIONS

4.1 Data quality dimensions in use

In RS applications it is frequent that, either for planning or for operational purposes, an organization or an individual has to take a decision that, according to its outcome, may lead to different possible utilities. Since the decision, among other factors, is influenced by the accuracy, completeness, currency, etc., of the available information, its final utility is indeed influenced by the quality of information.

The users of RS information have different requirements to the quality of the information depending on the purpose for which the RS data shall be used. Thereby, they define which subset of the possible quality dimensions becomes relevant. The QA4EO guidelines provide a framework that allows the development of an according documentation of the RS DQ that enables a user to judge the information's 'fitness for purpose', with the appropriate set of quality dimensions.

Applications can range from low to high levels of requirement to DQ. Low levels occur when the RS DQ is predominantly assessed by a subjective judgement of the user. An example is visualisation as the main purpose of the RS data like in the illustrated book 'Untouched Nature' (Eisl et al. 2013) that contains views of satellite images from many locations distributed over the world. Nevertheless, some readers may have an interest in information about the DQ, e.g. accurate location information and date of acquisition as documented in the book, if they want to search for more information about one of the presented RS datasets. Even if the general level of DQ requirement is low, a certain amount of DQ documentation can be valuable for users that want to follow an advanced purpose. Here the requirements to a DQ documentation can be considered as optional.

The other end of the range with high levels of requirement to RS DQ and associated documentation occurs where the user's purpose is the utilisation for safety-relevant decisions. Such high requirements to DQ generally occur when the consequences of using faulty information have a strong negative impact. Terrain information for aviation is an example of RS data where high standards of DQ are essential and according certification procedures are in place (Albrecht et al. 2014). Certification procedures of the European Aviation Safety Agency EASA clearly define what specific purpose the RS derived information serves, which quality levels the product has to adhere to and how its quality has to be documented (EASA 2014, EASA 2013).

Consequently, RS DQ and the need for its documentation with according dimensions depends on the specific purpose of the user and, where relevant, on existing standards in the user community.

4.2 Metrics of quality dimensions

We need to discuss models proposed for utility and for the utility assessment process in RS applications, observing that the definitions themselves of quality dimensions and metrics are deeply influenced by utility. Indeed, besides dimensions and their metrics that are called *objective* in (Batini & Scannapieco 2016), we discuss other dimensions and metrics whose definition and process of measurement inherently depend on the context of use of information, resulting in a new class of *contextual* dimensions and metrics.

5. DATA QUALITY IMPROVEMENT: ECONOMIC PERSPECTIVE

While in a perfect world all relevant dimension and the respective metrics mentioned above should be taken into account in a real world scientists working with RS data need to balance and trade off perfection and effort, i.e. costs. This may be very obvious when distinguishing precision and accuracy.

We may – very briefly – reason about the economic perspective in the area of RS projects. DQ standards in general are usually independent from an economically feasible perspective. They are standards and shall accommodate all instances. Nevertheless, DQ improvement is not a theoretical / hypothetical task.

In the following subsection we sketch out a methodology for the net benefit optimization of the information resource; here, the optimization corresponds to the maximization of economic benefits in comparison with costs of DQ management.

While in a perfect world all relevant dimension and the respective metrics mentioned above should be taken into account in a real world scientists working with RS data need to balance and trade off perfection and effort, i.e. costs. This may be very obvious when distinguishing precision and accuracy.

We may – very briefly – reason about the economic perspective in the area of RS projects. Data quality standards in general are usually independent from an economically feasible perspective. They are standards and shall accommodate all instances. Nevertheless, data quality improvement is not a theoretical / hypothetical task.

In the following subsection we sketch out a methodology for the net benefit optimization of the information resource; here, the optimization corresponds to the maximization of economic benefits in comparison with costs of data quality management.

Modelling the costs and benefits of any type of information is a critical task at global, regional and local level (Shapiro & Varian 1999) included the remote sensing data.

To identify and describe the most important variables of the set-up, operation, maintenance and service cost determined by the application sector, the stakeholder position, the technology and on the benefit side of the equitation by the economic, social and political value to increase the efficiency, to support the democracy (Codagone 2006, Longhorn & Blakemore 2008.).

It is a problematic question to define measurable indicators for cost/benefit modelling (Genovese & Roche 2010).

In the private sector the commercial cost/benefit of RS data for commercial vendors can be measured by the production cost, recovery, service, profit margins and return on investment. For the data user the value of the RS data typically reflected in the price that the consumer pay for the offered service.

In the public sector the commercial value for data owners and vendors evaluate by cost savings of service. For the data user has less quantifiable direct and indirect value of the usage of the RS data. But in the public sector there are big socio-economics and political value of the RS data for the whole society as a better decision-making, support the sustainable environment.

Attributes of RS data value:

Value of location attribute: location attribute provides spatial context to other attributes.

Time dependency value: time related value of data whether it is real-time requirement or historical.

Value due to legal or other mandatory requirements: the information is given an official or legal status for certain type of transaction.

Value due to network: some information has added network because it is used by large number of people.

Value due to quality of an information source: the factors related to the production of RS data: context, quality, timeliness, accuracy, completeness, history, etc.

Value determined by cost savings: reduce the duplication in data collection and transaction costs.

Cost of RS products:

Transaction cost: refer to the cost of measuring valuable attributes of what is being exchanged and the cost of protecting rights.

Data collection cost: is a very costly procedure represent a high percentage the total cost of producing RS data, includes the production and transformation costs.

Value pricing:

User's valuations and preferences: the producer sets the price according to the buyer's needs, related to the preferences of the potential users, introduced product differentiation for particular target groups, avoid price dispersion.

For the private organisations the revenues minus the cost give the added value, net benefits of RS data. For the public sector prices of RS data may be substantially below cost, but the final benefit must include the intangible benefits, which calculation is very complicated.

6. RS DQ DIMENSIONS AND METHODOLOGIES IN THE ERA OF BIG DATA

Do these RS DQ dimensions and methodologies debated so far require changes in the new era of 'big data'? (see Mayer Schönberger 2013 for a comprehensive discussion of the paradigm shifts from small to big data)

As briefly discussed, RS has always dealt with large amounts of data – although typically structured data and in most cases associated metadata. For many scientists this is not the same as the metaphorical search for the 'needle in the haystack' we discuss different terms for "big" remote sensing data, including datasets that originate from social networks and media, from the Internet of Things or from mobile computing.

Most recently, the term “Big Earth Data” has been introduced (see e.g. Guo 2013). Particularly in China, the notion of ‘Digital Earth’ has been promoted intensively, e.g. by the Chinese Academy of Sciences and the Society of Digital Earth. Still, it seems that – typical for new terms, maybe paradigms – a terminological inconsistency exists around big Earth data and whether or not this may go far beyond Earth observations (sometimes even reduced to remote sensing); it may reach out to large sensor nets on ground, in the air, in the ocean or a combination of all. Still, some authors mainly refer to ‘big Earth data’ particularly in the course of massive free data, Landsat, and in particular Sentinel.

A variety of “sensors” and “sensor networks” can be used to systematically assess and monitor dynamic geographic phenomena at different spatial and temporal scales. However, the monitoring is typically done for each phenomenon individually (e.g., for air temperature or mobility). In order to enhance—or at least verify—our understanding of both environmental and social processes for multidisciplinary studies, a more holistic monitoring framework is needed. One way to fully integrate the spatiotemporal dynamics of both environmental and social phenomena is the “adaptive geo-monitoring framework” (Blaschke et al. 2011, Sagl & Blaschke 2014) by adding the spatial dimension and the mutual context-awareness of dynamic geographic phenomena.

This short article aimed to bridge the views of the remote sensing community and the database community. In particular, the authors followed the work of Batini et al. (2009, 2015) and Batini and Scannapieco (2016) while comprising the benefits from the efforts of the remote sensing community, especially of CEOS and IEEE. Likewise, a next logical step will be to inform the computer science community about the efforts of the remote sensing community regarding data quality issues.

In particular, this ISPRS Working Group will discuss whether to extend the DQ dimensions to further processed data, and derived products as well. QA4EO and other quality guidelines (including the validation protocols of Copernicus core services) also address the entire processing chain, and uncertainty propagation through it. This opens new research avenues and needs going much beyond RS and GIS communities and will, for instance, necessitate to map quality from a cartographic point of view, or to ensure OGC compliance of generated vector data sets.

REFERENCES

Albrecht, F., F. Lipok, J. Dittrich, V. Logemann 2014. *Geo-Information in the Cockpit - Safe Emergency Landing in Alpine Terrain*. CERGAL 2014: International Symposium on Certification of GNSS Systems & Services, Dresden, Cuvillier.

Barna S., D. Srivastava, 2014. DQ: *The other face of big data*. Data Engineering (ICDE), 2014 IEEE 30th International Conference on. IEEE, 2014.

Batini, C., Cappiello, C., Francalanci, A., Maurino 2009. Methodologies for Data Quality assessment and improvement. *ACM Comput. Surv.*, 41(3): 16:1-16:52.

Batini C., A. Rula, M. Scannapieco, G. Viscusi, 2015. From DQ to Big DQ. *J. Database Manag.* 26(1): 60-82.

Batini C., M. Scannapieco, 2016. *Data and Information Quality - Dimensions, Principles and Techniques. Data-Centric Systems and Applications*, Springer 2016, pp. 449.

Blaschke, T., Hay, G.J., Weng, Q., Resch, B. 2011. Collective Sensing: Integrating Geospatial Technologies to Understand Urban Systems - An Overview. *Remote Sensing* 3(8): 1743-1776.

Chrisman, N. 2006. Development in the Treatment of Spatial Data Quality. In R. Devillers & R. Jeansoulin, editors, *Fundamentals of Spatial Data Quality*. Newport Beach, CA: ISTE, pp. 21-28.

Codagone C., Boccardeli P., Leone M. I. 2006. *Measurement Framework Final Version*, EC DG Information Society and Media eGovernment Unit.

Devillers, R. & R. Jeansoulin 2006. Spatial Data Quality. In R. Devillers & R. Jeansoulin, editors, *Fundamentals of Spatial Data Quality*. Newport Beach, CA: ISTE, pp. 17-20.

EASA, “*Certification specifications*” Retrieved 30 April, 2014. from <http://easa.europa.eu/agency-measures/certification-specifications.php>

EASA, “*Certification Specifications and Acceptable Means of Compliance for Large Aeroplanes*”, Amendment 14, Dec 2013, <http://easa.europa.eu/agency-measures/docs/agency-decisions/2013/2013-033-R/Annex%20to%20ED%20Decision%202013-033-R.pdf>

Eisl, M., G. Mansberger, P. Schreilechner 2013. *Untouched Nature - Naturlandschaften in Satellitenbildern*, eoVision.

Genovese E., Roche S., Caron C., Feick R. 2010. The EcoGeo Cookbook for Assessment of Geographic Information Value, *Int. Journ. of Spatial Data Infrastructures Research* 5: 120-144.

Guo, H. 2013. Digital Earth: Big Earth Data. *Int. J. Digital Earth* 7:1-2.

Goodchild, M.F. 2006. Foreword. In R. Devillers and R. Jeansoulin, editors, *Fundamentals of Spatial Data Quality*. Newport Beach, CA: ISTE, pp. 13–16.

Jacobsen, K., 2006. Calibration of imaging satellite sensors. *Int. Arch. Photogramm. Remote Sensing*, 36: 1.

Longhorn R.A. & Blakemore M. 2008. *Geographic Information Value, Pricing, Production and Consumption*, CRC Press, Boca Raton.

Lukoianova, T., V. Rubin, 2014. Veracity Roadmap: Is Big Data Objective, Truthful and Credible? *Advances In Classification Research Online*, 24(1).

Mayer-Schönberger V., K. Cukier 2013. *Big Data: a Revolution that will transform how we live, work and Think*. Houghton Mifflin, New York.

Sagl, G. & Blaschke, T. 2014. Integrated Urban Sensing in the 21st Century. In: Weng, Q. (Ed.). *Global Urban Monitoring and Assessment through Earth Observation*. Taylor & Francis/CRC Press, Boca Raton, FL, USA, 269-286.

Shapiro C. & Varian H.R. 1999. *Information Rules*, Harvard Business School Press, Boston.