

QUALITY MANAGEMENT OF REFERENCE GEO-INFORMATION

A. Jakobsson^{a,*}, A. Hopfstock^b, M. Beare^c, R. Patrucco^d

^a Development Centre, National Land Survey of Finland, P.O. Box 84, 00521 Helsinki, Finland - antti.jakobsson@nls.fi

^b Federal Agency for Cartography and Geodesy (BKG), Richard-Strauss-Allee 11, 60598 Frankfurt a.M., Germany-
anja.hopfstock@bkg.bund.de

^c 1Spatial Group Ltd., Tennyson House, Cambridge Business Park, Cambridge, CB4 0WZ, United Kingdom- matt.beare
@1spatial.com

^d Ordnance Survey, Adanac Drive, Southampton, SO16 0AS, United Kingdom- Ray.Patrucco@ordnancesurvey.co.uk

KEY WORDS: Standards, Quality Management, Quality Assurance, Automation, Usability, Trust, Quality Assessment

ABSTRACT:

This paper will introduce how quality of geo-information can be managed when the production environment is no longer inside one organization (e.g. collection of data is contracted out) or data is compiled from various sources like in case of Spatial Data Infrastructures (SDIs). The bases for quality management of reference geo-information are discussed using three viewpoints; data, process and organization and user centric viewpoints. These viewpoints can be met using ISO 19157 and ISO 19158 standards together with ESDIN developed Quality Model and Data Quality Services Framework (DQSF). Two different services are identified a *Data Quality Web Service* and a *Data User Web Service*.

We discuss how these principles and services are implemented now within EuroGeographics and Ordnance Survey of Great Britain. Further development will be done during the European Location Framework (ELF) project, which is providing a single source of reference geo-information for Europe during 2013-2016.

1. INTRODUCTION

1.1 Consequences of Supply Chain Change to Data Quality

In previous times determining dataset quality, 'fitness for purpose' or even ease of use for datasets had been a relatively simple exercise. The datasets themselves were simple (for example a single point feature with one or two associated attributes). These datasets appeared even simpler as those that had created them were very often the only organisations consuming the data. As a result downstream systems were designed to cope specifically with that dataset in mind; the dataset's limitations were well understood and accounted for by most if not all the users within the organisation.

This simple picture has of course become far more complex than many could have envisaged. The world is creating and consuming information at an alarming rate. The data has become far more complex (for example multi-feature object based datasets with multiple attributes with complex intra-relationships). This new data is being created (and even in some cases maintained) by multi-organisation supply chains. The data is being consumed by multiple organisations for at least as many different uses. Determining quality in this complex environment requires more than planning by one or two organisations. To ensure the most value is extracted from the data, if all are to succeed, it requires a more holistic approach to quality.

1.2 Changing Use of Reference Geo-information

The INSPIRE Directive of the European Union (European Union, 2007) establishes the basis for sharing and delivering the

geospatial data for environmental purposes. Annex I and II of the Directive define the reference geo-information part, which is important for the thematic data reference (Annex III of the Directive). Reference geo-information can be defined as series of datasets that everyone involved with geographic information uses to reference his/her own data to as part of their work. They provide a common link between applications and thereby provide a mechanism for sharing knowledge and information amongst people (FGDC, 2005; Rase et al., 2002; GSDI, 2009). In the previous chapter we noted how production of reference geo-information is changing. Supply chains (A supply chain being a collection of processes, some within the same organisation some outside organised together to produce one or more products) and reference geo-information are becoming more complex. For example, in INSPIRE context, one reference theme might contain data from several authorities.

Previously reference geo-information has mostly been used as a backdrop map and then other information has been overlaid onto it. With an introduction of Linked Data (Berners-Lee, 2009) concept and e.g. need for connecting more attribute information to spatial data, it is important to manage change and ensure that the latest data is used. Reference geo-information will come available through platforms, where users do not have to manage datasets, but can start integrating Data as a Service (DaaS) (Wikipedia, 2013) for their applications. This requires change in how data quality will be managed. One of the advantages of DaaS is that data quality can be managed as there is a single point for updates.

2. BASES FOR QUALITY MANAGEMENT OF REFERENCE GEO-INFORMATION

Quality Management of Reference Geo-Information must address; a) the provision of cost/time effective and standardised framework to measure and improve quality, b) meeting changing needs, c) increase of users trust and how to create confidence in the usage of available data to make informed decisions. We introduce here three aspects for quality management of reference geo-information that will have to be taken into account. These aspects are based on Jakobsson, Tsoulos (2007) and Jakobsson (2006).

2.1 Data Centric Approach

Geo-information quality standards ISO 19113, ISO 19114 and ISO 19138 have been replaced by one ISO 19157 Geographic Information – Data Quality – standard. It defines the data quality elements; completeness, logical consistency, positional accuracy, thematic accuracy and temporal quality, but also introduces an additional element – usability. Basically, this enables introduction of new measures that could meet changing user requirements. Metadata standards (like ISO 19115) are often considered quality standards as well because they contain information, which can be used for determining some aspects of quality. Elements like usage, lineage and date of last revision are good examples. These standards build an important basis for the quality management of reference geo-information and can be considered as data centric part of it. However, as pointed out by Devillers et al. (2010), metadata approaches have not really been a successful because of complexity of data quality.

2.2 Process and Organization Centric Approach

ISO 9000 is widely used quality management standard series which concentrates on process and organization centric part of the quality management. Recently accepted technical specification ISO19158 Geographic Information- Quality Assurance of Data Supply (2012) offers a framework in which a modern supply chain can understand the quality requirement of the data being produced (or maintained). In addition to data quality part, which is based on the ISO 19157, the technical specification considers other aspects of quality that would impact upon a supply chain: the schedule of delivery, the volume of delivery and the cost of delivery. This approach can then be used to assure that the entire supply chain is capable of producing the quality required in those terms. The framework allows the supply chain to be broken down into its constituent processes and with particular consideration for human interaction in the data production or maintenance processes.

Using ISO 19158 to gain assurance in any given supply chain the user requirements must first be understood. The relevant elements of quality must be identified followed by the assignment of measures and acceptable quality levels. These must relate directly to the real customer requirement or at least the perceived customer requirement. Once this requirement is identified it may be used by management elements within the supply chain to identify both the required outputs and expected inputs of individual processes. It enables user of the specification to understand the propagation of data error (through poor quality data being passed on to the next process) as well as the impact of poor scheduling and data volume management.

This level of understanding is achieved through a process that reviews, tests and assures each element of the production or maintenance processes. For any given process that impacts upon the data there are up to three levels of assurance (Basic, Operational and Full). Not all levels are mandated; the levels may be considered similar to risk mitigation and so with higher levels achieved comes greater assurance but at a higher implementation cost. The successful implementation of ISO 19158 is dependent on the relationship between customer and supplier as well as the customer's understanding of the processes undertaken by the supplier. This becomes even more critical with more complex data. If it is not possible for the customer to understand the process then a 'technical agent' should be used to act on their behalf. Note that the customer may or may not be internal to the supplier organisation.

The basic level of quality assurance ensures that a process appears to be capable of creating or maintaining product to the right quality. As this is predominantly a paper exercise it provides the lowest level of assurance. The supplier provides appropriate evidence to the customer (or their agent) which will identify their suitability for the production or maintenance of the dataset. For example documentation to be reviewed might include proposed management structure, quality plans, change control plans, training plans, tool specifications and high level process maps. The detail required at this level and others should be proportionate to the quality risk posed. The outcome of this assurance activity can provide information for the next level. For example it may identify areas where quality control is critical or where there are likely data flow restrictions 'bottlenecks' in the proposed production process.

The operational level of assurance comes from an assessment of a working process following implementation. Rather than the previous high level process review approach Operational assurance looks at all relevant processes in detail and breaks them down further as required (The requirement to do so is often identified at the basic level). At this level data outputs are checked for conformance to the agreed quality requirement. At this level individual operators' work is also assessed and assured. In this way all staff have a responsibility for quality not just quality control staff. The proportion of the staff that must have achieved an appropriate level of individual assurance is agreed between supplier and customer; allowing for staff churn and training. The responsibility for training and testing normally resides with the supplier however the customer (or agent thereof) is required to review achievements in terms of training records and the data quality results of their individual output. Once the appropriate number of staff has achieved the agreed standard and the output of the processes confirms that data quality, volumes achieved and schedule adherence is acceptable the supplier can be said to have achieved operational assurance. The testing of data (both product and individual output) may be reduced at this point as the risk diminishes.

The final level of assurance ensures that the supplier is capable of maintaining the quality achieved at the operational level over a period of time. This period will be agreed between Supplier and Customer. Data quality result trends will be analysed and reviewed by management of supplier and customer for the life of the process with the aim of continually improving the supply chain.

2.3 User Centric Approach

Usability is defined in ISO 9241 as “the effectiveness, efficiency, and satisfaction with which specified users achieve specified goals in particular environments”. It has its roots in engineering especially software development (e.g. Nielsen, 1993). Usability has been studied in connection of geo-information e.g. by Wachowicz and Hunter (2003). In the previous chapters usability is a key method for identifying the needed measures both in ISO 19157 and ISO 19158. However, in many cases these measures, while important for evaluation of quality, are not mentioned in user interviews. Users sometimes prefer verbal results or statements from other users how good the data is. From the user point of view trust is the key factor in selection of data or service to be used. How this trust is then created is an interesting question. Different methods of creating trust include for example certification (as in case of ISO 9000 certification), accreditation (e.g. ISO 17025 for laboratory testing) and now ISO 19158, which introduces assurance levels. Other examples include quality labelling e.g. Geo Label for Geoss (GeoViqua, 2013) and quality visualisation.

Using authoritative sources also creates trust for users and this is especially important for reference geo-information because most of the reference data is produced by public agencies. An authoritative source is “a managed repository of valid or trusted data that is recognized by an appropriate set of governance entities and supports the governance entity’s business environment” (Westman, 2009). The challenge here is that even if the source is considered authoritative it may lose users’ trust if it does not deliver good quality.

3. QUALITY MANAGEMENT OF REFERENCE GEO-INFORMATION IN A MULTI DATASET SDI ENVIRONMENT

Implementation of the above principles in to a multi dataset SDI environment is not yet implemented. In ESDIN project ISO 19157 principles were utilized into Data Quality Services Framework (DQSF see Figure 1). This framework contains services for data supplier and data user.

ISO 19158 has not yet been tested in an SDI context. ‘Full’ interoperability relies on all parties having the same perspective on data quality however the implementation of ISO19158 within an SDI would ensure that data quality is at least understood by all relevant parties. The proliferation of discovery quality metadata (aggregated from the data quality results identified in the assurance process) would provide this opportunity.

3.1 ESDIN work

Recently finished ESDIN project (ESDIN, 2011) made some fundamental findings in the quality management of reference geo-information. The project’s main focus was to study how to implement the INSPIRE directive for reference geo-information. Its central findings on quality were that a) an integrated model of quality and quality measures can be created for reference geo-information (Quality Model) and b) quality validation can be automated as a rule (DQSF). When these results are put into practice they will fundamentally change the quality management of reference geo-information.

3.1.1 Approach ESDIN approach is illustrated in Figure 1 (Beare et al., 2010, Jakobsson et al 2011). It uses ISO 19157 as a framework for evaluating data quality using quality measures. It includes parts that can be checked automatically like conformance rules and parts that require manual checking like completeness and positional accuracy. A Quality Model has to be defined for each dataset. This will set the quality requirements using the quality measures (from ISO 19157). Quality requirements should be set using users’ requirement studies.

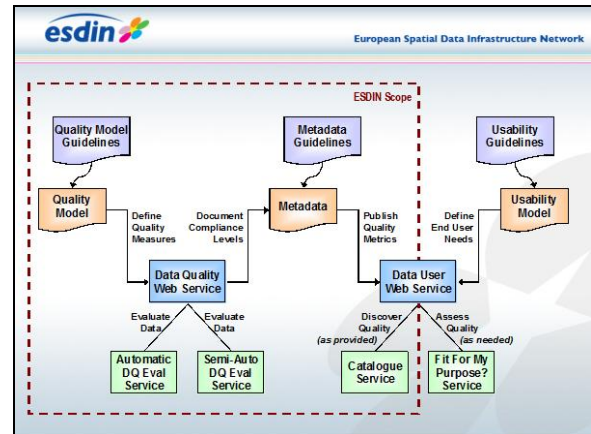


Figure 1 ESDIN DQSF

After running an evaluation using the *Data Quality Web Service* results may be reported in metadata. Two kinds of evaluation metadata may be provided. The first and the most common case would be a *dataset level metadata* for the feature types and attribute types reporting conformance levels set in the Quality Model. These conformance levels are validated by *Semi-Manual Service* typically through sampling but actual test results are not reported in metadata. Typically these measures are related to completeness, positional accuracy and thematic accuracy. For logical consistency and temporal accuracy actual test result may be provided as the whole dataset may be tested automatically.

In ESDIN project a need for *Data User Web Service* was recognized. This builds on setting the usability model based on user requirements and then *Data User Web Service* will give advice whether the data meets the user requirements or not.

4. USE CASES

4.1 EuroGeographics

EuroGeographics is a not-for-profit organization representing 56 national mapping, land registries and cadastral authorities (NMCAs) in 45 European countries. It has a long experience in building harmonized datasets based on its members’ data.

4.1.1 EuroRegionalMap (ERM) Quality Control: ERM is a topographic dataset covering 35 countries in Europe and based on NMCAs data at European regional level of details (1:250 000). While technical interoperability can be ensured by the use of a common data model (ERM data model) it is more challenging to provide comparable and harmonised data content. As shown in Pammer et al. (2009) the national production workflows vary due to national constraints, resources, and the availability and accessibility of suitable data sources. Mainly, national specifics cause deviations from the ERM specification with respect to the selection criteria, level of generalisation, and quality (Hopfstock et al., 2012).

ERM data specification provides a description of the content, accuracy, and data. The quality requirements are indicated as general requirements (absolute horizontal accuracy, data density level and selection criteria, dimensions: geometric resolution) and portrayal and quality criteria by Feature Type. In order to assure good quality of the resulting ERM product the ERM Validation Specifications details the validation procedure that should be carried out throughout the production process:

- The national producers are responsible for the validation of their national contribution using whenever possible the validation tools implemented for the final data validation and assessment phase. It is the responsibility of the data producer to ensure completeness of data collection.
- ERM Regional coordinators perform the final validation and QA of the national data components for final acceptance.
- A final validation and QA is carried out after the data assembly phase on the full European dataset by the Product Management Team.

The validation procedures consist of a series of checks to identify errors in the data's geometrical and topological structure as well as feature/attribute compliance with the current ERM specifications and in the consistency of data collection. The current process is mainly focused on supporting the production management. The validation results have been returned to each producer with recommendations as to how the national contribution can be improved for the next release.

4.1.2 ERM Quality Assurance: Since the successful completion of the ESDIN project the ERM production management team has started to apply the guidance of the ESDIN Quality Model for the ERM data set. With the assistance of ISpatial, the collation of national data contributions has been enhanced through the introduction of an automated data quality evaluation process. This process has enabled full data set evaluation for Transport, Hydrography and Settlement themes from all 32 national contributors.

Providing uniform assessment against a common set of (around 200) quality measures (business rules), quantitative and comparative metrics are automatically compiled for each national data set, with coverage of: geometric resolution; domain value integrity and topological connectivity (including cross-border consistency). These metrics provide objective viewpoints on the comparative quality of national contributions and awareness of the consistency of the data product across the whole of Europe. This knowledge helps to increase the confidence levels that the EuroGeographics ERM production management team have in the product quality prior to distribution to customers. Additionally, detailed non-conformance reports provide the management team with the information needed to advise data providers where they could best utilise their resources to improve data quality for future product releases, thereby instilling an informed continuous improvement process.

The discussion regarding conformance levels resulted in the definition of three conformance classes and corresponding acceptance levels based on the ERM quality criteria as shown in Table 1.

Conformance levels			
Criteria	Description	Class	Acceptance level
Shall be fulfilled	Basic requirements of ERM dataset	Class A	100%
		Class C	<100%
To be fulfilled	Basic requirements for user purpose (to be described for each item, see update plan requirements)	Class A	100%
		Class B	>90%
		Class C	<90%
Nice to have	General requirements of user - a goal that needs to be achieved (i.e. the attribute completeness rate for mandatory)	Class A	100%
		Class B	>75%
		Class C	<75%

Table 1. Proposed Conformance Classes

Further, a first ERM layer has been evaluated using the above conformance classes. This means the validation results for each ERM quality criteria have been assessed and assigned an acceptance level. Then the results have been grouped by data quality element to come up with an overall quality grading by country following the grade proposed by the ESDIN project (Table 2).

Grade	Data Quality description
Excellent	Only class A for all quality measures
Very good	A majority of A's, but also some B's
Good	A majority of B's, some A's, no C's
Adequate	Only a very few C's, the other B's and better
Marginal	A majority of C's but also some B's
Not good	No measure reached the class B (i.e. all measures on class C)

Table 2. Grading data (ESDIN 2011)

The discussion of the results in the ERM Production Management Team showed that the aggregation of data quality results poses some issues regarding:

- Aggregation where measurements at different scales and units
- Aggregation for inhomogeneous data
- Reporting details

The pilot implementation of ESDIN Quality Model proofed the applicability of the proposed ESDIN data model. However, a good understanding of the ISO standards on data quality is required. Further, the objective of the quality reporting, i.e. report to producers, management or users need to be clearly defined.

Also application of ISO 19158 has been started for the ERM producers. First basic level assurance has been achieved by the new NMCAs joining the EuroGeographics' production programme.

4.2 Ordnance Survey

Ordnance Survey implemented their approach to quality assurance (a forerunner to ISO19158) following the experiences it had gained from letting contracts over the years. Prior to implementation Ordnance Survey had let contracts to maintain its large scale database with limited input to the specific processes, tools and individuals that would be updating data on their behalf. As datasets become more complex Ordnance Survey and their suppliers started to experience data quality issues. Through the application of the approach outlined above Ordnance Survey was able to better support its suppliers and in return they received the quality of data that they required, to the appropriate volume and schedule. Realising the benefit of their approach they then applied it to all internal production and maintenance processes.

The approach has been successful in identifying data quality issues early in the process development cycle providing opportunity for Ordnance Survey to work with its suppliers to resolve those issues before they become unmanageable. As datasets have become even more complex there is greater opportunity for this approach to add value. With this approach the customer, supplier, individual operator have a good understanding of the data quality that is required and the quality that is being produced. As the relationship between the two is continually monitored it may be managed proactively and effectively.

There are challenges to be overcome. For example many consider that there is no customer value in quality metadata: to the end user the value lies in the data itself. As datasets become more complex more assurance is potentially loading processes and individuals with essential but 'non-value adding' costs and at the same time adding precious time to the process. As a result of this there is a tipping point at which individuals will become disenfranchised with the production and maintenance process which in turn will have a negative effect on data quality. This challenge may be mitigated with the investment in automated testing as discussed earlier.

5. FURTHER DEVELOPMENT

5.1 European Location Framework Project

The European Location Framework (ELF) project will during the next three years deliver the first implementation of the European Location Framework (Jakobsson, 2012) - a technical infrastructure which harmonises national reference data to deliver authoritative, up-to-date, interoperable, cross-border geospatial reference data for use by the European public and private sectors in a way that is easy to use by application developers and even end users.

The project will provide a critical mass of content and coverage as 15 Member States' national Elf/INSPIRE data will be made available from a single source (ELF platform) connecting the it to number of applications, the European Commission INSPIRE geo-portal, the Commission Internal portal run by Eurostat and ArcGIS Online, a commercial Cloud GIS platform. ELF platform will be implemented using an Open Source development made originally for the Finnish SDI, Oskari. Covering the full range of INSPIRE Annex I,II and III themes, these datasets will provide full national coverage of the rich content available from national and regional spatial data infrastructures.

In the ELF project quality evaluation based on ESDIN results will be operationalized using cloud based commercial services. The goal is also to introduce a standard way in which quality models can be expressed as rules, which enables using these in multiple software environments.

ELF – with its 'partnership' approach to the customer/supplier relationship the implementation of ISO19158 offers opportunities in organisational interoperability; encouraging organisation and process alignment. This alignment can lead to opportunities at the technical level (particularly around the resolution of data quality issues). It supports the requirement for quality metadata for discovery purposes and given the findings of the ESDIN project it also supports the approach, and provides a framework for, repetition of the quality evaluation process for all process steps in SDIs.

6. CONCLUSIONS

The main drivers for introducing better quality management to reference geo-information production are partly based on government policies like e-government, legislation (e.g. INSPIRE directive), cost effectiveness and then partly on users demands. We believe that introduction of the ESDIN *Data Quality Web Service* in the ELF project will decrease production cost and time. This will enable faster and more frequent release of reference geo-data at national, regional and global level. Further, introduction of ISO 19158 at national but

also at international level will increase users trust to reference geo-information.

REFERENCES

- Beare, M., Henriksson, R., Jakobsson, A., Marttinen, J., Onstein, E., Tsoulos, L., Williams, F., Mäkelä, J., De Meulenaer, L., Persson, I. and I. Kavadas, 2010b D8.4 ESDIN Quality Final Report – Part A. [pdf] ESDIN consortium. http://www.esdin.eu/sites/esdin.eu/files/D8-4_ESDIN_Quality_Final_Report.pdf. (accessed 27th April 2013)
- Berners-Lee, T., 2009. Linked Data, <http://www.w3.org/DesignIssues/LinkedData.html> (accessed 27th April 2013)
- Devillers, R., Stein, A., Bédard, Y., Chrisman, N., Fisher, P. and Shi, W. 2010. 30 years of research on Spatial Data Quality – Achievements, failures and opportunities. *Transactions in GIS*, 14(4): 387-400
- ESDIN (2011) ESDIN D1.10.3 Public Final Project Report, <http://www.esdin.eu> (accessed 27th April 2013)
- EuroGeographics (2012) <http://www.eurogeographics.org> (accessed 27th April 2013)
- European Union, 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). <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32007L0002:EN:NOT> (accessed 27th April 2013)
- FGDC (2005). *Framework Introduction and Guide* (Handbook), Digital version. <http://www.fgdc.gov/framework/frameworkintroguide/> (accessed 27th April 2013)
- GSDI (2009). *The SDI Cookbook*, http://www.gsdidocs.org/GSDIWiki/index.php/Main_Page (accessed 27th April 2013)
- Hopfstock et.al. (2012) EuroRegionalMap – A joint production effort in creating a European Topographic Reference Dataset. In: Jobst M. (ed.), *Service Oriented Mapping 2012*, JOBST Media Präsentation Management Verlag. Wien 2012, pp. 151-162
- Jakobsson A., 2006. *On the Future of Topographic Base Information Management in Finland and Europe*, Doctoral dissertation. Helsinki, University of Technology. <http://lib.tkk.fi/Diss/2006/isbn9512282062/> (accessed 27 April 2013)
- Jakobsson, A., Tsoulos, L., 2007. The Role of Quality in Spatial Data Infrastructures. In *Proceedings of the 23rd International Cartographic Conference*, Moscow, Russia, Cd-Rom
- Jakobsson, A., Mäkelä, J., Henriksson, R., Tsoulos, L., Beare, M., Marttinen, J., Lesage, N., 2009. Quality Beyond Metadata – Implementing Quality in Spatial Data Infrastructures. In: *Proceedings of the 24th International Cartographic Conference*, Santiago de Chile, Chile, Cd-Rom
- Jakobsson, A., 2012. Introducing a New Paradigm for Provision of European Reference Geo-Information – Case Study of the European Location Framework Concept, In: Jobst M. (ed.), *Service Oriented Mapping 2012*, JOBST Media Präsentation Management Verlag. Wien 2012, pp. 51-62
- Jakobsson, A., Beare, M., Marttinen, J., Onstein, E., Tsoulos L. and F. Williams, 2011 A Cohesive Approach Towards Quality Assessment of Spatial Data and its Automation. [cd-rom] In: *Proceedings of the 25th International Cartographic Conference*, Paris, France
- Nielsen, J. 1993. *Usability Engineering*. Boston:Academic Press
- Pammer et.al. (2009) EuroRegionalMap – How to succeed in Overcoming National Borders, In Gartner G. & Ortog F. (Eds.), *Lecture Notes in Geoinformation and Cartography: Cartography in Central and Eastern Europe*. CEE 2009 (1st ed., pp. 19–40), Heidelberg, Neckar: Springer Berlin
- Rase, D., Björnsson A., Probert, M. and M-F. Haupt, eds., 2002. Reference Data and Metadata Position Paper. Eurostat, <http://inspire.jrc.ec.europa.eu/> (accessed 26th January 2012)
- Wachowicz, M. and G. Hunter, 2003. Spatial Data Usability, Preface. *Data Science Journal* (Spatial Data Usability Special Section), Vol. 2, 26 February 2003, https://www.jstage.jst.go.jp/article/dsj/2/0/2_0_75/_article (accessed 27th April 2013)
- Westman R., 2009. What Constitutes an Authoritative Source? In: Durrant A. et al (Eds.) *Proceedings of the Privilege Management Workshop*, September 1-3,2009. U.S. Department of Commerce, National Institute of Standards and Technology. http://csrc.nist.gov/news_events/privilege-management-workshop/presentations/Roger_Westman.pdf (accessed 27th April 2013)