

A PRAXIS ON DATA QUALITY EVALUATION OF UNDERGROUND PIPELINE

Haitao Zhao¹, Jin Zhou^{1,*}, Changfeng Jing², Xuefang Li²

¹ National Quality Inspection and Testing Center for Surveying and Mapping Products, 28 West Lianhuachi Road, Beijing, China-1033486498@qq.com

² School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, 1 Zhanlanlu Road, Beijing, China

KEY WORDS: Data Quality, Underground Pipeline, Geospatial Data

ABSTRACT:

Underground pipelines are known as “life line”. With the rapid developing of city, more and more pipelines like power lines will move into underground. Facing the complex environment from underground and relationship with other kinds of pipeline, the data quality evaluation is very crucial for academic and business applications. This paper introduced our praxis on underground pipeline data quality on a real project. The datasets are mainly composing of vector data about 15 GB size, covers 3 counties, worked with 3 teams. The workflow, data sampling method and quality evaluation method were engaged in our work. This work can extend to other underground pipeline projects or similar spatial data quality evaluation projects.

1. INTRODUCTION

With the rapid development of cities, many pipelines are moving into underground to free more space and land for city re-planning, such as power line, utility lines and so on. Underground pipeline plays crucial role in city development, disaster management(Eskandari et al., 2017; Li et al., 2019, 2018). Underground pipeline is more complex owing to it is buried under the ground. Although the buried method can protect its transform and damage, but its shape and directory cannot inspect easily, therefore, its spatial data quality is crucial for the land management and its operation. As stated in many literature(Huang et al., 2019; Jing et al., 2019; Kilkenny and Robinson, 2018; Najafabadi et al., 2015), “garbage in, garbage out”, which is a big challenge for various data. The poor data quality will lead to the bias decision or analysis result. Many literature focus on the underground topic, such as pipeline change detection(Wang et al., 2019), the 3D data model for land administrative(Kalogianni et al., 2020; Yan et al., 2019), and so on. However, there are few work on underground pipeline spatial data quality.

With the development of information technology, some integrated data acquisition technologies were developed. The underground pipeline is more complex than the pipelines on the ground owing to its invisible in underground. Generally, the trajectory and location are the two important information for the underground pipeline. Therefore, integrated data acquisition technologies with conversional surveying equipment like Total station or Laser scanning photogrammetry, and the subsurface geophysical detection technologies like Ground Penetrating Radar or Electromagnetic Locators(Lagüela et al., 2018; Wai-Lok Lai et al., 2018; Yan et al., 2019)

There are many factors on underground pipeline data quality. The manual importing error is the major error in

underground pipeline data. Wrongly inputted or handled data are the most common form. The fusion error come from the multiple heterogeneous data is the second factor for quality. Perhaps, there are two or three sets of underground pipeline data or system, and which are heterogeneous in data model and file format. Therefore, it is easily to cause fusion error when updating data with them. Finally, the dynamic damage from external causes like construction work. When the underground pipeline is damage, it may be repaired with different trajectory and direction in old form. Therefore, it brings the data quality problem.

The remainder of this paper is structured as follows. In Section 2, the objectives of our work are introduced. Section 3 presents the praxis data quality evaluation work on one city’s underground pipeline spatial data. The simple result was shown in Section4. Finally, Section 5 provides some concluding remarks and proposes some future work.

2. OBJECTIVES

With the rapid developing of cities, many and many pipes will move into underground to free the above space for urban planning. Therefore, the underground pipeline data quality evaluation can provide valuable work for urban planning. The objective of this paper is to provide a praxis work on underground pipeline data quality evaluation. From the preparing work to giving the inspection score, the paper will give the key stages and introduce the politics of our work.

3. THE PRAXIS OF UNDERGROUND PIPELINE SPATIAL DATA EVALUATION

3.1 The guideline of work

The guideline file of our work is divided two parts, the

* Corresponding author

organization quality files and the geospatial data quality files. For the afore one, we referred the ISO quality files and domain files. These files are below.

- (1) ISO 9001:2015 quality management system
- (2) GB/T19000—2016 "Quality Management System—Basics and Terminology"
- (3) GB/T19001-2016 "Quality Management System-Requirements"
- (4) GB/T24001—2016 "Environmental Management System Requirements and Guidelines for Use"

For the geospatial data quality file, we used the specific files and standard files. These files are below.

- (1) GB/T24356-2009 "Quality Inspection and Acceptance of Surveying and Mapping Results";
- (2) GB/T18316-2008 "Quality Inspection and Acceptance of Digital Surveying and Mapping Results";
- (3) CH/T 1033-2014 "Technical Specification for Quality Inspection of Pipeline Measurement Results";
- (4) CJJ100-2004 "Technical Specification for Urban Basic Geographic Information System";
- (5) CJJ 7-2007 "Code for Urban Engineering Geophysical Exploration";
- (6) CJJ/T 73-2010 "Technical Specification for Satellite Positioning Urban Measurement";

3.2 The workflow of data quality evaluation

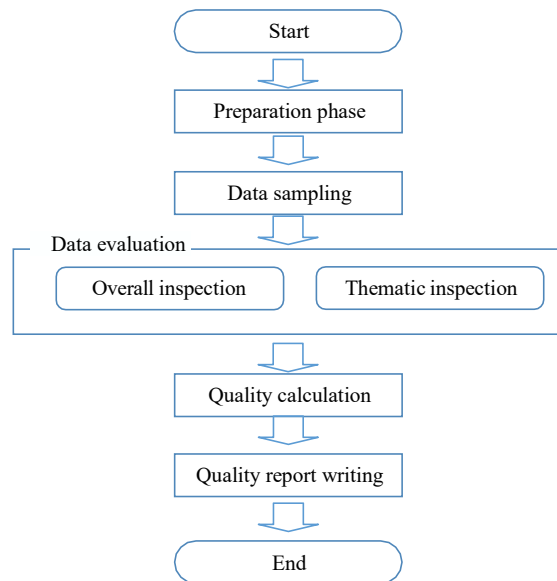


Figure 1. Data quality evaluation workflow

The data quality evaluation workflow is shown in Figure 1. The key stage is the data evaluation, which can be categorized as overall inspection and thematic inspection. The afore one is the general evaluation for dataset and the second one is the detail inspection for sampling spatial data.

3.3 Overall inspection

Overall inspection is an assessment of the overall quality

of the project. The inspecting items include the completion of project tasks, general inspection of non-sampling data, inspection of control measurement quality, and general inspection of factors affecting the quality of results. Data inspection uses the method of verification and analysis to conduct a general inspection of the compliance of the submitted results data. The inspection work requires that a second full general inspection be carried out after rectification of various quality errors.

(1) Inspection for completion of project tasks

The inspection of the completion of the project pays close attention to the progress of the contract period, the completion of the technical content and data of the contract. Mainly include the following three points.

- inspecting the performance of the operator's obligations and responsibilities stipulated in the contract
- whether the operating unit conducts a general survey in accordance with the contract requirements and the scope of the technical regulations, the content of the investigation and the selection criteria.
- Checking the standardization, completeness and correctness of the data submitted by the operating unit.

(2) General inspection of non-sampling data

The inspection of data content is the main content of data quality inspection. This work is mainly oriented to the general inspection of the overall quality of data, including data geometric expression, attribute value constraints, consistency of data logic concepts, consistency of data formats, and consistency of data topological relationships.

- The data geometric expression includes the geometric anomalies of pipeline elements; extremely short lines; extremely small faces; the correctness and compliance of the pipeline geometric edges.
- Data attribute value constraint refers to the uniqueness of all types of identification codes; the correctness and conformity of the correlation between the identification codes; the non-emptiness of each attribute item; the conformity of each attribute value constraint; the correctness and conformity of the attribute edge Sex and so on.
- Data consistency includes the rationality of data layer classification; compliance of attribute item definitions (such as name, type, length, ordinal number, decimal places, etc.);
- The consistency of the data format includes the compliance of the data file storage organization; the compliance of the data file format; whether the data file is missing, redundant, and the data cannot be read; the compliance of the data file name, etc.;
- Data topology consistency refers to the compliance of the definition of topological relationships; the same pipeline point, line, and surface layer include the compliance of the state. Does the pipeline point, line and surface overlap and cover? Are there wrong suspension points (compliance between points, lines, and surfaces)? Are there false pseudo-nodes (continuous pipeline compliance), interrupted, closed compliance?

(3) Inspection of control measurement quality

Controlling measurement data is the basis of all work, and its data quality is of great significance to the formation of project data quality. The quality inspection of control measurement data mainly includes the following contents.

- The compliance of the data coordinate system; the compliance of the elevation datum; the correctness of the projection parameters; the correctness of the starting point;
- Is the measurement method scientific? Compliance of observation methods (number or times of measurement, closure, attachment, branch); compliance of station tolerance; compliance of control point density; integrity and authenticity of measurement data;
- The correctness of the calculation method and calculation process; the completeness of the calculation data;
- Does it meet the tolerance requirements?
- The completeness of the information, the content and the clarity of the content.

(4) General inspection of factors affecting the quality

This work conducts a general inspection of factors that affect data quality, including primary data, control data, data structure, instrument verification and other major items and tendentious issues. Generally, only general problems are recorded after inspection.

3.4 Thematic inspection

According to the project situation, some thematic data can be selected for thematic inspection. In this project, we selected control measurement data, pipeline spatial data results, pipeline attribute data results, and ancillary facilities attribute results for thematic inspection. Here, we mainly introduce the inspection content of pipeline spatial data and attribute data. In addition to the pipeline spatial location information check, the spatial location relationship, such as the pipeline spatial connection relationship, the spatial hierarchical placement relationship of the pipeline, and the compliance of the boundary connection are the main contents of the pipeline spatial data inspection. The inspection of pipeline attribute data, the inspection content of this work mainly includes the data format, the integrity of the data content, and the correctness of the relationship of the data items. Of course, the logical relationship between pipeline diameter, flow direction, and pipeline points is also one of the important inspection contents.

3.5 Quality calculation

The quantitative calculation of quality is mainly based on the weighted average algorithm. The elements of quality inspection are formulated in accordance with the principles of GB/T 18316, and the weight distribution and classification of errors and omissions can be formulated in accordance with GB/T 24356. However, in project implementation, quality elements and weights are generally optimized and adjusted according to the actual characteristics of the project. In this project, quality elements and weights include control measurement (0.20), position accuracy (0.15), characterization quality (0.20), attribute accuracy (0.20), logical consistency (0.15), and data quality (0.10).

Quality element (weight)	Quality sub-element (weight)
control measurement (0.20)	Plane control survey (0.5), elevation control survey (0.5)
position accuracy (0.15)	Plane accuracy (0.35), elevation accuracy (0.35), exploration accuracy (0.30)
characterization quality (0.20)	Completeness (0.35), geographic expression (0.35), geometric expression (0.30)
attribute accuracy (0.20)	Attribute value binding (0.50), correctness of attribute content (0.50)
logical consistency (0.15)	Conceptual consistency (0.35), format consistency (0.35), topology consistency (0.30)
data quality (0.10)	Data completeness, correctness (0.70), finishing regularity (0.30)

Table 1. Quality elements and weights of pipeline data

The weighted average method is used to calculate the quality element score. The data quality score is calculated according to the following formula.

$$S_1 = \sum_{i=1}^n (S_{2i} \times p_i)$$

Where:

S_1, S_{2i} — quality element and corresponding quality sub-element score

p_i — the weight of the corresponding mass sub-element

n — the number of mass sub-elements contained in the mass element

3.6 Inspection for data item

The data item is the basic unit of data divided for inspection. It is often referred to as a "map sheet" in a topographic map or image plan of basic geographic information. Table 2 shows the inspecting item and quality elements.

Quality element	Inspecting item
Spatial reference system	Coordinate system, Elevation datum, Projection parameters, Starting data
Plane observation quality	Compliance and scientificity of measurement methods, observation methods, station tolerances, control point density; completeness and authenticity of measurement data.
Plane calculation quality	1. The correctness of the data calculation results; 2. Calculate the completeness of the data.
Plane math accuracy	1. Compliance of RTK measurement inspection; 2. The conformity of the error in

	the point position and the relative error in the side length.
Plane document quality	Control the completeness of the data, the content and the clarity of the content.
Elevation observation quality	Compliance and scientificity of measurement methods, observation methods, station tolerances, control point density; completeness and authenticity of measurement data.
Elevation calculation quality	1. The correctness of the calculated data; 2. Calculate the completeness of the data.
Mathematical accuracy of elevation	Compliance with the tolerances
Elevation document quality	Control the completeness of the data, the content and the clarity of the content.

Table 2. Quality elements

4. THE RESULT OF QUALITY CONTROL

4.1 Case study and dataset

The data collection for this project was completed in 2017. The pipeline data collected by the project includes information on main roads, secondary roads, branch roads and various underground pipelines within the scope of the suburban county and new towns, as well as various underground pipelines within the scope of the tenderer's requirements. The courtyards of factories, residential communities, parks, colleges, etc. are not included in the scope of this census. However, if there is a coherent main pipeline that crosses the above-mentioned area, the continuity of the pipeline should be maintained. The general survey of the project includes: channels, pipelines and cables buried underground. The types mainly include pipelines and auxiliary facilities such as water supply, drainage, gas, heat, electricity, communications, radio and television, industry, and various integrated pipe corridors.

The research area has many types of pipelines and complex pipeline orientation, which poses challenges for the development of production, quality inspection and other tasks.

The data is composed of more than 2000 map sheets covering more than 3000 thousand kilometer underground pipeline. This data collection work was collaborated among 3 teams.

4.2 Overview of inspection work

To evaluate the data quality, our work conducted on many indicators such as location accuracy, property precision, topologic inspection, and document quality. The weight of each element is preset by data type and expert knowledge. The data quality result is calculated by weighted sum algorithm. Then, the final data quality is graded according to the threshold value in the guideline files.

On the basis of GB/T 24356 "Quality Inspection and Acceptance of Surveying and Mapping Data" and GB/T18316 "Quality Inspection and Acceptance of Digital Surveying and Mapping Data", we refine the inspection content and methods of the data quality inspection. At the same time, we also stipulated the quality elements, weight division, quality judgment and other aspects. The inspection content and technical requirements meet the requirements of DB11/T 316-2015 "Technical Regulations for Underground Pipeline Detection".

The combination of manual verification analysis and program comparison analysis is the main method of data quality inspection. In the overall inspection, the method of manual verification and analysis is used to overall inspect the compliance of the data; the program software is used to comprehensively inspect the pipeline data; the inspection software provided by the owner is used to comprehensively inspect the pipeline data. For the thematic inspection, in-house inspections are mainly carried out by combining program inspections and human-computer interaction inspections to verify and analyze the pipeline's characterization quality, attribute accuracy, logical consistency, and data quality. Field inspections mainly use repeated measurement methods to compare and analyze the plane accuracy, elevation accuracy and exploration accuracy of pipeline results, and use on-site inspection methods to verify and analyze pipeline integrity, geographic expression, and correctness of attribute content.

4.3 Evaluation result

The data is divided into three areas according to administrative regions.

The inspection sample is determined by not less than 5% of the 1:500 standard framing pipeline diagram.

Area	Area 1	Area 2	Area 3
Control measurement (0.2)	100	100	88.8
Position accuracy (0.15)	92	89.9	86.8
Characterization quality (0.2)	73.5	70.8	90.8
Attribute accuracy (0.2)	74.4	77.6	86
Logical consistency (0.15)	80.7	80.4	94.7
Data quality (0.1)	78.6	78.6	87.6
weighted score	83.37	83.1	89.1

Table 3. Result of quality inspection

After inspection, no unqualified products were detected in the sample of the pipeline survey data in this project. The average sample quality score was 89.11 points, and the sample quality grade was good. After comprehensive judgment, the pipeline survey data quality was batch

qualified.

For example, one of data item in Area 1 was chosen to demonstrate the data inspection. Some of map sheets was chosen, meanwhile, the vertical axis is the score of inspection.

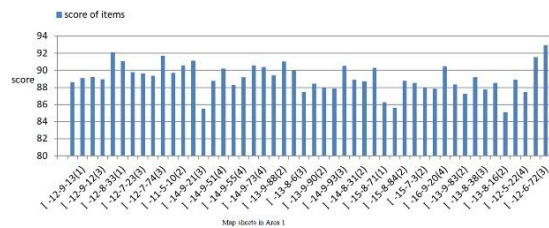


Figure 2. Example for data item inspection in Area 1

4.2 Quality problems

In the actual inspection work, some common quality problems were summarized in different inspection stages. These include:

(1) According to the inspection of control survey results, the interval between points measured by GNSS RTK is less than 60s. The difference between the plane and the elevation between the point measurement rounds exceeds the limitation.

(2) Main quality issues of sample data. The buried depth of the hidden pipe point is inconsistent in the two directions. The location of the pipeline point is inconsistent with the field. The attribute of the pipeline point is incorrectly filled. The accuracy level of pipeline point exploration is filled in incorrectly. Pipeline attachments and buildings (structures) are missing. Number of holes occupied by communication and power. The pipeline material is inconsistent with the actual site. Drainage and water supply pipe diameters are inconsistent with those on site

(3) The main quality problem of non-sample data. There is an attribute item definition error. There is a false pseudo node. There are different types of pipeline collisions. When individual attachments are wells, the required well information is missing. There are very short lines and two-point pipelines exceeding 75 meters. The outline of the auxiliary line layer is crossed and not closed. The drainage flow direction is inconsistent with the data vector direction. The direction of the pipeline points entering and leaving the measurement area is not remarked. Small rooms that meet the size requirements are not shown.

4.2 Potential problems analysis

Facing the existing data quality problems, we interviewed the relevant staff and got an in-depth understanding of their data production process. The following potential reasons are summarized for reference when the data production department revises the data.

(1) This is because the operator does not understand enough production regulations, such as attribute definition errors.

(2) In the process of data collection, the operation of the staff is not standardized, such as obvious pipeline point plane position and gross error in elevation, etc.

(3) The operators are not careful in their work, such as

the direction of the pipeline points entering and exiting the measurement area without remarks, and the small rooms meeting the size requirements are not shown.

5. CONCLUSION

Underground pipelines are the material basis for the survival and development of cities, and are known as the "lifelines" of cities. Urban underground pipeline survey and quality management is an important basic work of urban planning, construction and management. Data quality is crucial for the reliability and usage of underground pipeline data in urban planning and other smart cities projections. We conducted a praxis work on underground pipeline data quality evaluation. The whole dataset includes about 15 GB size vector data, covers 3 counties and 220 map sheets, worked with 3 teams. First, the overview data quality evaluation was executed. In order to evaluate data itself quality, hierarchy data sampling method was engaged to extract part from dataset. The detail inspection work was done on the sample dataset to get the score for each element. Then the item score was calculated by weighted algebraic algorithm. The paper work is the praxis work. For future work, we will put emphasis on the weight determination method research work in order to keep more precise data quality.

ACKNOWLEDGEMENTS

The authors would like to thank the valuable comments from anonymous reviewers.

REFERENCES

- Eskandari, M., Omidvar, B., Modiri, M., Nekooie, M.A., Alesheikh, A.A., 2017. Geospatial analysis of earthquake damage probability of water pipelines due to multi-hazard failure. ISPRS Int. J. Geo-Information 6. doi.org/10.3390/ijgi6060169
- Huang, J., Jing, C., Fu, J., Huang, Z., 2019. Uncertainty Analysis of Rainfall Spatial Interpolation in Urban Small Area. Springer, Cham, pp. 79–95. doi.org/10.1007/978-3-030-12971-2_5
- Jing, C., Du, M., Li, S., Liu, S., 2019. Geospatial dashboards for monitoring smart city performance. Sustainability 11, 5648. doi.org/10.3390/su11205648
- Kalogianni, E., van Oosterom, P., Dimopoulou, E., Lemmen, C., 2020. 3D land administration: A review and a future vision in the context of the spatial development lifecycle. ISPRS Int. J. Geo-Information 9. doi.org/10.3390/ijgi9020107
- Kilkenny, M.F., Robinson, K.M., 2018. Data quality: "Garbage in – garbage out." Heal. Inf. Manag. J. 47, 103–105. doi.org/10.1177/1833358318774357
- Lagüela, S., Solla, M., Puente, I., Prego, F.J., 2018. Joint use of GPR, IRT and TLS techniques for the integral damage detection in paving. Constr. Build. Mater. 174, 749–760. doi.org/10.1016/j.conbuildmat.2018.04.159

Li, S., Cheng, C., Pu, G., Chen, B., 2019. Qra-Grid: Quantitative risk analysis and grid-based pre-warning model for urban natural gas pipeline. *ISPRS Int. J. Geo-Information* 8. doi.org/10.3390/ijgi8030122

Li, W., Han, Y., Liu, Y., Zhu, C., Ren, Y., Wang, Y., Chen, G., 2018. Real-time location-based rendering of urban underground pipelines. *ISPRS Int. J. Geo-Information* 7. doi.org/10.3390/ijgi7010032

Najafabadi, M.M., Villanustre, F., Khoshgofaar, T.M., Seliya, N., Wald, R., Muharemagic, E., 2015. Deep learning applications and challenges in big data analytics. *J. Big Data* 2, 1–21. doi.org/10.1186/s40537-014-0007-7

Wai-Lok Lai, W., Dérobert, X., Annan, P., 2018. A review of Ground Penetrating Radar application in civil engineering: A 30-year journey from Locating and Testing to Imaging and Diagnosis. *NDT E Int.* 96, 58–78. doi.org/10.1016/j.ndteint.2017.04.002

Wang, S., Guo, Q., Xu, X., Xie, Y., 2019. A Study on a Matching Algorithm for Urban Underground Pipelines. *ISPRS Int. J. Geo-Information* 8, 352. doi.org/10.3390/ijgi8080352

Yan, J., Jaw, S.W., Soon, K.H., Wieser, A., Schrotter, G., 2019. Towards an underground utilities 3D data model for land administration. *Remote Sens.* 11, 1–21. doi.org/10.3390/rs11171957