Earth Sci Inform (2015) 8:919–935
DOI 10.1007/s12145-014-0196-9

SOFTWARE ARTICLE

A general framework for spatial data inspection and assessment

Yiliang Wan · Wenzhong Shi · Lipeng Gao · Pengfei Chen · Yong Hua

Received: 29 August 2014 / Accepted: 8 December 2014 / Published online: 25 January 2015

Abstract The quality aspects of spatial data are very important in the decision-making process. However, the quality inspection of spatial data is still dependent on manual checking, and there is an urgent need to develop an automatic or semi-automatic generic system for spatial data quality inspection. In this paper, we present a general framework that automatically copes with spatial data quality inspection based on various spatial data quality standards and specifications. The framework involves all descriptions of given spatial data, a data quality model characterized by quality elements, scheme batch checking and spatial data quality assessment based on quality control and assessment procedures. It is implemented in Unified Modeling Language with four main sets of classes: data dictionary, quality model, scheme checking and quality assessment. Accordingly, we have designed four structured Extensible Markup Language files for the framework to organize and describe the data dictionary, quality model, scheme check and quality assessment. It is very easy for users to describe the data requirements using the data dictionary file, and to extend the quality elements or check rules using the quality model file. Users can design the specified checks and quality assessment schemes without coding by configuring the scheme check files and quality assessment scheme files. The framework also incorporates a checking tool capable of solving the difficulties inherent in the diversity of spatial data quality standards and specifications. The proposed framework and its implementation, as a quality inspection system, will facilitate automatic multiple spatial data quality inspection and acceptance. As a result, the quality of diversified spatial data can be ensured and improved, which is extremely important in the era of spatial big data.

Keywords Spatial data quality · Inspection · Data assessment · Quality standards

Communicated by: H. A. Babaie

W. Shi (✉)
Joint Research Laboratory on Spatial Information,
The Hong Kong Polytechnic University and Wuhan University,
Wuhan, Hong Kong, China
e-mail: john.wz.shi@polyu.edu.hk

Y. Wan · L. Gao · P. Chen · Y. Hua
School of Remote Sensing and Information Engineering, Wuhan University, Luoyu Road 129, Wuhan, Hubei, China

Introduction

The quality aspects of geospatial data play a key role in promoting the development of geographical techniques and improving practical applications. Quality issues have attracted increasing attention from academics, industries and governments, resulting in increased requirements for the interoperability and integration of different data sources (McGranaghan 1993; Delavar and Devillers 2010; Li et al. 2012). Although many studies have explored the definition, propagation and reporting of spatial data quality, relatively little attention has been paid to its visualization or the assessment of its fitness for use (Guptill and Morrison 1995; Fisher 1997; Veregin 1999; Oort 2006).

Quality checking is currently dependent on manual operations, which are expensive, inefficient and error prone. An inspiring variety of related tools and software have emerged to help data producers and cartographers. Currently, these spatial data quality tools and software systems focus on one quality aspect or on checking specified data. Many studies have identified error propagation analysis as an important quality aspect, and the corresponding tools are being developed (Heuvelink 1993; Forier and Canters 1996; Burrough 2001; Crosetto and Tarantola 2001; Duckham 2002). Gong and Mu (2000) develop a tool for the detection of errors in a spatial database through consistency checking using logical relationships among spatial neighborhoods and attribute data from different sources. The Environmental Systems Research Institute (Esri 2007) provides some tools for inspecting the

topological and geometric errors in geographic information system (GIS) data. There are numerous commercial software programs designed to assess the quality of a specified data type, such as Digital Line Graphic (DLG) and Digital Elevation Model (DEM), or data that satisfy special data specifications. For instance, Wu et al. (2012) design and develop quality checking software for the Second National Land Inventory in China. Chen et al. (2005) implement a DEM quality detection tool based on a spatial index. The data quality indicators for the DLG data from different scales differ due to their varied data specifications and requirements, and there is responding quality software for the data of each scale (Cao and Song 2009; Zeng 2009; Zheng and Wang 2009; Fu 2010). Although these tools and systems can detect specified quality aspects of spatial data and check specified data, they only solve part of the quality problem and are not reusable. An example is the quality checking software for the Second National Land Inventory (Wu et al. 2012) in China, the sole function of which is the integration of all of the quality regulations. Cartographers and data producers must develop a data quality checking system that can detect all of the quality aspects and check all spatial data types.

However, there are many challenges in establishing automatic data quality checking methods and quality checking platforms, such as the difficulties inherent in the range of quality criteria specified for scale data in various projects. Because geospatial data are stored using a very different method from paper mapping, the relationships between map scale and data accuracy differ between geospatial data and paper maps (Goodchild and Gopal 1989; Devillers and Jeansoulin 2006). Accordingly, the quality criteria and model elements (Goodchild and Gopal 1989; NIST 1994; Guptill and Morrison 1995; Shi et al. 2003; Devillers and Jeansoulin 2006; ISO 2013) specified for different scale data vary. Taking the United States Geological Surveying Mapping Standards as examples, absolute horizontal accuracy for 1:1200 geospatial data should be less than 3.33 ft; that is, 166.67 ft for 1:100,000 geospatial data (USGS 1941). Generally, there are 5 to 11 quality elements (Aronoff 1989; Department of Commerce 1992; ISO 2013), such as lineage, position accuracy, attribute accuracy, logical consistency, completeness, etc. For instance, the Spatial Data Transfer Standard (SDTS) (Department of Commerce 1992) addresses five quality elements—lineage, positional accuracy, attribute accuracy, logical consistency and completeness—whereas ISO (2013) addresses those five, plus temporal quality.

We propose a general model and framework for solving the diversity of data specifications and quality standards for different scale data. The model covers a data dictionary, check rules, scheme checking and quality assessment. A set of structured Extensible Markup Language (XML; W3C 2008) files are designed to support the proposed model. We develop a quality checking system based on the model and framework. The system allows users to customize the data dictionary according to the data specifications, check rules, quality model and elements, based on the data quality standard. They can also customize the checking scheme based on checking requirements and assessment scheme flexibly. The flexible and customizable framework provides a possible solution for spatial big data (SBD) (Evans et al. 2014) inspection, which can be very difficult given the tremendous diversity of SBD sources.

The remainder of this paper is organized as follows. Section 2 reviews the literature on spatial data quality control and introduces a general spatial data quality control process. Section 3 presents the design objectives of the proposed system. Section 4 describes the global logical architecture of the system, and the detailed model design and implementation are presented in Section 5. Section 6 illustrates how to use the system to check spatial data. Section 7 reviews the application of the system and conclusions are given in section 8.

Review of general spatial data quality control process

As the most important part of quality assurance (QA), general data quality control (QC) is defined as a system of checks designed to assess and maintain the quality of the inventory being compiled (Whitney et al. 1998; Foken et al. 2005; ISO 2005). Early studies on spatial data quality control have focused on the spatial data quality indices and methods based on error and uncertainty theory. Shi (2008) presented research on quality control for three data and model types: object- and field-based spatial data and DEM regarding positional error; specifically, random positional error. Numerous researchers have contributed to other aspects of spatial data quality control such as attribute error, logical consistency and completeness error (Langran and Chrisman 1988; Shi et al. 2003; Burnicki et al. 2007; Heuvelink et al. 2007). Currently, more attention is being paid to spatial quality control methods as part of quality management systems and user quality requirements. Hegde and Hegde (2007) designed a QC plan based on the data flow in a GIS database update circle. Wu et al. (2010) proposed a tetrahedron quality control model and built a user assessment system to perform real-time quality control for digital chart cartography. Despite these significant contributions of spatial quality control methods to spatial data inspection and assessment, there has been a lack of focus on the spatial data quality control process. All spatial data quality control strategies coexist within the processes of spatial data production, maintenance and application.

The focus of this paper is the general spatial data quality control process within the spatial data production process, named "two checks, one acceptance." It is implemented strictly according to the spatial data quality control standards for spatial data production (ISO 2013; NASMG 1995; NASMG 2008).

Figure 1 illustrates a general spatial data quality control procedure comprising two checking procedures and one acceptance procedure. Only the spatial data that pass the quality test enter the next procedure. The spatial data products that do not pass are returned to their corresponding cartographic staff or department.

(1) Process check: This procedure aims to address the spatial data production link problem by assessing the production level of every cartographic staff member in the data production department. It involves the data producers performing self-checks, then inter-checking each other, followed by data quality inspectors checking department members. Checkers in every sub-procedure check the spatial data products to ensure data quality levels, and they do so manually or by using tools. In this procedure, the checked data are not necessarily the final data products; that is, they can be process data or raw data products, and every cartographic staff member is responsible for the spatial data if the quality test is failed.
(2) Final check: The final check involves the spatial data products of the checking departments, and it is executed by the specified spatial data quality control department. All final spatial data products must be checked completely according to specified quality acceptance standards during this procedure. If the data products fail to pass the quality test, they are returned to the data producer for quality error repair.
(3) Acceptance: The acceptance procedure determines whether the spatial data products can be accepted, and it is usually performed by the first party or a quality assessment outfit specified by the first party. A sampling strategy should be used before the acceptance test to determine the checked objects. After sampling, the checked objects are tested and evaluated by the proposed Spatial Data Quality Inspection and Assessment System (SDQIAS).
(4) Quality checking and assessing: Data producers and users check and evaluate spatial data products according to the specified quality control standard in this procedure, and a quality assessment report is finished automatically or manually according to quality test results. In this report, a comprehensive score is computed, with the spatial data product failing the quality test if the score is less than 60.

Throughout the checking process, the quality checking and assessment procedure is the key step affecting efficiency from process check to final acceptance. The next step is initiated when the spatial data pass through the quality checking and assessment procedure, the core of which is the checking and assessing method. Thus, we have designed and developed the Spatial Data Quality Inspection and Assessment System (SDQIAS), a general model and automatic data quality checking software package that improves productivity while decreasing the high risks induced by manual operation.

Objectives

The objective of the framework is to provide a general method for checking and assessing the spatial data. Through a common data quality control process, the solutions can be used for quality control with various spatial data in GIS. More specifically, the design has the following main functions.

(1) A completely automated system that enables data quality evaluations with various presentations to be generated,

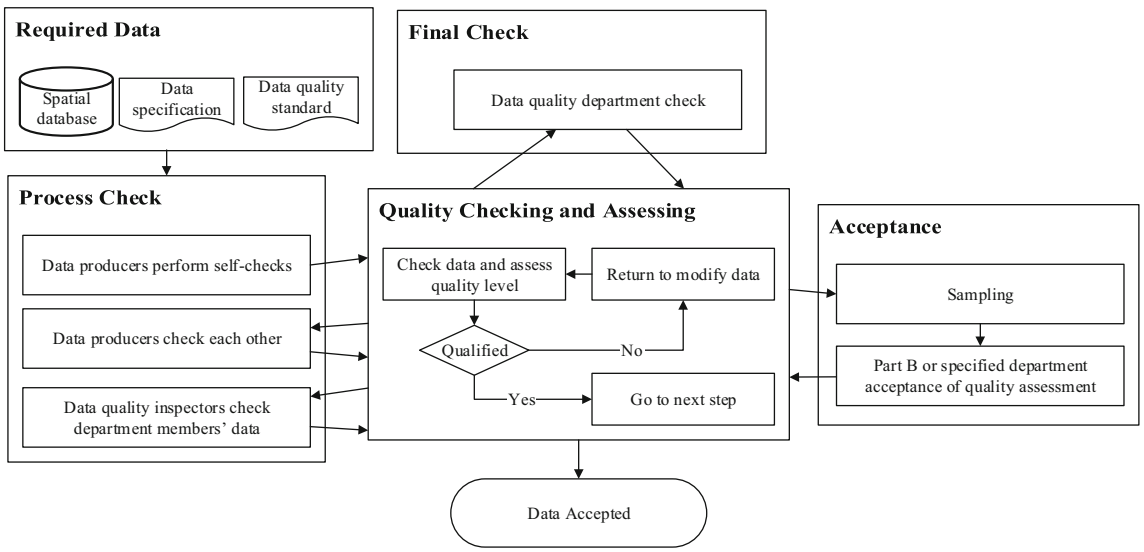


Fig. 1 General spatial data quality control process

satisfying the quality requirements of different applications, especially multi-scale data productions;

(2) Implementation of batch quality checking for spatial datasets collected for the same data source, data requirements or standards and devices;

(3) The ability to update the data quality model according to the newest data quality standards or concrete data quality requirements, and a flexible design capable of managing and updating the quality check rules.

Architecture of the spatial data quality control framework

Figure 2 shows the architecture of the system, which is built around an execution layer that implements all of the functionalities related to spatial data quality inspection and assessment, stores the information in the database and generates the user interface to display results and facilitate interaction. As the foundation of the architecture, the data layer provides the core and support layers with the information needed to build quality models, checking schemes and assessment schemes, which can then be designed in the execution layer. The core and support layers are abstraction layers for designing the quality models, checking schemes and assessment schemes.

The completeness of the data is crucial because it provides the user quality requirements, the information needed to design quality models and checking schemes and the checked spatial data. Three types of materials—geospatial databases, ancillary data including standards and specifications and user requirement documents—are collected in the data layer. Useful information about databases and data quality situations can be acquired from various geospatial database samples. For instance, a designed standard spatial data product can be used to design a checking scheme with checking rules using tools in the execution layer. However, databases are also target checking objects for the execution layer. The ancillary data and user requirement documents are the criteria and references for the quality models, checking schemes and assessment schemes designed in the core and support layers.

Although the quality elements are extracted based on ISO standards in the core of the framework, such as "ISO 19157 2013 Geographic information – Data quality" (ISO 2013), ISO standards still miss some key quality content that is important and compulsory to end-users. Therefore, we have designed the quality element framework by extending ISO standards to include all of the quality requirements proposed in user quality requirements reports, and all national standards such as the United States National Map Accuracy Standards (USGS 1941) and the

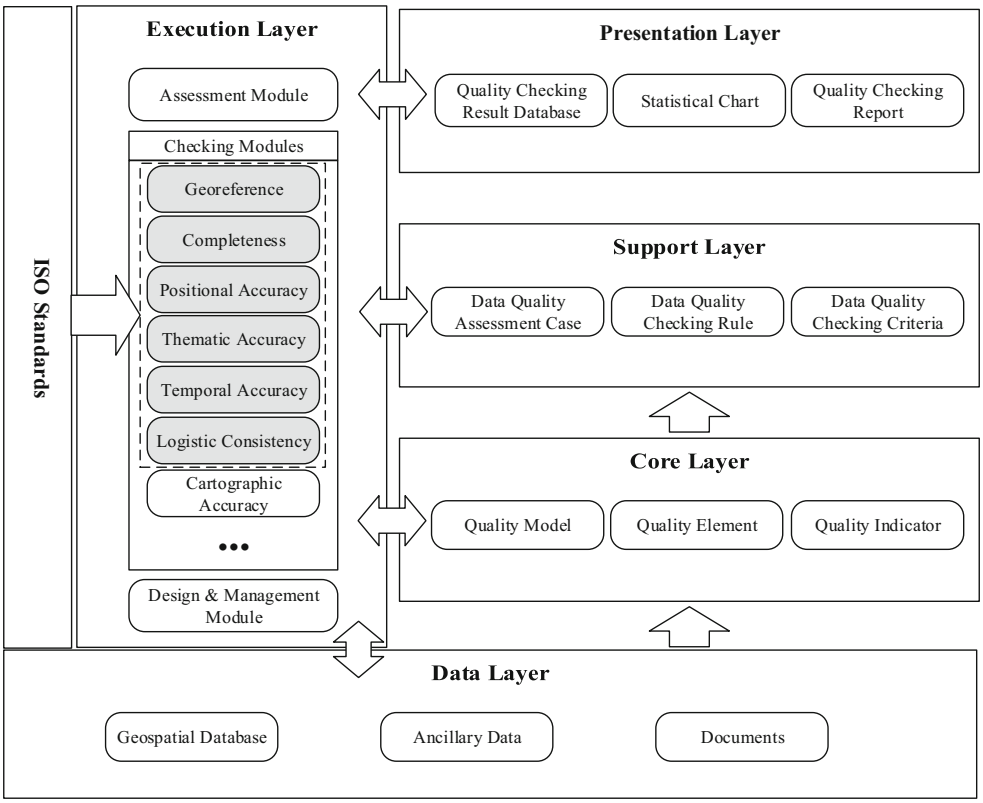


Fig. 2 Logical architecture of the framework

specifications for the inspection and acceptance of digital surveying and mapping achievement quality in China (NASMG 2008). The main quality elements under this framework can comprise georeference accuracy, completeness, positional accuracy, thematic accuracy, temporal accuracy, logistic consistency and cartographic accuracy. A hierarchical structure is established to extract quality elements. Thus, each quality requirement can be abstracted and grouped into one node as a quality indicator in this hierarchy. Finally, a theoretical quality model, which provides the abstract model for the support layer, is proposed in this layer and implemented in the execution layer.

We bring the theoretical achievements of the core layer into practical application in the support layer. There is a one-to-many relationship between a quality element and quality checking rules, between a quality model and quality elements and between a quality indicator and quality checking criteria. One quality indicator is described as a checking rule in this layer. Then, the theoretical quality model is translated into an executable model composed of quality elements, each of which contains one or more checking rules. In addition, this layer builds up an assessment scheme with various assessment cases extracted from spatial data quality assessment standards and specifications in the data layer. All of the checking rules and assessment cases in this layer are implemented in the execution layer.

In the execution layer, the software modules are programmed based on the support layer and organized according to the quality elements extracted from the core layer. The modules in this layer can be divided into three types: quality checking modules, which are the kernel modules in this system, design and management modules, which are used for batch quality checking such as final database quality checking, and assessment modules, which are used to assess and mark the spatial data. The execution layer is a customizable module framework that allows users to extend the functionalities of the system in accordance with quality requirements. The checking modules, which are designed based on the quality element framework in the core layer, cover the quality elements defined in the ISO standards and act as an extension of the 'cartographic accuracy' checking module.

The presentation layer uses various expression methods to display checking results for users, such as the quality checking result database, statistical charts and quality checking reports, which in turn provide users with a variety of GUI interfaces. The quality checking result database, which can be used to conveniently trace and locate bad quality features, is the main medium for storing and displaying quality checking results. By using the quality checking result database, we can also find which quality rule is disobeyed to facilitate error modification. Therefore, cartographers can modify the original database according to the quality checking results. Statistical charts are offered to users who want to view the details of disobeyed quality rules inherent in the database, such as quality distribution in the database.

Implementation

We have designed three subsystems based on the spatial data inspection and assessment processes: configuration maintenance, inspection and assessment. Figure 3 presents a use case diagram of the system, which defines the interactions between users who are named as actors in the Unified Modeling Language (UML) and the system. There are three types of users: spatial data quality control departments, checkers and cartographic staff members. The system provides different functionalities for different user types, but they can all inspect and assess spatial data through schemes. To ensure that a quality error has been corrected, a cartographic staff member must check the specified quality aspect of the spatial data. Only the spatial data quality controlling departments can manage the checking configurations, including data dictionaries, quality models, checking schemes and assessment schemes. The diagram illustrates the relationships among all functionalities. The configuration maintenance system, tasked with managing all of the configurations, is the base of the system and the subsystems are dependent on the system. The inspection functionality in the inspection system depends on the data dictionary and the checking scheme in the configuration maintenance system. Moreover, the checking of spatial data through the scheme is dependent on the checking scheme designed in the configuration maintenance system. All assessment processes must follow the assessment schemes in the configuration maintenance system.

As a system-level sequence diagram, Fig. 4 illustrates how users implement SDQIAS based on the framework for the inspecting and assessing spatial data, and the order of the processes and objects. First, the specified spatial data quality control department must prepare, including making the data dictionary and designing the quality model, checking schemes and assessment schemes. Then, all of the configuration files are sent to cartographic staff members and checkers. The SDQIAS provides some templates, designed according to standards, such as the data dictionary template for the Chinese national standards (NASMG 1994; NASMG 2007). When all of the configurations are ready, the SDQIAS can inspect the loaded spatial data using the specified checking scheme. In this process, each check item in the scheme is run, and the data dictionary is acquired as the spatial data specification and the inspection references in some check items, especially on attribute accuracy. The cartographic staff member can check the quality aspect to ensure the correctness of one quality error. The assessment process is followed by the

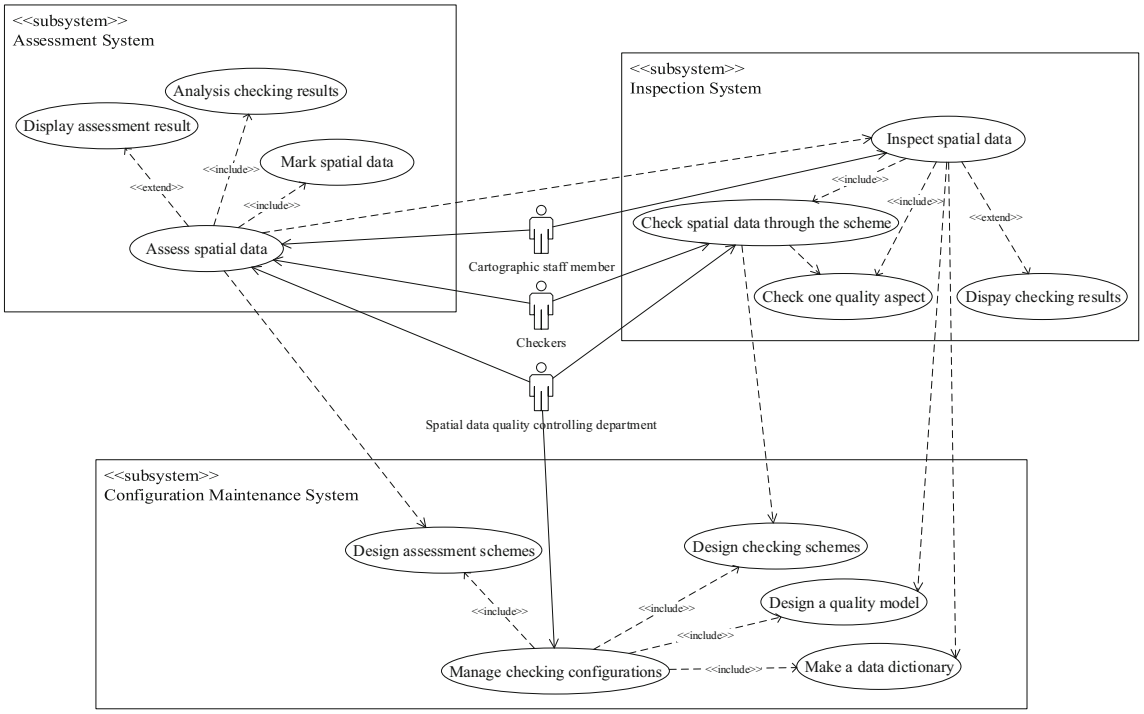


Fig. 3 Use case diagram of the framework

scheme inspection, which returns assessment results and the spatial data quality mark as a spatial data quality level.

The core design of the general framework comprises the following parts: a data dictionary, check rules, quality model, checking scheme and quality assessment. Figure 5 illustrates the overall UML diagram, which includes four packages, each of which solves one key problem of the general model. The detailed design and implementation of each package follows Fig. 5.

Data dictionary

The data dictionary describes the data in a database, and many related studies, publications and standards were proposed in the 1980s. Skidmore (2002) noted that some of the information about spatial accuracy in data dictionaries can be used to assess the accuracy of spatial data. In the SDQIAS, a data dictionary is defined as the description of a GIS data layer that includes the general description of the layer, the definitions of the fields and the range descriptions and geometric feature classifications based on data standards and specifications, such as some of the Chinese national standards (NASMG 1994; NASMG 2007). The designed data dictionary is the reference for all attribute accuracy check items, so it is very important that it be kept in accordance with the standards and data requirements.

The SDQIAS provides a very convenient and flexible editing tool for the data dictionary that has been used to design some data dictionary models according to the Chinese national standards. As is illustrated in the DataDictionary package (Fig. 5), the core main model is based on the general structure of a GIS data layer. The UML diagram includes IDataIO, IDataDictionary, ILayerInfo, IFieldInfo, IFeatureInfo, LayerInfo, FieldInfo, FeatureInfo and DataDictionaryXMLIO. The interfaces, including IDataIO, ILayerInfo, IFieldInfo and IFeatureInfo, define the attributes and functions of class objects. The IDataDictionary interface, which is the core interface of the design, defines all of the information including the data dictionary file path and name, all of the information on the layers and the feature code field name. The class DataDictionary implements the IDataDictionary. Each instance of the IDataDictionary interface is composed of one or more instances of the ILayerInfo interface. The relationship between the IFieldInfo, ILayerInfo and IFeatureCode interfaces is one-to-many, in which one instance of the ILayerInfo interface is composed of many instances of the IFeatureCode and IFieldInfo interfaces. The IDataDictionaryIO interface is used to write into or load from a data dictionary file. Each class, including DataDictionaryXMLIO, DataDictionary, LayerInfor, FieldInfo and FeatureInfo, implements the corresponding interface.

All of the information in a data dictionary is organized in XML, a markup language that defines a set of rules for encoding documents in a format that is both human- and machine-readable (W3C 2008). The dictionary file can be easily browsed using Notepad or an Internet browser. The dictionary editor provides user-friendly graphical interfaces through which to view and edit the data dictionary.

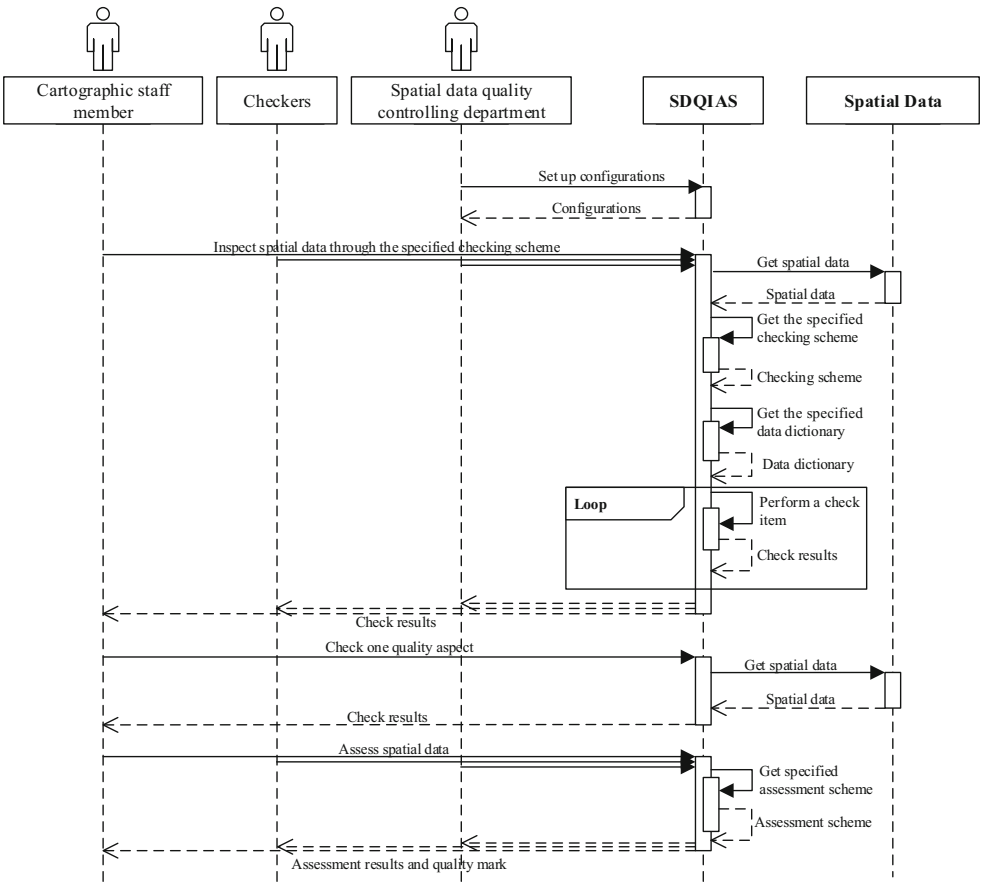


Fig. 4 System-level sequence diagram of the framework

Figure 6 shows the sequence diagram between the data dictionary editor and the main components, including the designed data dictionary interfaces and data dictionary file. First, the data dictionary editor can read the information in a data dictionary file through an instance of the IDataDictionaryIO interface, and return an instance of the IDataDictionary interface, which includes all of the information for the spatial data layers. The design of a data dictionary involves an iterative process in which layer information is created to compose another iterative process by which field and feature information is added to the layer. Then, the data dictionary, which includes all of the information from all of the layers, is returned. Subsequently, the data dictionary can save the designed data dictionary in a specified data dictionary file through the IDataDictionaryIO interface.

Check rules and quality model

The check rules, which are analyzed from the quality model, are the base in the SDQIAS. The key to a check rule is how to check its quality. In the CheckRules package (Fig. 5), the abstract class AbstractCheckItem is the super class of all check classes, which form the core of the check rules, and each instance of concrete derived check classes is related to only one instance of concrete derived check command, which inherits the abstract class AbstractCheckCommand. The abstract class AbstractCheckItem provides some basic information for a check rule that corresponds with the description of check rule class CheckItemDescriptor. One instance of a concrete derived class of AbstractCommand links the user interface with one instance of concrete derived check classes, and how the command runs in a batch scheme checking procedure is different from how it runs in a single check item. The check result for each instance of a concrete derived class of AbstractCheckItem is stored as part of a list in the IDLGQualityError interface and immediately displayed in the user interface.

One spatial data quality model comprises several quality elements that contain additional quality elements, each of which has one or more quality indicators and checking methods. The instance of various concrete derived classes inherited from the AbstractCheckItem class can implement the computation of the quality indicator and run the corresponding checking method. However, indicators and checking methods are not enough for a check rule. Some considerations, including how to schedule various check rules, how to identify one check rule from others and how to instantiate a check rule as a check item from a checking scheme file, must be involved in check rule design.

We have designed a static class CheckItemService and its related classes to solve these problems. The framework of the

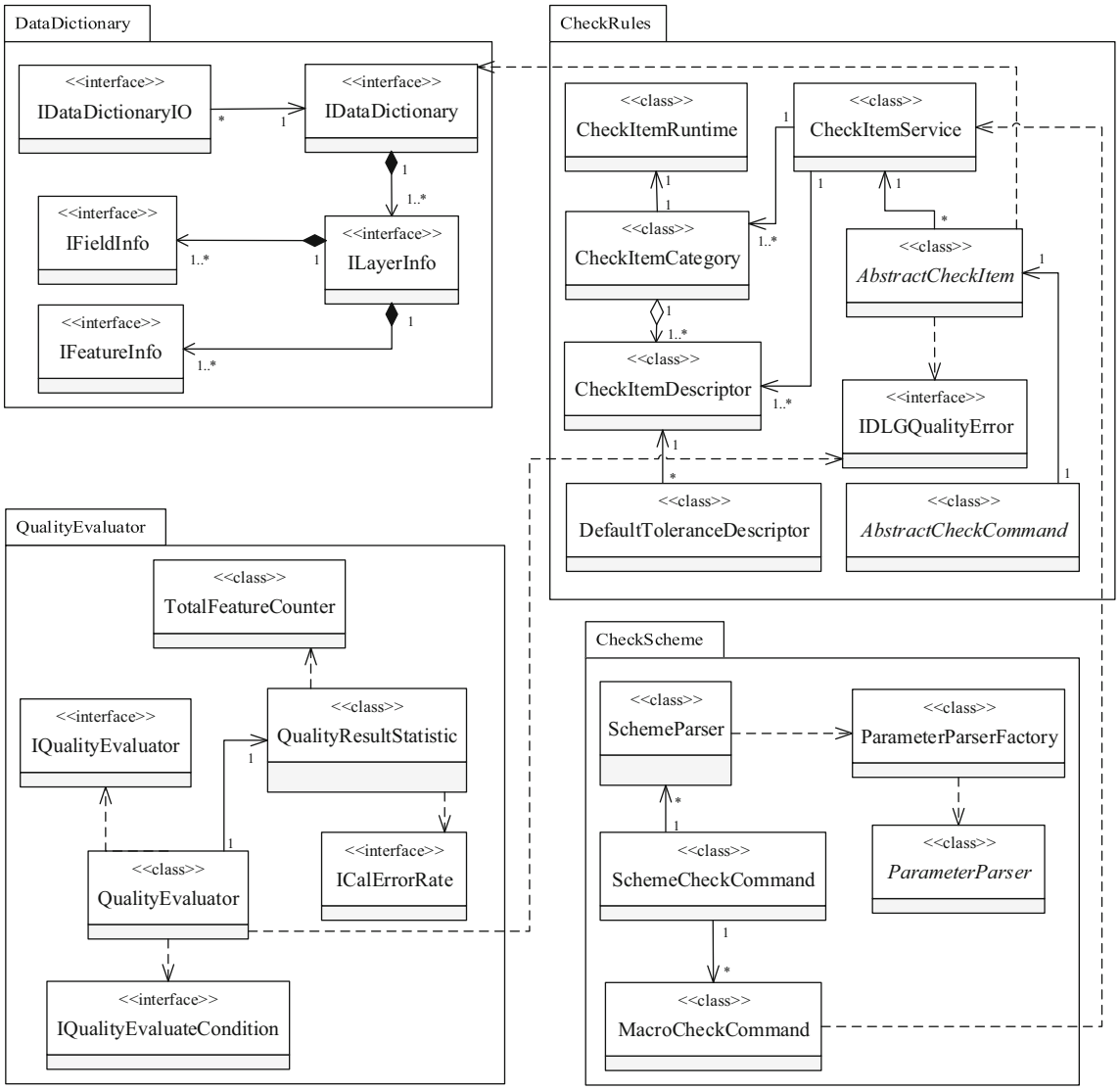


Fig. 5 Core modules of the general model

static class CheckItemService is based on the general architecture of a spatial quality model. The core classes of the design include CheckItemService, CheckItemCategory and CheckItemDescriptor. The CheckItemDescriptor class stores all of the necessary information about a check rule including name, priority, tolerance, checking class and command. Each instance of the CheckItemDescriptor class has a priority property to control its check sequence in a checking scheme. The check rule whose priority value is smaller would be run preferentially. The name attribute uniquely identifies the check rule. The CheckItemCategory class, which relates to one or more check rules, stores the information of one quality element in a quality model and can instantiate a CheckItemCommand or CheckItem class from a text description using reflection technology (Malenfant et al. 1996) that allows for the instantiation of new objects and the invocation of methods in object oriented programming (OOP) (De Champeaux et al. 1993; Beckert et al. 2007). The reflection runtime library sources are a list of instances of the CheckItemRuntime class that inherit the Runtime class, resulting from the one-to-many relationship between CheckItemCategory and CheckItemRuntime.

The data quality model may be different due to the diverse standards of different data products and scale data. Table 1 presents all of the quality elements and check rules for the land cover data products in the First General Survey Achievements of Geographic Conditions of China (FGSAGCC) (NMPQITC 2013), which are used as inputs in a variety of urban management and environmental studies applications (Wu et al. 2014). The SDQIAS includes 62 check rules and 6 quality models for different spatial data, including different scales for digital line graphics (DLG), metadata and remote sensed imagery interpretation samples. The design of the check rules and quality models is open and extensible. Users can create and develop new check rules under the framework and then customize a proper quality model for the spatial data product, which is then integrated into the SDQIAS for future applications. To customize and alter

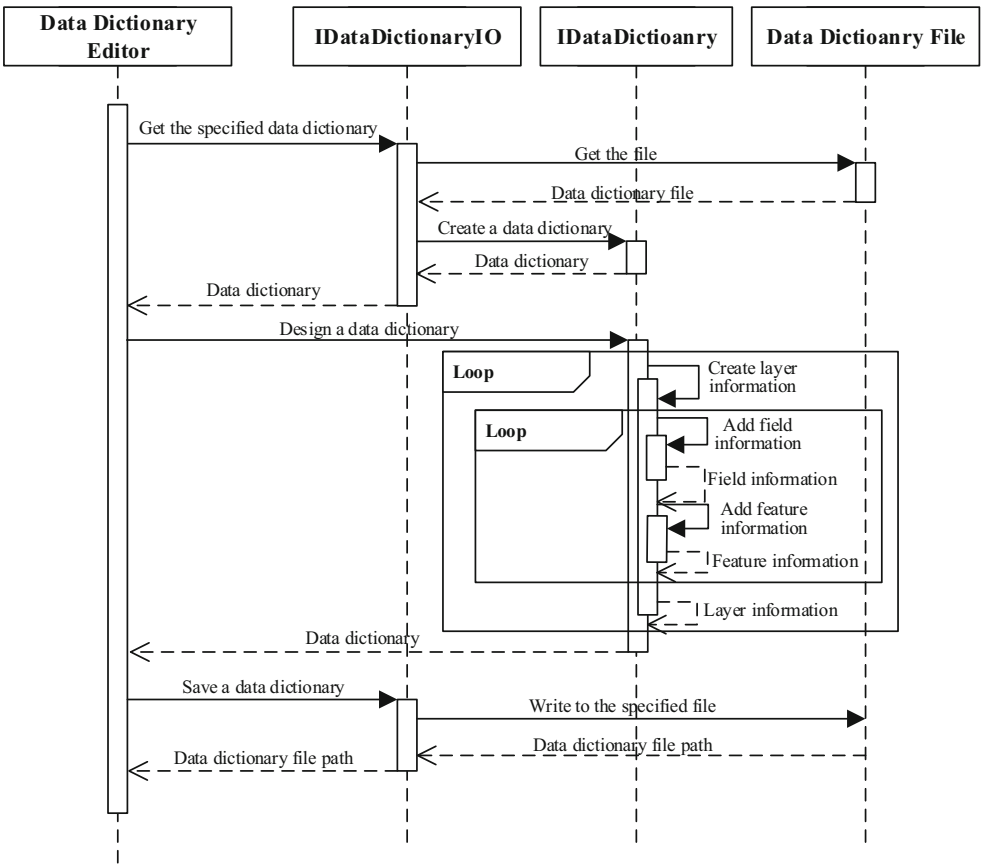


Fig. 6 Sequence diagram of data dictionary design

diverse spatial data quality models, a structured XML format file is designed to record and organize quality elements and check rules. In the XML file, each category node is one quality element and each check rule node is one check rule in one category node. The file structure is as follows:

```
<QualityModel>
<Category name="*" text="*">
<Import hintpath="*" assembly="*" />
…
<!-break line->
<Rule name="*" text="*" class="*" command="*" priority="*" description="*">
<Parameter name="*" type="*" description="*"/>
…
<!-break line->
<Tolerance name="*" text="*" value="*" description="*"/>
…
<!-Check Item->
</Rule>
…
</Category>
…
</QualityModel>
```

This structure is fairly rich in information, and users can extend any quality element node in a quality model and check rule node in an existing quality element node. The attributes of the rule node (Table 2) store some basic information of a check rule. We have designed two types of child nodes in which a rule node can store parameters and tolerance information. In the checking scheme part, the ParameterParserFactory class uses the type attribute of the parameter node to create a concrete ParameterParser class.

Checking scheme design

The core of the scheme classes comprises the MacroCheckCommand, SchemeCheckCommand and SchemeParser classes. The SchemeCheckCommand class links the SchemeParser and MacroCheckCommand classes in a one-to-many relationship. The SchemeParser class is dependent on the CheckItemCategory class to create one check rule object, and dependent on the ParameterParserFactory class to create the parameter values of one check rule. The abstract class ParameterParser is the super class of the LayerListParaParser, ValueParaParser, EnumParaParser and LayerParaParser classes.

Table 1 The data quality model for land cover data in the FGSAGCC

Quality element	Check rule	Description
GeoReference	Geographic coordinate system check	Check whether geographic coordinate system satisfies requirements.
	Vertical coordinate system check	Check whether vertical coordinate system satisfies requirements.
	Projected coordinate system check	Check whether projected coordinate system satisfies requirements.
Temporal accuracy	Currency check of the original data	Check the currency of the image, geographic information and special industry data.
	Currency check of the data product	Check the currency of the data product.
Logical consistency	Attribute item check	Check whether the definition of each field satisfies the requirements, such as name, data type, length and the order.
	Dataset check	Check whether the definitions of map layers satisfy the requirements.
	Format check	Check whether the data file format satisfies the requirements.
	File integrity check	Check whether the data files lack some files.
	File name check	Check whether the file names satisfy the requirements.
	Polygon gap check	Check whether there are gaps between polygons.
	Polygon overlap check	Check whether there are overlaps between polygons.
	Consecutive polygon check	Check errors in polygons whose positions are adjacent and attributes are the same.
Collection precision	Geometric displacement check	Check whether the fitness between polygon boundaries and orthophoto overlap.
	Vector edge match	Check whether the vector edge matches overlap.
Classification accuracy	Classification correctness check Integrity check	Check the classification correctness referring to orthophoto, survey and interpretation data.
Map representation	Abnormal geometry check	Check the abnormal geometry, such as unreasonably small polygons.

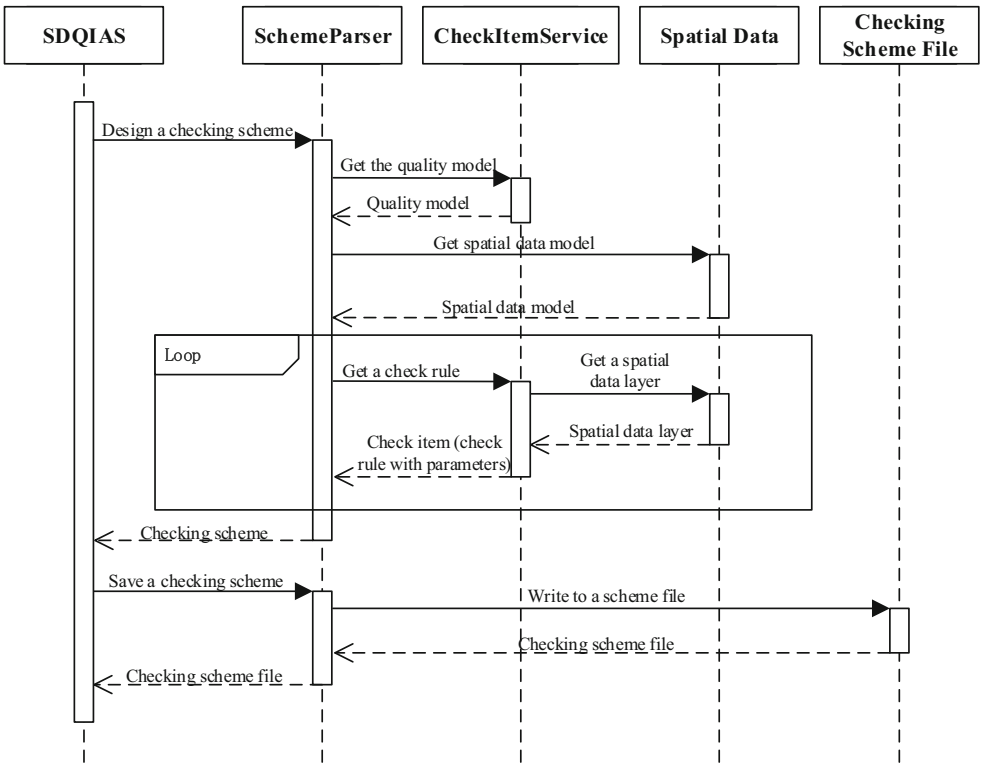
The design of the scheme class framework involves two patterns: command and factory. In the whole system, each check rule is designed as a check command and a check class, so one checking scheme is like a set of check commands with initialized parameters. The instance of the MacroCheckCommand class is a special command that runs a list of commands. The factory method pattern is applied in the design of parameter parsers. The ParameterParserFactory class creates a subclass instance of the abstract class ParameterParser according to the parameter type, such as ILayer, List (ILayer), string, int, double, etc.

All of the check items are written into a formatted XML file that serves as the scheme file. Each check item in one checking scheme includes the name of the check rule, which is used to inspect the data quality problems and parameter values. The check rule and parameter values are instantiated by the CheckItemCategory class and subclasses of the ParameterParser in the SchemeParser class. The SDQIAS provides a scheme manger to manage and configure the checking schemes.

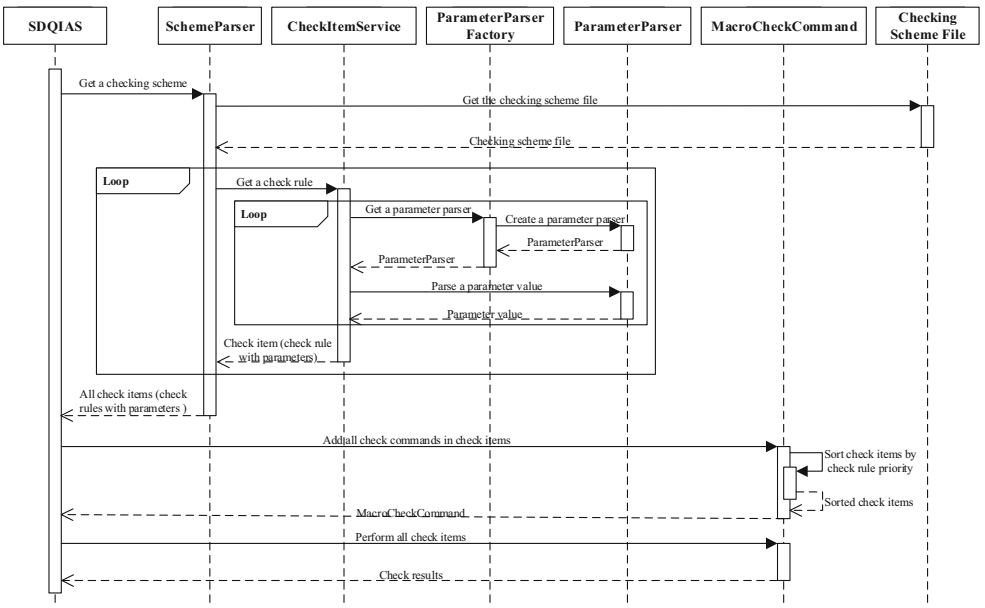
During the scheme configuration process, the SDQIAS connects the necessary check rules and parameters, including the spatial data model, as check items in the checking scheme. Figure 7a shows the process for designing a checking scheme. The SDQIAS provides the specified quality model with the necessary check rules and all parameter information, including the spatial data model and the data dictionary for a new checking scheme. Then, the SchemeParser class gets a check rule from the quality model through the CheckItemService class and sets the proper parameter values, such as the spatial

Table 2 Attributes of the rule node in a quality model file

Attribute	Description
Name	The name that uniquely identifies the check rule.
Text	The text description in local language.
Class	The full name of the check class, including namespace and class name, such as WHU.SDQIAS.TopologicalConsistency.CheckPseudoNode.
Command	The full name of the check command class, including namespace and command class name, such as WHU.SDQIAS.TopologicalConsistency.CheckPseudoNodeCommand.
Priority	The priority that decides which rule is preferred to run in the checking scheme.
Description	The description of the check contents.



(a) Sequence diagram of checking scheme design



(b) Sequence diagram of checking through a scheme

Fig. 7 Sequence diagrams of checking scheme design and application (**a**) Sequence diagram of checking scheme design (**b**) Sequence diagram of checking through a scheme

data layer name and the data dictionary, to create a check item. This continues until the checked contents of each spatial data layer have been considered. Ultimately, a checking scheme with all of the check items, including check rule names and parameter values, is returned and can be stored as a checking scheme file through the SchemeParser class.

The checking scheme file can be interpreted as check items to conduct the scheme checking process through the SchemeParser class. Figure 7b depicts that process. The SchemeParser class produces each rule of a check item, comprising a check rule name and a parameter values string, through the CheckItemService class, which can query the rule according

to the rule name in the check item. The instance of the ParameterParser class created by the ParameterParserFactory class can then be used to parse the parameter values string into real values. After the SDQIAS receives all of the check items, all of the commands with parameter values, which are wrapped in the check rules with parameter values, are sent to the instance of the MacroCheckCommand class. First, the MacroCheckCommand class sorts all commands by the priority of each rule and returns a list of check commands. Finally, each check command is run in the list order in the instance of the MacroCheckCommand. The SDQIAS performs this process.

Data quality assessment

The spatial data quality evaluation process is very complex and difficult to implement given the many factors and cases involved in following the standards (NASMG 2008; ISO 2013). We decompose the evaluation process into three sub-processes: gathering the statistics of the quality error data, evaluating whether one quality element of the quality model is qualified and calculating the score for each quality element and a comprehensive score for the data. The QualityEvaluator class is connected to the QualityResultStatistic class through a one-to-one relationship, and is dependent on the IQualityEvaluatorCondition and IQualityEvaluator interface to run the whole evaluation process under the guidance of the checking scheme file.

Some statistics and analyses must be completed before we evaluate the spatial data product. The QualityResultStatistic class sorts all of the quality errors by quality elements and stores some statistical information, including the quality error number in each quality element, the total number (N) and area of all checked features as the base of the error rate calculation, the feature number of each GIS layer and the error rate of each quality element. The total number of checked features depends on the TotalFeatureCounter class, and the error rate of each quality element depends on the concrete class implementing the ICalErrorRate interface, because the statistics of two datum are multi-situational. For example, in assessing the land cover data quality level the total number of checked features is calculated according to the following expression:

$$N = \begin{cases} 2000\ ,\ N' \le 2000 \\ N', 2000 < N' < 18000 \\ 18000\ ,\ N' \ge 18000 \end{cases} \quad (1)$$

where N is the total number of checked features as the base of the error rate calculation and N' is the real total number of all checked features.

The process for calculating the error rate of each quality model differs. Generally, there are two statistical methods: counting the number of errors and counting the area of the error features. The concrete class CalErrorAreaRate uses the former process and the CalErrorNumberRate class uses the latter.

The calculation of the qualified condition and the quality element mark varies in different assessment cases. Take the spatial data assessment of the FGSAGCC as an example. Table 3 provides an example of assessing each quality element and case of land cover data products in the FGSAGCC. The mark methods in the FGSAGCC include two different cases and the quality element named 'logical consistency' contains two different cases. By configuring the structured assessment scheme file, the framework can successfully solve the diversity problem in assessing cases.

The assessment process is shown in Fig. 8. There are two nested loop fragments in the whole process. The outer loop fragment executes until the mark of each quality element in the assessment scheme file is calculated. The inner loop fragment is part of the outer loop fragment and executes until the instance of the QualityEvaluator class calculates the quality mark of each assessment case in a quality element. The quality mark of a quality element is calculated by all assessment cases in the quality element. Ultimately, the instance of the QualityEvaluator class calculates the quality mark of the spatial data based on the marks of all of the quality elements, and returns the final quality mark to the SDQIAS.

GUI design

In the SDQIAS, seven default quality elements are designed and grouped into five main menus (Table 4). All of the check functions are automatically run according to the system settings and checked data.

Checking process through the system

The check workflow process using the SDQIAS includes some preparation before checking spatial data, loading checked data, checking data and assessing data quality level. Figure 9 presents an activity diagram of the process in the SDQIAS. In the UML, activity diagrams, constructed from a limited number of shapes and connected with arrows, show the overall flow of control (OMG 2007).

Before checking spatial data, some preparation work must be performed, including designing a data dictionary for spatial data, modifying the tolerance of the check rules and configuring global settings. The design of the data dictionary and the configuration of check rule tolerance are based on spatial data quality standards, data collection requirements and data

Table 3 The assessment methods for land cover data in the FGSAGCC

Quality element	Case	Check result	Mark condition	Case mark (s is quality mark)
GeoReference	All	Qualified/Failed	Qualified	$s=100$
Temporal accuracy	All	Qualified/Failed	Qualified	$s=100$
Logical consistency	Consecutive polygon	$r=n/N\times100\%$, where r is the error rate, n is error feature number and N is the total feature number.	$r\le r_0$, where r is the error rate and r_0 is the qualified limit.	$s=60+40/r_0\times(r_0-r)$
	Others	Qualified/Failed	Qualified	$s=100$
Collection precision	All	$r=n/N\times100\%$, where r is the error rate, n is the error feature number and N is total feature number.	$r\le r_0$, where r is the error rate and r_0 is the qualified limit.	$s=60+40/r_0\times(r_0-r)$
Classification accuracy	All	$r=n/N\times100\%$, where r is the error rate, n is the error feature number and N is total feature number.	$r\le r_0$, where r is the error rate and r_0 is the qualified limit.	$s=60+40/r_0\times(r_0-r)$
Map representation	All	$r=n/N\times100\%$, where r is the error rate, n is the error feature number and N is total feature number.	$r\le r_0$,, where r is the error rate and r_0 is the qualified limit.	$s=60+40/r_0\times(r_0-r)$

product requirement descriptions, such as ISO/TC 211 19157 (ISO 2013).

The SDQIAS provides two ways to load checked data: create a check task that supports single map data and batch checking, or add single map data into the system directly. The formats of the supported data include shape files (.shp), the Esri file geodatabase (.gdb), the Esri geodatabase (.mdb) and various raster file formats.

Users can check the spatial data product in relation to a single quality aspect or by a designed checking scheme including all check items. The former way is typically used in process check and data-making procedures, and the check

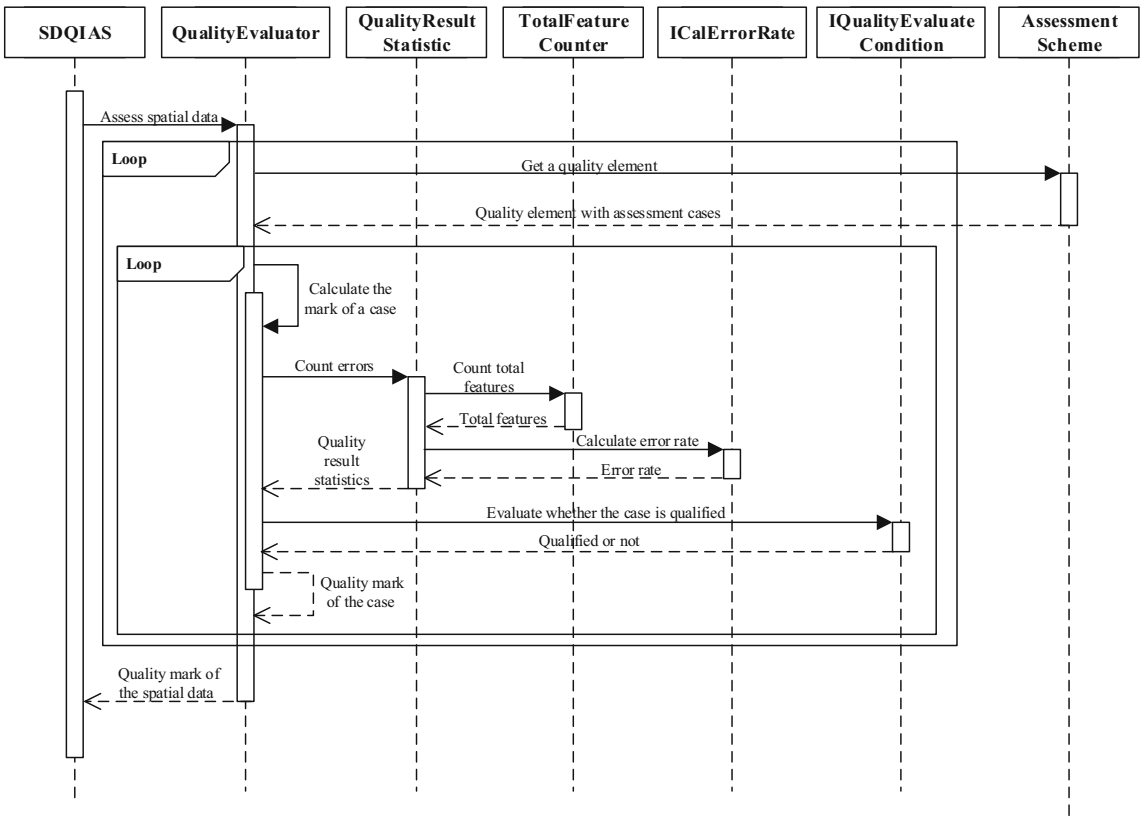


Fig. 8 Sequence diagram of data quality assessment scheme design

Table 4 SDQIAS menus and modules

Module type	Menu	Function
Quality checking modules	Temporal accuracy and coordinate system	Detect temporal quality errors and georeference correctness of the spatial data.
	Position accuracy	Detect position and geometric accuracy errors of the spatial data.
	Data integrity and attribute accuracy	Detect integrity and attribute accuracy errors of the spatial data.
	Topological consistency	Detect topological consistent quality errors of the spatial data.
	Map representation	Detect map representation quality errors of the spatial data.
Design and management modules	Scheme	Allow users to use schemes to check the spatial data and manage all checking schemes.
	Quality evaluation	Allow users to manage all assessing schemes and use one specified assessing scheme to assess the quality level of spatial data.
Other ancillary modules	File	Load checking tasks and checked spatial data.
	View	Control the visibility of all dock windows.
	Tools	Provide some convenient tools for configuring the system and create a supportive file.
	Help	Help users learn how to operate the system.

result is shown in the list immediately. The latter is most often used in final check and acceptance procedures.

After checking the spatial data products, it is necessary to verify the validation of the quality results and remove any false quality errors. In the process check and data-making procedures, the spatial data products are returned to the cartographer for quality error repair and then sent back for inspection until the quality errors are eliminated. In the final check and acceptance procedure, a full quality assessment is executed according to the specified quality and acceptance standards, and a comprehensive quality report giving a final score for the spatial data is automatically produced by the SDQIAS. If the score is less than 60, the spatial data are returned to the spatial data producer or cartographer for

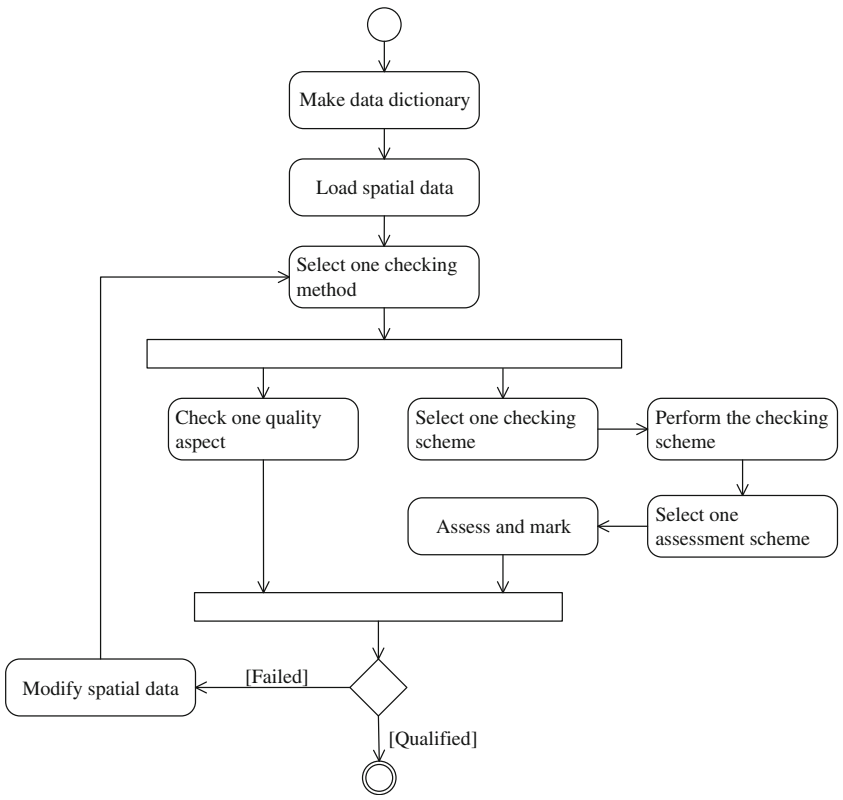


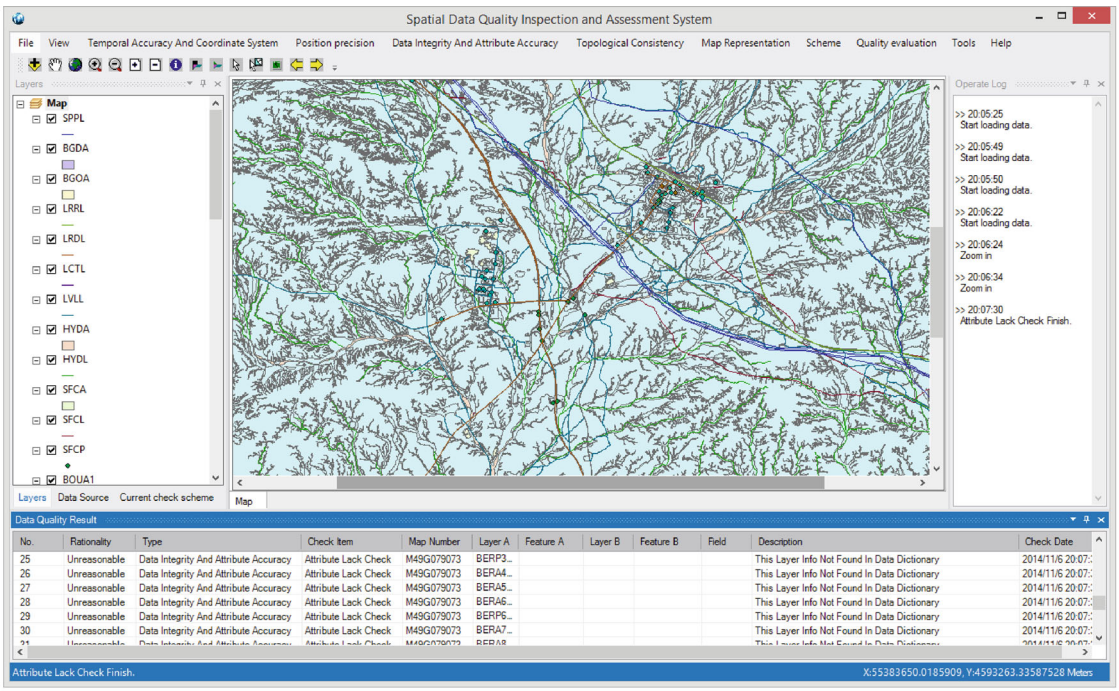
Fig. 9 Checking spatial data through the system

quality error repair. After modifications, the spatial data go through the quality test procedure until the data are qualified.

Application

Based on the framework, we developed the SDQIAS, which integrates basic check rules and provides some data dictionaries and quality models for various spatial data. The SDQIAS is implemented using C# language and processes spatial data through an ArcGIS Engine, which is a product of Esri. Moreover, we have designed a plug-in core for the system that allows users to develop new check rules and customize new user interfaces for novel functionalities.

The SDQIAS was used in the FGSAGCC, and it improved the efficiency of data quality testing and data acceptance. The data products include land cover data, state condition elements data, metadata, interpretation data, etc. Each data product takes a county as its minimum map boundary. Take Shandong Bureau of Surveying and Mapping as an application example. There are 17 cities, 49 municipal districts, 31 county-level cities and 60 counties, but only 17 departments making the FGSAG CC data products in the bureau. In the past, it took one or more weeks to check one data product for a county. Using the SDQIAS, the checking process for one data product from a county takes about 3 h. Figure 10 shows a checking example for a land cover data product.



(a) Overview of the SDQIAS

(b) Quality error statistics

(c) Data assessment results

Fig. 10 SDQIAS application example

Conclusions

Data quality is very important to spatial data applications and users. The related research has mainly focused on describing spatial data quality, and many studies, data quality specifications and standards have been published. Recently, some tools and software have been developed to specify data quality aspects, but they only inspect partial data quality and cannot be reused on the data from different data quality specifications and standards due to varied data types and map scales. In this study, we propose a general framework for inspecting and assessing the spatial data quality of various data types and scales. The result is the SDQIAS for spatial data quality control. The main conclusions can be summarized as follows.

(1) The general framework converts one spatial data quality specification or standard into one quality model that includes various quality elements. The quality model customization and check rules extraction are the core of the general framework for reviewing and analyzing the existing spatial data quality specifications and standards of different data.
(2) Under the general framework, we have designed and implemented four components: data dictionaries, checking rules, checking schemes and data quality assessments. We use structured XML; W3C 2008) files to store each model of the components.
(3) The SDQIAS is used to automatically inspect the spatial data products and assess their quality level, which promotes an efficient quality inspection procedure.

Although the SDQIAS, which is based on the framework, has been widely used in real-world applications that have resulted in more efficient data quality checking and data acceptance, there are still some aspects of the framework and software that can be improved. For the framework, the current data dictionary design for raster data products (e.g. digital elevation model) provides redundant information. For instance, the raster layer does not contain field information. For the SDQIAS, some check rules are still run manually, such as the match between vector data and the responding orthophotos. The system needs more data dictionaries and quality models for inspecting and assessing different data requirements, standards and quality standards.

Acknowledgments This research was sponsored by the National Key Technology R&D Program of China (Grant No. 2012BAJ15B04) and the Hong Kong Polytechnic University Joint Supervision Scheme with the Chinese Mainland (Grant No. G-UA35). The authors would also like to thank the Editor and the two anonymous reviewers whose insightful suggestions have significantly improved this paper.

References

Aronoff S (1989) Geographic information systems: a management perspective. Geocarto Int 4(4):58. doi:10.1080/10106048909354237
Beckert B, Hähnle R, Schmitt PH (2007) Verification of object-oriented software: The KeY approach. Springer, Berlin Heidelberg
Burnicki AC, Brown DG, Goovaerts P (2007) Simulating error propagation in land-cover change analysis: the implications of temporal dependence. Comput Environ Urban Syst 31(3):282–302
Burrough PA (2001) GIS and geostatistics: essential partners for spatial analysis. Environ Ecol Stat 8(4):361–377
Cao Y, Song W (2009) Research on RSDOM and DLG quality mutual check and evaluation technique. Urban Remote Sensing Event, 2009 Joint, 2009, pp 1–6. doi:10.1109/URS.2009.5137579
Chen X, Pan M, Wu H, Yang J, Zhu L (2005) DEM Data Quality Detecting Based on Spatial Index. Appl Res Comput (09):28–30
Crosetto M, Tarantola S (2001) Uncertainty and sensitivity analysis: tools for GIS-based model implementation. Int J Geogr Inf Sci 15(5):415–437
De Champeaux D, Lea D, Faure P (1993) Object-oriented system development. Addison-Wesley Longman Publishing Co., Inc., Boston
Delavar MR, Devillers R (2010) Spatial data quality: from process to decisions. Trans GIS 14(4):379–386. doi:10.1111/j.1467-9671.2010.01224.x
Department of Commerce (1992) Spatial Data Transfer Standard (SDTS). http://mcmcweb.er.usgs.gov/sdts/standard.html. Accessed 17 Sept 2013
Devillers R, Jeansoulin R (2006) Fundamentals of spatial data quality. ISTE, London
Duckham M (2002) A user-oriented perspective of error-sensitive GIS development. Trans GIS 6(2):179–193
Esri (2007) QA/QC for GIS Data. http://training.esri.com/gateway/index.cfm?CourseID=50099063_9.X&fa=catalog.courseDetail. Accessed 23 Feb 2014
Evans MR, Oliver D, Zhou X, Shekhar S (2014) Spatial big data. In: Karimi HA (ed) Big data: Techniques and technologies in geoinformatics. CRC Press, New York, pp 150–156
Fisher P (1997) Book reviews. Int J Geogr Inf Sci 11(4):407–408. doi:10.1080/136588197242356
Foken T, Göockede M, Mauder M, Mahrt L, Amiro B, Munger W (2005) Post-field data quality control. In: Lee X, Massman W, Law B (eds) Handbook of Micrometeorology. Springer, pp 181–208
Forier F, Canters F (1996) A user-friendly tool for error modelling and error propagation in a GIS environment. United States Department of Agriculture Forest Service General Technical Report Rm, pp 225–234
Fu H (2010) Quality control and assessment of 1:2000 DLG. China Place Name (03):70–71
Gong P, Mu L (2000) Error detection through consistency checking. Geogr Inf Sci 6(2):188–193
Goodchild MF, Gopal S (1989) The accuracy of spatial databases. Taylor & Fancis, London
Guptill SC, Morrison JLA (1995) Elements of spatial data quality. Elsevier Science, New York
Hegde NP, Hegde GL (2007) Quality control in large spatial databases maintenance. In: 5th International Symposium for Spatial Data Quality. ITC, Enschede, p 3
Heuvelink GBM (1993) Error propagation in quantitative spatial modelling: applications in geographical information systems. Dissertation, University of Utrecht
Heuvelink GBM, Brown JD, van Loon EE (2007) A probabilistic framework for representing and simulating uncertain environmental variables. Int J Geogr Inf Sci 21(5):497–513
ISO (2005) ISO 9000:2005: Quality management systems – Fundamentals and vocabulary. International Organization for Standardization (ISO), Geneva, p 30

ISO (2013) ISO 19157:2013: Geographic information – data quality. International Organization for Standardization (ISO), Geneva, p 164

Langran G, Chrisman NR (1988) A framework for temporal geographic information. Cartographica: Int J Geogr Inf Geovisualization 25(3): 1–14

Li D, Zhang J, Wu H (2012) Spatial data quality and beyond. Int J Geogr Inf Sci 26(12):2277–2290. doi:10.1080/13658816.2012.719625

Malenfant J, Jacques M, Demers FN (1996) A tutorial on behavioral reflection and its implementation. In: Kiczales G (ed) Proceedings of the Reflection '96 Conference, San Francisco, California, USA, 1996, pp 1–20

McGranaghan M (1993) A cartographic view of spatial data quality. Cartographica: Int J Geogr Inf Geovisualization 30(2):8–19. doi: 10.3138/310V-0067-7570-6566

NASMG (1994) GB 14804–93: Classification and codes for the features of 1:500,1:1000 and 1:2000 topographic maps. National Administration of Surveying, Mapping and Geoinformation (NASMG), Beijing, p 19

NASMG (1995) CH 1002–1995: Specifications for inspection and acceptance of surveying and mapping product. National Administration of Surveying, Mapping and Geoinformation (NASMG), Beijing, p 24

NASMG (2007) GB/T 20258.1-2007: Data dictionary for fundamental geographic information features – Part 1:Data dictionary for fundamental geographic information features of 1:500 1:1 000 1:2 000 scale. National Administration of Surveying, Mapping and Geoinformation (NASMG), Beijing, p 498

NASMG (2008) GB/T 18316–2008: Specifications for inspection and acceptance of quality of digital surveying and mapping achievements. National Administration of Surveying, Mapping and Geoinformation (NASMG), Beijing, p 27

NIST (1994) Federal Information Processing Standard Publication 173. (Spatial Data Transfer Standard Part 1. Version 1.1). National Institute Of Technology (NIST), U.S. Department of Commerce, p 193

NMPQITC (2013) GDPJ 09–2013: Specifications for inspection and acceptance of General Survey Achievements of Geographic. National Mapping Product Quality Inspection and Testing Center (NMPQITC), Beijing, p 33

OMG (2007) Unified Modeling Language: Superstructure, version 2.1.1 (non-change bar). Document formal/2007-02-05. http://www.omg.org/cgi-bin/doc?formal/2007-02-05. Accessed 16 Jan 2014

Oort PV (2006) Spatial data quality: from description to application. Dissertation, Wageningen University

Shi WZ (2008) From uncertainty description to spatial data quality control. Proceedings of the 8th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Vol II: Accuracy in Geomatics, Liverpool, 2008, pp 412–417

Shi WZ, Fisher P, Goodchild MF (2003) Spatial data quality. Taylor & Francis, London

Skidmore A (2002) Accuracy assessment of spatial information. In: Stein A, Meer F, Gorte B (eds), vol 1. Remote Sensing and Digital Image Processing. Springer Netherlands, pp 197–209. doi:10.1007/0-306-47647-9_12

USGS (1941) United States National Map Accuracy Standards. US Geological Survey (USGS), p 1

Veregin H (1999) Data quality parameters. In: Goodchild MF, Maguire DJ, Rhind DW (eds) Geographical information systems. Wiley, New York, pp 177–189

W3C (2008) Extensible Markup Language (XML) 1.0 (Fifth Edition). http://www.w3.org/TR/2008/REC-xml-20081126/. Accessed 24 Aug 2013

Whitney CW, Lind BK, Wahl PW (1998) Quality assurance and quality control in longitudinal studies. Epidemiol Rev 20(1):71–80

Wu D, Hu H, Yang XM, Zheng YD, Zhang LH (2010) Digital chart cartography: error and quality control. The international archives of the photogrammetry. Remote Sens Spat Inf Sci 38(Part II):255–260

Wu M, Zeng J, Li Q (2012) Development of quality checking software for the Second National Land Inventory. Land Resour Informatization (04):12–18

Wu H, Ye L, Shi W, Clarke KC (2014) Assessing the effects of land use spatial structure on urban heat islands using HJ-1B remote sensing imagery in Wuhan, China. Int J Appl Earth Obs Geoinformation 32: 67–78

Zeng J (2009) Quality control and assessment of 1:500 DLG. Beijing Surv Mapp (03):66–68

Zheng F, Wang X (2009) Quality control and detection for data production of 1:10000 topographic maps (DLG). Geospatial Inf (01):91–94