PROVIDING DATA QUALITY INFORMATION FOR REMOTE SENSING APPLICATIONS

F. Albrecht [a] *, T. Blaschke [a], S. Lang [a], H. M. Abdulmutalib [b], G. Szabó [c], Á. Barsi [c], C. Batini [d], A. Bartsch [e], Zs. Kugler [c], D. Tiede [a], G. Huang [f]

[a] Dept. of Geoinformatics, University of Salzburg, Austria – (florian.albrecht, thomas.blaschke, stefan.lang, dirk.tiede)@sbg.ac.at
[b] Dubai Municipality, Dubai, UAE – husseinma@dm.gov.ae
[c] Department of Photogrammetry and Geoinformatics Budapest University of Technology and Economics,
Budapest, Hungary – (szabo.gyorgy, barsi.arpad, kugler.zsofia)@epito.bme.hu
[d] University of Milano-Bicocca, Italy – batini@disco.unimib.it
[e] Austrian Polar Research Institute, Austria – annett.bartsch@polarresearch.at
[f], Chinese Academy of Surveying and Mapping, Beijing, China – huang.guoman@casm.ac.cn

ISPRS ICWG III/IVb

KEY WORDS: validation procedures, data quality dimensions, remote sensing lifecycle, quality criteria, disaster management, standardisation, ISO 19000

ABSTRACT:

The availability and accessibility of remote sensing (RS) data, cloud processing platforms and provided information products and services has increased the size and diversity of the RS user community. This development also generates a need for validation approaches to assess data quality. Validation approaches employ quality criteria in their assessment. Data Quality (DQ) dimensions as the basis for quality criteria have been deeply investigated in the database area and in the remote sensing domain. Several standards exist within the RS domain but a general classification – established for databases – has been adapted only recently. For an easier identification of research opportunities, a better understanding is required how quality criteria are employed in the RS lifecycle. Therefore, this research investigates how quality criteria support decisions that guide the RS lifecycle and how they relate to the measured DQ dimensions. Subsequently follows an overview of the relevant standards in the RS domain that is matched to the RS lifecycle. Conclusively, the required research needs are identified that would enable a complete understanding of the interrelationships between the RS lifecycle, the data sources and the DQ dimensions, an understanding that would be very valuable for designing validation approaches in RS.

1. INTRODUCTION

The last few years have seen a growth in availability and accessibility of remote sensing (RS) data. Several of the Sentinels, the Earth Observation (EO) satellite fleet of Copernicus, have become operational with more to come (European Commission, 2016). The development of Data and Information Access Services (DIAS) is on the way that will provide cloud-processing platforms as part of the Copernicus downstream services, and a large industry provides and develops new information products and services (ibid.). This leads to a larger and more diverse user community for RS data: new ideas for applications are implemented as shown by success stories from different initiatives, e.g. the Copernicus Masters competition (https://www.copernicus-masters.com/) and the users' testimonials on operational satellite applications gathered by Eurisy (https://www.eurisy.org/good-practices.php). The raised awareness of the benefits of RS data may help to address new research questions in a diverse set of scientific disciplines. Many of them are organized in the Group on Earth Observations (GEO) community activities and, with a global focus, target the UN 2030 Agenda for Sustainable Development, the Paris Agreement on Climate Change and the Sendai Framework for Disaster Risk Reduction (Group on Earth Observations, 2017). Researchers and developers accordingly have to come up with quality criteria to assess the suitability of their information products for the designated purpose. The increased accessibility of data of diverse sources and data types (Big Data), including RS data, has the potential for integration within new applications and needs coordinated validation approaches for assessing data quality (DQ). An opportunity for enhancing existing approaches for guidance arises from synergies between initiatives in EO following a processing view on providing information about data quality and initiatives in the domain of Geographic Information Systems (GIS) providing a structural and descriptive approach. The potential for synergies is investigated in the following with an application oriented perspective on RS DQ.

With the thorough investigation of remote sensing data quality, ISPRS ICWG III/IVb follows the same strategic objectives as previous/parallel initiatives from IEEE, CEOS and GEO, e.g. with the Quality Assurance Framework for Earth Observation (QA4EO; http://qa4eo.org/). These are to enhance trust in EO-derived information and to prevent wrong decisions based on EO-derived information by ensuring proper usage of EO data through the entire information production process. Ultimately, this will lead to increasing the benefit of RS for the society and the environment. Consequently, the existing initiatives, with the QA4EO guidelines in particular, offer a flexible approach of designing validation procedures that allow the effort for ensuring compliance to be commensurate with the final objectives. A

---

* Corresponding author

validation procedure assigns tasks for the quantitative assessment of uncertainty resulting in quality indicators (QI) for the outcome of a RS information production process. The used QIs ought to be based on the International Standards Organization (ISO) guide to the expression of uncertainty in information. Ultimately, it is for the final user (customer) of the information to determine if the information, with its associated QI, is suitable for their requirement. They must therefore have the ability to trace back to the origin of any resultant QI, or in practice have confidence that those responsible for accepting the results of each of the earlier steps in any process chain have undertaken an adequate review.

In the GIS domain, there are taxonomies available for data quality dimensions (QDs) that can be enriched for RS data (Batini et al. 2017). The concept of "structural characteristics" from GIS data in general comprise topological, geometrical, metric and thematic characteristics. Their application to RS data result in the following QDs: resolution, accuracy, completeness, redundancy, readability, accessibility, consistency and trust of sources. These dimensions are the categories that contain the available set of QIs that can be used for the validation of RS products within a specific application context. When designing a validation approach for a specific application, a set of Quality Criteria (QC) is designed based on the QIs by selecting appropriate uncertainty assessment methods and quality thresholds. The QC are the methods to measure the QDs.

In order to provide suitable guidance in designing validation approaches, there is a need for a better understanding of the role that QCs fulfil in the RS lifecycle and of the way how available standards accommodate this role. We investigate the RS lifecycle for highlighting how QC support decisions that guide the RS information processing chain towards producing geospatial information in alignment with the user requirements. This investigation enables us to relate the QC of the RS lifecycle to the QDs for RS data. Subsequently, we present an overview of the available standards that are valuable for the design of validation approaches for RS. Finally, we conclude with an investigation of the present gaps and the need for further research.

2. THE DATA QUALITY DIMENSIONS TAXONOMY WITHIN THE REMOTE SENSING LIFECYCLE

There is a strong interrelationship between RS QDs, the phases of the RS lifecycle and the types of data sources (Figure 1). Various data sources are processed within the RS lifecycle. Each of them requires a set of RS QDs that guides the phases of the RS lifecycle. In this research, we first introduce the (types of) data sources, then present the phased RS lifecycle, subsequently describe the RS QDs and finally we discuss the relationships among data sources, lifecycle and QDs.
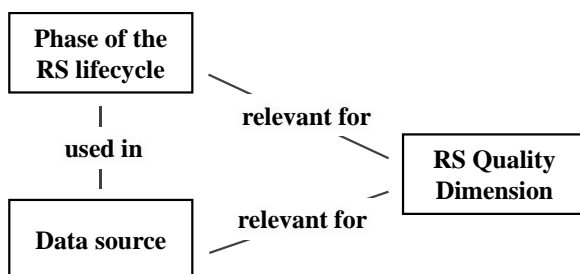

Figure 1: Dependencies between the RS lifecycle, data sources and RS Quality Dimensions.

The **data source** depends on the used sensors and provides the input data in an associated RS data type. As a complete categorisation of RS data types merits a publication on its own, the subsequent examples only serve as a brief introduction. There are active and passive sensors. Optical EO data originates from passive sensors and contains measurements of reflected (or emitted) energy signals stored in RS images, i.e. a raster with pixel values. RAdio Detection And Ranging (RADAR) data originates from active sensors and contains measurements of the backscatter. In case of Synthetic Aperture Radar, these signals are stored in RS images, as well. LIDAR active remote sensing technology emits laser pulse and measures the reflected pulse. RS analysis can happen on single RS images or RS time series but also on RS image cubes where a series of images (e.g. repeatedly collected for the same region) are integrated into one data structure.

The **RS lifecycle** structures all possible phases of the data production process, from its beginning of the data's coming to existence (that includes the sensor design prior to data collection) over storage, processing and use to archiving and deletion. As a representation of the RS lifecycle we adopt a *simplified* RS information production process (Figure 2).
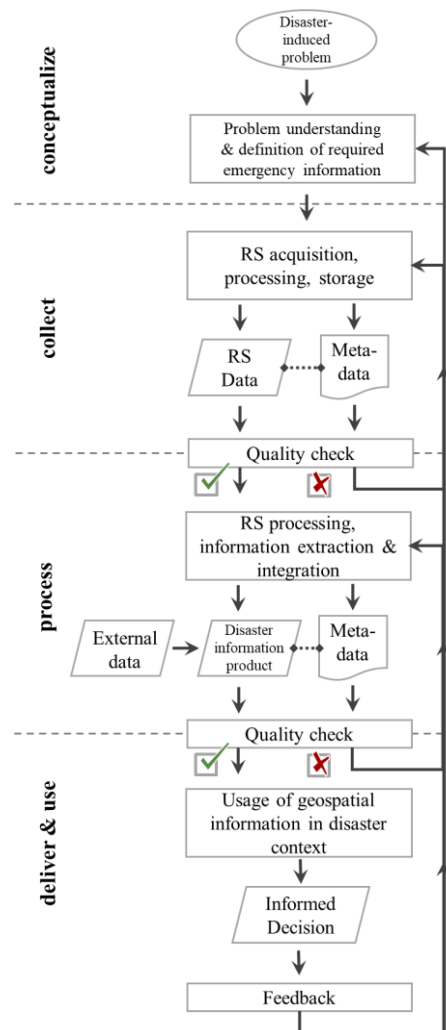

Figure 2: The RS information production process where delivery of RS data products is controlled by quality checks enhanced with data quality information stored in the metadata. The application context (disaster response) serves as an illustration, the process is otherwise considered generic.

The process is 'simplified' because some phases of the RS lifecycle have been excluded. In return we can cover the included phases in depth in order to highlight the way how QCs support decisions and guide the RS information production process. Despite the process description in Figure 2 is meant to be generic, for illustration purposes it refers to a specific application scenario in the context of emergency mapping in response to or for relief from natural or humanitarian disasters (Lang et al., 2017; Lang et al., 2018). Several operational services that target disaster mapping have formed the International Working Group on Satellite based Emergency Mapping (IWG-SEM) for developing common workflow descriptions. The addressed workflows also include the operational scenario of refugee camp mapping, to which the process description will occasionally refer.

The presented RS lifecycle describes the 'traditional' offline RS processing approach where a main operator, i.e. the RS analyst, implements most of the phases that may be supported by additional operators, e.g. the RS data provider or the distributer that delivers the RS information product to the beneficiary. Therefore, the phases along the RS information production process may not be under the control of a single operator but a chain of operators. The implementation of the RS information production process by multiple operators requires them to communicate data quality information and feedback in both directions and from end to end along the process. The QC are the basis for decisions in quality checks that guide the process.

Although the RS lifecycle as presented in Figure 2 is still practiced widely in the entire RS domain, it is subject to change, e.g. by improved tools for specific sub-phases that increase the share of automation and may even remove certain sub-phases from the hands of a human operator. For example for pre-processing, the sub-phases for radiometric calibration, resolution merge and ortho-rectification used to be part of the RS analyst's work. Today, they are included in highly automated processes of the EO data provider for providing products of a higher processing level. Current research in cloud (online) processing and in artificial intelligence (AI) algorithms may lead to significant change of the RS analyst's workflow that becomes more centred on 'information fusion' and information management. Despite the to-be expected changes, the presented simplified RS information production process is suitable for illustrating the relationship to RS Quality Dimensions.

The RS information production process leads a RS analyst (and supporting operators) through several phases that can be named 'conceptualize', 'collect', 'process' and 'delivery & use'. At the beginning of the RS information production process stands a specific problem to solve where information from RS data can be of value, because an appropriate RS data source is available that can detect relevant properties of the objects/features of interest. In some application scenarios, such as the mentioned emergency mapping, remotely sensed data may be the only (reasonable, accessible, economically feasible, etc.) data source and thereby considered highly mission-critical. The RS analyst has to invest resources in problem understanding and the setting of an objective for the RS workflow. This includes the compilation of a user requirements definition that clearly defines the outcome of the RS workflow. In addition, it contains the definition of system requirements that specify which RS data types and associated RS data properties qualify for meeting the user requirements. For example, if the objects to detect for the RS information user are tents and other dwelling types in a refugee camp, RS data that qualify for an effective detection require a *minimum spatial, spectral* and *radiometric resolution*, and in case of a monitoring of a refugee camp a *time resolution*.

The next phase of the RS information production process deals with data collection. For collecting the input RS data for their analysis, the RS analyst either queries images in RS data archives or orders images via an EO data tasking service. A RS analyst decides that RS data qualifies for further use if they can expect with sufficient certainty that they will be able to produce the geospatial information required by the RS information user. When deciding on an RS dataset, the RS analyst's judgement considers the user requirements (e.g. spatial and spectral resolution, temporal validity, area of interest AOI) as well as software routines and other available resources. This has implications on the data source, whether ordered from archives or being freshly acquired, i.e. tasked. Within EO data archives, the image QC serve as an input for querying images that fulfil the needs. In the new era of big EO data with archives like that of Sentinel-2 growing every day, more intelligent routines of image querying may become more and more important (e.g. Tiede et al., 2016). In EO tasking services, the image QC have the role of specifying the requirements placed in the order.

When browsing through RS archives or when tasking a RS image acquisition, among the main criteria that qualify a specific RS image for further use are its *temporal validity* and the *AOI completeness* that the RS image provides. For the sake of simplicity, we ignore other potentially relevant criteria for now. Search interfaces for EO data archives (USGS Earth Explorer, Copernicus Open Access Hub, Sentinel Hub provided by the company Synergise on the Amazon cloud, GeoStore of Airbus D&S, etc.) allow to specify an AOI as a spatial constraint. The search interface also allows giving a period as a temporal constraint that further reduces the set of potentially qualifying images in the database view. The *percentage of AOI area covered by a specific RS image* is a first indication of AOI completeness. In addition, the metadata provides a few valuable initial RS image quality indexes such as the *cloud cover percentage*. The quality of the values stored in the metadata depends on the underlying calculation methods used by the EO data provider. The cloud cover percentage indicates the amount of area in the RS image extent that is covered with clouds and therefore cannot provide information about the Earth surface conditions. Consequently, this RS image property affects the AOI completeness. Unfortunately, in the general case, the AOI and the RS image extent are not congruent. As the cloud cover percentage usually is reported as a global EO image parameter and as other difficult areas like the area of cloud shadow are disregarded, the expressiveness of this parameter as a QC for AOI completeness is limited. Depending on the size of the AOI, even a low cloud cover percentage may result in a large AOI share where the visibility of the Earth surface is blocked out. In lack of a more appropriate QC, the RS analyst will have to live with some uncertainty that their query may include unsuitable RS images or may exclude suitable RS images. Without additional tools, an improvement of their search result requires additional workload for visually checking each single RS image of a larger set of images identified with a relaxed cloud cover percentage. A more appropriate QC for the RS analyst would be an on the fly calculation of *AOI cloud cover percentage*, based on a RS image cloud mask. The cloud mask would need to be stored in the database of the RS data archive as a georeferenced and queryable spatial layer. The availability of such an operation within the search interface of the RS data archive would save time for the RS analyst. In a similar fashion, the integration of multiple cloud masks would be able to verify AOI completeness for a time series analysis. By checking each of the RS images in the database view on their AOI completeness with the available QCs, the RS analyst comes closer to understanding whether he will be able to obtain

appropriate RS data that can serve the user requirements to RS-derived information.

If the RS data archives cannot provide an RS dataset with adequate AOI completeness or *temporal match to a preferred acquisition date*, the tasking of an RS image acquisition may become an option. Tasking services require the customer to set an acceptable level of AOI cloud cover percentage for an ordered RS image. Lower levels of cloud cover may result in the need for an extended image collection period to ensure successful image acquisition. The RS image acquisition with low enough cloud cover percentage may take a long time, e.g. of sometimes more than half a year, especially in tropical regions where suitable weather conditions are sparse (Eberhardt et al., 2016). An alternative in such cases may be provided by RADAR data as its signals can penetrate clouds. For the production of RS-derived information within time, a trade-off in AOI completeness may be necessary. The possibility depends on the application case of the user. If for example, the RS analysis shall provide summarized information about the land surface of the AOI, an acceptable amount of uncertainty may result from the incomplete coverage of the AOI. However, if the information product requires that all locations within the AOI have been checked, the incomplete coverage may have a strong impact on the operations of the user. Obviously, there exists an interrelationship of chained quality requirements that apply to the service definitions of the RS data provider (to sell data), the RS analyst (to sell information products) and the RS information distributer (to meet information demand and to do reporting) at the same time. As a user's application context may be very specific, data demands may result for the RS data provider that would be achievable in theory but are very difficult to match with existing service definitions that have been developed with different users/customers in mind. Here, the business model of operators along the RS information production process comes into play. Even if permanently dropping, the current VHR EO data costs may still be prohibitive for some application scenarios. In addition, the business model includes further parameters like a minimum AOI size and maximum complexity of AOI (by number of vertices), as well as tasking priority, that ensure EO return of investments for data providers. Further information provides Batini et al. (2017) that discuss the cost-benefit relation in context of an economic perspective on RS DQ improvement.

The collection of RS input data can be a cyclical task where additional data is needed when the data search was not successful. Once the RS analyst is satisfied with the obtained EO dataset, they would move on to the next RS lifecycle phase of processing the RS data and producing a geospatial information product. Any pre-processing done at the EO data provider's side will need to be verified, for example the *spatial accuracy* dimension, i.e. the alignment of the RS data with external geospatial data to be used in subsequent processing. For spatial accuracy, a common metric is the *root mean squared error (RMSE)*. If necessary, the RS analyst can use ortho-rectification tools for correcting low spatial accuracy and improving the alignment with external data. Another potentially relevant pre-processing example concerns routines for image correction, i.e. image calibration, radiometric/atmospheric correction, topographic correction, etc. that transform the digital numbers of pixels (the radiance registered by the sensor) into radiance emitted/reflected by the target. The corresponding QC would be the corrected radiance value's *maximum deviation from the absolute radiance value*, acquired by reference ground

measurements/spectroscopy. The opportunity for the RS analyst to consider such a QC depends of course on the availability of reference data or qualifying substitutes.

Subsequently follows the information extraction. For the case of mapping a refugee camp, tents and other types of dwellings are identified in the RS data. Sticking to the example of refugee camp mapping, the RS analyst may use a (semi-)automated classification algorithm and assign specific threshold values on the parameters of a rule-set. Applied to the RS data, the algorithm generates a digital map of the camp area and reveals a certain number of dwelling categories. Based on external information about mean occupancy rates per dwelling type, the user is able to estimate the total number of inhabitants of the refugee camp. This is a critical information element shown by the map, valuable for aid organization providing humanitarian assistance. As the information extraction process contains multiple levels of uncertainty, the RS analyst performs an assessment of the *thematic accuracy* that compares the dwellings map to reference data to estimate the amount of *errors of omission and commission*. A common metric, calculated from the contingency matrix, is the *overall thematic accuracy* that measures the share of correctly classified dwellings in the map. As error propagates further in the RS information production process, the overall accuracy is a proxy for the reliability of population figures, which otherwise depend on constants (i.e. the occupancy rate per dwelling types). By the additional provision of uncertainty information the user organisation can take better informed decisions on any interventions to be taken in the camp. Later on, the map of the refugee camp may serve further purposes, e.g. for deciding where to locate specific facilities in relation to the population distribution within the camp.

When the RS information product has been produced, the RS analyst delivers the final geospatial information product to the distributor or user. Beyond QC related to structural properties of RS images and RS-derived information products (as handled by Batini et al. 2017), the implementation of a service that provides RS-derived information on a regular basis introduces process-related image QC like the *timeliness of information provisioning*. While the RS analyst is in control of the time they spend on producing RS-derived information from RS images, the time to acquisition is outside of their control and resides with the EO data provider. Especially for VHR satellite images, the time to delivery is part of the business model, i.e. fast acquisitions usually cost more and the customer may have to accept high cloud cover percentages. For the case of refugee camp mapping, timely arrival of map information may strongly support the decisions in planning facilities for humanitarian assistance. Therefore, a higher price for EO data acquisition can be justified.

Finally, the use of the RS information product within the tasks of the RS information user will show its actual value. Therefore, feedback from the usage of the RS information product is very valuable for informing all production steps, including the improvement of the conceptual understanding of the problem that may even lead to adjustments of the RS information product definition.

The above-described tasks within the simplified RS information production process identified several image QC within the phases and sub-phases that are matched to the QDs in RS within the Table 1.

Phases of RS lifecycle	Sub-phases of RS lifecycle	Relevant RS Quality Dimensions	Highlighted Quality Criteria
conceptualize	problem understanding, setting the objective	trust of sources, spatial resolution, spectral resolution, radiometric resolution, temporal resolution	successful peer-review of published concepts
collect	acquisition	AOI completeness,	percentage of AOI area covered by a specific RS image, cloud cover percentage of RS image, minimum AOI size, maximum complexity of AOI (by number of vertices)
		temporal validity	acquisition period, temporal match to a preferred acquisition date
	pre-processing	spatial accuracy, radiometric accuracy	root mean squared error (RMSE), maximum deviation from absolute radiance value
process	information extraction	thematic accuracy, consistency	omission and commission error, overall thematic accuracy
	production	timeliness	processing time
deliver & use	decision	timeliness	current provisioning of information, delivery time

Table 1. Highlighted QC along the simplified RS lifecycle and their relation to RS QD.

3. STANDARDS FOR DATA QUALITY DIMENSIONS IN REMOTE SENSING

In the last decades the standardization for geographic information data quality has developed rapidly. After the standardization for the vector world had reached maturity some organizations for standardization focused their activities on image data and gridded data. The major organizations involved are the ISO/TC 211 "Geographic information / Geomatics" which publishes the official standards, the Open Geospatial Consortium (OGC), which publishes industry standards, as well as the Quality Assurance Framework for Earth Observation which publishes best practice guidelines (Kresse 2004, 2010).

3.1 The Data Quality Standardization Bodies and Relevant Standards

QA4EO: The Quality Assurance Framework for Earth Observation compiled by the Group on Earth Observation and CEOS contributed to validation and overarching ideas (Lecomte et al., 2010). The Quality Assurance for Earth Observation (QA4EO) initiative has produced a number of guidelines that have been adopted as GEOSS best practice documents. The key principle of QA4EO is that all data and derived products must have associated with them a Quality Indicator (QI), which must be unequivocal and universal in terms of its definition and

derivation based on a statistically derived value. This value should be the result of an assessment of its traceability to an agreed reference standard as propagated through the data processing chain. Quality indicators may be entirely objectively derived by measurement and calculation or, if necessary, may be elicited by subjective expert judgement, but the data provider must specify how the quality indicator was derived. The scientific literature suggest that quality indicators can inform users of a global measure of quality without them having to examine the data in much detail.

ISO/TC 211: The most relevant standardization organizations for photogrammetry and remote sensing are the ISO/TC 211 Geographic information / Geomatics. Within the ISO/TC 211 the Working Group 6 "Imagery" integrates all related developments. The ISO focus on the theme specific metadata (Kresse 2004).

The following relevant standards are the outcome of ISO working groups:

ISO 19101-2 Reference model – Part 2: imagery: The central concept of the enterprise viewpoint is how the geographic imagery community interacts to enable imagery collected from different sources to become an integrated digital representation of the Earth widely accessible for humanity's critical decisions. The enterprise viewpoint provides the metric traceability

between this objective and the system design for distributed geographic imagery processing systems.

ISO 19115-2 Metadata – Part 2: Extensions for imagery and gridded data: This part of ISO 19115 extends the existing geographic metadata standard by defining the schema required for describing imagery and gridded data. It provides information about the properties of the measuring equipment used to acquire the data, the geometry of the measuring process employed by the equipment, and the production process used to digitize the raw data. This extension deals with metadata needed to describe the derivation of geographic information from raw data, including the properties of the measuring system, and the numerical methods and computational procedures used in the derivation.

ISO 19129 Imagery, gridded and coverage data framework: This standard provides a framework for applications dealing with imagery, gridded, coverage and/or raster data, such as remote sensing, photogrammetry, image processing, digital elevation and terrain models, and modelling using discrete surfaces (polygons with homogenous values) or continuous surfaces.

ISO 19130 Imagery sensor models for geopositioning: This standard identifies the information required to determine the relationship between the position of a remotely sensed pixel in image coordinates and its geoposition. It supports exploitation of remotely sensed images. It defines the metadata to be distributed with the image to enable user determination of geographic position from the observations.

ISO 19157 Data quality (The three formerly separate standards ISO 19113 "Quality principles", ISO 19114 "Quality evaluation procedures", and ISO/TS 19138 "Data quality measures"): The objective of this standard is to provide principles for describing the quality for geographic data and concepts for handling quality information for geographic data, and a consistent and standard manner to determine and report a dataset's quality information. It aims also to provide guidelines for evaluation procedures of quantitative quality information for geographic data.

ISO 19158 Geographic information – Quality assurance of data supply: This Technical Specification provides a framework for quality assurance specific to geographic information. It is based upon the quality principles and quality evaluation procedures of geographic information identified in ISO 19157 and is based upon general quality management principles as defined in ISO 9000. It is a framework that enables a customer to satisfy itself that its suppliers, both internal and external, are capable of delivering geographic information to the required quality. Fundamental to the framework is the assurance of the supplier's ability to understand and meet the quality requirements. Through accreditation both the customer and the supplier are able to consider the quality required at the earliest opportunity in the production/update process. Principles and responsibilities of the relationship between the customer and the supplier that facilitates the framework are provided. The responsibility for accreditation is shared between the customer and the supplier.

ISO 19159 Calibration and validation of remote sensing imagery sensors, **Part 1**: Optical sensors, **Part 2**: LIDAR, **Part 3**: SAR/InSAR: The standard shall define the calibration and validation of identified airborne and space borne remote sensing imagery sensors and data. The term calibration refers to geometry and radiometry, and includes the instrument calibration in a laboratory as well as in-situ calibration methods. The validation

methods are split into process- and product validation, and include the prerequisites for installing a validation environment.

OGC: The OGC is the worldwide leading consortium of GIS industries promoting the interoperability of geographic information across platform, system, and country borders. The main field of current activity is the complete integration of the sources of geographic information based on the Internet. The Open GIS Consortium (OGC) plays an important role on the implementation level. In theory, the ISO/TC211 develops the abstract standards and the OGC develops the implementations standards. The ISO/TC211 has created some of the ISO 19100 standards based on proposals from the OGC. The ISO 19100 standards that are not abstract have been adopted from the OGC. They are the ISO 19128 (Web map server interface), the ISO 19136 (Geographic Markup Language), and the ISO 19139 (Metadata - Implementation Specification).

3.2 The Standards in the Different Phases of the RS Lifecycle Process

Standard refer to the term 'quality' with different interpretations depending on the production and the lifecycle position (Jakobsson and Giversen, 2007).

Before production – conceptualize: The quality documentation focusing on the specification of the quality model. The main goal is the definition of the quality requirements of the customer. Based on the ISO 19131 "Geographic Information – Data Product Specifications" the concept of the quality model can be defined.

Production – collect, process: The certification process focuses on the documentation of the process history of produced data. The objective of the quality management is to inspect if the data meet the specification or not. During the production the quality management principles should be applied (ISO 9000 family of standards: ISO 19157, ISO 19115). Organisations that want to adhere to these principles have to undergo certification of their quality management process, record the lineage of each product, ensure appropriate use of production methods etc.

After production – deliver & use: After production, quality evaluation should be performed. The quality documentation addresses the metadata test reports. The main objective of the evaluation is to measure the conformance to quality requirements (ISO 19139, ISO 19157, ISO 19115).

3.3 The significance of standards for designing validation approaches in projects applying RS data and technologies

Standards play an important role when it comes to the quality assessment of individual projects and information products that apply remote sensing data or systems. In this case, users will look for understanding the specifications of all the quality measures involved in the project. These specifications can be understood from the used types of data, the applied process and the anticipated/achieved goals of the project. The QA4EO was established with a focus on such individual projects and with the objective to facilitate interoperability of GEO systems provided by EO.

An example that goes beyond the scope of individual application scenarios is NASA's Earth Science Data and Information System (ESDIS) project where standards were adopted based on defined EO mission requirements. Thereby, NASA was adopting and approving a list of standards including ISO 19115 Geographical

Information Metadata Standard, NASA Earth Science Data Preservation Content Specification, and Digital Object Identifiers (DOIs) for Earth Observing System Data and Information System (EOSDIS). NASA ensured to provide the users with the necessary required information to understand and use the data and products of their EO missions. Approved standards include data format standards, status of data and systems and specification document and user resources. Approved data formats are for example HDF EOS 5 and OGC KML. Additional information related to the specific formats can be found in the NASA Earth Science Community Recommendations for users. Thus a user of the data can refer to those approved standards to understand the quality aspects and the limitations of the types of data used for their project. Further, the community recommendations provide valuable information including the strength, weakness, applicability and limitations of specific data formats. Beyond data format standards, a list of metadata standards has been adopted that include, among others: the Unified Metadata Model (UMM) and NetCDF Climate and Forest (CF) Metadata Conventions (complete list available on the website https://earthdata.nasa.gov/user-resources/standards-and-references). Consequently, the designer of an RS validation process for a certain project is able to employ the metadata that provides standardized RS data parameters as candidates for the QC in the RS lifecycle. This includes temporal and positional information and many more factors relevant for the specific project at hand.

Some standards are formulated to add a value to other standards making sure to cover the elements related to RS, example of which is what the EOSDIS adds to Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata, it added metadata elements in the remote-sensing domain, such as platform information, sensor information, the history of data processing, and the spatial representation of sensor data (Di and Kobler, 2000). Thus, standards help regulate the broaden use of earth observation data and interoperability. Taking the EOSDIS as an example, the system provides structural and product metadata, covering relationships, data storing information, and the product information. Other types of standards exist that potentially have implications for designing the validation processes such as the FGDC Content Standard for Remote Sensing Swath Data from the context of mission planning. The standard defines the minimum data content of a swath and the relationships among its internal components with semantic definitions, this type of standard can be beneficial to certain applications of remote sensing as it helps in understanding the contents, coordinates, time scale of a swath data.

Thus, standards can provide the necessary support for designing the quality validation approaches, especially when detailed information are required. But initially following a unified process can be of benefit and guide the user to develop a process that is compliant and interoperable. Finally, the standards regulate the reporting process, being followed will enable the continuity and clearness of the produced products.

4. CONCLUSIONS

We have described the general relationships between the RS QDs, the phases of the RS lifecycle and the used types of data sources. A detailed description of the RS information production process followed that gave an overview how QC are used in highlighted phases. A summary table gives an overview of the illustrated QC and showed their relation to the RS QDs. We then provided an overview of the available standards, mostly provided by ISO and OGC, and the best practice guidelines of QA4EO that support the assurance of data quality in the RS domain. We matched them to the phases of the presented RS information production process and showed examples how they can provide the necessary support to designing adequate validation procedures.

This research intends to give a general overview how QC are employed in the RS lifecycle and what the role of standards is in support of validation procedures. However, it remained superficial in terms of several issues that require in depth coverage. There is a need to gather a more complete picture of the RS lifecycle that includes the most common phases and sub-phases. This would enable a generic lifecycle description that is applicable to the diversity of all RS application domains. A generic lifecycle would allow a more complete picture of the relations between QD and QC, as well as a more complete picture of RS data types. In addition, examples of validation procedures from different RS application domains may better illustrate how the extended concepts can support the design process. Further investigations of standards should take place by comparing practical examples of designed RS quality validation processes and referring them to existing and adopted standards. This will demonstrate how the extended concepts on RS DQ complement the available standards.

REFERENCES

Batini, C., Blaschke, T., Lang, S., Albrecht, F., Abdulmutalib, H. M., Barsi, Á., Szabó, G., and Kugler, Zs., 2017. DATA QUALITY IN REMOTE SENSING, In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W7, 447-453, https://doi.org/10.5194/isprs-archives-XLII-2-W7-447-2017.

Di, L., Kobler, B.: NASA Standards for Earth Remote Sensing Data, International Archives of Photogrammetry and Remote Sensing. Vol. XXXIII, Part B2. Amsterdam 2000

Eberhardt, I., Schultz, B., Rizzi, R., Sanches, I., Formaggio, A., Atzberger, C., Mello, M., Immitzer, M., Trabaquini, K., Foschiera, W., José Barreto Luiz, A., 2016. Cloud cover assessment for operational crop monitoring systems in tropical areas. Remote Sensing, 8(3), 219.

European Commission, 2016. Copernicus market report November 2016. In: *EU Law and Publications*, doi.org/10.2873/827100.

Group on Earth Observations, 2017. The 2017-2019 GEO Work Programme. In: *Earthobservations*, http://www.earthobservations.org/documents/work_programme/geo_2017_19_Work_Programme.pdf

Jakobsson A., Giversen, J., eds., 2007. Guideline for Implementing the ISO 19100 Geographic Information Quality Standards in National Mapping and Cadastral Agencies. Eurogeographics Expert Group on Quality.

Joint Committee for Guides in Metrology (JCGM), 2012. International vocabulary of metrology—Basic and general concepts and associated terms, BIPM, Sèvres, France.

Kresse, W., 2010. Status of ISO Standards for Photogrammetry and Remote Sensing, ISPRS – EuroSDR, EuroCOW 2010, Castelldefels, Spain.

Kresse, W., Fadaie, K., 2004. ISO Standards for Geographic Information. Springer.

Lang S., Füreder P., Rogenhofer E., 2018. *Earth observation for humanitarian operations*. In: Al-Ekabi C., Ferretti S. (eds) Yearbook on Space Policy 2016. Yearbook on Space Policy. Springer, Cham.

Lang, S., Schoepfer, E., Zeil, P., Riedler, B., 2017. Earth observation for humanitarian asssistance. GI_Forum 2017, 1, 157-165, doi.org/10.1553/giscience2017_01_s157

Lecomte, P. et al., 2010 A Quality Assurance Framework for Earth Observation (QA4EO). Key guidelines, version 4.0, Group on Earth Observations / Committee on Earth Observation Satellites.

Tiede, D., Baraldi. A., Sudmanns, M., Belgiu, M., Lang, S., 2016. ImageQuerying – Earth observation image content extraction and querying across time and space. ESA Big Data from Space BiDS 2016, Tenerife.