

5. DEVELOPING A SPATIAL DATA FRAMEWORK

The data compiled in this report have been organized into a spatial data framework to ease the delivery of the data collected and created by NOAA-CREP to our partners at the Timor-Leste Ministry of Agriculture and Fisheries (MAF). In lieu of specific capacity building activities, this spatial data framework is designed specifically for anticipated users and increases its relevance in resource management. This spatial data framework should therefore accompany the data structure provided to MAF at the workshop led by NOAA-CREP in June 2017.

Timor-Leste Project Portal

The framework is hosted on NOAA's Coral Reef Information Service (CoRIS) website, in a 'project portal' established specifically for this project with Timor-Leste (<https://www.coris.noaa.gov/activities/projects/timor-leste/>). The portal includes 3 tabs: 1) Project Overview, 2) Report Download, and 3) Data Download.

Project Overview

The Project Overview page includes a brief summary of the USAID-NOAA partnership in Timor-Leste, with the option to download this complete report in Adobe Acrobat (PDF) format.

Report Download

The Report Download page provides the contents of this report by chapter, also in PDF format, allowing a user to view or download only the chapters and/or appendices of interest. The appendices are organized by the corresponding chapter rather than in sequential order as they appear in this report.

Data Download

The Data Download page includes all datasets that have been collected or created by NOAA-CREP for Timor-Leste as part of this project, and the data are organized by the corresponding data chapters (Chapters 2–4). See the *Data Structure* section below for a detailed list and description of the datasets provided.

Data Documentation and Archival

Each dataset has been fully described in the NOAA Fisheries 'InPort' Enterprise Data Management Program, an online metadata (data documentation) catalog and repository (<https://inport.nmfs.noaa.gov/inport>). Additionally, to preserve the data in perpetuity, the data are archived and accessible online at the NOAA National Centers for Environmental Information. The data archive for each dataset is linked to the corresponding metadata record in the InPort catalog. Links for accessing the documentation in InPort and the data in the archive are included with each dataset in the project portal.

Data Structure

The data and information provided in the Timor-Leste project portal are organized similarly to the data structure delivered to MAF during the June 2017 workshop (Table 10).

Table 10. The folder structure of the data and information provided to MAF.

Folder name and hierarchy	Folder description
\TIMOR-LESTE\	The root folder of the entire data structure.
\Report\ \Full report\ \Report contents\	Contains the full report and the report contents in Adobe Acrobat format (PDF). The contents in these folders correspond to the contents available in the Project Overview and Report Download pages of the Timor-Leste project portal.
\Data\ \Satellite Mapping\ \Ecosystem Assessments\ \Climate Change Baselines\	Contains all datasets collected or created by NOAA-CREP and organized by the data chapters from the report. This corresponds to the Data Download page of the Timor-Leste project portal. See Table 11 for specific details.

Table 11 shows a list and description of the datasets available within the data structure, including the associated file formats and links to metadata for each dataset. A detailed description of the satellite mapping datasets is given in the *Satellite Mapping Data* section below, as these datasets have a complex data structure compared with the Ecosystem Assessment and Climate Change Baseline datasets.

Table 11. A list of the datasets available in the data structure and the Timor-Leste project portal. The folders are listed by chapter sequence from the report rather than in alphabetical order as found in the data structure. *The raw and georeferenced satellite imagery is not available on the Timor-Leste Project Portal because the DigitalGlobe license agreement prohibits public distribution of the source imagery (i.e., posting the imagery online is not allowed).

Folder name and hierarchy	Folder description
\Satellite Mapping\ \Image Catalog\ \Ground Truth\ \Raw Imagery\ \Georeferenced Imagery\ \Bathymetry\	Contains the satellite mapping datasets described in Chapter 2.
	Inventory of WorldView-2 satellite images purchased, as well as the image footprints and boundaries, and the regions of interest used to define the geographic areas to acquire the satellite images (Figure 3). Format: .GDB and .SHP Metadata: https://inport.nmfs.noaa.gov/inport/item/46151
	Ground-truth data collected by NOAA-CREP used to validate the depths derived from the WorldView-2 imagery (Figure 4). Format: .SHP Metadata: https://inport.nmfs.noaa.gov/inport/item/25307
	Raw WorldView-2 satellite imagery provided by DigitalGlobe, including the supporting metadata files (Appendix B). Format: .TIF
	WorldView-2 satellite imagery that was georeferenced to the ESRI basemap (Appendix B). Format: .TIF
	Bathymetry data derived from the WorldView-2 satellite imagery (Figure 5). Format: .TIF Metadata: https://inport.nmfs.noaa.gov/inport/item/46150

\Benthic Habitat\	Benthic habitat data derived from the WorldView-2 satellite imagery (Figure 6). Format: .TIF Metadata: https://inport.nmfs.noaa.gov/inport/item/29128
\Ecosystem Assessments\	Contains the datasets from the coral reef ecosystem assessment surveys in 2013 described in Chapter 3.
\Fish Surveys\	Reef fish survey data (Figure 8). Format: .CSV and .SHP Metadata: https://inport.nmfs.noaa.gov/inport/item/32998
\Benthic Images\	Benthic photographs collected during the fish surveys. Format: .JPG Metadata: https://inport.nmfs.noaa.gov/inport/item/46160
\Benthic Cover\	Benthic cover data derived from the analysis of the benthic images collected during the fish surveys (Figure 16). Format: .CSV and .SHP Metadata: https://inport.nmfs.noaa.gov/inport/item/46161
\Climate Change Baselines\	Contains the baseline datasets collected from the Climate Monitoring sites from 2012 to 2014 described in Chapter 4.
\Temperature\	Temperature data from STRs (Figure 27). Format: .CSV and .SHP Metadata: https://inport.nmfs.noaa.gov/inport/item/46164
\Seawater Chemistry\	Seawater chemistry data from seawater samples (Figure 28). Format: .CSV and .SHP Metadata: https://inport.nmfs.noaa.gov/inport/item/46163
\Calcification Rates\	Calcification rate data from the CAUs (Figure 31). Format: .CSV and .SHP Metadata: https://inport.nmfs.noaa.gov/inport/item/46162
\Biodiversity\	Marine invertebrate specimen and sequenced data, and species and plate photographs from the ARMS (Figure 38). Format: .CSV, .SHP, .FASTQ, and .JPG Metadata: https://inport.nmfs.noaa.gov/inport/item/46159
\Benthic Images\	Benthic photographs collected from the Climate Monitoring sites in 2012 and 2014 (see benthic photograph collages in Appendix E). Format: .JPG and .SHP Metadata: https://inport.nmfs.noaa.gov/inport/item/46160
\Benthic Cover\	Benthic cover data derived from the analysis of the benthic images collected at the Climate Monitoring sites in 2014 (Figure 43). Format: .CSV and .SHP Metadata: https://inport.nmfs.noaa.gov/inport/item/46161

The following file formats are included in the data structure (Table 11), along with links for more information about each format:

- ESRI File Geodatabase (.GDB), includes both vector and raster spatial data, <https://www.loc.gov/preservation/digital/formats/fdd/fdd000294.shtml>
- ESRI Shapefile (.SHP), vector spatial data, <https://www.loc.gov/preservation/digital/formats/fdd/fdd000280.shtml>
- GeoTiff (.TIF), raster spatial data, <https://www.loc.gov/preservation/digital/formats/fdd/fdd000279.shtml>

- JPEG Image Encoding Family (.JPG), optical data (e.g., photographs), <https://www.loc.gov/preservation/digital/formats/fdd/fdd000017.shtml>
- Comma Separated Values (.CSV), numerical/tabular data, <https://www.loc.gov/preservation/digital/formats/fdd/fdd000323.shtml>
- FASTQ (.FASTQ), raw sequence reads with corresponding quality scores, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2847217/>

Satellite Mapping Data

There are six folders (datasets) within the Satellite Mapping folder, including Image Catalog, Ground Truth, Raw Imagery, Georeferenced Imagery, Bathymetry, and Benthic Habitat. Below, we describe in more detail the data contents within each folder. The raw and georeferenced imagery, bathymetry, and benthic habitat folders contain four subfolders based on the regions of interest (Atauro, North shore, Oecusse, and South shore). Within the raw and georeferenced imagery folders, each region folder includes a subfolder for each image purchased, as each image is essentially a collection of files.

Folder: \TIMOR-LESTE\Data\Satellite Mapping\Image Catalog

The image catalog is provided as an ESRI file geodatabase (WV-2_Image_Catalog_Timor.GDB) and serves two purposes: 1) it is an inventory of all the images purchased by each region of interest (ROI), and 2) it contains features associated with the images, including the ROI, the boundary extent of the available images within each ROI, and the footprints of each image (Figure 44). The ESRI file geodatabase (.GDB) is a proprietary format for use with ArcGIS software; therefore, the datasets are also provided in shapefile format for use with other software programs.

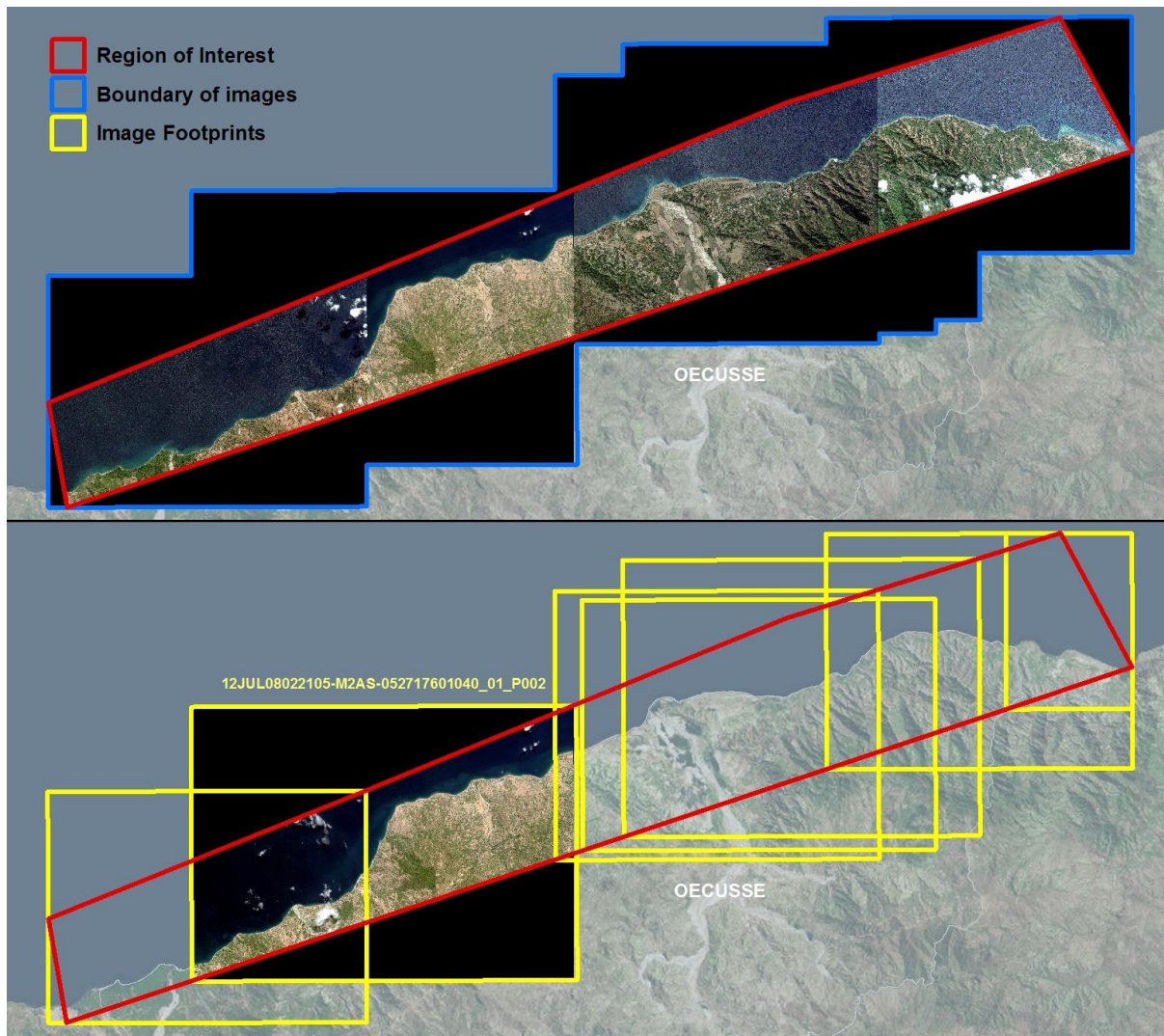


Figure 44. Maps showing the available features in the image catalog for the Oecusse region for the mosaic dataset (*top*) and individual images (*bottom*) including the region of interest (red), boundary of the available images (blue), and the footprints of the seven images (yellow) purchased for the Oecusse region.

The ROI (ROI in .GDB; WV-2_ROI_Timor.SHP) was used to define the desired geographic areas to purchase the satellite imagery from Digital Globe. A collection of raw images was acquired along the coasts of the identified four regions of interest (Atauro, North shore, Oecusse, and South shore). All images associated with each region are stored in what is called a mosaic dataset in the ESRI file geodatabase. Within each mosaic dataset is the boundary (extent) of the available images for each region (Boundary in GDB; WV-2_Image_Boundary_Timor.SHP) and the footprint (Footprint in GDB; WV-2_Image_Footprint_Timor.SHP) for each image within the boundary (Figure 44). Within GIS, the footprints can be used to quickly identify the image files that fall within a geographic area of interest. Furthermore, the associated attribute table indicates if the images were georeferenced and used to derive depths and habitat classes (Appendix B).

Folder: \TIMOR-LESTE\Data\Satellite Mapping\Ground Truth\

The ground-truth data include depth soundings collected by NOAA-CREP during the 2012 (WV-2_DepthSoundings_Timor_2012.SHP) and 2013 missions (WV-2_DepthSoundings_Timor_2013.SHP) in shapefile format. Soundings from 2012 were recorded every second and ranged in value from 0 – (-30) m. The 2013 soundings were recorded every 0.1–2 seconds and ranged in value from 0 – (-200) m. These data were used to validate the depths derived from the satellite imagery.

Folder: \TIMOR-LESTE\Data\Satellite Mapping\Raw Imagery\

Digital Globe provides an array of satellite products from a variety of sensors. The product that was purchased for this project was WorldView-2, multispectral (8 band), 16-bit, full swath imagery that was provided as a mosaic of individual images (sub-images) in GeoTIFF format, and the product type was Ortho Ready (level 2A) Standard (http://www.c-agg.org/cm_vault/files/docs/DigitalGlobe-Base-Product-FAQ.pdf). All images were projected to the Universal Transverse Mercator (UTM) coordinate system, zone 51S with 2-m pixel resolution.

Within the data structure, each image that was purchased has a corresponding Order ID folder that contains 1) the sub-images (.TIF), 2) a tile map file (.TIL), which is a stitched mosaic of the sub-images, and 3) the metadata files, which are collectively referred to as image support data (ISD; Figure 45).

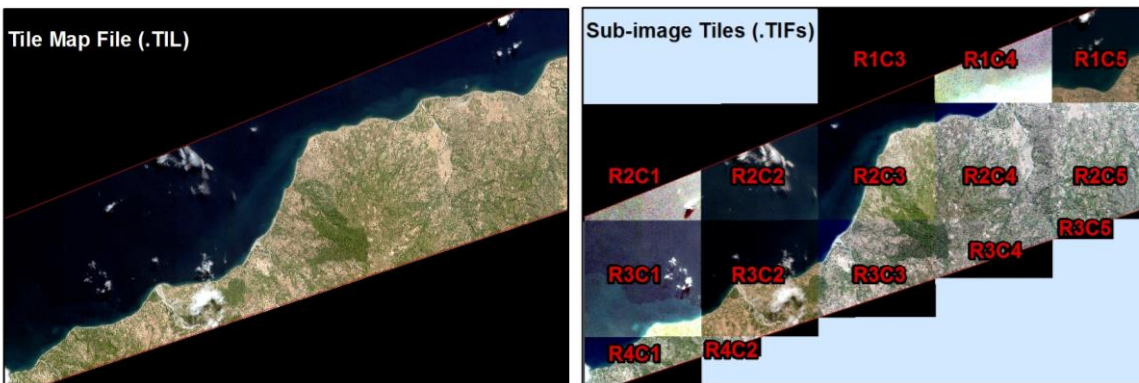


Figure 45. Example of a tile map file (.TIL) for Image ID 12JUL08022105-M2AS-052717601040_01_P002 (left) compared with the sub-image tiles (.TIF) for the same Image ID (right). The sub-image tiles are labeled by their respective row and column numbers.

The Order ID folder name is based on a unique ID assigned by Digital Globe that includes the **Delivery ID**, **Image Sequence**, and **Product Type**. The Delivery ID is a unique 15-character ID, the Image Sequence is the sequential number for each image in that delivery, and the Product Type is multispectral (MUL). For example:

Order ID: 052717601060_01_P001_MUL
Delivery ID: 052717601060_01
Image Sequence: P001
Product Type: MUL

The ISD files included with each image are listed in Table 12 along with an example. Refer to the Product Component-Level ISD documentation for further details about the naming conventions and ISD file descriptions (DigitalGlobe 2014).

Table 12. A list of the image support data (ISD) files provided by DigitalGlobe with each satellite image, and an example of each file name is provided for Image ID 11JUN07021210-M2AS-052717601060_01_P001.

Image Support Data File type	Example file name
Product Component Index Readme File (.TXT)	11JUN07021210-M2AS-052717601060_01_P001_README.TXT
License File (.TXT)	ENTERPRISE.TXT
Image Metadata File (.IMD)	11JUN07021210-M2AS-052717601060_01_P001.IMD
Product Browse File (.JPG)	11JUN07021210-M2AS-052717601060_01_P001-BROWSE.JPG
Tile Map File (.TIL)	11JUN07021210-M2AS-052717601060_01_P001.TIL
RPC00B File (.RPB)	11JUN07021210-M2AS-052717601060_01_P001.RPB
XML File (.XML)	11JUN07021210-M2AS-052717601060_01_P001.XML

All ISD files are named by their Image ID with the exception of the license file. The Image ID is based on the Acquisition Time, Product Info, and Order ID. The Acquisition Time includes the 2-digit collection year, 3-character month, 2-digit day and 6-digit time and the Product Info includes the band, product, and image types. Order ID, as previously described, is included minus the “MUL” suffix. For example:

Image ID: 11JUN07021210-M2AS-052717601060_01_P001.TIF
 Acquisition Year: 11 (2011)
 Acquisition Month and Day: JUN07 (June 7th)
 Acquisition Time: 021210 (2:12 am, and 10 sec)
 Product Info: M2AS (multispectral band, standard product, single/sub-scene image)
 Order ID: 052717601060_01_P001

For the sub-image .TIF files, the file name includes the Image ID plus the row and column number of the sub-image in relation to the mosaic (.TIL). For example, sub-image 11JUN07021210-M2AS_R1C4-052717601060_01_P001.TIF is found in row 1, column 4 of the corresponding tile map file (Figure 45). For mapping and visualization, the Tile Map File (rather than the TIFs) is a more user-friendly format to use in GIS.

Folder: \TIMOR-LESTE\Data\Satellite Mapping\Georeferenced Imagery\

To improve the geolocation accuracy, the raw images were georeferenced to an ESRI basemap.

The folder structure and naming conventions of the georeferenced imagery are similar to the raw imagery, with the exception of the suffix ‘REC’ (i.e., rectified) appended to the Order ID folder and the Image ID. For example,

Order ID: 052717601060_01_P001_MUL_REC
 Image ID: 11JUN07021210-M2AS-052717601060_01_P001_REC.TIF

Only images that fulfilled the evaluation requirements were georeferenced and further processed for deriving depths or habitat classes. Typically, only the tile map files were georeferenced and saved as .TIF

to the Order ID folder; the sub-image files were only georeferenced as needed, in addition to or instead of the tile map file for depth or habitat derivation. The georeferenced images are provided in the WGS 1984 geographic coordinate system (i.e., unprojected).

Folder: \TIMOR-LESTE\Data\Satellite Mapping\Bathymetry

The final product for the bathymetry data is provided as a mosaic for each region in TIF format, the preferred format to use in GIS. The individual bathymetry files that make up the mosaics, which were derived directly from the georeferenced imagery, are also provided in TIF format. The mosaic datasets are named Bathymetry_mosaic_<REGION>.TIF. The name of each bathymetry file includes the acquisition date (collection month, day, and year) and the image sequence (e.g., 11JUN07_P001.TIF). The individual bathymetry files are projected using UTM coordinate system zone 51S. Mosaic datasets are provided in the same projected coordinate system. Additionally, unprojected versions of the mosaics are also provided (WGS 1984). The bathymetry product has a ‘floating point’ pixel type (i.e., cell values are numbers with decimals), which is appropriate for continuous data that represent surfaces such as the seafloor.

\Satellite Mapping\Benthic Habitat

The final product for the benthic habitat data is provided in the same manner as the bathymetry product (as described above), with the exception of the pixel type. The pixel type of the benthic habitat product is ‘unsigned integer’ (i.e., cell values are positive whole numbers); therefore, unlike the bathymetry, it is discrete data that stores the cell values and associated attributes (habitat class and types) in an attribute table. The cell values are codes that represent the 13 habitat classes as shown in Table 13.

Table 13. Numeric cell values in the attribute tables of the benthic habitat data and the corresponding description for the habitat classes and types.

Cell value	Class	Type
1	Hard shallow	Hard substrate
2	Soft shallow	Soft substrate
3	Hard mid	Hard substrate
4	Soft mid	Soft substrate
5	Hard deep	Hard substrate
6	Hard soft	Soft substrate
7	Seagrass	Seagrass
8	Unknown	Unknown
9	Mangrove	Mangrove
10	Intertidal	Intertidal
11	Emergent rocks	Emergent rocks
12	Macroalgae	Macroalgae
13	Lagoon	Lagoon