Application of the object oriented quality management model to secondary data sources

Piet J.H. Daas and Peter W.M. van Nederpelt

Discussion paper (10012)

Statistics Netherlands

The Hague/Heerlen, 2010

Explanation of symbols

.	= data not available
*	= provisional figure
**	= revised provisional figure
x	= publication prohibited (confidential figure)
–	= nil or less than half of unit concerned
–	= (between two figures) inclusive
0 (0,0)	= less than half of unit concerned
blank	= not applicable
2008–2009	= 2008 to 2009 inclusive
2008/2009	= average of 2008 up to and including 2009
2008/'09	= crop year, financial year, school year etc. beginning in 2008 and ending in 2009
2006/'07–2008/'09	= crop year, financial year, etc. 2006/'07 to 2008/'09 inclusive

Due to rounding, some totals may not correspond with the sum of the separate figures.

Publisher
Statistics Netherlands
Henri Faasdreef 312
2492 JP The Hague

Prepress
Statistics Netherlands - Grafimedia

Cover
TelDesign, Rotterdam

Information
Telephone +31 88 570 70 70
Telefax +31 70 337 59 94
Via contact form: www.cbs.nl/information

Where to order
E-mail: verkoop@cbs.nl
Telefax +31 45 570 62 68

Internet
www.cbs.nl

ISSN: 1572-0314

6008310012 X-10

Application of the object oriented quality management model to secondary data sources

Piet J.H. Daas and Peter W.M. van Nederpelt

Summary: This paper describes the application of the Object Oriented Quality Management model to the object secondary data sources. The results obtained are compared to those of the, independently developed, Quality framework for Administrative Data Sources. An administrative data source is an example of a secondary data source. This exercise was performed to enable the evaluation of the strengths and weaknesses of the quality management model and the completeness of the quality framework.

Keywords: quality assurance, registers, quality management, secondary data

Contents

1. Introduction

During 2008 and 2009 a new quality management model was developed at Statistics Netherlands (SN). This model is called the Object Oriented Quality Management (OQM) model and is specifically suited to fully meet the statistical requirements of SN (Van Nederpelt, 2009a-d). Although the OQM-model was developed for and by SN it can be applied to all areas of quality assurance and all types of organizations (Van Nederpelt, 2009a-d).

In the same period and independent of the OQM-model, a quality framework for administrative registers was developed at SN (Daas et al., 2008b). The framework is, however, not limited to administrative data sources. It can be applied for the evaluation of the statistical quality of other secondary data sources, such as secondary survey data (Daas et al., 2008a), administrative registers (Daas et al., 2009a), and internet data (Ossen et al., 2010). In the remainder of this paper, the quality framework developed for administrative data sources will be abbreviated as the QADS-framework. This framework was developed by collecting all information on the study of the quality of administrative data sources published in the literature and the information available at SN (Daas and Arends-Tóth, 2007).

The approach followed for the development of the QADS-framework differed considerably from that proposed by the OQM-model. In the OQM-model quality aspects are derived for an object by an expert with the help of his/her domain knowledge aided by a list of characteristics. Subsequently, for the combined set of objects and characteristics selected by the expert, among other things, quality indicators and metrics are derived (Van Nederpelt, 2009c). The fundamental differences in the approaches followed and their independent development prompted the authors to compare the results of both approaches. Questions underlying the need for this study were: Can the QADS-framework be integrated in the OQM-model? And is it possible to obtain the quality aspects in the QADS-framework by following the OQM-approach? In an attempt to answer these questions, the OQM-model was applied to the object ‘secondary data sources’ and the results obtained were compared with those of the QADS-framework and visa-versa. The author’s interest particularly focussed on the strengths and weaknesses of the OQM-model and on the completeness of the QADS-framework.

1.1 For the reader

This report starts by explaining the OQM-model and the QADS-framework in chapter 2. In the next chapter the OQM-model is applied to the object ‘secondary data sources’. Here, special attention is paid to the selection of relevant characteristics and the object and sub-objects under study. In chapter 4 the results of both approaches are compared, discussed, and conclusions are drawn.

2. Quality management and the quality framework

2.1 The Object Oriented Quality Management model

The OQM-model was developed during the search for a SN quality management model that would comply to the Code of Practice and the Quality declaration of the European Statistical System (Van Nederpelt, 2009a). Since existing models did not fully meet the requirements of SN, it was decided to develop a new model (Van Nederpelt, 2009a,b). The OQM-model includes components of other well-known quality management models and can easily be combined with the European Foundation for Quality Management Excellence model (Van Nederpelt, 2009c,d). The most important focus of the OQM-model is that of the management of objects relevant for the organisation at hand.

The purpose of the OQM-model is to obtain a set of measures for one or more areas that require quality control by the organisation in a systematic way. The user selects those areas. The OQM-model is 'object oriented' which means that an organisation and its environment is seen as a collection of related objects (Van Nederpelt, 2009a). Examples of objects are customers, products, and processes. In fact, each noun where the words "the quality of …." can be added successfully can be seen as an object to which the OQM-model can be applied (Van Nederpelt, 2009c). Another important assumption of the OQM-model is that all objects have specific 'characteristics'. Examples of these are availability, completeness, and integrity. More examples of characteristics can be found in Annex A of this paper and Annex 5 of the paper by Van Nederpelt (2009c). Many of the names of the characteristics in those lists resemble those of quality dimensions and indicators identified for statistical data sources (Eurostat 2003, 2005; Daas et al., 2009a). This suggests that a comparison at the level of the characteristics for the object 'secondary data sources' (as identified by the OQM-approach) and those included in the QADS-framework is possible. This was the starting point for the study described in this paper.

In the OQM-model, the combination of an object and a single characteristic form a so-called quality area. These are the areas an organisation wants to control. Examples of quality areas are 'Completeness of the dataset' and 'Authenticity of units' (more on this topic below). Quality areas have the advantage that they compactly describe and identify an area at a level that both management and quality experts understand. It is also at this level where indicators and measures are defined to control a quality area and where links are made between the areas of different objects (Van Nederpelt, 2009c). Quality areas are therefore excellently suited to check the completeness and categorisation of quality indicators.

To control a quality area regular evaluation need to be performed. The OQM-model includes 12 steps to do this. Since not all of them are relevant for the topic studied and discussed in this paper, the reader is referred to the paper of Van Nederpelt (2009c) for more details. The most important thing to realize is that the sequence

and depth of elaboration of each step is not strictly defined by the OQM-model. This is controlled by the user and affected by the importance and complexity of the quality area under consideration. One of the twelve steps is the development of quality indicators to measure the status of each quality area. Another very important step is the construction of measures to control and manage the quality area. The OQM-model has been applied at SN to three cases so far (Van Nederpelt, 2009b).

2.2 The Quality framework for Administrative Data Sources

SN has developed a framework for the determination of the quality of administrative data sources because it is increasingly using those and other secondary data sources for the production of statistics (Daas and Arends-Tóth, 2007). As a result of this, SN is becoming more and more dependent on the quality of the data sources collected and maintained by others. It is therefore of vital importance that SN is able to determine the quality of those types of data sources in a systematic, objective, and standardized way (Daas et al., 2008b).

The QADS-framework is based on the results of a literature study in which all quality aspects, relevant for statistics, were identified for administrative data sources. In this study the quality aspect information available at SN (Daas and Fonville, 2007) and those mentioned in publications by others were combined. All quality aspects identified were combined into a single framework (Daas et al., 2008a-b). This approach aimed for a complete overview of the quality aspects of administrative (and other secondary) data sources relevant when used for statistics. This quite laborious exercise resulted in a framework composed of three high level views on quality, which were called the Source, Metadata, and Data hyperdimension (Daas et al., 2008a). These hyperdimensions contain 5, 4, and 10 quality dimensions, respectively. Each of these dimensions is composed of one or more quality indicators which are measured or estimated by one or more measurement methods.

The focus of the Source hyperdimension is the aspects of quality essential for the safe and secure use and the delivery of the data source. The Metadata hyperdimension focuses on the metadata aspects of the data source. In the Data hyperdimension the technical and accuracy related aspects of the data in the source are evaluated (Daas et al., 2009a; Kuijvenhoven and Schouten, 2008). To assist the evaluation of the quality indicators in the QADS-framework, a checklist was developed for the aspects included in the Source and Metadata part of the framework. Because of the large number of quantitative indicators in the Data hyperdimension, such an approach could not be applied for the Data part (Daas et al., 2009a-b). The QADS-framework has been successfully applied to eight different administrative registers so far (Daas and Ossen, 2010) and has also been used to evaluate price information on the internet and offline routing information (Ossen et al., 2010).

3. Application and comparison of the OQM-model

This chapter starts with a description of the application of the OQM-model to the object 'secondary data source'. During this exercise, it became apparent that quite some additional decisions had to be made. The most important were i) dividing quality in a metadata and a data domain, ii) identification of sub-objects for the object under study, and iii) selection of characteristics appropriate for the sub-objects. These intermediate results provided valuable information on the study of the quality of secondary data sources in general.

3.1 OQM-model approach

3.1.1 Object and sub-object selection

The OQM-model was applied to the object for which the QADS-framework was constructed for two reasons. These are: i) to test the strengths and weaknesses of the OQM-model and ii) to check the completeness of the QADS-framework.

The object chosen for in the OQM-approach was '*secondary data source*' and not 'administrative data source'. This was done because, in contrast to what its name suggests, the QADS-framework can and has been applied to much more data sources than administrative data sources alone. Examples of such non-administrative data sources are survey data collected by others (Daas et al., 2008a), registers (Daas et al., 2009a), offline routing information, and internet data (Ossen et al., 2010). These are all secondary types of data sources. The reader is referred to Daas and Beukenhorst (2008) for more information on the various types of data sources that SN distinguishes.

During the evaluation of the object 'secondary data source' it became apparent that there are in fact two domains of quality to which the OQM-model needed to be applied. These are the *data* and the *metadata* quality domain of a secondary data source. This observation corroborates the well established notion that there are (always) two different sides to quality, the quality of the *data* and the quality of the *metadata*. The latter is referred to as schema quality by Batini and Scannapieco (2006).

First evaluation results also revealed that in the object 'secondary data source' three sub-objects can and need to be distinguished. Each of these sub-objects is essential for obtaining a complete overview of the quality areas for the object under study. The sub-objects discerned from the start were: i) the dataset as a whole, ii) the units in the data source, and iii) the items (variables) in the data source. The three sub-objects will be referred to as dataset, units, and items in the remainder of this paper.

Combined with the data and metadata quality domains, a total of six sub-objects were discerned for which the quality areas would have to be evaluated. The sub-objects studied are shown in table 1.

Table 1. Overview of the sub-objects and objects evaluated in the OQM-model

Object:	Data of secondary data source	Metadata of secondary data source
Sub-objects:	Data of dataset in secondary data source	Metadata of dataset in secondary data source
	Data of units in secondary data source	Metadata of units in secondary data source
	Data of items in secondary data source	Metadata of items in secondary data source

3.1.2 Selection of relevant characteristics

Another important task was the identification of the characteristics applicable to the (sub)object under study. For an efficient evaluation the number of characteristics should be as limited as possible without any essential characteristics missing. The construction of such a list is something that would usually be performed by consulting an expert or experts in the field under study. In this case, another approach was followed deliberately. To show the strength of the OQM-model, the identification process started with a (very) long list of all conceivable characteristics. This approach had the additional advantage that it had the greatest change of demonstrating the possible absence of any quality aspects -or parts of it- in the QADS-framework. However, to prevent the evaluation of an extremely large amount of characteristics for all sub-objects discerned, additional intermediate characteristic reduction steps were included. In these steps the long list of characteristics was condensed by removing i) all characteristics irrelevant for the sub-object(s) under study and ii) all very similar or duplicate characteristics. This approach was chosen because it was, at that time, considered the most efficient, thorough, and unbiased one.

The construction of the long list of characteristics started with the 162 characteristics included in Annex 5 of the paper by Van Nederpelt (2009c). To this list the quality 'dimensions' mentioned in the Annex to the document of Daas (2009), those listed for epidemiological secondary data sources in the paper of Søresen et al. (1996), and those mentioned for internet data in the papers of Wang and Strong (1996), Pipino, Lee and Wang (2002), and Knight and Burn (2005) were added. This resulted in a total of 323 characteristics of which 221 remained after removal of duplicate names and not appropriate terms; such as 'unique keys'. The latter is not a characteristic but an object. The list of 221 characteristics remaining is included as Annex A in this paper.

The next step was the removal of any obvious non-applicable characteristics for the (sub-)objects studied (such as colour). This was followed by the removal of synonyms and antonyms. Eventuality a condensed list of 25 quality characteristics remained. These are shown in table 2. The 25 characteristics were considered the absolute minimum. They were used for the construction of quality areas. This was done by combining each of the six sub-objects in table 1 with the 25 characteristics of table 2. A total of 150 quality areas were formed that needed to be evaluated. Annex B and C give a complete overview of the quality areas produced for the sub-objects in the data and metadata domain.

Table 2. List of characteristics considered relevant for the (sub-)objects studied

Accuracy	Coverage	Replaceability
Authenticity	Dependency	Selectivity
Availability	Detailledness	Size
Clarity	Format	Stability
Coherence	Importance	Timeliness
Comparability	Linkability	Uniqueness
Completeness	Punctuality	Usefulness
Confidentiality	Relevance	
Correctness	Reliability	

3.1.3 Quality area evaluation and selection

For every quality area, e.g. the combination of a sub-object and characteristic, the authors looked at its meaning and relevance. Every quality area was described in the form of a question and rephrased to force the authors to make clear what the actual focus of the quality area was. For example, the quality area 'completeness of the data for the dataset' was written as the question: "How complete is the data in the dataset?" The question was rephrased as: "Is all data delivered and is all data accessible in the dataset?" The quality area descriptions were used to determine the relevance of the areas. For some quality areas it was immediately clear that they were relevant, for example the area 'correctness of the data for the items'. This area deals with the accuracy of the values for the variables in the data sources. This is relevant. Other quality areas were clearly not-relevant, such as 'clarity of the data for the units'. This 'area' has no real meaning at the data level. For quite some areas, however, it wasn't easy to determine their actual focus and relevance. It was therefore decided that only the undoubtedly meaningless areas were considered not relevant. The somewhat vaguely defined areas were not immediately discarded. Of the total of 150 quality areas discerned, 23 were found not to be relevant; 18 for data and 5 for metadata. The reader is referred to Annex B and C for more details.

By carefully reviewing the questions and the rephrased wordings it was found that quite some quality areas actually focussed on the same or a very similar aspect of quality. Examples of clearly identical quality areas are 'authenticity of the data for the items' and 'correctness of the data for the items'. These areas both point to the relation between the value of an item (variable) in the data source and the 'real' world value or state for that item. For example, if a person is registered as married is it indeed so that he or she is actually married? For both quality area examples correctness was selected as the characteristic of preference. Another example of identical areas is 'completeness of the data for the units' and 'coverage of the data for the units'. These areas both point to the completeness of the population of the units in the data source. Similarity was, however, not apparent for all quality areas. Some of the quality areas were found fairly similar but not completely identical. An example of this is 'relevance of the data in the dataset' and 'availability of the data in the dataset'. Both areas point to importance of the dataset for the NSI but the

latter obviously focuses more on the physical presence of the actual dataset then the former. To express these differences, for these types of quality areas, a first and a second characteristic of choice was identified. For the reader and as a future reference, the total result of the exercise described above can be found -for both data and metadata- in Annex B and C, respectively.

After all quality areas were carefully reviewed, the final choice of the preferred characteristic for the areas with two characteristics of choice was made. The selection between the first and second characteristic listed was made with the sole aim to *minimize* the total number of unique characteristics in the data domain *or* in the metadata domain. Let it be perfectly clear that the effect on the combination of both domains was ignored. This last reduction step eventually resulted in a total of 10 characteristics relevant for the data domain and 7 for the metadata domain of quality (table 3). A combined total of 12 unique characteristics were discerned for a secondary data source. How the 25 starting characteristics are linked to the remaining 10 of the data domain and 7 of the metadata domain is shown in figure 1 and 2, respectively.

The reason for the lower number of unique characteristics in the metadata domain is mainly the result of the fact that -at the metadata level- many quality areas ultimately refer to the clarity, comparability, completeness, or correctness of the metadata for the sub-object under study. Detailedness at the metadata level, for example, eventually refers to the completeness of the metadata (see Annex C).

The authors realize that -in some cases- the choices made during the reduction of the number of characteristics could be considered as somewhat subjective by some of the readers. The reader should, however, be aware that the whole procedure followed has been carefully thought-out, repeated, and reviewed several times to reduce the chance of any not well thought of decisions affecting this meticulous and laborious process. For transparency purposes and as a future reference, the starting points (Annex B and C) and intermediary results of the reduction process are all included in this paper (tables 3-5 and figures 1 and 2). The characteristics in table 3

Table 3.Characteristics remaining for the sub-objects identified in the data and metadata domains

Characteristics for data	Characteristics for metadata
Authenticity	
	Clarity
Coherence	
	Comparability
Completeness	Completeness
Confidentiality	Confidentiality
Correctness	Correctness
Detailedness	
Selectivity	
Stability	Stability
Timeliness	Timeliness
Uniqueness	

Figure 1. Overview of the relation between the starting and final characteristics in the data domain

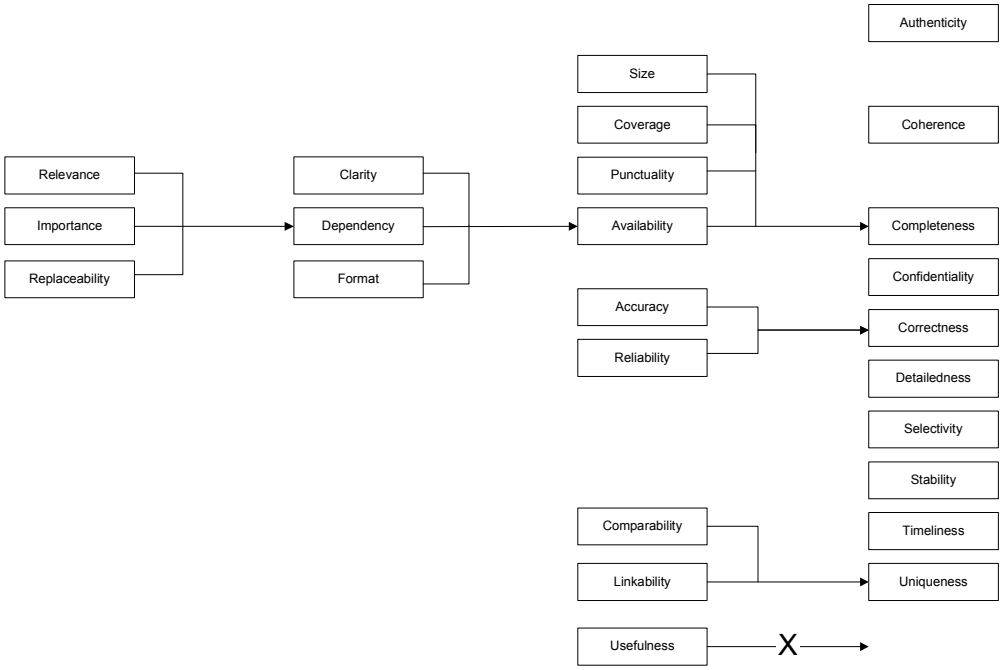
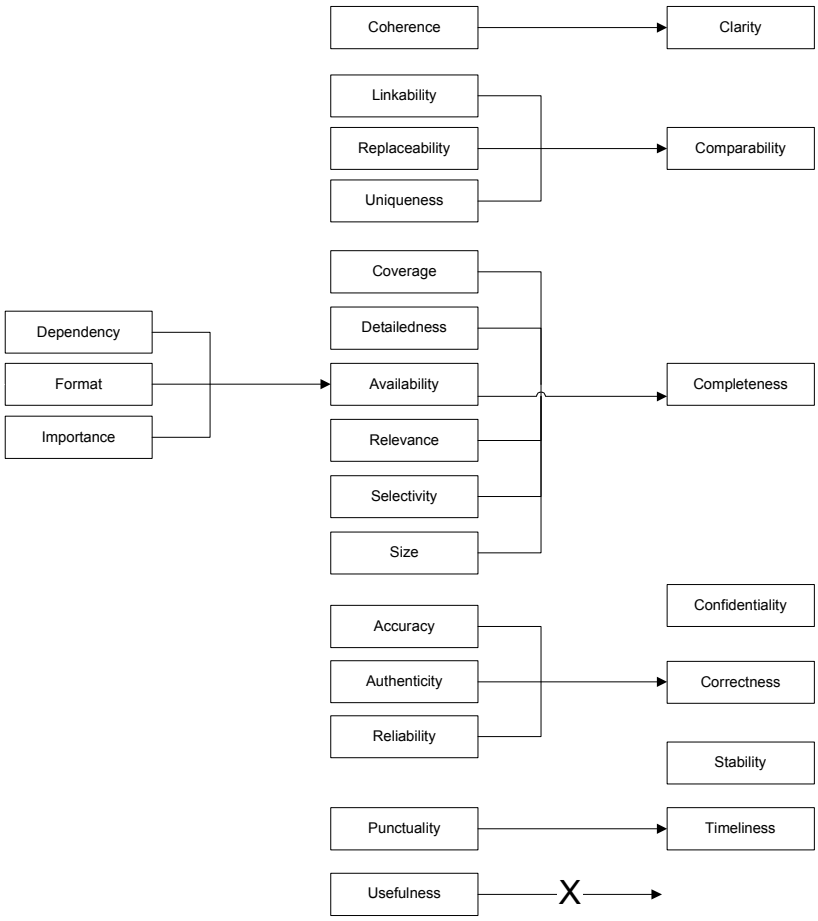


Figure 2. Overview of the relation between the starting and final characteristics in the metadata domain



are considered the absolute minimum set of characteristics that need to be distinguished for a secondary data source. It would be interesting to see if these characteristics apply to data sources in general.

As a result of the evaluation and characteristics condensation exercise described above the initial number of 150 quality areas started with (see Annex B and C) drastically reduced. A total of 39 unique quality areas remained; 20 for data and 19 for metadata. The final set of quality areas is shown in table 4 for both domains. With the identification of the minimal number of relevant quality areas for the sub-objects of secondary data sources the OQM-approach was stopped. These quality areas were all that was needed for the remainder of the work described in this paper. However, if the OQM-approach would have been continued, some of the things that would have to be done were the creation of quality indicators and the identification of causes and effects of problems for every quality area listed in table 4. This is still quite some work.

Table 4. Final quality areas identified for the data and metadata domains of the six sub-objects studied

Characteristic	Domain	Sub-object	Characteristic	Domain	Sub-object
Authenticity	data	units			
			Clarity	metadata	dataset
			Clarity	metadata	units
			Clarity	metadata	items
Coherence	data	items			
			Comparability	metadata	dataset
			Comparability	metadata	units
			Comparability	metadata	items
Completeness	data	dataset	Completeness	metadata	dataset
Completeness	data	units	Completeness	metadata	units
Completeness	data	items	Completeness	metadata	items
			Confidentiality	metadata	dataset
Confidentiality	data	units			
Confidentiality	data	items			
			Correctness	metadata	dataset
Correctness	data	units	Correctness	metadata	units
Correctness	data	items	Correctness	metadata	items
Detailedness	data	items			
Selectivity	data	units			
Selectivity	data	items			
Stability	data	dataset	Stability	metadata	dataset
Stability	data	units	Stability	metadata	units
Stability	data	items	Stability	metadata	items
Timeliness	data	dataset	Timeliness	metadata	dataset
Timeliness	data	units	Timeliness	metadata	units
Timeliness	data	items	Timeliness	metadata	items
Uniqueness	data	units			
Uniqueness	data	items			

3.2 Comparison of the OQM-results and the QADS-framework

The ultimate goal of the exercise described above is of course the comparison between the quality areas obtained by the OQM-approach and the quality aspects included in the QADS-framework. For a thorough comparison, first the quality areas

identified by the OQM-approach (in table 4) were compared with the dimensions and indicators in the QADS-framework (tables 1 and 2 in Daas et al., 2009b). Next, this comparison is reversed; i.e. the quality aspects included in the QADS-framework were compared to those identified by the OQM-results. This was done to ensure that the strengths and weaknesses of both approaches are revealed.

3.2.1 Comparison of OQM with QADS

Comparison of the quality areas identified in the OQM-approach with the quality indicators and dimensions in the QADS-framework demonstrated that the OQM-approach did not identify any complete new quality aspects; e.g. aspects absent in the QADS-framework. However, it did reveal that some of the OQM-areas were not completely covered by the QADS-framework. These areas are marked with a plus-minus sign (-/+) in table 5. For the data quality domain, the areas not fully covered are the stability and timeliness of units and items and the uniqueness of items. For the metadata domain, these areas are the correctness of units and items.

The incomplete coverage of some aspects in the QADS-framework can be solved easily by either adjusting or extending the measurement method(s) for the relevant quality indicators. In practice this would mean that the questions for the appropriate indicators in the checklist (Daas et al., 2009b) either have to be adjusted slightly or (when no other alternative is possible) an additional question might be added. Since the research on the determination of the indicators in the Data hyperdimension has not been completed yet (Daas and Ossen, 2010), adjustment for the latter is no problem. Let it be clear that in both cases these are all relatively small changes. The fact that it has become apparent that some quality areas are not fully covered in QADS clearly demonstrates the added value of the OQM-method.

Another intriguing discovery is the fact that not all of the OQM quality areas in the *data* domain relate to quality indicators in the *Data* hyperdimension of the QADS-framework (table 5). This contradicts what one would expect to find (see below). Characteristics of the areas for which this difference is observed are confidentiality (units and items), stability (all), timeliness (all), and uniqueness (items). Apparently, these characteristics were considered more metadata related issues (i.e. Source and Metadata hyperdimension) during the construction of the QADS-framework (Daas and Arends-Tóth, 2007). This finding certainly has to be taken into account in the current studies of the Data hyperdimension. For the OQM-areas in the metadata domain such a difference is much less apparent. The only quality areas in this domain not belonging to either the Source or the Metadata hyperdimension are the correctness of units and items. These areas are also not covered completely by the OQM-results for the metadata domain.

3.2.2 Comparison of QADS with OQM

When the quality dimensions and indicators in the QADS-framework are compared with the quality areas identified for the OQM-approach, it can be concluded that the majority of the dimensions and quality indicators in the framework are covered by

Table 5, OQM-results in relation to the QADS-framework

Data quality area Characteristic	Sub-object	In QADS framework	Identification nr. in QADS framework	In Data hyper-dimension
Authenticity	units	+	D7.1, via M4.1	+
Coherence	items	+	D7.1, 7.2, 8.1, 8.2, 8.3	+
Completeness	dataset	+	D1.1	+
Completeness	units	+	D2.1, 3.1, 5.1, via 5.3	+
Completeness	items	+	D6.1, D10.1, via D3.3, 4.4, 6.3	+
Confidentiality	units	+	S3.2, via S3.3	-
Confidentiality	items	+	S3.2, via S3.3	-
Correctness	units	+	D3.2, 4.1, 4.2, 4.3, via M4.1, D5.3	+
Correctness	items	+	D5.1, 6.1, 6.2, 7.1, 7.3 via M4.1, 4.2, via D3.3, 4.4, 6.3	+
Detailedness	items	+	D9.1, 10.3, via S2.3	+
Selectivity	units	+	D3.2, 4.3, 5.2	+
Selectivity	items	+	D6.2, 10.2	+
Stability	dataset	+	S5.2, 5.4, M1.5	-
Stability	units	-/+	via M1.5, S5.2	-
Stability	items	-/+	via M1.5, S5.2	-
Timeliness	dataset	+	S4.2, 4.3, 5.4	-
Timeliness	units	-/+	via M1.5, via S4.2, 4.3, 5.4	-
Timeliness	items	-/+	via M1.5, via S4.2, 4.3, 5.4	-
Uniqueness	units	+	D4.1, 4.2, 4.3, via M3.1	+
Uniqueness	items	-/+	via M3.2	-

Metadata quality area Characteristic	Sub-object	In QADS framework	Identification nr. in QADS framework	In Source hyper-dimension	In Metadata hyperdimension
Clarity	dataset	+	S4	+	-
Clarity	units	+	M1.1	-	+
Clarity	items	+	M1.2, 1.3, 1.4	-	+
Comparability	dataset	+	M2	-	+
Comparability	units	+	M2.1	-	+
Comparability	items	+	M2.2, 2.3	-	+
Completeness	dataset	+	S3.1, 5.4, via S4.2, 4.5	+	-
Completeness	units	+	M1.1	-	+
Completeness	items	+	M1.2, 1.3	-	+
Confidentiality	dataset	+	S3.2, 3.3	+	-
Correctness	dataset	+	S1.1, 4.2, D1.1, 1.2	+	-
Correctness	units	-/+	via D1.2	-	-
Correctness	items	-/+	via D1.2	-	-
Stability	dataset	+	M1.5, S5.2, 5.4	+	+
Stability	units	+	M1.5, S5.2	+	+
Stability	items	+	M1.5, S5.2	+	+
Timeliness	dataset	+	S5.2, M1.5	+	+
Timeliness	units	+	S5.2, M1.5	+	+
Timeliness	items	+	S5.2, M1.5	+	+

the OQM-approach (results not shown). However, a few quality indicators are not included. These indicators are: Purpose (S1.2), Usefulness (S2.1), Envisaged use (S2.2), Response burden (S2.4), Data collection (S5.1), and Feedback (S5.3). These numbers are missing from table 5. Interestingly, these indicators are all located in the Source hyperdimension. Particularly the indicators in the 'Relevance' dimension (S2) are covered very poorly.

The absent quality indicators also reveal something else. There appear to be two groups of indicators, of which the first focuses on requirements for the NSI (S1.2, S2.1, 2.2, S5.3). The second group looks at process related aspects of the data source keeper (S2.4, S5.1). One may wonder why those quality aspects were not (or very poorly) included in the OQM-results? After all, these are all very important aspects of quality. The absence of these indicators is very likely the consequence of a limitation in the selection of sub-objects made at the start of the OQM-approach. In the beginning of the OQM-approach (only) the sub-objects dataset, units, and items were distinguished for the data and metadata domain (see section 3.1.1). Apparently one or more other important sub-objects should have been included. When, at that point in time, the sub-objects referring to the actors involved in the collection and exchange of the data source (i.e. the data source keeper and the NSI) would have been thought of and included, it is very likely that the indicators now considered absent would have been detected by the OQM-approach. This observation clearly demonstrates the importance of a thorough review of the object en sub-objects selected in the OQM-model.

What the comparison between the results of the OQM-approach and the QADS-framework in addition reveals is the fact that in the QADS-framework the data and metadata aspects of quality are not distinguished in a strict formal way. The aspects included in the Source, Metadata, and Data hyperdimension are ordered in a way that they assist and guide the user. The latter is in accordance with the most important point-of-departure for the creators of the QADS-framework (Daas and Arends-Tóth, 2007, Daas et al., 2008a-b).

4. Discussion and conclusions

The top-down approach used in the OQM-model resulted in a complete coverage of all indicators belonging to the quality areas for the (sub-)object(s) identified at the start. Indicators that were part of other -not identified- sub-objects or quality areas are missed by default. The bottom-up approach used in the construction of the QADS-framework resulted in a complete overview of all quality aspects for which indicators were developed. Quality aspects missed, because they belong to new or not (yet) studied areas, are absent (Daas and Ossen, 2010). The comparison made between both methods, reveals that a combined approach seems the most fruitful way to assure the coverage of all quality areas for a particular object.

To obtain maximum coverage of the quality aspects of an object the following approach should be followed. First, quality areas should be identified top-down by the OQM-approach. Subsequently, any known indicators for the object studied should be linked to the quality areas identified; as shown in table 5 of this study. Quality areas which contain few, hardly any, or no indicators at all need to be investigated in depth to assure complete coverage. Indicators that can not be linked to any quality area should be looked at carefully to assure that all sub-objects and all relevant characteristics were considered in the OQM-approach. If this was not found to be the case the OQM-approach should be repeated for the combination of the new sub-object and/or new characteristics (i.e. quality areas). This combined dual approach will assure that all relevant areas of quality are covered completely for the object under study. The ultimate goal is an appropriate level of control by the NSI.

Apart from the methodological differences, the OQM-model and QADS-framework also differ in the hierarchical structure. The QADS-framework is composed of hyperdimensions, dimensions and quality indicators, while the OQM-model differentiates between objects, characteristics, quality areas (the combination of an object and a characteristic) and indicators, to name a few (Van Nederpelt, 2009c,d). Each approach has its pros and its cons. The major advantage of the OQM-model is that it can be applied to all objects. The QADS-framework is limited to one object, which is a secondary data source. The OQM-model has this advantage because it is based on one generally applicable idea; the quality area concept. This concept has the advantage that, it:

- Defines and specifies the scope of the area to be managed
- Is not limited to the development and identification of quality indicators
- Can be integrated into a general quality management framework for NSI's
- Is able to integrate other quality frameworks (such as the QADS-framework).

The results of this study demonstrate that the OQM-model is indeed able to replace the QADS-framework up to a certain point. This point demonstrates the weakness of the OQM-model. When an important sub-object of the object under study is not included from the start, the results of the OQM-model will contain a so-called 'blind-spot'; a quality area for which no indicators are constructed. However, the bottom-up approach of the QADS-framework is also not ideal. It has a -quite similar- disadvantage. The bottom-up approach used for QADS assures the coverage of all familiar quality areas by including all previously identified quality indicators. But when a quality area exists that no one has ever identified and published about, indicators for that area will be lacking. This results in a 'blind spot' as well. The best way to prevent any missing areas is applying both the object-oriented and the bottom-up approach to the object under study. With this combined method one may expect to identify almost all of the quality areas for the object being investigated.

Apart from the bottom-up approach described in this paper, i.e. using quality indicators identified by a literature study, another possibility to link existing knowledge with OQM-results is to use expert knowledge for the areas(s) and object under study. The combination of the approaches followed by OQM and QADS is the best way to assure a complete identification of all quality indicators for all quality areas of the object under study.

References

Batini, C., Scannapieco, M. (2006) *Data Quality: Concepts, Methodologies and Techniques*. Springer verslag, Berlin Heidelberg, Germany.

Daas, P.J.H. (2009). On the quality of data and data sources (in Dutch). Paper written for the Quality project group of the Payroll registration chain, Statistics Netherlands, Heerlen.

Daas, P.J.H., Arends-Tóth, J. (2007). Quality aspects of registers: conceptual framework (in Dutch), DMH-2007-10-30-PDAS, Statistics Netherlands, Heerlen.

Daas, P.J.H., Arends-Tóth, J., Schouten, B., Kuijvenhoven, L. (2008a). Proposal for a Quality Framework for the Evaluation of Administrative and Survey Data. *Paper for the workshop on the Combination of surveys and administrative data*, 29-30 May, Vienna, Austria.

Daas, P.J.H., Arends-Tóth, J., Schouten, B., Kuijvenhoven, L. (2008b). Quality Framework for the Evaluation of Administrative Data. *Paper for the Q2008 European Conference on Quality in Official Statistics*, 9-11 July, Rome, Italy.

Daas, P.J.H., Beukenhorst, D. (2008). Data sources of Statistics Netherlands: primary and secondary data sources (in Dutch). DMH-2008-10-24-PDAS, Statistics Netherlands, Heerlen.

Daas, P.J.H., Fonville, T.C. (2007). Quality control of Dutch Administrative Registers: An inventory of quality aspects. *Paper for the seminar on Registers in Statistics - methodology and quality*, 21-23 May, Helsinki, Finland.

Daas, P.J.H., Ossen, S.J.L. (2010). Quality Evaluation of Administrative Metadata. *Journal of Quality*, submitted for publication.

Daas, P.J.H., Ossen, S.J.L., Arends-Tóth, J. (2009a). Framework of Quality Assurance for Administrative Data Sources. *Paper for the 57th session of the International Statistical Institute*, 16-22 August, Durban, South Africa.

Daas, P., Ossen, S., Vis-Visschers, R., Arends-Tóth, J. (2009b). Checklist for the Quality evaluation of Administrative Data Sources. *Discussion paper* 09042, Statistics Netherlands, The Hague/Heerlen.

Eurostat (2003). Assessment of the quality in statistics, Item 4.2: Methodological documents, Definition of quality in statistics. Sixth meeting, 2–3 October, Luxembourg, Luxembourg.

Eurostat (2005). Quality in statistics, Standard quality indicators. Seventh meeting, 23-24 May, Luxembourg, Luxembourg.

Ivanov, K. (1972). *Quality control of information*. The Royal Institute of Technology and the University of Stockholm, Department of Information Processing Computer Science, Stockholm, Sweden.

Knight, S-A., Burn, J. (2005). Developing a Framework for Assessing Information Quality on the World Wide Web. *Informing Science Journal*, 8, pp. 159-172.

Kuijvenhoven, L., Schouten, B. (2008). Quality indicators of the Data hyperdimension. Internal CBS-paper (in Dutch), DMV-2008-03-31-BSTN-LKYN, Statistics Netherlands, Voorburg.

Ossen, S.J.L., Puts, M., Daas, P.J.H. (2010) Quality framework for registers applied to online price information and offline route information (in preparation). *Paper for the Q2010 European Conference on Quality in Official Statistics*, 4-6 May, Helsinki, Finland.

Pipino, L.L., Lee, W.L., Wang R.Y. (2002). Data Quality Assessment. *Communications of the ACM*. 45(4), pp. 211-218.

Søresen, H.T,, Sabroe, S., Olsen, J. (1996). A Framework for Evaluation of Secondary Data Sources for Epidemiological Research. *Int. J. Epidemiol*. 25(2), pp. 435-442.

Van Nederpelt, P.W.M. (2009a). Object Oriented Quality Management, a management model for quality. Report 29 April 2009, Statistics Nederlands, The Hague/Heerlen.

Van Nederpelt, P.W.M. (2009b) Statistics Netherlands develops a quality model (in Dutch). SIGMA, 3, pp. 18-21.

Van Nederpelt, P.W.M. (2009b). The creation and application of a new quality management model. *Discussion paper* 09040, Statistics Netherlands, The Hague/Heerlen.

Van Nederpelt, PWM. (2009c). The creation and application of a new quality management model. *Statistika* 05, pp. 385-395.

Wang, R.Y., Strong, D.M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4), pp. 5-34.

Annex A: List of characteristics

Ability
Acceptability
Acceptance
Access security
Accessibility
Accountability
Accuracy
Adaptability
Adaptively
Adequacy
Administrativeness
Advisability
Ambiguity
Amount (of data)
Applicability
Appreciably
Appropriateness
Approximation
Attractiveness
Authenticity
Availability
Awareness
Balance
Beauty
Believability
Benevolence
Brightness
Capacity
Changeability
Clarity
Clearness
Coherence
Colour
Comparability
Compatibility
Competence
Competitiveness
Completeness
Complexity
Comprehensibility
Concentration power
Conciseness
Confidentiality
Consistency
Contestability
Continuity
Controllability
Convenience
Correctness
Costs
Coverage
Creativity
Credibility
Creditworthiness
Currency
Data freshness
Degree of detail
Degree of filling
Delivery reliability
Dependability
Dependency
Detailedness
Disputability
Diversity
Docility
Dynamics
Ease of manipulation
Ease of understanding
Ease of use
Effectiveness
Efficiency
Eligibility
Employability
Empowerment
Enforceability
Enthusiasm
Environment
friendliness
Equivalence
Evenness
Exactness
Exclusivity
Existence
Expertise
Extensibility
Fairness
Familiarity
Faultlessness
Feasibility
Find ability
Flexibility
Format
Freedom from error
Friendliness
Functionality
Goodness
Growth
Health
Height
Helpfulness
Image
Impartiality
Importance
Incorrectness
Independence
Innovativeness
Integrity
Intensity
Interoperability
Interpretability
Inventiveness
Involvement
Legality
Legitimacy
Length
Level
Linkability
Loyalty
Maintainability
Market orientation
Measurability
Metadata compliance
Mobility
Mutation rate
Navigation
Necessity
Numerical consistency
Objectivity
Openness
Operability
Orientation to…
Over coverage
Passion
Perfection
Performance
Periodization
Pertinence
Plausibility
Portability
Power to ...
Precision
Predictability
Presence
Privacy
Productivity
Professionalism
Profitability
Proportionality
Provenance
Prudence
Punctuality
Purchasing power
Purity

Readability
Reasonableness
Recency
Record identifiability
Record matching ability
Redundancy
Reference time
Refinement
Relevance
Relevancy
Reliability
Repeatability
Replaceability
Representational consistency
Reproducibility
Reputation
Response
Responsibility
Responsiveness
Result orientation
Re-usability
Rightness
Robustness
Safety
Satisfaction
Scope
Secrecy
Security
Selectivity
Sensitivity
Severity
Shape
Size
Soundness
Speed
Stability
Stress resistance
Structure
Substitutability
Suitability
Sustainability
Tenacity
Thoroughness
Timeliness
Traceability
Transferability
Transparency
Trueness
Turnaround
Under coverage
Understandability
Uniformity
Uniqueness
Usability
Usefulness
Validity
Value
Value-added
Verifiability
Visibility
Voluntariness
Vulnerability
Weight
Wholeness
Width
Willingness to ...

Annex B: Quality area evaluation results for the data domain (part 1)

Characteristic (dimension)	Domain	Sub-object	Quality area description (as a question)	Rephrased wordings or question	Valid com-bination	First charact-eristic chosen	Second charact-eristic chosen	Selected characteristic
Accuracy	data	dataset	How accurate is the data in the dataset?	NOT APPLICABLE: Accuracy of data is meaningless at the dataset level	-			
Accuracy	data	units	How accurate is the data for the units in the dataset?	Correctness of identification keys for the units and correctness of referral to intended (real world) unit	+	Authenticity	Correctness	Authenticity
Accuracy	data	items	How accurate is the data for the items in the dataset?	Absence of measurement and other errors	+	Accuracy	Correctness	Correctness
Authenticity	data	dataset	How authentic is the data in the dataset?	NOT APPLICABLE: Authenticity of data is meaningless at the dataset level	-			
Authenticity	data	units	How authentic is the data for the units in the dataset?	Does the administrative unit refer to the intended (real world) unit?	+	Authenticity		Authenticity
Authenticity	data	items	How authentic is the data for the items in the dataset?	Does the data for the items in the data source correspond with the intended (real world) data for those items?	+	Correctness		Correctness
Availability	data	dataset	How available is the data in the dataset for SN?	Availability, readability, and completeness of the dataset	+	Availability	Completeness	Completeness
Availability	data	units	How available is the data for the units in the dataset for SN?	Availability, readability, and completeness of units information in the dataset	+	Availability	Completeness	Completeness
Availability	data	items	How available is the data for the items in the dataset for SN?	Availability, readability, and completeness of item information in the dataset	+	Availability	Completeness	Completeness
Clarity	data	dataset	How clear is the data in the dataset for SN?	Is SN able to access and read the information in the dataset?	+	Availability	Completeness	Completeness
Clarity	data	units	How clear is the data for the units in the dataset for SN?	NOT APPLICABLE: Already covered at the dataset level	-			
Clarity	data	items	How clear is the data for the items in the dataset for SN?	NOT APPLICABLE: Already covered at the dataset level	-			
Coherence	data	dataset	How coherent is the data in the dataset?	NOT APPLICABLE: Coherence is meaningless at the dataset level	-			
Coherence	data	units	How coherent is the data for the units in the dataset?	Uniqueness of the (identification keys used for the) units	+	Uniqueness	Coverage	Uniqueness
Coherence	data	items	How coherent is the data for the items in the dataset?	Coherence between the data of the items (per unit)	+	Coherence		Coherence
Comparability	data	dataset	How comparable is the data in the dataset?	NOT APPLICABLE: Comparability is meaningless at the dataset level	-			
Comparability	data	units	How comparable is the data for the units in the dataset?	Uniqueness of the identification keys of the units in the dataset	+	Uniqueness		Uniqueness
Comparability	data	items	How comparable is the data for the items in the dataset?	Uniqueness of the item information in the dataset, from an identification point of view	+	Uniqueness	Confidentiality	Uniqueness
Completeness	data	dataset	How complete is the data in the dataset?	Is all data delivered and can all data be accessed in the dataset?	+	Completeness		Completeness
Completeness	data	units	How complete is the data for the units in the dataset?	Coverage of the units in the dataset	+	Completeness	Coverage	Completeness
Completeness	data	items	How complete is the data for the items in the dataset?	Completeness of the information for the items in the dataset	+	Completeness		Completeness
Confidentiality	data	dataset	How confidential is the data in the dataset?	NOT APPLICABLE: Confidentiality of data is meaningless at the dataset level	-			
Confidentiality	data	units	How confidential is the data for the units in the dataset?	Confidentiality of the unit information in the dataset	+	Confidentiality		Confidentiality
Confidentiality	data	items	How confidential is the data for the items in the dataset?	Confidentiality of the item information in the dataset	+	Confidentiality		Confidentiality
Correctness	data	dataset	How correct is the data in the dataset?	Is SN able to correctly open and read the information in the dataset?	+	Availability	Completeness	Completeness
Correctness	data	units	How correct is the data for the units in the dataset?	Correctness of the identification key used for the units	+	Correctness	Uniqueness	Correctness
Correctness	data	items	How correct is the data for the items in the dataset?	Absence of measurement and other errors	+	Correctness		Correctness
Coverage	data	dataset	What is the coverage of the data in the dataset?	Does the data in the dataset cover the information demand of SN?	+	Completeness		Completeness
Coverage	data	units	What is the coverage of the data for the units in the dataset?	Does the data for the units in the dataset cover the unit information demand of SN?	+	Coverage	Completeness	Completeness
Coverage	data	items	What is the coverage of the data for the items in the dataset?	Does the data for the items in the dataset cover the item information demand of SN?	+	Completeness		Completeness
Dependency	data	dataset	How much does SN depend on the data in the dataset?	Dependency risk of SN on the availability and completeness of the data source	+	Dependency	Completeness	Completeness
Dependency	data	units	How much does SN depend on the data for the units in the dataset?	Dependency risk of SN on the availability and completeness of the unit information in the data source	+	Dependency	Completeness	Completeness
Dependency	data	items	How much does SN depend on the data for the items in the dataset?	Dependency risk of SN on the availability and completeness of the item information in the data source	+	Dependency	Completeness	Completeness
Detailedness	data	dataset	How detailed is the data in the dataset?	NOT APPLICABLE: Level of detail is meaningless at the dataset level	-			
Detailedness	data	units	How detailed is the data for the units in the dataset?	Level of detail and uniqueness of the unit information	+	Uniqueness	Detailedness	Uniqueness
Detailedness	data	items	How detailed is the data for the items in the dataset?	Level of detail for the item information in the dataset	+	Detailedness		Detailedness
Format	data	dataset	What is the data format of the data in the dataset?	Is SN able to access and read all the information in the dataset?	+	Availability	Completeness	Completeness
Format	data	units	What is the data format of the data for the units in the dataset?	NOT APPLICABLE: Already covered at the dataset level	-			
Format	data	items	What is the data format of the data for the items in the dataset?	NOT APPLICABLE: Already covered at the dataset level	-			

Annex B: Quality area evaluation results for the data domain (part 2)

Characteristic (dimension)	Domain	Sub-object	Quality area description (as a question)	Rephrased wordings or question	Valid combination	First characteristic chosen	Second characteristic chosen	Selected characteristic
Importance	data	dataset	How important is the data in the dataset for SN?	Importance of the availability and completeness of the data source for SN	+	Dependency	Completeness	Completeness
Importance	data	units	How important is the data for the units in the dataset for SN?	Importance of the availability and completeness of the unit information in the data source for SN	+	Dependency	Completeness	Completeness
Importance	data	items	How important is the data for the items in the dataset for SN?	Importance of the availability and completeness of the item information in the data source for SN	+	Dependency	Completeness	Completeness
Linkability	data	dataset	How linkable is the data in the dataset?	NOT APPLICABLE: Linkability of data is meaningless at the dataset level	-			
Linkability	data	units	How linkable is the data for the units in the dataset?	How well can the unit information be used to link the data to those of units in other datasets?	+	Linkability	Uniqueness	Uniqueness
Linkability	data	items	How linkable is the data for the items in the dataset?	How well can item information be used to link the data to those of units in other datasets?	+	Linkability	Uniqueness	Uniqueness
Punctuality	data	dataset	How punctual is the data in the dataset?	Is the dataset completely available (delivered) on time?	+	Punctuality	Completeness	Completeness
Punctuality	data	units	How punctual is the data for the units in the dataset?	NOT APPLICABLE: Already covered at the dataset level	-			
Punctuality	data	items	How punctual is the data for the items in the dataset?	NOT APPLICABLE: Already covered at the dataset level	-			
Relevance	data	dataset	How relevant is the data in the dataset for SN?	Importance of the availability and completeness of the data source for SN	+	Dependency	Completeness	Completeness
Relevance	data	units	How relevant is the data for the units in the dataset for SN?	Importance of the availability and completeness of the unit information in the data source for SN	+	Dependency	Completeness	Completeness
Relevance	data	items	How relevant is the data for the items in the dataset for SN?	Importance of the availability and completeness of the item information in the data source for SN	+	Dependency	Completeness	Completeness
Reliability	data	dataset	How reliable is the data in the dataset?	NOT APPLICABLE: Reliability of the data in the dataset is meaningless at the dataset level	-			
Reliability	data	units	How reliable is the data for the units in the dataset?	Correctness of the unit information in the dataset	+	Correctness	Accuracy	Correctness
Reliability	data	items	How reliable is the data for the items in the dataset?	Correctness of the item information in the dataset	+	Correctness	Accuracy	Correctness
Replaceability	data	dataset	How replaceable is the data in the dataset?	Dependency risk of SN on the availability and completeness of the data source	+	Dependency	Completeness	Completeness
Replaceability	data	units	How replaceable is the data for the units in the dataset?	Dependency risk of SN on the availability and completeness of the unit information in the data source	+	Dependency	Completeness	Completeness
Replaceability	data	items	How replaceable is the data for the items in the dataset?	Dependency risk of SN on the availability and completeness of the item information in the data source	+	Dependency	Completeness	Completeness
Selectivity	data	dataset	How selective is the data in the dataset?	NOT APPLICABLE: Selectivity of data is meaningless at the dataset level	-			
Selectivity	data	units	How selective is the data for the units in the dataset?	Selectiveness of the coverage of the units in the dataset	+	Selectivity	Coverage	Selectivity
Selectivity	data	items	How selective is the data for the items in the dataset?	Selectiveness of the item information in the dataset	+	Selectivity	Accuracy	Selectivity
Size	data	dataset	What is the size of the data in the dataset?	What is the size of the dataset (data file)?	+	Completeness	Size	Completeness
Size	data	units	What is the size of the data for the units in the dataset?	Does the data for the units in the dataset cover the unit information demand of SN?	+	Coverage	Completeness	Completeness
Size	data	items	What is the size of the data for the items in the dataset?	Does the data for the items in the dataset cover the item information demand of SN?	+	Completeness		Completeness
Stability	data	dataset	How stable is the data in the dataset?	Stability of the content of the dataset over time, changes in number of records	+	Stability	Size	Stability
Stability	data	units	How stable is the data for the units in the dataset?	Stability of the unit information in the dataset over time, changes in coverage	+	Stability	Coverage	Stability
Stability	data	items	How stable is the data for the items in the dataset?	Stability of item information in the dataset over time, comparability over time	+	Stability		Stability
Timeliness	data	dataset	How timely is the data in the dataset?	Recentness of the information in the dataset	+	Timeliness		Timeliness
Timeliness	data	units	How timely is the data for the units in the dataset?	Recentness of the unit information in the dataset	+	Timeliness		Timeliness
Timeliness	data	items	How timely is the data for the items in the dataset?	Recentness of the item information in the dataset	+	Timeliness		Timeliness
Uniqueness	data	dataset	How unique is the data in the dataset?	Importance of the availability and completeness of the dataset for SN	+	Dependency	Completeness	Completeness
Uniqueness	data	units	How unique is the data for the units in the dataset?	Presence and uniqueness of identification keys for the units	+	Uniqueness	Coverage	Uniqueness
Uniqueness	data	items	How unique is the data for the items in the dataset?	Uniqueness of the item information in the dataset, from an identification point of view	+	Confidentiality		Confidentiality
Usefulness	data	dataset	How useful is the data in the dataset for SN?	NOT APPLICABLE: Much to general characteristic, usefulness is a combination of other quality dimensions	-			
Usefulness	data	units	How useful is the data for the units in the dataset for SN?	NOT APPLICABLE: Much to general characteristic, usefulness is a combination of other quality dimensions	-			
Usefulness	data	items	How useful is the data for the items in the dataset for SN?	NOT APPLICABLE: Much to general characteristic, usefulness is a combination of other quality dimensions	-			

Annex C: Quality area evaluation results for the metadata domain (part 1)

Characteristic (dimension)	Domain	Sub-object	Quality area description (as a question)	Rephrased wordings or question	Valid combination	First characteristic chosen	Second characteristic chosen	Selected characteristic
Accuracy	metadata	dataset	How accurate is the metadata for the dataset?	Error free and complete metadata for the dataset	+	Correctness	Completeness	Correctness
Accuracy	metadata	units	How accurate is the metadata for the units for the dataset?	Error free and complete metadata for the units in the dataset	+	Correctness	Completeness	Correctness
Accuracy	metadata	items	How accurate is the metadata for the items for the dataset?	Error free and complete metadata for the items in the dataset	+	Correctness	Completeness	Correctness
Authenticity	metadata	dataset	How authentic is the metadata for the dataset?	Specificity of metadata for the dataset, availability of tailor made metadata	+	Correctness	Completeness	Correctness
Authenticity	metadata	units	How authentic is the metadata for the units for the dataset?	Error free and complete metadata for the units in the dataset	+	Correctness	Completeness	Correctness
Authenticity	metadata	items	How authentic is the metadata for the items for the dataset?	Error free and complete metadata for the items in the dataset	+	Correctness	Completeness	Correctness
Availability	metadata	dataset	How available is the metadata for the dataset?	Is metadata provided and complete for the dataset?	+	Availability	Completeness	Completeness
Availability	metadata	units	How available is the metadata for the units for the dataset?	Is metadata provided and complete for the units in the dataset?	+	Availability	Completeness	Completeness
Availability	metadata	items	How available is the metadata for the items for the dataset?	Is metadata provided and complete for the items in the dataset?	+	Availability	Completeness	Completeness
Clarity	metadata	dataset	How clear is the metadata for the dataset?	Is metadata for the dataset available and comprehensible?	+	Clarity		Clarity
Clarity	metadata	units	How clear is the metadata for the units for the dataset?	Is metadata for the units in the dataset available and comprehensible?	+	Clarity		Clarity
Clarity	metadata	items	How clear is the metadata for the items for the dataset?	Is metadata for the items in the dataset available and comprehensible?	+	Clarity		Clarity
Coherence	metadata	dataset	How coherent is the metadata for the dataset?	Is metadata for the dataset available, complete, and comprehensible?	+	Clarity	Completeness	Clarity
Coherence	metadata	units	How coherent is the metadata for the units for the dataset?	Is metadata for the units in the dataset available, complete, and comprehensible?	+	Clarity	Completeness	Clarity
Coherence	metadata	items	How coherent is the metadata for the items for the dataset?	Is metadata for the items in the dataset available, complete, and comprehensible?	+	Clarity	Completeness	Clarity
Comparability	metadata	dataset	How comparable is the metadata for the dataset?	Similarity of the metadata to those of other data sources	+	Comparability		Comparability
Comparability	metadata	units	How comparable is the metadata for the units for the dataset?	Similarity of the metadata for the units to those of other units (in other data sources)	+	Comparability	Linkability	Comparability
Comparability	metadata	items	How comparable is the metadata for the items for the dataset?	Similarity of the metadata for the items to those of other items (in other data sources)	+	Comparability		Comparability
Completeness	metadata	dataset	How complete is the metadata for the dataset?	Is the metadata completely available for the dataset?	+	Completeness		Completeness
Completeness	metadata	units	How complete is the metadata for the units for the dataset?	Is the metadata completely available for the units in the dataset?	+	Completeness		Completeness
Completeness	metadata	items	How complete is the metadata for the items for the dataset?	Is the metadata completely available for the items in the dataset?	+	Completeness		Completeness
Confidentiality	metadata	dataset	How confidential is the metadata for the dataset?	Safety measures taken for exchange of the data source	+	Confidentiality		Confidentiality
Confidentiality	metadata	units	How confidential is the metadata for the units for the dataset?	NOT APPLICABLE: Already covered at the dataset level	-			
Confidentiality	metadata	items	How confidential is the metadata for the items for the dataset?	NOT APPLICABLE: Already covered at the dataset level	-			
Correctness	metadata	dataset	How correct is the metadata for the dataset?	Error free metadata for the dataset	+	Correctness		Correctness
Correctness	metadata	units	How correct is the metadata for the units for the dataset?	Error free metadata for the units in the dataset	+	Correctness		Correctness
Correctness	metadata	items	How correct is the metadata for the items for the dataset?	Error free metadata for the items in the dataset	+	Correctness		Correctness
Coverage	metadata	dataset	What is the coverage of the metadata for the dataset?	Completeness of the metadata for the dataset	+	Completeness		Completeness
Coverage	metadata	units	What is the coverage of the metadata for the units for the dataset?	Completeness of the metadata for the units in the dataset	+	Completeness		Completeness
Coverage	metadata	items	What is the coverage of the metadata for the items for the dataset?	Completeness of the metadata for the items in the dataset	+	Completeness		Completeness
Dependency	metadata	dataset	How much does SN depend on the metadata for the dataset?	Is metadata available and complete for the dataset?	+	Availability	Completeness	Completeness
Dependency	metadata	units	How much does SN depend on the metadata for the units for the dataset?	Is metadata available and complete for the units in the dataset?	+	Availability	Completeness	Completeness
Dependency	metadata	items	How much does SN depend on the metadata for the items for the dataset?	Is metadata available and complete for the items in the dataset?	+	Availability	Completeness	Completeness
Detailedness	metadata	dataset	How detailed is the metadata for the dataset?	Is metadata available and complete for the dataset?	+	Completeness		Completeness
Detailedness	metadata	units	How detailed is the metadata for the units for the dataset?	Is metadata available and complete for the units in the dataset?	+	Completeness		Completeness
Detailedness	metadata	items	How detailed is the metadata for the items for the dataset?	Is metadata available and complete for the items in the dataset?	+	Completeness		Completeness
Format	metadata	dataset	What is the format of the metadata for the dataset?	Can the metadata be accessed completely and read?	+	Availability	Completeness	Completeness
Format	metadata	units	What is the format of the metadata for the units for the dataset?	Can the metadata for the units be accessed completely and read?	+	Availability	Completeness	Completeness
Format	metadata	items	What is the format of the metadata for the items for the dataset?	Can the metadata for the items be accessed completely and read?	+	Availability	Completeness	Completeness

Annex C: Quality area evaluation results for the metadata domain (part 2)

Characteristic (dimension)	Domain	Sub-object	Quality area description (as a question)	Rephrased wordings or question	Valid combination	First characteristic chosen	Second characteristic chosen	Selected characteristic
Importance	metadata	dataset	How important is the metadata for the dataset?	Is metadata available and complete for the dataset?	+	Availability	Completeness	Completeness
Importance	metadata	units	How important is the metadata for the units for the dataset?	Is metadata available and complete for the units in the dataset?	+	Availability	Completeness	Completeness
Importance	metadata	items	How important is the metadata for the items for the dataset?	Is metadata available and complete for the items in the dataset?	+	Availability	Completeness	Completeness
Linkability	metadata	dataset	How linkable is the metadata for the dataset?	Interoperability of the dataset in various systems	+	Comparability	Linkability	Comparability
Linkability	metadata	units	How linkable is the metadata for the units for the dataset?	Similarity of the metadata for the units to those of other units (in other data sources)	+	Comparability		Comparability
Linkability	metadata	items	How linkable is the metadata for the items for the dataset?	Similarity of the metadata for the items to those of other items (in other data sources)	+	Comparability		Comparability
Punctuality	metadata	dataset	How punctual is the metadata for the dataset?	Timely availability and recentness of the metadata for the dataset	+	Timeliness	Availability	Timeliness
Punctuality	metadata	units	How punctual is the metadata for the units for the dataset?	Timely availability and recentness of the metadata for the units in the dataset	+	Timeliness	Availability	Timeliness
Punctuality	metadata	items	How punctual is the metadata for the items for the dataset?	Timely availability and recentness of the metadata for the items in the dataset	+	Timeliness	Availability	Timeliness
Relevance	metadata	dataset	How relevant is the metadata for the dataset?	Completeness and correctness of metadata for dataset	+	Completeness	Correctness	Completeness
Relevance	metadata	units	How relevant is the metadata for the units for the dataset?	Completeness and correctness of metadata for the units in the dataset	+	Completeness	Correctness	Completeness
Relevance	metadata	items	How relevant is the metadata for the items for the dataset?	Completeness and correctness of metadata for the items in the dataset	+	Completeness	Correctness	Completeness
Reliability	metadata	dataset	How reliable is the metadata for the dataset?	Correctness of metadata for dataset	+	Correctness		Correctness
Reliability	metadata	units	How reliable is the metadata for the units for the dataset?	Correctness of metadata for the units in the dataset	+	Correctness		Correctness
Reliability	metadata	items	How reliable is the metadata for the items for the dataset?	Correctness of metadata for the items in the dataset	+	Correctness		Correctness
Replaceability	metadata	dataset	How replaceable is the metadata for the dataset?	Similarity of the metadata to those of other data sources	+	Comparability		Comparability
Replaceability	metadata	units	How replaceable is the metadata for the units for the dataset?	Similarity of the metadata for the units to those of other units (in other data sources)	+	Comparability	Linkability	Comparability
Replaceability	metadata	items	How replaceable is the metadata for the items for the dataset?	Similarity of the metadata for the items to those of other items (in other data sources)	+	Comparability		Comparability
Selectivity	metadata	dataset	How selective is the metadata for the dataset?	Completeness of the metadata for the dataset	+	Completeness	Correctness	Completeness
Selectivity	metadata	units	How selective is the metadata for the units for the dataset?	Completeness of the metadata for the units in the dataset	+	Completeness	Correctness	Completeness
Selectivity	metadata	items	How selective is the metadata for the items for the dataset?	Completeness of the metadata for the items in the dataset	+	Completeness	Correctness	Completeness
Size	metadata	dataset	What is the size of the metadata for the dataset?	Completeness of the metadata for the dataset	+	Completeness		Completeness
Size	metadata	units	What is the size of the metadata for the units for the dataset?	Completeness of the metadata for the units in the dataset	+	Completeness		Completeness
Size	metadata	items	What is the size of the metadata for the items for the dataset?	Completeness of the metadata for the items in the dataset	+	Completeness		Completeness
Stability	metadata	dataset	How stable is the metadata for the dataset?	Changeability of the metadata for the dataset	+	Stability	Timeliness	Stability
Stability	metadata	units	How stable is the metadata for the units for the dataset?	Changeability of the metadata for the units in the dataset	+	Stability	Timeliness	Stability
Stability	metadata	items	How stable is the metadata for the items for the dataset?	Changeability of the metadata for the items in the dataset	+	Stability	Timeliness	Stability
Timeliness	metadata	dataset	How timely is the metadata for the dataset?	Recentness of the metadata for the dataset	+	Timeliness		Timeliness
Timeliness	metadata	units	How timely is the metadata for the units for the dataset?	Recentness of the metadata for the units in the dataset	+	Timeliness		Timeliness
Timeliness	metadata	items	How timely is the metadata for the items for the dataset?	Recentness of the metadata for the items in the dataset	+	Timeliness		Timeliness
Uniqueness	metadata	dataset	How unique is the metadata for the dataset?	Similarity of the metadata to those of other data sources	+	Comparability		Comparability
Uniqueness	metadata	units	How unique is the metadata for the units for the dataset?	Similarity of the metadata for the units to those of other units (in other data sources)	+	Comparability	Linkability	Comparability
Uniqueness	metadata	items	How unique is the metadata for the items for the dataset?	Similarity of the metadata for the items to those of other items (in other data sources)	+	Comparability		Comparability
Usefulness	metadata	dataset	How useful is the metadata for the dataset?	NOT APPLICABLE: Much to general characteristic, usefulness is a combination of other quality dimensions	-			
Usefulness	metadata	units	How useful is the metadata for the units for the dataset?	NOT APPLICABLE: Much to general characteristic, usefulness is a combination of other quality dimensions	-			
Usefulness	metadata	items	How useful is the metadata for the items for the dataset?	NOT APPLICABLE: Much to general characteristic, usefulness is a combination of other quality dimensions	-			