

# Operational suitability assessment of information resources for a geospatial environmental database unity

Yulia Kroumova<sup>1</sup>, Tanya Trenkova<sup>1</sup>, Stoyan Nedkov<sup>1</sup>, Aleksandra Ravnachka<sup>1</sup>

<sup>1</sup> National Institute of Geophysics, Geodesy and Geography – Bulgarian Academy of Sciences, Sofia, Bulgaria

Corresponding author: Aleksandra Ravnachka ([ravnachka77@gmail.com](mailto:ravnachka77@gmail.com))

## Abstract

In the contemporary digital world, data is a pivotal capital considered an economic factor for generating digital products and services. The rapid growth in data volume and their heterogeneity requires optimization of the data management in order to extract higher added value from them. In the context of this issue, one of the goals of the National Scientific Program “Environmental Protection and Risk Reduction from Adverse Events and Natural Disasters” is the creation of a unified geo-informational environment integrating the data from the work packages. To realize this goal, the present study examines certain aspects of data quality related to their usability. By reviewing basic standards for geospatial data, the metadata paradigm, and previous experience in data quality assessment, a theoretical framework with the main stages of a methodological approach evaluating the suitability of primary data sources for building a unified geoinformation database has been developed. The research results reflect the application of developed criteria systems (detailed and generalized) in analyzing the quality and the study of the suitability degree of selected information resources for one of the Program’s working packages. The results of extensive surveys from work program packages have been generalized. Selected primary data sources have been catalogued, and a multicriteria methodological approach has been developed to assess their quality in terms of their suitability for the unity of the database.

**Key words:** Data quality assessment, geospatial database, interoperability



Academic editor: D. Varadzhakova

Received: 5 February 2024

Accepted: 3 April 2024

Published: 28 May 2024

**Citation:** Kroumova Y, Trenkova T, Nedkov S, Ravnachka A (2024) Operational suitability assessment of information resources for a geospatial environmental database unity. *GeoStudies* 1: 59–74. <https://doi.org/10.3897/geostudies.1.e120103>

**Copyright:** © Yulia Kroumova et al. This is an open access article distributed under terms of the Creative Commons Attribution License (Attribution 4.0 International – CC BY 4.0).

## Introduction

The dynamic development of geoinformation technologies in recent decades reveals increasingly complex challenges in the natural, social, and economic scientific fields, both in their specialized research and in their interaction, aiming at preservation of the natural resources and improvement of living conditions. The rapid advancement of technologies is accompanied by the constant growth and accumulation over time of vast and heterogeneous geospatial data sets (Gervais et al. 2002; Lush et al. 2012). These circumstances give rise to a range of problems in the research process, primarily related to the integration of geospatial data into a homogeneous, flexible, and operational GIS environment that will optimize the storage, organization, management, use, and sharing (Thakkar et al. 2007; Triglav et al. 2011).

The present study is focused on specific aspects of data quality within this scientific and technological issue and provides results to support the creation of a *unified geoinformation environment*, a pivotal direction within the activities in the work packages of the National Scientific Program “Environmental Protection and Risk Reduction from Adverse Events and Natural Disasters”. The research in this broadly interdisciplinary work is of both fundamental and applied scientific nature aimed at ensuring sustainable development and improving the living conditions of the population of Bulgaria.

The scientific program implementation covers the period from 2018 to 2023 and involves partnerships with nine scientific organizations and higher education institutions, including the Bulgarian Academy of Sciences (BAS) and Sofia University “St. Kliment Ohridski” being the leading institutions. The research process is organized into ten main and three administrative-technical work packages, divided into sub-packages with specific activities. The work packages cover selected natural and socio-economic scientific fields with an environmental focus corresponding to some main topics of EU environmental policy in the study of: climate (regional and local characteristics), water resources and water balance (quality and quantity), marine environment and coastal zones, biodiversity and ecosystem functions, quality of life, the assessment of the risk and the consequences of adverse natural events, and others (Stoyanova et al. 2023). From the perspective of fundamental science, the scientific program’s research focuses on the components of the natural environment, their interaction, the reflection of their characteristics on natural disasters, and their impact on the quality of life, health risks, and the state of ecosystems. From the perspective of applied science, the emphasis is on the risk assessment and mapping of adverse natural events and disasters, the development of systems for their prediction, early warning, and support for actions during their occurrence. Subsequent applied directions are related to the development of strategies to reduce the health risks for the population and mitigate changes in the natural environment of ecosystems due to harmful emissions into the atmosphere, etc.

The administrative-technical work packages encompass the creation of a unified geoinformation environment, international collaboration, public presentation, and communication. In this part of the program, the primary focus is on providing a contemporary research infrastructure, provisioning detailed and scientifically based documentation to administrative authorities supporting the development of sustainable management strategies, optimizing partnerships, and promoting multidisciplinary scientific research within the team activities (Koulov et al. 2021).

The main goal of achieving an effective geoinformation collateral for the program’s research is to establish a unified geoinformation environment for the integration of geospatial data and information from the thematic packages with the provision of specialized geoinformation services. The aim is to consolidate the geospatial information from all work packages, to develop services, tools, and users’ applications. The geospatial database uses an open-source management system (PostgreSQL) (Dimitrov 2019). In accordance with the program’s predefined goals and objectives, the development of the unified geoinformation environment focuses on quality assessment, integration, efficiency improvement in geospatial data processing, and application of innovative approaches. These specifics are synthesized into five main research tasks,

defining the individual sub-packages of activities in the process of the unified geoinformation environment establishment:

1. Study of the needs and availability of geospatial data and information resources;
2. Development of uniform requirements for the quality of geoinformation resources;
3. Development of a unified geodatabase;
4. Provision of a computational environment for modelling;
5. Development of a GIS server application for specialized geoinformation services (National Scientific Program, WP II 2021).

Based on the activities of the first two tasks, the current study presents fundamental concepts and experimental results related to the primary inventory of geoinformation resources and their categorization according to the level of their suitability for use in the establishment of a unified geospatial database and to interoperability achievement.

## Theoretical concepts

The diversity of topics within the National Science Program (NSP), the differences in the territorial scope of the primary sources, in the technologies for obtaining and storage of data, as well as in models, formats and management systems complicate the synchronization of geospatial information within the database. In this regard, an important factor for the effectiveness of geospatial data integration into a unified geoinformation environment is the assessment of their quality. From a theoretical point of view, among the many definitions of different authors, the quality of geospatial data is primarily associated with their usability and the degree of compliance of their characteristics with certain requirements (Popov 2019). In response to such requirements, geographic information standardization provides systems of rules, criteria, definitions, feature specifications, etc., encompassed in geographic information quality standards. The standards cover the discovery, management, generation, access, sharing, visualization and analysis of geospatial data. In the field of Internet services, standards are used to define the transfer of spatial data in the network and to provide remote access to data stored on web servers. The standards define data access and query methods, spatial data downloads, etc. (ISO/TC211).

The issues of geospatial resources quality and standardization are closely linked to the metadata and the interoperability. As a synthesized documentation for a given resource, metadata represents a description of the key aspects and characteristics of the data such as content, territorial and temporal scope, format, quality, authorship and date of creation, level of accessibility, etc. Their purpose is to assist in the identification and to provide opportunities to the users to make effective decisions about the suitability of a given information resource for the purposes of a particular task. By setting rules for structuring these features about specific topics, metadata standards contribute to the optimization of the use and management of the data. Metadata and their standards have a key role in the synchronization of data from different information sources and their integration into a homogeneous geoinformation environ-

ment. In the exchange and subsequent use of spatial information, they support the process of unifying geospatial information in terminological, technological and semantic terms, thereby contributing to achieve the interoperability.

Regarding internationally recognized standards, the conceptual framework for developing the unified geoinformation environment within the National Program in general refer to and use the series of standards for geographic information from the International Organization for Standardization (ISO), standards from the Open Geospatial Consortium (OGC), and the requirements of the INSPIRE Directive, particularly in their section on data quality (Popov 2019).

In the field of geospatial data quality, the most widely applied are the ISO 19000 series of standards of the International Organization for Standardization (Popov 2019). The standard 19115 is among the general ISO standards for geographic metadata, defining the description of geographic information and related services, including content, data quality, access and rights of use. Other specialized ISO standards cover topics such as: Methodology for properties cataloguing (ISO 19110); for transfer and storage of geographic information (ISO 19136); Metadata – application of the XML schema (ISO/TS 19139:2007); Services (ISO 19119) etc. The basic data quality standard is ISO 19157:2013, Geographic Information-Data Quality (International Organization for Standardization 2023) where the components for description, structure and content, general principles and procedures for determining and assessing data quality have been presented. This standard combines previous stand-alone standards in the field of: Quality principles (ISO 19113:2002); Quality Assessment Procedures (ISO 19114:2003) and Data Quality Measurements (ISO 19138:2006).

The development of spatial data quality standards is also within the scope of the Open Geospatial Consortium (OGC). (Open Geospatial Consortium (OGC) Standards) The organization creates, tests, and maintains standards for operational compatibility and quality of web-based services, aiming to optimize the sharing of geospatial information and improve its accessibility. Currently, a significant number of standards have been developed in the field of web-services, and they can be grouped as follows (Open Geospatial Consortium (OGC) Standards):

- Data Encoding Standards (e.g., GeoPackage, Geography Markup Language (GML)).
- Data Access Standards (e.g., Web Feature Service (WFS), Web Coverage Service (WCS), Sensor Observation Service (SOS), and more).
- Processing Standards (e.g., Web Processing Service (WPS), OGC API-Processes).
- Visualization Standards (e.g., Web Map Service (WMS), Web Map Tile Service (WMTS), KML, Styled Layer Descriptor (SLD), Symbology Encoding (SE), and more).
- Standards for metadata and services (Catalogue Services for the Web (CSW), Metadata).

The theoretical concepts in the development of the geospatial database for the National Science Program are in accordance with the requirements of the INSPIRE Directive. Aimed at building a unified infrastructure for spatial information for the benefit of the environmental policies of the EU countries, the Directive covers 34 topics for spatial data with environmental application. In order

to ensure the compatibility of the spatial data infrastructures of the member states, the Directive requires the adoption of general implementing rules (Implementing Rules – IR), formulated as decisions or regulations of the European Commission in five main areas: metadata, data specifications, network services, data and service sharing and monitoring and reporting (INSPIRE Directive 2024). According to the terminological explanations of the INSPIRE Directive, the concept of interoperability is perceived from the point of view of providing network services for access to data sets, as defined in Article 4 of the Directive. The purpose of such provision is aimed at achieving consistency when combining them with other geospatial data sets (Dimitrov and Popov 2020). In the context of the unity of the geoinformation infrastructure of the National Science Program and for the purposes of the present development, the interoperability is primarily associated with the achievement of effective exchange of information between the thematic packages. In this regard, the principles of the above-cited series of standards of the three organizations under the Program are accepted as a theoretical basis. At an initial stage of systematization of the input information resources, some specifics arising from the multidisciplinary nature of the Program necessitated the clarification of basic aspects of the quality of the data to outline and argue the framework of the research.

Among the variety of general theoretical concepts of geospatial data quality, the current study refers to those that concern the preparation of an initial inventory of the available geoinformation resources and their categorization according to selected quality criteria. A general understanding of these issues is contained in the research of Devillers and Jeansoulin (2006) and is adopted in some of the international standards. They distinguish two main groups of data qualities: *internal*, referring to the specific data sets and the presence of errors and omissions in them, and *external* – in accordance with certain user requirements. If the first group evaluates the *completeness*, *the logic of data structuring* and the *accuracy* of description of the real objects (positional, temporal and thematic), the criteria of the second group cover data qualities such as *definition*, *coverage* (spatial and temporal), *origin*, *legitimacy*, *accessibility* and their evaluation provides information about the satisfaction of specific contextual needs in their use (Popov 2019).

Similar groupings, with more detailed components of internal data quality, are also presented by other authors and publications. In these groupings, the quality of data *usability* is given a separate place. The requirements for this component, as defined by Ahmad (2009), are considered to be specific to individual organizations. The interpretation according to the basic standard for spatial data quality ISO 19157:2013 is analogous, where the element of *usability* is not detailed with sub-elements due to the different characteristics of the subject areas it applies to, which requires additional concretizations. These peculiarities are directly related to the suitability of the data for its use. (Popov 2019)

The current study of the quality of geoinformation resources and the assessment of their suitability for the unity of the database is based on the mentioned concepts of external qualities and usability of data. These statements concern determining the degree of data compliance with respect to the topics of the work packages content, the goals and tasks in the design of the database. In practice, in the multidisciplinary research, as is the present study, the results of such analysis are in the direction of categorizing the suitability of information resources to

optimize their use in the exchange of input information between packages and the achievement of homogeneity of the geoinformation database.

The presented understanding of data *usability* in the ISO 19157:2013 base standard has a key role in examining the quality of input spatial information in the present study. This argues for the private nature of the system of criteria proposed below for assessing the suitability of information resources, consistent with the specifics of the thematic areas of the Program and the connections between them.

## Materials and methods

In methodological terms the current study on the quality of primary spatial data sources uses some of the theoretical frameworks presented by Wang and Strong (1996), which are fundamentally important for information systems. Among the three approaches for studying data quality – *intuitive*, *theoretical*, and *empirical*, the authors give priority to the latter. In the *intuitive* approach, the selection of quality attributes is based on researchers' experience in determining the most significant attributes. In the *theoretical* approach, the focus is on revealing data errors and inconsistencies, while the *empirical* approach is directly related to the use of the information provided by data to satisfy specific research needs. Regardless of the advantages of the *intuitive* approach in choosing specific target quality attributes and of the *theoretical* approach in their completeness, these two approaches, according to the authors, focus on the characteristics of the information product itself. In contrast, the *empirical* approach emphasizes characteristics for data use. The authors highlight the advantage of this approach in capturing those quality attributes that are important to data users. These methodological approaches to the study of data quality and their attributes have a certain analogy with the grouping of quality characteristics by Devillers and Jeansoulin (2006). As a result, *intuitive* and *theoretical* approaches address *internal data quality*, while the *empirical* approach addresses *external* ones. The application of the *empirical* approach involves conducting surveys among data users regarding the priority of specific quality attributes and subsequent statistical analysis of the results to determine the suitability of data for use (Wang and Strong 1996).

In the context of the current study, the determination of data suitability for the program's objectives and the achievement of a unified geospatial database, at an initial stage of the inventory of the input information, relies primarily on the general principles of the *empirical* approach. The diversity of the external data sources and the specifics of the input databases of some of them required certain additions, which are reflected in the proposed stages of the methodological approach.

In terms of data quality assessment approaches, this study uses principles in Veregin and Hargitai's (1995) "Evaluation Matrix for Geographic Data Quality". It is based on a cross-analysis of the geographic information components (columns) and four of the data quality criteria (rows) with the definition of quality assessment tools for each cell. Although the presented methodological approach refers to the assessment of internal data quality, an attempt to adapt it for external data qualities and for the purposes of the present study is discussed in the Results section.

## First stage – Systematization of available information resources

The research process at this initial stage involves surveying the available information resources. As a result of the detailed review of the scientific documentation on work packages, sub-packages and activities, the input information sources cited in them and their metadata in the form of detailed descriptions of the content and characteristics of the spatial information were separated.

The next step of this stage was to conduct surveys on the needs and availability of geospatial data and information resources by work packages. The main objectives of this activity are primarily focused on the realization of a more detailed inventory of the geoinformation resources, clarifying additional details regarding their management, sharing and updating, as well as highlighting certain requirements and characteristics for the spatial data of the individual packages.

The survey questions are distributed into two main thematic groups. The first one includes questions related to the **use of geospatial data**, covering aspects such as data types, formats (for geospatial data/ for management and/ or distribution), estimated average size and volume, metadata (used standards), sharing (infrastructure/ tools), archiving, and access (estimated levels). The second group encompasses questions regarding the **use of models and software**, including spatial analysis models, data processing software, expected final products, and open-source software.

The results processing includes the systematization and generalizations of the answers and comments on the questions of the survey. Additional clarifications regarding the input databases from the preliminary discussions held with the work package managers have also been taken into account. The final part of this stage covers organizing and updating of the available information on the primary geospatial resources under the work packages of four thematic areas. The results of the survey studies primarily aim to achieve comparability of the input geospatial information from different data sources. An important place in this process and at this initial stage of research under the Scientific Program is the identification of needs and the determination of specifics in the available input spatial information.

In relation to the main objective of the present study, the systematization of information at this stage has an extended scope. In addition to the characterization of the input data, the questionnaire surveys include issues related to access, management and sharing of spatial information. In this sense, the application of the *empirical approach* is aimed at achieving the unity, first of all, of the geo-informational environment, laid down in the Program and directly affects the detailing of the primary inventory and the cataloguing of geospatial resources. The purpose of the results of this approach is to achieve uniformity and conformity in the characterization of information resources which will optimize the communication in an operational plan between the work packages and is part of the subject of the current development and the next stage of the methodological approach. At the same time, in relation to the investigation of the external qualities of the data, the application of the survey methods has an indirect nature. The combination of the results of such an exploration with the systematized technical documentation for the information resources by work packages is aimed at supporting the identification and clarifying the selection

of characteristics describing the data, as well as of the attributes for their quality. The application of such an interpretation of the empirical approach contributes to the determination of quality criteria that are consistent with the goals of the study and with the specific work packages' data requirements.

### **Second stage – Development of a basic structure of data catalogue model**

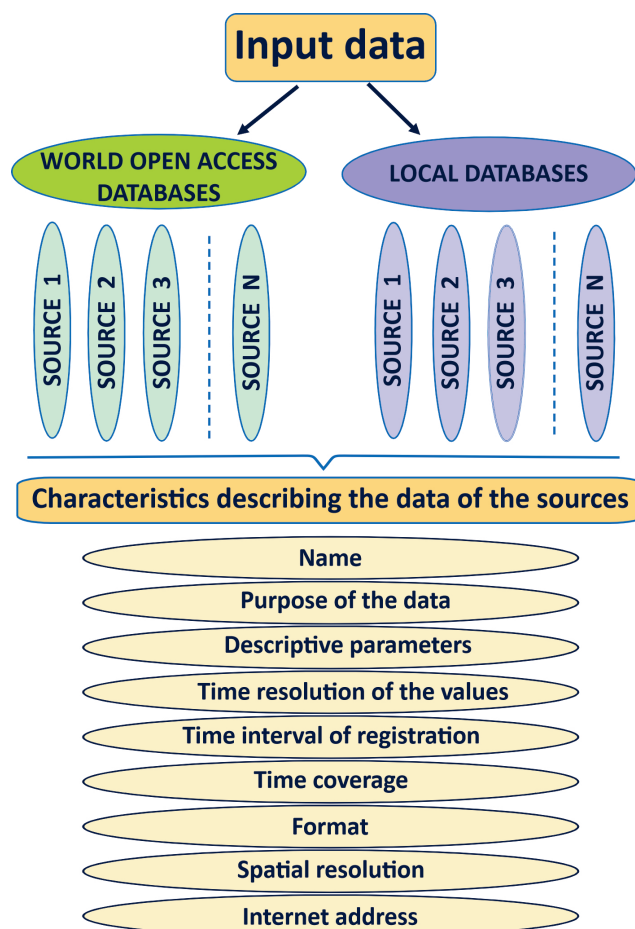
The main activities at this stage of the research of the information resources quality include a preliminary selection of main characteristics describing the data in the resources and the development of a test model of the information catalogue. The selection of descriptive characteristics for geospatial databases is made in accordance with the requirements outlined in the technical reports and the fundamental concepts from the section on Theoretical Concepts of this research, as well as the results from the first part of the survey. The requirements cover several topics, including the formats and standards used for geospatial data, the quality of geospatial data and metadata, their sharing, and the concept, structure, and content of the integrated geodatabase of the National Program.

Part of the systematized information about geospatial resources is structured in Excel format and has the character of an electronic catalogue model with options for additions and updates. The information sources in the model are organized according to their **data types** (rows in the table) and their belonging to the corresponding **work packages, sub-packages, and activities** (columns in the table). Data types are divided into input, derived, auxiliary, and output data. Input data, in turn, are categorized into two main groups: world open access databases and local databases, with the corresponding number of information sources for each sub-package.

The selection of the characteristics describing the data from the information resources is in accordance with the sources with the most detailed descriptions for the databases. As a result, each source in the table (information catalogue model) is described by nine main characteristics: 1) Name; 2) Purpose of the data; 3) Descriptive parameters; 4) Time resolution of the values; 5) Time interval of registration; 6) Time coverage; 7) Format; 8) Spatial resolution; 9) Internet address (Fig. 1).

The second part of this stage involves developing test samples for the electronic catalogue. This activity is based on the nine selected and mentioned above main characteristics describing geoinformation resources and the updated and systematized information available for part of the georesources of the WP.I.5 package, "Quality of Life in the Country," subsection "Meteorological, Natural, and Air Quality Databases – organization, current maintenance, and updating" (National Scientific Program 2019). The main thematic areas for which georesources provide information are natural geographical data, historical and contemporary meteorological data, and historical and contemporary air quality data. It is important to note that the choice of test sources from the "Quality of Life in the Country" package was based on broader scope and completeness of the information describing the available geoinformation resources. This gives reason for using these test sources as a template to entering the descriptive characteristics for georesources from the other packages.





**Figure 1.** Distribution scheme of the information resources input data types in the electronic catalog model.

### Third stage: Development of a system of criteria for analysing the quality of information resources

An important place in the initial inventory process of the available information resources for the achievement of the geospatial database unity in the National Program is occupied by the determination of the criteria and the levels of data quality regarding their suitability for joint use for the needs of the work packages. The selection of criteria for evaluating information sources is based on the system of general criteria for *external data quality* cited in the section on theoretical concepts. As it was emphasized, the contextual nature of the different subject areas in the use of geospatial information necessitates additional detailing of the general criteria for external data quality. Due to the diversity of the geoinformation sources from the different packages of the Program, the differences in the volume of the characteristics describing the databases and the specifics of the research areas, some modifications and additions were introduced in the general system of external quality attributes.

For more complete characterization of the qualities of the sources and in order to facilitate their use, some of the general criteria were additional concretized. For example, the general criterion *Accessibility* is divided in two criteria – 1. **Access** (to the information source) and 2. **Compatibility of data formats**; the general criterion *Coverage* corresponds to two criteria – 3. **Territorial coverage**

and 4. **Time range**; the general criterion *Definition* is represented by the temporal aspects – 5. **Actuality** (of the information source) and 6. **Data update frequency**; to the general criterion *Legitimacy* corresponds the criterion 7. **Metadata completeness**. Additionally, another criterion was introduced – 8. Level of the **Development of the resource**. During the study of the available test geoinformation sources from the WP.I.5 “Quality of Life in the Country,” presented in the information catalogue model, for the determination of assessment levels for the individual criteria, two types of criteria systems were developed – detailed and generalized. The detailed system is represented by four levels (Table 1).

The incompleteness in the assessment levels of some criteria in the detailed system determined its auxiliary character, for orientation when setting the levels thresholds of the separate criteria during the sources assessment of the remaining packages. These circumstances required merging of some of the levels to achieve unity in the evaluation. As a result of these transformations, primarily in levels 2 and 3 of the detailed system, the generalized model of the 3-level criteria system was obtained (Table 2).

A brief description of the threshold characteristics for the assessment levels of the individual criteria is made in the following sequence.

**Access** (to the information source) is presented in three levels, depending on whether specialized rights or registration are required. The latter level of free access is intended to be clarified during the process of analysis, as there may be free but partial access. The differentiation of the three levels of the **Compatibility of data formats** criterion is related to ESRI ArcGIS formats and the possibilities to apply additional transformation operations (Level I), the availability of connectivity of file formats (Level II), or complete compatibility (Level III). The completeness of the **Territorial Coverage** criterion was initially presented in four levels, the lowest of which – local refers to a settlement, neighbourhood and other urbanized territories, Regional I – for a municipality,

**Table 1.** Detailed criteria system (the subsequent merging of the assessment levels of some of the criteria for the generalized system is presented in colour).

Criteria/ levels	Criterion 1 Access	Criterion 2 Compatibility of data formats	Criterion 3 Territorial coverage	Criterion 4 Time range	Criterion 5 Actuality	Criterion 6 Data update frequency	Criterion 7 Metadata completeness	Criterion 8 Development of the resource
0	Lack of access	Lack of compatibility				not updated	lack of metadata	
1	Limited access with specialized rights	Partial compatibility (after transformation)	Local level (town, district)	Short-term (less than 10-year period)	Historical (20th century information only)	Long-term I (within up to 10 years/ multi-year)	Not completed (below 50%)	Initial stage (below 50% resource utilization)
2	Limited access with registration	Full compatibility (with the additional related file formats)	Regional level I (municipality, group of municipalities, natural geographical unit - subarea/region)	Medium-term I (from 10 to 20 year period)	Contemporary I (information for the period 2010-2019)	Long-term II (annual)	Partially filled (50-90%)	Intermediate stage (over 50% resource utilization)
3	Free/ full access	Full compatibility (with the supported file formats)	Regional level II (administrative or natural geographical units/ districts, areas)	Medium-term II (from 20 to 50 year period)	Contemporary II (21st century information to the present day)	Medium-term (weekly, monthly, seasonal)	Fully completed (90-100%)	Final stage (final version)
4			National level	Long-term (50 to 100 and over 100 year period)	Historical and contemporary (20th and 21st century information to the present)	Short-term (permanent, hourly, daily)		

Table 2. Generalized criteria system.

Criteria/ levels	Criterion 1 Access	Criterion 2 Compatibility of data formats	Criterion 3 Territorial coverage	Criterion 4 Time range	Criterion 5 Actuality	Criterion 6 Data update frequency	Criterion 7 Metadata completeness	Criterion 8 Development of the resource
0	Lack of access	Lack of compatibility				not updated	lack of metadata	
1	Limited access with specialized rights	Partial compatibility (after transformation)	Local level	Short-term (less than 10-year period)	Historical (20th century information only)	Long-term (annual, multi-year)	Limited completed (below 50%)	Initial stage (below 50% resource utilization)
2	Limited access with registration	Full compatibility (with the additional related file formats)	Regional level	Medium-term (from 10 to 50 year period)	Contemporary (21st century information to the present day)	Medium-term (weekly, monthly, seasonal)	Partially filled (50-90%)	Intermediate stage (over 50% resource utilization)
3	Free/ full access	Full compatibility (with the supported file formats)	National level	Long-term (50 to 100 and over 100 year period)	Historical and contemporary (20th and 21st century information to the present day)	Short-term (permanent, hourly, daily)	Fully completed (90-100%)	Final stage (final version)

group of municipalities, natural geographical unit (subarea/region), the next higher ranking Regional II, concerns an administrative districts or natural geographical area. Subsequently, these two levels were unified into one – **Regional**. The **Time range** criterion was also classified into the detailed system of four levels, depending on the most common periods of the sources in the tested work package. The transition to a 3- assessment level system required the consolidation of the two medium-term time frames into one in the second level. It is important to note that the periodization depends on the field of study and therefore it may be different for the sources in the different packages. This may exceptionally require the addition of other thresholds variations for the three levels. Similar changes were made also to the subsequent criteria, **Actuality**, and **Data update frequency**, which were initially categorized into four assessment levels. The **Metadata Completeness** and the **Development of the resource** criteria are categorized into three levels and presented in both systems with percentage ratios. Regarding metadata, they express the degrees of conformity according to the used standards in the Directive INSPIRE – ISO 19115-1:2013 и ISO 19115-2:2009 (Popov 2019). In the last criterion, three degrees of usability of the information source is presented – Initial stage (below 50% usability of the resource), in which single applications are finalized; Intermediate stage – in the development process (over 50% usability of the resource), where most of the applications are active and Final stage – final version of the information resource. For three of the criteria (Access, Data update frequency, and Metadata completeness), another – zero level was introduced to verify the complete absence of information on the respective quality characteristic.

#### Fourth stage: Selection of test information sources for application of criteria systems

Among the external geoinformation resources for the work package WP.I.5 “Quality of Life in the Country,” three of them have been selected, which was distinguished by the highest completeness of databases’ characteristics.

These are: two sources maintained by the National Centre for Environmental Prediction (NCEP) – the first is the System for Global Reanalysis of Meteorological Variables (NCEP/DOE Reanalysis 2 (R2)), and the second is the Global Tropospheric Analyses and Forecast Networks (NCEP GDAS/FNL 0.25 Degree Global Tropospheric Analyses and Forecast Grids), a database containing output fields from the Global Forecast System (GFS) model, with assimilated measurement data from all available sources. The third source is the CORINE 2018 Land Cover data. (Tables 3, 4).

## Results

The current study on the quality of data from the information sources for the National Program is experimental in nature, and its main results are in two directions. The first is related to the inventory and the development of an electronic catalogue model for primary sources of geospatial information from the different thematic areas to support the comparability and compatibility of the data to achieve unity in its integration into the overall database. On the other hand, an effectively organized arrangement of the primary geospatial resources is intended to optimize operational activities in the research process in the exchange of spatial information between the work packages of the Scientific Program. A central place in the process of cataloguing the geoinformation resources is the selection of the characteristics describing the data in them, to be valid for a larger part of the sources from the different work packages. Representative fragments of a proposed model are presented in Tables 3, 4.

The detailization of the field “key” covers the main varieties of data types in the information sources and the nine characteristics that describe them, as

**Table 3.** Information source – I System for Global Reanalysis of Meteorological Variables (NCEP/DOE Reanalysis 2 (R2)).

Question	Key				WPI.5.1-1 Meteorological, natural geographical and atmospheric air quality databases – organization, on-going maintenance & update	
					<i>Activity 1. Organization, support and update of natural geographical data for the region</i>	<i>Activity 2. Organization, support and update of historical and present-day meteorological data for the region</i>
1 Data types	Input Data	World Open Access DataBases (OADB)	Source 1	Name of the source	SRTM Digital Elevation Data	<b>NCEP/DOE Reanalysis 2 (R2)</b>
				Purpose of the data	The data is implemented in the WRF model	For the developed database
				Parameters, describing the data	Natural geographical elements	System for Global Re-Analysis of Meteorological Variables - Air Temperature and Winds of the Ground and Upper Atmospheric Layers, Boundary Layer Winds, Geopotential Height
				Type of values (time frame)		daily average and monthly average
				Frequency (time interval) of logging		every 6 hours (00 UTC, 06 UTC, 12 UTC, 18 UTC)
				Period/ year (time coverage)		from January 1979 to the present
				Format		GRIB
				Spatial resolution	~30 m	Gaussian T62 horizontal grid with 28 vertical levels
				Web-address	<a href="https://ita.cr.usgs.gov/SRTM1Arc">https://ita.cr.usgs.gov/SRTM1Arc</a>	<a href="https://rda.ucar.edu/datasets/ds091.0/#description">https://rda.ucar.edu/datasets/ds091.0/#description</a>

**Table 4.** Information sources – II Global Tropospheric Analyses and Forecast Grids (NCEP GDAS/FNL 0.25 Degree Global Tropospheric Analyses and Forecast Grids) and III – Corine Land Cover 2018 (CORINE2018 Land Cover data).

Question	Key			WP.I.5.1-1 Meteorological, natural geographical and atmospheric air quality databases – organization, on-going maintenance & update		
				Activity 1. Organization, support and update of natural geographical data for the region	Activity 2. Organization, support and update of historical and present-day meteorological data for the region	
1 Data types	Input Data	World Open Access DataBases (OADB)	Source 2	Name of the source	CORINE2018 Land Cover data	NCEP GDAS/FNL 0.25 Degree Global Tropospheric Analyses and Forecast Grids
				Purpose of the data	The integrated data in WRF model	For the developed database
				Parameters, describing the data	Land Cover (area/ha)	Database of output fields from the global GFS model with assimilated measurement data. The parameters include: surface pressure, sea level pressure, geopotential height, temperature, sea surface temperature, soil values, snow cover, relative humidity, u- and v-winds, vertical motion, vorticity, and ozone.
				Type of values (time frame)	Annually	Daily
				Frequency (time interval) of logging	6 years	every 6 hours (00 UTC, 06 UTC, 12 UTC, 18 UTC)
				Period/ year (time coverage)	2017-2018	from July 2015 to the present
				Format	GeoTiff, ESRI Geodatabase, SQLite Database	GRIB
				Spatial resolution	~90 m	0.5° (~55 km) for previous periods and 0.25° (~27 km)
				Web-address	<a href="https://land.copernicus.eu/pan-european/corine-landcover/clc2018?tab=download">https://land.copernicus.eu/pan-european/corine-landcover/clc2018?tab=download</a>	<a href="https://rda.ucar.edu/datasets/ds083.3/index.html#sfol-wl-/data/ds083.3">https://rda.ucar.edu/datasets/ds083.3/index.html#sfol-wl-/data/ds083.3</a>

selected and specified in the second stage in the development of the structure of the information catalogue model, presented above in the section 3.2.

The second direction is related to the selection and the argumentation of a criteria system for evaluating the external qualities of the information resources, which are directly related to the assessment of their suitability for subsequent joint operational activities for the work packages. An essential part of the analysis focused on a detailed study of the descriptive characteristics of the data from the sources for each criterion in one of the working packages. A primary purpose in determining the thresholds for assessment levels is to achieve relevant results with respect to the sources of the other packages. In that regard, similar to Veregin and Hargitai’s “Matrix for Assessing the Quality of Geographic Data” (Veregin and Hargitai 1995), two variants of a matrix (detailed and generalized) of the operational suitability of information resources were developed, but with a proposed new interpretation of the content in the columns and the rows. The columns reflect the eight criteria, and the rows represent the information sources (in this case, the three test sources mentioned above). Each cell contains the specific descriptive information about the data of the source according to the given criterion and its belonging to the corresponding assessment level according to the two criterion systems – detailed and generalized. The results of the two types of operational suitability categorization of the three test information sources are visualized by the diagrams in Fig. 2.

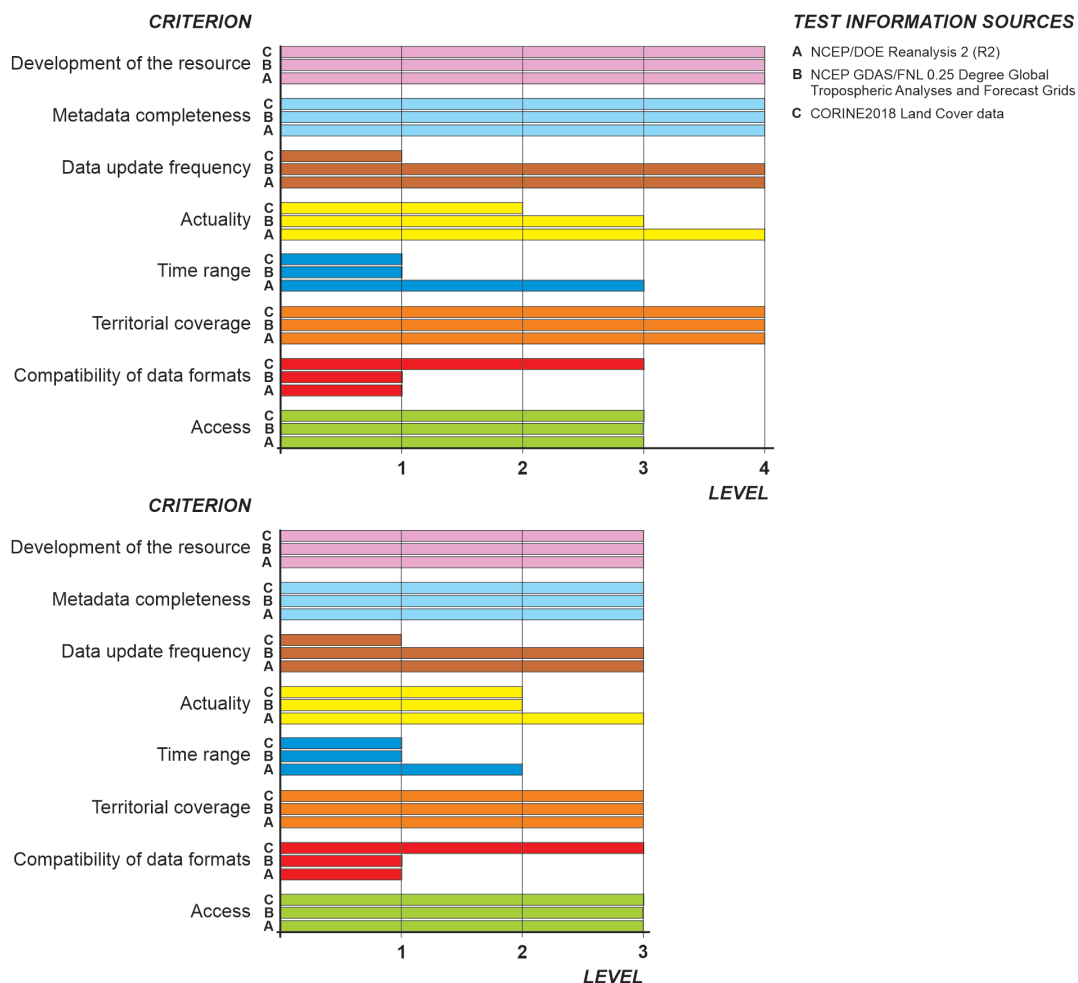


Figure 2. Criterion assessment of operational suitability for the three test data sources (according to the detailed four-level and the generalized three-level matrix schemes).

## Conclusions

The presented model of an information catalogue represents a structured form of detailed inventory of geoinformation resources. Its purpose has two aspects. On the one hand, it is aimed at increasing information awareness, improving and facilitating the communication between the working packages regarding input information. At the same time, such a systematization of information resources forms the basis of the data quality analysis of the information sources.

The proposed generalized criteria system is aimed at supporting the categorization of various geoinformation sources in multidisciplinary studies, necessary to optimize their application in the development of a unified and homogenous structure of the environmental database and achieving operational compatibility.

## Acknowledgements

The present study was conducted in connection with the implementation of the National Scientific Program “Environmental Protection and Risk Reduction from Adverse Events and Natural Disasters”, approved by Decision of the Ministry of the Interior No. 577/17.08.2018 and financed by the Ministry of Education and Science (Agreement No. D01-322/18.12.2019).

## Additional information

### Conflict of interest

The authors have declared that no competing interests exist.

### Ethical statement

No ethical statement was reported.

### Funding

This research is financed by the Ministry of Education and Science (Agreement No. D01-322/18.12.2019).

### Author contributions

All authors have contributed equally.

### Author ORCIDs

Yulia Kroumova  <https://orcid.org/0009-0006-3117-3504>

Stoyan Nedkov  <https://orcid.org/0000-0002-0052-9815>

### Data availability

All of the data that support the findings of this study are available in the main text.

## References

- Ahmad T (2009) Spatial data quality, I.A.S.R.I., New Delhi – 110012. In: Popov A (2019) A Review and critical analysis of the existing experience in assessing the quality of geospatial data and information. Technical report. NNP “Protection of the environment and reduction of the risk of adverse events and natural disasters”, 27–27.
- Devillers R, Jeansoulin R (2006) Spatial Data Quality: Concepts. In: Fundamentals of spatial data quality, GIS: Geographical Information Systems Series, London UK, Newport Beach USA, ISTE Ltd. In: Popov A (2019) A Review and critical analysis of the existing experience in assessing the quality of geospatial data and information. Technical report. NNP “Protection of the environment and reduction of the risk of adverse events and natural disasters”, 26–26. <https://doi.org/10.1002/9780470612156.ch2>
- Dimitrov St (2019) Design and development of an integrated geodatabase. Technical report. NSP “Protection of the environment and reduction of the risk of adverse events and natural disasters”, 3–3.
- Dimitrov St, Popov A (2020) Metadata requirements for geoinformation resources in the unified geoinformation system. Technical report. NNP “Protection of the environment and reduction of the risk of adverse events and natural disasters”, 3–4.
- Gervais DR, Bédard Y, Jeansoulin R (2002) Spatial Data Quality: From Metadata to Quality Indicators and Contextual End–User Manual. OEEPE/ISPRS Jt. Workshop Spat. Data Qual. Manag. 2002, 45–55.
- INSPIRE Directive (2024) INSPIRE Directive. <https://inspire.ec.europa.eu/inspire-directive/2>
- ISO (2013) International Organization for Standardization (ISO). [www.iso.org/standard/32575.html](http://www.iso.org/standard/32575.html)
- ISO (2022) ISO/TC211 Geographic Information/Geomatics, ISO 19115-1:2014 & ISO 19115-2:2009, Geographic Information - Metadata: Part 1 - Fundamentals; part 2-Ex

- tensions for imagery & gridded data. <https://committee.iso.org/sites/tc211/home/projects/projects---complete-list/iso-19115-2.html>
- Koulov B, Stoyanova V, Borisova B, Rachev N, Ganey K (2021) Review of local and regional aspects of natural hazards. In: Koulov B, Stoyanova V (Eds) Local and regional aspects of natural hazards, Az-Buki – National Publishing House for Education and Science, 7–14. [ISBN:987-619-7065]
- Lush V, Bastin L, Lumsden J (2012) Geospatial data quality indicators. In: Proceedings of the 10<sup>th</sup> International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Florianopolis, SC, Brazil, 10–13 July 2012, 121–126.
- National Scientific Program (2019) National Scientific Program “Environmental Protection and Risk Reduction from Adverse Events and Natural Disasters”, Report of the first year, 2019, Sofia, 219 pp. [https://nnpos.files.wordpress.com/2019/d0bed182d187d0b5d182\\_2019.pdf](https://nnpos.files.wordpress.com/2019/d0bed182d187d0b5d182_2019.pdf) [wordpress.com]
- National Scientific Program (2021) National Scientific Program “Environmental Protection and Risk Reduction from Adverse Events and Natural Disasters”, WP II: Horizontal (cross-cutting) work packages, 2021. [https://nnpos.files.wordpress.com/2021/01/tablica\\_nnp\\_oos\\_rp\\_ii\\_horizontal\\_2021.pdf](https://nnpos.files.wordpress.com/2021/01/tablica_nnp_oos_rp_ii_horizontal_2021.pdf)
- Open Geospatial Consortium (OGC) Standards. <https://www.ogc.org/docs/is>
- Popov A (2019) A Review and critical analysis of the existing experience in assessing the quality of geospatial data and information. Technical report. NNP “Protection of the environment and reduction of the risk of adverse events and natural disasters”, 2, 7, 8, 25, 26, 67, 69, 72, 84.
- Stoyanova V, Koulov B, Borisova B, Rachev N (2023) Review of Environmental and Natural Disaster Policies in the 21<sup>st</sup> Century. Problems of Geography, Bulgarian Academy of Sciences, vol. 1–2, Sofia, 107–129.
- Thakkar S, Knoblock CA, Ambite JL (2007) Quality-driven geospatial data integration. In: Proceedings of the 15<sup>th</sup> Annual ACM International Symposium on Advances in Geographic Information Systems (ACM, 2007), 16–16. <https://doi.org/10.1145/1341012.1341034>
- Triglav J, Petrovič D, Stopar B (2011) Spatio-temporal evaluation matrices for geospatial data. International Journal of Applied Earth Observation and Geoinformation 13(1): 100–109. <https://doi.org/10.1016/j.jag.2010.07.002>
- Veregin H, Hargitai P (1995) An evaluation matrix for geographic data quality. In: Guptill SC, Morrison JL (Eds) Elements of Spatial Data Quality, 167–188. Elsevier Science, Published on behalf of the International Cartographic Association, Great Britain. In: Popov A, (2019) A Review and critical analysis of the existing experience in assessing the quality of geospatial data and information. Technical report. NNP “Protection of the environment and reduction of the risk of adverse events and natural disasters”, 18–18. <https://doi.org/10.1016/B978-0-08-042432-3.50016-1>
- Wang R, Strong D (1996) Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems 12(4): 5–33. <https://doi.org/10.1080/07421222.1996.11518099>