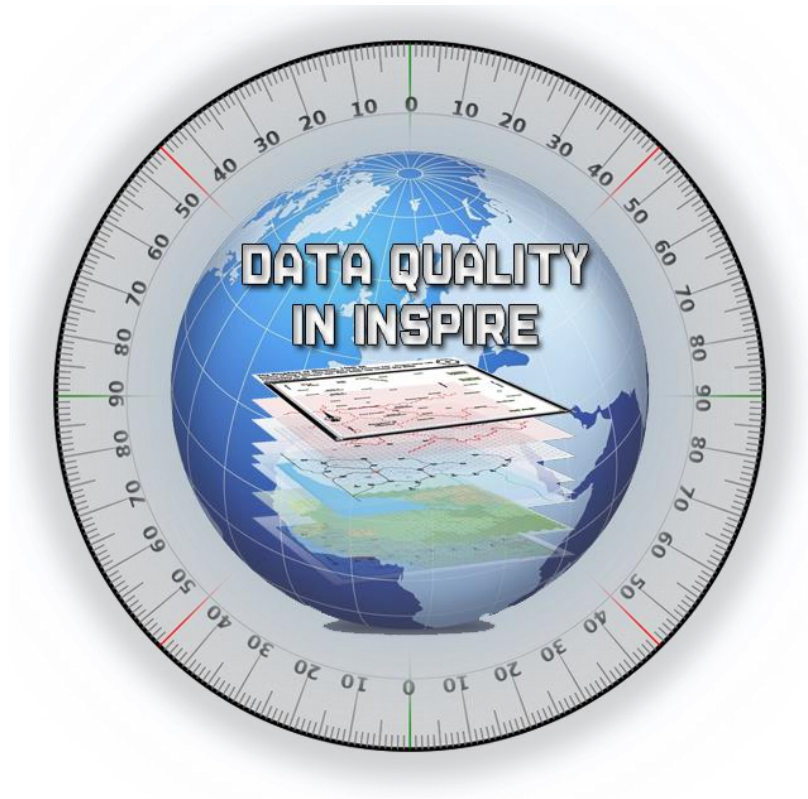


Data Quality in INSPIRE: Balancing Legal Obligations with Technical Aspects



Katalin Tóth
Robert Tomas
Vanda Nunes de Lima
Vlado Cetl

2013

European Commission
Joint Research Centre
Institute for Environment and Sustainability

Contact information

Robert Tomas

Address: Joint Research Centre, Via Enrico Fermi 2749, TP 262, 21027 Ispra (VA), Italy

E-mail: robert.tomas@jrc.ec.europa.eu

Tel.: +39 0332 78 5426

Fax: + 39 0332 78 6325

<http://ies.jrc.ec.europa.eu/>

<http://www.jrc.ec.europa.eu/>

This publication is a Technical Report by the Joint Research Centre of the European Commission.

Legal Notice

Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of this publication.

Europe Direct is a service to help you find answers to your questions about the European Union

Freephone number (*): 00 800 6 7 8 9 10 11

(*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed.

A great deal of additional information on the European Union is available on the Internet.

It can be accessed through the Europa server <http://europa.eu/>.

JRC83209

EUR 26097 EN

ISBN 978-92-79-32532-8 (pdf)

ISBN 978-92-79-32533-5 (print)

ISSN 1018-5593 (print)

ISSN 1831-9424 (online)

doi: 10.2788/9648 (pdf)

Luxembourg: Publications Office of the European Union, 2013

© European Union, 2013

Reproduction is authorised provided the source is acknowledged.

Printed in Italy



Data Quality in INSPIRE: Balancing Legal Obligations with Technical Aspects

Katalin Tóth
Robert Tomas
Vanda Nunes de Lima
Vlado Cetl

All from European Commission, Directorate General - Joint Research Centre

Keywords: data quality, spatial data infrastructure, metadata, conformance, interoperability

Executive summary

Geoinformation is increasingly being shared by many users across different fields and applications. The first milestone is marked by the possibility to combine and overlay different datasets in one spatial environment, and the second by the diffusion of web services that support the linking of information to a geographic location and sharing it with the widest possible audience (Tóth and Tomas, 2011).

Facilitating access to and reuse of spatial data is the main goal of spatial data infrastructures (SDIs). In Europe, the legal and technical framework for European SDI is defined by the INSPIRE (Infrastructure for Spatial Information in the European Community) Directive (2007/2/EC) and its related implementing rules and technical guidelines. It is assumed that INSPIRE is built on existing data originating from disparate sources.

Ideally, an SDI provides access to data in interoperable way, i.e. without the need for the specific *ad-hoc* interaction of humans or machines. The interoperability targets are formalised in **interoperability (data) specifications** that follow the structure of the data product specifications defined in the EN ISO 19131:2008 standard.

Even though the term “data quality” may seem to be self-explanatory, it is rather difficult to discuss because of the prevailing assumptions, incoherent terminology, and the diverging viewpoints to which it is subject. This report describes how data quality (DQ) was addressed during the development of the INSPIRE implementing rules and technical guidelines. This development process, which started in 2005 with the drafting of the conceptual framework, continued with the interoperability specification development for Annex I data themes in 2008-2010, and was finished with the definition of specifications for Annex II and III in 2013.

This report integrates the results of the work of the INSPIRE Data Specification Drafting Team, the JRC Data Specification Support Team, the INSPIRE Thematic Working Groups and INSPIRE the Data Quality Expert Group. This latter group was composed of specialists nominated as DQ contact points by the Member States. They were charged with organising the discussion on data quality at national level with the involvement of competent authorities.

The DQ contact points put forward the consolidated country results for INSPIRE were they were considered in the technical guidelines.

In order to avoid the difficulties in understanding that stem from the multidisciplinary environment of the stakeholders, this report also explains the similarities and differences between the notions of DQ in classical data production and in SDIs. In addition, because of the specific setup of INSPIRE, where Member States of the EU are obliged to fulfil requirements presented in legal and technical terms, the report also explains what is meant by the term ‘conformity’.

In conclusion, the report summarises the balance between legal requirements, stakeholders’ expectations, and willingness of data providers to share good quality data in an interoperable way. It also underlines the advantages of SDIs that take quality into consideration, and gives insight into potential further areas of research.

Contents

Executive summary	1
List of abbreviations.....	4
Introduction	5
1. Data Quality in production.....	8
2. Data Quality in SDIs	11
3. Data Quality in INSPIRE.....	15
3.1. Overview and challenges.....	15
3.2. Scope and user requirements	16
3.3. Interoperability specifications	17
3.4. Selection of datasets	21
3.5. Data and metadata transformations	23
3.6. Conformance testing.....	23
3.7. Metadata production and maintenance	25
3.8. Further improvements.....	25
4. Conclusions.....	27
5. Acknowledgments.....	28
Annex A : DQ requirements in INSPIRE and the corresponding formal DQ elements	29
Annex B Conformity of Hungarian Cadastral Parcel data.....	30
Annex C Structure of ATS proposed for INSPIRE.....	33
Annex D Data quality and metadata questionnaires used in INSPIRE.....	34
Annex E Contributors.....	38
Annex F Bibliography and References	39

List of abbreviations

ATS	Abstract Test Suite
DQ	Data Quality
EC	European Commission
EU	European Union
GEOSS	Global Earth Observation System of Systems
GCM	Generic Conceptual Model
GI	Geographic Information
GIS	Geographic Information Systems
GML	Geography Markup Language
INSPIRE	Infrastructure for Spatial Information in Europe
IR	Implementing rule (a.k.a. Commission Regulation)
ISO	International Standards Organisation
MD	Metadata
SDI	Spatial Data Infrastructure
TC	Technical Committee
TG	Technical Guidelines (a.k.a. Data/interoperability specification)
TWG	Thematic Working Group
UML	Unified Modelling Language
WG	Working Group

Introduction

Today, geoinformation is collected, processed, and used in domains as diverse as hydrology, disaster mitigation, spatial planning, statistics, public health, geology, civil protection, agriculture, nature conservation, and many others. The challenges regarding the lack of availability, quality, organisation, accessibility, and sharing of spatial information are common to a large number of policies and activities, and are experienced across the various levels of public authority in Europe. Spatial Data Infrastructures (SDIs) propose organisational and technical measures to search, find, and reuse spatial data collected by other organisations (Tóth et al., 2012).

It is easier to reuse spatial data when information about their quality and fitness for use is available, and when technical and legal barriers for integrating these into the user systems are removed. The first condition, quality, requires that rich and meaningful metadata be used, while ‘fitness for use’ requires the involvement of technical arrangements that ensure interoperability. INSPIRE provides the legal framework for reusing spatial data, and obliges public entities to share their data and metadata that fall within the scope of the Directive. Moreover, it requires interoperability, which is “the possibility for spatial datasets to be combined and for services to interact, without repetitive manual intervention, in such a way that the result is coherent and the added value of the datasets and services is enhanced” (European Commission, 2007).

ISO 9000 defines quality as the “degree to which a set of inherent characteristics fulfils requirements”. Since fitness for purpose and interoperability are fundamental user requirements, we can conclude that these aspects are closely related to data quality (DQ). A lack of quality that manifests as incompatibility, incomparability and different levels of accuracy of data may jeopardise interoperability. Therefore, a well-functioning SDI needs to incorporate clear and unambiguous quality requirements in order to ensure the accessibility of the maximum amount of data with reasonable quality.

Fitness for purpose is fundamental for data sharing. Each SDI must therefore provide the means for documenting data usability, which also incorporate DQ. For the convenience of the users, these descriptions should be easy to understand and similar across all stakeholder communities participating in a given SDI initiative. SDIs should also fix a common assessment methodology in order to be able to compare usability and DQ across these communities.

DQ in geoinformation is not at all a new concept. Topographic surveying, mapping, Global Navigation Satellite Systems (GNSS) measurements, cartography, etc. are fields in which quality has been taken into account from the very beginning. Under the continuous drive to produce better (more detailed, more complete, more accurate, etc.) data, it is not surprising that these fields largely contributed to rigorous product specification methodology, quality assurance, adjustment computations, statistics, and conformance testing. This experience has been embedded in standards dealing with the quality of geoinformation delivered by TC211 of ISO and the DQ Working Group of the Open Geospatial Consortium (OGC).

The significance of quality is also recognised within individual EU environmental Directives, where the importance of state-of-the-environment monitoring to ensure successful policy implementation is frequently underlined (European Commission, 2013a). However, documentation regarding the quality and usability of environmental data is often not available.

Nowadays, the wide spread of electronic spatial data, the Internet, and computer literacy allow data from diverse sources to be easily shared and combined. However, various communities approach DQ very differently. Some apply standardised methodology about the quality and usability of data production and publishing metadata (MD), while others do not share any such information. The situation is further complicated by the fact that there is a lack of shared terminology for the documentation of DQ. For example, various terms may be used for the same subject, or just the opposite: the same term may refer to completely different concepts.

Given that reusing data in an interoperable way is the main objective of an SDI, the question to be addressed is whether it is sufficient to address DQ and usability at the MD level, or whether it should be included at the stage of defining the goals for interoperability. The interoperability targets of SDIs may differ in their level of ambition, as manifested in the number of interoperability components they address (INSPIRE Data Specification Drafting Team, 2013)¹. Furthermore, the transformations that must be carried out to present data according to the interoperability agreements may also influence their original quality and usability. However, sufficient information is not yet available to properly assess the impact of these issues.

The Member States of the EU are legally obliged to fulfil the requirements of INSPIRE. This adds another layer of complexity in DQ conformity. On the one hand, conformity can be regarded as adherence to the legal obligation, while on the other it is closely linked to the technical content. Separating mandatory requirements from technical implementation recommendations presents another challenge to be resolved within framework of INSPIRE.

All in all, INSPIRE highlights the challenges that SDIs are likely to face in matters of DQ. These challenges are amplified by the increased heterogeneity of data, which is the natural consequence of different application domains, scientific methodologies, languages, technical standards, and national jurisdictions. Agreements on DQ can be achieved more easily in the framework of specific SDIs (e.g. those at the national or regional level, or within a thematic community).

This report describes how DQ was addressed in INSPIRE at each phase of the data specification development process. The work got underway in 2005 with the drafting of the conceptual framework. The interoperability specification development for Annex I themes, and later for Annexes II and III, defined the next iterative DQ cycles. The Thematic Working Groups (TWGs) responsible for the development of technical specifications were supported by the INSPIRE Data Quality Expert Group, composed of DQ specialists nominated by the Member States as contact points. This group was responsible for organising discussions at national level with the involvement of the competent national authorities.

In order to understand the INSPIRE approach, the difference in the role of DQ in traditional data production and SDIs must be explained. Sections 2 and 3 of this report address this aspect, paying special attention to clarifying the terminology used. Section 4 summarises the legal requirements of INSPIRE, the expectations of the stakeholders, and the dilemmas and problems to be resolved, and presents the agreements reached and how these are put into

¹ Refer fig. 2 in the document

practice. In conclusion, the report describes the advantages of quality-conscious SDIs and gives insights into potential further improvements that could be made in the area.

1. Data Quality in production

The classical process of geoinformation acquisition (e.g. topographic, cadastral, soil surveying/mapping, geodesy) is defined by rigorous workflows that can be divided into five distinct phases:

1. Study of user requirements;
2. Development of specifications;
3. Production;
4. Conformance testing;
5. Metadata publication.

Each of these phases should be accompanied by a quality assurance procedure, which documents each step as metadata (MD).

In the broad sense, Data Quality (DQ) covers usability, the technical characteristics of data, aspects of DQ evaluation, conformity, and the documentation of the findings as MD. Ideally, MD is incrementally accumulated after each phase of the production cycle (see Figure 1), in order to provide full information to the users about the properties of data.

As part of the production cycle, each phase provides input for the consecutive one, abstracting real world phenomena in a dataset and collecting information about the dataset in a set of MD that is shared with a wider community with a view to potential reuse.

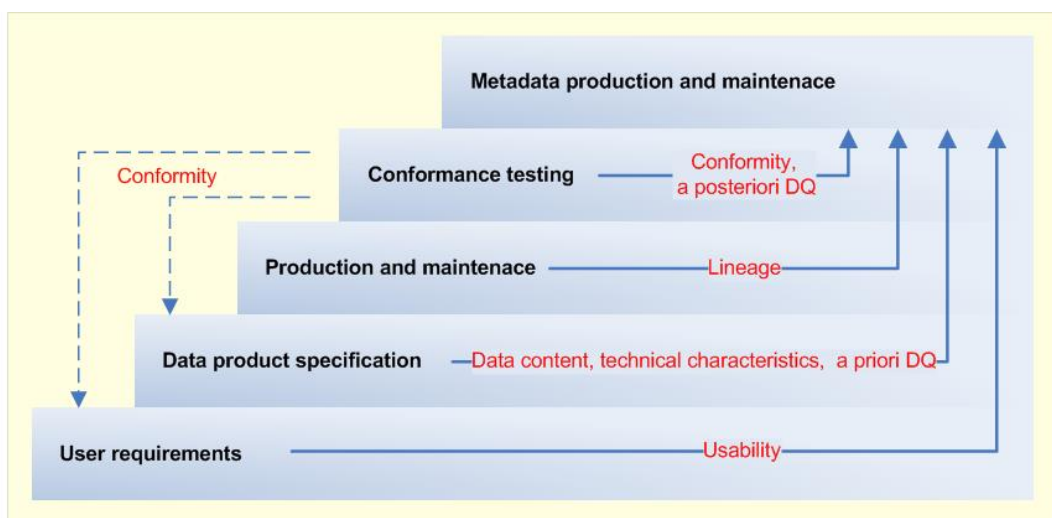


Figure 1: Incremental accumulation of data quality information in production processes

Studying **user requirements** is the initial phase, since every new dataset should cover gaps in the existing data supply or anticipate emerging use cases. Even the maintenance of traditional products – such as topographic or cadastral data – cannot skip this step, as modernisation, technological development, or new organisational setups may influence the needs of even the regular customers. The assessment of user requirements is frequently documented using formal templates², which helps to compare different products.

² See Annex B of the INSPIRE data specifications

In addition to giving a solid starting point for data specification development, a formal use case directly provides input for the *usability* MD element by describing, as a minimum, the default use case and the default group of users. Naturally, usability can be incremented when the data is reused by others and/or for other purposes.

The **data product specification** translates user requirements into data content and technical characteristics. It includes the data model, the identifiers of the datasets, the applicable coordinate reference and projection systems, encoding, portrayal, etc. EN ISO 19131:2008 defines the standard content and layout of the data product specification. In terms of DQ, the specification delivers the quality model, i.e. the

- definition of the DQ units (i.e. the combination of DQ scope and the applicable DQ elements,
- DQ measures and their parameters,
- *a priori* DQ result targets,
- value types of the results,
- methods for aggregating DQ results,
- scope of the quality evaluation procedure,
- accepted evaluation methodologies.

A priori data quality target results are very important: they set requirements in terms of completeness, consistency, accuracy, and of other user-defined aspects to be fulfilled in course of data production. For well-established products, standards or other regulations drive the data specification and production processes, fixing the technical characteristics of the datasets and setting strict target results for the selected DQ elements and the corresponding measures. Data specifications following EN ISO 19131:2008 contain a separate section on the MD content and documentation with a twofold purpose: to provide input for conformance testing, and to share DQ information with potential users.

The data specification phase contributes to the MD collection by specifying **MD for discovery** (identifiers of the dataset, producer, validity, etc.) and **MD for evaluation and use** (data content, scale, resolution, Coordinate Reference System (CRS), spatial representation, etc.). For the latter, some information items (such as feature concept dictionaries, feature catalogues, CRS, etc.) can be referenced, especially when they are maintained in registries. It is highly recommended to extract and document those MD values that are defined in the data specification and are invariant in the production phase (such as the identifier of the dataset, resolution, data content, etc.).

The values of other MD elements (such as those on data quality, or lineage) can be determined after the production process. Therefore, **MD publishes *a posteriori* quality results**, which are based on:

- de-facto evaluations and specific aggregation rules applied to the given dataset,
- other knowledge of the author/producer expressed as non-quantitative information.

In the first case, DQ metadata includes one or more DQ elements, each of which is expressed using selected DQ measures and the related DQ results. In the second case, descriptive notes or reports³ provide rather informal evaluation statements. These, frequently subjective, statements should not be underestimated; they are often better understood by the users than are formal DQ elements.

³ For this purpose ISO 19157 offers DQ_DescriptiveResult metadata element

The **data production**/acquisition and **updating processes** should be designed in such way that all the technical parameters and *a priori* DQ requirements can be fulfilled. **Quality assurance** (QA) plays crucial role, as it includes the selection of the appropriate data collection/measurement methodology, instruments, protocols, documentation, on-the-fly QA measures, quality control, and the participation of adequately trained personnel. The thorough information about what happened in course of data production/update should be published in the lineage MD element, or in an informal DQ report.

Conformance is defined by EN ISO 19105:2005 as the fulfilment of specified requirements. For geoinformation products, the specification that was used for the production is also the basis for conformance testing. In an ideal scenario, conformity with the data specification also implies conformity with the original user requirements. Therefore, as shown in Figure 1, conformance testing provides feedback to the specifications and the user requirements. In conformance testing, the specified **requirements** (expressed in data content, technical characteristics and DQ requirements) are compared with the *de-facto* values (results) achieved in the production phase, and are published as MD elements.

The scope of conformance evaluation may relate to a single specification element (e.g. the application schema, data capture rules, or selected data quality elements, etc.), to a conformance class (group of predefined elements, e.g. the complete set of DQ elements and their target results), or to all requirements included in the specification. The results of conformance testing, with reference to the specifications, are published as separate MD elements.

In addition to conformity with the data product specification, conformity with user requirements can be also assessed and communicated. To demonstrate that a dataset can be used for purposes other than those for which it has been produced is especially important for reusing the data. If user requirements are well documented, i.e. if they can be referenced, the value of the conformity MD element can be declared, as in the case of conformity with a specification. If not, it is better to use a free text description, as explained earlier.

To facilitate conformance testing and to ensure the comparability of the conformity MD values, the data product specification also contains the **abstract test suite** (ATS). Any product that claims conformance to the specification as a whole must pass all the tests described in the ATS. Each test refers to the corresponding requirements in the specification lists, the applicable tests, the quality measures and their threshold values.

A dataset can conform to one or more specifications at the same time. For example, a dataset that satisfies the product requirements specified in a national regulation may also fulfil the requirements of national and/or international standards. In order to fully inform users about the conformity of data, it is advisable to declare conformance with all the specifications against which the data has been tested.

Before concluding this section, it is necessary to point out that there is a lot of confusion in the terminology concerning DQ. In common day usage, DQ is not always distinguished from usability. When using DQ versus usability, professionals should use the correct terminology and should correctly map users' communications to specification elements. Similarly, MD is frequently used as synonym of data quality. As explained before, MD is a much broader term; it describes usability, technical, and data quality aspects. Unlike data product specifications, MD always communicates *a posteriori* values (European Commission, 2012).

2. Data Quality in SDIs

Until the 1980s, data were produced by professionals according to pre-defined data specifications, standards, or legal regulations, and the datasets produced were delivered to the specific (professional) users who ordered the work. In general, very little was done for reusing data for other purposes. The situation of scientific information collected for experimental or exploration purpose demonstrates this situation very well. In spite of the fact that such information might be relevant for reuse, in most the cases no documentation of quality or even specifications for data collection were available.

However, new technologies and increasing computer literacy now allow more and more people to get involved and use geoinformation for an ever growing number of applications. This tendency can be demonstrated, for instance, through “volunteered geographic information” (Goodchild M.F., 2007), where ‘non-professionals’ spontaneously collect and publish geographic data.

Data is also frequently used for decision making. Decision makers who may be well trained in the thematic area are less familiar with geoinformation (GI). Obviously, this new setup multiplies the demand both for easy-to-use data sharing and documentation methods.

In the case of INSPIRE, SDIs have two high-level objectives:

- To provide as wide **access to existing data** as possible,
- To enable **interoperability across spatial datasets**, regardless of their thematic scope and origin.

While data sharing is facilitated by the increasing number of geoportals (web services), proper documentation of DQ and data usability is lagging behind. Standardised MD is a smart solution for GI professionals. For other user communities, MD should be tailored according to their needs and conventions.

We will show later on that the two abovementioned objectives can be in conflict with each other. In order to balance them, it is necessary to clarify how the quality and MD aspects of SDIs differ from those that are well (better) known for data production.

We also assume that **data collection is not the (primary) purpose of an SDI**. Accepting this condition, the data production cycle is replaced by a transformation workflow, which represents original data and metadata in a way that is compliant with the agreed interoperability targets. This transformation workflow can be divided into six steps, as shown in Figure 2.

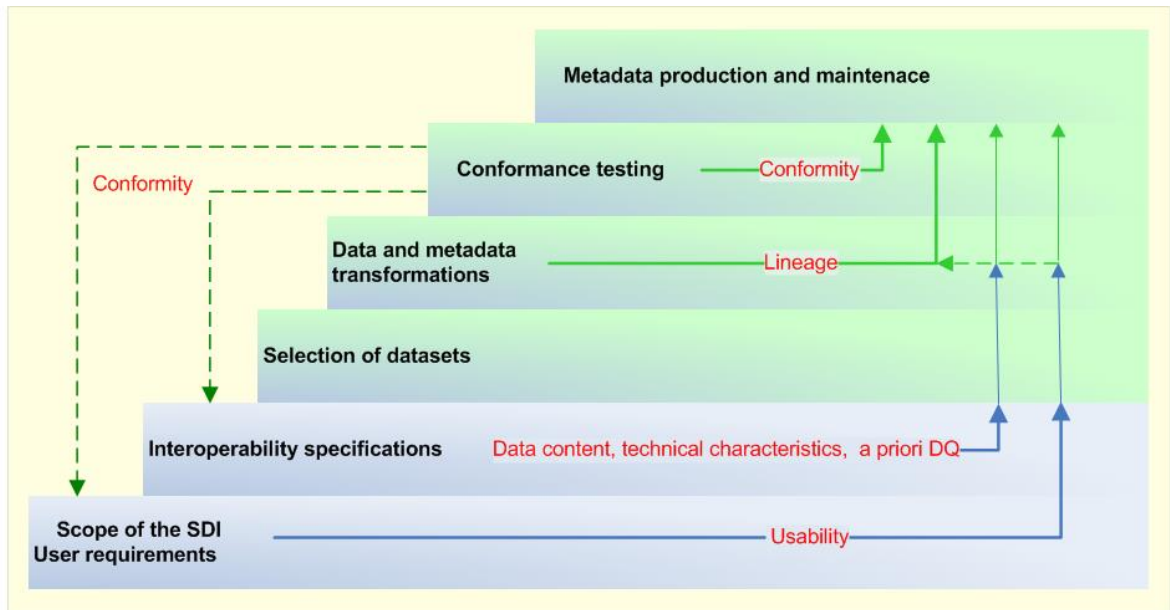


Figure 2: Data publication in SDIs

As in the data production cycle, each step in this workflow contributes to accumulating MD, which should be published in the infrastructure to promote possible reuse of data. Also here, the DQ can be understood in a broad sense, addressing usability, data content, technical characteristics, lineage, and conformity. While these terms are very similar to those used in data production, there are differences regarding their provenance. We will explain this in the next sections.

The development of the infrastructure should be driven by the requirements of the potential users. **User requirements** are important external benchmarks that help find an optimal balance between interoperability and the associated costs. Over-ambitious interoperability requirements may lead to excessive costs or the impossibility of data sharing. Weak requirements may not sufficiently promote interoperability or can create the possibility that a dominant data provider or user could force their particular technical solutions instead of seeking wider consensus (INSPIRE Data Specification Drafting Team, 2008).

High-level user requirements and consensus building play key roles in scoping the SDI, both in terms of data themes (which datasets will be affected) and the degree of harmonisation (the degree that best adds value for data sharing and balances likely costs and benefits). In general, it is easier to reach interoperability when the geographic or thematic scope of the SDI is narrow, allowing for more standardisation. As high-level interoperability requirements add value for the users, they should appear in the usability MD element of those datasets that are retrieved from the infrastructure.

As in data production, user requirements are formalised in technical specifications. In contrast to data product specifications, these technical specifications describe not specific products, but the *interoperability targets* to be achieved when data is delivered via the infrastructure. The **interoperability specifications** define the data content and structure, spatial representation and other technical characteristics agreed by the stakeholders. Even though the structure of the interoperability specification is very similar to that of the data product specification, some specification elements may be omitted. For example, scale and

resolution are not defined in INSPIRE, since the specified spatial object types and data types apply at all levels of detail, from local to national.

Naturally, all technical aspects documented in the interoperability specification provide input for the following steps, including MD documentation. It should be noted that the interoperability specifications only provide the MD model with the appropriate MD content and corresponding elements. The values of the specified MD elements should be assigned for each dataset after the necessary transformations are carried out.

In terms of *a priori* DQ, the interoperability specification can contain both the DQ model and the target results. It should be noted that strict quality targets might be discriminatory. A case can be made for including datasets that do not fulfil such targets in the infrastructure, as an overly strict approach goes against the aim of wide data sharing. In the end, it is the user who should decide whether data is fit for a particular purpose. Nevertheless, all SDIs must face the dilemma represented in Figure 3.

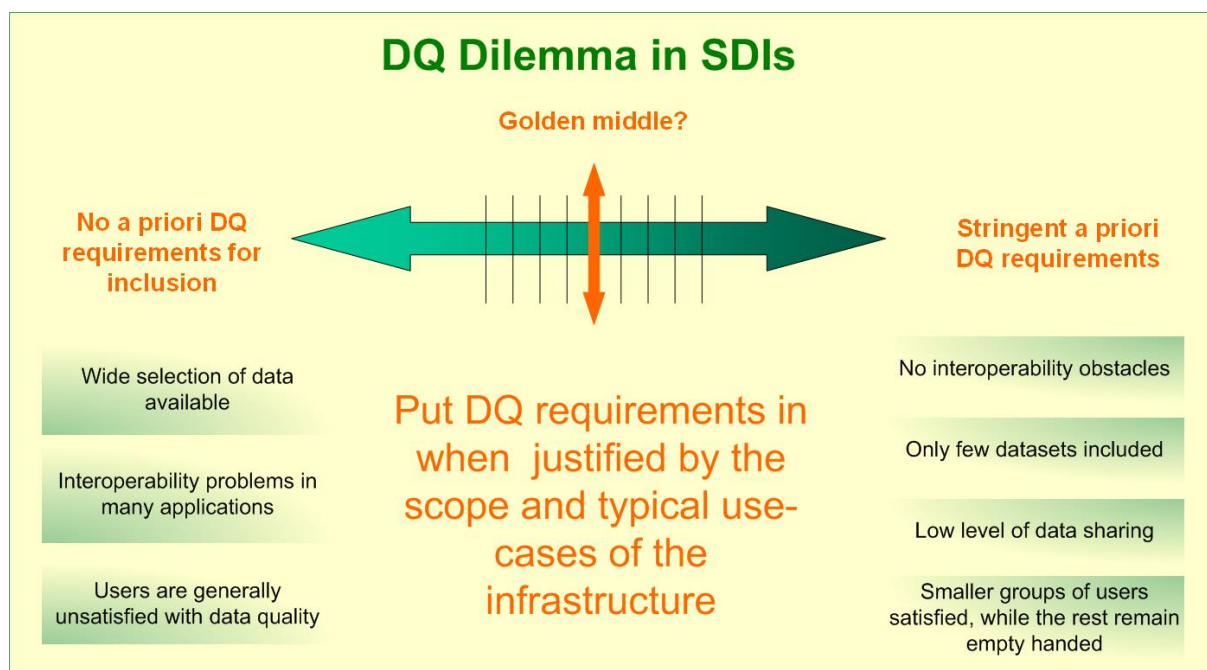


Figure 3: Balancing DQ with data sharing

The presence of *a priori* DQ in an SDI may also depend on the nature of the data (data theme). An *a priori* DQ can be easily justified for reference data, which can be reused for linking thematic data. A complete and accurate framework of reference data constitutes the basis for object referencing, which strongly contributes to the overall data consistency within the infrastructure (INSPIRE Data Specification Drafting Team, 2013).

High-level user requirements and defining interoperability specifications should be clarified through consensus building, in which every interested party, both users and producers, should be involved. When the interoperability specifications are ready, they serve as the basis for the next step to be executed by the data providers. Knowing the data content and technical characteristics, data providers have to *select whether / which of their datasets / spatial object types can be included* in the infrastructure.

The selected datasets / spatial object types need to be **transformed** according to the interoperability specifications. Naturally, this process requires adequate quality assurance (QA) procedures, which require appropriate methodology, software tools, protocols, documentation, on-the-fly QA measures, final quality control, and well-trained personnel.

In addition to data, the original MD should also be transformed, as the MD model of the original datasets does not follow the MD model defined in the interoperability specification. The most usual MD transformations are to change the level of granularity (from object level to dataset level), or to aggregate MD (if possible) whereby several datasets are integrated to cover the data content defined in the interoperability specification. The experience of INSPIRE implementation may help clarify difficulties encountered in carrying out such transformation procedures.

The transformation phase provides input for *lineage*: the original lineage MD element(s) (i.e. information about the data source before the transformation) should be complemented by the description of the necessary transformations carried out. Transformations can also be referred as “process step” elements in the lineage. It must be stressed that the original purpose of the dataset must be documented as it provides meaningful information when reuse is considered.

Conformance testing takes place against the interoperability specifications for each data theme / application schema. In this exercise, all *a priori* requirements (data content, technical characteristics and DQ requirements) are compared with the *de-facto* values achieved in the transformation phase. The testing process is guided by the ATS, which is part of the interoperability specification (EN ISO 19131:2008). As for data production, the final result can be conformant with a set of pre-defined requirements (conformance class) or with the specification as a whole. The results are published as *a posteriori* MD elements.

To enhance the reusability of datasets, conformity with other specifications (e.g. national, international, and domain-specific standards) can also be indicated. In addition, conformity with user requirements can be given, as explained in the data production section. Data providers are recommended to update the usability MD element whenever they learn about a new usage of a dataset, or provide a tool⁴ to allow the users to share their observations on the usability.

At the end of the transformation process, a **full set of metadata** should be made available for the users of an SDI. However, it should be noted that not all users can understand formally specified MDs; some may find descriptive reports more useful. It is therefore recommended that MD be published at multiple levels – a core descriptive set for the general public and a more detailed MD set for professional users.

⁴ Reviews of products and services available on the web.

3. Data Quality in INSPIRE

3.1. Overview and challenges

The INSPIRE Directive aims to establish a European SDI that underpins the implementation of policies that have environmental impact. It defines legal obligations whereby the public authorities of the Member States must share data belonging to 32 (+2)⁵ data themes in an interoperable way. In order to provide a sound basis for a “wide range of applications”⁶, INSPIRE adopted a two-step process for defining the technical data requirements.

First, the Generic Conceptual Model (GCM) (INSPIRE Data Specification Drafting Team, 2013) was developed. This model defined various aspects of interoperability that are common to all themes, called harmonisation elements. Three of these elements are directly related to DQ in a broad sense: data quality, MD, and conformity. In the second step, theme-specific technical provisions were defined. These latter were presented separately for each theme following the unified methodology based on EN ISO 19131:2008 (INSPIRE Data Specification Drafting Team, 2008). Even though they are commonly referred to as data specifications, it should be noted that these technical provisions are not data products, but are interoperability target specifications.

Recognising that the user and data provider communities follow very different approaches in matters of DQ for every theme, the GCM did not elaborate a common DQ model, but rather pointed to the related standards of ISO TC 211⁷. The Thematic Working Groups (TWGs) responsible for specification development were therefore in a rather difficult situation. They had to accommodate a wide divergence of opinions, ranging from introducing strict DQ requirements for all data to the complete omission of any requirement. The same was true for MD: some groups did not see why MD other than that specified in the Metadata Regulation (European Commission, 2008) were necessary; while others were working hard on developing MD for evaluation and use.

On 14 December 2010, the INSPIRE Committee unanimously approved the Regulation on the Interoperability of Data Sets and Services for INSPIRE Annex I data themes. However, it was pointed out that a common approach for DQ must be defined. To this end, the Commission organised a DQ working group (WG), consisting of the contact points nominated by the Member States. The WG was chaired by the JRC and was complemented on *ad hoc* basis by experts representing other relevant initiatives (e.g. the Group on Earth Observations, EuroGeographics).

In order to collect input from the widest possible source, a multi-level approach was followed. The data specification support team of the JRC played a facilitating role in the process. It prepared explanatory notes, the DQ discussion paper, (European Commission, 2012) and surveys (see Annex D) that were distributed to the relevant organisations in the Member States by the DQ contact points. The DQ contact points were responsible for consolidating opinions (answers to the questionnaires) at national levels. Having analysed the national

⁵ In fact, there are 32 themes. The additional two describe coordinate reference systems and geographic grids to be used in any theme.

⁶ See point (14) of the preamble in INSPIRE

⁷ At the beginning ISO 19113, 19114, and 19138 for DQ, 19115 for MD. Later those related to DQ were replaced by 19157 and 19105 for conformance testing was introduced.

responses, the JRC organised face-to-face meetings to discuss issues on which there was a divergence of opinion. These workshops took place in Krakow (June 2011 at the INSPIRE conference), in Brussels (February 2012), and in Istanbul (June 2012 at the INSPIRE conference).

The work of the Thematic WGs (TWGs) and the discussions in the DQ WG made it clear that the INSPIRE data specification process suffered from the following problems:

1. Lack of established terminology (confusion of *a priori* DQ requirements with MD, usability with DQ, data quality elements with measures, treating positional accuracy as a synonym for data quality, etc.);
2. No strategy was available on the role of DQ in INSPIRE;
3. Lack of understanding of how to apply the quality principles that were well known in the production environment, but less familiar in the context of SDIs;
4. No common DQ model existed;
5. Lack of vision as to how conformity in the legal sense could be described in technical terms.

In the subsequent sections we will summarise how these issues were addressed using the logical framework of steps described in section 2. For completeness, it should be noted that this framework is one of the results of the DQ work carried out under INSPIRE, which interlinks various aspects of quality with the steps necessary to publish data in an interoperable way.

3.2. Scope and user requirements

The INSPIRE Directive explicitly defines the thematic scope of the infrastructure in the Annexes, while the generic scope is indicated in high-level statements of various articles and the preamble. The verbal formulation “Member States shall...”⁸ points to the expectation not only in terms of the data content, but also in terms of usability, data quality, conformity and MD.

The main requirement for **usability** is interoperability, which, according to Art. 3(7), adds the value of ‘coherent representation’ of spatial data. Combining spatial data from different themes and sources in a coherent way is a DQ requirement that goes beyond the requirement for consistency within a data theme (European Commission, 2007):

- Art. 8(3) requires “consistency between items of information which refer to the same location or between items of information which refer to the same location at different scales”;
- Art. 10(2) requires that “a geographical feature, the location of which spans the frontier between two or more Member States shall [...] decide on mutual content on depiction and position of such common features.”

These provisions set obligations to be met by the data providers.

The consequent application of the data modelling elements and other provisions of the Generic Conceptual Model (GCM) enforces cross-theme consistency. The common types (shared application schemas, object referencing, and constraints) provide a sound basis for standardised modelling in each theme. They also define a firm starting point for the technical implementation of different algorithms (such as edge matching, automatic model

⁸ The verbal clause implies legal obligation for all addressees of the Directive

generalisation, conflation, etc.) which is useful for the transformation of spatial data within the infrastructure.

3.3. Interoperability specifications

Arts. 8(2) and 8(4) lay down the conditions to be fulfilled at the level of the *interoperability specifications*, which define the **data content**, **technical characteristics**, and only one generic *a priori* **DQ requirement: logical consistency**. The majority of these conditions are included in the application schemas of the data themes.

Chapter 7 of the interoperability specifications of INSPIRE deals with DQ. Even though no UML documentation has been prepared, DQ has been formally defined for the data component of INSPIRE through the common template that was iteratively developed in the course of the specification development process (INSPIRE Thematic Working Groups 2010, 2013).

Guided by ISO/DIS 19157, the data specifications contain the DQ elements to be used in INSPIRE for the evaluation process. Since the INSPIRE MD regulation requires reporting at dataset or dataset series levels, the default DQ addresses these levels. However, justified by common practice, the level of spatial object types is also defined. For example, according to the Cadastral parcels specification, the positional accuracy can be reported at any level of the hierarchy of spatial object types: at boundary, parcel, or cadastral zone level.

The data quality elements, sub-elements and their recommended measures were defined with a view to

- Evaluate and document DQ properties and constraints of spatial objects, where such properties or constraints are defined as part of the application schema(s);
- Evaluate and document DQ MD elements of spatial datasets foreseen in chapter 8 of the data specifications;
- Specify requirements or recommendations about the targeted DQ results applicable for datasets related to a specific data theme.

Annex A of this document establishes the mapping of the DQ-related statements of the Directive with DQ elements and sub-elements as defined in ISO/DIS 19157. The list of commonly used DQ elements in the INSPIRE data themes is given in Table 1. This list is not exhaustive; it could be extended according to the needs of a specific data theme.

Table 1: List of commonly used DQ elements in INSPIRE

Data quality element	Data quality sub-element	Definition	Evaluation Scope
Completeness	Commission	Excess data present in the dataset, as described by the scope.	dataset series dataset spatial object type
Completeness	Omission	Data absent from the dataset, as described by the scope.	dataset series dataset spatial object type
Logical consistency	Conceptual consistency	Adherence to rules of the conceptual schema.	dataset series dataset spatial object type spatial object
Logical consistency	Domain consistency	Adherence of values to the value domains.	dataset series dataset spatial object type spatial object
Logical consistency	Format consistency	Degree to which data is stored in accordance with the physical structure of the dataset, as described by the scope.	dataset series dataset spatial object type spatial object
Logical consistency	Topological consistency	Correctness of the explicitly encoded topological characteristics of the dataset, as described by the scope.	dataset series dataset spatial object type spatial object
Positional accuracy	Absolute or external accuracy	Closeness of reported coordinate values to values accepted as or being true.	dataset series dataset spatial object type spatial object
Positional accuracy	Relative or internal accuracy	Closeness of the relative positions of features in the scope to their respective relative positions accepted as or being true.	dataset series dataset spatial object type spatial object
Positional accuracy	Gridded data position accuracy	Closeness of gridded data position values to values accepted as or being true.	dataset series dataset spatial object type spatial object
Thematic accuracy	Classification correctness	Comparison of the classes assigned to features or their attributes to a universe of discourse.	dataset series dataset spatial object type spatial object
Thematic accuracy	Non-quantitative attribute correctness	Correctness of non-quantitative attributes.	dataset series dataset spatial object type spatial object
Thematic accuracy	Quantitative attribute accuracy	Accuracy of quantitative attributes.	dataset series dataset spatial object type spatial object
Temporal quality	Accuracy of a time measurement	Correctness of the temporal references of an item (reporting of error in time measurement).	dataset series dataset spatial object type spatial object

Temporal quality	Temporal consistency	Correctness of ordered events or sequences, if reported.	dataset series dataset spatial object type spatial object
Temporal quality	Temporal validity	Validity of data specified by the scope with respect to time.	dataset series dataset spatial object type spatial object
Usability	--	Degree of adherence of a dataset to a specific set of requirements.	dataset series dataset spatial object type spatial object

The DQ elements listed in Table 1 determine how the datasets delivered for INSPIRE should be described *after* the necessary transformations for interoperability are carried out. Since each DQ element can be quantified by applying different DQ measures, each interoperability specification prescribes which measures must be used for a specific theme. Using common DQ measures in the various data themes improves the comparability of datasets in terms of quality.

The DQ model presented in the data specification is also applicable for setting the targeted results. As described in section 2, setting *a priori* DQ results requires careful balancing of the widespread availability of data (data sharing obligation) with the benefits of interoperability. This is why **INSPIRE**, by exception of logical consistency, **does not set *a priori* DQ requirements**, but gives recommended preferences for DQ results. The role of *a priori* DQ result recommendations should not be underestimated; they help guide future developments for data provider communities.

Chapter 8 of the interoperability specifications deals with MD. In INSPIRE, MD for discovery and MD for first level evaluation of a spatial dataset or spatial data series as required by the Directive, including issues of quality, validity, and conformity, are mandated by Commission Regulation 1205/2008. This approach is justified by the different timelines of the Member States' obligations for data sharing and interoperability. According to Commission Regulation 268/2012⁹, the data sharing obligation entered into force in April 2010. The first datasets (newly collected or intensively restructured within the scope of Annex I) were made available in an interoperable way as of December 2012.

The MD section of the data specifications, apart from including theme-specific details, gives recommendations on how to use the MD elements already defined in the Metadata Regulation (e.g. lineage, keywords, conformity). It also defines MD elements for evaluation and use that are common to each theme and facilitate interoperability.

The Conformity MD element defined in the Metadata Regulation requires the reporting of conformance with the Implementing Rule for interoperability of spatial datasets and services. In addition, it may also be used to document conformance to any other specifications comprising the INSPIRE interoperability specifications (published as technical guidelines), standards, other international and national specifications, and user requirements.

The Conformity MD element includes two sub-elements: the Specification (a citation of the specification), and the Degree of conformity. The values of the latter can be "conformant" (if

⁹ <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:083:0008:0009:EN:PDF>

the dataset is fully conformant with the cited specification), “not conformant”, or “not evaluated”.

Lineage is an important source of information since it describes the history and/or the overall quality of the spatial dataset. Where appropriate, it may include a statement as to whether or not the dataset has been validated or quality assured, whether it is the official version (if multiple versions exist), and whether it has legal validity. In order to improve interoperability, proposed templates and instructions for using this MD element may be included in the theme specifications. Templates for lineage also help to avoid overloading this element, balancing formal MD (outside lineage) and free text descriptions of lineage (INSPIRE Data Quality Working Group, 2010).

After careful analysis of the user requirements and the interoperability specifications, the following common MD elements were determined for generic use (INSPIRE Thematic Working Groups, 2013):

- Coordinate Reference System: Description of the coordinate reference system(s) used in the dataset.
- Temporal Reference System: Description of the temporal reference system(s) used in the dataset. This element is mandatory only if the spatial dataset contains temporal information that does not refer to the default temporal reference system.
- Encoding: Description of the computer language construct(s) that specify the representation of data objects in a record, file, message, storage device, or transmission channel.
- Topological Consistency: Correctness of the explicitly encoded topological characteristics of the dataset as described by the scope. This element is mandatory only if the dataset includes types from the INSPIRE Generic Network Model to ensure centreline topology (connectivity of centrelines) for the network.
- Character Encoding: The character encoding used in the dataset. This element is mandatory only if an encoding language that is not based on UTF-8 is used.
- Spatial Representation Type: The method used to spatially represent geographic information.

These MD elements give essential information to users regarding the interoperable usage of a dataset delivered for the infrastructure. Of course, the descriptions included in these metadata elements can be replaced by appropriate references to freely accessible registers, web-pages, etc.

In addition to common MD elements, domain-specific metadata could be defined for each data theme. According to the general approach accepted in INSPIRE, when a specification contains *a priori* DQ requirements or recommendations, the results of their evaluations should be published using the DQ elements and measures defined in the quality model (Chapter 7 of the interoperability specification).

If the specification does not contain any provisions for a specific DQ element, Chapter 8 defines in general how to report the results of the evaluation of the logical consistency. Naturally, as explained earlier, any specification could opt for qualitative-descriptive forms - MD DQ statements - instead of formal elements.

In order to ensure the comparability of results and a unique approach for conformance testing, the interoperability specifications also contain an abstract test suite for each theme. This suite

includes a set of tests to be applied on a dataset in order to evaluate whether it conforms to the requirements of the specification.

The definition of the interoperability specifications of spatial data is the last step in the development phase of INSPIRE. Once these are accepted, the SDI enters into the implementation phase. While the development phase is a consensus-building process involving many actors, the main players of the implementation phase are the data providers.

3.4. *Selection of datasets*

In an SDI, datasets are generally selected for publication according to their content and the *a priori* requirements. Art. 4 of INSPIRE explicitly defines which datasets should be published. The deadlines for fulfilling data-sharing obligations for a specific theme differ from those for interoperability. There is therefore a period during which publication data that is not compliant with the Regulation on Interoperability of Spatial Datasets and Services is fully acceptable. Of course, compliance with the Interoperability Regulation marks a qualitative jump especially for “European” users, who need to integrate data across the Member State boundaries.

In order to fulfil the data sharing and the interoperability obligations, the responsible organisations of the Member States should define which datasets / spatial object types fall within the scope of the Directive, what data transformations are necessary, and how they can be best transformed. The following diagram (Figure 4) represents an overview of the four scenarios identified for conforming to INSPIRE’s data sharing and interoperability obligations.

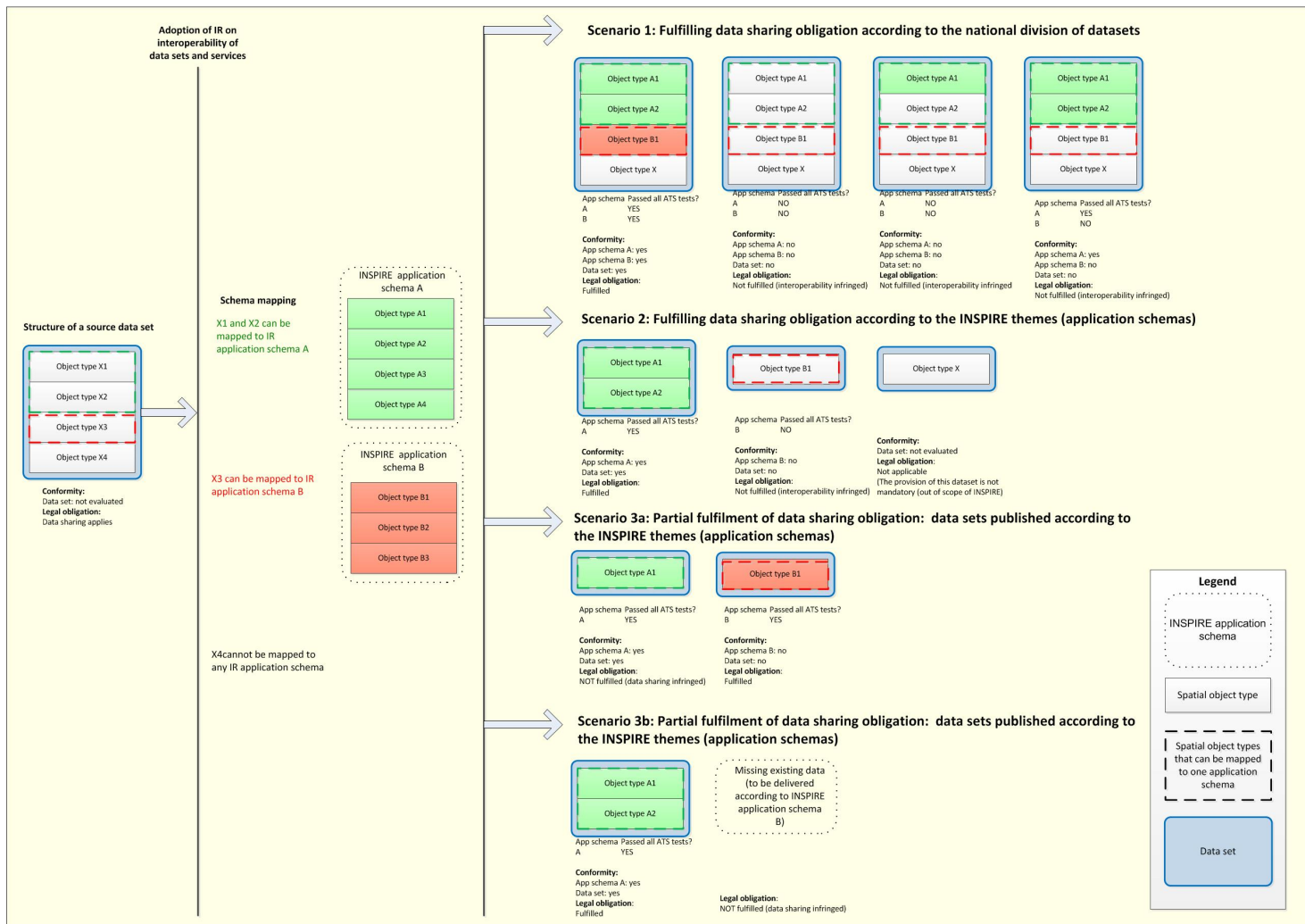


Figure 4: Conformity with data sharing and interoperability obligation in INSPIRE

3.5. Data and metadata transformations

Under the INSPIRE Directive, spatial data must be transformed in order to fulfil the technical content of the Interoperability Regulation. Even though the Directive does not set any requirement concerning quality assurance in the transformation process, according to the recommendations of the INSPIRE DQ WG (2012), it is good practice to preserve data integrity, consistency, and accuracy within the limits allowed by the available technology.

Indeed, as DQ experts reported (Östman et al., 2012), transformations necessary for INSPIRE may improve some aspects of data quality. In the course of mapping old to new application schema, the original dataset is revised, and misclassifications, broken associations and other semantic errors can be discovered and eliminated. However, the coordinate transformations, for example, may add additional errors/uncertainty to positional accuracy, but appropriate methods and algorithms can keep these to a minimum (INSPIRE Data Quality Working Group 2010, 2011, 2012).

The question of MD transformation is a delicate issue. Based on the feedback from the Member States, it was recognised that data transformation, especially in the initial phase of INSPIRE, may consume the resources of data providers. It may prevent them from working on MD transformation and updating the original DQ results¹⁰. In order to address this situation it was agreed that the original MD can be used, provided that the lineage element sufficiently documents what has happened to the data during the transformation process (INSPIRE Data Quality Working Group 2010, 2011, 2012).

3.6. Conformance testing

Art. 5, paragraph 2(a) of the INSPIRE Directive requires that conformity with the Implementing Rules be published as MD. The Metadata Regulation includes “conformity” as a mandatory MD element, and defines its evaluation scope and results. However, the open questions that stemmed from the interlinking of technical and legal aspects somewhat delayed the development of conformance-testing. As consequence, the related details were elaborated in a stream of Annex II-III specification developments only, resulting in the Abstract Test Suite (ATS) (European Commission, 2012, INSPIRE Thematic Working Groups, 2013).

The root of the problem was that the interoperability specifications contain various technical statements. Some are introduced in the Interoperability Regulation (IR requirements), while others are present only in the interoperability specifications (technical guidelines - TG). Those that remained in the TG were further classified as “TG requirements” and “recommendations”. It emerged that this classification had not been applied consistently. Therefore, the first step was to agree on what the subject of conformance testing was (i.e. the IR or the TG). The second decision was how to link conformity with the technical provisions. Finally, it was necessary to agree on a clear notation of different types or requirements and the recommendations in the TG. This latter is important, since experts, most probably, will use the TG for implementation purposes.

¹⁰ In strict sense all DQ results should be updated after the transformations based on new inspections of quality.

The solution of the situation is offered by the one-to-many (1..*) multiplicity of the conformity metadata element. Since the Directive legally requires a conformity statement, one conformity metadata element at dataset level was linked to the fulfilment of IR requirements. The second dataset-level conformity element is linked to the TG requirement, i.e. a dataset conformant with TG fulfils both IR and TG requirements. The other conformity metadata elements, as explained earlier, may refer to any standard, regulation, or user requirement. A practical example is given in Annex B.

Following the principles given in the related ISO standards, conformity with INSPIRE can be reported for each specification element and the results can be aggregated to the level of the dataset. The ATS links each requirement to one or more tests. The tests are grouped in several units that are referred to as a conformance classes. The ATS in INSPIRE consists of two parts: the first for checking conformance with the IR requirements and the second for checking conformance with the TG requirements. Conformance classes for TG recommendations were not defined.

In INSPIRE, the following conformance classes are defined:

- Application Schema;
- Reference Systems;
- Data Consistency;
- Data Quality Conformance;
- Metadata IR Conformance;
- Information Accessibility;
- Data Delivery;
- Portrayal;
- Technical Guidelines.

Depending on the content, the number of tests to be carried out in each conformance class varies from theme to theme. As none of the interoperability specifications contain *a priori* requirements on DQ, the corresponding conformance class plays the role of placeholder, which may be filled in course of implementation and maintenance of INSPIRE. The most frequently used tests are listed in Annex C.

The Application schema conformance class deserves special attention, as it contains the only INSPIRE requirement on DQ. The tests included in this class check for compliance with logical consistency.

From an implementation perspective, it could be asked why “not conformant” and “not evaluated” results are allowed in INSPIRE, given that there is a legal obligation for compliance. These values are used at different stages along the timeline for the data sharing and the interoperability obligation, as explained in section 3.3.

There is no requirement in the INSPIRE Directive to integrate or decompose original datasets in order to publish data according to the structure presented in the application schemas of interoperability specifications. Following the principle of compliant profiles and extensions presented in EN ISO 19106:2006, datasets that contain more or less spatial objects types that defined in the target application schema can be declared as being conformant. Conformance testing can be implemented either according to the national subdivision of datasets, or according to INSPIRE themes (see Figure 4).

3.7. *Metadata production and maintenance*

As stated in section 2, MD is accumulated step by step in each phase of transforming existing data, according to the implementing rules of INSPIRE. Data providers should therefore put in place appropriate on-the-fly documentation. Experience shows that producing “missing” MD later is more expensive.

Aside from producing, MD must also be maintained. Corrections of errors, certification, changes in validity, etc. need to be traceable. It is therefore important to keep the lineage MD element up to date.

Documenting new inputs for usability (new user requirements reflecting new use cases) is important not only for the users, but also for data providers, as it may multiply business opportunities of the latter. For instance, it is important to indicate whether a dataset fulfils the requirements of an EU thematic policy¹¹.

3.8. *Further improvements*

Even though the question of data quality has been systematically analysed and a number of issues have been clarified in context of SDIs for the first time, not all aspects were investigated in depth due to the limited period of INSPIRE data specification development. It is expected that the practical implementation and the consequent maintenance of the Interoperability Regulation and specifications will answer questions that have been raised but have not been answered. We list here the most important ones.

INSPIRE does not give any indication about the acceptable tolerance of the DQ results. The ATS, as currently presented, suggests that zero errors can be accepted. This has not been done on purpose; it is rather a gap that can only be filled in the future when the necessary information on implementation is collected. Nevertheless, the current formulation is overambitious for spatial data that is subject to uncertainty of observations, measurements, and human judgments.

The basis for comparison of the DQ elements defined in Chapter 7 is unclear; should they be compared to the source dataset, or to an external reference that is acknowledged as being true? If it is the source dataset, the recommended DQ results characterise the quality of the transformations. If it is the latter, the DQ results assess quality in a more “objective” way, whereas the result is dependent on the original quality and the quality of transformations.

Because of the data theme approach, only a few provisions were given for cross-theme consistency. Annex B of the Generic Conceptual Model (INSPIRE Data Specification Drafting Team, 2013) addresses generic issues of matching spatial objects along the boundaries and consistency between different scales. It is clear that practical action should be taken by the data providers. Nevertheless, it is necessary to clarify how this aspect affects conformity and provide guidance as to when (between which data themes and scales) cross-theme consistency makes sense. Moreover, the establishment of an agreed way to assess and document this process would add value for the users. These are tasks for the near future, as the Directive explicitly requires such consistency.

¹¹ Water Framework Directive, Noise Directive, just to mention a few.

INSPIRE data specifications do not currently contain instructions concerning the population of data used for conformance testing. Based on general practice we can assume that data providers will perform this task based on samples and not on a full inspection of the datasets. For comparability of conformity results it is necessary to propose sample sizes and methods, depending on the specificities of the theme and good practices of stakeholder communities.

Last, but not least, the conformity of extended schemas has remained open. The INSPIRE interoperability specifications fully respect cross-theme semantic consistency. This means that **a spatial object type defined in an application schema cannot conflict with the any other spatial object type in the infrastructure**, even when the latter is placed in another application schema. The INSPIRE Feature Concept Dictionary (FCD) and the Consolidated Model Repository define strict frameworks for reinforcing this principle. In the long term, this type of conformity should be achieved, but the principle of proportionality should be considered. As a first step, a targeted test (comparing definitions of spatial object types delivered by the data providers with the FCD) needs to be introduced in the ATS; however the legal basis of this solution has yet to be explored.

4. Conclusions

Despite the fact that data quality is one of the most important factors for data providers and users and that the topic is largely discussed from the research, standardisation, and implementation points of view, it was difficult to agree on a common approach in INSPIRE. The challenge mainly arose within the context of applying quality principles to SDI development processes, across very heterogeneous groups of stakeholders.

In order to outline a viable approach, a number of fundamental questions had to be clarified. For the first time in SDI development, these issues have been systematically reviewed outlining ways forward and the need for further investigations. Problems related to DQ terminology, to the role of *a priori* and *a posteriori* DQ, to conformance testing and quality documentation have been resolved, proposing an approach that can also be reused in other SDIs.

The work carried out so far gives a firm starting point for INSPIRE implementation, which will further contribute to accumulating knowledge and improving the quality of data delivered through INSPIRE in an interoperable way. Providing consistent data of documented quality is expected to attract responsible users that want to be confident in the decisions they make (e.g. improving the state of the EU environment, mitigation of global environmental problems) based on spatial analysis. Last, but not least, reusing good quality spatial data will leverage its wider use, creating an open space for new innovative applications and services.

5. Acknowledgments

The authors appreciate the help of Gilles Troispoux, Carol Agius, and Antti Jakobsson in drafting the initial DQ discussion paper and the questionnaires.

We thank all the DQ Contact Points for taking care of the questionnaires, organising discussions in their countries, and providing feedback to the Commission. They are listed in Annex E. We also thank all the participants of the two open DQ workshops at the INSPIRE conferences in Krakow and Istanbul for their interest and contributions.

Anne Ruas initiated the discussions on data consistency in the Data Specification Drafting Team and drafted Annex B of the Generic Conceptual Model. Jordi Escriu (TWG-EL) and Pierre-Yves Curtinot (TWG-OI) were tireless in testing and refining the content of the ATS. Gyula Iván (TWG-CP) reviewed Annex B of this document.

Our colleagues in the JRC Data Specification Support Team (listed in Annex E) were always willing to provide feedback. Special thanks to Tomáš Řezník and Michael Lutz for their input in conformance testing.

Annex A: DQ requirements in INSPIRE and the corresponding formal DQ elements

Art.	Citation	Related DQ (sub-)element in ISO 19157	Scope of DQ evaluation
5(2)	Metadata shall include information on the quality and validity of spatial datasets	All relevant to the dataset	Application schema of the data theme Implementing Rule / Data specification of the data theme
7(3)	Member States shall ensure that [...] spatial datasets and the corresponding spatial data services are available in conformity with the implementing rules [...]	Logical consistency	Application schema of the data theme
7(4)	Implementing rules [...] shall cover the definition and classification of spatial objects relevant to spatial datasets 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_Domain consistency DQ_ThematicClassificationCorrectness DQ_TopologicalConsistency	Application schema of the data theme
8(1), (2)	In the case of spatial datasets 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 DQ_DomainConsistency	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 (multiple-representation)
	or between items of information which refer to the same object represented at different scales	DQ_ConceptualConsistency (DQ_PositionalAccuracy)	Application schemas of the related data themes (Object referencing)
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_ConceptualConsistency (DQ_PositionalAccuracy)	Agreement between the interested parties

Annex B

Conformity of Hungarian Cadastral Parcel data

Conformity before transforming data according to INSPIRE

1. Conformity with the INSPIRE Implementing Rule

Title: Commission Regulation (EU) No 102/2011 of 4 February 2011 amending Regulation (EU) No 1089/2010 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards interoperability of spatial datasets and services

Date:

– date Type: publication

– date: 2011

MD value: not Evaluated

2. Conformity with the related INSPIRE Technical Guidelines

Title: D2.8.I.6 INSPIRE Data Specification on Cadastral Parcels –Guidelines

Date:

– date Type: publication

– date: 2010-04-26

MD value: not Evaluated

3. Conformity with the national standard MSZ 7772-1

Title: MSZ 7772-1 Digitális térképek. 1. rész: A digitális alaptérkép fogalmi modellje

Date:

– date Type: publication

– date: 1996

MD value: conformant

4. Conformity with ISO 19152

Title: ISO 19152 Geographic information — Land Administration Domain Model (LADM)

Date:

– date Type: publication

– date: -2012-11-01

MD value: conformant

Conformity after transforming data according to INSPIRE

1. Conformity with the INSPIRE Implementing Rule

Title: Commission Regulation (EU) No 102/2011 of 4 February 2011 amending Regulation (EU) No 1089/2010 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards interoperability of spatial datasets and services

Date:

- date Type: publication
- date: 2011

MD value: conformant

2. Conformity with the related INSPIRE Technical Guideline

Title: D2.8.I.6 INSPIRE Data Specification on Cadastral Parcels –Guidelines

Date:

- date Type: publication
- date: 2010-04-26

MD value: conformant

3. Conformity with the national standard MSZ 7772-1

Title: MSZ 7772-1 Digitális térképek. 1. rész: A digitális alaptérkép fogalmi modellje

Date:

- date Type: publication
- date: 1996

MD value: not Conformant

4. Conformity with ISO 19152

Title: ISO 19152 Geographic information — Land Administration Domain Model (LADM)

Date:

- date Type: publication
- date: 2012-11-01

MD value: conformant

NOTE 1: For sake of clarity, the abovementioned examples use citation to refer the regulation/standard/specification that is used for evaluating conformity. Naturally, as recommended in the TG, URI can be also used for referencing the specification when it exists. Example:

<http://inspire.ec.europa.eu/conformance-class/tg/cp/>

NOTE 2: Publishing value for conformity metadata element is mandatory for the Implementing Rule (point 1). The other statements are the choice of the data provider.

NOTE 3: After transforming Hungarian cadastral parcel data according to INSPIRE specification, the dataset was not conformant with the national standard because of the difference of

the coordinate reference system. The data provider may decide either to omit this conformity statement or to detail conformity according to the conformance classes specified in the ATS in the TG.

NOTE 4: Data model of Hungarian Land Registry (integrated land registry including cadastral maps and land records) acts as a Country Profile in ISO 19152 Geographic Information – Land Administration Domain Model.

Annex C

Structure of ATS proposed for INSPIRE

- 1 Application Schema Conformance Class
 - 1.1 Schema element denomination test
 - 1.2 Value type test
 - 1.3 Value test
 - 1.4 Attributes/associations completeness test
 - 1.5 Abstract spatial object test
 - 1.6 Constraints test
 - 1.7 Geometry representation test
- 2 Reference Systems Conformance Class
 - 2.1 Datum test
 - 2.2 Coordinate reference system test
 - 2.3 Grid test
 - 2.4 View service coordinate reference system test
 - 2.5 Temporal reference system test
 - 2.6 Units of measurements test
- 3 Data Consistency Conformance Class
 - 3.1 Unique identifier persistency test
 - 3.2 Version consistency test
 - 3.3 Life cycle time sequence test
 - 3.4 Validity time sequence test
 - 3.5 Update frequency test
- 4 Data Quality Conformance Class
 - 4.1 Data quality target results test
- 5 Metadata IR Conformance Class
 - 5.1 Metadata for interoperability test
- 6 Information Accessibility Conformance Class
 - 6.1 Code list publication test
 - 6.2 CRS publication test
 - 6.3 CRS identification test
 - 6.4 Grid identification test
- 7 Data Delivery Conformance Class
 - 7.1 Encoding compliance test
- 8 Portrayal Conformance Class
 - 8.1 Layer designation test
- 9 Technical Guideline Conformance Class
 - 9.1 Multiplicity test
 - 9.1 CRS http URI test
 - 9.2 Metadata encoding schema validation test
 - 9.3 Metadata occurrence test
 - 9.4 Metadata consistency test
 - 9.5 Encoding schema validation test
 - 9.6 Coverage multipart representation test
 - 9.7 Coverage domain consistency test
 - 9.8 Style test

Annex D

Data quality and metadata questionnaires used in INSPIRE

Questions from the first survey (May-June 2010)

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.

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 value	result	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 dataset / dataset series level, as required by the INSPIRE Metadata Regulation
 - In addition to dataset (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.

Questions from the second survey (May-June 2012)

9. **Do you agree that data quality target results should be specified for the mandatory (i.e. logical consistency) DQ elements?**
- Yes, for conceptual consistency
 - Yes, for domain consistency (where appropriate)
 - Yes, for topological consistency (where appropriate)
 - No, after the transformation the full dataset must be inspected and all errors have to be fixed.
 - No, once the systematic errors in the transformation process are fixed the remaining random errors are unlikely to jeopardise interoperability.

If no, please go to question 5. If yes, please answer questions 2, 3, and 4.

If you have crossed the last option, please provide your related experience or references

10. For conceptual consistency please give examples of target results that are used for different datasets. Please, include any additional explanation that might be useful for better understanding the reasoning behind. In case you know formally specified and well established requirements (e.g. NATO STANAGs, LPIS, national specifications, etc) please include it in the comment column.

Name of the data theme	DQ measure	Target result	Explanation / Comments

(Extend table if required.)

11. For domain consistency please give examples of target results that are used for different datasets. Please, include any additional explanation that might be useful for better understanding the reasoning behind. In case you know formally specified and well established requirements (e.g. NATO STANAGs, LPIS, national specifications, etc) please include it in the comment column.

Name of the data theme	DQ measure	Target result	Explanation / Comments

(Extend table if required.)

12. For topologic consistency please give examples of target results that are used for different datasets. Please, include any additional explanation that might be useful for better understanding the reasoning behind. In case you know formally specified and well established requirements (e.g. NATO STANAGs, LPIS, national specifications, etc) please include it in the comment column.

Name of the theme	DQ measure	Parameter (if applicable)	Targeted result	Comments

(Extend table if required.)

13. What other theme specific DQ elements are used in your country and what are the corresponding DQ target results? Please fill a separate table for each theme!

Name of the data theme

DQ element/ sub-element	DQ measure	Parameter (if applicable)	Targeted result	Comments

--	--	--	--	--

(Insert new tables and extend table if required.)

14. Conformity shall be declared at dataset level. The source datasets that the Member States offer for download, however, may contain less or more spatial object and object types. **Do you think that the datasets that have more/less types than specified in the INSPIRE application schema can be declared as conformant?**

- Yes, those that have more spatial object types and /or object types are conformant, unless the extensions do not break any rule of the “core” application schema.
- Yes, those that have less spatial object types and or object types are conformant unless the types specified in INSPIRE are conformant.
- No, any deviation from the specified schema breaks the rule of conformity. Users need to receive exactly what they expect. They should not be bothered neither by data integration, nor by the bigger and unnecessary data volumes.

15. Please provide justification and/or reference that might be useful for resolving question 6!

16. Declaration of conformity with well specified user requirements is also possible. **Do you have experience how this MD element is related to DQ_Usability, which is the “degree of adherence of a dataset to a specific set of requirements”?**

Annex E Contributors

DQ contact points contributing to INSPIRE

Austria	Georg Topf
Belgium	Geraldine Nolf, Leen De Temmerman
Czech Republic	Tomas Cajthaml
Denmark	Lars Storgaard
Estonia	Peep Krusberg
Finland	Aaro Mikkola
France	Gilles Troispoux,
Germany	Daniela Hogrebe, Andre Caffier
Hungary	Tamás Palya
Malta	Carol Agius
Netherlands	Ine de Visser
Norway	Erling Onstein
Poland	Marcin Grudzien
Romania	Daniela Docan
Slovakia	Ján Jendrichovsky
Slovenia	Irena Ažman,
Spain	Celia Sevilla-Sánchez, Dolores Barrot-Feixat,
Sweden	Christina Wasström
United Kingdom	Dan Haigh

The colleges of the INSPIRE Data Specification Support Team of the JRC who supported the authors with discussions and feedback

Andrej Abramić
Julien Gaffuri
Angel López
Michael Lutz
Tomáš Řezník
Alessandro Sarretta
Chris Schubert
Martin Tuchyňa

Annex F

Bibliography and References

EN ISO 19105:2005 Geographic information - Conformance and testing

EN ISO 19106:2006 Geographic information - Profiles

EN ISO 19115:2005 Geographic information – Metadata

EN ISO 19131:2008 Geographic information - Data product specifications

European Commission (2007). Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). (E. Commission, Ed.) Official Journal of the European Union, 50(L 108), 1–14. Retrieved from <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:EN:PDF>

European Commission (2008). Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata. Official Journal of the European Union, L(326/12), 1–19. Retrieved from <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:326:0012:0030:EN:PDF>

European Commission (2011). Commission Regulation (EU) No 102/2011 of 4 February 2011 amending Regulation (EU) No 1089/2010 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards interoperability of spatial data sets and services. Official Journal of the European Union, L 31, 13–34. Retrieved from <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:031:0013:0034:EN:PDF>

European Commission (2012). Data quality in INSPIRE: from requirements to metadata (pp. 3–17). Ispra.

European Commission (2013a). EU Shared Environmental Information System Implementation Outlook (pp. 1–10). Brussels. Retrieved from http://ec.europa.eu/environment/seis/pdf/seis_implementation_en.pdf

European Commission (2013b). INSPIRE Data specification template (pp. 1–102). Ispra.

Goodchild, M.F. (2007). "Citizens as sensors: the world of volunteered geography". *GeoJournal* 69 (4): 211–221.

INSPIRE Data Quality Working Group. (2010). Minutes of the INSPIRE Data Quality and Metadata Workshop. Krakow, 22 June 2010. Manuscript. (pp. 1–47). Ispra.

INSPIRE Data Quality Working Group. (2011). Minutes of the INSPIRE Data Quality and Metadata Workshop. Brussels, 28 February 2011. Manuscript. (pp. 1–8). Ispra.

- INSPIRE Data Quality Working Group. (2012). Minutes of the INSPIRE Data Quality and Metadata Workshop. Istanbul, 24 June 2012. Manuscript. (pp. 1–7). Ispra.
- INSPIRE Data Specification Drafting Team. (2013). INSPIRE Generic Conceptual Model (pp. 1–157). Ispra. Retrieved from http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/D2.5_v3.4rc3.pdf
- INSPIRE Data Specifications Drafting Team. (2008). INSPIRE Methodology for the development of data specifications (pp. 1–123). Retrieved from http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/DataSpecifications/D2.6_v3.0.pdf
- INSPIRE Metadata Drafting Team. (2009). INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119 (pp. 1–74). Retrieved from http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/metadata/MD_IR_and_ISO_20090218.pdf
- INSPIRE Thematic Working Groups. (2010). INSPIRE Data Specifications – Technical Guidelines (Annex I). Retrieved from <http://inspire.jrc.ec.europa.eu/index.cfm/pageid/2>
- INSPIRE Thematic Working Groups. (2013). INSPIRE Data Specifications - Draft Technical Guidelines (Annex II&III). Retrieved from <http://inspire.jrc.ec.europa.eu/index.cfm/pageid/2>
- ISO TC 211. (2013). ISO/FDIS 19157 Geographic information - Data quality
- Östman, A., Hedefalk, F., & Eriksson, H. (2012). Effects on the geodata quality when adapting to INSPIRE. INSPIRE conference 2012 “Sharing environmental information, sharing innovation”. Istanbul. Retrieved from http://inspire.jrc.ec.europa.eu/events/conferences/inspire_2012/presentations/62.pdf
- Ramage, S. (2010). INSPIRE, Data Quality and SDIs. Directions Magazine, January 18th 2010. Retrieved from <http://www.directionsmag.com/articles/inspire-data-quality-and-sdis/122424>
- Tóth, K., Portele, C., Illert, A., Lutz, M., & Nunes de Lima, V. (2012). A Conceptual Model for Developing Interoperability Specifications in Spatial Data Infrastructures (pp. 1–56). Luxembourg. doi:10.2788/21003
- Tóth, K., Tomas, R. (2011). Quality in Geographic Information – Simple Concept with Complex Details. Proceedings of the 25th International Cartographic Conference (pp. 1-11). Paris: Edited by Anne Ruas on behalf of the French Committee of Cartography. ISBN: 978-1-907075-05-6.
- World Wide Web Consortium (W3C). (2005). Quality Assurance glossary. Retrieved February 14, 2013, from <http://www.w3.org/QA/glossary>

European Commission

EUR 26097 – Joint Research Centre – Institute for Environment and Sustainability

Title: Data Quality in INSPIRE: Balancing Legal Obligations with Technical Aspects

Authors: Katalin Tóth, Robert Tomas, Vanda Nunes de Lima, Vlado Cetl

Luxembourg: Publications Office of the European Union

2013 – 44 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1018-5593 (print), ISSN 1831-9424 (online)

ISBN 978-92-79-32532-8 (pdf)

ISBN 978-92-79-32533-5 (print)

doi: 10.2788/9648 (pdf)

Abstract

Nowadays the wide spread of electronic spatial data, the internet, and computer literacy allow to data from diverse sources to be easily shared and combine. However, various communities approach Data Quality (DQ) very differently. Some of them apply standardised methodology for producing data and publishing metadata (MD) about quality and usability, while others do not share any such information.

This report integrates the results of the work of the INSPIRE Data Specification Drafting Team, the JRC Data Specification Support Team, the INSPIRE Thematic Working Groups and the INSPIRE Data Quality Expert Group. This latter was composed of specialists nominated by the Member States as contact points for DQ. They were charged with organising discussions at national level with the involvement of the competent national authorities, and with delivering the results for INSPIRE. The consolidated results (disseminated as recommendations) were considered in the technical guidelines of the INSPIRE data themes.

In order to avoid the difficulties in understanding stemming from the multidisciplinary environment of the stakeholders, this report explains the similarities and differences between the notions of DQ in classical data production and in Spatial Data Infrastructures (SDI). In addition, because of the specific setup of INSPIRE, whereby Member States of the EU are obliged to fulfill requirements presented in legal and technical terms, the meaning of conformity was explained.

In conclusion, the report summarises the technical aspects that must be balanced between legal requirements, stakeholders' expectations, and preparedness of data providers to share good quality data in interoperable way. It also underlines the advantages of quality conscious SDIs and gives insight in potential further work.

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new standards, methods and tools, and sharing and transferring its know-how to the Member States and international community.

Key policy areas include: environment and climate change; energy and transport; agriculture and food security; health and consumer protection; information society and digital agenda; safety and security including nuclear; all supported through a cross-cutting and multi-disciplinary approach.