Flexible Access to a Harmonised Multi-resolution Raster Geodata Storage in the Cloud

Lassi Lehto, Jaakko Kähkönen, Juha Oksanen and Tapani Sarjakoski
Finnish Geospatial Research Institute (FGI)
National Land Survey of Finland
Masala, Finland
e-mail: lassi.lehto@nls.fi, jaakko.kahkonen@nls.fi, juha.oksanen@nls.fi, tapani.sarjakoski@nls.fi

***Abstract*—A viable approach for tackling the challenges of integration and analysis of geospatial raster data is to pre-process datasets into a common framework and store them into a cloud repository, accessible through a set of well-defined access protocols. This paper describes an initiative called GeoCubes Finland, where the aim is to provide a number of country-wide raster geodatasets in a common schema. In addition to more traditional access methods, a custom Application Programming Interface (API) has been designed for supporting the various tasks related to retrieval, use, visualisation and analysis of the contained raster datasets.**

Keywords—raster data; multi-resolution; harmonisation; cloud service; RESTful access.

I. INTRODUCTION

Geospatial datasets are increasingly being managed in cloud service platforms. Improved performance of wired and mobile networks has made it a viable approach to centralise data maintenance procedures. At the same time, one can recognize a steady shift from simple data file downloads to the use of flexible content access APIs. Similarly, a gradual shift can be noticed from data analysis run on a local computer to the use of centralised computing resources, possibly allowing access to High-Performance Computing (HPC) platforms with extensive parallel computing capabilities.

Raster-formatted geospatial datasets have qualities that ease the integration and analysis tasks considerably. Raster data is usually simple to manage and store. Effective parallelisation of the analysis problem is usually relatively straightforward for local and neighbourhood functions. Visualisation of raster-formatted content is effective.

A few challenges remain though. In many cases raster datasets are stored in individual spatial reference systems and in differing resolutions. It is often difficult to find out the explanation for the coded cell values, as there are no standardised mechanisms available for presenting this information to the user. In visualisation, it is often a challenging task to produce reliable and visually pleasing representations for small scales. In analysis processes, it might be difficult to achieve consistent results across a range of resolution levels.

One possible approach for tackling the challenges related to the integration, analysis and visualisation of geospatial raster data is to pre-process the datasets into a common harmonised framework and store the resulting representations into a cloud platform, accessible through a set of well-defined access protocols. A development aimed at building this kind of raster cloud storage is described in this paper. The initiative is called GeoCubes Finland and is carried out by a consortium of Finnish Universities and governmental research institutions [1]. The predominant goal of the initiative is to facilitate the use of raster-formatted geospatial datasets in academic research.

The paper is organized as follows. In Section II the existing approaches comparable with the proposes access API are described. In Sections III and IV harmonization aspects and implementation details of the data storage are discussed. Section V deals with access methods generally and Section VI details the developed custom API. In Section VII the applications of the proposed multi-resolution approach are discussed and Section VIII concludes the paper.

II. EXISTING APPROACHES

The predominant standardised access mechanism for raster datasets in network service-based architecture is the Web Coverage Service (WCS) interface specification of the Open Geospatial Consortium (OGC) [2]. A WCS is supposed to provide access to raster datasets via its GetCoverage operation. This operation allows the calling application to indicate the requested coverage by name, limit the requested area by a bounding box, and optionally ask the service to produce the result in certain resolution by setting the SCALEFACTOR query parameter. For performance reasons, in many WCS implementations the maximum allowed size of the requested dataset is set to a rather low value.

III. STORAGE - HARMONISATION, MULTI-RESOLUTION

In the GeoCubes Finland raster data storage, the individual content layers are pre-processed during the ingestion process into the common spatial reference grid (the national standard grid). GeoCubes Finland is inherently a multi-resolution data storage. A fixed set of resolution levels (10 values) has been selected and all ingested datasets are pre-processed into those resolutions. Individual, dataset-specific generalisation procedures, suggested by the original data provider, are applied in the process. This way the best possible consistency among the resolution levels can be achieved.

 ISBN: 978-1-61208-687-3

IV. Implementation in Cloud-Optimized GeoTIFF

GeoCubes Finland data storage is implemented in the form of Cloud-Optimized GeoTIFF (COG) files (TIFF: Tagged Image File Format), each representing a 100 km * 100 km block [3]. The area of the country is divided into 60 such blocks. The resolution levels are stored both as internal GeoTIFF overview layers and as individual resolution-specific GeoTIFF files. The set of 60 blocks is aggregated into a single content representation by using the GDAL's (Geospatial Data Abstraction Library) Virtual format (VRT) mechanism [4]. VRT files also combine together the files on individual resolution levels. GDAL's Python API is extensively used in the data ingestion and data provision procedures. Parallelisation of the computing processes is done applying Python's subprocess mechanisms.

V. Access methods - HTTP GET Range, VRT, REST

A set of different access methods are supported in the GeoCubes Finland's raster data repository. The traditional standardised methods for raster data access and visualisation are supported. These include OGC-specified interfaces WCS, Web Map Service (WMS) and Web Map Tile Service (WMTS). Individual block-wise GeoTIFF files can be accessed using a standardised URL (Uniform Resource Locator) scheme and the conventional HTTP (Hypertext Transfer Protocol) based data transfer. In addition to that, a HTTP GET Range request can be applied for partial file downloads. This process is efficient, because of the optimised organisation of the contents in the COG files.

The GDAL Virtual format (VRT) mechanism is used to combine together the 60 blocks covering the country. VRT is a light-weight XML (Extensible Markup Language) formatted text file describing the parts that belong to the merged dataset. Data transmission is optimised, as only the actually needed part of the raster data content - on the requested resolution level - is transferred over the network.

A Django-based service platform has been developed to facilitate flexible access to the GeoCubes content. A RESTful API (REST: Representational State Transfer) supports the various tasks related to the retrieval, use, visualisation and analysis of the contained raster datasets.

VI. RESTful access API

The designed RESTful API is based on the following general semantic structure of the path components:

```
/ what to do / on which resolution level
/ with which content layer / where / how
```

The API contains operations for accessing basic metadata of the contained raster layers. Based on the provided information, the client application can find out the names of the theme layers in the storage and form the URLs needed for accessing the individual files.

The API contains operations for efficient and flexible retrieval of raster content. The area of interest can be selected by a bounding box, a set of block identifiers, and by a list of codes or names of administrative units of the country (three administrative levels supported). The layer contents can be downloaded either as GeoTIFF content or as a VRT file. When using the VRT alternative, the user can determine, if only the requested resolution level should be returned or if all the coarser levels have also to be included.

Once retrieved, one of the first tasks in using a raster dataset is to understand the meaning of the cell values. In some cases, it is rather straightforward to interpret the values, for instance, if data represents a continuous variable. However, for categorical data, it is often difficult to find meaning for the values. In the GeoCubes API, there is an operation for finding the explanation for a given code value. It is also possible to ask for the category description by layer name, resolution level and a coordinate point. This enables dynamic lookup of category information by moving the cursor on top of a visualised map.

The API also contains demonstrative examples of analysis functions performed on server side, using GeoCubes content layers as input data. Examples include value distribution calculations by administrative unit, change detection between two epochs of datasets with time series, and aggregation between theme layers.

VII. Applications of multi-resolution storage

There are some recognised use cases, where the multi-resolution approach of the GeoCubes Finland repository can be utilised. Sometimes an analysis process has to be carried out on certain resolution level, because another input data set is only available on that resolution. In addition, there are long-standing conventions for analysing certain phenomenon on a specific resolution level. Furthermore, while developing an analysis procedure, it is often useful to first test the analysis on coarser levels of resolution, before launching the real long-running procedure on a fine-grained resolution level.

An interesting application of multi-resolution raster data storage is a geospatial analysis task, where results are explored visually in the form of a map. When the user explores the resulting visualisation in various zoom levels, the background analysis can always be interactively run, utilising data from the corresponding resolution level in the data storage. This way the analysis can be run in roughly constant time over the whole range of the visualisation scale. This approach makes it also possible to configure a visualisation of an analysis result as a new content layer for the data storage.

VIII. Conclusion

GeoCubes Finland is a new initiative to provide academic sector users with an easy-to-use raster geodata repository in the cloud, with a set of flexible content access methods and support for server-side analysis procedures. Modern raster data management techniques, like Cloud-Optimized GeoTIFF, HTTP GET Range protocol and GDAL Virtual Format, are applied in the data access process. A new RESTful API has been designed for facilitating data retrieval, interpretation and analysis. The data repository is currently under construction. First user tests will be carried out in the next coming months.

 ISBN: 978-1-61208-687-3

ACKNOWLEDGMENT

The work described in this paper has been carried out in the context of the project 'Open Geospatial Information Infrastructure for Research' (oGIIR, urn:nbn:fi:research-infras-2016072513), a part of Finland's Roadmap for Research Infrastructures. The project is funded by the Academy of Finland, grant number 306536. The computing infrastructure used in the work is funded by the Academy of Finland through 'Finnish Grid and Cloud Infrastructure', urn:nbn:fi:research-infras-2016072533, grant number 283818.

REFERENCES

[1] oGIIR, Open Geospatial Information Infrastructure. http://ogiir.fi [retrieved: Jan, 2018]

[2] OGC, Web Coverage Service. http://www.opengeospatial.org /standards/wcs [retrieved: Jan, 2018]

[3] GeoTIFF, GeoTIFF home page. http://trac.osgeo.org/geotiff/ [retrieved: Jan, 2018]

[4] GDAL, Geospatial Data Abstraction Library. http://gdal.org [retrieved: Jan, 2018]

 ISBN: 978-1-61208-687-3