# A FULL LIFE DATA QUALITY WORKFLOW RESEARCH AND PROJECT PRACTICE

Haitao Zhao [1], Wenchao Gao [1,*], Changfeng Jing [2], Xuefang Li [2]

[1] National Quality Inspection and Testing Center for Surveying and Mapping Products, 28 West Lianhuachi Road, Beijing, China
[2] School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, 1 Zhanlanlu Road, Beijing, China

**KEY WORDS:** Data Quality, Full Life Workflow, Geospatial Data, Work Organization.

**ABSTRACT:**

With the development of technology and geospatial equipment, more and more work, particularly covering a big area survey and mapping work, are conducted with different worker and equipment, or heterogeneous time variation. With the traditional data quality focus on the geospatial data quality itself, which is not include the work files and organization. Facing the heterogeneous characteristics of geospatial data and work organization, a full life data quality workflow was proposed to manage and control geospatial data quality. The proposed workflow is extended to include the file preparation, quality evaluation, and quality calculation phrases. It is demonstrated in one real data quality project, which include vector data, raster data and other geospatial data covering 160 thousands square kilometers and 300 work zones finished by 8 teams. The usability and reliability were validated in our work.

## 1. INTRODUCTION

Data quality is crucial for the reliability and usage of data, which provides benefit for many application and services such as urban planning, land sources and natural disaster management(Basiri et al., 2019; Birigazzi et al., 2019; Kitchin and McArdle, 2016; Li et al., 2020). 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. The volunteered geographic information (VGI) data and government authorized data are facing this challenge. Since the born of VGI (Goodchild, 2007; Zhang and Zhou, 2018; Yamashita and Seto et al. 2019), as a crowdsourcing collected data, it impended the academic research and business. In the VGI research field, the data quality is a buzz word. Many works dedicated to this research topic, such as earth system science data quality management and control(Sun and Wang, 2010), reviewing on measures and indicators of VGI data quality(Antoniou and Skopeliti, 2015; Fonte and Antonio et al. 2017; Wu and Clarke et al. 2020) and challenge study on crowdsourced geospatial data quality(Basiri et al., 2019; Adhinugraha and Rahayu et al. 2020). These work focused on the data overall or relative data quality. The word "uncertainty" is well proper to descript its quality. That is, the different data accuracy may be allowed for neighbour spatial objects or the same area.

However, for the government authorized geospatial data, the absolute data quality is very important. Meanwhile, the same quality indicators are necessary thing for the whole data even if collected with different worker and equipment, or heterogeneous time variation. Therefore, for one geospatial data quality evaluation work covering a big area such as one or two states, how to manage and control data quality is a hard work in practice.

With the sharply development of information communication technology and internet application, the usage of geospatial data has been changed increasingly. Excepting for the professional users, it may be used by personal users for their interests with non-professional scenario. It is also possible used in many professional or non-professional software (Elshaw Thrall and Thrall, 1999; Goodchild, 1995; Devillers 2002). Geospatial data can provide services for decision-making and other system. Consequently, its quality is very crucial.

Generally, the data quality evaluation work is more like a working project without academic topic. In fact, how to organize the all procedures of work, data sampling method and data indicators are all including the research value. Therefore, in our work, facing the heterogeneous massive geospatial data collected by different workers and equipment at various time, a full life data evaluation workflow was proposed, and a praxis in one project was demonstrated in our work.

The remainder of this paper is structured as follows. In Section 2, the objectives of our work are introduced. Section 3 presents our contributions, a full life geospatial data evaluation workflow. The research application was shown in Section 4. Finally, Section 5 shows the simple result, furthermore provides some concluding remarks and proposes some future work.

## 2. OBJECTIVES

The objective of this paper is to provide a full life data evaluation workflow that support massive government authorized data quality control and management. This goal is

* Corresponding author
E-mail addresses: 397595381@qq.com (Wenchao Gao)

achieved based on the work preparing, key technologies validation, and the data itself quality evaluation. Finally, by a praxis on one project, the usability and reliability of this workflow was demonstrated.

## 3. FULL LIFE DATA EVALUATION WORKFLOW AND PRACTICE

### 3.1 Concept of Workflow

The proposed workflow under the guideline of data quality principle, "two-level data inspection, one-level data acceptance". For the two-level inspection, which is carried in the data procedure stages by work unit and department. This work is a kind of self-check by some criteria. The one-level acceptance is a third-party inspection of their all work, including geospatial, the organization of data collection, and the key technologies validation work. Our work is belonging to this one. From aforementioned description, we can see this work is a wide concept outside the geospatial data itself.
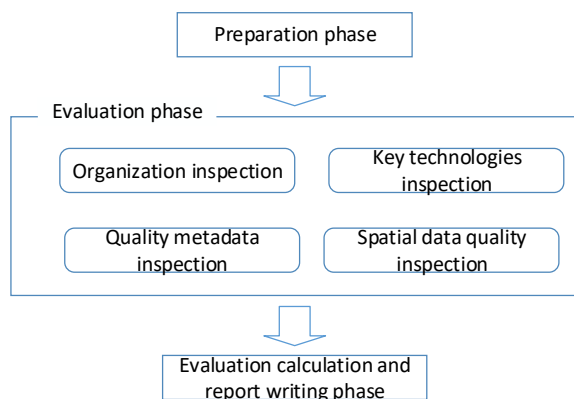


**Figure 1**. the data quality evaluation workflow

Figure 1 gives the proposed full life workflow. The workflow includes three phases as shown in Figure 1. For the **preparation** phase, the documents and files about the evaluation geospatial data were collected, about including project files and technological documents. The **evaluation** phase is the most important work, which covers all business work of evaluation. The data sampling, data accuracy calculation and other related works were conducted. The third phase is the data quality synthetic calculation and report writing work. In where, the calculation of data quality is almost based on the weighted calculated. For the weight of each element is decided by data type and experts' experience.

This assessment workflow is applied to a major provincial engineering project, including 1 project management unit, 1 project lead unit, and 8 production units. The data results include the surveying and mapping benchmark system (the buried and observed GNSS observation points, the second-class levelling route with a length of about 2000 kilometres), the digital Orth-image (about 160,000 square kilometres), and the digital elevation model (about 160,000 square kilometres). Meters), terrain element data (about 300 survey areas).

### 3.2 Preparation

The evaluation preparation work includes three aspects: data collection, personnel organization and plan formulation.

(1) Data collection is the work of sorting out data sources for quality inspection. The required data includes, but is not limited to, organizational management documents, technical regulations, and quality management regulations. Based on these data, the inspectors can quickly understand the project background, progress, and related results to facilitate subsequent project evaluation.

(2) The determination of the technical plan is the core content of the preparation for quality inspection. The specific evaluation content and evaluation method should be included in the technical plan. For a certain inspection item, it is necessary to design the corresponding evaluation indicator and construct a table to record the inspection result.

(3) Personnel organization is the establishment of related work groups or institutions for quality inspection. The needs of personnel are mainly based on technical solutions to achieve a perfect match between personnel and work. After determining the personnel, carrying out targeted personnel training is also one of the very important tasks in personnel organization.

### 3.3 Organization Inspection

The purpose of organizational management evaluation is to investigate the organization and operation of the project. This work is mainly based on the organization and management documents collected in the early stage, related rules and regulations, task arrangements, equipment maintenance and other aspects. The method of work development is mainly to refine the specific content of the inspection items, establish evaluation indicators, and comprehensively analyze the project organization and management model and the operating mechanism of the organization by means of quantitative scoring. Through the above steps, we can evaluate the scientific, rationality and applicability of organization management.

A table is given below to show some of the specific content and indicators of the organizational structure inspection.

| Inspecting Item | Item detail information and some indicators |
|---|---|
| Organization | 1. Whether an organization and management institution with clear responsibilities has been established; 2. Whether the management organization is operating effectively and performing its duties due diligence; 3. Whether the project organization unit has implemented its responsibilities in terms of implementation plan formulation, technical design preparation, quality management, results acceptance, results submission, and data archiving. |
| Management guidelines | 1 Whether it has formulated management rules and regulations involving project management, technical management, safety production management, quality management, results inspection and acceptance, results submission, and data archiving; 2. Whether the established management rules and regulations can be applied to the needs of management during project implementation, and provide management |

| | |
|---|---|
| | system guarantee for the orderly and standardized implementation of the project. |
| Competitive mechanism | 1. Whether a competition mechanism has been introduced in the task implementation; 2. Whether the contract has been signed with the bid-winning unit through government procurement methods such as bidding and bidding, specifying responsibilities, rights, and benefits. |
| Logistics support | Logistics support 1. Whether the equipment can meet the production needs of domestic and foreign industries; 2. Whether to carry out the calibration and appraisal of the equipment according to the requirements. |
| Education and training | 1. Whether the project participants have carried out training related to professional technology, production safety, confidentiality education, etc.; 2. Whether the production personnel have passed the relevant assessments, and have the qualifications and ability to work. |

**Table 1**. Organization inspection item

## 3.4 Key Technologies Inspection

Technical management evaluation is based on technical management documents, project design documents, professional design documents, data regulations, technical standards, and new process development records. Among them, key technology research and application demonstration reports are key documents, because these represent the advancement of technology and the applicability of applications. The purpose of this evaluation is to investigate the innovation and operation of technology, evaluate the scientific level of technology management, and evaluate the applicability of new technologies.

The technical evaluation work includes:
(1) Check whether the project has established a technical management system by consulting relevant technical problem management documents and communicating on-site. In the implementation of the project, whether the technical management related work is effectively implemented, and whether the technical problem handling mechanism is effectively operating.

(2) Conduct interviews with production personnel by consulting technical materials such as formulated standards and technical regulations, project design documents, and professional technical design documents to assess whether actual production is carried out in accordance with standards and technical designs, and whether technical standards and technical designs are standardized, Complete, whether the approval procedure is complete.

(3) Evaluate whether the design and production process match, whether the new technology and new equipment are complete. These tasks are generally carried out by consulting technical documents, trial production reports, and on-site inspection of production processes.

(4) Through expert review opinions on the key technology of the project, watch the demonstration on the spot, and check whether the key technology research and promotion and application have been carried out in accordance with the overall design requirements.

The detail contents for technological inspection are shown below in Table 2.

| Items | Content |
|---|---|
| Technology Management System | 1.Whether a technical management and technical problem handling mechanism has been established? 2.If established, whether the technical management and technical problem handling mechanism are operating effectively? 3. Whether the technical design preparation is standardized and complete, and whether the approval procedures are complete? |
| Standards and technical regulations formulation | 1. According to the dataset situation, whether the relevant standards and technical regulations have been formulated? 2. Whether it can guide the data update and quality inspection work of the basic geographic information? |
| Technical process | 1. Has the project design and professional technical design been carried out in accordance with the overall design technical requirements? 2. Does the production process meet the technical design requirements? 3. Is the production process popularized and applied on the basis of new technology tests, formation of technical routes and methods, and trial production verification? |
| key technology research | 1. Has the key technology research been carried out in accordance with the overall design requirements, and is it in line with expectations? 2. Has the key technology been promoted and applied in production? |

**Table 2**. Technical management inspection content

## 3.5 Quality metadata inspection

Quality evaluation is an output evaluation. The quality assessment materials mainly come from the quality management document regulations, the quality inspection process, the approval form for the use of surveying and mapping results, the inspection record form for the results, and the data inspection report. The work is mainly to evaluate the scientific and rationality of the quality management system and the quality characteristics of the data by comparing quality inspection standards with production processes and data products.
This work includes:

(1) Check whether the contents of relevant quality management documents are complete, whether the formulation is

standardized and reasonable, whether a quality assurance system has been established, and whether a quality management operation mechanism has been established.

(2) Check whether the table item settings are complete and reasonable, whether the records are complete, and whether the storage is standardized. This includes the meeting sign-in form, training personnel record form, surveying and mapping results use approval form, quality inspection record form, etc.

(3) Check whether the sampling ratio of inspections at all levels meets the requirements, whether the record content is complete and standardized, whether the quality problem description is clear, whether the score is reasonable, whether the inspection report is standardized, etc., conduct inspections to evaluate whether the quality management system is operating effectively. The detail contents for quality inspection are shown below in Table 3.

| Items | Content |
|---|---|
| Management manuals and other program files | 1. Is the system management-related reference document information accurate, complete and applicable?<br>2. Are the formulated quality policy and quality objectives reasonable?<br>3. Is the quality policy documented?<br>4. Are the countermeasures for risks and opportunities correct and feasible?<br>5.Do the resources of the support system meet the management needs?<br>6. Are the contents of environment, safety (including data confidentiality), occupational health, etc. complete, scientific and applicable?<br>7. Is the quality assessment reference and process quality control standardized? |
| Quality assurance System | 1. Has a quality management and quality issue responsibility mechanism been established?<br>2. Can it guide the effective operation of quality management and quality problem responsibility mechanisms?<br>3. Has a clear quality problem handling mechanism been formed? |
| Quality management operation mechanism | 1. Are the quality level responsibilities clear, without overlap or overlap?<br>2. Is the definition of the two-level inspection and acceptance system clear?<br>3. Are inspection and acceptance responsibilities and objectives clear?<br>4. Is there a mechanism for handling unqualified results? |
| Record form and inspection record | 1. Is the type of record form complete and applicable?<br>2. Is the setting of the record table items complete and reasonable to meet the characteristics of controllable, traceable, and convenient management?<br>3.Does the first-level inspection check ratio |

meet, and are the inspection contents and records complete?
4.Does the secondary inspection inspection ratio comply with the inspection content, whether the inspection content and records are complete, and whether the result quality evaluation and inspection report are standardized?
5. Is the acceptance inspection ratio consistent, whether the inspection content and records are complete, and whether the result quality evaluation and acceptance report are standardized?

**Table 3**. Quality inspection content

### 3.6 Spatial data quality inspection

The quality evaluation of the results adopts a sampling inspection mode. Among the completed results of the project, the samples to be inspected shall be selected and quality inspections shall be carried out in accordance with the relevant standards and regulations of the current quality inspection of the results. At present, the sample extraction work is mainly implemented in accordance with the provisions of GB/T 24356-2009 "Quality Inspection and Acceptance of Surveying and Mapping Results" and GB/T 18316-2008 "Quality Inspection and Acceptance of Digital Surveying and Mapping Results". At the current stage, the result quality assessment work is mainly applied to several achievements such as surveying and mapping datum (geodetic datum, elevation datum), digital orthophoto data, digital elevation model data, and topographic element data.

Quality elements are the key components in data quality inspection. For an instance, the quality elements in geodetic datum are listed in Table 4.

| Quality elements | Sub-element | Inspection item | Method |
|---|---|---|---|
| Data quality | Mathematical accuracy | Accordance of the error of the point position and the relative error of the side length | manual |
| | Observation quality | The rationality of instrument inspection items, observation methods, data and records, etc. | |
| | Calculated quality | The correctness of the starting point, data usage, compliance of various indicators, etc. | |
| Document quality | Finishing quality | Point record and custody procedures, observation handbook, calculation results, technical summary, inspection report and other standardization and regularity | manual |
| | Document integrity | Completeness and completeness of technical summary, inspection report, and submitted materials | manual |

**Table 4**. Quality elements in geodetic datum

## 4. APPLICATION

### 4.1 Dataset

This assessment involves a major provincial engineering project, with 1 project management unit, 1 project lead unit, and 8 production units. The data set involves a surveying and mapping benchmark system (40 GNSS B-level points and 30 C-level points that have been buried and observed, and a second-class level route with a length of about 2000 kilometres that has been observed), digital ortho photos (about 16 10,000 square kilometres), digital elevation model (about 160,000 square kilometres), topographic feature data (about 300 survey areas).

### 4.2 Result of data inspection

In terms of organizational structure, the organization has a sound organization, a sound management system, sophisticated technical equipment, strong logistics support, and significant effects of education and training. This earned an excellent score.

For the inspection of technical management, the data production unit has established a complete technical management system; the content is complete. The data producer has very well constructed standards and technical regulations that are in line with regional characteristics, which provides non-technical guarantees for data quality. The production process developed by the project can meet the needs of basic geographic information data update and application. In addition, the key technology of data production is forward-looking, in line with the overall design requirements of the project, and has been promoted and applied in production. This earned an excellent score.

In the production of the data set, the project party has established a complete quality management system, which effectively guarantees the data quality. The file classification system of the data set is complete, the classification and content are standardized, all kinds of records are complete, and the preservation is standardized. This gets a good score.

In summary, the data quality of this project meets the technical design requirements, and the overall quality of the results is excellent.

## 5. CONCLUSION

In our work, a full life data quality evaluation workflow was proposed. The usability and reliability of the workflow was validated through the praxis on one real project. The whole datasets include vector data, raster data and other geospatial data covering 160 thousands square kilometres and 300 work zones finished by 8 teams' collaboration work. The organization quality is evaluated by the ISO standard and domain specific files. For the data quality calculation, the weighted concept was adopted and the quality of dataset was evaluated. For the reliability of weigh for each element, we will continuous work on it in future.

## ACKNOWLEDGEMENTS

## REFERENCES

Adhinugraha, K., Rahayu, W., Hara, T., & Taniar, D. (2020). Dealing with noise in crowdsourced GPS human trajectory logging data. Concurrency and Computation: Practice and Experience, e6139.

Antoniou, V., Skopeliti, A., 2015. Measures and indicators of vgi quality: An overview. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. 2, 345–351. doi.org/10.5194/isprsannals-II-3-W5-345-2015.

Basiri, A., Haklay, M., Foody, G., Mooney, P., 2019. Crowdsourced geospatial data quality: challenges and future directions. Int. J. Geogr. Inf. Sci. 33, 1588–1593. doi.org/10.1080/13658816.2019.1593422.

Birigazzi, L., Gregoire, T.G., Finegold, Y., Cóndor Golec, R.D., Sandker, M., Donegan, E., Gamarra, J.G.P., 2019. Data quality reporting: Good practice for transparent estimates from forest and land cover surveys. Environ. Sci. Policy 96, 85–94. doi.org/10.1016/j.envsci.2019.02.009.

Devillers, R., Gervais, M., Bédard, Y., & Jeansoulin, R. (2002, March). Spatial data quality: from metadata to quality indicators and contextual end-user manual. In OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management (pp. 21-22).

Elshaw Thrall, S. and Thrall, G. I., 1999, "Desktop GIS software". In Geographical Information Systems, Vol. Vol. 1 - Principles and Technical Issues (Eds, Longley, P. A., Goodchild, M. F., Maguire, D. J. and Rhind, D. W.) John Wiley & Sons, New-York, pp. 331-345.

Fonte, C. C., Antoniou, V., Bastin, L., Estima, J., Arsanjani, J. J., Bayas, J. C. L., ... & Vatseva, R. (2017). Assessing VGI data quality. Mapping and the citizen sensor, 137-163.

Goodchild, M. F., 1995, "Sharing Imperfect Data". In Sharing Geographic Information (Eds, Onsrud, H. J. and Rushton, G.) Rutgers University Press, New Brunswick, NJ, pp. 413-425.

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

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.

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.

Kitchin, R., McArdle, G., 2016. Urban data and city dashboards: Six key issues. Data City 1–21. doi.org/10.17605/OSF.IO/K2EPN.

Li, H., Herfort, B., Huang, W., Zia, M., Zipf, A., 2020. Exploration of OpenStreetMap missing built-up areas using twitter hierarchical clustering and deep learning in Mozambique. ISPRS J. Photogramm. Remote Sens. 166, 41–51. doi.org/10.1016/j.isprsjprs.2020.05.007.

Najafabadi, M.M., Villanustre, F., Khoshgoftaar, 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.

Sun, C., Wang, J., 2010. Study on the data quality management and the data quality control-a case study of the earth system science data sharing project. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch. 38, 208–211.

Wu, H., Lin, A., Clarke, K. C., Shi, W., Cardenas-Tristan, A., & Tu, Z. (2020). A comprehensive quality assessment framework for linear features from Volunteered Geographic Information. International Journal of Geographical Information Science, 1-22.

Yamashita, J., Seto, T., Nishimura, Y., & Iwasaki, N. (2019). VGI contributors' awareness of geographic information quality and its effect on data quality: a case study from Japan. International Journal of Cartography, 5(2-3), 214-224.

Zhang, F., Zhou, B., Liu, L., Liu, Y., Fung, H. H., Lin, H., & Ratti, C. (2018). Measuring human perceptions of a large-scale urban region using machine learning. Landscape and Urban Planning, 180, 148-160.