INSPIRE
Infrastructure for Spatial Information in Europe

Data quality in INSPIRE: from requirements to metadata

Discussion paper

Title	Data quality in INSPIRE: from requirements to metadata - Discussion paper
Creator	European Commission
Date	2010-10-04
Subject	Data quality and metadata
Publisher	European Commission
Type	Text
Description	The aim of this discussion paper is to outline the process of addressing the topic of data quality, to provide background information for the discussion, and to invite the INSPIRE (Data Quality) Member States Points of Contact to answer a number of questions which will help structuring the discussion.
Contributor	Katalin Tóth, Robert Tomas, Antti Jakobsson, Gilles Troispoux, Carol Agius, Vanda Nunes de Lima, Paul Smits
Format	MS Word (doc)
Source	INSPIRE_DQ_MD_v1 8 sent for discussion on 19 May 2010
	Minutes of the INSPIRE Data Quality Workshop (Krakow, 26 June 2010)
Rights	INSPIRE Member States Contact Points
Identifier	INSPIRE_DQ_MD_v1.9.doc
Language	En
Relation	n/a
Coverage	Project duration

Table of contents

List of Abbreviations

DQ	Data Quality
EC	European Commission
GCM	Generic Conceptual Model
INSPIRE	Infrastructure for Spatial Information in Europe
ISO	International Standards Organisation
MD	Metadata
NMCA	National Mapping and Cadastre Agencies
SDI	Spatial Data Infrastructure
TWG	Thematic Working Group

Foreword

In the process of the development and adoption of the draft INSPIRE Implementing Rules for interoperability of spatial data sets and services, it became apparent that further discussion is needed to better understand and address the aspects of data quality in the context of INSPIRE. The Commission agreed to initiate and lead this discussion.

At the meeting of the INSPIRE Member States Contact Points on 10 March 2010, the contact points were requested to inform the Commission as to who will represent the countries in the discussion on data quality. The first meeting of the data quality experts took place on 22 June 2010 in Krakow, Poland that was opened to the boarder audience of the INSPIRE Conference too. The quality experts agreed that further discussions are necessary both at national and European level.

The aim of this discussion paper is to guide the process of addressing the topic of data quality, to provide background information for the discussion, and to invite the INSPIRE (Data Quality) Member States Points of Contact to answer a number of questions which will help structuring the discussion.

1 Introduction

On 14 December 2009 the INSPIRE Committee approved unanimously the draft Regulation on Interoperability of Data Sets and Services for INSPIRE Annex I data themes. In the process of developing this draft Regulation, which is based on the INSPIRE data specification guidelines developed by the Thematic Working Groups (TWGs), the question of data quality was a re-occurring issue, both during the data specifications development and the consultations.

This interest in data quality can be explained by the fact that quality is one of the data harmonisation components underpinning interoperability. The peculiarity of the discussion was the wide divergence of opinions, ranging from introducing strict data quality requirements for all data included in the infrastructure, to complete omission of requirements.
During the discussions it became clear that the terminology is often understood and used differently. As a result the *a-priori* data quality requirements need to be carefully distinguished from metadata.

The draft data specifications of each Annex I theme (v 2.0) have been consulted with stakeholders' communities. The comments related to data quality and metadata parts have been addressed by the Thematic Working Groups (TWGs) responsible for the specification process. Based on outcome of the consultation and the recommendation of the Data specification Drafting the idea of introducing minimal a priori data quality requirements for INSPIRE has been dropped. This approach is reflected in v 3.0 that constituted the technical basis of the draft Regulation on Interoperability of spatial data sets and services. The INSPIRE Committee has raised two concerns connected to data quality:

- How the comparability of information derived from different spatial data sets is ensured in the draft Regulation? In its answer of the Commission pointed to the rigorous and common data modelling principles that addresses all interoperability/ data harmonisation components of the Generic Conceptual Model (GCM). Never the less the need for reviewing the data quality and metadata parts have been agreed.

- Need to introduce a minimal absolute geometric accuracy (to be less then 2/1000 of the distance resolution). The Commission did not accept this general approach. The agreement was that the issue should be further explored in relation to the different data themes.

Following these concerns the European Commission committed to organise a broader consultation with the Member States on data quality with the leadership of the EC INSPIRE Team. This updated paper prepares and guides the discussions on the above topics, clarifying the details and giving an initial position to stimulate the exchange of views. It is expected that the results of this process will be useful for developing data specifications of Annex II and III data themes, and that they will be considered in future updates of other INSPIRE documents.

The remainder of this paper is organized as follows. Section 2 gives an overview of the main objectives of and the process for the discussions. Section 3 introduces the topic of data quality and metadata. Section 4 explains the difference between data quality concepts in the data production and in the SDIs. Section 5 reports on previous experience on data quality in INSPIRE, followed by the discussion points and preliminary conclusions (section 6).

2 Objectives and process

The INSPIRE data quality and metadata discussions are expected to reach the following objectives:

1. Clarify the context and terminology related to data quality in SDIs

2. Find evidence whether specifying data quality requirements are appropriate for INSPIRE;
3. If yes, propose data quality elements, measures, and target values;
4. Fix how metadata on data quality has to be presented;
5. Provide guidance on DQ requirements and Metadata for the specifications of Annex I, II and III data themes
6. Formulate proposals amending, if appropriate, different INSPIRE documents
7. Raise awareness about the role of data quality and metadata in spatial data infrastructures.

For reaching these objectives it is necessary to channel information exchange through a few steps, which are expected to reach agreements in a bottom up manner. This process is halfway now already outlining some progress after the DQ workshop mentioned in point 5.

1. **Drafting the discussion paper**
   The first version of the discussion paper was developed by the European Commission supported by a small group of experts with aims to scope the subject, clarify the terminology, review the initiative already in place, and propose an initial position.

2. **Consultations in the Member States (first round until 11 June 2010)**
   The discussion paper was sent for consultation in the Members States via their nominated data quality or INSPIRE contact points. They were invited to organise a review in their countries to achieve an agreed and consolidated position. In order to reach a structured result, specific questions were asked. The Commission received back 15 answers from 12 countries. It was clear that the time planned for national discussions was too short to collect and consolidate the answers coming from different organisations.

3. **Analysis of the results of the consultation (14-18 June 2010)**
   The answers received by the Commission were analysed and presented at the Data Quality Workshop held in frame of the INSPIRE conference in Krakow. Even the small number of consolidated “national” answers confirmed that the initiative was well-received and the participants agreed to continue this activity.

4. **Face-to-face discussion (22 June 2010)**
   The DQ contact points have been invited to the workshop in Krakow on 22 June 2010. The workshop has been opened up to the general audience of the INSPIRE Conference The national DQ contact points presented the status of the data quality discussion in their countries, gave feedback both on the discussion paper and the process. They could raise other relevant aspects too. Since INSPIRE has to be based, whenever appropriate, on international initiatives and standards a short presentation about the upcoming ISO 19157 Data quality standard was given by its editor.

5. **Updating the discussion paper**
   The participants of the ISNPIRE DQ workshop agreed that further discussions are necessary. For this purpose the discussion paper has been updated paying special attention to the upcoming ISO 19157 standard. The EC has agreed to deliver the new version of the discussion paper at the beginning of October 2010.

6. **Consultation in the Member States (second round until 30 November 2010)**
   The Member States will organise (further) discussions in their countries. Their national positions shall be forwarded till the end of November to the EC JRC data specification support team.

7. **Analysis of the result of the consultation (January 2011)**
   The answers given to the questionnaire of the discussion paper will be analysed by the EC JRC data specification support team with views of preparing a draft report on the subject. The report will summarise the points of agreement and highlight the

issues where further work is needed. It will propose preliminary recommendations how to handle data quality in INSPIRE.

8. **Data quality contact point meeting (February 2011)**
   The draft report will be distributed to the data quality contact points of the Member States for review, which will be followed by a face to face meeting to discuss open issues and agree the final text of the report.

9. **Final report**
   The Commission will update the report with the results of the February meeting and will be disseminated to a wider public. It will provide recommendations for possible updates of INSPIRE documents and how data quality and metadata should be addressed in spatial data infrastructures in general. In may also involve creating a discussion platform that will help to keep the recommendations of the report up to date.

3 Quality and its role in geographic information

Quality is a degree to which a set of inherent characteristics fulfils requirements [ISO 9000]. Geographic data sets describe the real world from different viewpoints giving direct or indirect reference to a location. They are increasingly being shared, interchanged and used for other purposes that they were originally produced for. Information about the quality of geographical data is indispensable for selection and proper use of data. Data providers are interested in the widest reuse of their datasets that is only possible when

- the data sets satisfy the requirements of the targeted users
- the degree to which the requirements are fulfilled is documented.

Both aspects lead to formalising data quality. In the first, based on users' requirements specific targets on data quality are established that have to be achieved in the course of data production or transformation. The second aspect corresponds to documenting the quality of the data that is eventually delivered to the users. For each of these tasks a common way of expression is necessary that comprises an agreed terminology, evaluation, and reporting methods. The upcoming ISO/CD 19157 Geographic Information – Data Quality[1] Standard fulfils this need giving conceptual framework for the issue.

Based on this upcoming standard data quality of geographical data can be described using three main types of components:

- **Data quality elements** and their descriptors are used to describe how well a dataset meets the criteria set forth in product specification or user requirements
- **Non-quantitative overview statements** such as purpose, usage and lineage are reported as metadata in accordance of ISO 19115
- **Metaquality** that describes the quality of the data quality results in terms of the defined characteristics

On its turn, a data quality element is described using

- **Data quality measure** (test applied to evaluate a specific data quality element)
- **Evaluation** (set of operations applied in the data quality evaluation method)
- **Result** (outcome of the data quality evaluation, which can be a descriptive, quantitative, or conformance result)

It is necessary to point out that quality is relevant to each phase of data production and sharing. **In the data production cycle** data quality is present in the specification phase (dedicated part data product specification according to ISO 19131), in production phase (quality assurance), quality checking (sampling), and documentation (metadata, and quality evaluation reports). From the point of view of **data sharing** the focus is on documentation and proper *communication of* the quality of available data. However, in **Spatial Data Infrastructures** the interoperability principle brings back both the specification and quality assurance aspects. The interoperability target to be achieved within the infrastructure has to

[1] See the status on http://www.iso.org/iso/catalogue_detail.htm?csnumber=32575

be specified; likewise the necessary data transformations have to be underpinned by quality assurance. In addition, the original metadata may not be valid for the transformed data; consequently the related metadata and the quality evaluation reports have to be updated.

In the following chapters our discussion paper will deal only with the specification and documentation aspects of data quality, quality assurance and checking will be mentioned only in the context of the first two.

4 Two faces of data quality: quality requirements and metadata

4.1 Data quality in data production

Before starting discussing data quality it is necessary to clarify its role in spatial data/information production and in SDIs in general.

Till the latest decade in ideal case data was produced by professionals according to pre-defined data product specifications, standards, or legal regulations. However the new technologies allow more and more the spontaneous data/information collection and sharing, giving way to "volunteered geographic information"[2] In consequence the data quality requirements, which are normally the basis for the production of the data, the quality control, the quality evaluation and conformance testing sometimes remain unknown. While on one hand there is an increasing need to get informed about the quality of the data, on the other hand the undocumented requirements somewhat hamper the delivery of good and meaningful metadata.

Ideally data is being produced to fulfil specific use-cases, i.e. to satisfy the requirements of users connected to well-defined tasks. These requirements are optimally, formalised in data product specifications, which is the basis for data production. As a rule they contain specific parts related to ***'a priori'*** **requirements** on data quality to be followed during the production process. Frequently standards or other regulations drive the data specification and production processes giving strict target values for selected data quality measures.

DQ **Metadata**[3] gives ***'a posteriori'*** statement about the data quality based on the de-facto measurements, specific aggregation rules applied to the given dataset, or the knowledge of the author about the data expressed as non quantitative information. Metadata for evaluation and use may include one or more data quality elements, each of them expressed by a selected data quality measure and the data quality result. It should be noted that conformance statement, which is a specific data quality result can be related either to a selected data quality element or to all data quality elements applicable to the inspected dataset.

Conformance statement against a data product specification can be defined as *internal conformance*. Internal conformance statements, like the quantitative data quality results are useful to expert users that know the contents of the referred specifications or have the necessary knowledge to understand and interpret the data quality measures and their eventual values. Other non-specialist users rather benefit from a conformance statement against user requirements or a standalone quality evaluation report. A conformance statement against user requirements can be defined as *external conformance*. .However this latter has not been widely used so far mainly because of the lack of formalisation of such requirements. Using "DQ_DescriptiveResult" offered in the new draft of the ISO 19157 will give the possibility for such "subjective" evaluations.

The connection between a priori data quality requirements and metadata for evaluation and use is clear. From the point of view of theory they share the same concepts, i.e. the same data quality elements can be used both for specifying data quality requirements and reporting

[2] Goodchild

[3] For simplicity, later on when metadata is mentioned metadata for evaluation and used has to be understood only.

data quality as metadata. Consequently it is logical/practical the same DQ elements possibly with the same DQ measures for metadata that were used for specifying the requirements.

4.2 Data quality in SDIs

An SDI provides the technical and legal framework for accessing and reusing spatial data produced in a defined geographical zone (global, national, and sub-national), or thematic field. It is assumed that it is built on existing data that are produced by different data providers. The flow of data from the data producers to their provision within an SDI is shown in Figure 1.

Ideally an SDI provides access to data in interoperable way, i.e. without the need for specific ad-hoc interaction of humans or machines. The interoperability target is formalised in **data specifications** that have a similar structure to the data product specifications of the production process.

As compared to data production the role of '*a priori*' data quality requirements in SDI is different. When establishing the data component of an SDI, two aspects need to be balanced:

1. Giving access to the widest selection of data;
2. Achieving interoperability.

The background of the first aspect is that the final decision whether a data set is useful should belong to the users. This might lead to providing data in the infrastructure without any minimal *a priori* requirements for the data quality. The underlying principle of this approach is that any data is better than no data.

The second condition, promoting interoperability, implies that data from disparate sources can be combined without specific efforts. However, when the quality of data is very different, some data harmonisation measures, for example edge-matching, become meaningless and the integrated use of data is jeopardised. In addition, SDIs are expected to provide the future coherent basis for the development of new applications. From these points of view minimal *a priori* requirements on data quality are relevant.

Requirements on data quality are usually more stringent for reference data[4] especially in terms of positional accuracy, as their geometries are frequently used for object referencing. In addition, in thematic SDIs some fundamental use-cases such as hazard or flood risk mapping may also justify specific requirements for quality; otherwise the infrastructure does not reach its objective.

In summary, data quality requirements may play a discriminative role when deciding whether to include a specific data set in an SDI. Balancing the wide spread publication of data sets with requirements against the quality of data is a delicate decision that directly influences the content of specifications for interoperability. Consequently data quality requirements may be completely absent from the target specifications if the main purpose of the infrastructure is to make available every existing data set.

Naturally users have to be informed about the quality of the data that they retrieve from the infrastructure. Contrary to data quality requirements, metadata on data quality is an indispensable content of every SDI[5], which has bigger visibility[6]. However, before publishing metadata in the infrastructure each data provider has to analyse whether the transformations necessary to fulfil the interoperability targets have lead to deterioration of data quality. If yes, metadata has to be updated accordingly.

The data and metadata flow in data production and SDI implementation presented in fig.1 shows the similarities, differences, and interconnections of the two working chains.

---

[4] Data that can be used for linking other types of (thematic) information.

[5] In the Quality Assurance Framework for Earth Observation of GEO (QA4EO) this is expressed as "Data and derived products shall have associated with them a fully traceable indicator of their quality".

[6] This sometimes leads to confusions – some authors misuse the notion of metadata when they speak about a priori data quality requirements.

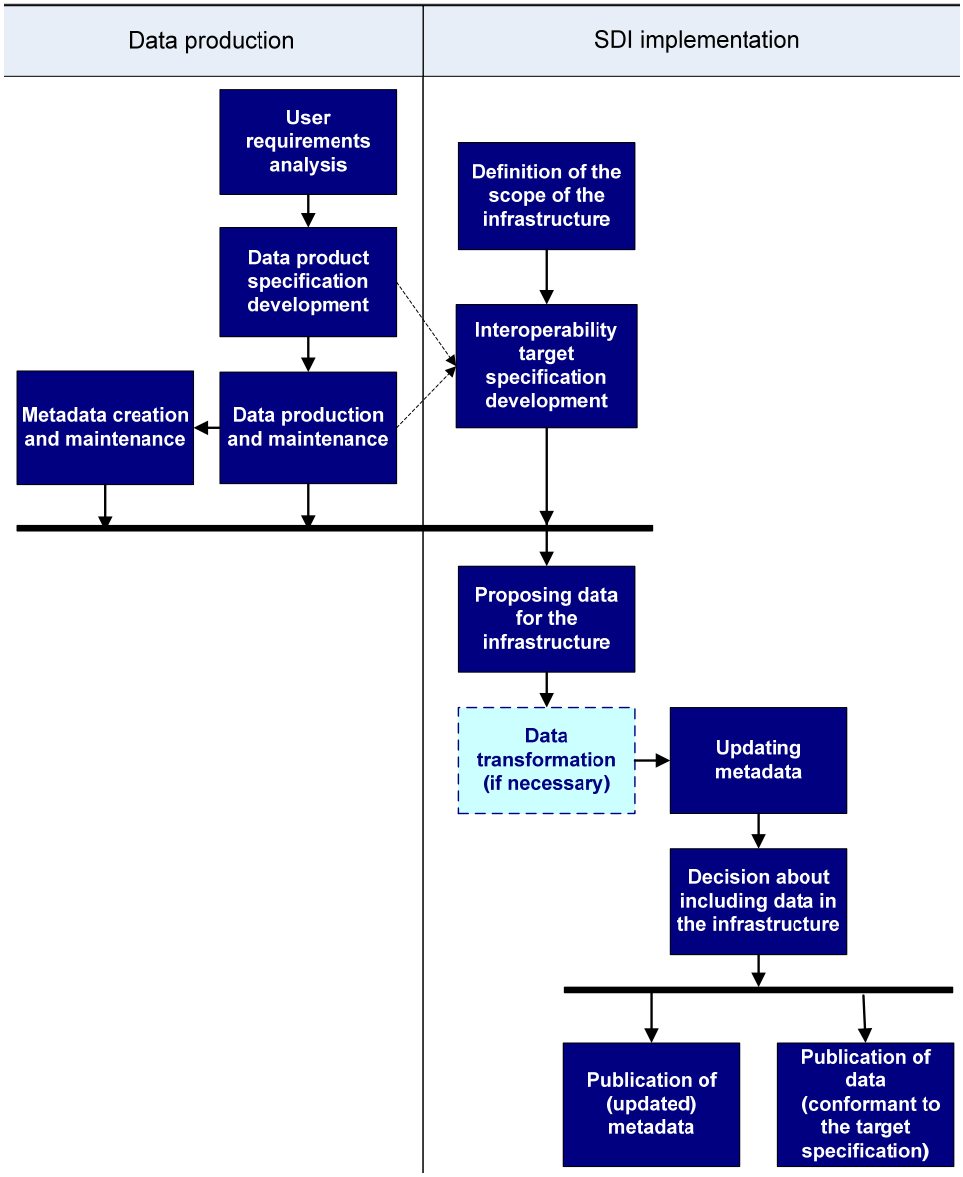

Figure 1: Data and metadata flow in spatial data infrastructures

In SDI design and implementation the 'a priori' data quality requirements stem from the scoping phase, when the high level use-cases underpinning the infrastructure are defined. In this process a wider range of individual use-cases are analysed with views of finding analogies and the right balance of data harmonisation needs[7]. Like to the 'User requirement analysis' of the data production process the high-level use-cases are translated in requirements that may contain specific a priori data quality requirements for the infrastructure. In this case, the interoperability specification contains a specific data quality section where the appropriate DQ elements are defined with the corresponding DQ measures and their target results.

Having compared the interoperability specifications with the specifications of the existing data products and the corresponding metadata the data providers may propose their data for the infrastructure. The comparison of specifications answers the question whether the data has to be transformed to reach the interoperability target in terms of spatial schema, while comparing the original metadata elements with the a priori data quality requirements results in preliminary checking of fitness for use in the given SDI. In case when no data transformation

[7] See details in Annex A of INSPIRE D2.6 – Methodology for Data Specification Development http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/DataSpecifications/D2.6_v3.0.pdf

is necessary the original metadata can be directly published provided that the SDI does not requires additional metadata elements.

Data transformation may lead to deterioration of original data quality; consequently after the data transformations the appropriate parts of the metadata have to be updated and once again compared to the data quality requirements of the infrastructure. In case when the quality of the data after the transformations is still satisfactory, the data can be included in the infrastructure.

Updating the results of different metadata elements might be cumbersome. In topographic data production, for example, it can be based on calculations/aggregations, or on quality inspection based on appropriate sampling[8]. A practical alternative can be to use the original metadata with information on the process step describing the transformation methods and the possible associated errors, expressed in the MD_Lineage element[9] or express qualitatively in DQ_DescriptiveResult metadata sub-element as a report. The shortfall of this simplified process is that the fitness for the infrastructure after the transformations is not inspected.

Interoperability also requires that data quality is measured and reported in an agreed way; otherwise it is not possible to compare the metadata associated to different datasets. Therefore the metadata part of the interoperability target specification has to fix the data quality elements, evaluation methods, and measures to be used in a data theme. If possible, these aspects should be harmonised across the data themes too.

From point of view of users the significance of the same DQ element is different. As said before, positional accuracy is more important for reference data than for other thematic data, while thematic classification correctness is prime importance for some coverage data (e.g. land cover). Therefore specific attention should be paid in the data specification development process as to which are the most meaningful or expressive elements to describe the quality of data.

However detailed the metadata descriptions of data quality are they may have a “fairly limited impact on user’s ability to understand the possible uses of data,” which means there is still a gap between what the quality assessment mapping experts can produce and the information users can understand and use.[10]. In order to cover this gap the concept of usability has been proposed. A specific DQ element DQ_Usability has been included in the upcoming ISO 19157[11] standard, which defines it as a “degree of adherence to a specific set of data quality requirements”, or “adherence to a particular application”. While this standard finds usability especially useful when other data quality elements do not sufficiently address a component of quality, the Quality Assurance Framework for Earth Observation of GEO (QA4EO) emphasises that this information should be based on a quantitative assessment to an agreed reference or measurement standard, which can be presented as numeric or a text descriptor. There is no divergence between ISO and QA4EO in using this element to declare conformance with a particular specification.

[8] A model based on ISO data sampling methods is proposed by the ESDIN project.

[9] Report or lineage role metadata element is mandatory when the scope of the DQ element is the dataset (ISO 19115)

[10] 30 years of research on Spatial Data Quality – Achievements, failures and opportunities by Rodolphe Devillers

[11] In QA4EO this concept is referred by “Quality Indicators”.

5 Data quality in INSPIRE

5.1 DQ in the legal acts

INSPIRE, like any Spatial Data Infrastructure, targets to use data from different providers by multiple users and applications. The Directive contains explicit and implicit requirements related to data quality[12]. Table 1 establishes mapping between the DQ related statements of the Directive and DQ elements and sub-elements as defined in the upcoming ISO 19157. The table also indicates the data specification elements that serve for the basis of the data quality evaluation process.

[12] Even though our discussion paper deals with the data component of the infrastructure, for completeness it is worth noting that the requirement in article 16 (a) stating the minimum performance criterion has been translated into “Quality of Service” criteria in the NS regulation.

Table 1: Data quality related parts of the Directive and the possibly related DQ elements and sub-elements

Art.	Citation	Related DQ (sub-)element	DQ evaluation against the
5(2)	Metadata shall include information on the quality and validity of spatial data sets	All relevant to the data set	Application schema of the data theme Implementing Rule / Data specification of the data theme
7(3)	Member States shall ensure that […] spatial data sets and the corresponding spatial data services are available in conformity with the implementing rules […]	DQ_ConceptualConsistency	Application schema of the data theme
7(4)	Implementing rules […] shall cover the definition and classification of spatial objects relevant to spatial data sets related to the themes listed in Annex I, II or III and the way in which those spatial data are geo-referenced.	DQ_ConceptualConsistency DQ_ThematicClassificationCorrectness DQ_TopologicalConsistency	Application schema of the data theme
8(1), (2)	In the case of spatial data sets corresponding […] the themes listed in Annex I or II […] the implementing rules shall address the following aspects		
	- framework for the unique identification	DQ_DomainConsistency	GCM
	- the relationship between spatial objects	DQ_ConceptualConsistency	Application schema of the data theme
	- key attributes […]	DQ_ConceptualConsistency DQ_NonQuantitativeAttributeAccuracy DQ_QuantitativeAttributeAccuracy	Application schema of the data theme
	- information on the temporal dimension	DQ_TemporalConsistency DQ_TemporalValidity	Application schema of the data theme GCM
8(3)	[…] consistency between items of information which refer to the same location	DQ_ConceptualConsistency	Application schema of the data theme (multi-scale representation)
	or between items of information which refer to the same object represented at different scales	(DQ_PositionalAccuracy)	Data specifications of the related data themes
10(2)	In order to ensure that […] a geographical feature, the location of which spans the frontier between two or more Member States, are coherent, Member States shall, […] decide by mutual consent on the depiction and position of such common features.	DQ_LogicalConsistency DQ_PositionalAccuracy	Agreement between the interested parties

5.2 DQ status after Annex I

The '**Conformity**' metadata element specified in Commission Regulation (EC) 1205/2008 shall be defined at data set level. Since the multiplicity of this metadata element is 1..*, it is possible to report conformity with different specifications. It stems from Art 5(2) that reporting conformity with the (draft) Regulation on Interoperability of Spatial Data Sets and services is mandatory. The possible results of conformity are defined with the following enumeration:

- Conformant
- Not conformant
- Not evaluated

According to the related ISO standards conformance can be reported for each specification element and the results can be aggregated for the level of data set. The metadata regulation requires conformity statement at the level of data sets or data set series. However, neither the metadata, nor the data interoperability regulations contain rules for the aggregation process[13] and the related technical documents miss to link conformity with the appropriate (data quality) specification elements.

For evaluating conformity not only the elements, but also the applicable DQ measures and the related results and tolerance values have to be specified. Guided by the "non exclusion" principle the TWGs responsible for Annex I data specifications have selected DQ measures from ISO 19138, but as a rule have not established targets for their values. However the Directive mandates logical consistency. The missing DQ requirements on it constitute a situation when all the data provided in INSPIRE should be 100% correct to be conformant with the Directive. Needless to say that this contradicts to the uncertainty associated with the measurements and classifications present in the GI technology.

Combining spatial data from different themes and sources in a consistent way represents a strong data quality demand that go beyond the logical consistency within a data theme. The consequent application of the data modelling elements and other provisions of the Generic Conceptual Model (GCM) enforces cross theme consistency with such modelling methods as object referencing and constraints. The GCM lists data quality among the data harmonisation elements, but does not give a generic data quality model and other details of specifying data quality requirements and metadata descriptions.

The "non exclusion" approach is in line with the recommendation of "D2.6 Methodology for development of data specifications" INSPIRE framework document, which lead to dropping the idea of uniform mandatory minimum data quality requirements. Only a few of the data specifications introduced a priori data quality requirements or recommendations. They are listed bellow:

Cadastral parcels:

- Rate of missing items should[14] be 0% for cadastral parcels and cadastral zonings (if any).
- Mean value of positional uncertainties should be 1 meter or better in urban areas and 2,5 meters or better in rural/agricultural areas. Cadastral data may be less accurate in unexploited areas.
- Edge-matching between cadastral parcels in adjacent data sets should be done. Ideally, there should be no topological gaps or topological overlaps between cadastral parcels in adjacent data sets. Status of edge-matching should be reported as metadata, under lineage element
- There should be no topological overlaps between cadastral parcels.
- There should be no topological gaps between cadastral parcels.

Transport networks:

[13] D2.6 "Methodology for Data Specification Developments" states that ideally the data quality information has to be collected at the level of spatial object types and has to be aggregated to the dataset (series) level metadata

[14] The verbal close "should" stands for recommendations, while "shall" for requirements.

- Guarantee that a continuous transport network can be built from the elements provided in the transport network datasets, by assessing their conformance to some basic topological consistency rules aimed at ensure at least clean connections between features.

These requirements are focused on enhancing interoperability and informing about the fitness for use of the datasets. They are based on the analyses of user requirements, use cases and especially on existing good practices.

Due to the natural diversity of the data themes, the user requirements, and the slightly different approach in the data specification process, each TWG proposed different set of data quality elements and MD elements. At the end of the specification process an effort of harmonisation took place. The result of this exercise is presented in the Table2.

In field of metadata the Annex I data specifications adopted all the mandatory metadata elements of 1205/2008/EC Regulation. In addition new mandatory metadata elements have been introduced and accepted:

- Coordinate Reference System
- Temporal Reference System (Mandatory, if the data set or one of its feature types contains temporal information that does not refer to the Gregorian Calendar or the Coordinated Universal Time)
- Encoding
- Character Encoding (Mandatory, if a non-XML-based encoding is used that does not support UTF-8)
- DQ_LogicalConsistency / DQ_TopologicalConsistency that apply only to Hydrography and Transport networks themes that is mandatory when the data set does not assure centreline topology (connectivity of centrelines).

The metadata part of data specifications have been complemented by recommendations related to the use of metadata elements specified in 1205/2008/EC. Those that apply to all themes are listed bellow:

- Conformity. In order to report conceptual consistency with a given INSPIRE data specification the Conformity metadata element should be used. The value of Conformant should be used for the Degree element only if the dataset passes all the requirements described in the abstract test suite[15]. An example is given bellow:
    - title: “INSPIRE Data Specification on Transport Networks – Guidelines”
    - date:
    - dateType: publication
    - date:
- Lineage.[16] It should describe the process history, if feasible, with a free text. The overall quality of the dataset (series) should be included in the Lineage metadata element. This statement should contain any quality information required for interoperability and/or valuable for use and evaluation of the data set (series).[17]
- Temporal reference. If feasible, the date of the last revision of a spatial data set should be reported using the Date of last revision metadata element.

These additional metadata elements can be displayed together with the obligatory content given in the Metadata regulation. However, it is not clear if and how these additional metadata elements can be searched (cf. Art. 11(2)). It is expected that further harmonisation between Data, Metadata, and Network services components of INSPIRE may take place upon completing their technical drafting.

[15] Abstract test suit is the fundamental tool for conformance testing, which defines platform independent steps to verify a data set against the specified requirements. For data quality it collects the data quality elements to be tested, the targeted DQ results with the appropriate tolerances, and defines the acceptable quality evaluation and aggregation methods.

[16] According to ISO 19115 lineage is information about the events or source data used in constructing the data.

[17] This statement seems to misuse “lineage”. Quality results should be reported by using appropriate quality elements.

The list of data quality and metadata elements that are currently in the Guidelines of Annex I themes is presented in the Table 2, which also reflects the result of the harmonisation effort that took place at the end of the data specification development process.

Table 2: Summary table of data quality information and related MD elements described/used in Annex I data specification

Data quality element	Data quality sub-element	DQ / Metadata element	AD	AU	CP	GN	HY	PS	TN
Completeness	commission	DQ_CompletenessCommission	X	X			X	X	2X
	omission	DQ_CompletenessOmission	X	X	X	X	X	X	X
Logical consistency	conceptual consistency	DQ_ConceptualConsistency	X	X			X		X
	domain consistency	DQ_DomainConsistency	X				X		X
	format consistency	DQ_FormatConsistency							X
	topological consistency	DQ_TopologicalConsistency		2X	3X**		9X		6X
Positional accuracy	absolute or external accuracy	DQ_AbsoluteExternalPositionalAccuracy	2X	X	X	X	X	2X	X
	relative or internal accuracy	DQ_RelativeInternalPositionalAccuracy					X		
Temporal accuracy	temporal consistency	DQ_TemporalConsistency	X						
Thematic accuracy	classification correctness	DQ_ThematicClassificationCorrectness							X
	non-quantitative attribute correctness	DQ_NonQuantitativeAttributeAccuracy	X				X		X
	quantitative attribute accuracy	DQ_QuantitativeAttributeAccuracy					X		
Maintenance*		MD_MaintenanceInformation	X	X	X	X	X	X	X

** The reason for inclusion of Maintenance information is the fact that information about the update frequency and the scope is related to the temporal accuracy of a resource. Thus, it is related to the data quality of a dataset in general.*

*** It is only expressed as recommendations without given corresponding MD-DQ element / measure (Lineage template)*

Legend:

- X Recommended MD element with one DQ measure
- X Mandatory MD element with one mandatory measure at least
- 2X The coefficient shows the number of specified DQ measures
- AD Addresses TWG
- AU Administrative units
- CP Cadastral parcels
- GN Geographical names
- HY Hydrography
- PS Protected sites
- TN Transport networks

In addition to the harmonised DQ and MD element **theme specific metadata elements** have been defined as well.

Cadastral parcels:

- Lineage. Main specificities of cadastral data should be published in the element "description of a data set", using the relevant template.
- Maintenance. Frequency with which changes are made for INSPIRE should be as close as possible to the frequency with which changes are made in a national cadastral register or equivalent. Moreover, frequency with which changes are made for INSPIRE should be one year or better.
- Positional Accuracy. A cadastral data provider may give information about absolute accuracy:
    - o at spatial object level, as attribute "estimatedAccuracy"on CadastralZoning or on CadastralBoundary
    - o at spatial object type, as metadata element "positional accuracy".

In case none of these solutions are feasible, the cadastral data provider should give information about positional accuracy under the "lineage" metadata element[18]. This may occur, for instance, if the information about positional accuracy does not provide from quality measures but is just estimated from the knowledge of source data and of production processes. More generally, absolute positional accuracy should be function of the density of human activities. This recommendation may be adapted to the specific context of each Member State).

Hydrography:

- For evaluation purpose the Data quality measure and Metadata element Rate of missing items (Completeness Omission) should be included for all spatial object types apart from the following list of types: HydroPointOfInterest and ManMadeObjects.
- Keywords should be taken from the GEMET – General Multilingual Environmental Thesaurus where possible.
- When publishing metadata for any dataset conforming to this specification; it shall have the topic category 'Inland Waters' for the corresponding metadata element (requirement).

6 Points for discussion

6.1 Direct questions

In order to propose a balanced and agreed way how DQ requirements and Metadata need to be approached in INSPIRE, we invite the (DQ) Contact Points of the Member States to organise national discussions and provide agreed and consolidated answers to the questions bellow. These questions were first published in v1.8 of the discussion paper, distributed prior to the workshop in Krakow. Based on the feedback received in the first round of consultation some of the questions have been modified for sake of better focusing on the target and quality.

1. **Is there a need to include a priori data quality targets (elements, measures, and values) in INSPIRE data specifications?**
    - ☐ Yes, for each dataset addressing the same set of requirements
    - ☐ Yes, but only for those datasets where achieving interoperability requires so
    - ☐ No

    If no, please go to question 4. If yes, please answer questions 2 and 3.

[18] This needs revision, whether the right place for reporting positional accuracy is the "lineage". DQ-Usability seems better fit for the purpose.

2. **Please, indicate the theme and whether these targets should be addressed by mandatory requirements (M) or recommendations (R)? Please, include justification if necessary**. In case you know formally specified and well established user requirements (e.g. NATO STANAGs, LPIS, etc) please include in the justification line.

Name of the data theme	Condition M/R	Justification / Comments

(Extend table if required.)

3. **Please, indicate the data quality elements, measures, and the target results to be used** (add as many lines as needed). Please fill a separate table for each data theme to which *a priori* DQ requirements/recommendation apply.

Name of the data theme

DQ element	DQ measure	Targeted result value	Comments

(Extend table if required.)

4. **Do you think that further theme specific mandatory metadata elements have to be specified in INSPIRE?**

☐ Yes ☐ No

If no, please go to question 6. If yes, please answer question 6

5. **Please, indicate the theme and justify these mandatory requirements from user's point of view.**

Name of the data theme	Metadata element	Justification / Comments

6. **What is the best way to generate DQ metadata** about the data that has been made conformant to the INSPIRE data specifications, i.e. after the necessary data transformations? Please mark more, if appropriate.
   - ☐ Keep the original metadata
   - ☐ Generate new metadata based on calculations, quality inspection by appropriate sampling.
   - ☐ Keep the original metadata and described as process step in MD_lineage (transformations performed with their possible effect on data quality)
   - ☐ Generate new stand alone qualitative report about the data quality

7. **How conformity should be reported?** (Please, mark more, if appropriate)
   - ☐ Only at data set / data set series level, as required by the INSPIRE Metadata Regulation
   - ☐ In addition to data set (series) level, conformity has to be reported for a set of key specification elements defined by the related implementing rule and/or data specification
   - ☐ In addition to conformity with the implementing rule / related data specification conformity with well-known user requirements has to be recorded.

8. **Do you think that reporting on data quality in INSPIRE should follow a template?** (Please, mark more, if appropriate)
   - ☐ Yes, for lineage. The template should be part of the data specification.
   - ☐ Yes, for usability. The template should be part of the data specification.
   - ☐ Yes, for a standalone evaluation report. The template should be part of the data specification.
   - ☐ No. Data providers know better what is important to communicate about their data.

6.2 Preliminary conclusions

The first round of discussions in the Krakow workshop and further studies in the field have lead to updating the discussion paper and arriving to preliminary conclusions where a consensus is being outlined.

1. The terminology used for describing data quality has to be harmonised. Whenever possible, the terms have to be taken from the related ISO 19xxx standards. When new terms, especially data quality elements and measures have to be introduced other authentic sources (standards on metrology, DQ terms defined by international organisations) have to be used. The DQ terms have to be introduced in the INSPIRE Glossary.

2. In order to establish a common approach in INSPIRE the GCM has to be complemented, if possible, with a quality model with appropriate references to the related ISO standards.

3. A priori data quality requirements and recommendations have to be based on appropriate use-cases and user requirements. DQ requirements have to be set with caution; they may exclude data from the infrastructure. Requirements should stem directly from the Directive, other binding European legal acts, or very strong / well-accepted user requirements. Logical groupings of DQ requirements or recommendations can constitute basis for (user-defined) conformance classes.

4. *A priori* requirements on data quality and related MD should be more stringent than those on thematic data themes. This is justified by the commonly used object referencing that plays important role in geographical referencing of thematic data and achieving overall consistency in INSPIRE.

5. The potential of DQ_ConformanceResult should be better understood and exploited. Beyond the mandatory conformance class stemming from the ISDSS Regulation (conformity with the implementing rule) other classes can be specified with more ambitious targets. In line with the INSPIRE MD regulation and the upcoming ISO 19157 only two values (pass: Boolean) are allowed when conformance is evaluated. When a data specification element is very significant from the point of usability conformity with that element should be reported separately.

6. Conformance testing should be based on objective criteria that are traceable, quantified, and linked to specific DQ elements. The target results for a priori DQ requirements (DQ_QuantitativeResult) have to be accompanied, when appropriate, with values for tolerances. The assembly of the necessary steps in conformance testing has to be included in the abstract test suits.

7. The number of metadata elements for evaluation and use should be kept reasonable. The TWGs have to pick the most representative metadata and specify which the best way of reporting is selecting between the DQ_ConformanceResult, DQ_QuantitativeResult, DQ_DescriptiveResult, or QE_CoverageResult options.

8. The transformations necessary to reach interoperability within the infrastructure may lead to deterioration of data quality, which requires updating metadata about the data quality. This can happen with full inspection, sampling, or indirect evaluation. Since the first two methods may require substantial investments, indirect evaluation using original metadata and description of the process step of the transformation can be used.

9. Lineage should not be misused. According to ISO 19115 lineage is information about the events or source data used in constructing the data specified by the scope. Even tough data quality is an aggregate of lineage and DQ element, lineage cannot be the container of data quality elements that are not specified elsewhere in the data specification. Lineage should be limited to the lineage source and process step.

10. Additional DQ information not specified by the DQ elements of ISO 19157 should go in DQ_Usability (with full definition of the element), or in a stand alone DQ report.

11. It is highly desirable to structure descriptive reports. The eventual theme specific templates both for lineage and usability should be harmonised across the data themes whenever possible. This may help user's orientation and machine translation.

12. Positional accuracy is of prime interest only for a couple of users and use-cases. When this DQ element is set in the a priori requirements it should be related to the scale.

The points above are only first steps towards the final resolutions and recommendations that the discussion process is expected to yield. In addition to the questions in the previous chapter comments on these points are welcome from the contact points of the Member States.

7 References and Further Readings

References

INSPIRE Directive
http://eur-lex.europa.eu/JOHtml.do?uri=OJ:L:2007:108:SOM:EN:HTML

INSPIRE Generic Conceptual Model
http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/D2.5_v3_3.pdf

INSPIRE Methodology for the Development of Data Specifications
http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/DataSpecifications/D2.6_v3.0.pdf

INSPIRE Data Specifications – Guidelines (Annex I)
http://inspire.jrc.ec.europa.eu/index.cfm/pageid/2

Further readings

In order to facilitate discussion on the data quality and metadata here are some references from different theme communities:

GEO Task DA-09-01 Data Management Subtask a.: GEOSS Quality Assurance Strategy:
http://www.grouponearthobservations.org/cdb/geoss_imp.php
http://qa4eo.org/

World Meteorological Organisation:
www.wmo.int/pages/prog/www/WDM/wdm.html

Global Spatial Data Infrastructure Association:
www.gsdi.org/gsdiconf/gsdi11/papers/pdf/283.pdf

Quality Assurance Framework for Earth Observation:
http://lpvs.gsfc.nasa.gov/PDF/qa4eo_guide.pdf
http://qa4eo.org/documentation.html

INSPIRE, Data Quality and SDIs:
www.directionsmag.com/article.php?article_id=3380

***Q-KEN - EuroGeographics WG on quali*ty:**
www.eurogeographics.org/about/quality

ESDIN – EU eContentPlus project:
www.esdin.eu