Large Scale Map Data Contents and End User Needs: Do they meet? – Case: City of Helsinki

Kari Ingberg

City of Helsinki, City Survey Department, Field Survey Office
Postal address: P.O. Box 2205, 00099 City of Helsinki, Finland
Email: kari.ingberg@hel.fi

Abstract. Differences in views what should be collected to large scale map as features lead inevitably to quality variation in end products and enforces uncertainty. Research hypothesis in this study is the assumption that for large scale maps some features are collected unnecessarily and some features that are needed are not collected at all. Research problem is approached both from municipal data collection and end user perspectives. The objective is to identify possible inconsistencies in large scale map data collection features and reflect them to end user needs. Research methods in this case study are; semi structured interviews, participating observations and a semi structured group interview. Results indicate notably information content mismatch what is collected and what should be collected. Conclusions can be made that terrain, building and transportation related feature information collection should be extended and at the present collected additional feature information is not contributing to end user needs.

Keywords: Large Scale Map, Feature Classification, Uncertainty, Quality

1. Introduction

Large scale map is here defined as a map produced to the scale form 1:500 to 1:2 000 with general topographic mapping principles, which creates base data set for planning, real estate management and municipal technical planning. (NLS 1997) Large scale map data set is usually produced by means of aerial imagery, field survey, digitizing, real estate management operations and building inspection measurements. Large scale maps correspond to city plan maps (Kostet 1992) which is prescribed by a decree in Finland by national mapping authority i.e. National Land Survey of Finland (NLS). Differences in views what should be collected as map features lead inevitably to quality variation in map products and it is obvious that quality of any spatial data is crucial for its effective use. (Goodchild 2006) In addition to general importance of uncertainty and quality issues (Devillers & Jeansoulin 2006) more attention should be paid to large scale maps in municipalities. It is verified that large scale maps in municipalities are important data sources in built-up areas and those can be used in some extent to compile national level e.g. scale 1:10 000 data sets (Jakobsson 2006). Our environment in real world changes rapidly especially in urban environment (Servigne et al. 2006) and population is concentrated (Kostet 1992, Jakobsson 2004) where large scale map data is initially collected and mostly used.

Research hypothesis in this study is the assumption that for large scale maps some features are collected unnecessarily and some features that are needed are not collected at all. The research problem is approached from two perspectives; 1) from local data collection and national guidance relation view point and 2) from end user needs view point.

Research questions from internal and external quality (Devillers & Jeansoulin 2006) perspectives arise: "What geographic information has been collected?", "Which geographic features are important to the end users?" and finally "Are those map features collected that are really needed?"

Objective is to identify possible inconsistencies in large scale map features, first in reference to national guidance (NLS 1997) and secondly to end user needs and finally to get adjustments to data collection feature scheme.

It is assumed that two universes of discourses (UoD) (Servigne et al. 2006) exists; Helsinki (HEL UoD) and National Land Survey (NLS UoD). National Land Survey of Finland as a national mapping authority provides official guidance (NLS 1997) to municipalities. For HEL UoD there exists no official written consistent documented feature catalog (Lagerstedt et al. 2011) instead map features collected are based on earlier documentation (HEL 1987), technical environment feature identification code lists and tacit knowledge. Official guidelines to municipalities give a lot of freedom defining the content of large scale map. (Jakobsson 2006) On the other hand City of Helsinki has a strong will to integrate to national spatial data infrastructure (GINorden 2011).

In the earlier research related to feature catalogs in municipalities (Jakobsson 2006) research project studied the possibility of using municipality data for the production of national level topographic database. In that study it was reported that feature classification systems used in the municipalities are not sufficiently compatible to national level and this prevents feasible national usage of municipality data. The justification to this particular research into information contents semantics is more straightforward research approach from local City of Helsinki level mapping to official national level not using regional level classification system in between provided by Association of Finnish Local and Regional Authorities as in earlier research. End user need prioritization issues in international research often relate to generalization (Elzakker & Berg 2010, Foerster et al. 2012) where similar problem area can be found.

The framework of this research related to geographic information uncertainty and quality will be presented next in Chapter 2. Research arrangements and procedures for data collection and end user needs are presented as material and methods in Chapter 3. Results from both data collection and end user needs are presented in Chapter 4. Discussion on results is presented in Chapter 5 and conclusions in Chapter 6.

2. Framework and scope of this study

Both internal and external quality perspectives (Devillers & Jeansoulin 2006) are present in this case study. Quality internally in part 1: *3.1 Data Collection* concentrates on present data collection feature catalogs and their interpretations in data collection process. Quality externally in part 2: *3.2 End User Needs* concentrates on user needs through research on end user prioritization of collected features and possible adjustments on importance of what should be collected.

In the conceptual model of uncertainty (Devillers & Jeansoulin 2006) the research perspective is on poorly defined objects and mostly in the field of non-specificity. In part 1 (data collection) non-specificity is inherently present in data collection process because it involves assigning an object to a certain hierarchical classification based on interpretation of different observers. Vagueness is present in the form of loose feature catalog definitions and lack of consistent data collection specification. Discord is present in form of multiple incompatible feature catalog versions used by different data collectors. In part 2 (end user needs) discord is present in end user perspective because discord may arise if data collection process uses different feature catalog from end users feature catalog i.e. in this research HEL UoD versus NLS UoD. Prioritization of end users needs for features according to official NLS UoD feature classes may also be affected by vagueness in form of loose feature catalog definitions and end users possible unfamiliarity to official feature classification.

The core issues in uncertainty arise from the basic nature of geographic information as objects and attributes. Uncertainty can be present in various forms and in different life stages of geographic information in form of definition problem. (Devillers & Jeansoulin 2006) In part 1 (data collection) classification problem exists in form of multiple feature catalogs used in data collection process. Object definition problem exists in form of scarce object definitions in feature catalogs and lack of proper documented data collection specification in data collection process. Properties measurement problem exists in the form of different practices used in different data collection methods (photogrammetry, field survey, digitizing) and their inherent limitations. In part 2 (end user needs) classification problem and object definition problem exists from the end user perspective in the form of understanding geographic information contents presented in end product. Uncertainty in measurements (Devillers & Jeansoulin 2006) over three principal dimensions (attributes, space and time) is present here only in part 1 (data collection).

3. Research Arrangements and Procedures

This research is a case study (Stake 1995) with City of Helsinki. According to recent statistics (AFLRA 2011a) and survey (AFLRA 2011b) it is justified to say that City of Helsinki is the only large urban city in Finland that produces and updates 1:500 large scale vector map digitally with long city survey tradition. (Kostet 1992)

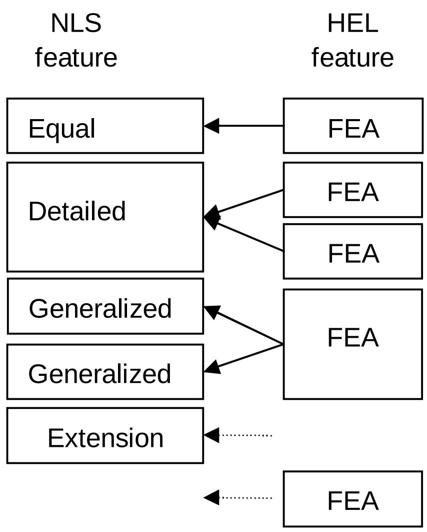
3.1. Data Collection

Data collection is studied with semi structured interview and supplementing participating observations. Practical arrangements for the semi structured interview for data collection personnel includes following guide lines. Interviewees are selected from different branches in data collection; aerial imagery, field survey, building inspection and base map construction. These four main interviewee categories contribute the content of a large scale map. Selection criterion for interviewees is that they have been working in that particular field at least for 10 years. Reasonable long working experience requirement for interviewees is motivated by the existing non documented state and the role of tacit knowledge in working environment (Lagerstedt et al. 2011). Selection criterion also ensures proper balance between real world and universe of discourse i.e. interviewees know and are familiar with the actual local terrain that they are mapping. Interviewees are in their natural working environment so that technical data collection environment is available including present map data sets. Natural working environment enforces semantic correctness of information content mapping and promotes efficiency by short answering times in interview situation. Interviews are carried out with 2 hour sessions maximum to ensure proper concentration.

Material for interview is based on existing governmental NLS feature catalog (NLS 1997) and HEL feature catalog. NLS is presenting map features as feature classes with attributes. HEL is presenting map features with identification codes. NLS feature class and attribute notations are converted into HEL like identification codes. This conversion is done to form an interview form used in interview situation as matrix presentation. In this matrix rows present NLS features, columns interviewees and cells HEL feature identification codes. Interview matrix contains all together 220 rows; 88 terrain, 7 elevation, 44 building, 35 transportation, 46 network related features.

The key issue and question to data collectors here for every feature in row vise is: "What HEL feature identification code do you use in your data collection environment to collect this information content?" Definition of the information content is taken from NLS feature catalog (NLS 1997). If information has been collected it is marked down with appropriate HEL feature code in interview matrix. Interviewer assures on the basis of national and municipal definitions (NLS 1997, HEL 1987) that possible synonymy and homonymy in conceptual level from existing data collection environment is mapped properly to official governmental feature catalogue with supporting open discussion.

Information content mapping indications from HEL UoD to NLS UoD is presented formal as look-up-table descriptions in *Figure 1*.



Feature collected with more detailed classification **“Detailed”**
Feature collected as defined **“Equal”**
Feature collected with more generalized classification **“Generalized”**
Feature not collected **“No”**
Feature collected as additional class **“Extension”**

Figure 1. A look-up-table presentation for HEL to NLS mapping with features.

Mapping information is gained from interview matrix with following definitions

- “Detailed” information is gained from interview matrix row referring to feature. If some cell contains more than one feature codes the indication is valid for that row.
- “Equal” information is gained from interview matrix row referring to feature. If some row contains only one kind of feature code the indication is valid for that row.
- “Generalized” information is gained from matrix by a search. If certain feature code is used in multiple rows the indication is valid for those rows.
- “No” information is gained from interview matrix row referring to feature. If some row contains no feature code the indication is valid for that row.
- “Extension” information is gained from interview matrix and feature code list by a search. If feature code list contains feature codes that are not in interview matrix the indication is valid for that feature code.

3.2. End User Needs

Semi structured group interview is used to find out end users needs. End user target group is selected based on previous study by Economic and Planning Center inside City of Helsinki. (EPC 2010) As a conclusion of this study it is appropriate to select representative group of end users from most active organizational units which in this case is City Planning Agency. Another equally important criterion for the selection of end users target group is base map's original primary usage which is initially city planning (NLS 1997).

Material for interview is based on existing governmental guidance (NLS 1997) which presents official universe of discourse for large scale map in Finland as feature catalog. Interviewees are office chiefs from City Planning Agency's City Plan Department which covers all city planning activities inside City of Helsinki. All together interviewees consist of 15 production offices. Interview starts with common meeting where the research is presented in general and instructions are given. To ensure equal level of knowledge of interviewees about contents of feature catalog is presented in general. Structure of the semi structured group interview can be presented as follows in *Figure 2*.

Prioritization of existing features (**+**, **0**, **-**)
Additional feature group suggestion and Additional feature class suggestion
Open comments on feature catalog contents

Figure 2. Structure of semi structured group interview form.

In prioritization of existing features section interviewees are asked to indicate for every feature on their point of view and special attention to their planning activities the coarse prioritization as follows:

+ (plus) important feature (can't be removed)
0 (null) neutral feature (good to be present)
- (minus) meaningless (can be removed)

Rule is not to answer if feature definition is unclear to interviewee. Feature definitions are opened up by interviewer as needed to ensure common understanding of features. Answering is made easy with answering sheet presented in *Figure 3*.

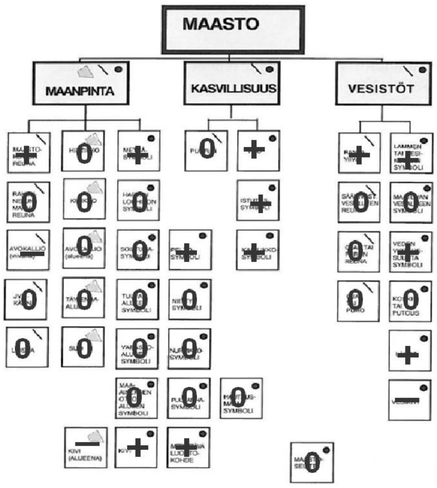


Figure 3. Example of answer sheet for end users needs.

In additional feature group suggestions and additional feature class suggestions section two questions is presented.

- "Is there a feature group (i.e. theme) that is missing and that is very important from city planning point of view?"
- Is there a feature class that is missing and that is very important from city planning point of view?"

In open comments section additional free text is allowed about feature catalog contents.

As a backup method to group interview and in order to ensure as much answers as possible, in addition to pre-prepared paper material, all material is provided also by email in electronic form. Interviewees that are not able to answer all the questions in group session are allowed to return answers afterwards.

4. Results

4.1. Data Collection

Semi structured interviews and participating observations were conducted between autumn 2011 and spring 2012 in several sessions. Stereo-operators (aerial imagery data collection) were interviewed as pairs in 6 separate sessions 11 hours total, building inspection (additions to buildings) in 2 separate sessions 4 hours total, field survey (supplements to aerial imagery data collection) in 4 separate sessions about 7 hours total, base map (end product refinement) in 5 separate sessions about 10 hours total. Interview times vary based on the amount of contribution to map contents.

Results from mapping HEL UoD to NLS UoD summarized with number of feature classes in *Figure 4*.

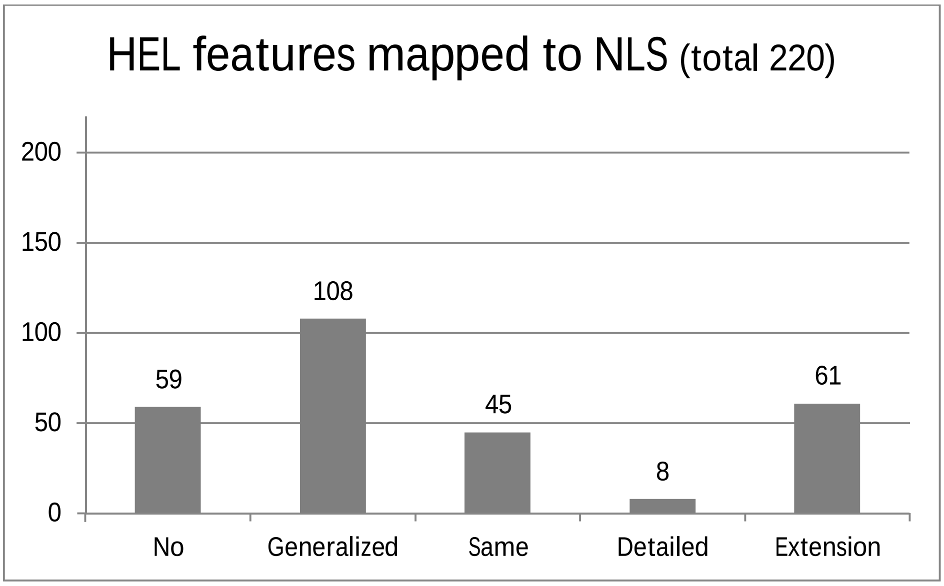


Figure 4. Mapping from HEL UoD to NLS UoD as number of feature classes.

Note that additional information collected as “Extension” is not included in the total number of feature classes (220).

4.2. End User Needs

Semi structured group interview was conducted in spring 2012. 11 answers were received out of 15 offices. 4 of the answers were returned in form of answering sheet and 7 in non-structured format as written text. After a group interview session participating interviewees were reminded several times about returning answering sheets.

As a result to prioritization of existing features interviewees indicated for every feature; + (plus) important feature (can’t be removed), 0 (null) neutral feature (good to be present), - (minus) meaningless (can be removed). Summary results by feature groups about prioritization in *Figure 5.*

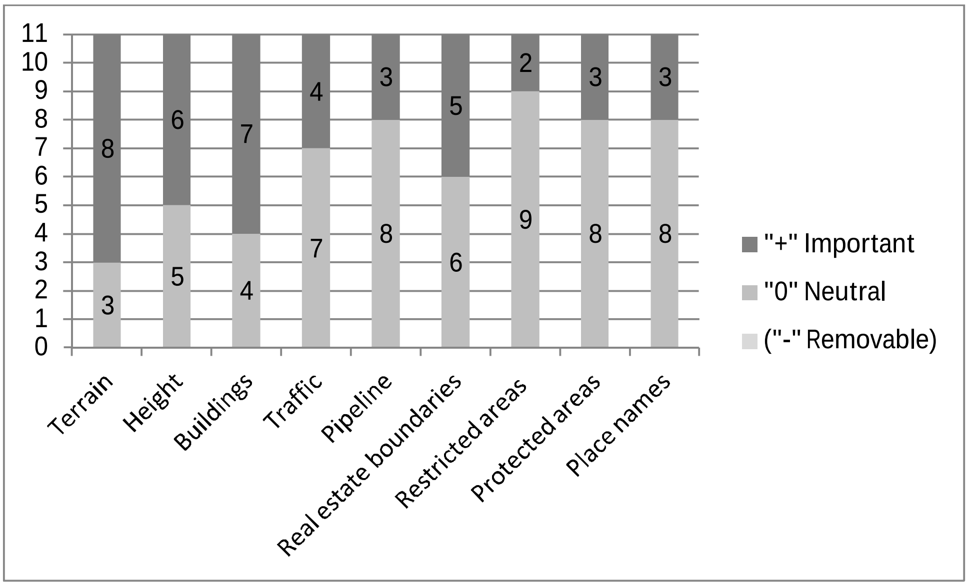


Figure 5. Summary results about feature group prioritization.

Note that none of the feature group or individual feature classes was indicated as meaningless i.e. removable with minus sign.

As a result to additional feature group suggestions and additional feature class suggestions; bathymetric contour lines and depth information on water areas were suggested more than once.

As a result to open comments about feature catalog contents; building heights (frequently), number of floors (frequently), height differences in buildings (frequently), entries to the buildings, tilt angel to steep roads. General comments included; “more information content the better“ (frequently), “good as it is now“ (few), “more up to date content“ (few) and few remarks to individual features visibility in technical environment was made.

4.3. Combined results from data collection and end user needs

Matrix representation of combined results from data collection and end user needs in *Table 1.*

Importance **“+” count**	No	Generalized	Equal	Detailed	Extension
Terrain **8**	23	34	28	3	6
Buildings **7**	2	34	6	2	17
Elevation **6**	2	-	5	-	1
Boundaries **5**	-	-	-	-	-

Transportation **4**	1	28	4	2	19
Network **3**	31	12	2	1	10
Protected areas **3**	-	-	-	-	2
Place names **3**	-	-	-	-	6
Restricted areas **2**	-	-	-	-	-

Table 1. Combined results from data collection and end user needs.

Columns represent data collection (part 1) feature class mapping from HEL to NLS and rows represent end user needs (part 2) with prioritization importance ("+") count.

5. Discussion

In the summary matrix representation (additional feature classes as "Extensions" not included and only most significant terrain, elevation, building, transportation, and network related features included) meaning of combined results can be derived from *Table 2.*

Importance **"+" percent** (feature count)	No	Generalized	Equal	Detailed
Terrain **73%** (88)	26%	39%	32%	3%
Buildings **64%** (44)	5%	76%	14%	5%
Elevation **55%** (7)	29%	-	71%	-
Transportation **36%** (35)	3%	80%	11%	6%
Network **27%** (45)	67%	27%	4%	2%

Table 2. Combined results from data collection and end user needs in percents.

As a general note, mismatch what is collected and what should be collected is notable, except in case of elevation information. Terrain as most important (73% importance weight) and largest (88 features) information content area suffers from the fact that information content is not collected at all (26%). Network's high mismatch rate (67%) is explained due to fact that separate network chart is collected in another process and measurement method for information content is not accurate in this case. Elevation's high mismatch rate (29%) and on the other hand high match rate (71%) is not significant due to relatively small feature count (7). Buildings (76%) and transportation (80%) suffer from too generic feature classification. More detailed classification is used relatively little. Additional features collected i.e. “Extension” information is presented in *Table 3.*

“+” Important percent (feature count)	Extension
Terrain 73% (88)	**7**% (6)
Buildings 64% (44)	**39**% (17)
Elevation 55% (7)	**14**% (1)
Transportation 36% (35)	**54**% (19)
Network 27% (45)	**22**% (10)

Table 3. Additional HEL features collected.

Notably amount of additional extra information is collected to transportation and building information as a supplement. This refers to a unique urban mapping environment and supports also the usage of case study research method in unique large scale mapping environment. Most of the end user’s wishes on additional features were related to building height information (metric height, floor count) although none of these additional feature classes (“Extension”) responded to this end user need in more detailed analysis.

6. Conclusions

The main conclusions of this study can be summarized as follows: (i) Findings indicate a significant uncertainty in large scale map feature information contents due to feature classification differences between local municipal implementation and official national guidance. End user needs do not meet in all feature groups. (ii) Results of this study agree well to the previous research findings that municipal large scale map feature information contents diverge from national uniformity (Jakobsson 2006). (iii) Terrain related feature information collection should be extended to enforce compatibility to national official guidance. (iv) Building related information should be collected with more detailed classification to enforce compatibility to national guidance and end user needs. (v) Transportation related information should be collected with more detailed classification to enforce compatibility to national guidance. (vi) Additional features collected to large scale map are not contributing to end users wishes.

Although mismatches and uncertainty exists in large scale map information content the present situation should be seen as an opportunity to enforce interoperability and the role of large scale maps as base data sets contributing to national spatial data infrastructure.

References

AFLRA, 2011a, Association of Finnish Local and Regional Authorities (Suomen kuntaliitto), Kaupunkien ja kuntien lukumäärä , http://www.kunnat.net/fi/tietopankit/tilastot/aluejaot/kuntien-lukumaara/Sivut/default.aspx accessed 12.10.2012

AFLRA, 2011b, Assosiation of Finnish Local and Regional Authorities (Suomen kuntaliitto), Kartastotietoja 2007, http://www.kunnat.net/fi/asiantuntijapalvelut/yty/kiinteisto-mittaustoimi/kimiti/Documents/Kartastotietoja%202007.pdf accessed 12.10.2012

Devillers, R., Jeansoulin, R., 2006, (ed. Devillers and Jeansoulin) in Fundamentals of Spatial Data Quality, ISTE Ltd, ISBN 10: 1-905209-56-8, p.17-20, p.31-42, p.43-59

Elzakker, C.,Berg, W., 2010, ASPRS/CaGIS, ISPRS Proceedings Topographic base maps for physical planning maps: user research for generalization http://www.isprs.org/proceedings/XXXVIII/part4/files/van%20Elzakker.pdf, accessed 12.10.2012

EPC, 2010, Economic and Planning Center in City of Helsinki, Geographic Information Survey – summary report 25.1.2011, Outi Hermans

Foerster T., Stoter J., Oosterom P. 2012: On-demand base maps on the web generalized according to user profiles, International Journal of Geographical Information Science, 26:1, 99-121

GINorden, 2011, Conference presentation - GIS Architecture, The City Of Helsinki, City of Helsinki, http://www.lounaispaikka.fi/paikkaoppi/presentations/GI_Norden_Hermans.pdf, accessed 12.10.2012

Goodchild, M., 2006, Foreword, (ed. Devillers and Jeansoulin) in Fundamentals of Spatial Data Quality, ISTE Ltd, ISBN 10: 1-905209-56-8, p.13-16

HEL, 1987, City of Helsinki, Kantakartta 1:500 (Base map 1:500) Kuvaustekniset ohjeet, Painatus Hankintakeskus 1987

Jakobsson, A., 2004, Maanmittaustieteiden Seuran julkaisu 41 - Maanmittaustieteiden päivät 2004, http://mts.fgi.fi/paivat/2004/mts41_jakobsson.pdf, accessed 12.10.2012

Jakobsson, A., 2006, On Needs for Harmonization and Joint-Use fo Topographic Information Produced by the National Land Survey and Municipalities in On the future of topographic base information management, Publications of the National Land Survey of Finland No.101, ISBN 951-48-0192-X, p.46, 97-142, 116

Kostet, J., 1992, Helsingin kaupunkimittauksen vaiheita - 100 vuotta ensimmäisestä kaupungin geodeetista 1892-1992 , Gummerus Kirjapaino Oy, Jyväskylä, ISBN 951-772-184-6, p.245, 158

Lagerstedt, K., Hårdh, J., Vento, R., Raudaskoski, M., Gröhn, J., Suppola, J., 2011, Interviews in Helsinki City Survey Department, nonstructured interview: on base map feature catalog in City of Helsinki (during September 2011)

NLS, 1997, Kaavan pohjakartta, Maanmittauslaitoksen julkaisuja n:o 85, National Land Survey of Finland, ISBN 951-48-0152-0

Servigne, S., Lesage, N., Libourel, T., 2006, Quality Components, Standards and Metadata, (ed. Devillers and Jeansoulin) in Fundamentals of Spatial Data Quality, ISTE Ltd, ISBN 10: 1-905209-56-8, p. 179-210

Stake, R., E., 1995, The Art of Case Study Research, SAGE Publications Inc, ISBN 0-8039-5766-1