

Investigations towards enabling a Web-based environmental geospatial information system (Web-EGIS) in EnviDat

Ionuț Iosifescu Enescu^{a, *}, David Hanimann^a, Dominik Haas-Artho^a, Marius Rüetschi^a, Dirk Nikolaus Karger^a, Gian-Kasper Plattner^a, Martin Hägeli^a, Rebecca Kurup Buchholz^a, Lucia de Espona^a, Niklaus E. Zimmermann^a, Loïc Pellissier^{a, b}

^a Swiss Federal Research Institute WSL, ionut.iosifescu@wsl.ch, david.hanimann@wsl.ch, dominik.haas@wsl.ch, marius.rueetschi@wsl.ch, dirk.karger@wsl.ch, gian-kasper.plattner@wsl.ch, martin.haegeli@wsl.ch, rebecca.kurup@wsl.ch, lucia.espona@wsl.ch, niklaus.zimmermann@wsl.ch

^b ETH Zurich, loic.pellissier@usys.ethz.ch

* Corresponding author

Keywords: Web-EGIS, time series, environmental raster data, very large geodata, gridded data, HTML5, EnviDat

Abstract:

Web-based environmental geospatial information systems (Web-EGIS) are a distinctive case of Web-GIS. The purpose behind Web-EGIS is to integrate the functionalities of geographic information systems (GIS) with large numbers of environmental geo-referenced data and services, using generic hypercube-based data organization and visualization as previously described in Iosifescu et al. (2015). The user requirements for basic Web-EGIS functionalities are: (1) automated display of the environmental data contents on an interactive 2D map and/or on a 3D globe, (2) spatial, thematic and temporal navigation through a large number of environmental data layers of a time series, and (3) extracting and displaying data subsets over an entire time series for any interactively selected location on the map.

Climate model outputs pose significant challenges for a functional and performant implementation of such a Web-EGIS, since many terabytes of georeferenced raster data are being produced in complex directory structures containing a large number of individual large files. Examples are the “High resolution climate data for Europe” (<https://www.envidat.ch/#/metadata/eur11>) or the “Climatologies at high resolution for the Earth’s land surface areas (CHELSA)” (<https://www.envidat.ch/#/metadata/chelsa-climatologies>) that contain gridded time-series with many thousands of layers for a single climate variable. CHELSA provides free climate model data such as temperature and precipitation patterns at 1km resolution for various time periods as documented in Karger et al. (2017). Similarly, time series of remote sensing data such as the ones produced by Landsat and Sentinel missions are continuously growing in size. In contrast to the climate model outputs, the remote sensing data contain multiple high resolution data bands in the same file (organized as a data cube), thus further increasing the amount of storage required.

Currently, such massive numbers of environmental data layers cannot be efficiently configured and interactively delivered by traditional Spatial Data Infrastructures (SDIs) and conventional Web Map Services (WMS). Furthermore, sampling, clipping and subsetting smaller areas of interest from larger global datasets would take a significant amount of time with an SDI’s standard geoprocessing services. This prevents any real-time parallel interactions with a multitude of concurrent users, as required by interactive Web applications. Finally, well-known Web-mapping frameworks are not adapted for providing user-friendly navigation along a time axis for such massive numbers of data layers.

Possible solutions to the above challenges emerge when trying to take advantage of: (a) the intrinsic characteristics of such large gridded timeseries, (b) the use of modern video encoding and decoding algorithms for compressing gridded data and (c) the distribution of geospatial processing to the client’s Web browser through modern HTML5 features.

The intrinsic characteristic of gridded data timeseries is that the individual data layers cover the same area of the Earth along an evenly spaced matrix/grid. Therefore, if we consider the timeseries a whole, all data layers have an identical spatial extent and resolution. These characteristics can be used to better compress the time series in its entirety instead of storing a series of independent data layers, thus substantially reducing the overall data size.

Changes in environmental variables and their spatial parameters can be considered a series of successive individual frames like in a movie. Modern video codecs take advantage of these intrinsic characteristics to only store specific keyframes and then perform motion compensation, by only storing and compressing the differences between successive frames. Therefore, the same principles of encoding and decoding video as implemented in modern codecs such as the open and royalty-free AV1 codec can also be applied to gridded data in order to drastically reduce the storage and transmission size of massive time series.