



This document is a **general guide for quality checking geospatial vector datasets to be published in the UTU Geospatial data service**. Quality of each dataset should be checked carefully prior sharing the data for public use – high-quality data is a corner stone for reliable further data usage.

Due to varying nature of geospatial datasets, many of the steps in this guide must be applied case-specifically.



## METADATA

### 1. All compulsory metadata is filled

*Check the completeness of the metadata. At least all compulsory metadata fields should be filled in in a satisfactory manner. Use the [Metadata instruction \(PDF\)](#) document as a reference.*



<b>Title</b>	<b>Topic category</b>	<b>Data quality statement</b>
<b>Abstract</b>	<b>Language</b>	<b>Restrictions</b>
<b>Keywords</b>	<b>License</b>	<b>Supplemental information (raster)</b>
<b>Date of creation</b>	<b>Regions</b>	<b>Responsible parties</b>

### 2. Metadata is consistent with the data

*Check the consistency of the metadata by comparing the metadata information to the data itself. For example, does the coordinate system, regions and cell size stated in the metadata match with the data? Utilize the **Properties** function of a GIS software.*



## GEOMETRIC QUALITY

### 3. Spatial geometry of the dataset has good connectivity

*Run a topology check for the dataset. Different types of vector data have different connectivity rules, which can vary case-specifically. Most often used connectivity rules determines that the geometries must not be invalid, and there should not be gaps or overlapping between individual objects.*

*A quick guide to QGIS Topology checker –tool can be found from [QGIS documentation](#). More comprehensive instruction can be downloaded [here](#).*



#### 4. Absolute locations of the features are accurate and precise



*Visually check that the vector features have accurate and precise absolute locations. Compare the features to a high-resolution third-party satellite image and visually evaluate whether the objects are where they should be, and that the geometries of the features are precise enough. This means, for example, that a feature representing a road should follow the real-world roadway, and not cut through buildings or water areas.*

*Checking absolute locations of vector objects is case-specific, and possible only for those datasets where the features can be found from a reference data (i.e., satellite imagery). For example, municipality borders are impossible to find from a satellite imagery, and thus absolute locations of features (municipalities) cannot be evaluated, and this step must be skipped.*

*For a dataset with multiple features, use sampling to ease the work.*



## ATTRIBUTE QUALITY

#### 5. Attributes of the data set are relevant



*Evaluating the relevancy of attributes is challenging, since the relevancy is always case-specific. Use your own expertise in the evaluation process.*

#### 6. Attribute table is complete



*Check the attribute table for empty cells and empty fields. Both can be done by using e.g. the **Show statistical summary** -tool of QGIS.*

*If empty cells can be found, look for explanation for them in the metadata or attribute metadata – high-quality metadata should explain the reason for NULL values.*

#### 7. There are no unwanted duplicates in the attribute table



*Check the attribute table for duplicate values in fields that should only consist of unique values, such as the FID-field. Use **DB Manager** in QGIS to find duplicate values, download instructions [here](#).*

#### 8. All compulsory attribute metadata is filled



*Go through the attribute metadata and check that all attributes have description and a label.*



## TEMPORAL QUALITY

### 9. Metadata of the dataset indicates when the data was produced and updated (if relevant)



*Information of temporal aspects of the data is crucial when data users evaluates the data set's fitness for their user activities, and therefore it is important to double-check the temporal information is provided correctly. Check that the sections Date, Edition, Temporal extent and Maintenance frequency are filled in properly (however, note that not all of these might be relevant for all cases). Also, look for more details from the Abstract or Data quality statement about the time period when the data was collected, or recommendations for updating frequency.*



## VISUAL QUALITY

### 10. Visualisation is comprehensible



*Assess whether the visualization of the dataset is understandable. Ponder, for example, is the theme or subject of the data possible to understand without too much effort? Are the colors or symbols intuitive? Are the objects of the data set distinguishable? Is the visualization appropriate for the data type? Evaluation of the visualization is rather subjective but use your expertise and trust your own view.*

### 11. Visualisation is appropriate



*Evaluate whether the visualisation gives an honest image of the dataset, and the data hasn't been manipulated via choice of colors or classifications.*



## FILE FORMAT

### 12. Data is in acceptable format



*Check that the vector data files are in acceptable machine-readable format, ready to be uploaded to the service:*

- *ESRI shapefile, Zipped shapefile, GeoJSON, or GML*