

# The Ocean Health Index Toolbox Manual

## Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
1.1	The Ocean Health Index . . . . .	4
1.2	The Ocean Health Index Toolbox . . . . .	4
<b>2</b>	<b>Overview of the OHI Toolbox App</b>	<b>5</b>
2.1	Data page . . . . .	6
2.1.1	Overview of display options . . . . .	6
2.1.2	Overview of variable options . . . . .	7
2.2	Calculate page . . . . .	8
2.3	Report page . . . . .	9
<b>3</b>	<b>Explore Global Results with the Toolbox App</b>	<b>9</b>
3.1	Map . . . . .	9
3.2	Histogram . . . . .	11
3.3	Table . . . . .	11
<b>4</b>	<b>Introduction to OHI Regional Assessments</b>	<b>13</b>
4.1	Before getting started . . . . .	14
<b>5</b>	<b>Conducting a Regional Assessment</b>	<b>14</b>
5.1	What to expect when conducting a regional assessment . . . . .	14
5.1.1	Timeline . . . . .	14
5.2	What is provided . . . . .	15
5.3	Points to remember . . . . .	15
5.4	Checklist: To Do Before Using the Toolbox . . . . .	16
5.5	Discovering and Gathering Appropriate Data and Indicators . . . . .	17
5.5.1	Data sources . . . . .	17
5.5.2	Data gathering responsibilities . . . . .	17
5.5.3	The process of data discovery . . . . .	18
5.5.4	Requirements for data layers . . . . .	18
5.5.4.1	Relevance to ocean health . . . . .	19
5.5.4.2	Reference point . . . . .	19
5.5.4.3	Appropriate spatial scale . . . . .	19
5.5.4.4	Appropriate temporal scale . . . . .	19

5.5.5	Notes about data and regions . . . . .	19
5.5.6	Example: US West Coast data discovery . . . . .	19
5.5.6.1	Reasons data were excluded . . . . .	19
5.5.6.2	Creative approaches to using data . . . . .	21
5.6	Formatting Data for the Toolbox . . . . .	21
5.6.1	Introduction . . . . .	21
5.6.2	Gapfilling . . . . .	22
5.6.2.1	Temporal gapfilling . . . . .	22
5.6.2.2	Spatial gapfilling . . . . .	25
5.6.3	Long formatting . . . . .	27
<b>6</b>	<b>Installing the Toolbox for a Regional Assessment</b>	<b>28</b>
6.1	Overview . . . . .	28
6.2	Getting started . . . . .	30
6.3	Accessing repositories with GitHub . . . . .	30
6.3.1	Overview . . . . .	30
6.3.2	Install R and RStudio . . . . .	31
6.3.3	Create a GitHub account . . . . .	31
6.3.4	Install <i>git</i> software . . . . .	31
6.3.5	Set up your Git Identity . . . . .	31
6.3.6	Clone your repo to your computer . . . . .	32
6.3.7	Working locally . . . . .	35
6.3.8	Committing and pushing changes locally . . . . .	35
6.3.9	Keeping synchronized . . . . .	39
6.3.10	More Information . . . . .	39
6.4	Accessing repositories without GitHub . . . . .	40
6.4.1	OHI regional assessments without GitHub . . . . .	40
6.4.2	Install the latest version of R . . . . .	40
6.4.3	Download the repo to your computer . . . . .	40
6.4.4	Working locally . . . . .	40
6.5	OHI Toolbox File System . . . . .	41
6.5.1	Assessments and scenarios . . . . .	41
6.5.2	<i>layers.csv</i> . . . . .	41
6.5.3	<i>layers</i> folder . . . . .	43
6.5.4	<i>conf</i> folder . . . . .	43
6.5.4.1	<i>config.r</i> . . . . .	44
6.5.4.2	<i>functions.r</i> . . . . .	44

6.5.4.3	<i>goals.csv</i>	.....	44
6.5.4.4	<i>pressures_matrix.csv</i>	.....	44
6.5.4.5	<i>resilience_matrix.csv</i>	.....	44
6.5.4.6	<i>resilience_weights.csv</i>	.....	44
6.5.5	spatial folder	.....	44
6.5.6	<i>calculate_scores.r</i>	.....	45
6.5.7	<i>scores.csv</i>	.....	45
6.5.8	Relaunching the Toolbox	.....	45
<b>7</b>	<b>Using the Toolbox for a Regional Assessment</b>		<b>45</b>
7.1	Modifying pressures and resilience matrices	.....	45
7.2	Modifying and creating data layers	.....	45
7.2.1	Template data layers	.....	46
7.2.2	Create data layers with proper formatting	.....	46
7.2.3	Save data layers in the <i>layers</i> folder	.....	46
7.2.4	Register data layers in <i>layers.csv</i>	.....	46
7.2.5	Check pressures and resilience matrices	.....	48
<b>8</b>	<b>Update pressures_matrix.csv</b>		<b>48</b>
<b>9</b>	<b>Update resilience_matrix.csv</b>		<b>48</b>
9.1	Modifying goal models	.....	49
9.1.1	Update <i>functions.r</i>	.....	49
9.1.2	Check and possibly update <i>goals.csv</i>	.....	49
9.2	Removing goals	.....	50
9.3	Example modifications	.....	50
9.3.1	Adding a new layer to a goal model	.....	50
9.3.2	Adding a new pressure layer	.....	52
9.3.2.1	Register the new layers in <i>pressure_matrix.csv</i>	.....	52
9.3.2.2	Create the new layers	.....	52
9.3.2.3	Register the new layers in <i>layers.csv</i>	.....	53
9.3.3	Updating resilience matrix with local habitat information	.....	53
9.3.4	Removing a goal	.....	56
9.4	Notes about R	.....	57

<b>10 Frequently Asked Questions (FAQs)</b>	<b>57</b>
10.1 Overall . . . . .	57
10.2 Conceptual . . . . .	57
10.3 Timing and Resources . . . . .	58
10.4 Structure . . . . .	59
10.5 Reference points . . . . .	59
10.6 Appropriate data layers . . . . .	59
10.7 Food Provision . . . . .	60
10.8 Livelihoods & Economies . . . . .	61
10.9 Tourism & Recreation . . . . .	61
10.10 Natural Products . . . . .	61
10.11 Sense of Place . . . . .	62
10.12 Pressures . . . . .	62
<b>11 Toolbox Troubleshooting</b>	<b>62</b>
11.1 Loading RWorkspace on Restart . . . . .	62
11.2 Calculating Pressures. . . . .	64
11.2.1 ‘The following components for [goal] are not in the aggregation layer [layer]...’ . . . . .	64
11.2.2 ‘Error in matrix...’ . . . . .	66
11.3 Calculating Resilience . . . . .	66
11.3.1 ‘Error in match(x, table, nomatch = OL) : object id_num not found’ . . . . .	66

# 1 Introduction

## 1.1 The Ocean Health Index

The **Ocean Health Index (OHI)** project was created with a strong foundation in communication and collaboration. Following the initial publication by Halpern et al. 2012 in *Nature* (available at [ohi-science.org/](http://ohi-science.org/)), we have conducted annual global assessments to track changes in global ocean health. At the same time, we have been developing the **Ocean Health Index Toolbox** software that will enable others to conduct assessments on their own using the OHI framework. Here, the **Ocean Health Index Manual** describes how to use the Toolbox software.

OHI scores by goal are calculated at the scale of the reporting unit, which is called a **region** and then combined using a weighted average to produce the score for the overall area assessed, called a **study area**.

## 1.2 The Ocean Health Index Toolbox

The **Ocean Helath Index Toolbox Application (App)** is open-source software developed to facilitate Index calculations at any scale. The Toolbox App runs using the statistical programming language R and displays input data and calculated scores graphically with maps, histograms, and tables. Goal scores are calculated for each region separately and then combined using a weighted average to produce the score for

the overall study area assessed. Modifications can be made to the global studies, allowing calculation of scores with updated data that can then be visualized with the Toolbox App.

The Toolbox can also be used to calculate scores in smaller-scale **regional assessments** after finer-scale data has been gathered and decisions have been made regarding goal models, pressures and resilience, and reference points. The Toolbox App therefore enables the Ocean Health Index framework to be customized to an area of interest, incorporating the data, indicators, and priorities regarding ocean-derived benefits that are relevant to the chosen area. The App can additionally be used to compare how different management scenarios could affect overall ocean health, which can inform effective strategies for ocean resource management at a local scale.

We recommend first becoming familiar with the Toolbox App and exploring the results from the global assessment before beginning a regional assessment.

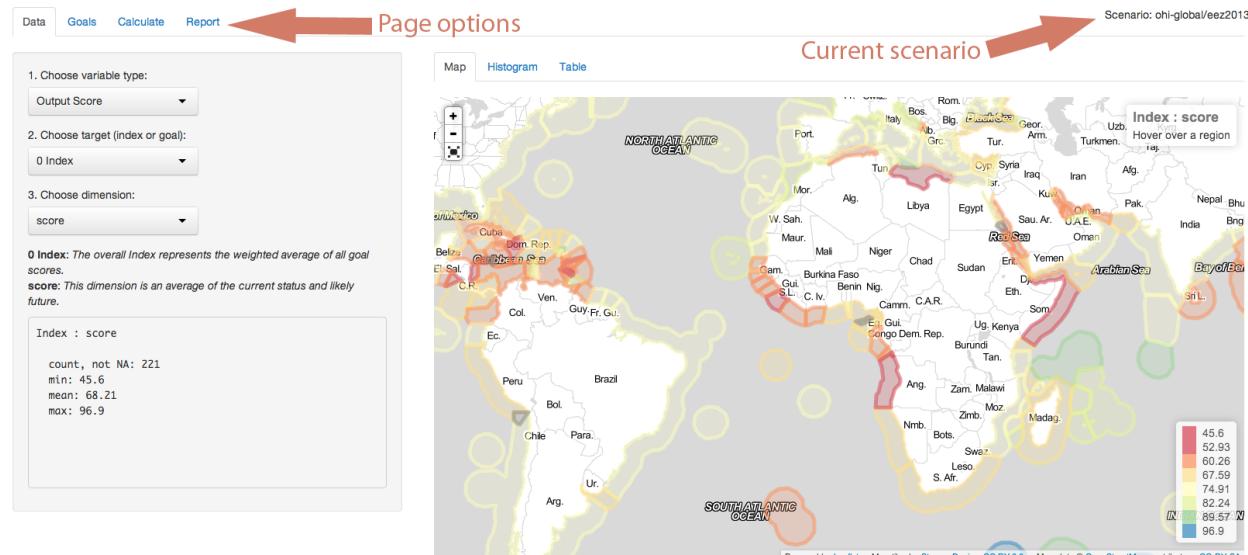
**Requirements:** an Internet connection for the initial Toolbox App installation. The Toolbox runs on both a Mac and PC.

1. Install the latest version of R (<http://cran.r-project.org/>).
2. Install and launch the Toolbox Application (<http://ohi-science.org/pages/install.html>).

## 2 Overview of the OHI Toolbox App

The Toolbox Application (App) runs through a web browser and has three **Page options**: *Data*, *Calculate*, and *Report*. The *Data* page provides several ways of viewing the data (*Map*, *Histogram*, *Table*), and is the default home-screen when the Toolbox is opened. The *Calculate* page is used to calculate goal scores based on the information in the file system (data can be changed and tested). The *Report* page provides different methods of reporting the data.

The **Current scenario** is displayed in the upper-right corner of the Data page. This identifies the scale (example: global) and year (example: 2013) of the information displayed.

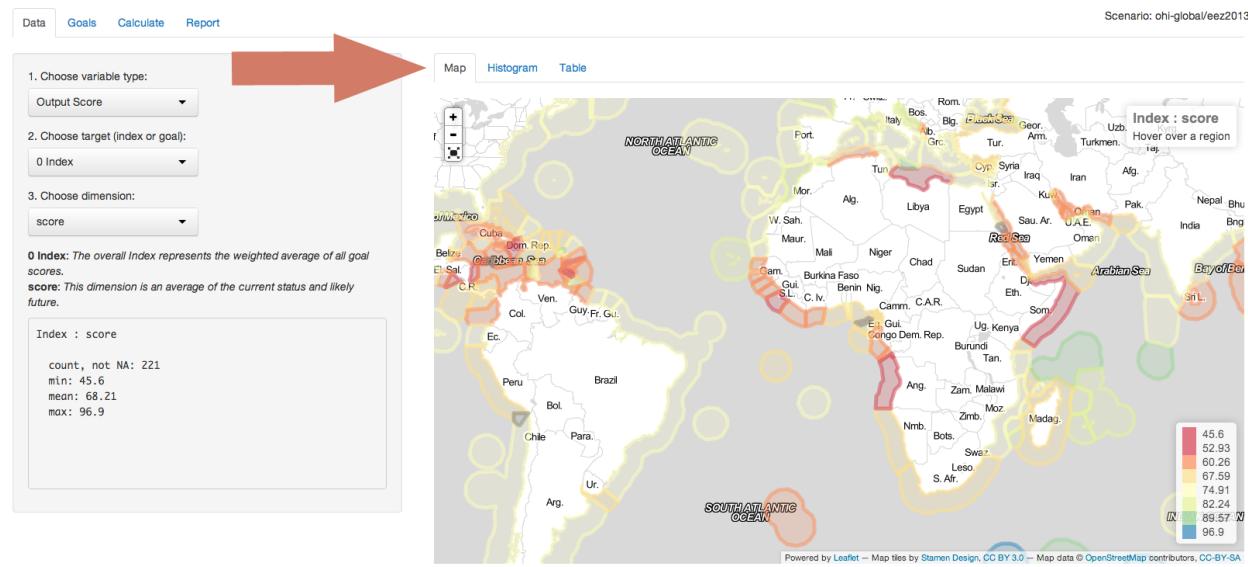


## 2.1 Data page

### 2.1.1 Overview of display options

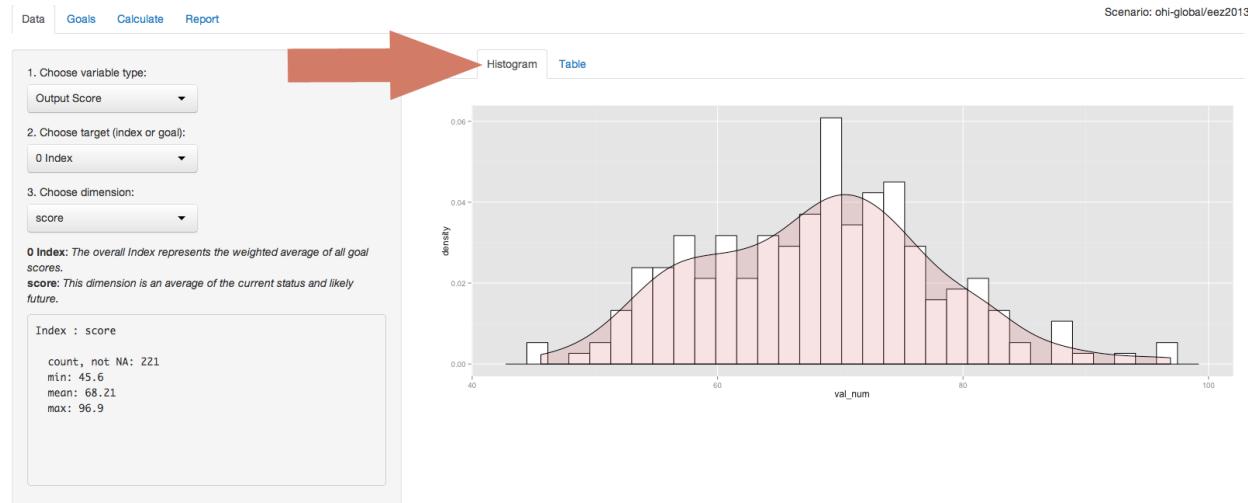
The Data page displays input data or calculated scores for each goal parameter, and presents the information as a Map, Histogram, or Table. These options are presented as tabs located above the map (Map view is the default display option for the Data page).

#### Data displayed in Map form:



The map displays data for every region as reported in the scenario. A color legend is displayed in the lower right corner of the map that provides a linear colormap of the data. The range of values will change as different variables are selected.

#### Data displayed in Histogram form:



The histogram shows the distribution of the selected variable as the number of observations per value bin (white bars) and a smoothed density function (pink shading).

#### Data displayed in Table form:

Data Goals Calculate Report Scenario: chi-global/eez2013

1. Choose variable type: Output Score

2. Choose target (index or goal): 0 Index

3. Choose dimension: score

**0 Index:** The overall Index represents the weighted average of all goal scores.  
**score:** This dimension is an average of the current status and likely future.

Index : score

```
count, not NA: 221
min: 45.6
mean: 68.21
max: 96.9
```

Table

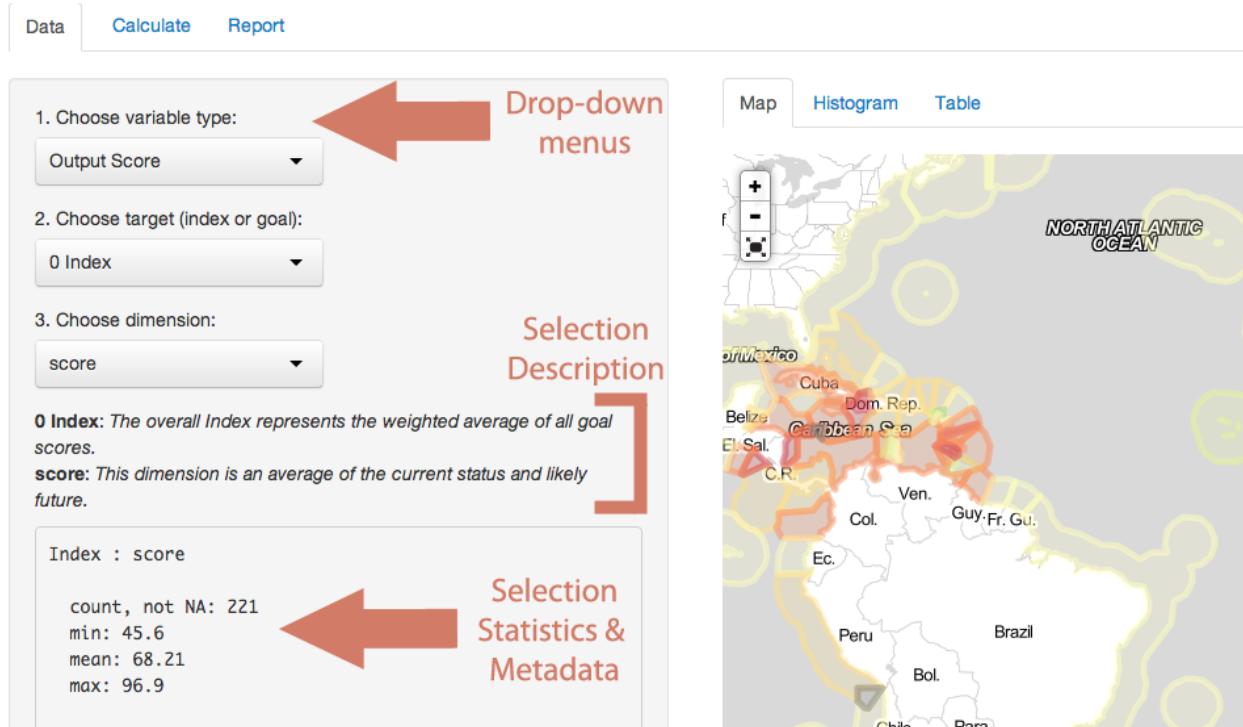
25 records per page Search:

rgn_id	rgn_name	value
0	GLOBAL	70.06
1	Cocos Islands	70.31
2	Christmas Island	69.05
3	Norfolk Island	77.91
4	Macquarie Island	87.40
5	New Caledonia	70.59
6	Vanuatu	71.51
7	Solomon Islands	66.34
8	Palau	65.11
9	Micronesia	68.85
10	Nauru	78.57
11	Marshall Islands	77.43
12	Wake Island	70.27

The table displays the target value for each region and the overall study area. It provides an identifying code (rgn\_id), name (rgn\_name), and value (value) for each.

### 2.1.2 Overview of variable options

The Data page has drop-down menus from which the user chooses the data to be displayed by the Toolbox. The data selected from the pull-down menus can be viewed in Map, Histogram, or Table form as described in the ‘Overview of display options’ section above. Descriptions, statistics and metadata for the chosen fields are also displayed below the drop-down menus on the left side of the page.



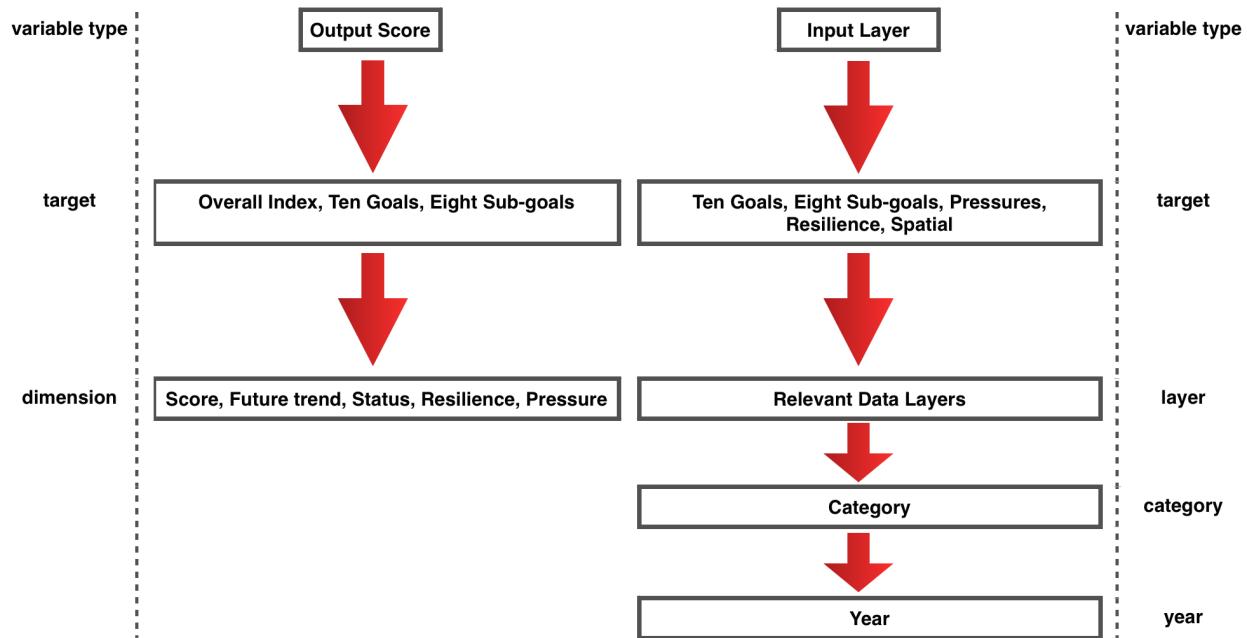
The first selection to be made from the drop-down menus is variable type, in which the user can choose the **Output Score** that will show a calculated score (for a particular target chosen subsequently), or **Input**

**Layer** that will show the data layer used to calculate the score of a particular target.

In short: **Output Score = Calculated Score** and **Input Layer = Data**

For example, if the user selects ‘Output Score’ as the variable type (which is the default), they will then be able to choose a goal/sub-goal (target), and the goal/sub-goal score calculation (dimension) to be reported.

On the other hand, if the user selects “Input Layer” as the variable type, they will be able to choose a target and a specific data layer that is associated with that target. If that layer has multiple categories or years available, the user will be able to select a preference, with a default being the first category alphabetically and the most recent year.



To further familiarize yourself with the display and variable options of the Toolbox App, please see the ‘Explore Global Results with the Toolbox App’ section below.

## 2.2 Calculate page

The Calculate page is used to calculate the Index. The Calculate page displays the “Scenario path” in the user’s home directory where the folders are located.

<a href="#">Data</a>	<a href="#">Goals</a>	<a href="#">Calculate</a>	<a href="#">Report</a>
Scenario path exists:			
TRUE			
Scenario path			
/Users/julialowndes/ohi-global/eez2013			
<a href="#">Calculate</a>			

The screenshot shows a user interface for generating a report. At the top, there are tabs: Data, Goals, Calculate, and Report. The Report tab is selected. Below the tabs, there are several input fields and checkboxes:

- Reports directory:** /Users/julialowndes/ohi-global/eez2013/reports
- Report filename to output:** report.html
- Include:**  Flowers  Tables
- Options:**
  - Open in new window
  - Global only (vs all regions which takes time)
  - Overwrite existing figures
- Generate Report** button

## 2.3 Report page

The Report page allows the user to create a report of results as .html and .pdfs files.

It is currently possible to create flower plots and tables with several options, including saving global figures or figures per country. In future versions of the Toolbox App, more options for visualizations, including maps, histograms, and equations, will be possible. It will also be possible to compare different scenarios.

## 3 Explore Global Results with the Toolbox App

The following section will be devoted to exploring the Ocean Health Index Global Results using the Toolbox App. The section will further familiarize the user with the organizational structure and functionality of the Toolbox, and will also help the user understand the results that were obtained for the global assessment, including the types of data that were used to calculate the Ocean Health Index.

To begin exploring results, the user must first select the data he wishes to display with the Toolbox (i.e. variable type > target > dimension/layer > category > year).

**For example:**

**Output Score > Fisheries (FIS) > score** shows the final calculated score for the fisheries sub-goal in the current scenario (2013).

**Input Layer > Mariculture > Mariculture species harvested** shows the different harvested species that were used to calculate the mariculture sub-goal in the current scenario (2013).

As mentioned previously, the selected data can then be viewed as a Map, as a distribution on the Histogram tab or numerically in Table form.

### 3.1 Map

Suppose for instance that **Output Score > Index > score** is selected and viewed as a map (default selection).

The target (Index) and dimension (Score) chosen from the drop-down menu is displayed as text on the top right corner of the map. Below this information, the region name, (region number), and value appear if the

**Screenshot 1:**

- 1. Choose variable type: Output Score
- 2. Choose dimension: score
- 3. Index : score
  - count, not NA: 221
  - min: 45.6
  - mean: 68.21
  - max: 96.9

**Screenshot 2:**

- 1. Choose variable type: Output Score
- 2. Choose target (index or goal): 1.1 Fisheries (FIS)
- 3. FIS : score
  - count, not NA: 205
  - min: 1.01
  - mean: 58.88
  - max: 99.5

**Screenshot 3:**

- 1. Choose variable type: Output Score
- 2. Choose target (index or goal): 1.1 Fisheries (FIS)
- 3. Choose dimension: score
  - score
  - pressures
  - resilience
  - status
  - trend
  - future

**Screenshot 1:**

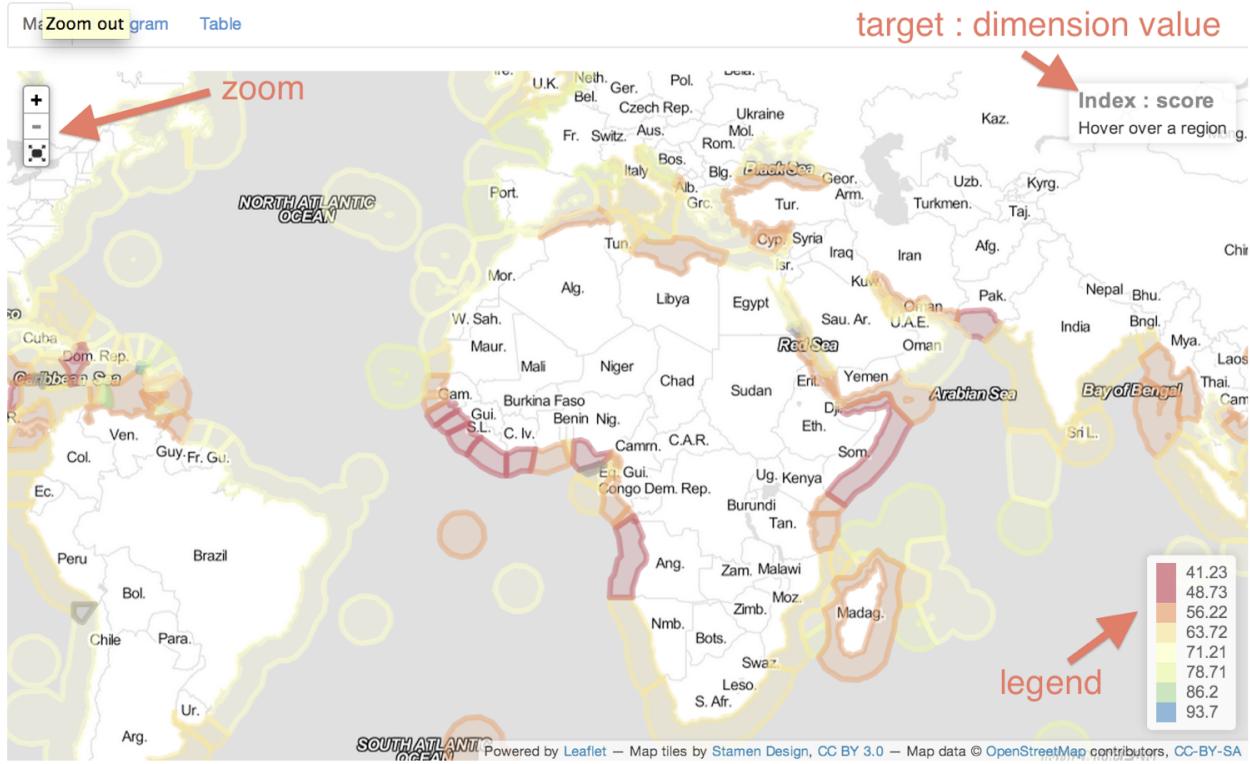
- 1. Choose variable type: Input Layer
- 2. Choose layer: Fisheries weighting factor (fp\_wildcaught\_weight)
- 3. Choose year:
  - Fisheries weighting factor (fp\_wildcaught\_weight): Proportional yield of wild caught fisheries
  - Fisheries weighting factor (fp\_wildcaught\_weight)
    - count, not NA: 206
    - min: 0.1181
    - mean: 0.9148
    - max: 1
  - targets: FP
  - layer: fp\_wildcaught\_weight

**Screenshot 2:**

- 1. Choose variable type: Input Layer
- 2. Choose target (goal, pressures, resilience or spatial): 1.2 Mariculture
- 3. Coastal population inland 25 miles
  - Coastal population inland 25 miles (mar\_coastalpopn\_inland25mi): Coastal population inland 25 miles by year (2005-2015)
  - Coastal population inland 25 miles (mar\_coastalpopn\_inland25mi) : 2015
    - count, not NA: 202
    - min: 0
    - mean: 8.85e+06
    - max: 2.221e+08

**Screenshot 3:**

- 1. Choose variable type: Input Layer
- 2. Choose target (goal, pressures, resilience or spatial): 1.2 Mariculture
- 3. Choose layer: Mariculture species harvested (mar\_harvest\_species)
  - Mariculture species harvested (mar\_harvest\_species)
    - Mariculture sustainability score (mar\_sustainability\_score)
      - Mariculture sustainability score (mar\_sustainability\_score)
        - count, not NA: 202
        - min: 0
        - mean: 8.85e+06
        - max: 2.221e+08



user's mouse hovers over a specific region. When the value of that region is displayed, the region will darken on the map. If the mouse is not indicating a specific region, this area will instruct to 'hover over a region'. Zooming in and out of the map is possible with the plus/minus in the upper left of the map. A color legend is also displayed in the lower right corner of the map that provides a linear colormap of the data. The range of values will change as different variables are selected.

### 3.2 Histogram

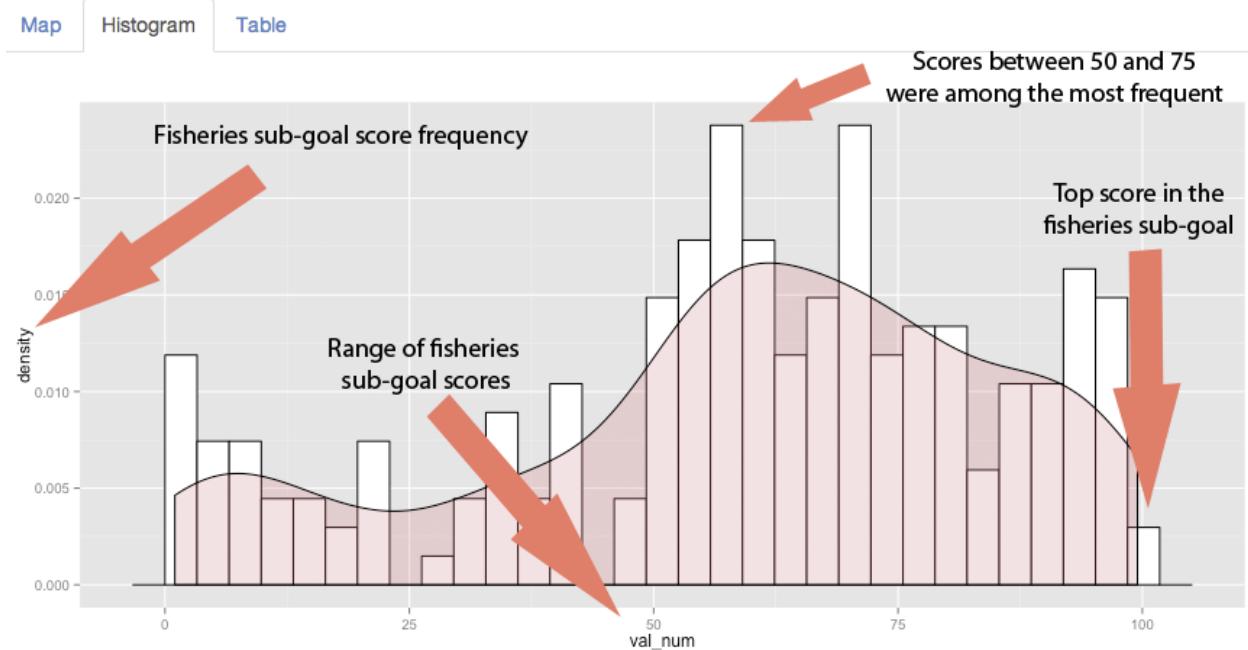
For example, if the **Output Score > Fisheries > score** sub-goal is selected, the histogram shows the frequency of each fisheries score:

As noted in the figure above, the Histogram view allows the user to see the range of scores (lowest and highest) as well as the frequency of each (most vs. least common). The latter may be helpful for instance if the user wishes to compare to score of a particular country to all the other scores.

### 3.3 Table

The table displays the target value for each region and the overall study area, with an identifying code (rgn\_id), name (rgn\_name), and value (value) for each.

It is possible to search the data using the search bar at the upper right of the table (caps-sensitive). It is also possible to search within individual columns by entering the search term at the bottom of each column; this option is more easily displayed by changing the number of records per page to 10 using the drop-down menu. Data may also be sorted the arrows at the left of each column name. Sorting by rgn\_id is the default.



Map   Histogram   Table

10 records per page

sort alphabetically

Search:

rgn_id	rgn_name	value
82	Albania	60.01
84	Algeria	52.51
151	American Samoa	61.45
92	Amsterdam Island and Saint Paul Island	73.59
26	Andaman and Nicobar	55.98
200	Angola	44.93
118	Anguilla	66.33
213	Antarctica	
120	Antigua and Barbuda	73.33
172	Argentina	61.82

rgn\_id rgn\_name value

Showing 1 to 10 of 222 entries

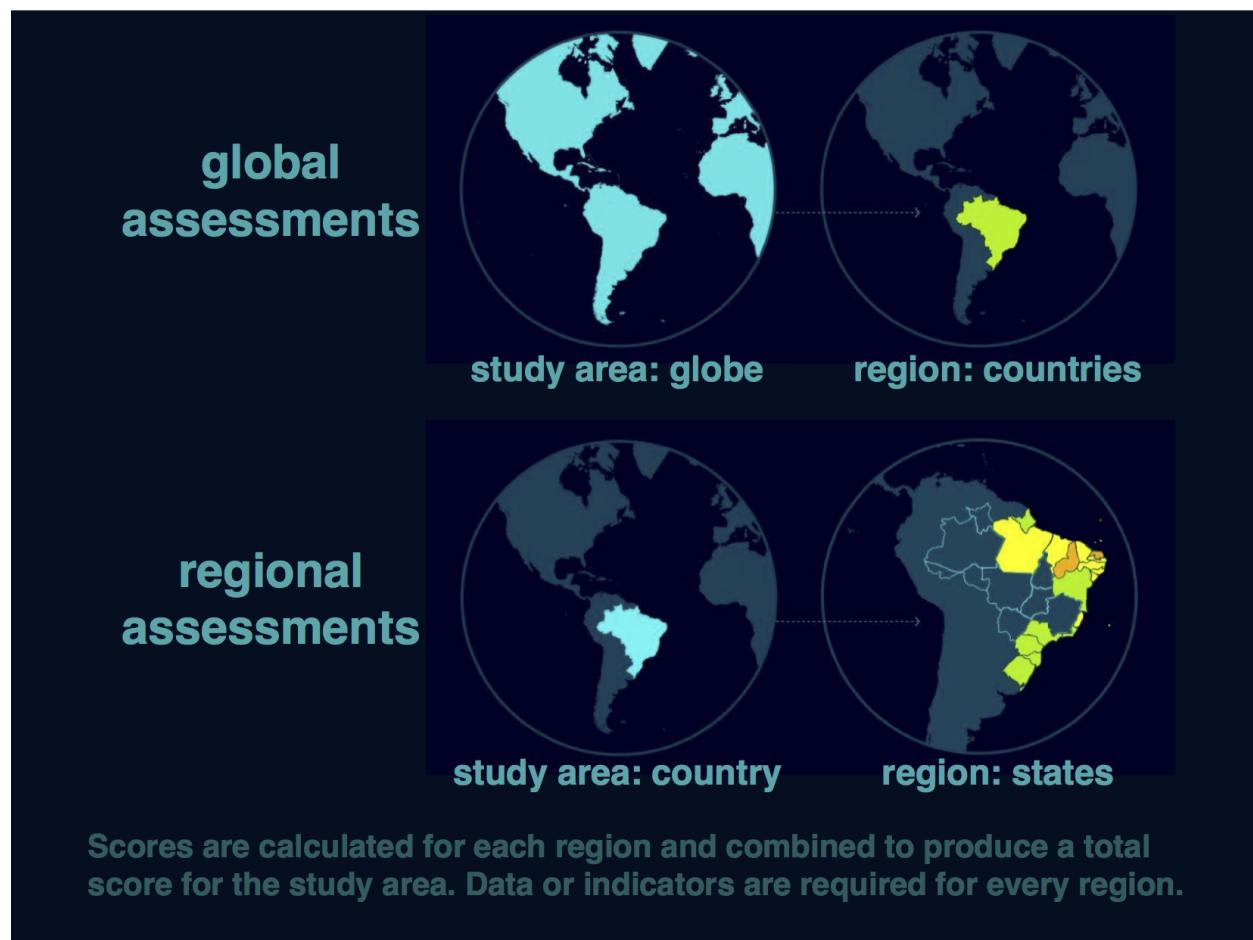
search specific columns

← Previous 1 2 3 4 5 Next →

## 4 Introduction to OHI Regional Assessments

Regional assessments use the Ocean Health Index framework to study smaller spatial scales, often where policy and management decisions are made. Using ten criteria (goals), the Index scores how well coastal regions optimize their potential ocean benefits and services in a sustainable way relative to self-established reference points (targets), on a scale of 0 to 100. Regional assessments incorporate local priorities, higher-resolution data and indicators, and use tailored goal models and reference points, which produce scores better reflecting local realities. The OHI is designed to combine existing work into a comprehensive assessment of ocean health. And because the assessments are developed at the scale of decision-making using local data and parameters, the findings can help inform decision-making and management actions aimed at maximizing sustainable productivity while preserving vital natural capital.

Scores by goal are calculated weighted average to produce the score for the overall area assessed, called a **study area**. In regional assessments, coastal states or provinces are often the **regions** within a country, which is the **study area**.



The Ocean Health Index framework is flexible to accommodate regional priorities: goal models should be modified to incorporate the best available local data and indicators, or excluded entirely to best represent regional attributes. Similarly, pressures and resilience measures can be refined using local data and indicators. Calculated Index scores are only as ‘good’ as the data on which they are based, so finding the best data or indicators available is fundamental.

The process for developing a regional assessment can be as valuable as the final calculated scores, since it creates an ocean alliance that combines knowledge and cultural values from

**many different perspectives and disciplines.** Calculating Ocean Health Index scores at regional scales requires engagement from as many different groups as possible, including research institutions, government agencies, policy groups, non-governmental organizations, and the civil and private sectors.

Regional assessments have been published in peer-review journals for the U.S. West Coast (Halpern et al. 2014) and for Brazil (Elfes et al. 2014), and more are underway (available at ohi-science.org/).

## 4.1 Before getting started

Before beginning a regional assessment, it is important to have planned several things:

- Develop strategic plan and timeline
  - determine resources available
  - outline timeline with any meetings and workshops necessary
  - engage management early for results to be most useful (optional)
- Assemble a qualified team with diverse skills
  - have a broad scientific understanding and experience with environmental policy
  - are comfortable with handling large data sets, making decisions, and thinking creatively
  - collaborate well in a multidisciplinary team, remotely and in person
  - can work with the software program R and user-created packages (at least one person)
  - can work with ArcGIS or other spatial analysis software (at least one person)
  - are fluent in English (at least one person)
- Funding
  - greatly depends on the local context
  - potentially needed for a management and scientific team, workshops and meetings (including travel), communications, policy engagement, and operating costs
- Policy and management interest (optional)
  - engage policy makers early: informing government policies to improve ocean health is most effective if there is early interest and engagement from government agencies
  - requires ongoing communication during Index development to best inform management actions that could have measurable impacts.
  - repeated assessments as new data become available enable tracking ocean health through time and evaluating management priorities

# 5 Conducting a Regional Assessment

## 5.1 What to expect when conducting a regional assessment

There are key considerations and processes that will be a part of every regional assessment, however, how each regional assessment is conducted will be unique depending on the local context. For example: what data and indicators are available will determine how goal models are able to be tailored to the region, and what skillsets and resources are involved will affect the time it takes to complete the assessment.

### 5.1.1 Timeline

The time required to complete a regional assessment depends on the local context and available resources, but the proportion of time required for different phases of the assessment is more predictable.

**Expect for strategic planning, discovering and gathering data, and developing reference points and models, to comprise > 80% of the time allotted for the regional assessment.**

Tasks	PHASE 1 Create Framework for a Regional Assessment	PHASE 2 Calculate the Index Scores	PHASE 3 Influence Decision-Making
Develop strategic plan & timeline; assemble skilled team			
Data discovery & acquisition	■	■	
Assign reference points (targets) & modify goal models	■	■	
Use Toolbox Application		■	■ ■
Document decisions & results interpretation of scores; publish findings	■	■	■

## 5.2 What is provided

The Ocean Health Index Toolbox can be used to calculate scores for regional assessments. When you begin a regional assessment, you will be provided with a **repository**: a set of Toolbox files and software scripts specific to your study area. The data in the files provided serve as a template: they are the data used in the global assessments, extracted for your study area. These data files will need to be updated with locally available data and indicators so as to give a more fine-scale assessment of ocean health in your study area. However, if better data are not available, it is possible to use the files provided.

Our team of scientists and managers is prepared to provide guidance for regional assessments, from initial meetings to discussions about model development to disseminating results.

## 5.3 Points to remember

We recommend to remember the following as you develop your approach:

- People are part of ocean health
  - economic and social data are included, and certain terrestrial data near the coast
- Document all decisions made
  - document data sources, urls, date of access, reference points set, why models were modified the way they were (and not other potential ways)...
  - important for transparency, describing methods and explaining results in reports and publications, and for reproducibility (for any future comparable assessments in your study area)
- Goal models should be modified to capture local characteristics and priorities of the region
  - how goal models are developed depends on what is important locally and what data/indicators are available.
- Assessments can use a mixture of regional-, country- and global-scale data
- Scores are calculated by region
  - comparing scores between subcountry regions is a main reason of conducting a regional assessment
  - regional scores are combined by weighted average to calculate scores for the study area
  - goal models are the same across all regions; data are specific to the region. For example, the carbon storage goal model is calculated in the same way for each region, but which habitats are present in each region, and the area and condition of those habitats are specific to each region

- All data do not need to be at the same spatial or temporal scale
  - each assessment should represent the best understanding of ocean health at the point in time. For example if fisheries data are available from 1980-2011 and tourism from 2008-2012; these can be used together to calculate Index scores, which will be the representation of current ocean health

## 5.4 Checklist: To Do Before Using the Toolbox

Most of the time spent while conducting a regional assessment occurs before using the Toolbox, although being familiar with the Toolbox's composition and structure can help guide you. The following will not be checked-off in sequence: there is a lot of back and forth as you discover data, develop reference points and models, revisit other data possibilities.

Here is a checklist of decisions and steps that must be done before the Toolbox can calculate regional goal scores:

- Understand the philosophy of the Ocean Health Index
  - understand what goals represent and how they are modeled
  - understand what pressures and resilience are included
  - understand how reference points are set
- Identify local characteristics and priorities
  - should any goals be removed?
  - should any goals be redefined?
  - should other goals be added?
  - should goals be weighted unequally?
  - what are pressures to the local system?
  - what resilience measures (laws, regulations) are in place locally?
  - what local cultural preferences or priorities should be captured in the assessment?
- Decide the spatial scale for regions within the study area
  - we recommend multiple regions
  - regions within the study area should be at the smallest scale the data allow
  - in what spatial scale are data most frequently reported?
  - in what spatial scale are policy decisions made (optional)?
- Be familiar with the global inputs to your Toolbox repository
  - understand the data and models used in global assessments
  - prioritize which data and models should be updated
  - become familiar with the Toolbox framework for your study region
  - example: ohi-science.org/ecu
- Discover and gather appropriate data and indicators
  - what local data and indicators are available for goal models?
  - what local data and indicators are available pressures and resilience?
  - how would reference points be set for local data?
  - do local data and indicators capture the philosophy of the Ocean Health Index?
  - are local data and indicators at the appropriate spatial and temporal scales?
  - process and format data and indicators to create data layers\* for the Toolbox
  - use the Toolbox repository as a registry to organize data layers
- Modify pressures and resilience matrices

- categorize any new pressures layers
- example: pollution, fishing, etc
- set pressure weighting/ranking based on literature, expert opinion
- identify responsive resilience measures
- any pressure layer with a weight of 2 or 3 needs a complementary resilience layer
- categorize any new resilience layers
- example: regulations, ecological, etc
- Modify goal models and set reference points:
  - can goal models be refined using locally available data and indicators?
  - can reference points be refined using locally available data and indicators?

\* Each data component that is included in the OHI is called a **data layer** because it will be combined with others to create the most complete picture of ocean health. Many data layers are rescaled from 0-1 to be combined with others on the same unitless scale.

## 5.5 Discovering and Gathering Appropriate Data and Indicators

The OHI spans disciplines and integrates diverse data and sources to give a comprehensive assessment of ocean health. A hallmark of the OHI is that it uses freely-available data to create models that capture the philosophy of individual goals, and finding appropriate data requires research and creativity. There are many decisions to make when gathering from disparate sources, deciding reference points, and developing goal models.

**Index scores are a reflection of data quality, and thus accessing the best data and indicators available is of highest importance.**

### 5.5.1 Data sources

Existing data and indicators can be gathered from many sources across environmental, social, and economic disciplines, including:

- government reports
- government websites
- academic literature
- masters and PhD theses
- university websites
- non-profit organizations

All data will be rescaled to specific reference points (targets) before being combined; therefore setting these reference points at the appropriate scale is a fundamental component of any OHI assessment. This requires the regional assessment team to interpret the philosophy of each Index goal and sub-goal using the best available data and indicators.

### 5.5.2 Data gathering responsibilities

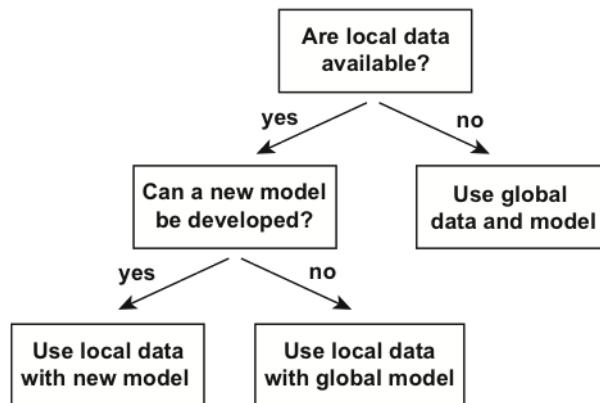
Gathering appropriate data requires searching for and accessing freely-available data. It is important that team members responsible for data discovery make thoughtful decisions about whether data are appropriate for the regional assessment, and that they also get feedback from the full team to discuss the merits of different data sources. Data discovery and acquisition can be quite an iterative process, as there are both practical and philosophical reasons for including or excluding data, in addition to requiring access to the data.

Index scores can be recalculated annually as new data become available. This can establish a baseline of ocean health and serve as a monitoring mechanism to evaluate the effectiveness of actions and policies in improving the status of overall ocean health. This is good to keep in mind while looking for data: will it be available again in the future? It is also important to document the sources of all data so that it is both transparent where it came from and so that you will be able to find it again in the future.

### 5.5.3 The process of data discovery

In terms of philosophical considerations, the most important thing to remember when gathering data is that the data must contribute to measuring ocean health. Many data sources that enhance our knowledge of marine processes may not directly convey information about ocean health and may not be appropriate within the OHI framework. Because of this, compiled indicators can sometimes be more suitable than raw data measuring single marine attributes.

Begin by understanding and comparing the best approaches used in assessments that have been completed, including global assessments, Brazil, and the US West Coast. If finer-resolution local data were available in the study area, they could be used either in a newly developed regional goal model using locally appropriate and informed approaches, or in the existing global goal model. When local data were not available, the same global-scale data were used with the original global goal model, which is least desirable because it does not provide more information than the global study. When looking for data, the following decision tree may be useful. This should be a goal-by-goal process:



Searching for data requires exploring data sources beyond any single discipline, and a good place to start is with an internet search. Internet searches can lead to published data in government and non-governmental organization reports, peer-reviewed articles, and masters and doctoral dissertations. Not everything will be freely available online but it is sometimes possible to request access.

It is good practice to keep detailed notes of attributes of each potential data layer, since there may be different options to work with. Searching for data by goal is a good approach, although some data layers will be used for multiple goals.

### 5.5.4 Requirements for data layers

Four requirements to remember when investigating (or ‘scoping’) potential data layers are:

1. relevance to ocean health
2. how to the reference point
3. spatial scale
4. temporal scale.

**5.5.4.1 Relevance to ocean health** There must be a clear connection between the data and ocean health, and determining this will be closely linked to each goal model.

**5.5.4.2 Reference point** As each data layer must be scaled to a reference point, as you consider different data sources it is important to think about or identify what a reasonable reference point may be. Ask the following types of questions as you explore data possibilities:

- Is there a known relationship associated with these data?
- Have policy targets been set regarding these data?
- Would a historic target be appropriate?
- Could a region within the study area be set as a spatial target?

**5.5.4.3 Appropriate spatial scale** Data must be available for every region within the study area.\*

**5.5.4.4 Appropriate temporal scale** Data must be available for at least five years to calculate the trend. Longer time series are preferable because this can be used to set temporal reference points.\*

\* It is not always possible to meet the spatial and temporal requirements with each data layer. In these cases it can still be possible to use these data if appropriate gap-filling techniques are used. It is important that data satisfy as many of these requirements as possible, and in cases where creative ways of working with such data are not possible, it might be better to exclude these data from the analyses and try a different approach.

## 5.5 Notes about data and regions

Final calculated scores by region will be represented on a map in addition to the flower plot. The map displays scores in the exclusive economic zone (EEZ) of the region. This is a visualization that exactly maps the available information of some data layers, but not all. Therefore, available data do not need to have been collected for all of the region's EEZ, but they need to be available for the region itself.

## 5.5.6 Example: US West Coast data discovery

Below are examples of some decisions made when exploring available data for the US West Coast regional assessment. Determining whether certain data could be included started with a good understanding of the data layers and models included in the global assessment, and because the US West Coast is a data-rich region, finer-resolution local data could be used in place of many of the global data layers.

**5.5.6.1 Reasons data were excluded** There are a lot of existing data that contribute to our scientific understanding of ocean processes and interactions but that are not ideal for the OHI. Reasons to exclude data occur both on a practical level (do data adhere to the requirements above?) and on a philosophical level, which requires reflecting on the relationship with ocean health. Some common reasons for excluding data are listed below:

- **The data do not cover the entire area of the reporting region.** The state of California had excellent, long-term data on public attendance at state parks that would have been quite useful in the calculation of the tourism and recreation goal. However, Oregon and Washington did not have these same data so they were not used.
- **There is not a clear and scientifically proven connection between the metric described by the data and ocean health.** Along the US West Coast, kelp beds are a very important habitat because of their contribution to biodiversity and coastal protection. However, kelp coverage is quite



Figure 1: Scores will be presented in the EEZ of each region, even if individual data layers do not all extend to the EEZ

variable and is driven primarily by abiotic natural forcing (wave/storm disturbance and temperature) and thus kelp coverage is not a good metric of ecosystem health. For these reasons kelp coverage was not included in the assessment.

- **The feature being measured may provide benefits to people, but this feature is not derived from the ocean.** Sea walls and riprap provide coastal protection to many people along the US West Coast. However, these structures are not a benefit that is derived from the ocean itself, so only biogenic habitats were included in the calculation of this goal. These data can be included as a pressure due to habitat loss.
- **Data collection is biased and might misrepresent ocean health.** The US Endangered Species Act identifies a species list focused on species of concern within the US. As such, these data are biased in the context of ocean health since they assess only species whose populations may be in danger. For the calculation of the biodiversity goal, using these data would be inappropriate because this goal represents the status of all species in the region, not just those that are currently of conservation concern. Using these data may have shown the status of biodiversity to be lower than it really is because the selection of species to assess was already biased towards species of concern.

#### 5.5.6.2 Creative approaches to using data

- **Time series data are not long enough to calculate a trend or a reference point (when a historical reference point is most appropriate).** For the US West Coast, available data measure the current extent of seagrass habitats, however, these only exist for one time point in most areas so could not be used to calculate the trend or set a historical reference point. As these were the best data available for habitat coverage, we built a model to calculate the status and trend of seagrass habitats using other data that were available over time. A reasonable approach was to model the pressures exerted on seagrasses over time as a proxy for seagrass health.

## 5.6 Formatting Data for the Toolbox

### 5.6.1 Introduction

The Ocean Health Index Toolbox App is designed to work in the programming language **R** using input data stored in text-based *.csv* files (*csv* stands for ‘comma-separated value’; these files can be opened as a spreadsheet using Microsoft Excel or similar programs). Data layers (data input) each have their own *.csv* file that are combined within the Toolbox in model calculations. These data layers are used for calculating goal scores, meaning that they are inputs for status, trend, pressures, and resilience. In the global analysis, there were over 100 data layer files included, and there will be nearly as many in regional applications, no matter what the spatial scale. This document describes and provides examples of how to format data for the Toolbox App.

Ocean Health Index goal scores are calculated at the scale of the reporting unit, which is called a ‘**region**’ and then combined using a weighted average to produce the score for the overall area assessed, called a ‘**study area**’. The OHI Toolbox App expects each data file to be in a specific format, with data available for every region within the study area, with data organized in ‘long’ format (as few columns as possible), and with a unique region identifier (*rgn\_id*) associated with a single *score* or *value*. In order to calculate trend, input data must be available as a time series for at least 5 recent years (and the longer the time series the better, as this can be used in setting temporal reference points).

The example below shows information for a study area with 4 regions. There are two different (and separate) data layer files: tourism count (*tr\_total.csv*) and natural products harvested, in metric tonnes (*np\_harvest\_tonnes.csv*). Each file has data for four regions (1-4) in different years, and the second has an additional ‘categories’ column for the different types of natural products that were harvested. In this example, the two data layers are appropriate for status calculations with the Toolbox because:

1. At least five years of data are available,
2. There are no data gaps
3. Data are presented in ‘long’ or ‘narrow’ format (not ‘wide’ format).

#### Example of data in the appropriate format:

rgn_id	year	count
1	2005	177.14
1	2006	201.39
1	2007	199.81
1	2008	212.99
1	2009	228.81
2	2005	580.98
2	2006	730.18
2	2007	717.00
2	2008	851.44
2	2009	836.68
3	2005	129.69
3	2006	173.45
3	2007	229.86
3	2008	231.44
3	2009	221.95
4	2005	397.00
4	2006	566.00
4	2007	591.00
4	2008	1154.00
4	2009	1570.00

rgn_id	product	year	tonnes
1	ornamentals	2005	10327
1	ornamentals	2006	10389
1	ornamentals	2007	10897
1	ornamentals	2008	9985
1	ornamentals	2009	9001
2	shells	2005	6179
2	shells	2006	6823
2	shells	2007	8239
2	shells	2008	8819
2	shells	2009	9205
3	coral	2005	22079
3	coral	2006	25297
3	coral	2007	25361
3	coral	2008	23817
3	coral	2009	23623
4	shells	2005	7500
4	shells	2006	9700
4	shells	2007	8600
4	shells	2008	9400
4	shells	2009	9300

#### 5.6.2 Gapfilling

It is important that data prepared for the Toolbox App have no missing values or ‘gaps’. Data gaps can occur in two main ways: 1) **temporal gaps**: when several years in a time series in a single region have missing data, and 2) **spatial gaps**: when all years for a region have missing data (and therefore the whole region is ‘missing’ for that data layer).

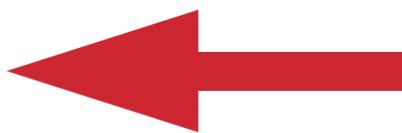
How these gaps are filled will depend on the data and regions themselves, and requires thoughtful, logical decisions to most reasonably fill gaps. Each data layer can be gapfilled using different approaches. Some data layers will require both temporal and spatial gapfilling. The examples below highlight some example of temporal and spatial gapfilling.

All decisions of gapfilling should be documented to ensure transparency and reproducibility. The examples below are in Excel, but programming these changes in software like R easily enables transparency and reproducibility.

**5.6.2.1 Temporal gapfilling** Temporal gaps are when some data are available for a region, but there are missing years. The Toolbox requires data for each year for every region. It is important to make an informed decision about how to temporally gapfilling data.

Many times, creating a linear model is the best way to estimate data and fill temporal gaps. If data do not fit a linear framework, other models may be fit to help with gapfilling. Here we give an example assuming linearity.

rgn_id	year	count
1	2005	177.14
1	2006	
1	2007	
1	2008	212.99
1	2009	228.81
2	2005	580.98
2	2006	730.18
2	2007	717.00
2	2008	851.44
2	2009	836.68
3	2005	129.69
3	2006	173.45
3	2007	229.86
3	2008	231.44
3	2009	
4	2005	397.00
4	2006	566.00
4	2007	
4	2008	1154.00
4	2009	1570.00



In this data layer, there are three regions that have missing values for one or more years and will require temporal gapfilling: Regions 1, 3, 4.

No regions in this example require spatial gapfilling.



Using a linear model can be done in most programming languages using specific functions, but here we show this step-by-step using functions in Excel for Region 1.

\*\* Temporal gapfilling example (assumes linearity):\*\*

There are four steps to temporally gapfill with a linear model, illustrated in the figures with four columns.

### 1. Calculate the slope for each region

The first step is to calculate the slope of the line that is fitted through the available data points. This can be done in excel using the **SLOPE(known\_y's,known\_x's)** function as highlighted in the figure below. In this case, the x-axis is *years* (2005, 2006, etc...), the y-axis is *count*, and the Excel function automatically plots and fits a line through the known values (177.14 in 2005, 212.99 in 2008, and 228.81 in 2009), and subsequently calculates the slope (12.69).

Steps to temporally gapfill data:			
1	2	3	4
Slope	Intercept	$y = mx + b$	Value (final)
		176.70	177.14
		189.39	189.39
		202.08	202.08
		214.78	212.99
12.69	-25273.89	227.47	228.81

### 2. Calculate the y-intercept for each region

The next step is to calculate the intercept of the line that is fitted through the available data points. This can be done in excel similarly as for the slope calculation, using the **INTERCEPT(known\_y's,known\_x's)** function that calculates the y-intercept (-25273.89) of the fitted line.

Steps to temporally gapfill data:			
1	2	3	4
Slope	Intercept	$y = mx + b$	Value (final)
	176.70	177.14	
	189.39	189.39	189.39
	202.08	202.08	202.08
	214.78	214.78	212.99
12.69	-25273.89	227.47	228.81

### 3. Calculate y for all years

The slope and y-intercept that were calculated in steps 1 and 2 can then be used along with the year (independent variable) to calculate the unknown 'y-values'. To do so, simply replace the known three values into the  $y = mx + b$  equation, to calculate the unknown 'count' for a given year (189.39 in 2006, and 202.08 in 2007).

### 4. Replace modeled values into original data where gaps had occurred

	B	C	D	E	F	G	H	I	J
39									
40									
41	<b>rgn_id</b>	<b>year</b>	<b>count</b>						
42	1	2005	177.14						
43	1	2006							
44	1	2007							
45	1	2008	212.99						
46	1	2009	228.81						

Formula       $(\$G\$46 \times C43) + \$H\$46$

Steps to temporally gapfill data:			
1	2	3	4
Slope	Intercept	$y = mx + b$	Value (final)
		176.70	177.14
		189.39	189.39
		202.08	202.08
		214.78	212.99
12.69	-25273.89	227.47	228.81

Substitute these modeled values that were previously gaps in the timeseriew. The data layer is now ready for the Toolbox, gapfilled and in the appropriate format.

**5.6.2.2 Spatial gapfilling** Spatial gaps are when no data are available for a particular region. The Toolbox requires data for each region. It is important to make an informed decision about how to spatially gapfilling data.

To fill gaps spatially, assumptions must be made that one region is like another, and data from another region will be substituted in place of the missing data. This will depend on the type of data and the properties of the regions requiring gapfilling. For example, if a region is missing data but has similar properties to a different region that does have data, the missing data could be ‘borrowed’ from the region with information. Each data layer can be gapfilled using a different approach when necessary.

**Characteristics of region regions requiring gapfilling that can help determine which type of spatial gapfilling to use:**

1. proximity: can it be assumed that nearby regions have similar properties?
2. study area: are data reported for the study area, and can those data be used for subcountry regions?
3. demographic information: can it be assumed a region with a similar population size has similar data?

#### **Spatial gapfilling example:**

For a certain data layer, suppose the second region (rgn\_id 2) has no data reported, as illustrated in the figure above. How to spatially gapfill rgn\_id 2 requires thinking about the properties and characteristics of the region and the data, tourist count.

Here are properties that can be important for decision making:

rgn\_id 2:

- is located between rgn\_id 1 and 3
- is larger than rgn\_id 1
- has similar population size/demographics to rgn\_id 3
- has not been growing as quickly as rgn\_id 4

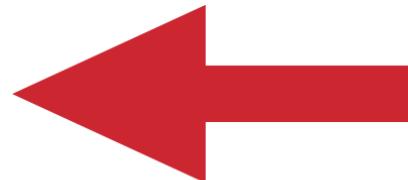
There is no absolute answer of how to best gapfill rgn\_id 2. Here are a few reasonable possibilities:

Assign rgn\_id 2 values from:

rgn_id	year	count
1	2005	177.14
1	2006	201.39
1	2007	199.81
1	2008	212.99
1	2009	228.81
2	2005	
2	2006	
2	2007	
2	2008	
2	2009	
3	2005	129.69
3	2006	173.45
3	2007	229.86
3	2008	231.44
3	2009	221.95
4	2005	397.00
4	2006	566.00
4	2007	591.00
4	2008	1154.00
4	2009	1570.00

In this data layer, Region 2 is missing from this dataset and requires spatial gapfilling.

No temporal gapfilling is required in this example.



- rgn\_id 1 because it is in close proximity to rgn\_id 2
- rgn\_id 3 because it is in close proximity to rgn\_id 2 and has similar population size/demographics
- rgn\_id 1 and 3 averaged since they are in close proximity to rgn\_id 2

Suppose the decision was made to gapfill rgn\_id 2 using the mean of rgn\_id 1 and 3 since this would use a combination of both of those regions. Again, other possibilities could be equally correct. But some form of spatial gapfilling is required so a decision must be made. The image below illustrates this in Excel.



A	B	C	D	E	F	G	H
rgn_id	year	count			rgn_id	year	count
1	2005	177.14			1	2005	177.14
1	2006	201.39			1	2006	201.39
1	2007	199.81			1	2007	199.81
1	2008	212.99			1	2008	212.99
1	2009	228.81			1	2009	228.81
2	2005				2	2005	153.42
2	2006				2	2006	187.42
2	2007				2	2007	214.84
2	2008				2	2008	222.22
2	2009				2	2009	225.38
3	2005	129.69			3	2005	129.69
3	2006	173.45			3	2006	173.45
3	2007	229.86			3	2007	229.86
3	2008	231.44			3	2008	231.44
3	2009	221.95			3	2009	221.95
4	2005	397.00			4	2005	397.00
4	2006	566.00			4	2006	566.00
4	2007	591.00			4	2007	591.00
4	2008	1154.00			4	2008	1154.00
4	2009	1570.00			4	2009	1570.00

Formula: AVERAGE(H43, H53)

The data layer is now ready for the Toolbox, gapfilled and in the appropriate format.

### 5.6.3 Long formatting

The Toolbox expects data to be in ‘long’ or ‘narrow’ formatting. Below are examples of correct and incorrect formatting, and tips on how to transform data into the appropriate format.

#### Example of data in an incorrect format:

With ‘wide’ format, data layers are more difficult to combine with others and more difficult to read and to analyze.

#### Transforming data into ‘narrow’ format:

Data are easily transformed in a programming language such as R.

In R, the `reshape` package has the `melt` command, which will melt the data from a wide format into a narrow format. It also can `cast` the data back into a wide format if desired. R documentation:

		Year									
Region	DataLayer	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
A	GDP_USDx1000	8	7	30	26	69	39	108	92	261	151
B	GDP_USDx1000	13	9	13	14	10	12	14	10	6	5
C	GDP_USDx1000	2132	2325	2963	3214	2942	2910	1759	2029	2077	2453
D	GDP_USDx1000	21	5	14	2	11	3	26	14	15	100
A	governance_indicator						0.8545	0.5400	0.7062	1	1
B	governance_indicator						0.8564	0.7794	0.8615	1	1
C	governance_indicator						0.8779	1	1	0.8986	1
D	governance indicator						0.8537	0.5373	0.7044	1	1

- <http://cran.r-project.org/web/packages/reshape2/reshape2.pdf>
- <http://www.slideshare.net/jeffreybreen/reshaping-data-in-r>
- <http://tgmstat.wordpress.com/2013/10/31/reshape-and-aggregate-data-with-the-r-package-reshape2/>

Example code using the *melt* command in the *reshape2* library. Assume the data above is in a variable called *data\_wide*:

```
install.packages('reshape2')
library(reshape2)
data_melt = melt(data=data_wide, id.vars=c('Region', 'DataLayer'), variable.name='Year')
data_melt = data_melt[order(data_melt$DataLayer, data_melt$Region),]
```

This will melt everything except any identified columns ('Region' and 'DataLayer'), and put all other column headers into a new column named 'Year'. Data values will then be found in a new column called 'value'.

The final step is optional: ordering the data will make it more easy to read for humans (R and the Toolbox can read these data without this final step):

#### Example of data in the appropriate format:

With 'narrow' format, each row of data provides complete and unique information, and does so with as few columns as possible. Data layers in this format can be easily combined with other data layers: the range of years available can be different for each data layer, and there are minimal column names.

## 6 Installing the Toolbox for a Regional Assessment

### 6.1 Overview

The Ocean Health Index (OHI) Toolbox is software written in R that can be used to conduct regional assessments using the OHI framework. Ocean Health Index assessments have a specific organization of data files, registry files, and R scripts that work together to do calculations and present the final scores graphically. The files and scripts for any given assessment are packaged together in a folder called a **repository** ('repos'), and stored online through GitHub, an open-source development platform that enables easy collaboration with versioning capabilities (See: <http://en.wikipedia.org/wiki/GitHub>).

It is possible to conduct a regional assessment by downloading a template repository from GitHub. We recommend having a GitHub user account and syncing with the online repository: this facilitates collaboration within your team but also allows us to provide support when you need it. However, it is possible to work independently without GitHub (after the initial download), although we cannot offer much support in this case.

GDP_USDx1000.csv		
Region	Year	value
A	2000	8
A	2001	7
A	2002	30
A	2003	26
A	2004	69
A	2005	39
A	2006	108
A	2007	92
A	2008	261
A	2009	151
B	2000	13
B	2001	9
B	2002	13
B	2003	14
B	2004	10
B	2005	12
B	2006	14
B	2007	10
B	2008	6
B	2009	5

governance_indicator.csv		
Region	Year	value
A	2000	
A	2001	
A	2002	
A	2003	
A	2004	
A	2005	0.854599407
A	2006	0.540059347
A	2007	0.706231454
A	2008	1
A	2009	1
B	2000	
B	2001	
B	2002	
B	2003	
B	2004	
B	2005	0.856410256
B	2006	0.779487179
B	2007	0.861538462
B	2008	1
B	2009	1

## 6.2 Getting started

**Requirements:** an Internet connection for the initial Toolbox App installation and continued access if syncing with GitHub (highly recommended); proficiency with data management and the coding language **R**; proficiency in spatial software such as ArcGIS (potentially, depending on the regional assessment strategy).

1. Set up: \* Install R and RStudio \* Clone regional assessment repository on Desktop
2. Become familiar with the OHI Toolbox File System

Detailed instructions for step 1 are provided below for users who want to either:

- work with GitHub (See: Accessing repository with GitHub section), or
- work without GitHub (See: Accessing repository without GitHub section).

Once the user has completed steps 1 and 2, they will then be able to make changes to the Toolbox to calculate regional scores (See: ‘Using the Toolbox for a Regional assessment’ section below).

## 6.3 Accessing repositories with GitHub

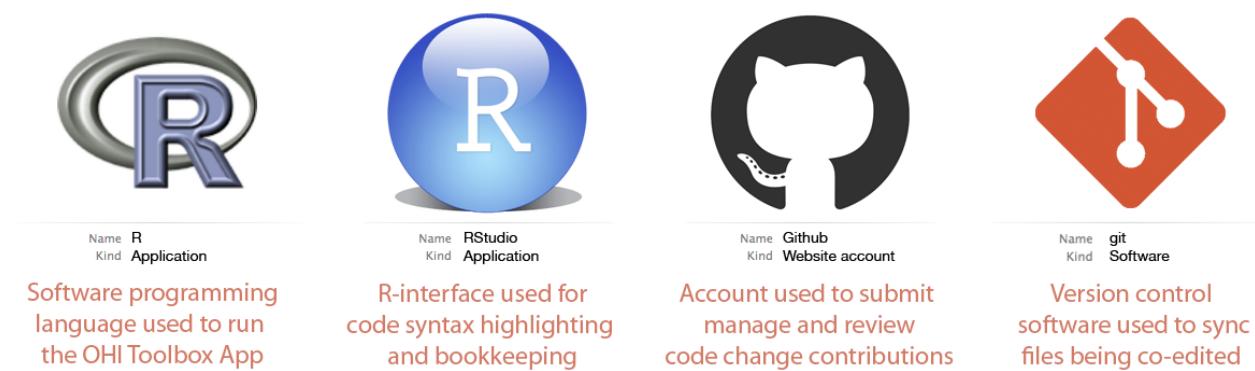
### 6.3.1 Overview

We recommend that groups interested in conducting OHI assessments do so using GitHub, which is an online interface for version-control software called *git*. GitHub has powerful versioning capabilities, which allows changes to be archived and tracked by each user. This is incredibly useful to not only to document what work has been done, but how it differs from work done in the past, and who is responsible for the changes: it provides a record of all changes made. Conducting an OHI regional assessment using GitHub will therefore enable collaboration and transparency, and will provide access to the latest developments in the Toolbox software.

It is also possible to conduct a regional assessment without GitHub. We do not recommend this because it is more difficult for you to track progress and decisions, and for us to help or advise you. However, if this is preferable, see the instructions in the ‘Accessing repositories without GitHub’ section below. We can always get your filesystem onto GitHub later.

To get started on an OHI assessments using GitHub, follow the steps below:

Software required for regional assessments with version control:



### 6.3.2 Install R and RStudio

Make sure you have the most current version of R and RStudio. (Hint: type `sessionInfo()` into the R console to see which version of R you have installed, and update it if this is not the most recent version available on CRAN at <http://cran.r-project.org/>).

### 6.3.3 Create a GitHub account

GitHub is an online interface for *git*, and to use it you must create a GitHub account at <http://github.com>. You will use this username and password when you install *git* and access your GitHub repository (next two steps below).

### 6.3.4 Install *git* software

*git* is version-control software that you will need to install on your computer.

Download *git* at <http://git-scm.com/downloads> and follow the install instructions. Note that there are specific settings to select for **Windows** and **Mac**.

### 6.3.5 Set up your Git Identity

After downloading and installing, you will then need to set up your Git Identity, which identifies you with any changes made. You will use the command line:

- **Mac:** launch the Terminal application (Applications > Utilities > Terminal)
- **Windows:** go to command line in Windows (Start > Run > cmd)

Substitute your GitHub username instead of jdoe:

```
git config --global user.name jdoe
```

and then: substitute the email address you used to create your GitHub account:

```
git config --global user.email john.doe@example.com
```

You can check settings with the following:

```
git config --list
```

Quit the Terminal after typing:

```
exit
```

### 6.3.6 Clone your repo to your computer

OHI-Science is the ‘owner’ of the OHI repositories stored on GitHub at <https://github.com/OHI-Science>, and individual users contribute to these repositories when they have permission. The OHI team will create a repository for your assessment once you provide your GitHub username: email your username to [bbest@nceas.ucsb.edu](mailto:bbest@nceas.ucsb.edu) or [lowndes@nceas.ucsb.edu](mailto:lowndes@nceas.ucsb.edu).

The repository will be backed up and stored on github.com and you will also have a local version on your computer.

**\*\* Important:** Please create a folder called `github` in your root directory and save the repository there. This will make collaborating possible, particularly between Macs and Windows machines. The directory for the assessment `~/github/[assessment]` will therefore be:

- on a Mac: `/Users/[User]/github/[assessment]`
- on Windows: `C:\Users\[User]\Documents\github\[assessment]`

Downloading a repository from GitHub is called **cloning** and there are multiple ways to do it. Cloning allows you to make changes on your local computer and sync them to the online repository.

#### Cloning options:

There are several options to clone the repository to your local machine:

1. **GitHub App**
2. **Command line** (Windows) or **Terminal** (Mac)
3. **RStudio**

#### 1. The GitHub App

(a) For Mac (freely available at <https://mac.github.com/>):

Once the GitHub App is installed and opened, the following page will appear:



- 1 Click “Continue” to connect to GitHub and enter your GitHub.com credentials and/or your GitHub Enterprise credentials, and click “Sign in”. If you have 2 factor authentication enabled on your account (and we’d suggest that you do!) – you’ll be prompted to enter your authorisation code. Once you’re signed in, click “Continue” to move onto the next step.
- 2 GitHub for Mac autofills your Git configuration from your GitHub name and email address. Make sure that these are correct so that Git can correctly attribute all of your commits and then click “Continue”.
- 3 Select any local repositories on your machine to add to the GitHub for Mac app. Don’t worry if you don’t already have any local repositories – you can clone or create them later!

Follow the instructions for cloning on a Mac (copied from: <https://mac.github.com/help.html>)

#### (b) For Windows:

Follow the instructions for cloning on Windows (copied from: <https://windows.github.com/help.html>)

- 1 Download the latest version directly from [windows.github.com](https://windows.github.com).
- 2 When you start the app, you’ll be given the option to either sign into your GitHub account, or create a new one.
- 3 On the left, you’ll see your GitHub account, as well as any organizations you’re a part of. Clicking on a name will show you which repositories are available. Clicking on **clone** brings the repository to your computer.
- 4 Alternatively, you can click on **+ add** at the top of the program, and create a new repository locally.



## 2. The Command line (Windows) or Terminal (Mac)

Below is an example for a repository called ‘Spoon-Knife’:

(Instructions copied from: <https://help.github.com/articles/fork-a-repo#keep-your-fork-synced>)

### 3. RStudio

This is best after the initial clone, since RStudio occasionally has trouble with setting the username / password. Open RStudio, and follow the instructions below copied from: <https://github.com/OHI-Science/ohiprep/wiki/Setup#rstudio>.

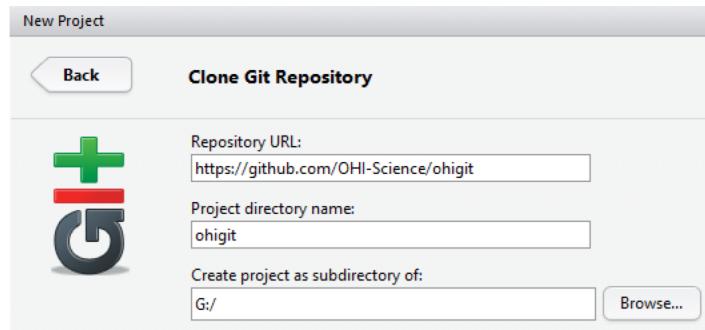
- 1 On GitHub, navigate to **your fork** of the Spoon-Knife repository.
- 2 In the right sidebar of your fork's repository page, click  to copy the clone URL for your fork. This will either be an HTTPS or SSH clone URL, depending on how you authenticate to GitHub.  

- 3 Open Terminal (for Mac and Linux users) or the command line (for Windows users).
- 4 Type `git clone`, and then paste the URL you copied in Step 2. It will look like this, with your GitHub username instead of `YOUR-USERNAME`:
 

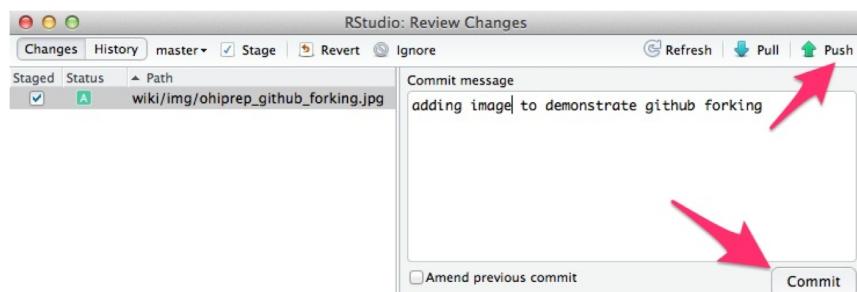
```
$ git clone https://github.com/YOUR-USERNAME/Spoon-Knife
```
- 5 Press **Enter**. Your local clone will be created.
 

```
$ git clone https://github.com/YOUR-USERNAME/Spoon-Knife
Cloning into `Spoon-Knife'...
remote: Counting objects: 10, done.
remote: Compressing objects: 100% (8/8), done.
remove: Total 10 (delta 1), reused 10 (delta 1)
Unpacking objects: 100% (10/10), done.
```

Go to menu File > New Project... > Version Control > Git and enter URL, directory and containing folder. (In example below be sure to substitute 'OHI-Science' with your GitHub username and 'ohigit' with the forked github repository.)

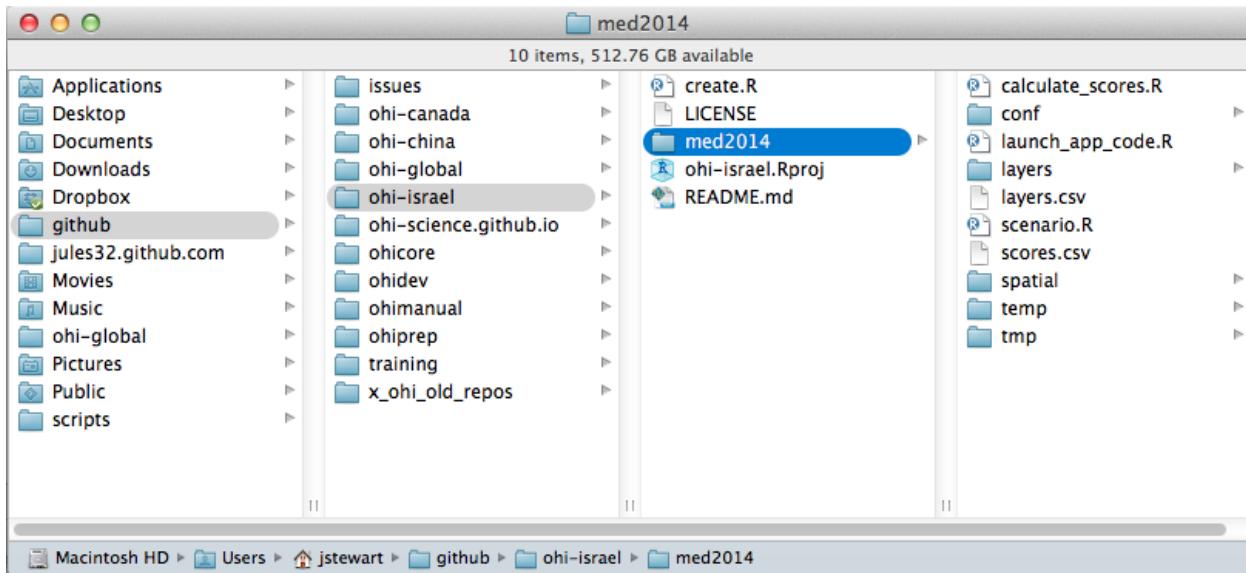


Once you're done, you should then see an extra **Git** pane in RStudio like:



### 6.3.7 Working locally

All of the files available from the online server repository [https://github.com/OHI-Science/\[assessment\]](https://github.com/OHI-Science/[assessment]) are now cloned on your local computer, in the location you specified. We recommend cloning the repository to this file path: `~/github/[assessment]`.



You will then work locally on your own computer, changing the cloned files in `~/github/[assessment]` to reflect the desired modifications your team has identified for your regional assessment. Because there can be multiple users working on the same repo, there are specific steps involved, and GitHub has specific words for each of these steps, including **cloning** a repo from the server to a local repository, **committing** any changes made with a description, **pulling** any updates from the server's repo, and finally **pushing** committed changes back to the server.

All changes within this folder will be tracked by GitHub when you commit and sync these changes back to the repo on the GitHub server, *even if you make these changes outside of R or RStudio*. This means that you can delete or paste files in the Mac Finder or Windows Explorer, edit `layers.csv` and `goals.csv` in Excel or a text editor, and modify `functions.r` in RStudio, and commit all these changes to GitHub to track the changes you make.

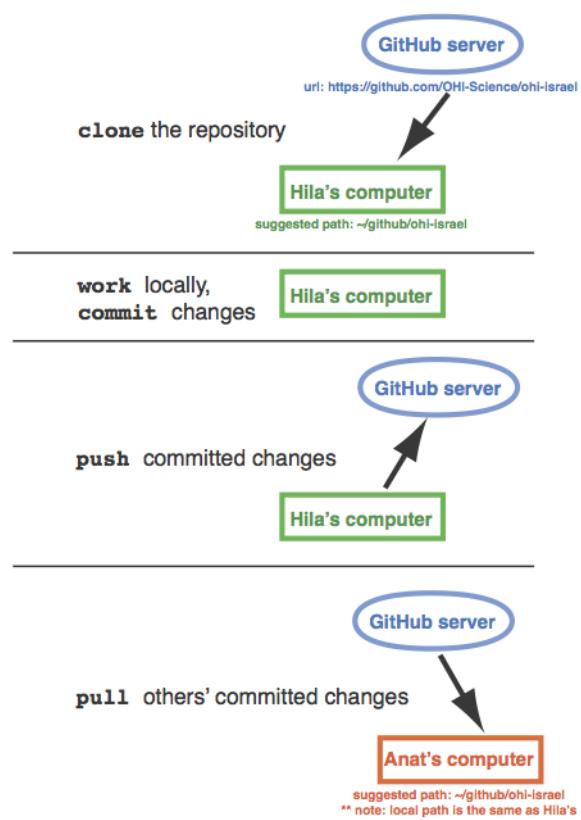
This example illustrates GitHub's collaborative workflow with the `ohi-israel` repo owned by `OHI-Science`:

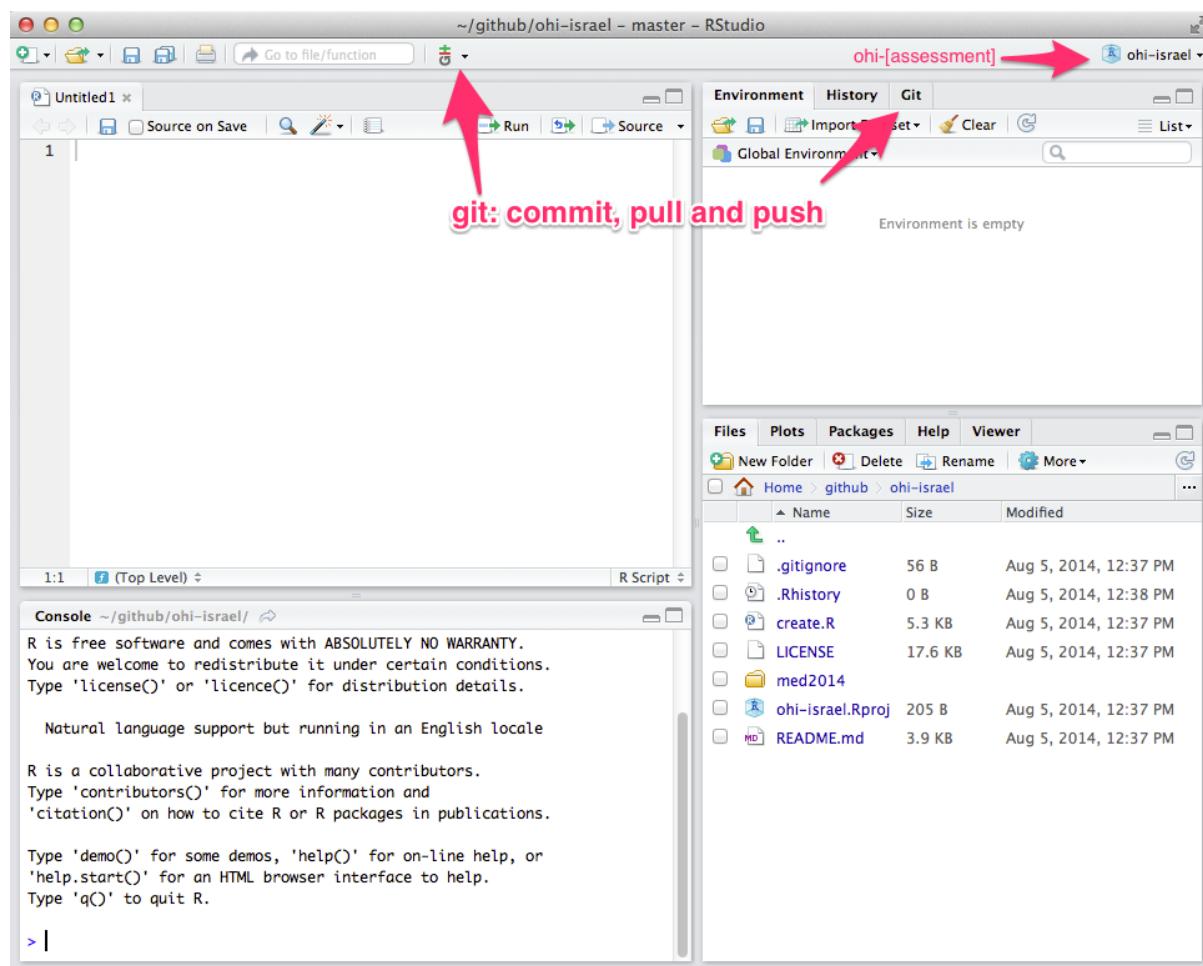
### 6.3.8 Committing and pushing changes locally

Committing and pushing changes can be done using either RStudio or the GitHub App.

#### Using RStudio to commit and push changes to GitHub repository

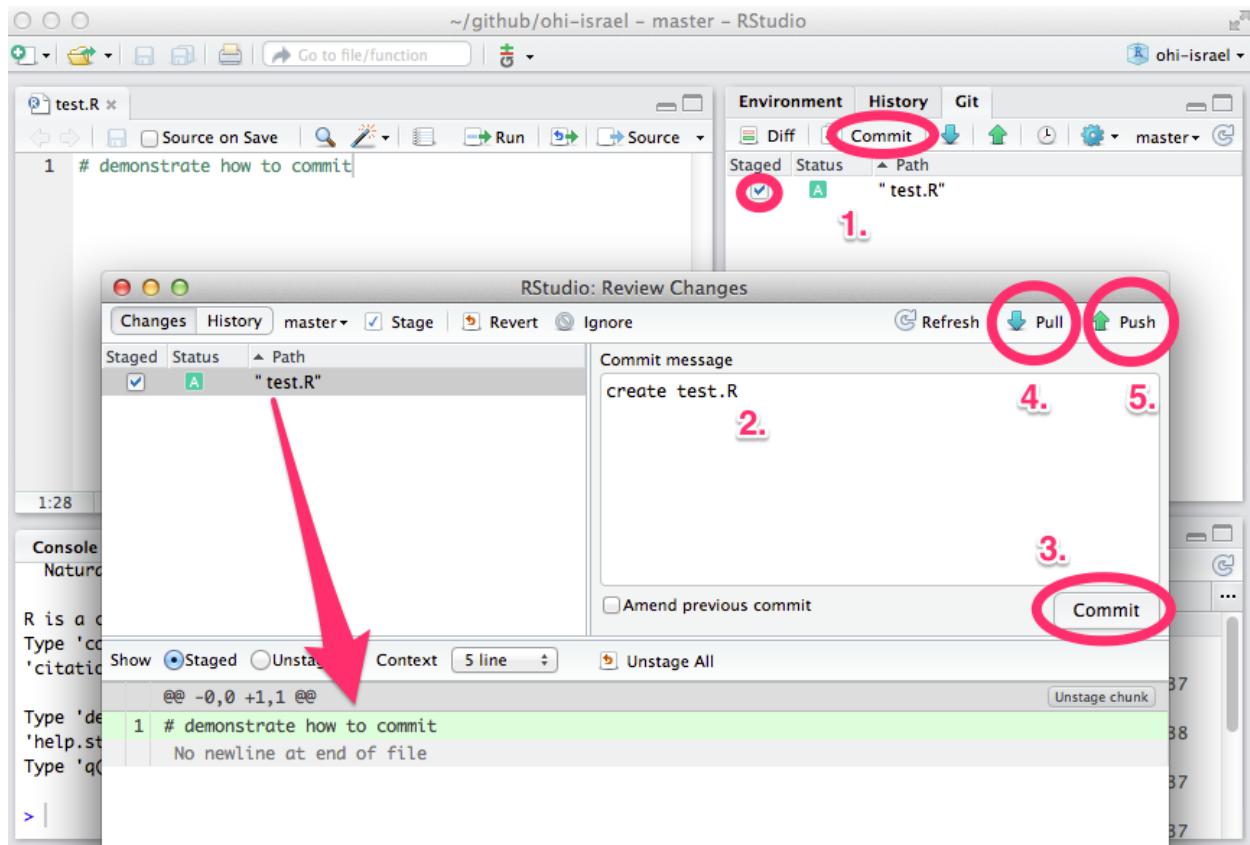
Launch your project in RStudio by double-clicking the `[assessment].Rproj` file. From RStudio, Commit locally, associating a message with each set of changes.





When you modify or add a file, the file will appear in the ‘Git’ window. In the example below, the file `test.R` was created.

1. Clicking the ‘Staged’ box and the ‘Commit’ button opens a new window where you can review changes.
2. Type a commit message that is informative to the changes you’ve made.
- Note 1: there will often be multiple files ‘staged’ at the same time, and so the same commit message will be associated with all of the updated files. It is best to commit changes often with informative commit messages.
- Note 2: clicking on a staged file will identify additions and deletions within that file for your review
3. Click ‘Commit’ to commit the changes and the commit message
4. Pull any changes that have been made to the online repo. This is important to ensure there are no conflicts with updating the online repo.
5. Push your committed changes to the online server at [github.com/OHI-Science/\[assessment\]](https://github.com/OHI-Science/[assessment]). Your changes are now visible online.



### Using GitHub App to push commits to GitHub repository

The GitHub App is freely available for download for both Mac ([mac.github.com](https://mac.github.com)) and PC ([windows.github.com](https://windows.github.com)), and can be used instead of RStudio to commit and sync any changes made locally in the `github` folder. To do so, write a summary and description in the respective message bars in the GitHub App window, then click on commit and sync.

Write a summary/description about the changes you made, then commit and sync

Repositories

Changes made in the github folder appear here

Preview of changes

### 6.3.9 Keeping synchronized

Always pull before pushing commits. This is the best way to avoid merging errors due to other users in your group modifying files.

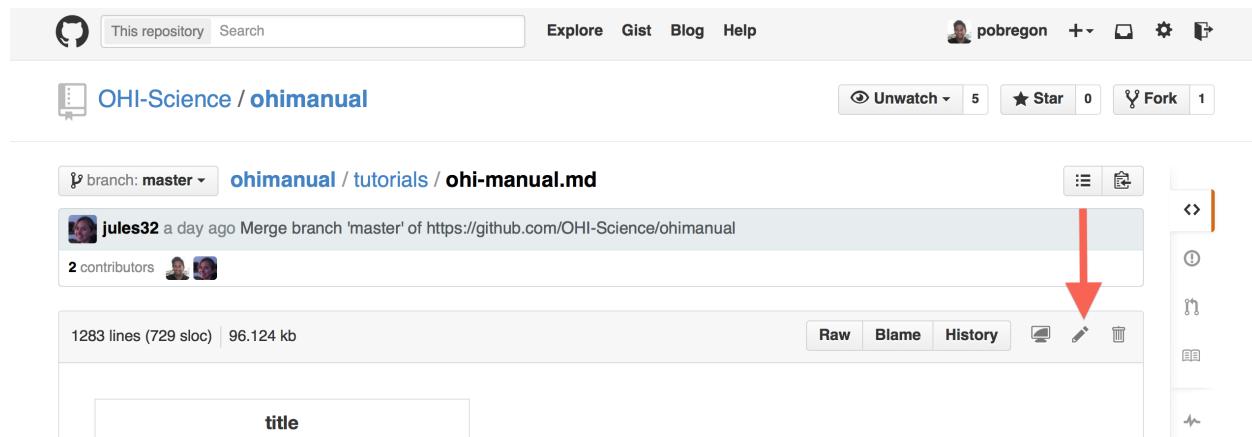
### 6.3.10 More Information

The following is a diagram describing three ways in which files can be edited on GitHub. The third column refers to the ‘clone’, ‘push’ and ‘commit’ model described above:

<code>github.com/ohi-science/[repo]</code> (web)	<code>github.com/[user]/[repo]</code> (web)	<code>~/github/[repo]</code> (local)
->	-> <code>fork</code>	-> <code>clone</code>
<- <code>merge pull request {admin}</code> <-	<- <code>pull request</code>	<- <code>push</code> , <-> <code>commit</code>

where: \* [repo] is one of OHI-Science repositories: ohicore, ohiprep, etc. \* [user] is your GitHub username

Note however that users can also change files directly in the repository on the GitHub website (column 1), by clicking on the button indicated by the red arrow:



See instructions at: <https://help.github.com/articles/fork-a-repo/#keep-your-fork-synced>.

For more information about GitHub, see:

- ‘GitHub flow in the browser’ at: <https://help.github.com/articles/github-flow-in-the-browser>
- A presentation about ‘Reproducible science with the Ocean Health Index’ can be found at: [http://bbest.github.io/talks/2014-06\\_OHI-repro-sci/#1](http://bbest.github.io/talks/2014-06_OHI-repro-sci/#1)
- A wiki about ‘Using GitHub’ can be found at: <https://github.com/OHI-Science/ohiprep/wiki/Using-GitHub>

## 6.4 Accessing repositories without GitHub

### 6.4.1 OHI regional assessments without GitHub

It is possible to conduct a regional assessment without having a GitHub account. You can do a one-time download from the GitHub repository and then work locally on your computer without syncing to the GitHub server. We do not recommend this because it is more difficult for you to track progress and decisions, and for us to help or advise you. We recommend accessing a repository with GitHub but provide instructions here of how to access a repository without GitHub. We can always get your filesystem onto GitHub later.

### 6.4.2 Install the latest version of R

Download R at <http://cran.r-project.org/> and install on your computer. If you already have R installed, check the website for updates. There are frequent updates to the R software, and the current version is identified on the website. Compare what is available from their website with what you already have on your computer by typing `sessionInfo()` into your R console. (This will also identify packages you have installed).

While not required, we highly recommend working with RStudio, which is an interface that makes working with R much easier. RStudio also has great compatibility with GitHub so you can collaborate with your team directly from your R code. RStudio does not get updated as often as R does, but it is good to check for updates regularly.

### 6.4.3 Download the repo to your computer

Contact the OHI team ([bbest@nceas.ucsb.edu](mailto:bbest@nceas.ucsb.edu) or [lowndes@nceas.ucsb.edu](mailto:lowndes@nceas.ucsb.edu)) to create a repository for your group. The repository will be stored on [github.com](https://github.com) and called `OHI-Science/[assessment]`: for example, [github.com/OHI-Science/ohi-israel](https://github.com/OHI-Science/ohi-israel). When your repository has been created, enter the url in to an internet browser (for example, [github.com/OHI-Science/ohi-israel](https://github.com/OHI-Science/ohi-israel)). Click `Download ZIP`.

### 6.4.4 Working locally

Unzip the downloaded folder `[assessment]-master` and save the folder on your computer. We recommend creating a folder called `github` in your root directory and saving your repository to this file path: `~/github/[assessment]`. This will make collaborating much simpler since everyone will work with the same file path.

You will work on your computer to change the files in `~/github/[assessment]` to reflect the desired modifications your team has identified for your regional assessment.

Israel's Ocean Health Index — Edit

24 commits 1 branch 0 releases 1 contributor

branch: master → ohi-israel / +

Update README.md

		latest commit 9c8c0d9f2d
bbest authored 8 days ago	updated tbx screen with scenario ohi-israel/med2014	13 days ago
med2014	ignoring OS-specific launch_app.*	13 days ago
.gitignore	Initial commit	2 months ago
LICENSE	Update README.md	8 days ago
README.md	moved scenario to med2014 subfolder	13 days ago
create.R	initialized Israel scenario files. resolves OHI-Science/ohicore#143	2 months ago

README.md

## ohi-israel: Israel's Ocean Health Index

### OHI Toolbox App with Israel Scenario

HTTPS clone URL  
<https://github.com/>

You can clone with [HTTPS](#), [SSH](#), or [Subversion](#).

Clone in Desktop Download ZIP

## 6.5 OHI Toolbox File System

This section describes the files within the `ohi-[assessment]` folder that you have accessed by either cloning through GitHub and RStudio or downloading to your computer from GitHub.

### 6.5.1 Assessments and scenarios

Within the `ohi-[assessment]` folder is the **scenario** folder. The scenario folder contains all the data, functions and other files required to calculate the Ocean Health Index. To calculate the Index for a different region or with new data or models, you will modify the files within this folder (default data is from the global assessment).

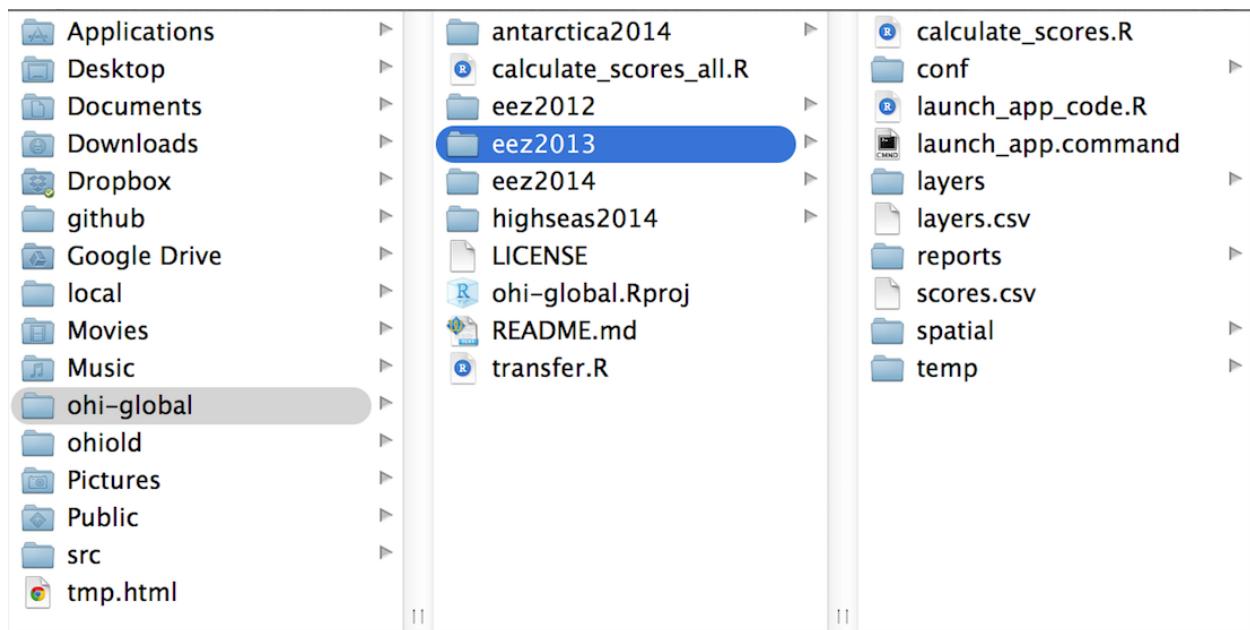
In this example, **ohi-global** is the assessment folder and **eez2013** is the scenario.

### 6.5.2 *layers.csv*

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

Each row of information 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.

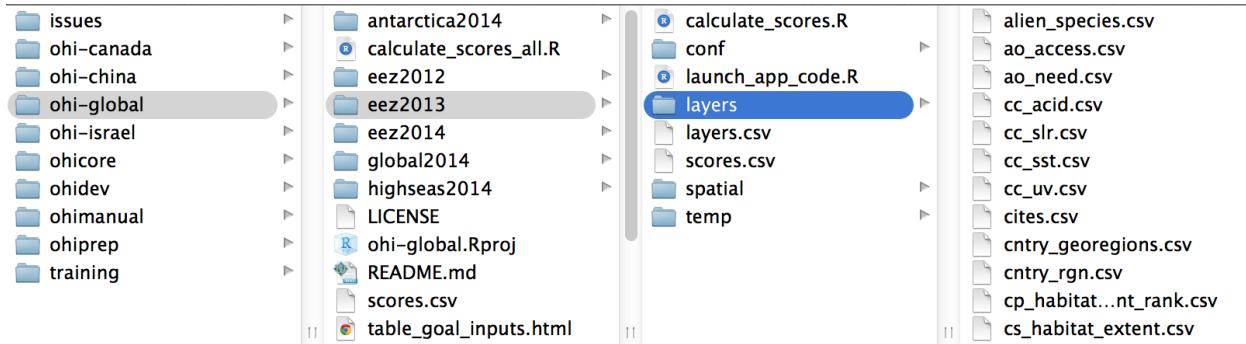


1	targets	layer	name	description	fld_value	units	filename
2	AO	ao_access	Fisheries management	The opportunity for value	value	value	ao_access.csv
3	AO	ao_need	Purchasing power	The per capita purchasing power	value	value	ao_need.csv
4	CW	cw_coastalpopn_trend	Coastal human population	Coastal population trend	trend	score	cw_coastalpopn_trend
5	CW	cw_fertilizer_trend	Fertilizer consumption	Statistics on fertilizer consumption	trend	score	cw_fertilizer_trend
6	CW	cw_pathogen_trend	Trends in access	Trends in percent	trend	score	cw_pathogen_trend
7	CW	cw_pesticide_trend	Pesticide consumption	Statistics on pesticide consumption	trend	score	cw_pesticide_trend
8	FIS	fis_b_bmsy	B/Bmsy estimates	Estimated B/Bmsy values obtained using the B/Bmsy model	B / B <sub>msy</sub>	metric tons	fis_b_bmsy.csv
9	FIS	fis_meancatch	Catch data for each	Reported data including mean catch	mean_catch	metric tons	fis_meancatch.csv

- **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’.

### 6.5.3 *layers* folder

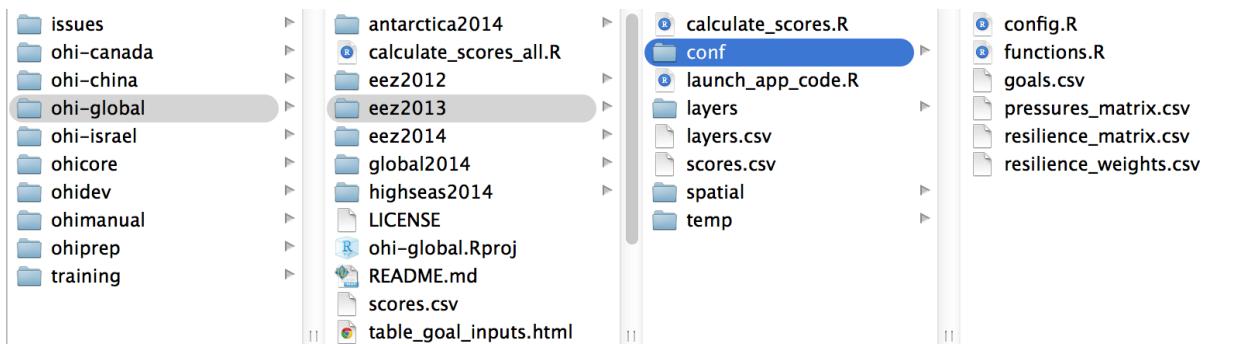
The `layers` folder contains every data layer as an individual `.csv` file. The names of the `.csv` files within the `layers` folder correspond to those listed in the `filename` column of the `layers.csv` file described above. All `.csv` files can be read with text editors or with Microsoft Excel or similar software.



Note that each `.csv` file within the `layers` folder has a specific format that the Toolbox expects and requires. Comma separated value files (`.csv` files) can be opened with text editor software, or will open by default by Microsoft Excel or similar software. Open a `layers/*.csv` file: 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. Please see the section on *Formatting Data for the Toolbox* for further details and instructions.

### 6.5.4 *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`).



**6.5.4.1 config.r** config.r is an R script that configures labeling and constants appropriately.

**6.5.4.2 *functions.r*** *functions.r* contains functions for each goal and sub-goal model, which calculate the status and trend using data layers identified as ‘layers’ in *layers.csv*.

**6.5.4.3 *goals.csv*** *goals.csv* is a list of goals and sub-goals and their weights used to calculate the final score for each goal. Other information includes the goal description that is also presented in the Toolbox App. *goals.csv* also indicates the arguments passed to functions.R. These are indicated by two columns: *preindex\_function* (functions for all goals that do not have sub-goals, and functions for all sub-goals) and *postindex\_function* (functions for goals with sub-goals).

**6.5.4.4 pressures\_matrix.csv** `pressures_matrix.csv` describes the layers ('layers' column in `layers.csv`) needed to calculate pressure categories. The matrix has weights assigned that were determined by Halpern *et al.* 2012 (*Nature*) based on scientific literature and expert opinion (3=high pressure, 1=low pressure). These weights are relative across the rows. Setting these weights will take deliberation with your regional assessment team.

Pressures are grouped by category, indicated by a prefix (for example: *po\_* for the pollution category). Each category is calculated separately before being combined with the others, so it is important to register the new pressure with the appropriate category prefix decided by your regional assessment team.

**6.5.4.5 *resilience\_matrix.csv*** *resilience\_matrix.csv* describes the layers ('layers' column in *layers.csv*) needed to calculate resilience categories.

**6.5.4.6 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.

### 6.5.5 spatial folder

The spatial folder contains a single file, *regions\_gcs.js*. This is a spatial file in the JSON format; it has the appropriate study area and regions for the assessment. This file will be created by the OHI team for all regional assessments.

### **6.5.6 *calculate\_scores.r***

`calculate_scores.r` will run the Toolbox calculations 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`.

### **6.5.7 `scores.csv`**

`scores.csv` is a record of the calculated scores for the assessment (Global 2013 scores). Scores are reported for each dimension (future, pressures, resilience, score, status, trend) for each reporting region, and are presented in ‘long’ format.

### **6.5.8 Relaunching the Toolbox**

After the initial Toolbox setup, further launches of the Toolbox Application can be done without the software program R. Instead, PC users can double-click the `launchApp.bat` file and Mac users can double-click the `launchApp.command` file.

## **7 Using the Toolbox for a Regional Assessment**

This page explains how to incorporate all of the pre-Toolbox decisions your team has made for your regional assessment into the OHI framework and your repository. Having a good understanding of how the Toolbox is structured can also help identify what must be modified for a regional assessment, particularly with data and models. This page assumes a good understanding of the Toolbox file system.

The most common modifications you will make to your repository are changes with:

- updating or adding new data layers for status, trend, pressures and resilience
- modifying pressures and resilience matrices
- modifying goal models
- removing goals

### **7.1 Modifying pressures and resilience matrices**

[develop JSL]

Even before data layers are created and formatted, it is important to modify `pressures_matrix.csv` and `resilience_matrix.csv` (located in the `[assessment]/[scenario]/conf` folder)

snag material from `update_matrices.md`

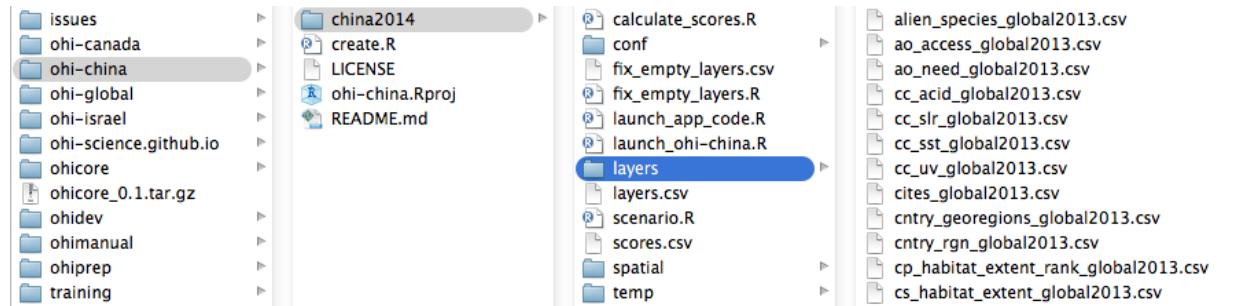
### **7.2 Modifying and creating data layers**

There are several steps to follow when working with data layers:

1. Modify or create data layer with proper formatting
2. Save the layer in the `layers` folder
3. Register the layer in `layers.csv`
4. Check (and update when appropriate) `pressures_matrix.csv` and `resilience_matrix.csv` (located in the `[assessment]/[scenario]/conf` folder)

### 7.2.1 Template data layers

Data layers are `.csv` files and are located in the `[assessment]/[scenario]/layers` folder. All layers provided in your regional assessment repo are the global values from the 2014 assessment.



- Layers with the suffix `_gl2014.csv` ('gl' for 'global') have been exactly copied from the global assessment and applied equally to each region, and therefore the values will be the same across all subcountry regions.
- Layers with the suffix `_sc2014.csv` ('sc' for 'subcountry') have been spatially-extracted from global data or adjusted with spatially-extracted data so that each subcountry region has a unique value. For example, gross domestic product (GDP) used in the global assessment was reported at the national (most often country) level. Instead of being applied equally across all subcountry regions (which would greatly increase the nation's GDP), national GDP was down-weighted by the proportion of coastal population in each region compared with the total coastal population.

Both types of layers are data are at coarse-resolution and should be exchanged for local, high-resolution data when possible. The priority should be to replace as much of the `_gl2014.csv` data as possible.

### 7.2.2 Create data layers with proper formatting

[develop JSL, with Rscript examples]

### 7.2.3 Save data layers in the `layers` folder

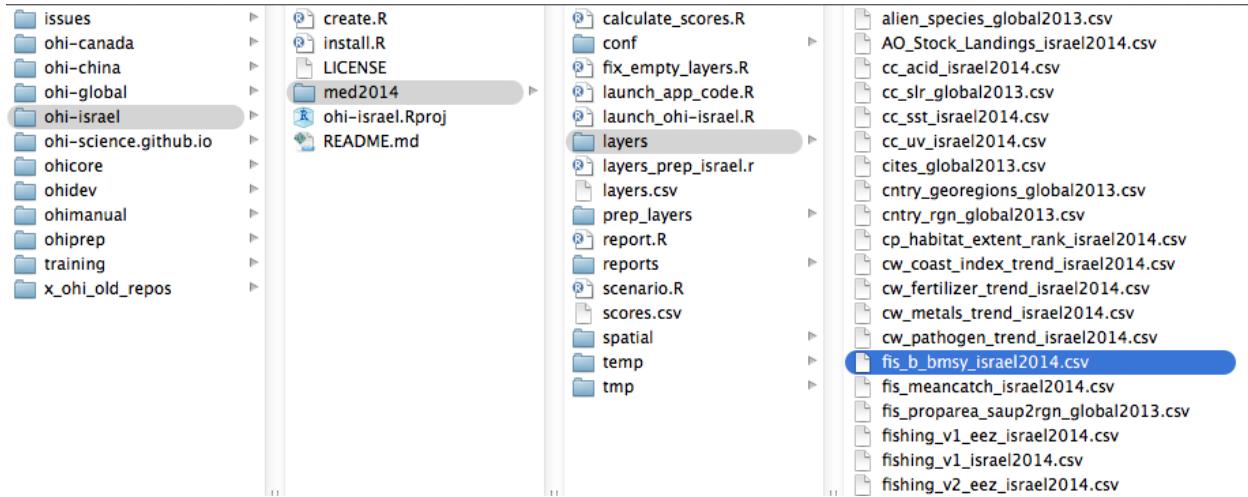
When you modify existing or create new data layers, we recommend saving this as a new `.csv` file with a suffix identifying your regional assessment (example: `_israel2014.csv`). Modifying the layer name provides an easy way to track which data layers have been updated regionally, and which rely on global data. Template layers (`_gl2014.csv` and `_sc2014.csv`) can then be deleted.

### 7.2.4 Register data layers in `layers.csv`

When there are new filenames associated with each layer, they will need to be registered in `[assessment]/[scenario]/layers.csv`. If a layer simply has a new filename, only the `filename` column needs to be updated:

However, if a new layer has been added (for example when a new goal model is developed), you will need to add a new row in the registry for the new data layer and fill in the first eight columns (columns A-H); other columns are generated later by the Toolbox App as it confirms data formatting and content:

- **targets:** Add the the goal/dimension that the new data layer relates to. Goals are indicated with two-letter codes and sub-goals are indicated with three-letter codes, with pressures, resilience, and spatial layers indicated separately.



The screenshot shows an Excel spreadsheet titled 'layers.csv' with the following data:

	A	B	C	D	E	F	G
1	targets	layer	name	description	fid_value	units	filename
2	AO	rk_y ao stock	Fisheries stock landings	The opportunity for a value	value		AO_Stock_Landings_israel2014.csv
3	CW	cw_coast_index_trend	"Clean coast index" trend as a proxy	"Clean coast index" trend	trend.score		cw_coast_index_trend_israel2014.csv
4	CW	cw_fertilizer_trend	Estimates of point-source nitrate a	Estimates of point-so	trend.score		cw_fertilizer_trend_israel2014.csv
5	CW	cw_pathogen_trend	Trend in pathogen indicators statu	Trend calculated from trend	trend.score		cw_pathogen_trend_israel2014.csv
6	CW	cw_metals_trend	Heavy metals as a proxy for trend	Heavy metals from fis	trend.score		cw_metals_trend_israel2014.csv
7	FIS	fis_b_bmsy	B/BMSY estimates obtained using the	B / B_msy			fis_b_bmsy_israel2014.csv
8	FIS	fis_meancatch	Catch data for each Taxon average	mean_catch	metric tons		fis_meancatch_israel2014.csv
9	FIS	fis_proarea_saup2rgn	area of each saup/total OHI report	For converting report	prop_area		fis_proarea_saup2rgn_global2013.csv
10	FP	fp_wildcaught_weight	Fisheries weighting factor	Proportional yield of w_fis	value		fp_wildcaught_weight_israel2014.csv
11	HAB CS	hab_extent	Habitat extent	Sandy shore accordir	km <sup>2</sup>		hab_extent_israel2014.csv

- **layer:** Add an identifying name for the new data layer, which will be used in R scripts like functions.R and .csv files like **pressures\_matrix.csv** and **resilience\_matrix.csv**.
- **name:** Add a longer title for the data layer: this will be displayed in the Toolbox interface.
- **description:** Add a longer description of the new data layer this will be displayed in the Toolbox interface.
- **fld\_value:** Add the appropriate units for the new data layer (which will be referenced in subsequent calculations).
- **units:** Add a description about the ‘units’ chosen in the ‘fld\_value’ column above.
- **filename:** Add a filename for the new data layer that matches the name of the csv file that was created previously in the ‘layers’ folder.
- **fld\_id\_num:** Area designation that applies to the newly created data layer, such as: *rgn\_id* and *fao\_id*.

### 7.2.5 Check pressures and resilience matrices

Next, check **pressures\_matrix.csv** and **resilience\_matrix.csv** to make see if the new or modified layer should be registered or if anything should be altered.

1. New pressure layer

## 8 Update pressures\_matrix.csv

[add content]

## 9 Update resilience\_matrix.csv

When you update layers in the **layers** folder and in **layers.csv**, you will need to update the **resilience\_matrix.csv** file as well.

A few things about these updates: 1. There must be at least one field for each goal. Correct:

goal	component	alien_species	cites	fishing_v1	fishing_v1_ez	fishing_v2_ez	fishing_v3	fishing_v3_ez	habitat	habitat_combi	habitat_combi_ll_gci	ll_sector_even	mariculture	msi_gov	species_diver	species_diver	tourism	water	wgi_all
AO																species_diver	tourism	water	wgi_all
HAB	coral only	alien_species								habitat_combo	habitat_combo_eez	mariculture	species_diversity		species_diversity	tourism	water	wgi_all	
HAB		alien_species								habitat_combo	habitat_combo_eez	mariculture	species_diversity		species_diversity	tourism	water	wgi_all	
HAB	soft_bottom	alien_species								habitat_combo	habitat_combo_eez	mariculture	species_diversity		species_diversity	tourism	water	wgi_all	
HAB	soft_bottom	alien_species								habitat_combo	habitat_combo_eez	mariculture	species_diversity		species_diversity	tourism	water	wgi_all	
SPP		alien_species	cites							habitat_combo	habitat_combo_eez	mariculture	species_diversity		species_diversity	tourism	water	wgi_all	
CS																water	wgi_all		
CW																water	wgi_all		
FIS																water	wgi_all		
MAR																water	wgi_all		
ECO																water	wgi_all		
LIV																water	wgi_all		
NP																	wgi_all		
OP																water	wgi_all		
ICO																water	wgi_all		
LSP																water	wgi_all		
TR																water	wgi_all		

Incorrect:

goal	component	alien_species cites	fishing_v1	fishing_v1_ee	fishing_v2_ee	fishing_v3	fishing_v3_ee	habitat	habitat_combi	habitat_combo_ll_gci	ll_sector_ever	mariculture	msi_gov	species_diver	species_diver_tourism	water	wgi_all
AO								habitat_combo		habitat_combo_eez		mariculture		species_diver	species_diver_3nm	water	wgi_all
HAB	coral only	alien_species							habitat_combo	habitat_combo_eez	mariculture			species_diver	species_diver_tourism	water	wgi_all
HAB	alien_species								habitat_combo	habitat_combo_eez	mariculture			species_diver	species_diver_tourism	water	wgi_all
HAB	soft_bottom	alien_species							habitat_combo	habitat_combo_eez	mariculture			species_diver	species_diver_tourism	water	wgi_all
HAB	soft_bottom	alien_species							habitat_combo	habitat_combo_eez	mariculture			species_diver	species_diver_tourism	water	wgi_all
SPP		alien_species cites							habitat_combo	habitat_combo_eez	mariculture			species_diver	species_diver_tourism	water	wgi_all
CS																	
CW																water	wgi_all
FIS																water	wgi_all
MAR																water	wgi_all
ECO																water	wgi_all
LIV																water	wgi_all
NP																water	wgi_all
CP																water	wgi_all
ICO																water	wgi_all
LSP																water	wgi_all
TR																water	wgi_all

## 9.1 Modifying goal models

In the discussion on data layers above, when an existing layer is still used as before but has a new *filename*, nothing further needs to be done for the Toolbox to incorporate this updated layer. However, if a new layer has been added to the `layers` folder and registered in `layers.csv` (and potentially added to the pressures or resilience matrices), the Toolbox will still not use it unless it is incorporated into a goal model.

There are several steps to follow when working with goal models:

1. Update `functions.r`
2. Check and possibly update `goals.csv`

### 9.1.1 Update `functions.r`

To incorporate a new data layer into a goal model, open `functions.R`: this script contains all the models for each goal and sub-goal. In RStudio, there is a navigation pane that can be used to navigate between them:

```

1 Setup = function(){
2
3   extra.packages.required = c('zoo') # zoo for MAR(), NP()
4   Setup
5   FIS
6   score
7   MAR
8   FP
9
10  AO
11  NP
12  CS
13  CP
14
15  TR
16  LIV_ECO
17  LE
18

```

Navigation options

### 9.1.2 Check and possibly update `goals.csv`

`goals.csv` provides input information for `functions.r`, particularly about goal weighting and function calls. It also includes descriptions about goals and sub-goals, which is presented in the Toolbox Application.

Changing goal weights will be done here by editing the value in the *weight* column. Weights do not need to be 0-1 or add up to 10; weights will be scaled as a percentage of the goal totals. `goals.csv` also indicates the arguments passed to `functions.r`. These are indicated by two columns: *preindex\_function* (functions for all goals that do not have sub-goals, and functions for all sub-goals) and *postindex\_function* (functions for goals with sub-goals).

order_color	order_hierarchy	order_calculate	goal	parent	name	name_flower	description	weight	preindex_function	postindex_function
1	1.2	1	15 FP	Food Provision	Food Provision		This goal measures the amount of seafood sustainably harvested.	1		
3	1.1	1.1	1 FIS	FP	Fisheries	Fisheries	This subgoal model aims to assess the amount of wild-caught fish.	0.5	FIS(layers, status_year=2012)	
4	1.3	1.2	2 MAR	FP	Marculture	Marculture	This subgoal measures the ability to obtain maximal seafood.	0.5	MAR(layers, status_years=2005:2011)	
5	2	2	3 AO	Artisanal Fishing Opport.	Artisanal Fishing Opportunities	Natural Products	This goal captures the access people have to coastal resources.	1	AO(layers, year_max=2012)	
6	3	3	4 NP	Natural Products	Natural Products		This goal model calculates overall status by weighting the subgoals.	1	NP(scores, layers, year_max=2011)	

When updating layers or goal models, it is important to ensure that information called from *goals.csv* is correct:

- check the years
- etc...

## 9.2 Removing goals

If a goal is not relevant in your region, it is possible to remove the goal completely from the calculation. There are four places where you will need to remove the reference to this goal:

1. *functions.r*
2. *goals.csv*
3. *pressures\_matrix.csv*
4. *resilience\_matrix.csv*

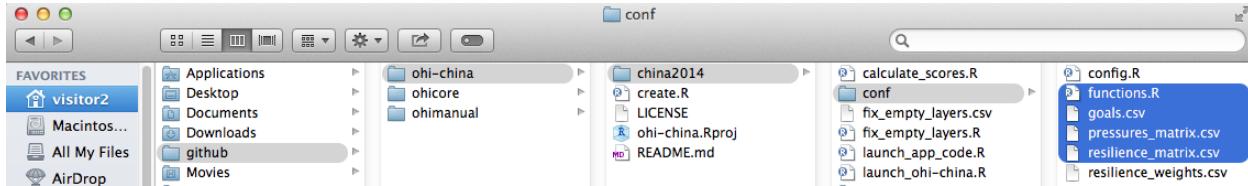


Figure 2: Failing to delete all referenced layers after the goal is deleted will prompt a number of error messages.

## 9.3 Example modifications

### 9.3.1 Adding a new layer to a goal model

In this example we will walk through the following steps:

1. decide to add artisanal access component to the model because of locally available data
  2. prepare the data file; save layer *ao\_access\_art*
  3. register in *layers.csv*
  4. update goal model in *functions.r*
  5. update goal call in *goals.csv*
- 1) and 2) is done outside of the Toolbox
3. register in *layers.csv*
  4. update goal model
  5. [develop]

	A	B	C	D	E	F	G	H
1	targets	layer	name	description	fid_value	units	filename	fid_id_num
2	AO	ao_access	Fisheries management	The opportunity	fid value	value	ao_access_china2014.csv	rgn_id
3	AO	ao_access_art	Example data	Made-up data	value	value	ao_access_art_china2014.csv	rgn_id
4	AO	ao_need	Purchasing power	The per capita pu	value	value	ao_need_global2013.csv	rgn_id
5	CW	cw_coastalpopn_trend	Coastal human popn	Coastal population	trend	trend score	cw_coastalpopn_trend_global2013.csv	rgn_id
6	CW	cw_fertilizer_trend	Fertilizer consumption	Statistics on fertil	trend score	trend score	cw_fertilizer_trend_global2013.csv	rgn_id

referencing new layer



```

325
326 AO = function(layers,
327   year_max=max(layers_data$year, na.rm=T),
328   year_min=max(min(layers_data$year, na.rm=T), max(layers_data$year, na.rm=T)-10),
329   Sustainability=1.0){
330
331 # cast data
332 layers_data = SelectLayersData(layers, targets='AO')
333
334 ry = rename(dcast(layers_data, id_num ~ layer, value.var='val_num',
335   subset = .(layer %in% c('ao_need'))),
336   c('id_num'='region_id', 'ao_need'='need')); head(ry); summary(ry)
337
338 r = na.omit(rename(dcast(layers_data, id_num ~ layer, value.var='val_num',
339   subset = .(layer %in% c('ao_access'))),
340   c('id_num'='region_id', 'ao_access'='access'))); head(r); summary(r)
341
342 ra = na.omit(rename(dcast(layers_data, id_num ~ layer, value.var='val_num',
343   subset = .(layer %in% c('ao_access_art'))),
344   c('id_num'='region_id', 'ao_access_art'='access_art'))); head(r); summary(r)
345
346 ry = merge(ry, r)
347 ry = merge(ry, ra); head(ry); summary(ry); dim(ry)
348
349 # model
350 ry = within(ry,{
351   Du = (1.0 - need) * (1.0 - (access + access_art)/2 )
352   status = ((1.0 - Du) * Sustainability) * 100
353 })
354
355 # status
356 r.status = subset(ry, year==year_max, c(region_id, status)); summary(r.status); dim(r.status)
357
358 # trend

```

modifying 'AO' model



### 9.3.2 Adding a new pressure layer

In this example there are two new pressure layers to include: the effects of desalination operations. The new layers are called *po\_desal\_in*, and *po\_desal\_out*.

Adding these pressure layers to the Toolbox requires the following steps:

1. register (add) pressure layer(s) in **pressures\_matrix.csv**
  - 1a set the pressure category
  - 1b identify the goals affected and set the weighting
  - 1c modify the resilience matrix (if necessary)
2. create pressure layer(s); save in the **layers** folder
3. register pressure layer(s) in **layers.csv**

**9.3.2.1 Register the new layers in *pressure\_matrix.csv*.** **1a.** This step should simply be transferring previous decisions made by your team into the Toolbox format. Create a data layer name with a prefix that signifies the pressure category (for example: *po\_* for the pollution category). Each category is calculated separately before being combined with the others, so it is important to register the new pressure with the appropriate category prefix decided by your regional assessment team.

**1b.** This step is also transferring prior decisions into the Toolbox format. Mark which goals are affected by this new pressure, and then set the weighting. Pressures weighting by goal should be based on scientific literature and expert opinion (3=high pressure, 1=low pressure).

TOTAL PRESSURES																
ECOLOGICAL															SOCIAL	
Pollution					Habitat destruction			Species pollution		Fishing pressure			Climate change		Social	
Chemicals <i>po_chemicals</i>	Chemicals; 3nm <i>po_chemicals_3nm</i>	Pathogens <i>po_pathogens</i>	Nutrients <i>po_nutrients</i>	Nutrients; 3nm <i>po_nutrients_3nm</i>	Trash <i>po_trash</i>	Subtidal softbottom <i>hd_subtidal_sb</i>	Subtidal hardbottom <i>hd_subtidal_hb</i>	Intertidal <i>hd_intertidal</i>	Invasive species <i>sp_alien</i>	Genetic escapes: <i>sp_genetic</i>	Commercial high bycatch <i>fp_com_hb</i>	Commercial low bycatch <i>fp_com_lb</i>	Artisanal low bycatch <i>fp_art_lb</i>	Artisanal high bycatch <i>fp_art_hb</i>	Targeted harvest <i>fp_targetharvest</i>	
CLEAN WATERS	3	3		3	3											1
World Governance Indicators <i>ss_wgi</i>																

**1c.**

**9.3.2.2 Create the new layers** Suppose that there is information

- for each region
- compliance reports
- rescale from 0 to 1
- save in **layers** folder

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	goal	component	component_name	po_desal_in	po_desal_out	po_chemicals	po_chemicals	po_pathogens	po_nutrients	po_nutrients	po_trash	hd_subtidal_si	hd_subtidal_h	hd_intertidal
2	FIS			3	2	2		1		3		2	2	1
3	MAR			3	2	2				1		3		
4	AO						1			2		1	3	1
5	NP	corals		1	2	1				1		1	3	
6	NP	fish oil		1	2	2				1		2		
7	NP	ornamentals		1	2	2				1			3	
8	NP	seaweeds		1	2	2				2				1
9	NP	shells		1	2					1		2		
10	CS	sponges		1	2					1				3
11	CS	mangrove		2	2		1			2				3
12	CS	saltmarsh		2	2		1			3				3
13	CS	seagrass		2	2		2							
14	CP	corai		2	2		1			2			3	
15	CP	mangrove		2	2		1			1			3	
16	CP	saltmarsh		2	2		1			2			3	
17	CP	sealce_shoreline		2	2		1							
18	CP	seagrass		2	2									3
37	LSP						2			3				
38	CW						2			3			2	3
39	HAB	coral			2		1			3		3		
40	HAB	mangrove			2		1			2				3
41	HAB	saltmarsh			2		1			1		2		3
42	HAB	sealce_edge			2									
43	HAB	seagrass			2					3				3
44	HAB	soft_bottom			2					2		3		
45	SPP				2						1	3	2	2

	A	B	C	D	E	F	G	H
1	targets	layer	name	description	fld_value	units	filename	fld_id_num
60	TR	tr_sustainabil	Sustainability in Tourism	Comp score	score		tr_sustainability_global2013.csv	rgn_id
61	TR	tr_unemploym	Percent unemployment	percent	percent	unemplc	tr_unemployment_global2013.csv	rgn_id
62	pressures	po_desal_in	Example data	Made-up data value	pressure score	po_desal_in_china2014.csv	rgn_id	
63	pressures	po_desal_out	Example data	Made-up data value	pressure score	po_desal_out_china2014.csv	rgn_id	
64	pressures	cc_acid	Ocean acidificati	Modeled distri	pressure scor	cc_acid_global2013.csv	rgn_id	
65	pressures	cc_slr	Sea level rise	Modeled sea l	pressure scor	cc_slr_global2013.csv	rgn_id	

### 9.3.2.3 Register the new layers in layers.csv

### 9.3.3 Updating resilience matrix with local habitat information

In this example we will borrow from the experience of `ohi-israel`, where they assessed habitats in the Habitats (HAB) sub-goal that were not included in global assessments `ohi-global`. Therefore, the resilience matrix may need some revision.

The habitats assessed for `ohi-israel` are:

`rocky_reef, sand_dunes, soft_bottom`

#### *Layers affected:*

- `resilience_matrix.csv`
- `resilience_weights.csv` (only if adding new resilience layers)

#### *Scripts affected:*

- none (but may need to create a simple code to generate modified layers)

#### *Default resilience layers:*

The full list of layers used to calculate resilience in `ohi-global` are:

```
alien_species, cites, fishing_v1, fishing_v1_eez, fishing_v2_eez, fishing_v3,
fishing_v3_eez, habitat, habitat_combo, habitat_combo_eez, li_gci, li_sector_evenness,
mariculture, msi_gov, species_diversity, species_diversity_3nm, tourism, water,
wgi_all
```

Some of these layers capture general aspects of governance that apply to the protection of any habitat. These are:

```
alien_species, cites, msi_gov, water, wgi_all
```

Two layers only apply to the livelihoods and economies goal (LE), so they should be excluded from HAB resilience:

```
li_gci, li_sector_evenness
```

The remaining layers will apply to certain habitats, but not others. We focus on these to determine how to adapt the HAB resilience calculation for **ohi-israel**. They are:

```
fishing_v1, fishing_v1_eez, fishing_v2_eez, fishing_v3, fishing_v3_eez, habitat,
habitat_combo, habitat_combo_eez, mariculture, species_diversity, species_diversity_3nm,
tourism
```

#### To determine how to modify these resilience layers:

- If the new habitat occurs only along the coast, we should use **tourism** and **species\_diversity\_3nm**, otherwise, only use **species\_diversity**.
  - **sand\_dunes** should use **tourism** and **species\_diversity\_3nm**,
  - **soft\_bottom** should use **species\_diversity**,
  - is **rocky\_reef** mainly coastal? if so it should use **tourism** and **species\_diversity\_3nm**.
- If the habitats can be affected by mariculture plants (e.g. eutrophication and decreased water quality can occur if mariculture plants are close by and have poor wastewater treatment), then the **mariculture** resilience score should be added.
  - are there any mariculture plants in Israel? If yes, on which habitats do they occur?
- The remaining layers are the **fishing\_v...** and **habitat...** layers, these are composite indicators that we call ‘combo’ layers, obtained from different combinations of the following datasets:

**Mora, Mora\_s4, CBD\_hab, MPA\_coast, MPA\_eez,**

where:

- **Mora** is a fisheries governance effectiveness indicator by Mora *et al* (2009)
- **Mora\_s4** is another indicator from Figure S4 of the supplementary material of the same publication that focuses on regulations of artisanal and recreational fisheries
- **CBD\_hab** is a questionnaire compiled by countries that committed to Rio’s Convention on Biodiversity (CBD) to establish their progress towards habitat biodiversity protection
- **MPA\_coast** is the proportion of coastal (3nm) waters that are in a marine protected area (MPA), with the maximum being 30% of coastal waters
- **MPA\_eez** is the proportion of the whole EEZ that is in a marine protected area, with the maximum being 30% of the whole EEZ.

This table shows which data-sets are used by each combo layer:

Layer	Mora	Mora_s4	CBD_hab	MPA_coast	MPA_eez
fishing_v1	Mora		CBD_hab	MPA_coast	
fishing_v1_eez	Mora		CBD_hab		MPA_eez
fishing_v2_eez	Mora	Mora_s4	CBD_hab		MPA_eez
fishing_v3		Mora_s4	CBD_hab	MPA_coast	
fishing_v3_eez		Mora_s4	CBD_hab		MPA_eez
habitat			CBD_hab		
habitat_combo			CBD_hab	MPA_coast	
habitat_combo_eez			CBD_hab		MPA_eez

### Questions to consider:

- 1) For which habitats should you use both a fishery and a habitat combo, or just use a habitat combo?
  - fisheries regulations do not affect the conservation of sand-dunes, so this habitat should not use any of the fisheries combos. Also, this is a strictly coastal habitat, so choose the habitat layer that uses the **MPA\_coast** instead of the **MPA\_eez**, i.e. **habitat\_combo** (and, as mentioned above, choose the coastal version of biodiversity, i.e. **species\_diversity\_3nm**).
  - The rocky reef and soft bottom, on the other hand, should definitely include fisheries regulations. So you'll need to choose a fisheries and a habitat combo for these two habitats.
- 2) Which fisheries and habitat combos for **rocky\_reef** and **soft\_bottom**? The choice depends on two things:
  - whether they are coastal habitats (within 3nm of the coast) or EEZ-wide habitats
    - if coastal, use the fisheries and habitat combos with **MPA\_coast** (**fishing\_v1**, **fishing\_v3**, **habitat\_combo**), and the **species\_diversity\_3nm** layer
    - if EEZ-wide, use the fisheries and habitat combos with **MPA\_eez** (**fishing\_v1\_eez**, **fishing\_v2\_eez**, **fishing\_v3\_eez**, **habitat\_combo\_eez**), and the **species\_diversity** layer
  - whether the fisheries occurring on that habitat are mainly artisanal, mainly commercial, or both
    - if only commercial fisheries, use a layer that only uses the **Mora** data **fishing\_v1..**)
    - if only artisanal/small-scale fisheries, use a layer that only uses the **Mora\_s4** data (**fishing\_v3..**)
    - if both, use a layer that uses both **Mora** and **Mora\_s4** data (**fishing\_v2..**)
- 3) Are the existing combo layers appropriate or do you need an ad-hoc version for any of the Israel habitats?
  - if rocky reef is mainly coastal, and it is fished by both commercial and artisanal methods, then we need a new combo, specifically, we need a combo that uses **Mora**, **Mora\_s4**, **CBD\_hab**, and **MPA\_coast** (this is the same as **fishing\_v2\_eez**, but we use the **MPA\_coast** layer instead of the **MPA\_eez**). All other combinations are already present.
- 4) Are there local data to be used?

- if there are local data on Marine Protected Areas (MPAs) and any areas with special regulations, this should be used to generate the `MPA_coast` and `MPA_eez` layers. \*\*NOTE: these are the same datasets used to calculate the status of Lasting Special Places (LSP).

## 5) How to update `resilience_matrix.csv`?

- write the complete list of layers you want to use for each habitat. Based on the above, for example, `soft bottom` in Israel matches the combination of layers called `soft bottom, with corals` in the default `resilience_matrix.csv`. But the `rocky_reef` and `sand_dunes` don't seem to match any existing combination, so you'll probably need to delete some of the rows, e.g. the `coral only`, and replace with new ad-hoc rows.

### 9.3.4 Removing a goal

#### 1) Remove the carbon storage (CS) goal model from `functions.r`:

```

627     return(scores_NP)
628 }
629
630 CS = function(layers){
631
632   # layers
633   lyr$ = list('rk' = c('hab_health' = 'health',
634                 'hab_extent' = 'extent',
635                 'hab_trend' = 'trend'))
636   lyr_names = sub('^\\w*$', '', names(unlist(lyr$)))
637
638   # cast data
639   D = SelectLayersData(layers, layers=lyr_names)
640   rk = rename(dcast(D, id_num + category ~ layer, value.var='val_num', subset = .(layer %in% names(lyr$[['rk']])),
641               c('id_num'='region_id', 'category'='habitat', lyr$[['rk']])))
642
643   # limit to CS habitats
644   rk = subset(rk, habitat %in% c('mangrove','saltmarsh','seagrass'))
645
646   # assign extent of 0 as NA
647   rk$extent[rk$extent==0] = NA
648
649   # status
650   r.status = ddply(na.omit(rk[,c('region_id','habitat','extent','health')]), .(region_id), summarize,
651                   goal = 'CS',
652                   dimension = 'status',
653                   score = min(1, sum(extent * health) / sum(extent)) * 100)
654
655   # trend
656   r.trend = ddply(na.omit(rk[,c('region_id','habitat','extent','trend')]), .(region_id), summarize,
657                   goal = 'CS',
658                   dimension = 'trend',
659                   score = sum(extent * trend) / sum(extent) )
660
661   # return scores
662   scores = cbind(rbind(r.status, r.trend))
663   return(scores)
664 }
665
666
667 CP = function(layers){
668 }
```

630:1 | CS ▾ R Script ▾

Figure 3: Delete the highlighted text that references the CS layers and calculates CS goal status, trend, and scores

#### 2) Remove the CS row from `goals.csv`:

#### 3) Remove all CS rows from `pressures_matrix.csv`:

	A	B	C	D	E	F	G	H	I	J
1	order_color	order_hierarch	order_calculat	goal	parent	name	name_flower	description	weight	preindex_func
2	1.2	1	15	FP		Food Provision	Food Provision	This goal mea	1	
3	1.1	1.1	1	FIS	FP	Fisheries	Fisheries	This subgoal n	0.5	FIS(layers, sta
4	1.3	1.2	2	MAR	FP	Mariculture	Mariculture	This subgoal n	0.5	MAR(layers, st
5	2	2	3	AO		Artisanal Fishi	Artisan	This goal capti	1	AO(layers, yea
6	3	3	4	NP		Natural Produ	Natural Produ	This goal mod	1	NP(scores, laye
7	4	4	5	CS		Carbon Storag	Carbon Storag	This goal capti	1	CS(layers)
8	5	5	6	CP		Coastal Protec	Coastal \nProte	This goal mea	1	CP(layers)
9	6	6	7	TR		Tourism & Rec	Tourism & \nR	This goal capti	1	TR(layers, year
10	7.2	7	16	LE		Coastal Liveli	Coastal Liveli	This goal aims	1	

Figure 4: Delete the highlighted row that contains the CS goal

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	goal	component	component_n	po_desal_in	po_desal_out	po_chemicals	po_pathogens	po_nutrients	po_nutrients	po_trash	hd_subtidal_si	hd_subtidal_h	
9	NP	shells		1	2				1			2	
10	NP	sponges		1	2				1			3	
11	CS	mangrove		2	2		1			1			
12	CS	saltmarsh		2	2		1			2			
13	CS	seagrass		2	2		2			3			
14	CP	coral		2	2		1			2			3

Figure 5: Delete the highlighted rows that contain CS pressures

- 4) Remove all CS rows from `resilience_matrix.csv`:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	goal	component	alien_species	cites	fishing_v1	fishing_v1_ee:fishing_v2_ee:fishing_v3	fishing_v3_ee:habitat	habitat_combo	habitat_combx_ll_gci	habitat_combx_ll_gci	habitat_combo_eez	habitat_combo	habitat_combo
7	SPP		alien_species	cites		fishing_v2_eez							
8	CS												
9	CS	mangrove only											
10	CW						fishing_v2_eez						habitat_combo_eez
11	FIS												

Figure 6: Delete the highlighted rows that contain CS resilience

## 9.4 Notes about R

The Toolbox is written in R, and relies heavily on a few R packages created to facilitate data handling and manipulation. 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: <https://github.com/hadley/dplyr#dplyr> for documentation.

## 10 Frequently Asked Questions (FAQs)

This document provides answers to some frequently asked questions about conducting regional assessments using the Ocean Health Index. A few questions are related to general concepts in the Ocean Health Index, but mostly those topics are covered at <http://www.oceanhealthindex.org/About/FAQ/>. Here, the FAQ are primarily technical questions regarding regional assessments and using the OHI Toolbox. This document will be updated continually as we have more questions. Questions are arranged by theme, and have the format Q: (question) and A: (answer).

### 10.1 Overall

### 10.2 Conceptual

**Q:** Are regional assessment scores comparable with global assessment scores?

A: Regional Index scores cannot be directly compared to global Index scores, or to other regional Index scores calculated through separate efforts. This is because data and indicators (both what they measure and their quality), reference points (set using local knowledge and priorities), and specific goal models are often different for the areas being compared.

However, because scores for each goal are scaled to a reference point, qualitative comparisons can be made. For example, a score of 71 in the US West Coast compared to 66 in Brazil says that the US West coast is closer to fully meeting its sustainable goals (i.e., meeting regional reference points). Furthermore, use of the same Ocean Health Index framework across regional assessments permits fruitful discussion and general comparisons even if data inputs differ. Ocean Health Index assessments at any scale always work within a standardized definition of ocean health, using information to capture the philosophy of the ten goals that have been identified (and undergone scientific peer-review) prior to compiling relevant data. Use of the ten-goal framework is important both to ensure that all aspects of ocean health are captured and to allow better comparison across regional assessments than would be possible if the different regions used different methods.

**Q: Where is climate change measured in the Index?**

A: Four different aspects of climate change – increases in sea surface temperature (SST), sea level rise (SLR), ultraviolet radiation (UV), and ocean acidification (OA) – are included as pressures to many goals in the Index, including Natural Products, Carbon Storage, Coastal Protection, Sense of Place, Livelihoods & Economies and Biodiversity. Mitigation of climate change through carbon storage is one of the ten goals.

**Q: Why are food provision and artisanal fishing opportunities goals separated?**

A: These goals measure different aspects of how people relate to fishing. The catch of fish made by artisanal (=small-scale, subsistence type) fisheries is captured in the food provision goal. Jobs, wages and income from both the food provision and artisanal fishing goals are captured in the livelihoods & economies goal. The purpose of the artisanal fishing opportunity goal is to evaluate the opportunity for people to pursue this fishing in relation to their need to do so.

### 10.3 Timing and Resources

**Q: How much does it cost to produce a regional assessment?**

A: Regional assessments can be completed at(varying costs depending on the local context.(Funds are needed for a management and scientific team, workshops and meetings (including travel), communications, policy engagement, and operating costs. Therefore, securing funding is an important component to satisfactorily complete the assessment. We encourage the development of a local proposal or strategic action plan that details a timeline of activities and the resources needed to accomplish them.

**Q: How many people are required in a team?**

A: rather than a specific number of individuals, what is required are specific skillsets. For example, if the scientific analysts were capable of effectively conducting the R analysis, then a dedicated R analyst would not be required. In current assessments, teams range between 2 and 8 people.

**Q: How long does it take to calculate OHI at a regional scale?**

A: This depends on many things: budget and number of people involved, the scale of the study area and whether new regions will need to be created, how easily data can be acquired, how much local data can be incorporated, how many goal models need to be changed. Additionally, decisions about setting reference points require input from experts.

**Q: How much time will modifications by an R analyst take?**

A: This will depend on if you are changing any models, and potentially data layers—but a lot of changing data layers just requires registering them properly in layers.csv (and maybe pressures\_matrix.csv and resilience\_matrix.csv if they are pressures or resilience files) and having the functions.r file call those layers. That is more ‘bookkeeping’ than actual R programming.

**Q: How much time will modifications by a GIS analyst take?**

A: this will depend on how many layers you are processing: you are clipping spatial data? That will take some time because there are quite a few files, but maybe not too long since it is pretty small scale and once there is a clipping mask created I think you apply it to other files.

**Q: Which goals require a GIS analyst?**

A: All goals using spatial data could potentially require a GIS analyst. These goals are commonly: habitat-based goals and sub-goals: (Coastal Protection, Carbon Storage, Habitats—a sub-goal of Biodiversity), Food Provision, Sense of Place, Species—a sub-goal of Biodiversity, Clean Waters

## 10.4 Structure

**Q: Can we remove or add goals to the OHI?**

A: A lot of deliberation went into defining the ten goals, and they seem to do a pretty good job of covering many if not most ocean uses, so additional goals may not be necessary. But it could be that they eclipse or replace an existing goal.

## 10.5 Reference points

**Q: Can planning targets can be used as the reference points?**

A: Yes, planning targets can be used as reference points. This won't be appropriate for every goal, but there are cases where this seemed best (example: iconic species sub-goal in the global assessment, mariculture sub-goal in the US West Coast assessment).

**Q: What is sector evenness?**

A: Sector evenness (also called a diversity index) is an economic concept that is included in OHI to enable comparison across many different sectors included in the Livelihoods & Economies goal. This goal evaluates jobs, wages and revenues for nine marine employment sectors. The distribution of employment across these nine sectors is an effective indicator of resilience. If total employment within a community is primarily based in one or two sectors, the overall economic system will be excessively vulnerable to downturns in those sectors. Conversely, if employment is spread relatively evenly throughout all nine sectors, the overall system will be more robust and resistant to such disturbances. Overall revenue within the community will remain more stable during such downturns, and workers displaced by a downturn in their sector may be able to find employment in another sector without leaving the community.

## 10.6 Appropriate data layers

**Q: Shipping and port activity are hardly affected by the health of the ecosystem. Why are these included in the Index?**

A: Shipping and port activity are included as pressures only

**Q: Can oil spills be included in OHI?**

A: Yes, oil spills could be included as a pressure and in the Clean Waters goal.

**Q: Is seasonal (non-permanent) sea ice included in OHI habitats?**

A: No, sea ice only includes permanent sea ice.

**Q: Can seaweeds be included in the Carbon Storage goal?**

A: Because they store carbon for less than 100 years, seaweeds and corals are not included in the carbon storage goal. While the pelagic oceanic carbon sink (phytoplankton) plays a large role in the sequestration

of anthropogenic carbon, the pelagic ocean mechanisms are not amenable to local or regional management intervention. Phytoplankton contribute to carbon fixation when they die and sink to the sea bottom at sufficient depth, because it is effectively out of circulation. However, if those phytoplankton are eaten, the carbon is cycled back into the system and not sequestered. Something that could potentially be included in the carbon storage goal is mollusc shells, if they are added to a landfill and not recycled in the sea. So if information on mariculture production and waste disposal are available, this could be an interesting addition to carbon storage at a regional scale.

**Q: Is coastal engineering included in Coastal Protection? What if it reduces erosion?**

A: We did not include an assessment of the protection afforded by man-made structures, such as jetties and seawalls, because these structures cannot be preserved without maintenance, may have other negative side effects (e.g. alter sedimentation rates causing erosion in new locations), thus they do not constitute long-term sustainable services. Coastal engineering (jetties, harbours, marina and breakwater) is not natural, and is mostly seen as a pressure. It will also be evident in the status of due to decreased natural habitat. It gets tricky when structures are built to help reduce coastal erosion—they are still manmade and therefore not a natural benefit that the ocean provides. But if available data allow, it might be possible to include tradeoff effects: maybe in areas where natural habitats are degraded and man-made structures have been built to reduce erosion, we could reduce the pressure that would otherwise be applied.

**Q: How is seawater used for cooling on-shore power plants incorporated into OHI?**

A: The use of cooling water for on-shore power plants would be a pressure on the ocean, since it causes entrapment of fishes, larvae, etc, and usually is circulated back into the ocean at higher temperatures (and maybe other chemicals, minerals, etc). Since the energy is coming from land-based activities, there isn't a service that the ocean is providing that 'benefits' people, it is only a pressure from the OHI perspective.

**Q: How is freshwater production through desalination incorporated into OHI?**

A: Desal would be incorporated into OHI in several places. The benefit is that there is freshwater produced, which could be incorporated into the Natural Products goal (or potentially into its own goal). Data required would be the volume of freshwater created based on the volume of seawater involved and spatial extent. Setting the reference point would not be based on how much can be produced, but some other targets perhaps set by government (percentage of the population served). Similar to the mariculture sub-goal and tourism goals, any negative effects caused by desal that affect other goals (example: species) do not influence the ability to obtain desalination targets now and in the future. Therefore, the sustainability coefficient only measures the ability to sustain that goal, but not the impacts on other goals: instead, they are taken into account as pressures when calculating the other goals. Desal should be included as a pressure similar to cooling on-shore power plants since the discharge brine is dense, doesn't plume very well and there are chemicals involved.

## 10.7 Food Provision

**Q: Could the culture of marine fish in closed pools on-shore be included in the Mariculture sub-goal?**

A: This should not be included because onshore aquaculture does not require a marine environment.

**Q: Can aquaculture farms that receive seawater supply and return seawater back to the sea be included in the food provision goal?**

A: This would be more appropriately included in the Mariculture sub-goal, and with finer-scale data additional pressures due to the intake pipes and the processed brine back into the marine system could be incorporated as well. Natural Products

**Q: If natural products are all produced through on-land aquaculture, should this goal be removed?**

A: In this case you would probably have good reason to exclude the natural product goal due if this was defendable through discussions with experts and any reports/papers on the topic. This would also depend on the origin of these natural products—are they from the region's waters? Habitat-based goals

**Q: How is coral health calculated?**

A: Coral health was estimated by compiling point data from multiple studies of percent live coral cover. In other words, estimates of coral cover within transects of certain sites were repeated in time and we used that rate of change in time as an indication of health of the reefs in the whole region. The difficulty lies in 1) having enough different locations sampled that you can say something about the whole region and 2) finding studies that did repeated measures in time, in the same location, over at least 20 years. In the Global 2013 assessment, there were so few datasets that satisfied this condition that we had to pool observations from different locations.

**Q: Is it possible to calculate habitat goals when there is only one year of habitat data?**

A: With only one year of habitat data, it is not possible to calculate the trend (which requires 5 years of data). Instead, it might be best to use the available habitat data to calculate the current status and then to overlay pressures for the last 5 years to calculate trend.

## 10.8 Livelihoods & Economies

**Q: Benefits gained from Wild-caught fisheries, Mariculture, Tourism & Recreation are included in specific goals. Why are these counted again in Livelihoods & Economies?**

A: The quantity of fish, mariculture, and participation in T&R are considered separately in goals whereas the monetary component is captured in L&E.

**Q: Why are revenue data from shipping, boat building, ports and harbors included as revenue? Do these activities rely on a healthy ocean?**

A: These sectors are included in the Ocean Health Index because the demand for some of those boats (fishing boats, sailboats, yachts) is dependent on a healthy ocean.

**Q: Why isn't oil and gas industries included in revenue?**

A: The Natural Products goal does not include non-living items such as oil, gas, and mining products, because these practices are not considered to be sustainable. They are also done at such large scales that including them would essentially make OHI an index for oil and mining—and they are not truly an ocean product. Because these products are not included in terms of quantity extracted, it did not seem appropriate to include information regarding jobs, wages or revenue.

## 10.9 Tourism & Recreation

**Q: How do I calculate the sustainability term for TR?**

A: The best way is to use a local indicator or measure of tourism sustainability or competitiveness, otherwise use the TTCI value from the Global 2013 assessment for the study area (applied evenly across all regions).

## 10.10 Natural Products

**Q: Where do Natural Products come from?**

A: In the global assessments, Natural Products data come from the UN's Food and Agriculture Administration ([www.fao.org/fishery/statistics/software/fishstatj/en](http://www.fao.org/fishery/statistics/software/fishstatj/en)). These data are compiled and reported by product for each country, and available by downloading the FishStatJ software.

## 10.11 Sense of Place

**Q:** Data are only available for marine protected areas, not terrestrial protected areas. Can we still calculate the Lasting Special Places sub-goal?

A: Yes, it is possible to calculate only the marine component of this sub-goal: this is not ideal but OHI is flexible to work with the data available.

## 10.12 Pressures

**Q:** How are single ecological pressures (si in Equation S8) calculated?

A: Data included in pressures calculations are accessed in the same manner as any other data layer, and rescaled from 0-1 with an appropriate reference point. For further information, see HowTo\_GatherAppropriateData and HowTo\_CalculatePressures from ohi-science.org.

**Q:** Does the pressures matrix need to be changed?

A: It is likely that the pressures matrix will not need to be changed. The weights assigned in the matrix were set using information from the literature and by experts; the matrix was created by Halpern et al. 2012.

**Q:** How is commercial high and low bycatch calculated?

A: Commercial high and low bycatch are categorical values that were set based on fishing gear type. This began as a list of gear types used, producing a range of potential bycatch frequencies (from local reports when possible), which can be rescaled.

# 11 Toolbox Troubleshooting

The Toolbox prints messages during its processing to help guide error checking and debugging. Here are a few troubleshooting tips. This section will be updated frequently; please share any problems that you encounter.

## 11.1 Loading RWorkspace on Restart

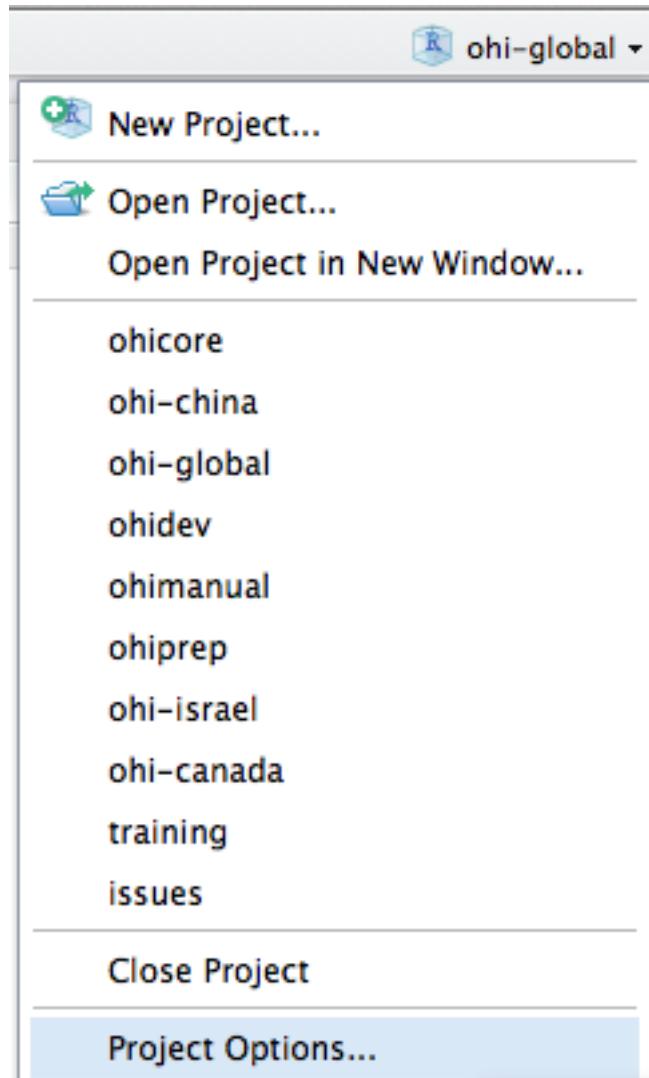
When you restart your R Session (Session > Restart R on a Mac), if you see that it is trying to load `ohicore`, it may give you an error:

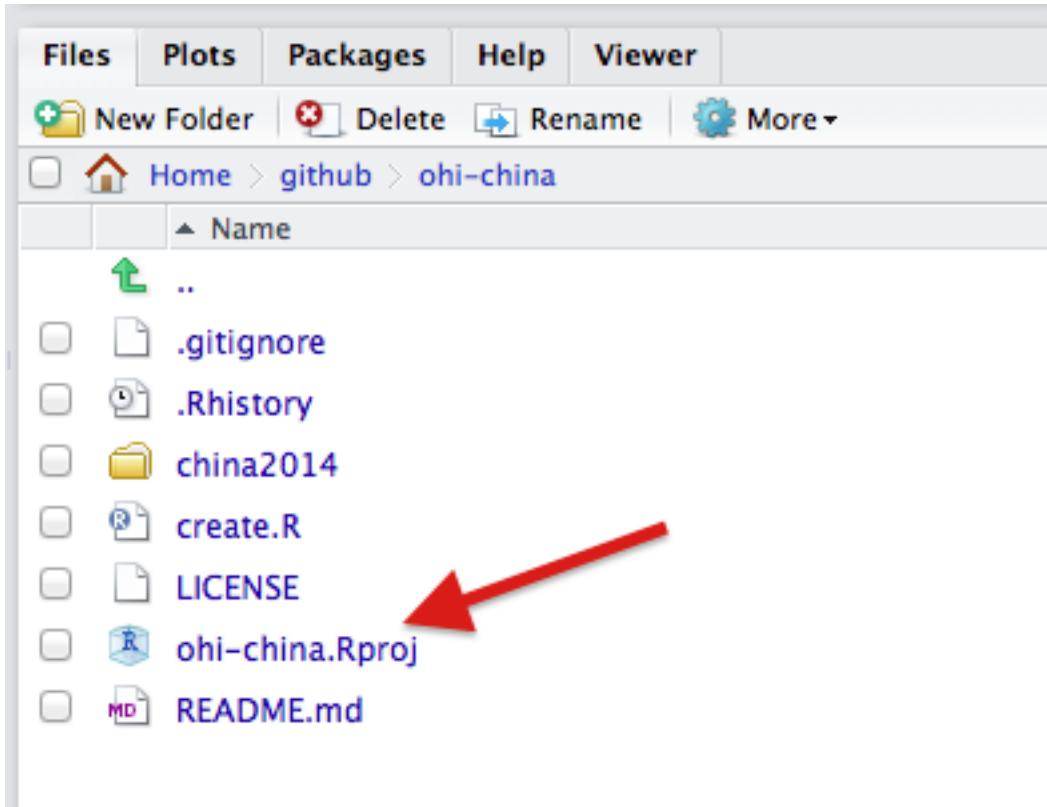
You do not want it to load `ohicore` or to save anything in your workspace. You will need to change the default setting from your `.Rproj` file. Steps to do this:

1. Go to Project Options, either in the pull-down menu or by double-clicking the `.Rproj` file:
  
  
  
2. Change all options to **No**:

```
Console ~/github/ohi-china/china2014/ ↵
Restarting R session...

Loading required package: ohicore
Error in .requirePackage(package) :
  unable to find required package 'ohicore'
In addition: Warning message:
In library(package, lib.loc = lib.loc, character.only = TRUE, logical.return = TRUE,  :
  there is no package called 'ohicore'
Loading required package: ohicore
Error in .requirePackage(package) :
  unable to find required package 'ohicore'
In addition: Warning message:
In library(package, lib.loc = lib.loc, character.only = TRUE, logical.return = TRUE,  :
  there is no package called 'ohicore'
>
```





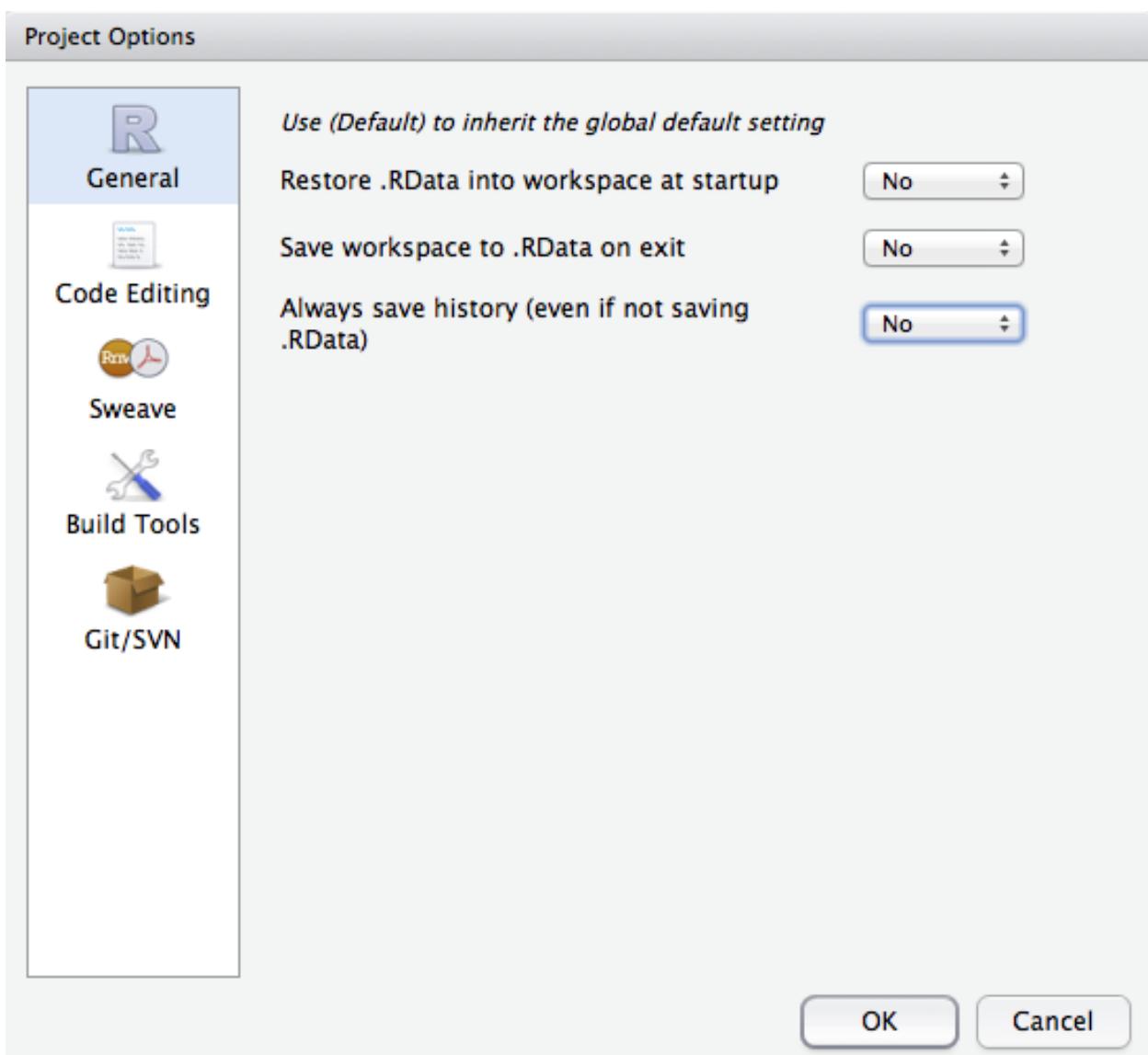
## 11.2 Calculating Pressures...

### 11.2.1 ‘The following components for [goal] are not in the aggregation layer [layer]...’

Example:

```
Running Setup()...
Calculating Pressures...
The following components for NP are not in the aggregation layer np_harvest_product_weight categories (:
Error in data.frame(names(P), P) :
  arguments imply differing number of rows: 0, 1
```

This error means you should update your pressures matrix because it expects there to be components that your region does not have.



	A	B	C	D	E	F	G
1	goal	component	component_n	po_chemicals	po_chemicals	po_pathogens	po_nutrients
2	FIS			1			1
3	MAR			2			
4	AO				1		
5	NP	corals		1			2
6	NP	fish_oil		2			1
7	NP	ornamentals		2			1
8	NP	seaweeds		2			2
9	NP	shells					1
10	NP	sponges					1

### 11.2.2 ‘Error in matrix...’

	A	B	C	D	E	F	G
1	goal	component	component_ni	po_chemicals	po_chemicals	po_pathogens	po_nutrients
2	FIS			1			1
3	MAR			2			
4	AO				1		
5	NP	corals			1		2
6	NP	fish_oil			2		1
7	NP	ornamentals			2		1
8	NP	seaweeds			2		2
9	NP	shells					1
10	NP	sponges					1

Example: >

This error means there is an empty column in pressures\_matrix.csv, and the Toolbox cannot handle empty columns.

## 11.3 Calculating Resilience ...

### 11.3.1 ‘Error in match(x, table, nomatch = 0L) : object id\_num not found’

```
tr_sustainability
tr_unemployment
Running Setup()
Calculating Pressures...
Calculating Resilience...
Error in match(x, table, nomatch = 0L) : object 'id_num' not found
In addition: There were 18 warnings (use warnings() to see them)
> |
```

This error means you should check that there is at least one entry for each goal (for each row) in resilience\_matrix.csv.