SIDS GIS DATA CATALOGUE

RISING UP FOR SIDS

August 2022

Small Island Developing States

INTRODUCTION

As most of the decisions within governmental organizations and industries become increasingly data driven, there exists a plethora of data that grows exponentially every day. Thus, there exists a perennial quest for new sources of data and means of data analysis and visualization. Geospatial data plays a key role when the rendering of information is in question. Since the location is found associated with the attribute information, large sets of spatial data are gleaned from many diverse sources in varying formats. The additional location context makes it content-rich and provides one with better insights that may otherwise be difficult to infer.

In the recent years, technological advancements in data acquisition systems have made a majority of the geospatial technology affordable and accessible to the global community. The sheer volume of this geospatial data necessitates specialized knowledge and expertise for pre-processing. Unless the raw geospatial data is pre-processed, it is not deemed fit for targeted use.

In addition to this, there has been a significant movement in recent years within academic research to promote open data. This new movement towards open data also presents many challenges in geospatial research, as the process of finding insights from this open data is not simple. For example, every dataset can vary in projection, resolution, file format, spatial alignment, and in other key ways that raise significant barriers to data scientists to extract meaningful insights. Open data especially presents its own challenges as well, as there is a varying degree of quality and metadata standards. These datasets are often spread across many different geospatial portals, hidden in academic journals, and distributed across the internet in a way that is rarely optimized for discovery or access.

Although there have been several portals designed to begin to address this issue such as OpenDRI, ResourceWatch, or Google Earth Engine, there has never been any effort to compile geospatial resources specifically focused on the SIDS. There is a definite need for a broad survey, compilation, and release of geospatial datasets that are global and publicly available with a coverage and focus on SIDS in order to provide cohesive data across each Sustainable Development Goal (SDG).

Hence, this data catalogue brought out by the SIDS team aims to fill this exact void. In total, there are 82 datasets, predominantly rasters, curated from various sources such as ResourceWatch, Socioeconomic Data and Applications Center (SEDAC), World Resources Institute (WRI). Datasets pertaining to a wide range of developmental aspects such as Agriculture, Infrastructure, Biodiversity, Ocean Health, Soil, Land Use/Land Cover find place in this assortment.

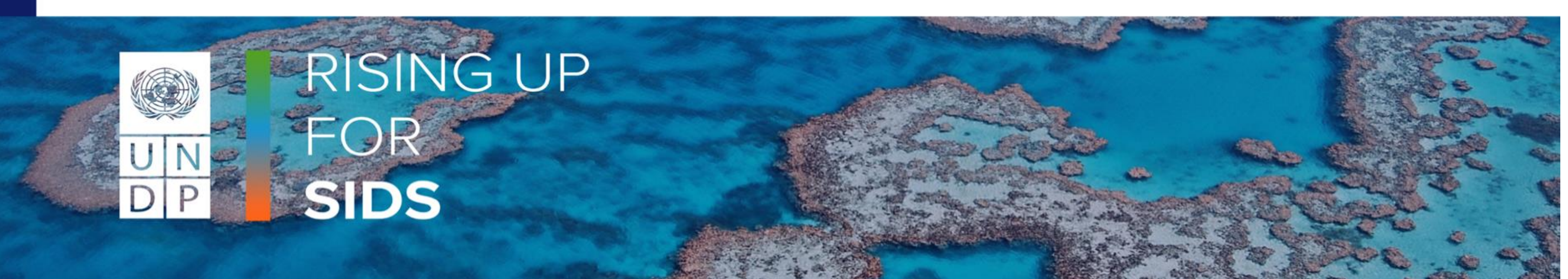
The raster datasets have been standardized to several resolutions (1km, 5km, and 10km) for hexagonal and square grids, as well as two levels of admin regions. Each of these raster datasets has been processed with our custom data pipeline which reprojects and resamples the original datasets using a zonal statistics method for these pre-generated grids. There are three basic functions that the pipeline does:

1. **Standardization**

Ensures that all the datasets are in a common projection system (EPSG:4326) with data clipped to the boundaries of SIDS.

2. **Zonal Statistics Computation**

Zonal statistics (mean and count) are computed for each feature in a given vector from every available raster.



3. Conversion to MapBox Vector Tiles

The results from zonal stats are added to the vector geometries in each layer in attribute columns. Depending on the provided arguments the attributes are either added to the same vector layer or new layers are created for every raster layer. Once this pre-processing is done, the datasets are exported as MapBox Vector Tiles (MVT).

The subsequent section comprises the list of datasets with relevant metadata.