UNIVERSITY OF TURKU

UTU Geospatial data service

DATA QUALITY CHECKING
RASTER

This document is a **general guide for quality checking geospatial raster 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.

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

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 resolution is correct

Check that the spatial resolution (cell size) of the data matches with the information provided in the metadata. If this information is not provided in the metadata, add it to the Supplemental information *-section.*

You may also evaluate how well the cell size is able to project the phenomenon the raster represents: is the cell size small enough to detect and pass on the desired information?

4. Absolute location of the data is accurate and precise

☑

If the raster layer has values that can be detected from a satellite imagery (or from another reference data) compare some of the most distinct values in the raster to your reference. Visually evaluate whether the values are truthful and that they have accurate and precise absolute locations.

For example, a raster cell of a LULC data that represents forests should lay on such area that is a forest in the reality as well. Use your own judgement when evaluating whether this step can be done to the dataset in your hands.

Check at least 10 different locations to get a good understanding of the positional quality of the dataset.

5. Geometries are complete and there are no holes

☑

Visually check that the raster layer has no unwanted holes. The layer should be geometrically complete in the area of interest, i.e. all areas of interest should be covered with raster cells with corresponding values.

VALUE QUALITY

6. Value range is appropriate

☑

Check that the actual value range of the raster data matches with the range information given in metadata. If the metadata states that values range between 0 and 200 meters (e.g. in a Digital Elevation Model), no negative values should be detected. ***Raster layer statistics*** *in QGIS is a handy tool for this.*

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).

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 raster data is in acceptable machine-readable format, ready to be uploaded to the service:

- *GeoTIFF, GZIP*

☑