# Startup Instructions

## Contents



# 1 Introduction

Welcome to !repo.

!repo is repository: a folder of files and scripts required to run the Ocean Health Index (OHI) Toolbox and will be necessary to complete your regional assessment. Although the Toolbox is used to calculate Index scores after all data have been gathered and decisions have been made, it is important to understand the structure of the Toolbox file system, as this aids in bookkeeping, which is important for transparent and reproducible science. You will find that although data preparation and model modification will require knowledge of the software program R, most data management and bookkeeping within the Toolbox file system can be done without any programming knowledge.

Please refer to the **Ocean Health Index Toolbox Manual** for instructions on how to prepare to use the Toolbox, including how to download this repository using GitHub to facilitate collaboration.

# 2 !repo has been extracted from the global repository

This repository is a starting point for your assessment. Information for !study_area has been extracted from the global assessment and will serve as a template to modify to better represent the local charactistics of !study_area. This is to say that all data within this repository are based on global databases where !study_area is represented at the national scale.

With !repo, you will calculate scores for each region within !study_area and combine them into a single score for !study_area. You will be able to substitute much of the global data provided with locally-available data and indicators, and you will be able to modify the goal models to better reflect local data and priorities. This will provide a finer-scale understanding of ocean health in !study_area.

## 2.1 `!x` regions within !study_area

Whereas in the global assessments, !study_area was one region combined within the study area (the entire world), here, !study_area is the study area, which contains `!x` subcountry coastal regions: `!listed_here`.

These regions were identified by Global Administrative Areas (www.gadm.org) as being the largest subcountry unit within !study_area. The boundaries of these regions have been extended offshore to divide the exclusive economic zone (EEZ) of !study_area into portions for each subcountry region. These regions have been provided because in previously completed regional assessment, this is often the scale at which data are available and policy decisions are made. However, it is possible to use different regions than the ones provided here. See ohi-science.org/pages/create_regions.html for more details.

## 2.2 !repo online display

An online version of the Toolbox App display has been created for !study_area, showing the `!x` regions. This display is easy to share with your colleagues. You will find this at this url:

**https://ohi-science.shinyapps.io/!study_area/**

This website will display the Input Data and Output Scores from the !repo on GitHub: github.com/OHI-Science/!repo. Originally, these data are This displays the information on GitHub on a webpage that is easy to share.

# 3 File System

Within the !repo repository is **subcountry2014**, the scenario folder. This contains all the data, functions and other files required to calculate the OHI scores for the `!x` regions within !study_area. You will modify the files within this folder to complete your assessment.

## 3.1 Data

All data used to calculate OHI scores are contained within the `layers` folder, and the `layers.csv` file acts as a registry for managing all data. In most cases, data provided within the `layers` folder are only place-holders and should be replaced by local, finer-resolution data for your assessment.

Values within the *.csv* files are based on values for all of !study_area and if data are available locally by region, this will allow for a much more precise assessment of ocean health in !study_area. Any data layers that did not have available values for !study_area have place-holder values based on the global mean and are listed in `layers-empty_swapping-global-mean.csv`.

### 3.1.1 *layers.csv*

`layers.csv` is the registry that manages all data to be used in your assessment.

Each row in this file represents a specific data layer that has been prepared and formatted properly for the Toolbox. The first columns contain information inputted by the user; other columns are generated later by the Toolbox App as it confirms data formatting and content. The first columns have the following information:

- ***targets*** indicates how the data layer related goals or dimensions. Goals are indicated with two-letter codes and sub-goals are indicated with three-letter codes, with pressures, resilience, and spatial layers indicated separately.

| goal | name |
|---|---|
| FP | Food Provision |
| FIS | Fisheries |
| MAR | Mariculture |
| AO | Artisanal Fishing Opportunity |
| NP | Natural Products |
| CS | Carbon Storage |
| CP | Coastal Protection |
| TR | Tourism & Recreation |
| LE | Coastal Livelihoods & Economies |
| LIV | Livelihoods |
| ECO | Economies |
| SP | Sense of Place |
| ICO | Iconic Species |
| LSP | Lasting Special Places |
| CW | Clean Waters |
| BD | Biodiversity |
| HAB | Habitats |
| SPP | Species |

- ***layer*** is the identifying name of the data layer, which will be used in R scripts like `functions.R` and *.csv* files like `pressures_matrix.csv` and `resilience_matrix.csv`. This is also displayed on the Toolbox App under the drop-down menu when the variable type is 'input layer'.
- ***name*** is a longer title of the data layer; this is displayed on the Toolbox App under the drop-down menu when the variable type is 'input layer'.
- ***description*** is further description of the data layer; this is also displayed on the Toolbox App under the drop-down menu when the variable type is 'input layer'.
- ***fld_value*** indicates the units along with the units column.
- ***units*** some clarification about the unit of measure in which the data are reported
- ***filename*** is the *.csv* filename that holds the data layer information, and is located in the folder 'layers'.

### 3.1.2 *layers* folder

The `layers` folder contains every data layer as an individual comma separated value (*.csv*) file. The names of the *.csv* files within the layers folder correspond to those listed in the *filename* column of the `layers.csv` file described above. *.csv* files can be opened with text editor software, or will open by default by Microsoft Excel or similar software.

Open any *.csv* file within `layers` and note two important things:

1. There are `!x` numbers represented within the data file: these are unique region identifiers *rgn_id*s for each coastal region in !study_area. The layer called `rgn_layers.csv` identifies which *rgn_id* is associated with which number.
2. There is a specific format that the Toolbox expects and requires that every *.csv* file within the `layers` folder has. Note the unique region identifier (*rgn_id*) with a single associated *score* or *value*, and that the data are presented in 'long format' with minimal columns. See the *Formatting Data for the Toolbox* section of the OHI-Manual for more details.

### 3.1.3 notes on georegionally-gapfilled data

If information for !study_area was not available for a certain data layer, it was often gapfilled using an average of other values within its georegion. Georegions are defined by the United Nations (unstats.un.org/unsd/methods/m49/m49regin.htm). The other countries within the georegion with !study_area are: Ecuador, Peru, Venezuela, Guyana, Suriname, French Guiana, Brazil, Argentina, Uruguay, Chile.

***layers-empty_swapping-global-mean.csv*** is a list of data layers that did not have values for !study_area and were not able to be gapfilled georegionally. This is because calculated scores of other nearby countries were used to gapfill values for !study_area instead of gapfilling occurring at the regional level. These layers especially should be substituted with local data.

## 3.2 *conf* folder

The `conf` folder includes includes R functions (*config.R* and *functions.R*) and *.csv* files containing information that will be accessed by the R functions (*goals.csv*, *pressures_matrix.R*, *resilience_matrix.csv*, and *resilience_weights.csv*).

### 3.2.1 *config.r*

`config.r` is written in the software program R and has important information for the proper calculation and display of goal scores. Briefly, it sets several constants, sets the proper map orientation, and provides labeling for dimension descriptions for the Toolbox Application.

`config.r` also identifies how to aggregate resilience and pressures matrices for goals with categories (natural products, livelihoods & economies, and habitat-based goals). This portion of `config.r` may need to be modifed if entire goals with categories are removed from the assessment. Otherwise, you will likely not modify this file at all.

### 3.2.2 *goals.csv*

`goals.csv` provides information for each goal and sub-goal. Much of the information in this file is for displaying the goals in the Toolbox Application and will not be modified. However, a few columns are important for proper calculation and display in the Toolbox App:

- *description* describes the goal and is the text that will be displayed in the Toolbox App. If you modify what the goal captures, this description will need to be updated.
- *weight* indicates how the goals are weighted when they are combined into an Index score. Currently, all 10 goals are weighted equally (1), and subgoals are weighted as half of a full goal (0.5). If you choose to weight the goals unequally, you will need to modify this column. Weights are set proportionally to each other; goals do not need to add to 10. Goal weights were set equally in the global assessment because no information was available at the global scale to indicate otherwise; in your region there may be local priorities that could inform how to change the weights.
- *preindex_function* indicates how models are called in `functions.r` for all sub-goals and goals that do not have sub-goals. Each goal is calculated with a model identified with the goal's 2- or 3-letter code. The model's inputs are identified in this column, so as you modify goal models or inputs it is important to check this column and modify accordingly. This is called 'preindex' because it refers to .
- *postindex_function* indicates how models are called in `functions.r` for all goals that have sub-goals.

#### 3.2.3 *functions.r*

`functions.r` is written in the software program R and contains models for each goal and sub-goal model, which are called using the inputs in `goals.csv`. Each model calculate the status and trend for the goal or sub-goal using data layers identified in `layers.csv`.

Models will be modified individually, by goal or sub-goal, if you incorporate new data layers or interactions. To do this will require proficiency in the software program R, and ease with learning new user-created packages. The primary R package used is called `dplyr` by Hadley Wickham. The `dplyr` package allows for 'chaining' between functions, which is represented with a `%>%`. See github.com/hadley/dplyr#dplyr for documentation.

#### 3.2.4 Pressures and resilience

There are three files that the Toolbox uses to calculate calculate pressures and resilience components for goal scores. The pressures and resilience matrices are probably more complicated for the global assessment than for any regional assessments, so it is likely that you will simplify these matrices as you complete your study.

Pressures and resilience are calculated for goals without subgoals and for subgoals, but NOT calculated for goals with subgoals. Notice that there are not pressures or resilience scores for LE, SP, FP or BD, but there are scores for LIV, ECO, LSP, ICO, FIS, MAR, HAB, and SPP. This is because combining the pressures scores from subgoals to 'supra' goals don't really make sense conceptually (although obviously very easy technically).

**Based on the local characteristics and priorities of your study area, you must consider several things, in this order:**

1. whether all local pressures are represented in `pressures_matrix.csv`, and if not, add them
2. which pressure ranks/weights are appropriate (i.e. which pressures matter and how much)
3. whether all necessary resilience measures have been included: you will need a resilience measure for each of the pressures ranked as 2 or 3 in the`pressures_matrix.csv`
4. find data to measure additional pressures and resilience; substitute local data for existing data layers (e.g. a data source for chemical pollution, how to assess fishing regulation, etc.)
5. convert all the values to a scale from 0 to 1. To do this, you need to set a reference point (e.g. for chemical pollution, when there is no pollution, the pressure is 0, and you might decide that the highest pollution, that gets a 1, is the highest value measured across all regions in your study area).

#### 3.2.4.1 *pressures_matrix.csv*

`pressures_matrix.csv` describes the layers (from `layers.csv`) needed to calculate pressure categories. The matrix has ranks (weights) assigned that were determined by Halpern *et al.* 2012 (*Nature*) based on scientific literature and expert opinion.

Pressures (columns in `pressures_matrix.csv`), are matched with different goals and subgoals (rows) to indicate which pressures will be included when goal scores are calculated. In some cases the goals are further divided into components (e.g. habitats are divided by habitat type, natural products by product type).

Ranks (weights) are assigned on a scale from 1-3, based on expert judgment of how important the pressure in that column is for the delivery of the goal, sub-goal, or component, in that row. For example, coastal nutrient pollution (`po_nutrients_3nm`) affects both Mariculture (MAR) and Artisanal Fishing Opportunity (AO) But it affects MAR more: MAR has a 3 and AO just a 1. Another example is intertidal habitat destruction (`hd_intertidal`), which has a 3 for *seagrasses* in the Habitats sub-goal because such a pressure has a strong effect on the health of seagrasses.

It is important to note that the matrix identifies the pressures relevant to each goal, and which weight it will get. But each pressure is a data layer, located in the `subcountry2014/layers` folder. This means that pressures layers need information for each region in the study area, and some layers will need to be updated with local data.

Here is some background on the reasoning behind nutrient and chemical pollution in the global `pressures_matrix.csv`. Nutrient and chemical pollution were calculated from the global cumulative impact maps (spatial data). These data were clipped to each global region's EEZ: 200 km from the coast.

- For some goals, these data clipped to the EEZ would affect goals that far from shore, so `po_chemicals` apply to goals relevant offshore: FIS, MAR, ECO, and SPP.
- However, some goals are really only nearshore, and so we clipped the spatial data to 3nm from shore and used this as a separate input. So `po_chemicals_3nm` apply to goals nearshore: AO, CS, CP, TR, ICO, LSP, HAB.

These distinctions don't always apply for smaller-scale assessments. For example, in the US West Coast study (Halpern et al. 2014), only a single `po_chemicals` layer was used: we did not distinguish between offshore and 3nm.

#### 3.2.4.2 *resilience_matrix.csv*

`resilience_matrix.csv` describes the layers (from `layers.csv`) needed to calculate resilience categories.

`resilience_matrix.csv`, in similar way to `pressures_matrix.csv`, shows which kind of regulations and social measures apply to each of the goals and goal components. There must be a resilience measure that could directly affect any pressure with a rank/weight above 1.

#### 3.2.4.3 *resilience_weights.csv*

`resilience_weights.csv` describes the weight of various resilience layers, were determined by Halpern et al. 2012 (Nature) based on scientific literature and expert opinion.

### 3.3 spatial folder

The spatial folder contains two files: *regions_gcs.js* and *regions_gcs.geojson*. These are spatial files, one in JSON format and one in GeoJSON format. Both are maps displaying the appropriate study area and regions for the assessment. It is possible to view the regions on GitHub, which is able to render the GeoJSON format. Go to this url: github.com/OHI-Science/!repo/blob/master/subcountry2014/spatial/regions_gcs.geojson.

### 3.4 *launch_app_code.R*

Running `launch_app_code.r` from within the `subcountry2014` folder will launch the Toolbox Application, which will be displayed in a web browser. The Toolbox App will display all input data from the `layers` folder as well as the calculated scores from `scores.csv`.

### 3.5 *calculate_scores.r*

You will run `calculate_scores.r` when you are ready to calculate scores with your local data and goal models. `calculate_scores.r` calls other functions within the Toolbox to calculate goal scores using the *.csv* files in the *layers* folder that are registered in *layers.csv* and the configurations identified in *config.r*. Scores will be saved in *scores.csv*.

### 3.6 scores.csv

`scores.csv` reports the calculated scores for the assessment. Scores are reported for each dimension (future, pressures, resilience, score, status, trend) for each reporting region. The scores here were calculated using the data in the repository (extracted from the global assessment) simply to demonstrate the Toolbox has functioned properly. Note that in `scores.csv` there are `!x + 1` region identifiers (*rgn_id*). The combined score for the !study_area is reported in *rgn_id* 0.