The method for the quality evaluation of open geospatial data for creation and updating of datasets for National Spatial Data Infrastructure in Ukraine

Abstract. The purpose of the article is to present the research on method of the quality evaluation of published open geospatial data and its implementation in Ukraine. The method of the quality evaluation of open geospatial data considers the international standard ISO 19157 “Geographic information. Data quality”. This method is to determine the number of points or levels (maximum – 5). The research was carried out for the evaluation of open geoinformation resources for production of geospatial datasets, as defined in the Ukrainian Law on NSDI. The authors evaluated the quality of 142 open geoinformation resources and other information resources (materials) for the production and updating of 34 geospatial datasets for the development of NSDI in Ukraine. The authors present the example of the quality evaluation of geospatial data for datasets: “State Geodetic Reference Coordinate System UCS-2000”, “State Geodetic Network”, “Geographical Names” and “Administrative Units” because they are the components of the Core Reference Dataset of NSDI. Limitations of the research were determined by the adopted the Law of Ukraine “On National Spatial Data Infrastructure” and the Order for NSDI functioning in Ukraine and the requirements of the international standard ISO 19157 “Geographic information. Data quality”. The results of the research will be employed to evaluate the quality of NSDI implementation in Ukraine. The proposed method allows evaluating the quality of open geospatial dataset before using them for analysis and modeling of terrain, phenomena. This method takes into account the quality of geospatial data, and its related requirements for their production, updating and publication.

Keywords: open data, data quality, GKI, NSDI, RDF

1. Introduction

The Spatial Data Infrastructure (SDI) is evolving into the Geospatial Knowledge Infrastructure (GKI), which is a hallmark of the Fourth Industrial Revolution (4IR), where people exploit to the maximum the possibilities of the digital world. Data is not the end product and its true value lies in gaining knowledge for satisfying the needs of society. Thus, geospatial knowledge is crucial to addressing the world’s greatest challenges, and it is important to understand the role of geospatial knowledge in the future of a stable digital society (Geospatial World and the United Nations Statistics Division, 2020; UN-GGIM, 2019). The openness of official data will ensure information interaction between data owners and users, between states and international organizations (Lindgren et al., 2016). It will reduce the time to identify problems in various fields and will allow to respond to them faster, to prevent their dissemination. It is essential that open spatial data is official and high quality for use in other information and geoinformation systems for processing, analysis, monitoring of the state of the environment and production of new models for forecasting phenomena.

The important aspect of modern national spatial data infrastructures is the production and distribution of available open data for direct download through geoportals.
There is a tendency for public sector geo-spatial data to be given the status of open government official data for free download from official government web portals. The latest trend is recorded as a requirement in national strategies for the development of geospatial information for the quality of geospatial data for decision-making systems on the principle of AAA (Accurate, Authoritative and Assured).

Adoption of the Law of Ukraine “On National Spatial Data Infrastructure” (NSDI) on April 13, 2020, and its implementation emphasized the urgent need and urgency in creating Core Reference Datasets, which form a unified digital coordinate-spatial basis for production, integration and other activities with different thematic spatial datasets (The Law of Ukraine “On The National Spatial Data Infrastructure”, bill no. 2370, 2020). Also, the Order for the NSDI functioning in Ukraine was adopted on May 26, 2021, which establishes a mechanism for the creation, functioning and development of national spatial data infrastructure and the organization of production, updating, processing, storage, publication, visualization, supply and use of spatial data, metadata and other activities with them (The Order of Functioning of The National Spatial Data Infrastructure, no. 532, 2021).

1.1. Formulation of the problem

The geospatial datasets used in the NSDI must correspond to the Law of Ukraine “On National Spatial Data Infrastructure” and the Order for NSDI functioning in Ukraine and the requirements of the national standard ISO 19157 “Geographic information. Data quality”. Quality description information is included in the metadata description and the spatial data specification.

Annexe 2 to the Order for NSDI functioning precise the geospatial dataset and executive authorities, local governments and other stakeholders responsible for the creation and updating of geospatial datasets and metadata.

So, there are two questions: 1. Is it possible to use available open data, geoinformation resources and other information resources (materials) to create and update geospatial datasets? 2. Is there a method for the quality evaluation of open geospatial data?

1.2. The purpose of the research

The purpose of the article is to present the research on the method of the quality evaluation of published open geospatial data and its implementation in Ukraine.

2. Analysis of recent research

The problem of using open geospatial data in the process of development of spatial data infrastructures of states was studied e.g. by Strobl and Nazarkulova (2014), Hanečák et al. (2017), Lazorenko-Hevel (2017) and Corti et al. (2018). The quality of these data was pondered by Devillers et al. (2010), Usery et al. (2018) and Hamad (2020). The tendency to follow requirements is defined to geospatial data according to a set of standards ISO 19100, even for volunteer geoinformation systems. The international project of the Open Knowledge Foundation (Open Knowledge International, 2017) has also been taken into consideration. This project evaluates data sets of countries around the world according to the following criteria:

- openly licensed,
- an open and machine-readable format,
- downloadable at once,
- up-to-date,
- publicly available,
- available free of charge.

The Resolution of the Cabinet of Ministers of Ukraine of October 21, 2015, № 835 “Regulations on datasets to be published in the form of open data” was adopted regarding the evaluation of open data sets, which describes the method of the quality evaluation of open data, separately for each dataset. The evaluation of the dataset is formed by adding the numerical weights of the answers for each criterion. This Resolution does not take into account the peculiarities of the quality of geospatial data.

In 2010, Berners-Lee proposed a method for estimating data openness by five stars (Berners-Lee, 2006; Janowicz et al., 2014):

**The One-Star Open Data.** The open data with one star is defined as data available on the Internet, in any format, but with an open license. The users can search, store, modify data and share data with anyone. The organization as a data owner knows where and how to publish them.
The Two-Stars Open Data. The open data should be available in a machine-readable structured form, such as data in an Excel spreadsheet instead of a scanned table image. The users of two-star open data can do whatever they do with one-star data, plus process it directly with their own software and export it to another structured format. However, this type of data is still blocked because users can get data from the document, depending on their software.

The Three-Stars Open Data. The open data with three-stars is defined when users do not need their own software package for data analysis. There is one example of this data. It is the comma-separated value format (CSV), which stores tabular data as plain text.

The Four-Stars Open Data. The open data with four-stars is defined as the data that uses open standards from the W3C, such as Framework Description Framework (RDF) and SPARQL Protocol and RDF Query Language (SPARQL), to identify entities. RDF is a standard that uses in a semantic graphs database. This semantic graph database (also known as RDF triplestore) is a type of semantic technology for storing and managing interconnected data and understanding that interconnected data. In contrast to a relational database, the triplestore displays various relationships between entities in graph databases. SPARQL is a standardized W3C query language for the RDF database. The basic concept of triplestore and the basic principle of connected data is a unique resource identifier (URI). URI is a unique identifier for all related things. The user can reference it from anywhere else or reuse parts of the data by presenting data in a graph database.

The Five-stars related to open data. The data owners associate their data with other user’s data to provide context using W3C standards and related data principles. A semantic graph database is capable of processing different datasets and associates’ links to related open data sources, such as DBpedia or GeoNames, for example. The five-stars data users can discover more and more interconnected information using this data. As the semantic graph database can derive new references from existing facts, users can find more relationships in their related data.

According to the United Nations Department of Economic and Social Affairs (2019) the five principles for the Global Fundamental Geospatial Data Themes are:

1. Use of fundamental geospatial infrastructure and geocoding;
2. Geocoded unit record data in a data management environment;
3. Common geographies for the dissemination of statistics;
4. Statistical and geospatial interoperability;
5. Accessible and usable geospatially enabled statistics.

The quality of open geospatial data affects the quality of the created geospatial datasets, which will ensure the implementation of SDI and provide a basis for the formation of knowledge in the Geospatial Knowledge Infrastructure.

3. Results

3.1. The requirements for NSDI datasets production

The adopted Order for NSDI functioning in Ukraine determined the general requirements for the organization of production/updating of geospatial data, which should minimize duplication of work using geographic information services such as WMS and WFS (The Order of Functioning of The National Spatial Data Infrastructure, no. 532, 2021).

It is necessary to follow uniform requirements to ensure the interoperability of geospatial data for:

- coordinate system for the coordinates of spatial features;
- compliance with the rules of topological relations and topological consistency of spatial features;
- classification of features of Core Reference and thematic geospatial datasets using feature catalogues of classes and their attributes;
- presentation of feature identifiers (Karpinskyi et al., 2020);
- data presentation format.

The national standard DSTU ISO 19111:2017 (ISO 19111: 2007, IDT) Spatial referencing by coordinates should be used for the presentation of geospatial data in the NSDI.

It is important to use Core Reference data when creating thematic spatial data and to ensure regulated access of individuals involved
3.2. The requirements for interoperability of spatial relations descriptions and topological consistency of spatial features

The integrity of the spatial database is ensured by compliance with all the rules of relational integrity (restriction of attribute domains, rules of entity integrity and rules of link integrity), and the rules of topological consistency of features.

The topological constraints at the spatial feature class level define the rules of topological relationships between features of the same class, such as constraints on overlays, common geometry if the features form a continuous coverage, and so on.

All restrictions on spatial relationships and coordinate-topological consistency of spatial data must be described and documented in the specification of spatial data, developed following the requirements of the national standard DSTU ISO 19131: 2019 (ISO 19131: 2007; Amd 1: 2011, IDT) Specification data-product.

3.3. The requirements for the classification of spatial features using feature catalogues of classes and their attributes

The mandatory condition for production of spatial data in NSDI by stakeholders is the presence of classifiers in the form of feature catalogues and their attributes, developed according to national standards of Ukraine such as DSTU ISO 19110:2017 (ISO 19110: 2016, IDT) “Geographical information. Methodology for feature cataloguing” and DSTU 8774:2018 “Geographical information. Rules for spatial data modelling”.

Table 1. The criteria for the quality evaluation of open geospatial data

| №  | Criteria                                                        | Evaluation* |
|----|-----------------------------------------------------------------|-------------|
|    |                                                                | 1  | 2  | 3  | 4  | 5  |
| 1  | Internet access in an open license                               | M  | M  | M  | M  | M  |
| 2  | The presence of a geographical identifier                       | O  | M  | M  | M  | M  |
| 3  | Data structure                                                  | O  | M  | M  | M  | M  |
| 4  | Ability to obtain the coordinates of features in the coordinate system that registered in EPSG | O  | M  | M  | M  | M  |
| 5  | OGC data formats                                                | O  | O  | M  | M  | M  |
| 6  | 0 level of spatial data topology                                | M  | M  | O  | O  | O  |
| 7  | 1 level of spatial data topology                                | O  | O  | M  | O  | O  |
| 8  | 2 level of spatial data topology                                | O  | O  | M  | M  | O  |
| 9  | 3 level of spatial data topology                                | O  | O  | O  | O  | M  |
| 10 | The transfer of the value of the feature ID to the URL is referenced on other systems | O  | O  | M  | M  | M  |
| 11 | Publishing data using geographic information services           | O  | M  | M  | M  | M  |
| 12 | RDF data format                                                 | O  | O  | O  | O  | M  |

* M – mandatory, O – optional
There are approved state feature classifiers for the subject areas, and their use is mandatory.

The feature identification using address data and geographical identifiers should be performed following the standard DSTU ISO 19112: 2017 (ISO 19112: 2003, IDT) “Geographic information. Spatial reference by geographic identifiers”.

3.4. Requirements for spatial data exchange formats

The exchange format of spatial data to ensure interoperability must correspond to the following requirements:
- to have publicly available documentation describing the data format;
- the description must be sufficient to enable existing converters to be used or new data converters to be developed in open exchange formats;
- in the case of submission of data in a format for which the documentation is not publicly available, such data must be accompanied by converters that provide conversion of data into open exchange formats.

The basic formats of vector data exchange in NSDI define the following open and neutral to GIS platforms formats:
- formats based on the use of GML following national standard DSTU ISO 19136:2017 (ISO 19136: 2007, IDT);
- formats based on use, such as GeoJSON, which is an extension of JSON notation (JavaScript Object Notation).

These formats provide the most complete representation of the feature structure of spatial datasets, the topological relationships between features defined in application data schemas, and feature class catalogues as part of spatial data specifications.

3.5. Method for the quality evaluation of open geospatial data

The authors propose a modified method of Berners-Lee for open data according to the requirements for the creation of geospatial datasets for NSDI (tab. 1).

The method of the quality evaluation of open geospatial data takes also into the consideration the international standard ISO 19157 “Geographic information. Data quality”. This method is to determine the number of points or level (maximum – 5, as in Berners-Lee):
- 1st level – to make available data that contain geo-identifiers, such as COATUU, COATTG, name of the settlement, address, kilometers mark for roads, etc., on the Internet (regardless of the data format, representation of the geometry of feature) under an open license.
- 2nd level – to make data available in a structured form (Excel spreadsheet instead of a scanned document or figure), which contains the geometry of features, representation in the form of coordinates in the coordinate system registered in the EPSG register; to ensure the data publication using geographic information services (WMS, WFS, WCS, CSW, etc.).
- 3rd level – to use non-proprietary formats (CSV, JSON, XML) that contain the geometry of features in OGC formats (WKT, WKB, GML, KML, GeoJSON, etc.) in the coordinate system that registered in the EPSG Geodetic Parameter Dataset (register); to provide the ability to transfer the feature ID value to a URL for reference it in other systems.
- 4th level – to use URIs to indicate the entity so that users can reference them, where the geometry of the features should be reduced to levels 1 to 2 of the topology.
- 5th level – to link the data to others and to provide a context where the geometry of feature should be reduced to level 3 topology.

3.6. The quality evaluation of open geospatial data for Ukraine

The research was performed for the evaluation of open geoinformation resources for the creation of geospatial datasets, as defined in the Law of Ukraine on NSDI (Kin & Lazorenko-Hevel, 2021). The article presents the example of the quality evaluation of geospatial data for datasets “State Geodetic Reference Coordinate System UCS-2000”, “State Geodetic Network”, “Geographical Names”, “Administrative units” because they are one of the components of the Core Reference Dataset of NSDI (tab. 2). These geospatial datasets are separate, as well as part of the Digital Topographic Map of Ukraine at a scale of 1: 100,000 (fig. 1).
Table 2. The example of evaluation of open geoinformation resources for the creation of the geospatial datasets

| №  | The geospatial dataset                                      | The geoinformation resources for the creation of the geospatial dataset | The quality evaluation of open geospatial data | URL                                                                 |
|----|------------------------------------------------------------|------------------------------------------------------------------------|---------------------------------------------|----------------------------------------------------------------------|
| 1  | State Geodetic Reference Coordinate System UCS-2000       | Geodetic databank                                                      | ★★★                                        | 2 https://dgm.gki.com.ua/dgm_punkt_list/205                           |
| 2  | State geodetic network                                     | Geodetic databank                                                      | ★★★                                        | 2 https://dgm.gki.com.ua/dgm_punkt_list/205                           |
| 3  | Geographical names                                        | State Register of Geographical Names                                    | ★                                           | 1 https://land.gov.ua/info/informatsiia-pro-derzhavnyi-reiestr-heohrafichnykh-nazv/ |
|    |                                                            | Regular Reference Map of Ukraine at the scale 1:100,000                | Unavailable                                | 0 Unavailable                                                        |
| 4  | Administrative units                                      | Reference dataset of administrative units                               | ★★★                                        | 4 https://atu.gki.com.ua/ua/karta                                    |
|    |                                                            | State Register of Geographical Names                                    | ★                                           | 1 https://land.gov.ua/info/informatsiia-pro-derzhavnyi-reiestr-heohrafichnykh-nazv/ |
|    |                                                            | State Land Cadastre                                                    | ★                                           | 1 https://map.land.gov.ua                                            |
|    |                                                            | Regular Reference Map of Ukraine at the scale 1:100,000                | Unavailable                                | 0 Unavailable                                                        |

The authors evaluated the quality of 142 open geoinformation resources and other information resources (materials) for the creation and updating of 34 geospatial datasets for the development of NSDI in Ukraine (fig. 2).

According to the results of the research it was found that most open resources correspond to the first level of quality (95 resources out of 142). The 20 resources were not available on the Internet, but are limited to data owners. The only one resource corresponds to the fourth level of quality of open geospatial data. It is the data of the geoportal "Administrative units of Ukraine" (Ministry of Development of Communities and Territories of Ukraine, n.d.).

4. Conclusions

The geospatial knowledge infrastructure depends on the following of the principles of NSDI, the quality of implementation and realization of this policy in the state. Geospatial knowledge will have its value, using qualitative and official data from national data owners (authorities), which must ensure the relevance and reliability of this data, as well as their metadata.

The preliminary results of research on open data and volunteer GIS are an example of the feasibility and relevance of the proposed method of evaluating the quality of open geospatial data for the creation and updating of geospatial datasets in NSDI.

The proposed method allows evaluating the current state of geospatial data sets in NSDI and takes into account the quality of geospatial data, which also must meet the requirements for their production, updating and publication.
The method for the quality evaluation of open geospatial data for creation and updating...

Fig. 1. The Digital Topographic Map of Ukraine at a scale of 1:100,000 (reduced);
https://dgm.gki.com.ua/map/shema-dgm#map=6/47.98256841921405/32.01416015625001&layer=181154
1161957918378-1,100/1957486367818647401-0,100/12052842560826521-0,100/12058222771252388-0,100

Fig. 2. Results of the quality evaluation of open geospatial data for Ukraine
Literature

Berners-Lee, T. (2009). Linked data – design issues. Retrieved May 31 2021 from https://www.w3.org/DesignIssues/LinkedData.html

Corti, P., Kralidis, A. T., & Lewis, B. (2018). Enhancing discovery in spatial data infrastructures using a search engine. PeerJ. Computer Science, 4, e152. https://doi.org/10.7717/peerj-cs.152

Devillers, R., Stein, A., Bédard, Y., Chrisman, N., Fisher, P., & Shi, W. (2010). Thirty years of research on spatial data quality: Achievements, failures, and opportunities. Transactions in GIS, 14(4), 387–400. https://doi.org/10.1111/j.1467-9671.2010.01212.x

Geospatial World and the United Nations Statistics Division. (2020). Geospatial Knowledge Infrastructure White Paper. Retrieved May 31 2021 from https://www.geospatialworld.net/gw-assets/pdf/GKI-White-Paper.pdf

Hamad, S. (2020). Updating topographic maps at scale 1:250000 for Libyan territory using quantum gis (qgis) and open geospatial data: Libya topo-project. Journal of Geographical Studies, 4(1), 22–34. https://doi.org/10.21523/gcj5.20040103

Hanečák, P., Schleidt, K., & Tomáš Kliment. (2017). SHMU INSPIRE & OpenData. DanubeHack 3.0. https://doi.org/10.13140/RG.2.2.17939.37926

Janowicz, K., Hitzler, P., Adams, B., Kolas, D., & Vardeman II, C. (2014). Five stars of linked data vocabulary use. Semantic Web, 5(3), 173–176. https://doi.org/10.3233/SW-140135

Karpinskyi, Y., & Liashchenko, A. (2006). Strategy for the formation of the national infrastructure of geospatial data in Ukraine. NDIKH.

Karpinskyi, Y., Lazorenko-Hevel, N., & Kin, D. (2020). Inspireid implementation in the topographic database of the main state topographic map of Ukraine. ISTCGCAP “Geodesy, Cartography and Aerial Photography”, 91(91), 20–27. https://doi.org/10.23939/istcgcapi2020.91.020

Kin, D., & Lazorenko-Hevel, N. (2021). Quality assessment of geospatial datasets NSDI in Ukraine: (Version 1.0) [Data set]. Retrieved May 31 2021 from Zenodo. https://doi.org/10.5281/zenodo.4892847

Lazorenko-Hevel, N. (2017). EuroSDR and Eurogeographics initiatives by the use of modern methods of geospatial data collection for topographic cartography. Engineering Geodesy, 64, 67–77.

Lindgren, T., Ekenberg, L., Nouri, J., Talantsev, A., & Hansson, K. (2016). Open government ideologies in post-soviet countries. International Journal of Electronic Governance, 8(3), Article 10002287, 244. https://doi.org/10.1504/IJEVG.2016.10002287

Ministry of Development of Communities and Territories of Ukraine, State Service of Ukraine for Geodesy, Cartography and Cadastre. Geoportal of the administrative-territorial structure of Ukraine. Retrieved May 31 2021 from https://atu.gki.com.ua/

Open Knowledge International. (2017). Global open data index. Retrieved May 31 2021 from https://index.okfn.org/

Štrobl, J., & Nazarkulova, A. (2014). Open geospatial data: New opportunities for gis and giscience in central asia? In GISCA 2014, Urumqi, China.

The Law of Ukraine “On The National Spatial Data Infrastructure”, bill no. 2370, 13.04.2020. (2020). Retrieved May 31 2021 from http://w1.c1.rada.gov.ua/pls/zweb2/webproc34?id=&pf3511

The Order of Functioning of The National Spatial Data Infrastructure, no. 532, 26.05.2021. (2021). Retrieved May 31 2021 from https://zakon.rada.gov.ua/laws/show/532-2021-%D0%BF?lang=uk#Text

UN-GGIM. (2019). Integrated Geospatial Information Framework (IGIF). A Strategic guide to develop and strengthen national geospatial information management.: Part 2: Implementation Guide. Retrieved May 31 2021 from https://ggim.un.org/IGIF/part2.cshtml

United Nations Department of Economic and Social Affairs. (2019). The Global Fundamental Geospatial Data Themes. Retrieved May 31 2021 from https://ggim.un.org/documents/Fundamental%20Data%20Publication.pdf

Usery, E. L., Varanka, D. E., & Davis, L. R. (2018). Topographic mapping evolution: From field and photographically collected data to gis production and linked open data. The Cartographic Journal, 55(4), 378–390. https://doi.org/10.1080/00087041.2018.1539555