

The Ocean Health Index Conduct Phase

Contents

1	Introduction	5
1.1	Where you are in the OHI+ process	5
1.2	Best practices for OHI assessments	7
1.2.1	Incorporate core values and characteristics into the OHI assessment framework before gathering information	7
1.2.2	Maintain core values and characteristics within the assessment framework regardless of limited information quality	8
1.2.3	Strategically define spatial boundaries balance information availability and decision-making scales	8
1.2.4	Carefully document and share all decisions in writing and computational code	8
1.3	The Toolbox software and WebApp	9
1.4	Outcomes of conducting an assessment	9
2	Defining spatial boundaries	10
2.1	Drawing spatial boundaries	10
2.2	Buffers	11
2.3	Updating the map in your WebApp	11
3	Discovering and gathering input information	12
3.1	Thinking creatively	12
3.2	Data sources	12
3.3	Gathering responsibilities	13
3.4	Requirements for data and indicators	13
3.4.1	Relevance to ocean health	13
3.4.2	Accessibility	13
3.4.3	Quality	14
3.4.4	Reference point	14
3.4.5	Appropriate spatial scale	14
3.4.6	Appropriate temporal scale	14
3.4.7	The process of information discovery	14
3.4.8	Example: U.S. West Coast data discovery	15

4 The Ocean Health Index Toolbox	16
4.1 File system organization	17
4.1.1 Assessment repositories	17
4.1.2 Scenario folders	18
4.1.3 <code>subcountry2014</code> contents	20
4.2 Formatting Data for the Toolbox	25
4.2.1 Introduction	25
4.2.2 Uploading and formatting raw data files	26
4.2.3 Gapfilling	26
4.2.4 Long formatting	29
4.2.5 Rescaling your data	31
5 Installing the Toolbox	31
5.1 Overview	31
5.2 GitHub	32
5.2.1 Learning GitHub	32
5.3 Accessing GitHub Repositories	32
5.3.1 Create a GitHub account	33
5.3.2 Install <i>git</i> software	33
5.3.3 Set up your Git Identity	36
5.3.4 Create a folder called <i>github</i> on your computer	37
5.3.5 Update permissions	37
5.3.6 Work locally with R and RStudio	37
5.4 GitHub repository architecture	41
6 Using the Toolbox	42
6.1 Layer preparation workflow	42
6.2 Modifying and creating data layers	42
6.2.1 Create data layers with proper formatting	43
6.2.2 Save data layers in the <i>layers</i> folder	43
6.2.3 Register data layers in <code>layers.csv</code>	45
6.2.4 Check pressures and resilience matrices	46
6.3 Modifying pressures matrices	46
6.3.1 Create the new pressure layers and save in the <code>layers</code> folder	46
6.3.2 Register the new pressure layers in <code>layers.csv</code>	47
6.3.3 Register the new layers in <code>pressure_matrix.csv</code>	47
6.3.4 Modify the resilience matrix (if necessary)	47

6.4	Modifying resilience matrices	48
6.4.1	Updating resilience matrix with local habitat information	48
6.5	Modify goal models	51
6.5.1	Setup	51
6.5.2	Model modification	51
6.6	Removing goals	55
6.7	Modifying the pressures matrix for goals with categories	56
6.7.1	Background	56
6.7.2	Example 1: Pressures	57
6.7.3	Example 2: Resilience	58
6.8	Calculate overall OHI Index Scores	58
6.8.1	Additional data layers for pressures and resilience calculations	59
6.8.2	Final calculation	60
6.9	Other example modifications	60
6.9.1	Preparing the fisheries sub-goal	60
7	Toolbox Troubleshooting	64
7.1	General Software Errors	64
7.1.1	rpostback-askpass error	64
7.1.2	Git not detected on system path	67
7.1.3	Loading RWorkspace on Restart	70
7.2	Errors when Using the Toolbox	72
7.2.1	Useful Errors when Calculating Scores	72
7.2.2	Calculating Pressures...	74
7.2.3	Calculating Resilience	75
8	Overview of the OHI WebApp	76
8.1	Defining and displaying regions	78
8.2	Exploring inputs and outputs with the WebApp's App page	78
8.2.1	The App's Data tab	78
8.2.2	The App's Compare tab	81
8.3	Updating the WebApp's pages	82
8.3.1	Regions	83
8.3.2	Layers	83
8.3.3	Goals	83
8.3.4	Scores	84

9 Appendix 1: Goal Models, Reference Points, and Pressure and Resilience	84
9.1 Developing multiple goal models at the same time	84
9.2 Keeping Reference Points in Mind	84
9.3 Identifying pressures and resilience	85
9.3.1 Artisanal Fishing Opportunities	86
9.3.2 Biodiversity	87
9.3.3 Coastal protection	90
9.3.4 Carbon storage	91
9.3.5 Clean Waters	92
9.3.6 Food Provision	94
9.3.7 Livelihoods and Economies	98
9.3.8 Tourism and Recreation	103
9.3.9 Sense of Place	104
9.4 Pressures and resilience	106
9.4.1 Pressure	107
9.4.2 Resilience	109
9.4.3 Incorporating local resilience measures in your assessment	113
10 Appendix 2: Instructions for preparing data files for the OHI Toolbox	114
10.1 1. Description document for each goal	115
10.2 2a. <i>.csv</i> files for each data layer	115
10.2.1 Contents of data layers	115
10.2.2 Naming data layers	116
10.3 2b. Description files to accompany each data layer	117
10.4 3. Pressures and resilience	117
10.4.1 Pressures	117
10.4.2 Resilience	117
11 Appendix 3: Instructions for preparing model descriptions	118
12 Appendix 4: Develop a record-keeping spreadsheet	119
13 Appendix 5: R Tutorials for OHI	119
13.0.3 R Very Basics:	120
13.0.4 <i>tidyverse</i> functions	120
13.0.5 <i>dplyr</i> functions	122
13.0.6 Coding style	127

14 Appendix 6: Frequently Asked Questions (FAQs)	128
14.1 Conceptual	129
14.2 Timing and Resources	131
14.3 Structure	131
14.4 Appropriate data layers	132

Welcome to the Conduct Phase of your OHI+ assessment! This manual contains all the essential information on how to complete your own independent OHI assessment. It should be used by *goal keepers* (ie. experts tasked with gathering data and developing goal models), and the *toolbox master* (ie. technical expert(s) responsible for calculating the scores using the Toolbox).

The first four sections will be helpful for all participants. It contains information on OHI philosophy, what to expect when conducting an OHI+ assessment, best practices, and an introduction to the toolbox. Details on **goal model development and Pressure and Resilience are housed in Appendix 1**, which includes the meaning of the goal, reference point, ideal approach, practical guidance, and examples. **Appendix 2 and 3 provides guidelines on how to report data layers and model descriptions to the toolbox master(s)**. In addition, **Conceptual and technical Frequently Asked Questions are included in the Appendix 6**.

The remaining sections focus on technical guidance on how to use the toolbox and troubleshoot and will be most useful to the toolbox master. **Appendix 4 provides an example of a record-keeping spreadsheet** which will be helpful for tracking progress. Appendix 5 includes tutorials on frequently used R commands in the Toolbox.

1 Introduction

1.1 Where you are in the OHI+ process

The OHI+ process consists of four phases. In the first phase, you **learned** about the OHI to understand the philosophy behind the goals and the motivation for conducting a study. In the second phase, you actively **planned** to conduct your OHI+ assessment. Now you will actively **conduct** the assessment by engaging with the work of finding the data, preparing the goal models, and taking the necessary steps to learn how to use the **OHI Toolbox** and related software to produce the final results. This is where the science of data discovery and goal model development comes in. In the final phase, you will **communicate** the findings and results of your assessment with others.

The OHI framework allows you to synthesize the information and priorities relevant to your local context and produce comparable scores. Because the methods of the framework are repeatable, transparent, quantitative, and goal-driven, the process of carrying out an OHI+ assessment is as valuable as the final results.

The first completed assessment for a study area is valuable because it establishes a baseline and highlights the state of information quality and availability in an area. Any subsequent assessments carried out through time are also valuable because they can be used to track and monitor changes in ocean health. Your assessment will require careful thought and consideration along the way, and we encourage documentation and scripting to be done within the OHI Toolbox to facilitate collaboration and transparency, as well as the reproducibility for future assessments.

Each OHI+ assessment should have a clear purpose. One of the typical reasons for conducting an independent assessment is to inform policy and management decisions. Assessments can be more relevant to management when they are conducted at the spatial scales at which policy decisions are made, such as states, provinces, or counties. The **regions** and the overall **study area** are definitions that will be used throughout the assessment. The study area is the entire spatial boundary of your assessment, while the regions are the

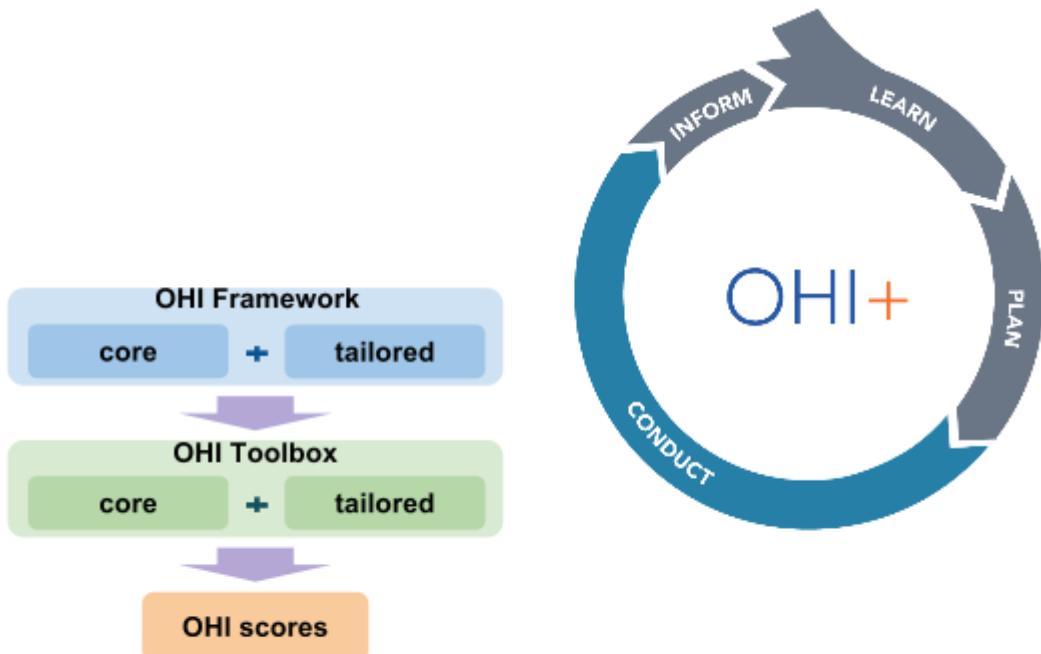


Figure 1: four phases of OHI Assessments

smaller subdivisions within the study area. In the OHI framework, goal scores are calculated for regions separately and then combined to produce an overall OHI score for each study area. The number of regions varies with each assessment's study area; completed assessments have had between one and 221 regions.

The process of conducting an OHI+ assessment is as valuable as the final results. Documenting decisions made, as well as the challenges and successes encountered along the way, can lead to better understanding of the system, help inform management decisions, and guide future assessments to track changes through time.

When conducting an OHI+ assessment, it is important to include information that best represents your study area, and to make science-driven decisions and clearly document what was done and why. Your team should be as creative and insightful as you can be while working within the bounds of informational and technical limitations.

There are **key processes and considerations** that will be a part of every assessment. Every assessment should ideally build from the lessons learned of previously completed assessments and identify what local characteristics need to be included in a study. This is done partly by comparing the local situation to situations in previous assessments; it is also done by comparing the default information provided in the WebApp to what is known about local realities. After you have outlined and identified local characteristics and priorities, you should prepare to develop models, set reference points, gather information, and use the Toolbox software to analyze data and calculate scores. You will also visualize the outputs in the WebApp's maps and flower plots that can be shared with other partners and collaborators. Above all, you should be prepared to know that **this process takes time and is iterative, meaning that you often return to previous steps**.

How long does an assessment take? Past assessments have taken between two and three years, with the time varying depending the size and composition of the team, the challenges encountered in discovering

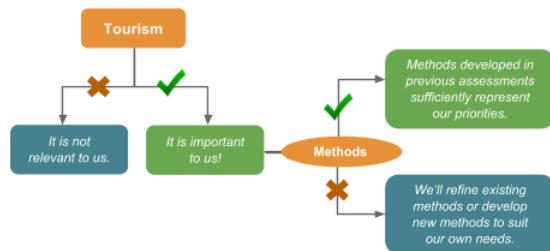
and gathering information, and how many models are redeveloped. The amount of data processing and goal model development needed before you will be able to use the Toolbox also affects the amount of time it takes to conduct the assessment. The skill sets of the team members and the amount of technical resources available are also hugely important factors. You should think about which team members are needed at what stage of the process, including an R programmer and a spatial analyst. It will take time for the technical team to become familiar with the OHI Toolbox and GitHub.

1.2 Best practices for OHI assessments

Conducting an assessment requires both an understanding of how past assessments have been completed and the innovation to capture important characteristics of your study area using the information available. You can start by understanding the structure of completed assessments at global and smaller scales and the models that were created. Understanding the approaches in different contexts will help you think about what should be done similarly and differently in your local context. Information, publications, and websites for completed OHI+ assessments are listed on the projects page of ohi-science.org, and example approaches for each goal are listed in Appendix 1 of the manual.

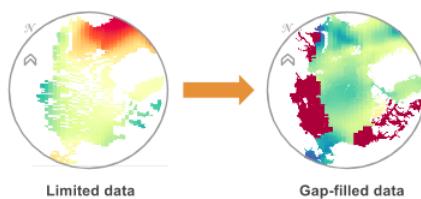
Build the conceptual framework

Which benefits does the ocean provide in your local area? What are key ecological, social, and economic characteristics and priorities? Identify these and build a conceptual framework before gathering existing information.



Remain true to the framework

Existing, available data can be limited or not ideal. But it is possible to fill gaps, use proxy data, or incorporate intermediate models. But stick to your conceptual framework and priorities to get a complete picture of ocean health.



Define spatial boundaries

The boundaries of your assessment should be driven by the boundaries where information is reported and policy decisions are made.



Document and share the process

Methods developed and lessons learned through the assessment process should be shared with future assessments. Methods should be reproducible in one location through time and repeatable in different locations. This requires transparency and communication throughout the assessment.



Figure 2: Best practices of OHI+ assessments

1.2.1 Incorporate core values and characteristics into the OHI assessment framework before gathering information

Begin your assessment by identifying local socio-cultural-economic characteristics and priorities related to ocean health, and how they would ideally be captured with the existing or modified OHI framework. This means understanding the rationale behind the components of the OHI framework and identifying what must be added or removed or redefined to ensure that it best represents the local context. Are all goals relevant to your study area? What should be added, removed, or redefined? In

this process it is important to identify not only characteristics that could be included in goal models, but also the important stressors (pressures) and resilience elements within the study area. What are the key issues that should be included for your assessment to be credible, useful, and meaningful? How do people typically relate to the ocean in your area in terms of social and cultural patterns? These are the kinds of questions you should consider prior to assembling the available information.

The OHI framework should guide your assessment, but you should not be constrained by it. If a goal is not relevant, it should be removed. If there are elements important to your study area that are not present within the existing framework, how could they be included? Having a clear picture of how the framework should be restructured and what the assessment should include is very important before moving on to assemble information, because otherwise the assessment could be biased by what information is available instead of what is important to include. When specific information is limited there are ways to capture them with indirect measures.

1.2.2 Maintain core values and characteristics within the assessment framework regardless of limited information quality

The assessment framework can be implemented using the best freely-available existing information, even if the information available is ‘limited’ or not ‘ideal’. ‘Limited’ information may be of low quality, have gaps, or be indirectly obtained through modeling instead of being directly measured. Different methods can be used to work with limited data, such as gap filling, incorporating indirect (proxy) or place-holder information, or using intermediate models.

Remaining true to the conceptual framework by using those methods, hence developing less-than-ideal goal models, provides a fuller picture than redesigning it to only include characteristics where ideal information is available. This is because all key characteristics in the system should be represented somehow in a comprehensive assessment, even if assumptions must be made to compensate for missing information. If these methods, including assumptions and rationales, are clearly considered and explained, completed assessments will not only provide the best possible picture of the current system but will also identify information gaps and highlight areas for improvement. Such scrutiny of available knowledge could be lost if important elements were simply excluded from the assessment due to imperfect representation.

1.2.3 Strategically define spatial boundaries balance information availability and decision-making scales

Identifying the spatial boundaries of the Regions within the Assessment Area is extremely important because OHI scores are calculated for each unique Region, and the boundaries will be used to aggregate or disaggregate input information reported at different spatial scales. Spatial boundaries should be defined with geographic information system (GIS) mapping software, ideally per management jurisdiction (see *Defining spatial boundaries* section for technical guidance). **Jurisdictional boundaries are optimal because it is often at these scales where management and policy decisions are made, cultural priorities and management targets are identified, and information is collected in standardized and therefore comparable ways.**

Within the OHI framework, there is no limit to the number of Regions that can exist within the Assessment Area; the number is only constrained by data availability and the utility of having scores calculated for a particular Region. Although it is possible to assess only one region in the study area (i.e. the region is the assessment area), this might not be ideal because it eliminates the possibility of making comparisons or identifying geographic priorities within the study area.

1.2.4 Carefully document and share all decisions in writing and computational code

Documentation of all aspects of an assessment process is paramount to ensuring transparency of the decisions made, reproducibility and interpretation of results, repeatability to facilitate and compare future assessments,

and the ease of communication throughout the process.

Detailed information about how the assessment is conducted will enhance its credibility and reproducibility. Decisions of why information was included and why models were developed in a certain manner are of great importance so that future assessments can incorporate the same logic and understanding of the system — or make improvements. Further, the type of workflow developed and software used to organize and process information will greatly affect the efficiency, transparency, and reproducibility of subsequent assessments. Providing public access to all such information, as well as input data and computational code is becoming the standard for scientific inquiry, so every effort should be made to achieve those aims.

This means documenting and sharing not only the tools and methods used but also the knowledge gained through the process based on decisions made, what was decided against (e.g. why information was included or excluded, and how it was processed), challenges encountered, and recommendations for improvement. Frank documentation about the successes and shortcomings makes for greater scientific credibility, enables others to replicate what was done, and allows for the most appropriate interpretations of the results, as well as the highest potential for future improvement of assessment approaches, and ultimately, management towards ocean health.

1.3 The Toolbox software and WebApp

The **OHI Toolbox** is the main software used for organizing and processing information, documenting decisions made, calculating scores, and visualizing results. It was created to facilitate score calculations as well as the organization of information and transparency of the approach. The Toolbox is built with open-source software, meaning that it is freely available for you to use and can be modified to meet your needs. You will access the software from online and will use several free software tools to conduct your assessment in a collaborative, transparent, and reproducible manner.

The Toolbox software should be used when your team has gathered the information necessary for the goal models you have developed. The steps of carefully preparing and gathering data layers, indicators, and developing goal models can be done independently of the actual software use; however, if done within the Toolbox structure these steps will be traceable and collaborative. But of course the assessment can be done without all members of your team becoming familiar with the technical aspects of the Toolbox.

Accompanying the OHI Toolbox is the **OHI+ WebApp**, which is a visualization tool that displays input information and calculated scores in several ways, including interactive maps and flower plots. Most coastal countries have a WebApp that was created to facilitate planning and communication during assessments. The WebApp visually presents inputs, goal models and calculated scores for each defined region through interactive maps, histograms, and tables. All inputs presented in a country's WebApps were extracted for each country from global assessments, and scores were calculated using goal models from global assessments.

The WebApp is a widely useful communication tool. Not all team collaborators may be involved with the technical aspects of the Toolbox, but the WebApp enables everyone to explore inputs and calculated scores.

1.4 Outcomes of conducting an assessment

Your completed assessment will produce OHI scores for each goal for every region in your study area, and scores within the assessment can be compared with each other. These scores will not be quantitatively comparable to those of other OHI assessments because they differ in the underlying inputs, goal models, and reference points. The only *quantitative* comparisons can be made within an assessment's study area, whether between regions or through time. However, *qualitative* comparisons between different OHI assessments can be made because the scores are an indication of how far a region is to achieving its own targets. For instance, if two study areas have scores of seventy and sixty-five, it should be interpreted that the first study area is closer to its management targets than the second is, but since these management targets are different (in addition to the underlying data and models), they cannot be quantitatively compared.

While the OHI scores are valuable information, **the process of conducting an OHI assessment can be as valuable as the final results.** This is because during an OHI assessment you will bring together meaningful ocean health information from many disciplines. In doing so, you will have a census of existing information and will also identify knowledge and data gaps. Further, conducting an OHI+ assessment can engage many different groups, including research institutions, government agencies, policy groups, non-governmental organizations, and both the civil and private sectors.

2 Defining spatial boundaries

Boundary definitions should match the purpose of the assessment and be informed by the scale at which information is available. The boundaries displayed in your WebApp are provided by default using subcountry region definitions from Global Administrative Areas (GADM: www.gadm.org). It is possible to redefine the spatial boundaries for your study area and regions.

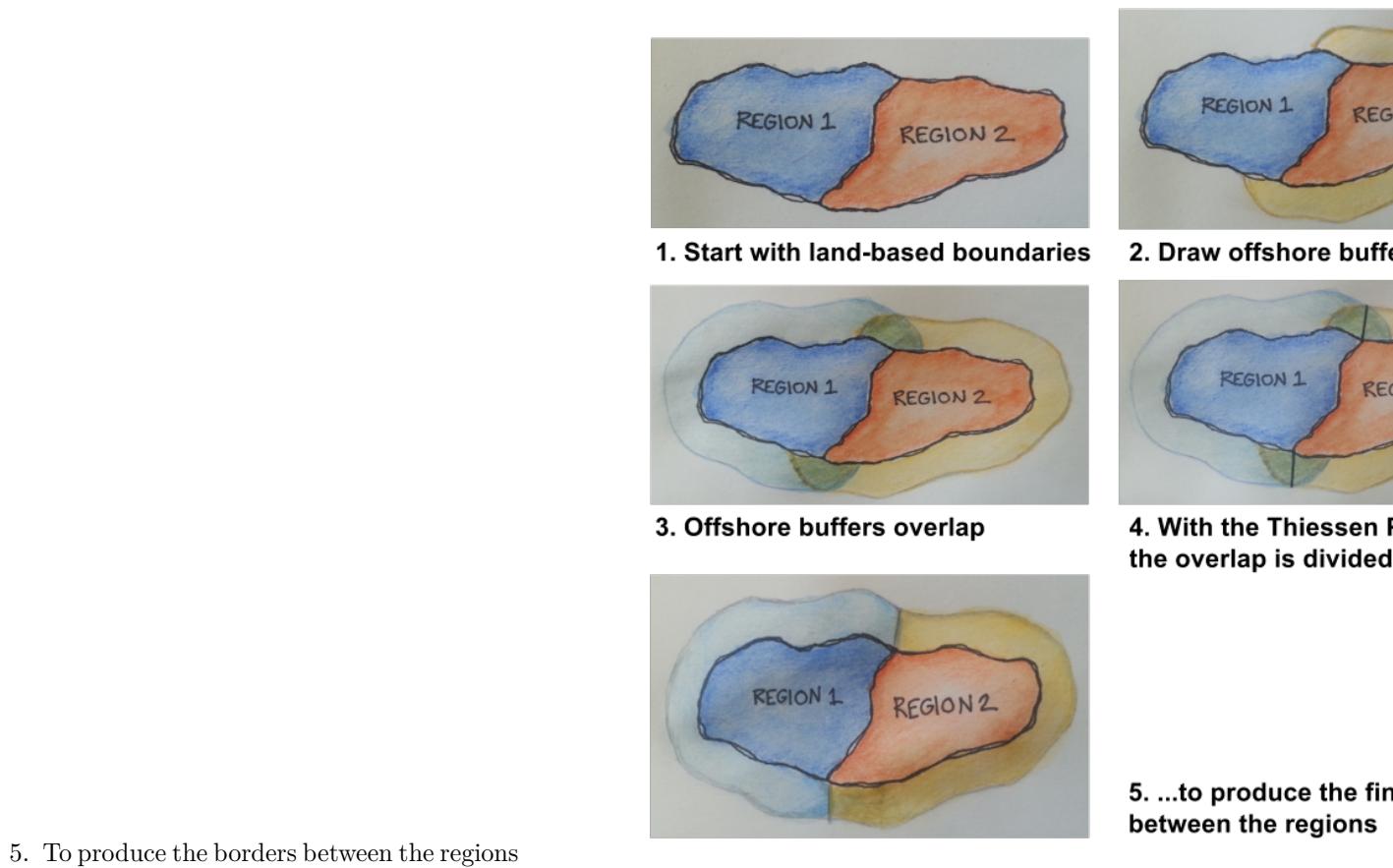
Note that the OHI does not take a stance on disputed territories. The boundaries are defined by the original map data providers.

2.1 Drawing spatial boundaries

Spatial boundaries must be drawn with geographic information system (GIS) mapping software such as ArcGIS, QGIS, or GRASS. You will need someone with GIS skills to create a **shapefile** that will be used by the Toolbox to display your information. The shapefile will also be used to extract information for each of your defined regions when data are reported in raster format for a different area. For more information see https://en.wikipedia.org/wiki/Geographic_information_system and <http://en.wikipedia.org/wiki/Shapefile>.

Regions must be unique (non-overlapping), and boundaries must be drawn offshore, extending to the exclusive economic zone (EEZ) edge in most cases. Offshore boundaries should be made with spatial methods in order to extend boundaries from those designated on land. One possible method is to create boundaries with Thiessen Polygons, and we provide a Python script that can be used, but it requires ArcGIS. The Python script and further details can be found at http://ohi-science.org/pages/create_regions.html. Using Thiessen Polygons, offshore boundaries are created with the following steps.

1. Start with land-based boundaries
2. Draw offshore buffers for each region
3. But the buffers overlap
4. For the Thiessen Polygon approach, the overlap is divided



2.2 Buffers

When drawing your regions, it is also a good idea to create *inland* and *offshore* buffers that will be used to extract data in your assessment. Buffers are not necessary for display in the WebApp but they will be important for later layer preparation. For example, the global assessment used coastal population information, and raster data were available for entire countries. This meant that ‘coastal’ had to be defined: for global assessments it was defined as 25 miles from the coast. To extract just the coastal population from the population raster file, we created a 25 mile inland buffer for each reporting region. But to extract mangrove data for each region from raster files, global assessments used 1km inland and 1km offshore as the buffer.

At this point, you may not know which buffers you will need, as they depend on the data available, your goal models and definitions. Some buffers used in the global assessments were 1km inland, 25miles inland, 1km offshore, 3nm offshore.

2.3 Updating the map in your WebApp

Once you have created your boundaries with GIS software, you will need to send them to us. Please send us a .zip file of all files produced. Files with the following extensions are required (but you can send all files):

- .dbf
- .shp
- .shx

- .prj

The .dbf file needs the following in its attribute table:

- **rgn_id** (unique numeric region identifier)
- **rgn_name** (unique named region identifier)
- **area_km2** or **area_hectare** (area in km² or hectares)

Note that we can send you the shapefiles used to create your WebApp if you would like to modify them.

3 Discovering and gathering input information

A hallmark of the OHI is that it uses freely-available existing information (data and indicators) to create the models that capture the philosophies of individual goals. The quality of the inputs are important because calculated Index scores area only as good as the inputs on which they are based. Assembling the appropriate input information, which means both discovering and gathering data and indicators, is an important part of any OHI assessment.

Once your team has tailored the OHI framework appropriately for your study area and identified the information that ideally would be included, the data discovery and gathering process can begin. There are many decisions to make when deciding which data are available and appropriate to include in your assessment. Finding appropriate data requires problem-solving abilities and creativity, particularly when ideal data are unavailable. You will need input information to calculate status models as well as pressures and resilience.

3.1 Thinking creatively

Humans interact with and depend upon the oceans in complex ways, some of which are easy to measure and others of which are harder to define. More familiar measurements include providing seafood, or disposing of waste. A less familiar measurement is how marine-related jobs affect coastal communities, or how different people receive or perceive benefits simply from living near the ocean. Thinking creatively and exploring the information available can make your assessment more representative of reality.

Data used in OHI assessments spans a wide array of disciplines beyond oceanography and marine ecology. It is important to think creatively and beyond the interests of a specific institution or one particular field of study. Therefore, it is necessary to look beyond the most known or obvious data sources to find data relevant for the goals in the study area. Discussions with colleagues, literature searches, emails to experts, and search engines are good ways to understand what kinds of data are collected and to hunt for appropriate data. Investigate what kinds of information are available from government and public records, scientific literature, academic studies, surveys and reports etc.

3.2 Data sources

Existing data and indicators can be gathered from many sources across environmental, social, and economic disciplines. This includes government reports and project websites, peer-reviewed literature, masters and PhD theses, university websites, and information from non-profit organizations, among others.

All data must be rescaled to specific reference points (targets) before being combined with the Toolbox; therefore setting these reference points at the appropriate scale is a fundamental component of any OHI assessment. This requires your assessment team to capture the philosophy of each Index goal and sub-goal using the best available data and indicators. Indicators that are already scaled (e.g., from 0-1 or 0-10) can easily be incorporated into your assessment; they have already been scaled to some kind of identified target

or reference point. Data that have not been scaled in most cases will need to be, whether this is by scaling to the maximum value in the range or to some other understood value. You should think about how you would rescale data during your search.

Because data and indicators will come from different sources, they will also have different formatting. To include these data and indicators in your assessment, you will need to process these files into the format required by the Toolbox, which is explained in the section **Formatting Data for the Toolbox**. When data have been prepared and formatted for the Toolbox, they are called **layers**. Because creating layers can be quite time-intensive, data should only be prepared for the Toolbox after final decisions have been made to include the data or indicator in your assessment, and after the appropriate goal model and reference points have been finalized.

3.3 Gathering responsibilities

Gathering appropriate data requires identifying and accessing existing data. It is important that team members responsible for data discovery make thoughtful decisions about whether data are appropriate for the assessment. Data discovery and acquisition are typically an iterative process, as there are both practical and philosophical reasons for including or excluding data.

3.4 Requirements for data and indicators

There are six requirements to remember when investigating (or ‘scoping’) potential data and indicators that are presented in this section. It is important that data satisfy as many of these requirements as possible. To meet these requirements, you may have use appropriate methods to fill gaps in the data set. Data sources may need to be excluded from the analyses if requirements are not met and gap-filling solutions are not possible. If data cannot be included, you may elect to use layers from the global assessment or identify other data or modeling approaches.

3.4.1 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.

3.4.2 Accessibility

The two main points regarding accessibility are whether the source is open access and whether the data or indicators will be updated regularly.

The Index was created in the spirit of transparency and open-access, using open-source software and online platforms such as GitHub, is to ensure as much accessibility and open collaboration as possible. Data and indicators included should also follow these guidelines, so that anyone wishing to understand more about the Index may be able to see what data were used and how. For this reason we emphasize the importance of using data that may be made freely downloadable, as well as the importance of clearly documenting all decisions and reasons for the choices made in selecting data, indicators, and models.

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 you will be able to find it in the future.

3.4.3 Quality

Understanding how the data or indicators were collected or created is important. Are they collected by a respected organization with quality control? Are there any protocol changes to be aware of? For instance, were there changes in the collection protocol to be aware of when interpreting temporal trends?

3.4.4 Reference point

Most data will need to 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:

- Has past research identified potential targets for these data? *For example, fisheries goal require a Maximum Sustainable Yield (MSY).*
- Have policy targets been set regarding these data? *For example, maximum levels of pollutants allowed in beaches.*
- Would a historic reference point be an appropriate target? *For example, the percent of habitat coverage before coastal development took place.*
- Could a region within the study area be set as a spatial reference point? *For example, a certain region is regarded as the leader in creating protected areas.*

3.4.5 Appropriate spatial scale

Data must be available for every region within the study area. It is not always possible to fully meet the spatial and temporal requirements with each source. In these cases, provided that the gaps are not extensive, it can still be possible to use these data if appropriate gap-filling techniques are used (See: **Formatting Data for Toolbox** section).

3.4.6 Appropriate temporal scale

Data must be available for ideally the five most-recent years to calculate the recent trend. For some goals, where temporal reference points are desirable, longer time series are preferable.

3.4.7 The process of information discovery

The most important thing to remember when gathering data and indicators is that they must contribute to measuring ocean health. Not all information that enhances our knowledge of marine processes 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.

Whether you are working goal-by-goal, or layer by layer, it is important to consider where you can find synergies in data discovery. For example, while you are looking for information for the fisheries goal, you may also find data layers for fishing pressures, such as metrics on bycatch or trawling intensity. This will save you time and allow you to start thinking about how to rank pressures and resilience weights on your goals as well. Conceptually, it will help your team build a picture of how your goals are interlocking in a way that is reflective of the actual linkages that exist in the connected systems you are studying. Some key examples are listed below, and are further explained in the following sections.

You should begin by understanding and comparing the best approaches used in assessments that have been completed, including the global assessments (Halpern *et al.* 2012; 2013), Brazil (Elfes *et al.* 2014), Fiji (Selig *et al.*, 2014), and the U.S. West Coast Assessment (Halpern *et al.*, 2014). For OHI+ assessments, if finer-resolution local data were available in the study area, these data were either incorporated into modified

goal models that used locally appropriate and informed approaches or into the existing global goal model. When local data were not available, the global-scale data and global goal models were used, 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 when going goal-by-goal for discovering data and developing models:

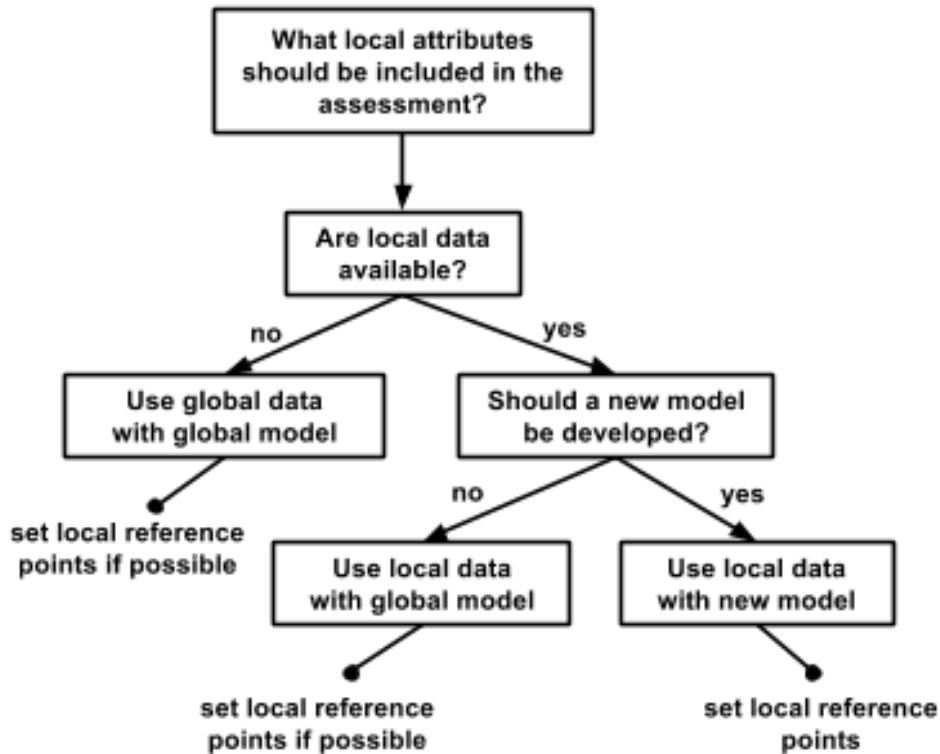


Figure 3:

3.4.8 Example: U.S. West Coast data discovery

Below are examples of some decisions made when exploring available data for the U.S. West Coast assessment. Determining whether certain data could be included began with a solid understanding of the layers and models included in the global assessment. Since the U.S. West Coast is a data-rich region, finer-resolution local data could be used in place of many of the global data layers. The U.S. West Coast assessment had five regions: Washington, Oregon, Northern California, Central California, and Southern California.

There are a lot of existing data that contribute to our scientific understanding of ocean processes and interactions but are not ideal for the OHI. Reasons to exclude data are both due to practical requirements (e.g., spatial or temporal resolution) and philosophical requirements (i.e., they do not help capture the attributes of interest for assessing ocean health). **Some common reasons for excluding data** are:

- **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, data were only available for three of the five regions (the three California regions but not Oregon and Washington), so they could not be used.

- **There is not a clear and scientifically observed relationship between the data and ocean health.** Along the U.S. West Coast, kelp beds are a very important habitat because of their contribution to biodiversity and coastal protection. However, kelp coverage variation is driven primarily by abiotic natural forcing (wave or storm disturbance and temperature) and thus is not a good indicator of kelp forest health, particularly in the case of anthropogenic impacts. 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 marine or coastal ecosystems.** Sea walls and riprap provide coastal protection to many people along the U.S. West Coast. However, these structures are not a benefit that is derived from the marine ecosystems, so only coastal habitats were included in the calculation of this goal. These data can be included as a pressure due to habitat loss. They were not used as a resilience measure because they can often have negative side effects (e.g., by altering sedimentation dynamics), and because they have limited long-term sustainability (i.e., they need maintenance).
- **Data collection is biased and might misrepresent ocean health.** The U.S. 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 only assess 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.
- **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 U.S. West Coast, the current extent of seagrass habitats was available, however, these do not exist for previous points in time in most areas, so could not be used to calculate the trend or set a historical reference point. Therefore, we estimated the trend in health of seagrass habitats using as a proxy the trend in the main stressor (i.e., turbidity). In other words, we assumed that the rate of seagrass loss was directly proportional to the rate of increase in turbidity. Similar solutions may be used to estimate trends in your own assessment, if there is scientific support for assuming that the trend of what we want to assess (or the relationship between the current state and the state in the reference year) has a strong relationship with the trend of the proxy data available.

4 The Ocean Health Index Toolbox

In this section, you will learn the basics of how to use the OHI Toolbox to conduct your assessment. You will be introduced to files the Toolbox requires, how you will modify them, and how they interact to calculate the final output scores.

The OHI Toolbox is an ecosystem of small files and scripts that are the tools needed to calculate OHI scores. These files and scripts are stored in two Github ‘*repositories*’: folders that are synchronized with collaborators. The first folder is your **assessment repository** that has a three-letter code, such as *esp* for Spain or *ecu* for Ecuador. You will edit this repository with your data, goal models, and updated pressures and resilience matrix tables. The second repository is called **ohicore** and it contains core functions for combining your data and goal models to calculate OHI scores. You will not edit **ohicore**, but you are able to explore it to understand the calculations.

Your **tailored assessment repository** contains data input layers, configuration files, and scripts. These files are organized in the same way within a *scenario folder* called **subcountry2014**, with data layers, goal model equations, and configuration files from the global 2014 assessment. Files within the scenario folder are comma-separated-value (*.csv*) files and scripts written in the programming language *R*.

The repository will be provided for you by our OHI science team *after* you have planned all your goal models, identified good data, and finalized your spatial boundaries. When you request a repository, send

OHI Toolbox

calculations and organization, using open source collaborative software

core
standardized, familiar
R package on github:
[ohi-science/ohicore](#)



tailored
flexible to the local context
repository on github: [ohi-science/xxx](#)

Figure 4: Toolbox = your repo + ohicore

your [prepared shapefiles](#) that have the specific features described. You will not need a repository before this. Before then, it might be helpful to look at existing OHI+ Github repositories on the [Projects page](#) of our website and get familiar with their structure.

Each OHI+ assessment repository has inputs and goal models based on the 2014 global assessment. This means that each assessment repository isolates the information used for each region in the global assessment and stores it in a separate OHI+ assessment repository. Therefore, it can be an easy way to explore the inputs used in the global assessment for one specific place. When conducting an assessment, you will replace and modify as much of this information with local information that better represents your study area, since information reported at a national scale cannot always be attributed to its subcountry regions, as has been done in most cases in each OHI+ assessment repository. See more details in the discussion of the [layers folder](#).

The Toolbox is open-source and can be downloaded and installed for free. You are able to navigate through these files both at www.github.com/OHI-Science and on your own computer once you have cloned the repository to your computer. GitHub is an online platform used by the OHI that facilitates collaboration and archives past versions of all files with the author identified. It can be accessed remotely by all members of your team and enables team members to synchronize their work together. Because all versions are saved, you can return to previous work and also compare different points in history to track how changes you make affect the output scores. Instruction on how to access your assessment repository is in the [Installing the Toolbox](#) section.

The following sections will describe the files included in the Toolbox. You will then learn what is required for data layers for your assessment and how to change goal models.

4.1 File system organization

This section is an orientation to the files within your assessment repository. The file system organization is the same for all assessment repositories, and can be viewed at github.com/OHI-Science or on your computer. While reading this section it is helpful to explore a repository at the same time to become familiar with its contents and structure. The following uses the assessment repository for Ecuador (*ecu*) as an example, available at www.github.com/OHI-Science/ecu.

Most of your time will be spent preparing input layers and developing goal models. You will also register prepared layers to be used in the goal models. This all will be an iterative process, but generally speaking you will work goal-by-goal, preparing the layers first, registering them, and then developing the goal models in *R*. to calculate the scores.

4.1.1 Assessment repositories

Assessment repositories are identified by a three-letter code; Ecuador's assessment repository is called '*ecu*'. Assessment repositories contain several things:

- The **scenario folder** is the most important folder within the repository; by default it is named **subcountry2014**. It contains all of the inputs needed to calculate OHI scores, and you will modify these inputs when conducting your assessment. The scenario folder is explained in detail in this section.
- All other files in the assessment repository are accessory files. Files with names beginning with a ‘**‘** are required for versioning capabilities by GitHub and do not appear when the assessment repository is viewed on your computer.

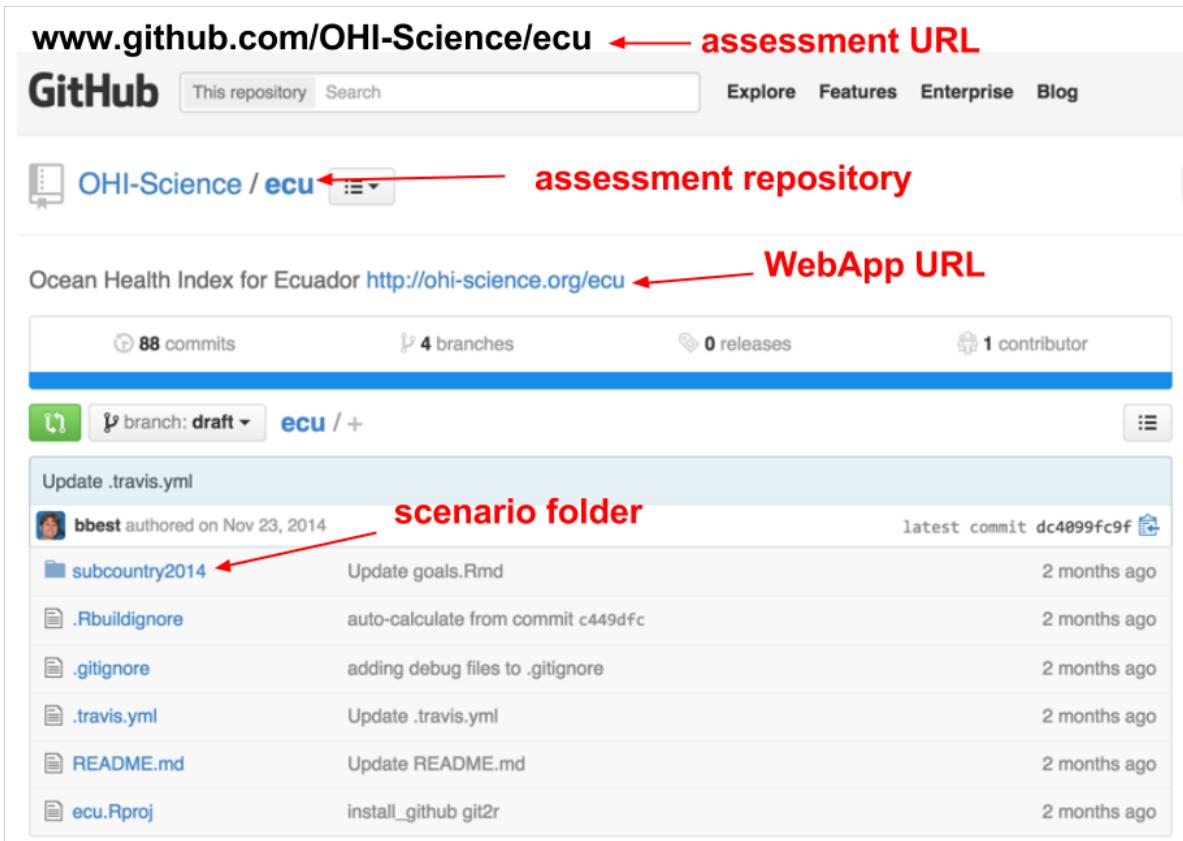


Figure 5:

4.1.2 Scenario folders

Scenario folders contain all files and scripts necessary to calculate OHI scores. There are two file types:

- ** **.csv** files** contain data inputs or configuration information.
- ** **.R** scripts** are written in the programming language R and use data inputs for processing and calculations.

There is one scenario folder in your assessment repository and it is called **subcountry2014** to indicate that the assessment is conducted at the subcountry scale (province, state, district, etc.), based on input layers and goal models used in the 2014 global assessment. When conducting your assessment, you can rename your scenario folder to reflect the subcountry regions in your study area and year the assessment was completed.

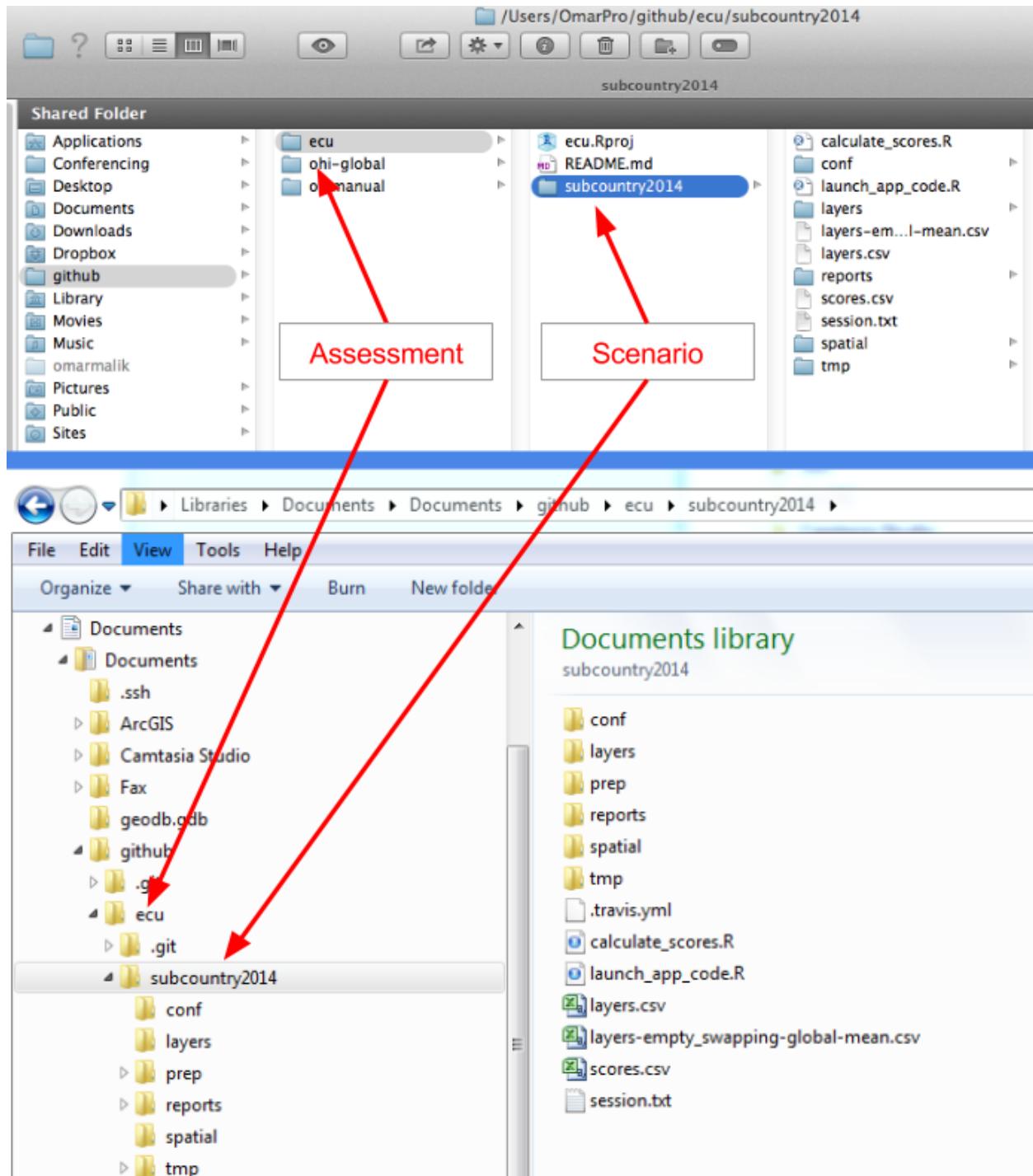


Figure 6: Navigating the assessment repository. The figure shows Mac folder navigation above and Windows navigation below.

For example, `province2015` would indicate the assessment was conducted for coastal provinces in the year 2015.

Once you complete your assessment with the `subcountry2014` (or equivalent) scenario, further assessments can be done simply by copying the `subcountry2014` folder and renaming it. This can be done for future assessments, for example `subcountry2016` or `subcountry2018`, which eventually would enable you to track changes in ocean health over time. You can also copy scenario folders to explore different policy and management scenarios, for example `subcountry2014_policy1`.

4.1.3 `subcountry2014` contents

This figure illustrates the files contained within the assessment repository's `subcountry2014` scenario folder. Important files are either `.csv` text files or `.R` script files. Files are organized into different folders within the `subcountry2014` folder, and you will modify some of these files while leaving others as they are.

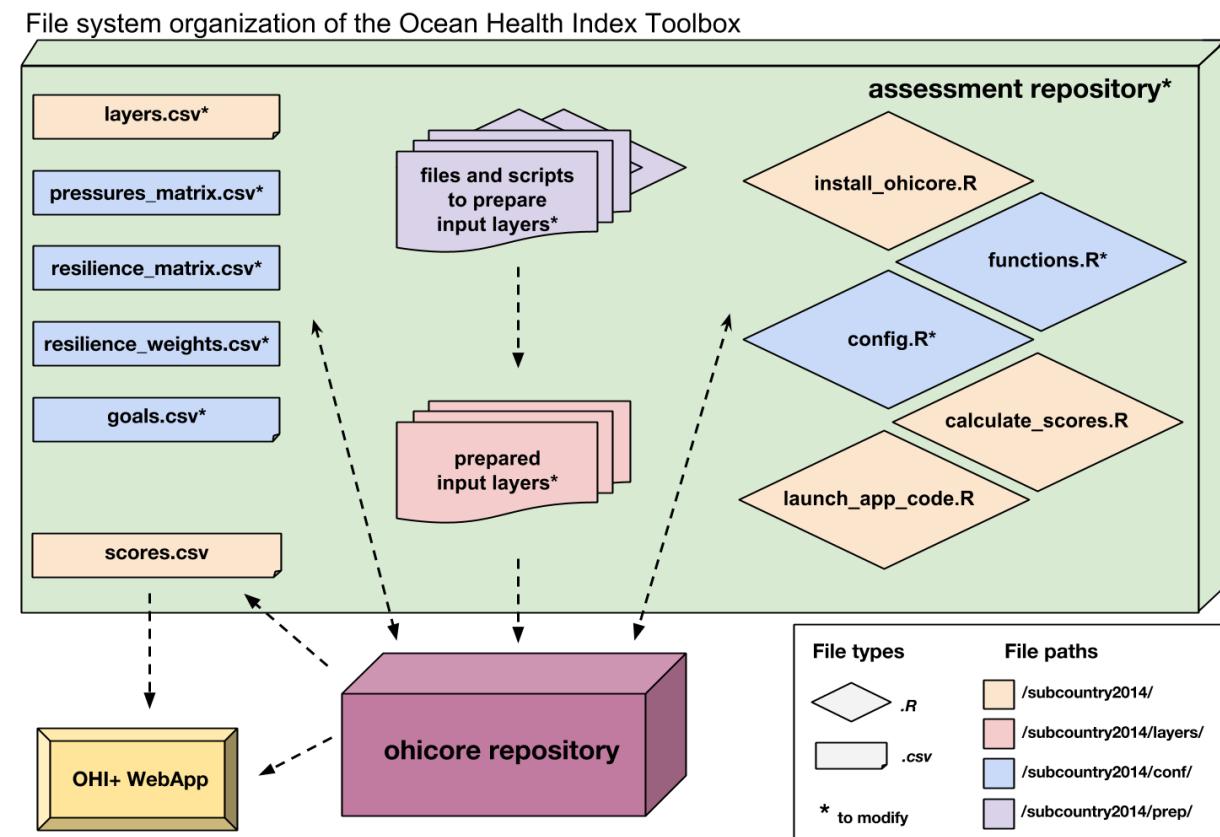


Figure 7: File system organization of the Ocean Health Index Toolbox

Files and folders are presented here in alphabetical order. See the **Using the Toolbox** section for the workflow of how you will use the files.

4.1.3.1 `calculate_scores.R` `calculate_scores.R` is a script that runs everything required to calculate OHI scores using the prepared layers the `layers` folder that are registered in `layers.csv`. Scores will be saved in `scores.csv`.

4.1.3.2 *conf* folder The *conf* folder includes important configuration files required to calculate OHI scores. There are both *.R* scripts (*config.R* and *functions.R*) and *.csv* files (*goals.csv*, *pressures_matrix.csv*, *resilience_matrix.csv*, and *resilience_weights.csv*).

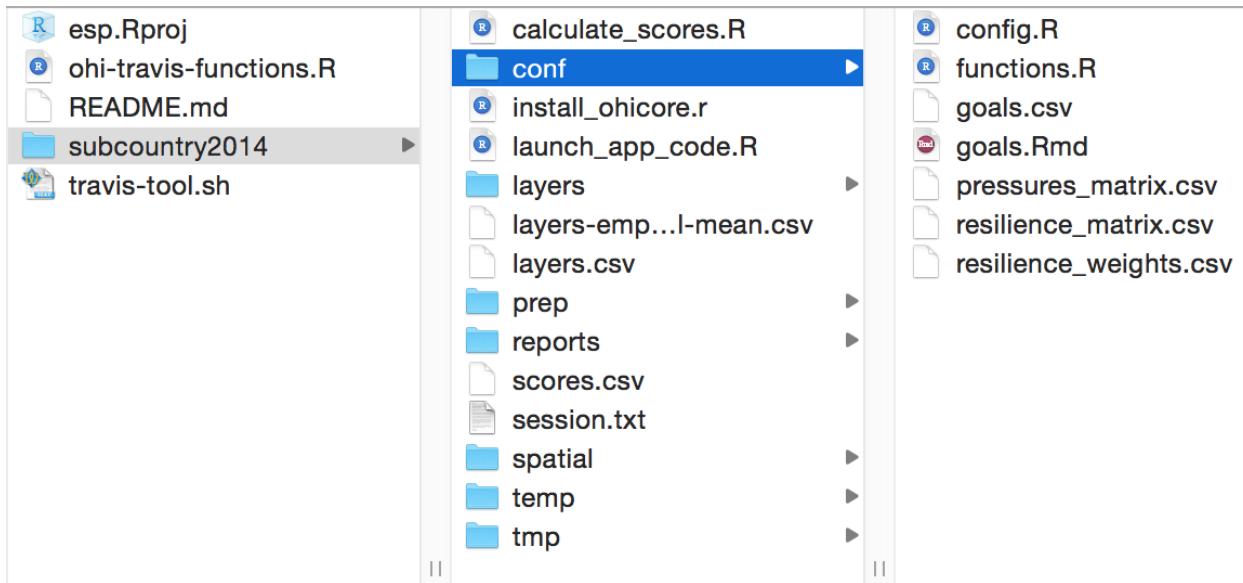


Figure 8: The *conf* folder contains important R functions and *.csv* files.

4.1.3.3 *config.R* *config.R* is an R script that configures labeling and constants appropriately. You will only need to modify this file when working with goals that have categories (example: habitat types or economy sectors) that are affected differently by pressures and resilience measures.

4.1.3.4 *install_ohicore.R* *install_ohicore.R* is a script that will install *ohicore*, which is the second repository required for the Toolbox and is the engine behind all OHI calculations. You will need to run this script only once when you begin.

4.1.3.5 *functions.R* *functions.R* is an R script containing the equations for each goal and sub-goal model. Each goal and sub-goal equation is stored as a separate function within the script. These functions calculate the status and trend using prepared layers saved in the ‘layers’ folder and registered in *layers.csv*. You will need to code in R to modify or develop new models, and it is best to work on one goal at a time.

4.1.3.6 *goals.csv* *goals.csv* is a table with information about goals and sub-goals. This includes the weight of each goal that is used to calculate the final Index scores when all goals are combined. Other information includes the goal description that is also presented in the WebApp. *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).

TIP: It’s important to check *goals.csv*’s weightings and preindex functions when you change goal or sub-goal model equations in *functions.r*.

4.1.3.7 *launch_app_code.R* *launch_app_code.R* will launch a version of the App on your computer so that you can visualize any edits you make before synching to *github.com*.

4.1.3.8 *layers* folder The *layers* folder contains all layers required to calculate goal scores, and each layer is 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*. All *.csv* files can be read in R, or with text editors or spreadsheet editors like Microsoft Excel.

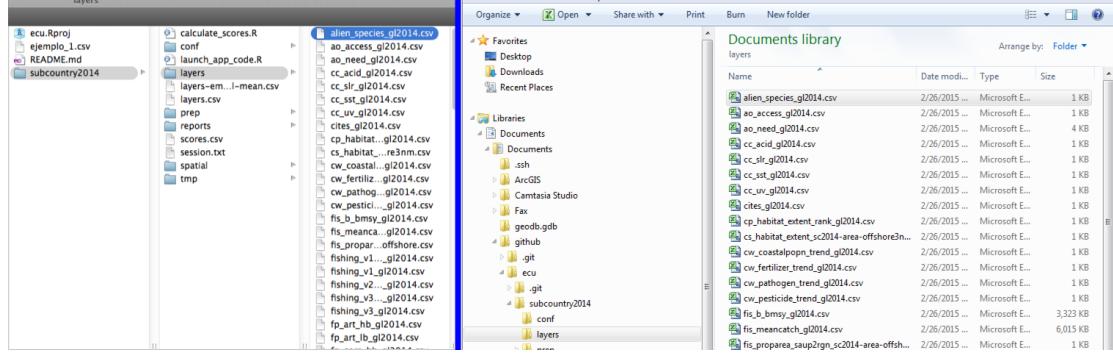


Figure 9: The *layers* folder contains every data layer as an individual *.csv* file. Mac navigation is shown on the left and Windows navigation is shown on the right.

Note that each *.csv* file within the *layers* folder has been formatted consistently. The Toolbox expects all data layers to be in the correct ‘long format’ and in separate files (see **Using the Toolbox**). Each file also has a column with unique region identifier (*rgn_id*). These numeric region identifiers have region names associated with them, that are set in *rgn_labels.csv* and can be modified.

TIP: You can check your region identifiers (*rgn_id*) in the *rgn_labels.csv* file in the *layers* folder.

4.1.3.8.1 **_gl2014 and **_sc2014** suffixes** In your repository, layers are provided for your country based on input information from the 2014 global assessment. The global assessment had information for your country at the spatial scale of the entire country, whereas your assessment has information for each subcountry region within your country. In most cases, layers from the global assessment was allocated equally to all regions in your study area (country). When this occurred, the layer was given a suffix of *_gl2014* to indicate that information is equal across all regions in the study area. While these layers may not provide much useful information to your assessment, the proper input structure is provided in these layers. Some layers contain information such as km² of habitat that could not be equally allocated across all regions since this would provide a sum much greater than reality. In these cases, layers were down-weighted based on the proportion of offshore area or coastal population density. These layers have the suffix *_sc2014* with an indication of what was used to downweight. While this method removes any error of inflated sums and provides the Toolbox with functioning layers, the allocation of these values may not be sensible to your study (i.e. if corals are only present in some regions of your study area but they are allocated to all).

4.1.3.9 *layers-empty_swapping-global-mean.csv* *layers-empty_swapping-global-mean.csv* contains a list of layers where information for your country was not available for the global assessment. For the Toolbox to be able to run, these layers were filled with averages from all other countries included in the global assessment. This file is not used anywhere by the Toolbox but is a registry of layers that should prioritized to be replaced with your own local layers if you require these layers for the models you develop.

4.1.3.10 *layers.csv* The *layers.csv* file is the registry and directory that manages all data required for your assessment. All relevant input information is prepared as individual data layers and then registered in this file. The Toolbox uses *layers.csv* to access the proper input information and display information on the WebApp. You will update some of the columns in *layers.csv*, and some of them will be auto-generated by the Toolbox code when it is running.

Figure 10: Differences in input layers with equal information for each region (suffixed with `_gl2014`) and weighted information for each region (suffixed with `_sc2014`).

TIP: `layers.csv` is a very useful reference throughout the assessment process.

`layers.csv` is easiest to view in spreadsheet software (i.e. Microsoft Excel). When you open it, you will see that each row of information represents an individual input layer that has been prepared for the Toolbox. The first columns contain information that will be updated by your team as you incorporate modified or new layers; all other columns are generated later by the Toolbox as it confirms data formatting and content and alerts you of any formatting inconsistencies. The columns you will update are: *targets*, *layer*, *name*, *description*, *fld_value*, *units*, *filename*.

1	targets	layer	name	description	fld_value	units	filename
2	AO	ao_access	Fisheries management	The opportunity for value creation	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	trend score	cw_coastalpopn_trend.csv
5	CW	cw_fertilizer_trend	Fertilizer consumption	Statistics on fertilizer consumption	trend.score	trend score	cw_fertilizer_trend.csv
6	CW	cw_pathogen_trend	Trends in access	Trends in percent of access	trend	trend score	cw_pathogen_trend.csv
7	CW	cw_pesticide_trend	Pesticide consumption	Statistics on pesticide consumption	trend.score	trend score	cw_pesticide_trend.csv
8	FIS	fis_b_bmsy	B/Bmsy estimates	Estimated B/Bmsy values	B / B_msy	B / B_msy	fis_b_bmsy.csv
9	FIS	fis_meancatch	Catch data for each species	Reported data including mean catch	mean_catch	metric tons	fis_meancatch.csv

Figure 11:

Columns you will update

- **targets** indicates which goal or dimension uses the layer. Goals are indicated with two-letter codes and sub-goals are indicated with three-letter codes (see the table just below). Pressures, resilience, and spatial layers indicated separately.
- **layer** is the identifying name of the input layer that 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 WebApp under the drop-down menu when the variable type is ‘input layer’.
- **name** is a longer title of the input layer; this is displayed on the WebApp under the drop-down menu when the variable type is ‘input layer’.

- **description** is further description of the input layer, including the source of the original data. This is also displayed on the WebApp under the drop-down menu when the variable type is ‘input layer’.
- **fld_value** the values’ units in the input layer. The information in this column must match the column header in the input layer.
- **units** the values’ units in the input layer. This differs from *fld_value* above as the *units* column is displayed on the WebApp and can have more descriptive naming.
- **filename** is the input layer itself. This file has input information for each region within the study area, and is located in the `subcountry2014/layers` folder.

Goal	Subgoal	2- or 3- letter code
Food Provision		FP
	Fisheries	FIS
	Mariculture	MAR
Artisanal Fishing Opportunity		AO
Natural Products		NP
Coastal Protection		CP
Carbon Storage		CS
Livelihoods and Economies		LE
	Livelihoods	LIV
	Economies	ECO
Tourism and Recreation		TR
Sense of Place		SP
	Lasting Special Places	LSP
	Iconic Species	ICO
Clean Waters		CW
Biodiversity		BD
	Habitats	HAB
	Species	SPP

4.1.3.11 *prep* folder The `prep` folder is included in your repository so that layer preparation can be collaborative and version controlled. It is not necessary to use this folder but you may find it useful as other assessments have.

4.1.3.12 *pressures_matrix.csv* `pressures_matrix.csv` is a table that indicates which individual pressures (stressors) affect which goal, sub-goals, or components, and weights them from 1-3 (a weight of 0 is shown as a blank). These weights are relative to each row of the matrix (goal, sub-goal, or component). These weights are used in global assessments based on scientific literature and expert opinion, and you can modify the weightings if necessary for your assessment. The pressures matrix is the same as Table S25 in the Supplementary Information for Halpern *et al.* 2012.

Each pressure (column) of the pressures matrix is the layer name of the pressures layer file that is saved in the `layers` folder and is registered in `layers.csv`. Pressures layers have values for every region in the study area and the filenames have prefixes to indicate the pressure category (for example: `po_` for the pollution category). Pressures values are scaled such that all values range from 0-1.

4.1.3.13 *reports* folder The `reports` folder contains flower plots and tables for every region in the study area and for the study area itself, which by convention is called ‘GLOBAL’ in these files.

4.1.3.14 *resilience_matrix.csv* `resilience_matrix.csv` is a table that indicates which individual resilience measures affect which goal, sub-goals, or components. Like the pressures matrix, the resilience

matrix also has weights, but these weights depend on the level of information available. These weights are stored in a separate file in the `conf` folder: `resilience_weights.csv`. The resilience matrix is the same as Table S26 in the Supplementary Information for Halpern *et al.* 2012.

Each resilience measure (column) of the resilience matrix is the layer name of the resilience layer file that is saved in the `layers` folder and is registered in `layers.csv`. Resilience layers have values for every region in the study area. Resilience values are scaled such that all values range from 0-1.

4.1.3.15 `resilience_weights.csv` `resilience_weights.csv` is a table that indicates the weight of each resilience layer based on the level of information available.

4.1.3.16 `scores.csv` `scores.csv` is a text file containing the calculated scores for each dimension (future, pressures, resilience, score, status, trend) for each region in the study area. Regions have the numeric identifiers set in `subcountry2014/layers/rgn_labels.csv` and the study area has the numeric identifier of 0. Scores are calculated with registered layers in `layers.csv`: when you begin an assessment this will be information for your country from the global 2014 assessment and goal models from the global 2014 assessment. Scores from `scores.csv` are viewed through the WebApp using the ‘Output Score’ pulldown menu on the ‘App’ page.

4.1.3.17 `session.txt` `session.txt` is not used in OHI calculations but stores information about how the Toolbox was installed which may be useful for debugging purposes.

4.1.3.18 *spatial* folder The *spatial* folder contains two spatial files: `regions_gcs.geojson` and `regions_gcs.js`. These files spatially identifies the study area and regions for the assessment and are stored in the JSON and GeoJSON formats that can be displayed by the App. If you plan to redefine the spatial boundaries for your assessment, you will need to provide a shapefile to the OHI+ development team and we will create the proper `regions_gcs.geojson` and `regions_gcs.js` files for you. You will need a spatial analyst to do this: see the **Defining spatial boundaries** section for instruction.

4.1.3.19 `temp` or `tmp` folders Contents within the `temp` or `tmp` folders are not used to calculate scores but can be used for temporary organization for your assessment.

4.2 Formatting Data for the Toolbox

4.2.1 Introduction

The OHI Toolbox 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). Each data layer (data input) has its own `.csv` file, which is combined with others within the Toolbox for the model calculations. These data layers are used for calculating goal scores, meaning that they are inputs for status, trend, pressures, and resilience. The global analysis included over 100 data layer files, and there will probably be as many in your own assessments. This section describes and provides examples of how to format the data layers for the Toolbox.

OHI goal scores are calculated at the scale of the reporting unit, which is called a ‘**region**’ and then combined using an area-weighted average to produce the score for the overall area assessed, called a ‘**study area**’. The OHI Toolbox expects each data file to be in a specific format, with data available for every region within the study area, with data layers 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 – see “**Long Formatting**” section).

4.2.2 Uploading and formatting raw data files

Unformatted data files can be uploaded to the `pre-proc` folder in your github repository and processed with R. Saving raw data in the same repository helps to keep track of how the data has been treated. Raw files can be uploaded as `.csv` or `.xlsx`. However, formatted data has to be saved as `.csv` in the `layers` folder.

In addition to `pre-proc`, a `prep` folder has been set up for data formatting. Inside the folder: - several sub-folders exist to house formatted data files for each goal/sub-goal - `prep.r` is where formatting occurs for each goal/sub-goal. - `README` is where you can record information on raw data files and processing for future reference

Example of data in the appropriate format:

<code>rgn_id</code>	<code>year</code>	<code>count</code>
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

<code>rgn_id</code>	<code>product</code>	<code>year</code>	<code>tonnes</code>
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

Figure 12:

4.2.3 Gapfilling

It is important that data prepared for the Toolbox 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 is preferred because it promotes easy transparency and reproducibility.

4.2.3.1 Temporal gapfilling Temporal gaps occur when a region is missing data for some years. The Toolbox requires data for each year for every region. It is important to make an informed decision about how to temporally gapfill data.

The screenshot shows a Microsoft Word document with a table of data and two callout boxes with red arrows pointing to specific parts of the table.

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.

Three red arrows point from the right side of the document towards the table: one arrow points to the first row of the table, another to the third row, and a third to the ninth row.

Figure 13:

Often, regression models are the best way to estimate data and fill temporal gaps. Here we give an example that assumes a linear relationship between the year and value variables within a region. If data do not fit a linear framework, other models may be fit to help with gapfilling. Here we give an example assuming linearity.

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: able to be represented by a straight line on a graph)):

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).

The screenshot shows a Microsoft Excel spreadsheet with a table of data and a formula bar at the bottom.

rgn_id	year	count
1	2005	177.14
1	2006	
1	2007	
1	2008	212.99
1	2009	228.81

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

The formula bar at the bottom shows the formula `SLOPE(D42:D46, C42:C46)`. A red arrow points down to the value 12.69 in the first row of the 'Slope' column.

Figure 14:

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 the **INTERCEPT(known_y's,known_x's)** function that calculates the y-intercept (-25273.89) of the fitted line.

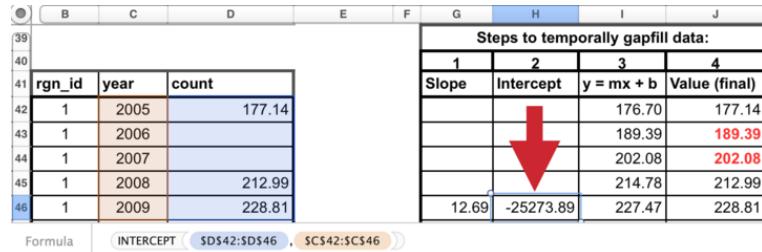


Figure 15:

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 (m =slope, x =year, b =intercept), to calculate the unknown ‘count’ for a given year (189.39 in 2006, and 202.08 in 2007).

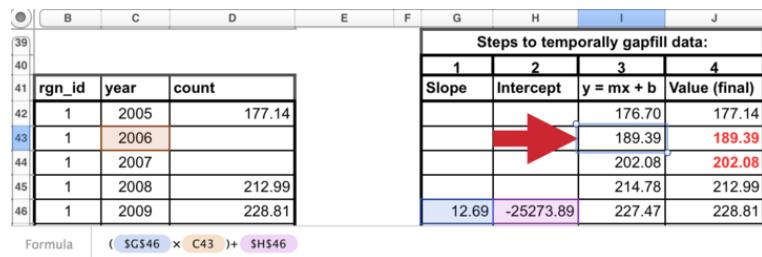


Figure 16:

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

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.*

4.2.3.2 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, you must assume that one region is like another, and data from another region is adequate to 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 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:



<i>rgn_id</i>	<i>year</i>	<i>count</i>
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

Figure 17:

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, in this case, 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* 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.

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

4.2.4 Long formatting

The Toolbox expects data to be in ‘long’ or ‘narrow’ format. 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:

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

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	153.42
2	2006	187.42
2	2007	214.84
2	2008	222.22
2	2009	225.38
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

Figure 18:

Region	DataLayer	Year									
		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	0.706	1	1
B	governance_indicator						0.8564	0.7794	0.861	1	1
C	governance_indicator						0.8779	1	1	0.8986	1
D	governance_indicator						0.8537	0.5373	0.7044	1	1

Figure 19:

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

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

- <http://blog.rstudio.org/2014/07/22/introducing-tidyr/>
- <http://www.rstudio.com/resources/cheatsheets/>

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),]
```

Figure 20:

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 easier for humans to read (R and the Toolbox can read these data without this final step):

Example of data in the appropriate (long) format:

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

Figure 21:

4.2.5 Rescaling your data

An important consideration is how to rescale your data when preparing it for use in the Toolbox. Rescaling involves turning a distribution of data into a value from zero to one. This is based on finding a highest observed or theoretical point in the distribution of the data, and from there, the relative value of the data can be calculated.

4.2.5.1 Example: Global Data Approach You should base your decision on whether your consider it more appropriate to decide the reference point based on the data distribution of all data points, be they observed or interpolated, or whether we think we should only consider the observed data. If the interpolation covers large areas, and these get assigned values that aren't very frequent in the observed data, then the two distributions will be very different, and what value is in the 99.99th percentile is different too.

In theory, one would favor deciding the reference point based on as many observations as possible (i.e., interpolate first, then obtain the percentile). In practice, if we think that large interpolated areas are very unreliable, we might prefer to use real observations only (i.e., percentile first, then interpolate).

5 Installing the Toolbox

In this section, you will learn how to successfully download, install, and use the software required to conduct an assessment. You will create a GitHub account and install R, RStudio, git, and the Github desktop app. OHI assessments are conducted through open-source platforms that allow you to make real-time changes with collaborators, and to track progress so that errors can be corrected and new insights can be shared in the future.

5.1 Overview

Your assessment repository is located at github.com/OHI-Science and we recommend saving it to your computer so that you can sync changes back online to save versions and facilitate collaboration. Conducting an OHI assessment using GitHub enables collaboration and transparency, and will provide access to the latest developments in the Toolbox software, allowing the OHI team to provide support remotely if necessary.

This section explains the GitHub workflow and how to access and setup required software. You can use GitHub to upload any modifications you make so that you can work collaboratively with your team.

Required:

1. GitHub
2. git
3. R
4. RStudio



Figure 22:

5.2 GitHub

GitHub, together with the version control system **git**, is an open-source development platform that enables easy collaboration and versioning, which means that all saved versions are archived and attributed to each user. It is possible to revert back to any previous version, which 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.

GitHub Vocabulary:

- **clone** ~ download to your computer from online version with syncing capabilities enabled
- **commit** ~ message associated with your changes at a point in time
- **pull** ~ sync a repo on your computer with online version
- **push** ~ sync the online repo with your version, only possible after committing
- **sync** = pull + commit + push

5.2.1 Learning GitHub

The following section describes how to use GitHub to access and sync your assessment repository. There are also many great resources available online with more in-depth information:

- **Git and GitHub** by Hadley Wickham: r-pkgs.had.co.nz/git.html
- **Collaboration and Time Travel: Version Control with Git, GitHub and RStudio** video tutorial by Hadley Wickham: www.rstudio.com/resources/webinars
- **Good Resources for Learning Git and GitHub** by GitHub: help.github.com/articles/good-resources-for-learning-git-and-github/

5.3 Accessing GitHub Repositories

GitHub has an online interface and a desktop application for the version-control software called ** git**, where your project repository and any changes done to it are kept and recorded. All changes within your

local, desktop repository will be tracked by GitHub regardless of the software you use to make the changes, and saved shared via its online interface. This means that you can delete or paste files in the Mac Finder or Windows Explorer and edit *.csv* files in Excel or a text editor, and still sync these changes with the online repository. To do so, you will need to create a GitHub account online and install git, R, and RStudio on your local computer. All are freely available.

5.3.1 Create a GitHub account

Create a GitHub account at <http://github.com>. Choose a username and password. You will use this username and password when you install and set up *git* on your computer.

5.3.2 Install *git* software

How you install *git* will depend on whether you are working on a Windows or Mac computer. It will also depend on your operating system version. If you have problems following these instructions, it is likely because your operating system requires a previous version of *git*. Previous versions are available from <http://www.wandisco.com/git/download> (you will need to provide your email address).

For Windows:

- Download *git* at <http://git-scm.com/downloads> and follow the install instructions.
- When running the Windows installer, use all default options except “Adjusting your PATH environment”: instead, select “***Run Git from the Windows Command Prompt***”. This will allow later compatibility with RStudio.

For Mac:

- Download *git* at <http://git-scm.com/downloads> and follow the install instructions.
- Apple’s **Xcode** has a command line tools option during install which can override the preferred *git* command line tools. To ensure you are using the latest preferred version of *git*, you will need to launch Terminal and type the following few lines of code:
- Access Terminal from the Applications folder: **Applications > Utilities > Terminal**. When you launch Terminal a window will appear with your computer’s name followed by a \$. When you type, your commands will appear after the \$.

Add access your ‘bash profile’ by typing:

```
pico ~/.bash_profile
```

You are now able to edit your ‘bash profile’ and your screen will change; it will look like this:

Now type:

```
export PATH=/usr/local/git/bin:$PATH
```

Exit pico by entering:

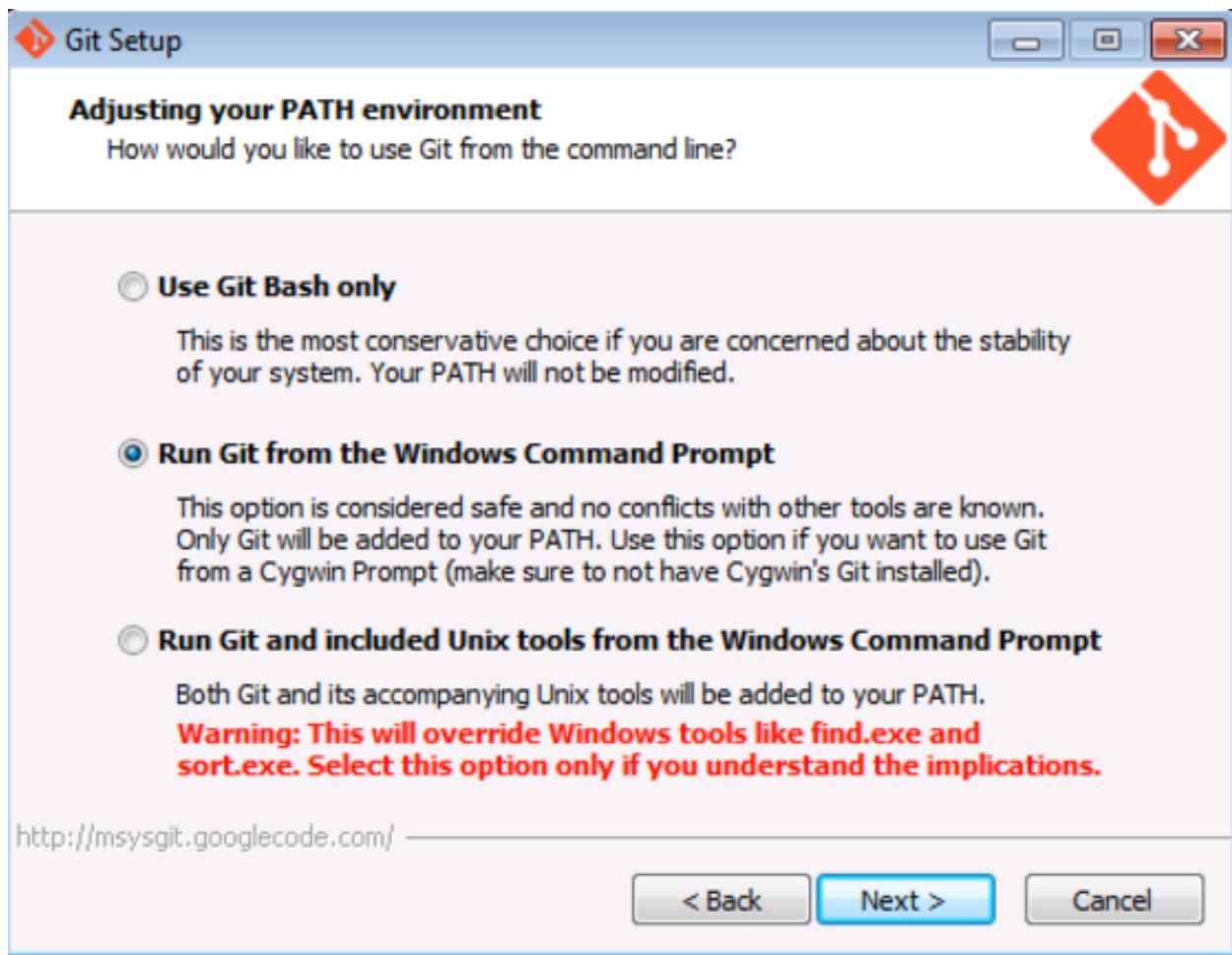


Figure 23:

```
GNU nano 2.0.6      File: /Users/julialowndes/.bash_profile

export PATH=/usr/local/git/bin:$PATH

[ Read 1 line ]
^G Get Help  ^O WriteOut  ^R Read File  ^Y Prev Page  ^K Cut Text  ^C Cur Pos
^X Exit     ^J Justify   ^W Where Is   ^V Next Page  ^U UnCut Text ^T To Spell
```

Figure 24:

```
control-X  
y  
return/enter
```

Exit Terminal by typing:

```
exit
```

Finally, quit Terminal.

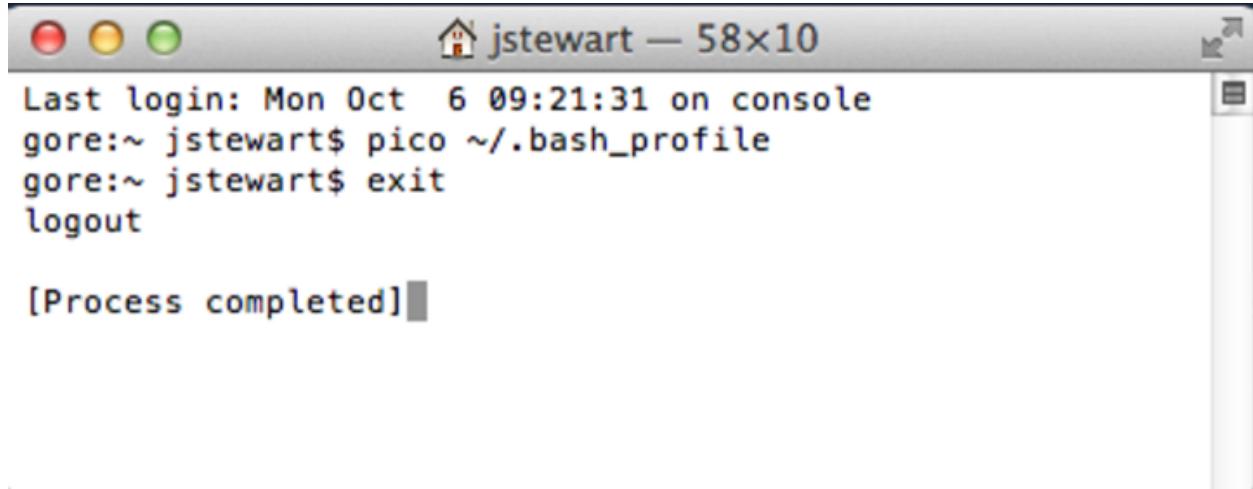


Figure 25:

5.3.3 Set up your Git Identity

After downloading and installing *git*, you will need to set up your **Git Identity**, which identifies you with your work. *Note:* if you have any problems with the following instructions, it is likely because of incompatibility between the version of your operating system and the version of git you downloaded in the previous section. In this case, find and download a compatible version at www.wandisco.com/git/download and then follow the instructions below.

You will set up your GitHub identity using the command line specific to Windows or Mac:

- **Windows:** Start > Run > cmd
- **Mac:** Applications > Utilities > Terminal

In the window, you will see a cursor where you are able to type. Type the following and press return (or enter) at each step. Make sure all spaces and symbols are identical to the example below, including all spaces () and dashes (-).

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
```

5.3.4 Create a folder called *github* on your computer

Because you will use GitHub to collaborate with your team or request support from the OHI team, it is important you save files in places where the file path that is universal and not specific to your computer. When team members save files in different places, this will create a lot of problems when collaborating, particularly between Macs and Windows machines.

Please create a folder called *github* in your root directory. The file path for this folder will be:

- **Windows:** `Users\[User]\Documents\github\`
- **Mac:** `Users/[User]/github/`

This folder can be identified by any computer as `~/github/`.

TIP: You can check the location of your *github* folder by right-clicking the folder icon and selecting ‘Get Info’ on a Mac or ‘Properties’ on Windows.

5.3.5 Update permissions

You need to **email your username to ohi-science@nceas.ucsb.edu** for permission to upload modifications to your GitHub repository (you only need to do this once). Only team members who will be modifying files will need to do this; all other members can view online and download the repository without these permissions.

5.3.6 Work locally with R and RStudio

You will then work locally on your own computer, modifying the files in the repository to reflect the desired modifications your team has identified for your assessment. Multiple users can work on the same repository at the same time, so there are steps involved to ‘check in’ your modifications so they can merge with the work of others without problems. GitHub has specific words for each of these steps. You have already successfully **cloned** an online repository to your local machine. After making modifications, you will **commit** these changes with a description before being able to sync back to the online repository. **Synching** involves both **pulling** any updates from the online repository before **pushing** committed changes back to the server.

TIP: While you can edit files in the online GitHub repository, we do not recommend this. It is good practice to track changes through commits and syncing.

The example below illustrates GitHub’s collaborative workflow:

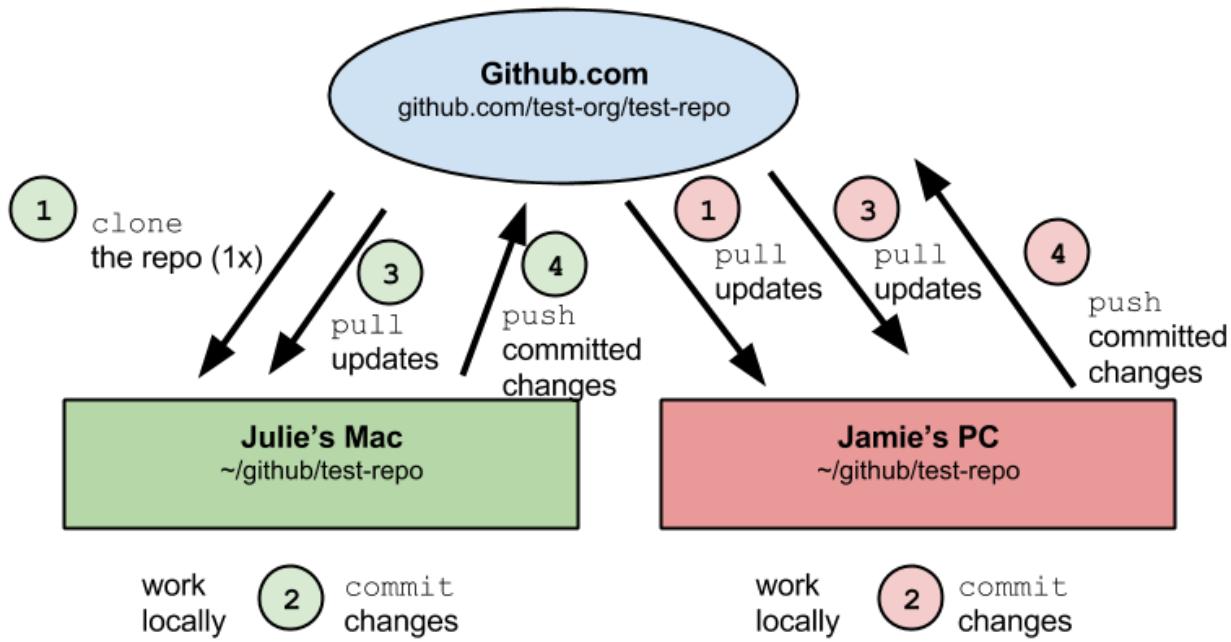


Figure 26:

All changes within your local repository will be tracked by GitHub regardless of the software you use to make the changes. This means that you can delete or paste files in the Mac Finder or Windows Explorer and edit .csv files in Excel or a text editor, and still sync these changes with the online repository. We recommend doing as much data manipulation as possible in a programming language like R, to maximize transparency and reproducibility.

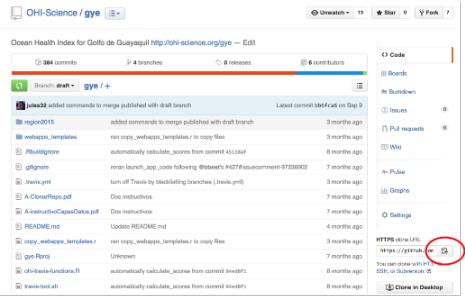
- **R** is a free computing and graphical software where all the modifications to your OHI repository are done. Download the latest version of R appropriate for your operating system at <http://cran.r-project.org/> and follow the instructions to install it on your computer.
- **RStudio** is a user-friendly R interface that can be used to synchronize any modifications you make to files in your assessment's repository. It is also where model modifications and calculations occur. Install the latest version of RStudio at <https://www.rstudio.com/>.

R and RStudio update frequently. If you already have those softwares installed, check on the websites for the most recent version. In general, it's good practice to check periodically for updates. When you run into unexplained errors while modifying contents in RStudio, check first if your versions of R or Rstudio are out-of-date. For R, you can see what you already have on your computer by typing “sessionInfo()” into your R console.

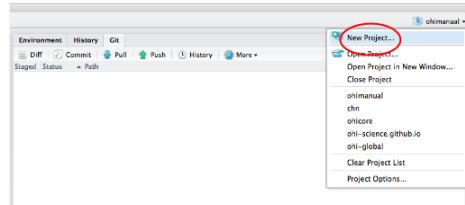
If you are working on a *Mac*, you will need to tell RStudio to use the proper version of Git by doing the updating the preferences for ‘Git executable’:

RStudio > Preferences... > Git/SVN > Git executable: /usr/local/git/bin/git

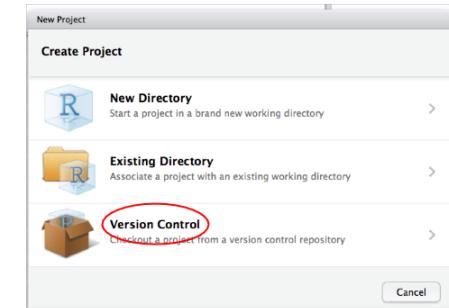
5.3.6.1 Cloning a repository to your local computer In order to sync GitHub and RStudio, you need to clone your GitHub repository to RStudio. Here is a step-by-step guide on how to do so:



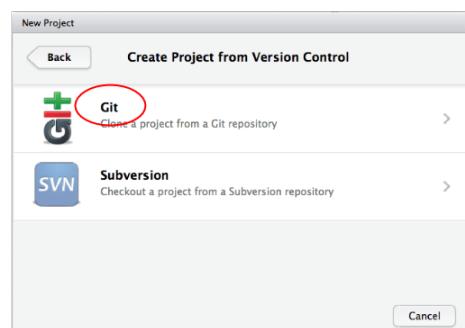
Step 1: Copy the link by clicking the highlighted button, or select and copy the link inside the text box.



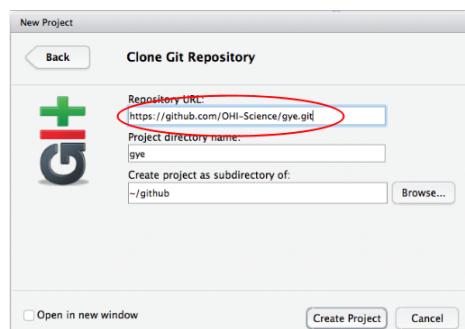
Step 2: Open Rstudio interface, click the project button (eg. ohimanual) at the upper right corner, select "New Project".



Step 3: Click "Version Control".



Step 4: Click "Git".



Step 5: Paste the link you copied from step 1 in the circled box. Hit the "tab" key on your keyboard and the following two boxes will be filled automatically. Click "Create Project". Once downloading is complete, the proper project R page will open automatically.

Figure 27: How to clone repository to Rstudio
39

5.3.6.2 Syncing your repository with RStudio RStudio can sync files with GitHub directly. It will capture the changes made to any files within the repository, no matter which software was used to modify them. In RStudio, you sync by first pulling and then pushing (separately); in the GitHub App these two functions are done together.

Launch your project in RStudio by double-clicking the .Rproj file in the assessment folder on your local hard drive.

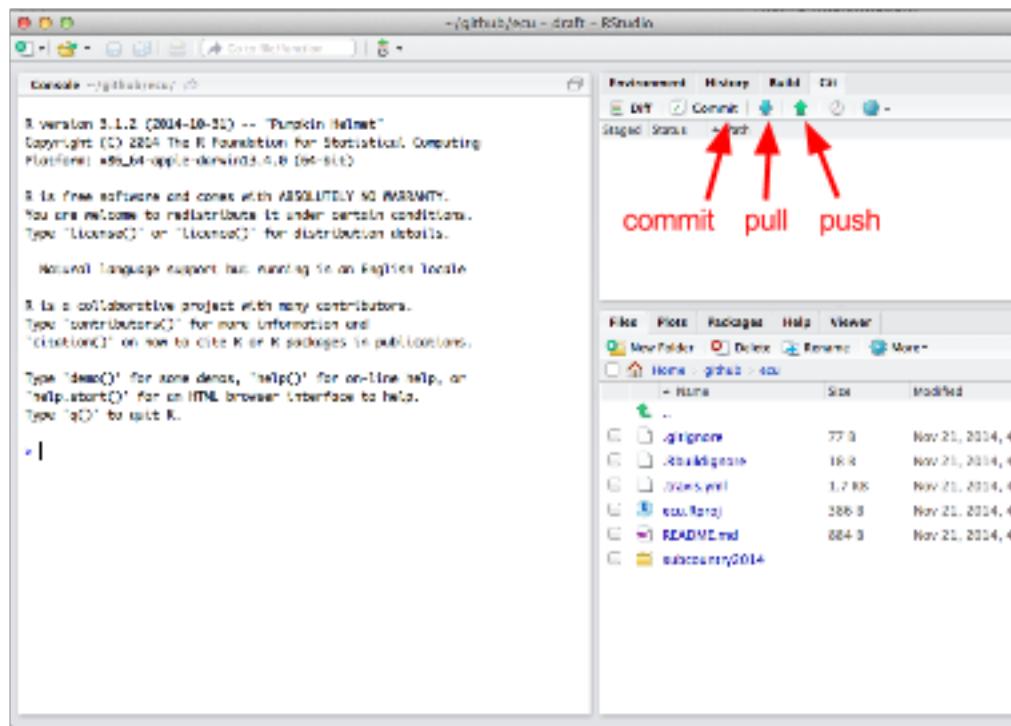


Figure 28:

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

1. Clicking the ‘*Staged*’ (checked) 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 repository. This is important to ensure there are no conflicts with updating the online repository, especially if you are working with collaborators who might be working on the same files as you are.

5. “Push” your committed changes to the online repository. Your changes are now visible online.

TIP: If you aren’t seeing your changes in the ‘Git’ window, try saving the file again.

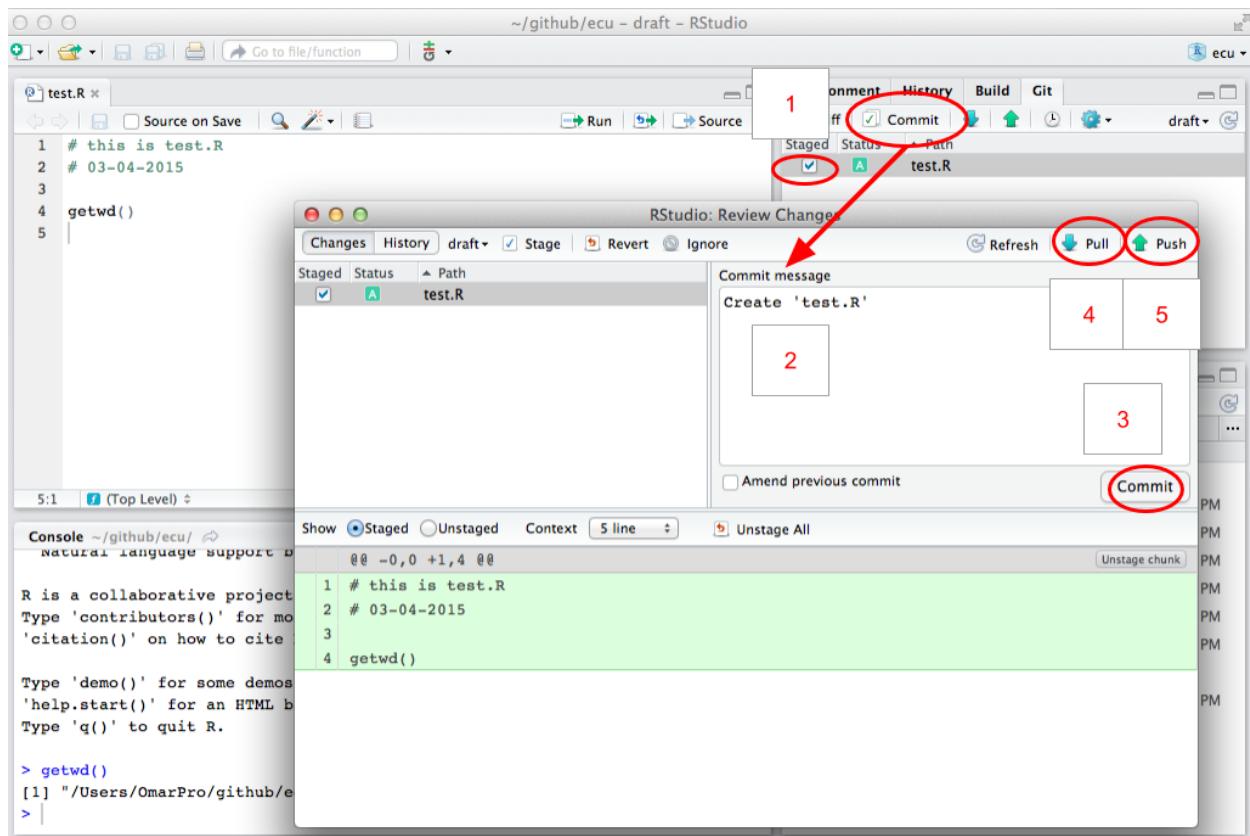


Figure 29: Figure showing RStudio when syncing. After first staging your changes, click the ‘commit’ button to open a new window where you can enter a ‘commit message’ and then pull and push new changes.

Note: Another option to syncing your edits on a repo with the online version is to use **Command Lines**, if you are familiar with it. There are resources available online on how to do so.

5.4 GitHub repository architecture

GitHub stores all data files and scripts for your assessment in a repository (a folder). Different copies or complements to these folders, called *branches* can also exist, which aid with versioning and drafting. Your repository has four branches, two of which are displayed on your website (e.g., ohi-science.org/ecu):

1. **draft** branch is for editing. This is the default branch and the main working area where existing scenario data files can be edited and new scenarios added.
2. **published** branch is a vetted copy of the draft branch, not for direct editing. This branch is only updated by automatic calculation of scores if:
 1. no errors occur during the calculation of scores in the draft branch, and
 2. publishing is turned on. During the draft editing and testing phases of development, it is typically desirable to turn this off.

3. **gh-pages** branch is this website. The results sections of the site (regions, layers, goals, scores per branch/scenario) are overwritten into this repository after automatic calculation of scores. The rest of the site can be manually altered.
4. **app** branch is the interactive layer and map viewer application. The user interface and server-side processing use the [Shiny](#) R package and are deployed online via [ShinyApps.io](#) to your website. Once deployed, the WebApp pulls updates from the data branches (draft and published) every time a new connection is initiated (i.e., browser refreshes).

TIP: When looking at files on GitHub, note that the timestamps are associated with the ‘commit’ time rather than the ‘push’ time.

6 Using the Toolbox

In this section, you will learn about the most common modifications made to repositories. You will be given examples to follow to help with your own assessment. The most common modifications are changing the pressures and resilience matrices, changing or creating data layers, and changing or removing goals models.

As your team finalizes which data should be included in the assessment and begins developing goal models, you can incorporate this information into your repository. Input information must be properly formatted into **layers**, which are registered with the Toolbox for use. Layers for the Toolbox can be prepared with any software that handles *.csv* files, but goal models must be updated in R. It is recommended that layer preparation occurs within your repository’s **prep** folder as much as possible, as it will also be archived by GitHub. Calculations can be done locally and offline by running `subcountry2014/calculate_scores.R`.

6.1 Layer preparation workflow

It is recommended that you construct a useful workflow with your team to incorporate local information into the Toolbox. Adding layers to the Toolbox will require working with GitHub and the Toolbox file system structure. The overall process involves preparing the layers (which can be done in the **prep** folder), saving them in the **layers** folder, and registering the layers. The layer preparation process can occur in tandem with the model modification process.

6.2 Modifying and creating data layers

Data layers are *.csv* files and are located in the `[assessment]/subcountry.year/layers` folder (example: `gye/subcountry2014/layers`). Remember that all data layers provided in your repository are extracted from the global 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 region in your assessment 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 incorrectly increase the nation’s GDP several times), national GDP was down-weighted by the proportion of coastal population in each region compared with the total coastal population.

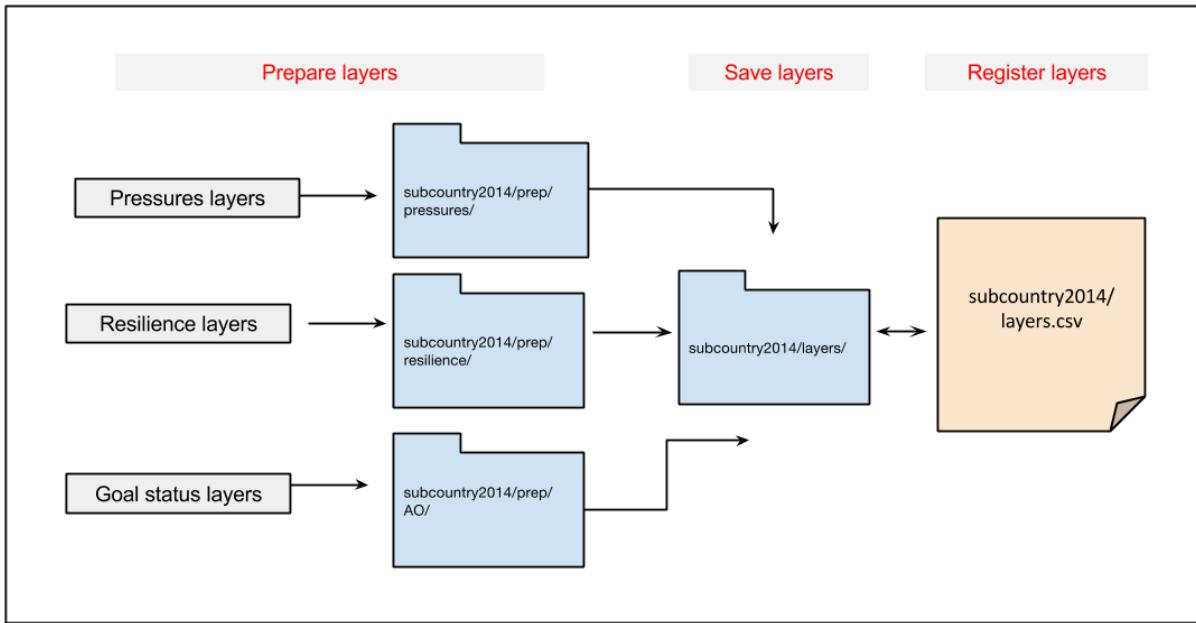


Figure 30: Diagram of OHI Toolbox data preparation workflow. You should start by prepping the files, loading them into the `layers` folder when they're ready for the Toolbox, and then registering them in `layers.csv`

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

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: `[assessment]/subcountry2014/conf`)

6.2.1 Create data layers with proper formatting

The OHI Toolbox expects each data layer to be in its own `.csv` file and 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. See the ‘Formatting data for the Toolbox’ section for more information.

6.2.2 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 assessment (example: `_sc2014.csv`). Modifying the layer name provides an easy way to track which data layers have been updated regionally, and which rely on global data. Then, the original layers (`_gl2014.csv` and `_sc2014.csv`) can be deleted.

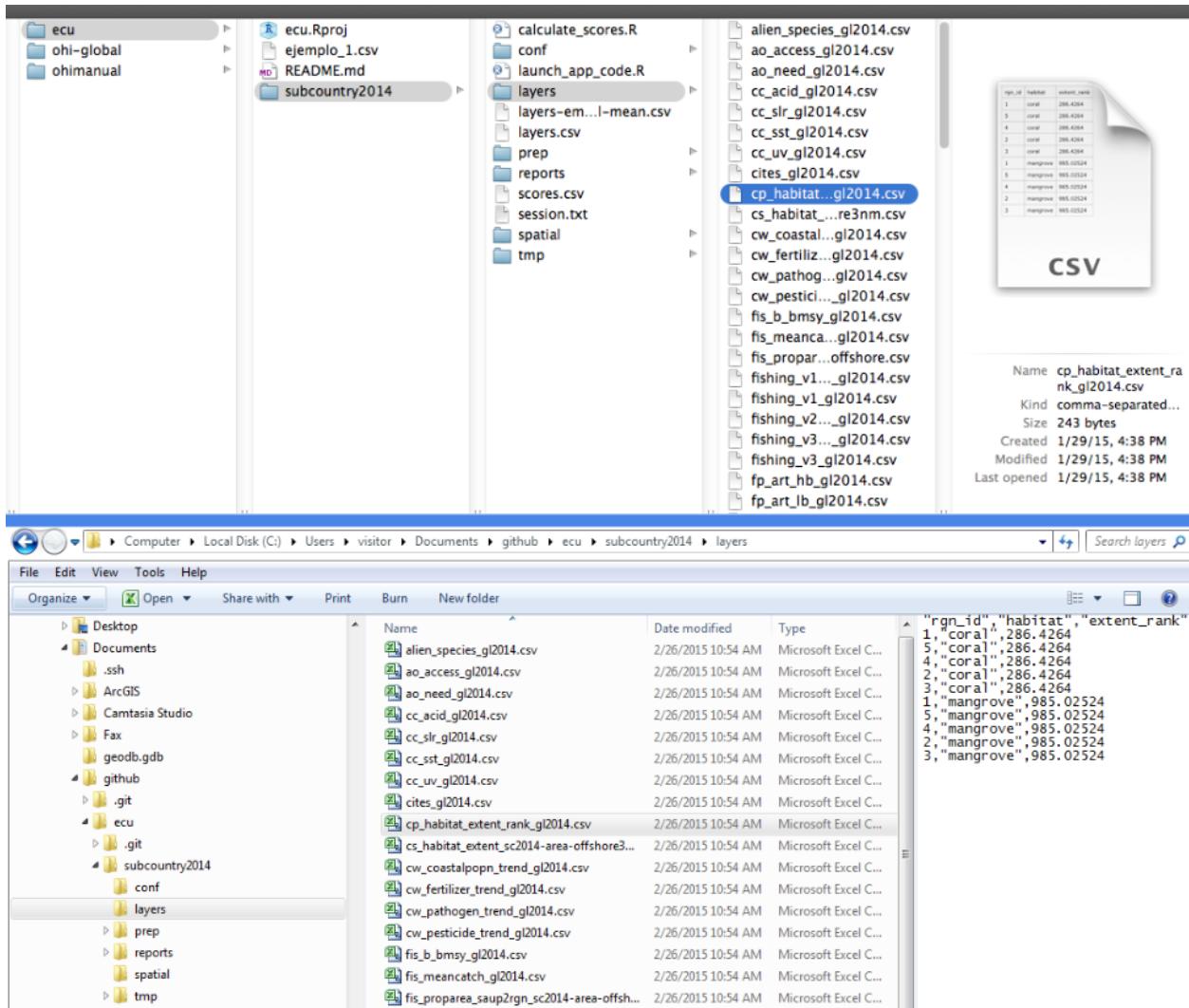


Figure 31: This figure shows the location of your data layers. Mac navigation is shown above and Windows is shown below.

* Note: filenames should not have any spaces: use an underscore (' ') instead. This will reduce problems when R reads the files.

6.2.3 Register data layers in `layers.csv`

When there are new filenames associated with each layer, they will need to be registered in `[assessment]/subcountry2014/layers.csv`. If a layer simply has a new filename, only the *filename* column needs to be updated:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V		
1	targets	layer	name	description	fld_value	units	filename	clip_n_shi	clip_n_ship	dislayer_gl	path_in	rgns_in	fld_id	r_fld_id	fld_catgex	fld_ye	fld_val_nf	fld_val_ch	file_exist	year_min	max_val_min		
2	AO	ao_access	Fisheries	The	value	value	ao_access_g2014.csv	equal	global	manager	ao_access	/github/global	rgn_id		value	TRUE				1990	2013	0.055772	
3	AO	ao_need	Purchasing	The	per	value	ao_need_g2014.csv	equal	global	purchasi	ao_need	/github/global	rgn_id	year	value	TRUE				1990	2013	0.063467	
4	CW	cw_coasta	Coastal	trend	trend	score_cw_coasta	trend_g2014.csv	equal	global	trends	score_cw_coasta	/github/global	rgn_id	trend	score	TRUE						0.077218	
5	CW	cw_fertil	Fertilizer	Statistics	trend	score_cw_fertilizer	trend_g2014.csv	equal	global	trends	score_cw_fertilizer	/github/global	rgn_id	trend	score	TRUE						-1	
6	CW	cw_patho	Trends in	Trends	in	trend	trend_cw_pathogen_trend_g2014.csv	equal	global	trends	score_cw_pathogen_trend	/github/global	rgn_id	trend	score	TRUE						0.002294	
7	CW	cw_pestic	Pesticide	Statistics	trend	score_cw_pesticide	trend_g2014.csv	equal	global	trends	score_cw_pesticide	/github/global	rgn_id	trend	score	TRUE						0.878753	
8	FIS	fi_b_bmsy	B/Bmsy et	The ratio	b_bmsy	B/B_msy	fi_b_bmsy_g2014.csv	equal	global	mean	ratio	/github/global	fao_id	taxon_nar	year	b_bmsy	TRUE			1954	2011	0.054352	
9	FIS	fi_mean	Catch data	Mean	mean_cat	metric	tor	fi_meancatch_g2014.csv	equal	global	mean	cat/fis_mean	/github/global	fao_id	taxon_nar	year	mean	catch	TRUE	2006	2011	8.18E-18	
10	FIS	fp_preparea	area	Lookups	prop_area	proprio	fp_preparea_saup_g2014-area-offshore.csv	equal	global	prop/area	prop/area	/github/global	rgn_id	saup_id	prop_area	TRUE						0.001827	
11	FP	fp_wildca	Fisheries	Proprio	w_fis	value	fp_wildcaight_weight_g2014.csv	equal	global	weights	p_wildca	/github/global	rgn_id	w_fis	value	TRUE						0.469561	
12	HAB CS	CF_hab	exterior	Habitat	Model	ed km2	km2	hab_extent_g2014.csv	raster	area	offshore	hab_extent	/github/global	rgn_id	habitat	km2	TRUE						71.6066
13	HAB CS	CF_hab	health	Habitat	Model	health	value	hab_health_g2014.csv	equal	global	habitat	hab_health	/github/global	rgn_id	habitat	health	TRUE						0.741379
14	HAB CS	CF_hab	trend	Habitat	Model	trend	trend	score_hab_trend_g2014.csv	equal	global	trends	score_hab_trend	/github/global	rgn_id	trend	score	TRUE						-1
15	ICO	ico_spp_IUCN	Extinction	International	category	category	ico_spp_extinction_status_g2014.csv	equal	global	sciname	category	/github/global	rgn_id	sciname	category	TRUE							
16	ICO	ico_spp_IUCN	pop	International	popn_trt	trend	score_icospopn_trend_g2014.csv	equal	global	sciname	category	/github/global	rgn_id	sciname	category	TRUE							
17	LE	le_gdp	GDP	Gross	usd	2010	USD	le_gdp_sc2014-popn-inland25km.csv	population	gdp	year	/github/global	rgn_id	year	usd	TRUE				1998	2011	1327122	
18	LE	le_jobs	Jobs	gapfilled	value	jobs	le_jobs_sector_year_g2014.csv	equal	global	sector	year	/github/global	rgn_id	year	value	TRUE				1997	2010	15.32	
19	LE	le_unemp	Unemployment	gapfilled	percent	percent	le_unemployment_g2014.csv	equal	global	sector	year	/github/global	rgn_id	year	percent	TRUE				1997	2010	6.065547	
20	LE	le_wage	Wages	gapfilled	value	2010	USD	le_wage_sector_year_g2014.csv	equal	global	sector	year	/github/global	rgn_id	year	value	TRUE				1993	2006	307.6993
21	LE	le_workfrc	Modelled	adjusted	jobs	jobs	le_workforce_size_adj_sc2014-popn-inland25km.csv	population	gdp	year	jobs	/github/global	rgn_id	year	jobs	TRUE				1990	2012	112.473	
22	LE	pressure_secto	Jobs	weig	jobs	weight	le_sector_weight_g2014.csv	equal	global	sector	weight	/github/global	rgn_id	sector	weight	TRUE						1	
23	LIV ECO	le_popn	Total	Populatio	count	count	le_popn_g2014.csv	equal	global	popn	count	/github/global	rgn_id	year	count	TRUE				1960	2012	4514593	
24	LSP	lsp_prot_Coastal	pr Coastal	area_km2	km2	lsp_prot_area_inland1km	lsp_prot_area_g2014.csv	raster	area	inland1km	lsp_prot	/github/global	rgn_id	year	area_km2	TRUE				1959	2007	1	
25	LSP	lsp_prot_Coastal	m Coastal	area_km2	km2	lsp_prot_area_offshore3nm	lsp_prot_area_g2014.csv	raster	area	offshore3nm	lsp_prot	/github/global	rgn_id	year	area_km2	TRUE				1959	2007	3	

Figure 32: Register new layers in `layers.csv`. Be sure to note if there is a change in the filename.

TIP: This part is done manually. If you prefer not to manipulate your file by hand, you can generate a script that automates this.

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):

- **targets:** Add 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.
- **layer:** Add an identifying name for the new data layer, which will be referenced 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 on your WebApp.
- **description:** Add a longer description of the new data layer–this will be displayed on your WebApp.
- **fld_value:** Add the appropriate units for the new data layer. It is *the same as the column name in the data file*, which will be referenced in R scripts in subsequent calculations. (example: `area_km2`)
- **units:** Add a description about the *units* chosen in the `fld_value` column above. Think about what units you would like to be displayed on the WebApp when filling out “units.” (example: `km^2`)
- **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`.

It is important to check that you have filled you the fields correctly, for instance, if “`fld_value`” does not match the header of the source data layer, you will see an error message when you try to calculate scores. Other columns are generated later by the Toolbox as it confirms data formatting and content.

6.2.4 Check pressures and resilience matrices

If the new or modified layer is a pressures layer, check that `pressures_matrix.csv` and `resilience_matrix.csv` have been properly modified to register the new data layers.

6.3 Modifying pressures matrices

Your team will identify if any pressures layers should be added to the pressures matrices, and if so, which goals the pressure affects and what weight they should have (see Appendix 6 for guidance on Pressure and Resilience). You can transfer this information in `pressures_matrix.csv` (located in the `[assessment]/subcountry2014/conf` folder). It is important to note that the matrix identifies the pressures relevant to each goal, and which weight will be applied in the calculation. Each pressure is a data layer, located in the `subcountry2014/layers` folder. This means that pressure layers need information for each region in the study area, and some layers will need to be updated with local data. In modifying pressures, you will need to consider whether data layers can be updated or added, and whether data layers map onto goals appropriately in the local context.

Adding a new pressure to the pressures matrix requires the following steps:

1. Create new pressure layer(s) and save in the `layers` folder
2. Register pressure layer(s) in `layers.csv`
3. Register pressure layer(s) in `pressures_matrix.csv`
 - a. Set the pressure category
 - b. Identify the goals affected and set the weighting
 - c. Modify the resilience matrix (if necessary)

The following is an example of adding two new pressures layers.

6.3.1 Create the new pressure layers and save in the `layers` folder

If you create a new data layer, give it a short but descriptive name that also includes a prefix that signifies the pressure category (for example: `po_` for the pollution category). There are five physical categories and one social category:

- `po_` = pollution
- `hd_` = habitat destruction
- `fp_` = fishing pressure
- `sp_` = species pollution
- `cc_` = climate change
- `ss_` = social pressure

So for example, `po_trash` is a pollution layer with trash on beaches, and `sp_alien` is species pollution due to alien (invasive) species.

In the current example, the two new layers created to account for the input and output effects of desalination operations will be called `po_desal_in`, and `po_desal_out`.

These new layers will have scores from 0 to 1, with values for each region in your study area, and will be saved in the `layers` folder.

6.3.2 Register the new pressure layers in `layers.csv`

Add two new rows in `layers.csv`, and register the new pressure layers by filling out the first eight columns for `po_desal_in`, and `po_desal_out`.

	A	B	C	D	E	F	G	H
	targets	layer	name	description	fid_value	units	filename	fid_id_num
60	TR	tr_sustainability	Sustainability in Tourism	Comp score	score		tr_sustainability_global2013.csv	rgn_id
61	TR	tr_unemployr	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 acidification	Modeled distri pressure_scor	pressure score		cc_acid_global2013.csv	rgn_id
65	pressures	rr_slr	Sea level rise	Modeled sea lr pressure_scor	pressure score		rr_slr_global2013.csv	rnn_id

Figure 33:

6.3.3 Register the new layers in `pressure_matrix.csv`

`pressures_matrix.csv` identifies the different types of ocean pressures (columns) with the goals that they affect (rows). Adding a new pressures layer to `pressures_matrix.csv` requires adding a new column with the pressure layer name.

6.3.3.1 Set the pressure category This step requires transferring previous decisions made by your team into `pressures_matrix.csv`. Each pressure 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.

6.3.3.2 Identify the goals affected and set the weighting This step also requires transferring prior decisions into `pressures_matrix.csv`. 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 = highly influential pressure, 2 = moderately influential pressure, 1 = not very influential pressure). Remember that the rankings in the pressures matrix are separate from the actual data within the pressures data layers. The rankings ensure that within a particular goal (e.g. within a row of the pressures matrix), the stressors that more strongly influence the goal's delivery have a larger contribution to that goal's overall pressure score. Therefore, the rankings are assigned independently of the actual pressure scores, and only determine their importance within the calculations.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	goal	component	component_name	po_desal_in	po_desal_out	po_chemicals	po_chemicals	po_pathogens	po_nutrients	po_tides	hd_susidal	shd_susidal	hd_intertidal	
2	FIS			3	2	1			1		3	2	2	1
3	AD			3	2	2			1		1	3	1	
4	AO						1							
5	NP	corals		1	2	1					2			
6	NP	fish_oil		1	2	2					1			
7	NP	ornamentals		1	2	2					1			
8	NP	seaweeds		1	2	2					2			
9	NP	shells		1	2	2					1			
10	NP	sponges		1	2	2					2			
11	GP	mangrove		2	2	1					1			3
12	GP	saltmarsh		2	2	1					2			3
13	GP	seagrass		2	2	2					3			3
14	GP	corals		2	2	1					2			3
15	GP	mangroves		2	2	1					1			3
16	GP	salterns		2	2	1					2			3
17	GP	seagrass_shoreline		2	2	1					2			3
18	GP	seagrass		2	2	1					3			3
19	SP						2				2			
20	CW						3				3			2
39	HAB	coral					3				2			3
40	HAB	mangroves					1				1			3
41	HAB	saltmarsh					1				2			3
42	HAB	seagrass_didge					1				1			3
43	HAB	seagrass					2				2			3
44	SP	soft_bottom					2				1			2

Figure 34:

6.3.4 Modify the resilience matrix (if necessary)

Resilience is included in OHI as the sum of the ecological factors and social initiatives (policies, laws, etc.) that can positively affect goal scores by reducing or eliminating pressures. The addition of new pressure layers may therefore warrant the addition of new resilience layers that were not previously relevant. Similarly, the removal of pressure layers may warrant the removal of now irrelevant resilience layers.

6.4 Modifying resilience matrices

Previous decisions made with your team will identify if any resilience layers should be added to the resilience matrices, and if so, which goals and/or pressures the resilience affects and what weight they should have. You can then transfer this information into `resilience_matrix.csv` (located in the [assessment]/subcountry2014/conf folder).

`resilience_matrix.csv` maps the different types of resilience (columns) with the goals that they affect (rows). New resilience layers may be added to `resilience_matrix.csv` based on finer-scale local information either in response to a new pressures layer, or as a new independent measure. Any added layer must be associated with a pressures layer that has a weight of 2 or 3 in the OHI framework so that resilience measures can mitigate pressures in each region.

Each goal must have a resilience measure associated with it. In the figure below, the Toolbox would give an error because there are no resilience layers indicated for the natural products (NP) goal.

goal	component	alien_species	cites	fishing_v1	fishing_v1_eez	fishing_v2_eez	fishing_v3	fishing_v3_eez	habitat	habitat_combi	habitat_combo	habitat_li_gci	li_sector_even	mariculture	msi_gov	species_diver	species_diver_tourism	water	wgi_all
AO									habitat_combo							species_diver	species_diver_3nm	water	wgi_all
HAB	coral only	alien_species								habitat_combo_eez				mariculture		species_diver	tourism	water	wgi_all
HAB	alien_species									habitat_combo_eez				mariculture		species_diver	tourism	water	wgi_all
HAB	soft_bottom	alien_species				fishing_v2_eez				habitat_combo_eez				mariculture		species_diver	tourism	water	wgi_all
HAB	soft_bottom	alien_species			fishing_v1_eez					habitat_combo_eez				mariculture		species_diver	tourism	water	wgi_all
SPP						fishing_v2_eez				habitat_combo_eez				mariculture		species_diver	tourism	water	wgi_all
CS									habitat_combo								water	wgi_all	
CW																	water	wgi_all	
FIS										habitat_combo_eez				mariculture		species_diver		water	wgi_all
MAR														msi_gov				water	wgi_all
ECO																		water	wgi_all
LIV																		water	wgi_all
NP																			
CP									habitat_combo							species_diver		water	wgi_all
ICO										habitat_combo_eez								water	wgi_all
LSP			cites			fishing_v2_eez			habitat									water	wgi_all
TR																		water	wgi_all

Figure 35:

6.4.1 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 needed some revision.

The habitats assessed for `ohi-israel` are:

```
rocky_reef, sand_dunes, soft_bottom
```

Updates are required for the following files:

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

6.4.1.1 Global resilience layers The first step is to determine which resilience layers from the global assessment are relevant to your assessment, and whether others need to be added. The full list of layers included in the global resilience matrix 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_diver, species_diver_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 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`

6.4.1.2 Determining how to modify these resilience layers

- To determine whether `species_diversity_3nm` or `species_diversity` should be used:
 - `sand_dunes` should use `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, which are composite indicators obtained from different combinations of the following indicators:

`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 score assigned based on answers to a questionnaire compiled by countries that committed to Rio's Convention on Biological Diversity (CBD) to establish their progress towards habitat biodiversity protection
- `MPA_coast` is an indicator obtained as 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 an indicator obtained as 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 indicators are used by each combo layer:

Layer	Mora	Mora_s4	CBD_hab	MPA_coast	MPA_eez
<code>fishинг_v1</code>	Mora		CBD_hab	MPA_coast	
<code>fishинг_v1_eez</code>	Mora		CBD_hab		MPA_eez
<code>fishинг_v2_eez</code>	Mora	Mora_s4	CBD_hab		MPA_eez

Layer	Mora	Mora_s4	CBD_hab	MPA_coast	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:

The first objective is to determine whether the general `fishing_v...` or `habitat_...` categories are relevant to each of the habitats. For example, fisheries regulations do not affect the conservation of sand dunes, so this habitat should not use any of the fisheries combos. If the general resilience categories are relevant to the habitat, the next step is to select one resilience layer within the `fishing_v...` and `habitat...` categories that most adequately captures the suite of combined resilience variables that affect the habitat. For example, the sand dune habitat is a strictly coastal habitat, so the most appropriate resilience layer would be the one that uses the `MPA_coast` (i.e., `habitat_combo`). The rocky reef and soft bottom, on the other hand, should definitely include fisheries and habitat regulations. So, you'll need to choose a fisheries and a habitat combo for these two habitats. To do so, 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) It may also be that the existing global combo layers are not appropriate for your habitats. For example, if rocky reef is mainly coastal, and it is fished by both commercial and artisanal methods, then we need a new 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) Another issue to consider is whether local data are available to improve the pressure layers (that are based on global data). For example, 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. You may know that only certain types of protected areas are closed to fisheries, and may want to only include those. Also, local datasets may be more accurate and regularly updated. **NOTE: in the global study, 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.

6.5 Modify goal models

Before getting started on the R codes, make sure that you have finished these steps:

- Check that you have installed the latest versions of R, RStudio, and GitHub
- Synchronize GitHub and Rstudio
- Update data layers, pressure, and resilience in both layers folder and layers.csv

6.5.1 Setup

Now you are ready to modify your goal models and calculate scores. Here are a few steps to set it up. We will use CHN Carbon Storage goal as an example:

1. Open your project (eg. CHN) in RStudio, and open the folder of your specific assessment (eg. province2015).
2. Run `install_ohicore.R`. *This only needs to be done once to load all the background functions for OHI.*
3. Run `calculate.scores.R` from line 1 all the way through `Load Scenario Layers` section. *This will load all the data layers and call the background functions for your next step.*
4. In `conf` sub-folder, open `functions.R`. *This is where all the status and trend calculations occur.*
5. Go to the appropriate goal section.

6.5.2 Model modification

Your repository is pre-loaded with r codes for calculations from the 2014 Global assessment. Regardless of how you have changed your models, the basic sequence of events are similar for all goals. In `functions.R`, each goal is set up as a function (eg. `HAB = functions(layers) {...}`) and you will make modifications for each goal within its function. Below is the step-by-step instruction on how to modify CS goal model and calculate its status and trend, as an example.

6.5.2.1 Load data

1. Identify and select the data layers we need. (*Note that the layer names are what was set up in layers.csv. Now the toolbox will look for those layers*)

```
lyrs = c('cs_condition',
        'cs_contribution',
        'cs_extent',
        'cs_extent_trend')

D = SelectLayersData(layers, layers=lyrs)
# SelectLayersData is an `ohicore` function that will call the layers from layers folder you just nam
```

```
head(D); summary(D)
````
```

It is good practice to use `head()` and `summary()` after each step to make sure the data looks the way it should.

```
````

id_num      category val_num      layer id_name val_name category_name
1          1 saltmarshes    0.8 cs_condition rgn_id   value      habitat id_num | category | val_num
2          2 saltmarshes    0.8 cs_condition rgn_id   value      habitat id_num | category | val_num
... ...

id_num      category val_num      layer      id_name      val_name
Min. : 1.000 mangroves :20 Min. : -0.1 Length:84      Length:84
1st Qu.: 4.000 saltmarshes:44 1st Qu.: 0.4 Class :character Class :character
Median : 8.000 seagrasses :20 Median : 0.7 Mode  :character Mode  :character
Mean   : 6.952                   Mean   : 70215.0
3rd Qu.:10.000                  3rd Qu.: 25.8
Max.   :11.000                  Max.   :2513980.0
```

2. Combine all the data layers into one formatted data file. Select only the columns we need with `select`, change the row format to columns with `spread`, and change the column names to something easier to use with `rename`.

```
rk = D %>%
  select(region_id = id_num,
         layer,
         habitat = category,
         val_num) %>%
  spread(layer, val_num) %>% # spread is a tidyr function
  rename(contribution = cs_contribution, # rename is a dplyr function
         condition     = cs_condition,
         extent        = cs_extent,
         extent_trend = cs_extent_trend); head(rk)
```

This is what `head(rk)` looks like:

```
region_id      habitat contribution condition  extent extent_trend
1          1 saltmarshes    1.0      0.8 1188600      -0.1
1          1 seagrasses     0.5      0.8     100       0.0
2          2 saltmarshes    1.0      0.8  81551       -0.1
...
```

Note: the `%>%` is a chain operator from `dplyr` used to simplify coding writing. To read more about it: <http://cran.rstudio.com/web/packages/dplyr/vignettes/introduction.html> on chaining

3. Select only the habitats that contribute to CS (*Not all habitats included in the raw data files are used for carbon storage*). You can select specific rows with `filter`.

```
rk = rk %>%
  filter(habitat %in% c('mangroves', 'saltmarshes', 'seagrasses'))
```

6.5.2.2 Status Calculation for easy reference, you can write down the equation as a comment before calculations.

```
## status model calculations
# xCS = sum(ck          * Cc/Cr      * Ak) / At
#      = sum(contribution * condition * extent_per_habitat) / total_extent_all_habitats
```

1. Calculate status for all reported years. Most frequently used functions are *mutate*, *group_by*, and *summarize*. To learn more, see Appendix 5.

```
StatusData = rk %>%
  mutate(c_c_a = contribution * condition * extent) %>% # mutate adds a new column
  group_by(region_id) %>%                                # signifies the following calculations are done within each region
  summarize(sum_c_c_a = sum(c_c_a),                         # summarize also adds a new column, but gives one aggregated value per region
           total_extent = sum(extent)) %>% # for each region
  ungroup() %>%                                         # always a good practice to ungroup before next operation
  mutate(xCS_calc = sum_c_c_a/total_extent,
         score = pmax(-1, pmin(1, xCS_calc)) * 100)      #score can't exceed 100
```

2. Select only the status of the most recent year, and add a column for dimension “status”. For final reporting, the toolbox will need four pieces of information: goal, region_id, dimension, and score, although they don’t need to be listed in a certain order at this step.

```
status <- StatusData %>%
  filter(year==status_year) %>%
  mutate(score      = round(Status*100),                      # score is 0-100
         dimension = 'status') %>%
  select(region_id=rgn_id, dimension, score) %>%           # select the correct columns
  data.frame()
```

6.5.2.3 Trend Calculation For CS, a variable `extent-trend` has been prepared to calculate the trend:

```
trend = rk %>%
  group_by(region_id) %>%
  summarize(trend_raw = sum(extent * extent_trend) / sum(extent),
            score = max(min(trend_raw, 1), -1)) %>%
  mutate(dimension = "trend")
```

However, for most other goals, trends are calculated in a regression model based on the most recent 5 years of status:

```
trend = StatusData %>%
  filter(year > (max(year)-5)) %>%                      # select the most recent 5 years of data
  group_by(rgn_id) %>%
  do(mdl = lm(Status ~ year, data = .)) %>%            # regression model
  summarize(region_id = rgn_id,
           score = coef(mdl)[['year']] * 5) %>%          # trend is the coefficient of year x 5
  ungroup() %>%
  mutate(score = round(score, 2),
         dimension = "trend") %>%
  select(region_id, dimension, score) %>%
  data.frame()
```

6.5.2.4 Combine Status and Trend Scores To report the results, you'll assemble status and trend scores you just calculated above into one data frame. Now the your results would contain region_id, score, dimension, and goal, which will be combined with the results of other goals and produce one results table:

```
scores = rbind(status, trend) %>% mutate(goal='CS')
```

6.5.2.5 Update goal call in goals.csv goals.csv in the conf folder provides input information for functions.R, particularly about function calls. 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).

In the *preindex_fuction*, you could specify variables such as *status_year* and *trend_year*, which you can call in your goal function. Note that it is not necessary to specify those variables. If you do not use them in your function as in the CS example, you could delete those variables in *preindex_fuction*.

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 proportion of the number of goals assessed. The ten goals are weighted equally by default.

goals.csv also includes descriptions about goals and sub-goals, which is presented on the WebApp.

The screenshot shows two versions of the goals.csv file in RStudio. The top version is the original CSV with 18 observations and 11 variables. It includes a 'description' column with detailed text for each goal. A red annotation 'Check the information in goals.csv' is overlaid on this part. The bottom version is a modified version of the same CSV, where the 'description' column has been removed, leaving only the 'weight' and 'preindex_function' columns. A red arrow points from the 'description' column in the top table to the 'preindex_function' column in the bottom table, indicating that the information in the descriptions is mapped to the function calls in the modified CSV.

| order_color | order_hierarchy | order_calculate | goal | parent | name | name_flower | description |
|-------------|-----------------|-----------------|------|--------|---------------------------------|---------------------------------|---|
| 1 | 1.2 | 1.0 | 15 | FP | Food Provision | Food Provision | This goal measures the amount of seafood sustainably harvested in a given region th |
| 2 | 1.1 | 1.1 | 1 | FIS | Fisheries | Fisheries | This subgoal model aims to assess the amount of wild-caught seafood that can be sus |
| 3 | 1.3 | 1.2 | 2 | MAR | Marculture | Marculture | This subgoal measures the ability to obtain maximal seafood yield from farm-raised |
| 4 | 2.0 | 2.0 | 3 | AO | Artisanal Fishing Opportunity | Artisanal Fishing Opportunities | This goal captures the access people have to coastal resources, whether or not they |
| 5 | 3.0 | 3.0 | 4 | NP | Natural Products | Natural Products | This goal model calculates overall status by weighting the status of sustainable ha |
| 6 | 4.0 | 4.0 | 5 | CS | Carbon Storage | Carbon Storage | This goal captures the carbon stored in natural coastal ecosystems that absorb and |
| 7 | 5.0 | 5.0 | 6 | CP | Coastal Protection | C | tain coastal habitats provide by mea |
| 8 | 6.0 | 6.0 | 7 | TR | Tourism & Recreation | T | xperiencing and taking pleasure in co |
| 9 | 7.2 | 7.0 | 16 | LE | Coastal Livelihoods & Economics | C | spended livelihoods and productive c |
| 10 | 7.1 | 7.1 | 8 | LIV | Livelihoods | L | hin from the oceans in the form of jo |
| 11 | 7.3 | 7.2 | 9 | ECO | Economics | E | ated with marine industries. |
| 12 | 8.2 | 8.0 | 17 | SP | Sense of Place | S | for aspects of the coastal and ma |
| 13 | 8.1 | 8.1 | 18 | ICO | Iconic Species | Iconic\nSpecies | is subgoal captures the species that are important to a region because they are r |
| 14 | 8.3 | 8.2 | 11 | LSP | Lasting Special Places | Lasting Special\nPlaces | is subgoal captures the conservation status of geographic locations that hold sig |
| 15 | 9.0 | 9.0 | 12 | CW | Clean Waters | Clean Waters | nificance. |
| 16 | 10.2 | 10.0 | 18 | BD | Biodiversity | Biodiversity | This goal captures the degree to which waters are polluted. |
| 17 | 10.1 | 10.1 | 13 | HAB | Habitats | Habitats | This goal captures the conservation status of marine species. |
| 18 | 10.3 | 10.2 | 14 | SPP | Species | Species | This subgoal measures the condition of habitats that are important for supporting a |

| weight | preindex_function | postindex_function |
|--------|--|--------------------|
| 1.0 | FIS(layers, status_year=2011) | FP(layers, scores) |
| 0.5 | MAR(layers, status_years=2007:2012) | |
| 0.5 | AO(layers, year_max=2013) | |
| 1.0 | NP(scores, layers, year_max=2011) | |
| 1.0 | CS(layers) | |
| 1.0 | CP(layers) | |
| 1.0 | TR(layers, year_max=2012) | |
| 1.0 | LE(scores, layers) | |
| 0.5 | LIV_ECO(layers, subgoal='LIV') | |
| 0.5 | LIV_ECO(layers, subgoal='ECO') | |
| 1.0 | ICO(layers) | SP(scores) |
| 0.5 | LSP(layers, status_year=2013, trend_years=2008:2013) | |
| 1.0 | CW(layers) | BD(scores) |
| 1.0 | HAB(layers) | |
| 0.5 | SPP(layers) | |

Figure 36: Check the information in goals.csv. It provides input information for functions.R.

Check goals.csv
when changing goal or
sub-goal models

| | B | C | D | E | F | G | H | I | J | K |
|----|--------------|---------------------|----|----------------|----------------|----------------|---------------|--------|--|--------------------|
| 1 | order_hierar | order_calculat_goal | | parent | name | name_flowe | description | weight | preindex_function | postindex_function |
| 2 | 1 | 15 FP | | FP | Food Provisi | Food Provisi | This goal me | 1 | | FP(layers, scores) |
| 3 | 1.1 | 1 FIS | | FP | Fisheries | Fisheries | This subgoal | 0.5 | FIS(layers, status_year=2011) | |
| 4 | 1.2 | 2 MAR | | FP | Mariculture | Mariculture | This subgoal | 0.5 | MAR(layers, status_years=2007:2012) | |
| 5 | 2 | 3 AO | | Artisanal Fish | Artisan | Artisan | This goal cap | 1 | AO(layers, year_max=2013) | |
| 6 | 3 | 4 NP | | Natural Prod | Natural Prod | Natural Prod | This goal mo | 1 | NP(layers) | |
| 7 | 4 | 5 CS | | Carbon Stora | Carbon Stora | Carbon Stora | This goal cap | 1 | CS(layers) | |
| 8 | 5 | 6 CP | | Coastal Prot | Coastal Prot | Coastal Prot | This goal me | 1 | CP(layers) | |
| 9 | 6 | 7 TR | | Tourism & Rec | Tourism & Rec | Tourism & Rec | This goal cap | 1 | TR(layers, year_max=2012) | |
| 10 | 7 | 16 LE | | Coastal Liveli | Coastal Liveli | Coastal Liveli | This goal aim | 1 | | |
| 11 | 7.1 | 8 LIV | LE | Livelihoods | Livelihoods | Livelihoods | This subgoal | 0.5 | LIV_ECO(layers, subgoal=LIV') | |
| 12 | 7.2 | 9 ECO | LE | Economies | Economies | Economies | This subgoal | 0.5 | LIV_ECO(layers, subgoal=ECO') | |
| 13 | 8 | 17 SP | | Sense of Plac | Sense of Plac | Sense of Plac | This goal aim | 1 | | SP(scores) |
| 14 | 8.1 | 10 ICO | SP | Iconic Specie | Iconic Specie | Iconic Specie | This subgoal | 0.5 | ICO(layers) | |
| 15 | 8.2 | 11 LSP | SP | Lasting Speci | Lasting Speci | Lasting Speci | This subgoal | 0.5 | LSP(layers, status_year=2013, trend_years=2008:2013) | |
| 16 | 9 | 12 CW | | Clean Water | Clean Water | Clean Water | This goal cap | 1 | CW(layers) | |
| 17 | 10 | 18 BD | | Biodiversity | Biodiversity | Biodiversity | This goal cap | 1 | | BD(scores) |
| 18 | 10.1 | 13 HAB | BD | Habitats | Habitats | Habitats | This subgoal | 0.5 | HAB(layers) | |
| 19 | 10.2 | 14 SPP | BD | Species | Species | Species | This subgoal | 0.5 | SPP(layers) | |
| 20 | | | | | | | | | | |
| 21 | | | | | | | | | | |
| 22 | | | | | | | | | | |
| 23 | | | | | | | | | | |

Figure 37: A screenshot of goals.csv, used to modify goal model

6.6 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. Failing to delete all referenced layers after the goal is deleted will result in errors. To remove goals from your assessment, you will have to do the following:

1. Remove the goal model from functions.R
2. Remove the goal's row from goals.csv
3. Remove the goal's row from pressures_matrix.csv
4. Remove the goal's row from resilience_matrix.csv

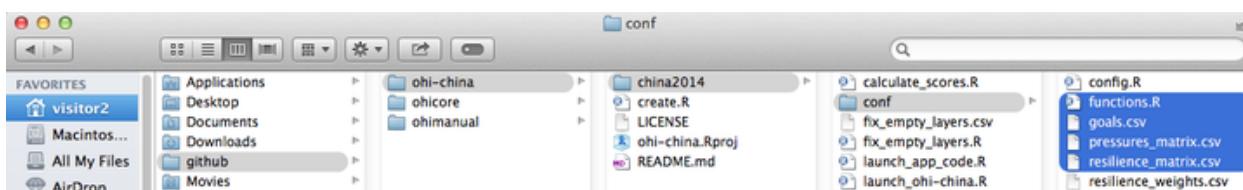


Figure 38:

Example: Removing carbon storage (CS) goal

To completely remove the carbon storage goal from Index calculations, you will do the following.

- 1) Remove the carbon storage (CS) goal model from functions.R. Delete the highlighted text in the figure below that references the CS layers and calculates CS goal status, trend, and scores.
- 2) Remove the CS row from goals.csv. Delete the highlighted row in the figure below that contains the CS goal.

```

627 return(scores_NP)
628 }
629
630 CS = function(layers){
631   # layers
632   lyr_rk = list("rk" = c("rb_health" = "health",
633                 "rb_extent" = "extent",
634                 "rb_trend" = "trend"))
635   lyr_rk$names = sub("^(NP|CP)", "", names(unlist(lyr_rk)))
636
637   # count area
638   D = SelectLayersData(layers, layers=lyr_rk$names)
639   rk = renames(dossID, (d.num = category ~ layer, value.var='val_num', subset = .(layer %in% names(lyr_rk[[ "rk"]]))),
640             c("dnum"="region_id", "category"="habitat", "lyr_rk[[ "rk"]])))
641
642   # limit to CS habitats
643   rk = subset(rk, habitat %in% c('mangrove', 'saltmarsh', 'seagrass'))
644
645   # assign extent of 0 as NA
646   rk$extent[rk$extent==0] = NA
647
648   # status
649   r.status = dplyr::na.omit(rk[,c("region_id", "habitat", "extent", "health")], .byregion_id, summarize,
650                           goal = "CS",
651                           dimension = "status",
652                           score = min(C, sum(extent * health) / sum(extent)) * 100)
653
654   # trend
655   r.trend = dplyr::na.omit(rk[,c("region_id", "habitat", "extent", "trend")], .byregion_id, summarize,
656                           goal = "CS",
657                           dimension = "trend",
658                           score = sum(extent * trend) / sum(extent))
659
660   # return scores
661   r.scores = bind(bind(r.status, r.trend))
662   return_scores
663 }
664
665
666
667 CP = function(layers){
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2177
2178
2179
2180
2181
2182
2183
2184
2185
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2287
2288
2289
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2298
2299
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2387
2388
2389
2389
2390
2391
2392
2393
2394
2395
2396
2397
2397
2398
2399
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2487
2488
2489
2489
2490
2491
2492
2493
2494
2495
2496
2496
2497
2498
2499
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2548
2549
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2559
2559
2560
2561
2562
2563
2564
2565
2566
2567
2568
2569
2569
2570
2571
2572
2573
2574
2575
2576
2577
2578
2578
2579
2580
2581
2582
2583
2584
2585
2586
2587
2587
2588
2589
2589
2590
2591
2592
2593
2594
2595
2596
2597
2597
2598
2599
2599
2600
2601
2602
2603
2604
2605
2606
2607
2608
2609
2609
2610
2611
2612
2613
2614
2615
2616
2617
2618
2619
2619
2620
2621
2622
2623
2624
2625
2626
2627
2628
2629
2629
2630
2631
2632
2633
2634
2635
2636
2637
2638
2638
2639
2639
2640
2641
2642
2643
2644
2645
2646
2647
2648
2648
2649
2649
2650
2651
2652
2653
2654
2655
2656
2657
2658
2659
2659
2660
2661
2662
2663
2664
2665
2666
2667
2668
2669
2669
2670
2671
2672
2673
2674
2675
2676
2677
2678
2678
2679
2680
2681
2682
2683
2684
2685
2686
2687
2687
2688
2689
2689
2690
2691
2692
2693
2694
2695
2696
2697
2697
2698
2699
2699
2700
2701
2702
2703
2704
2705
2706
2707
2708
2709
2709
2710
2711
2712
2713
2714
2715
2716
2717
2718
2719
2719
2720
2721
2722
2723
2724
2725
2726
2727
2728
2729
2729
2730
2731
2732
2733
2734
2735
2736
2737
2738
2738
2739
2739
2740
2741
2742
2743
2744
2745
2746
2747
2748
2748
2749
2749
2750
2751
2752
2753
2754
2755
2756
2757
2758
2759
2759
2760
2761
2762
2763
2764
2765
2766
2767
2768
2769
2769
2770
2771
2772
2773
2774
2775
2776
2777
2778
2779
2779
2780
2781
2782
2783
2784
2785
2786
2787
2787
2788
2789
2789
2790
2791
2792
2793
2794
2795
2796
2797
2797
2798
2799
2799
2800
2801
2802
2803
2804
2805
28
```

| | 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 | fishing_v2_eez | fishing_v3_ee_habitat | | | | habitat_combo | habitat_combo_eez | i |
| 2 | SPP | | alien_species_cites | | | | | | | | habitat_combo | habitat_combo_eez | |
| 8 | CS | | | | | | | | | | habitat_combo | habitat_combo_eez | |
| 9 | CP | | mangrove only | | | | | | | | habitat_combo | habitat_combo_eez | |
| 10 | CW | | | | | | | | | | | | |
| 11 | FIS | | | | | | fishing_v2_eez | | | | | habitat_combo_eez | |

Figure 42:

pressures and resilience scores, the Toolbox uses `config.r` to identify which categories to expect in the matrix tables, and will give a warning if they do not match. `config.r` relies upon the data layers identified in the `pressures_components` and `resilience_components` variables.

In global assessments, there are several goals that have categories indicated in the matrix tables and `config.r` file:

| Goal | Category | layer indicated in <code>config.r</code> |
|------|------------------|--|
| NP | product types | <code>np_harvest_product_weight</code> |
| CS | habitat types | <code>cs_habitat_extent</code> |
| CP | habitat types | <code>cp_habitat_extent_rank</code> |
| HAB | habitat types | <code>hab_presence</code> |
| LIV | industry sectors | <code>le_sector_weight</code> |
| ECO | industry sectors | <code>le_sector_weight</code> |

If you have modified any of the category types in the matrix tables of the above goals, or added new category types to any goals, you will likely need to update the layer indicated in `config.r`. It is also possible to identify individual categories in other goals than those listed above. For example, in the mariculture sub-goal, you could specify the pressures on nearshore mariculture separately from offshore mariculture.

It is important that the file identified in `config.r` does not contain any NA values.

6.7.2 Example 1: Pressures

Here is an example of how to modify existing category types for the natural products goal.

In the China OHI+ assessment there are three natural product types (seasalt, sea chemicals, and sea medicine), which differ from those assessed in the global assessments (corals, fish_oil, ornamentals, seaweeds, sponges). After modifying and registering the appropriate data layers and updating the NP function in `functions.r`, it is time to update the natural product types in `pressures_matrix.csv`, `resilience_matrix.csv`, and `config.r`.

- to update `pressures_matrix.csv` and `resilience_matrix.csv`, make sure that each product type has a separate row, with the appropriate pressures identified and weights attributed.
- to update `config.r`, check that the data layer identified in the `pressures_components` and `resilience_components` has the same category types.

When you run `calculate_scores.r`, the following warning will alert you that there is a mis-match between category types identified in the matrix and `config.r`:

Calculating Pressures...

The following components for NP are not in the aggregation layer `np_harvest_product_weight` categories
(corals, fish_oil, ornamentals, seaweeds, sponges):
seasalt, sea_chemicals, sea_medicine

This message indicates that the `np_harvest_product_weight` layer identifies five categories (corals, fish_oil, ornamentals, seaweeds, sponges) but the `pressures_matrix.csv` indicates three (seasalt, sea_chemicals, sea_medicine).

To ensure that pressures are calculated correctly for the categories in your assessment, you will need to change the layer identified in `config.r`.

Note that more subtle examples of these mismatch between the categories identified in `pressures_matrix.csv` and `config.r` can also occur. For example, after updating the carbon storage layers and goal model in the China OHI+ assessment, the following warning message appeared when running `calculate_scores.r`:

```
Calculating Pressures...
The following components for CS are not in the aggregation layer
cs_extent categories (saltmarshes, seagrasses, mangroves):
mangrove, saltmarsh, seagrass
```

The problem here is that the categories identified in `config.r` (saltmarshes, seagrasses, mangroves) are plural, whereas the categories identified in the pressures matrix (mangrove, saltmarsh, seagrass) are singular, and the Toolbox needs exact matches. To fix this warning, you need to update the pressures matrix with the plural names.

6.7.3 Example 2: Resilience

For resilience calculations, the proper categories also need to be identified both in `resilience_matrix.csv` and `config.r`. If there is a mismatch, you will see the following message:

```
Calculating Resilience...
Note: each goal in resilience_matrix.csv
must have at least one resilience field
Based on the following components for NP:
corals
fish_oil
ornamentals
seaweeds
shells
sponges
```

With resilience, if we update only the `resilience_matrix.csv` but not `config.r`, we get the following error message instead of the warning message we saw for pressures above.

```
Based on the following components for NP:
seasalt
sea_chemicals
sea_medicine
Error in subset.default(SelectLayersData(layers, layers = lyr),
id_num == : object 'id_num' not found
In addition: Warning messages:
1: Grouping rowwise data frame strips rowwise nature
2: In left_join_impl(x, y, by$x, by$y) :
joining factors with different levels, coercing to character vector
```

This error can be fixed by updating `config.r` with a layer identifying the appropriate categories.

6.8 Calculate overall OHI Index Scores

Congratulations if you've finished all goal model modifications as they are the most time-consuming part of the Index calculation process! In the goal model modification step, you've calculated goal status and trend, there are only a couple of more steps to do to calculate the overall index score, including calculating pressures and resilience.

6.8.1 Additional data layers for pressures and resilience calculations

So far you would have prepared the pressure and resilience matrices, as well as the appropriate data layers, a few more data layers are needed to complete the calculation. They are identified in `sub-country/conf/config.R` as follows:

```
# components describe the layer and level with which to aggregate resilience and pressures matrices for
resilience_components = list('NP' = c('layer'='np_harvest_product_weight' , 'level'='region_id-category'),
                             'CS' = c('layer'='cs_habitat_extent' , 'level'='region_id'),
                             'CP' = c('layer'='cp_habitat_extent_rank' , 'level'='region_id'),
                             'HAB' = c('layer'='hab_presence' , 'level'='region_id'))
pressures_components = list('NP' = c('layer'='np_harvest_product_weight' , 'level'='region_id-category'),
                            'CS' = c('layer'='cs_habitat_extent' , 'level'='region_id'),
                            'CP' = c('layer'='cp_habitat_extent_rank' , 'level'='region_id'),
                            'LIV' = c('layer'='le_sector_weight' , 'level'='region_id'),
                            'ECO' = c('layer'='le_sector_weight' , 'level'='region_id'),
                            'HAB' = c('layer'='hab_presence' , 'level'='region_id'))
```

`np_harvst_product_weight` is also used in NP and CS status calculations, and thus do not require special preparations. The rest of the data layers need additional preparations, which can be done in the `prep` folder).

`cs_habitat_extent` is calculated as `habitat_extent * rank`. Rank refers to relative contributions of each type of habitats to carbon storage. Here is an example calculation:

```
# cs_habitat_extent = Habitat extent * rank, per Carbon Storage habitat
#           = extent * contribution

extent = layers$data[['cs_extent']] %>%
  select(rgn_id, habitat, extent = hectare)

contribution = layers$data[['cs_contribution']] %>%
  select(rgn_id, habitat, contribution = value)

result = full_join(extent, contribution, by = c('rgn_id', 'habitat')) %>%
  mutate(hectare = extent*contribution) %>%
  select(rgn_id, habitat, hectare)
```

`cp_habitat_extent_rank` is calculated as `habitat_extent * weight`. Weight refers to relative contributions of each type of habitats to coastal protection. See this example:

```
# cp_habitat_extent_rank = Habitat extent * rank, per Coastal Protection habitat
#           = extent * weight

habitat.wt = c('saltmarshes' = 3,
              'mangroves' = 4,
              'seagrasses' = 1,
              'coral reef' = 4)

m = layers$data[['cp_extent']] %>%
  group_by(rgn_id, habitat) %>%
  filter(year == max(year)) %>% #choose the most recent year's data
  select(-layer,
         -year,
```

```

    extent = hectare) %>%
mutate(weight = habitat.wt[habitat],
       extent_rank = extent * weight) %>%
select(rgn_id, habitat, extent_rank)

```

`hab_presence` is calculated on a binary basis. All regions with a habitat is assigned “1”. Example:

```

# hab_presence: 1 for presence

m = layers$data[['hab_extent']] %>%
  group_by(rgn_id, habitat) %>%
  filter(year == max(year)) %>% #choose the most recent year's data
  select(-layer,
         -year,
         extent = hectare) %>%
  mutate(boolean = 1) %>%
  select(rgn_id, habitat, boolean)

```

`le_sector_weight` assigns relative importance of each sector listed in Livelihood & Economies goal. By default, all sectors are considered equal and assigned a weight of “1”. One thing to pay special attention to is the `sector` names, which should match `component` names that were listed in the pressures matrix.

6.8.2 Final calculation

Go back to `calculate_scores.r`, now you are ready for the final calculations:

```

# calculate scenario scores
scores = CalculateAll(conf, layers, debug=F)

# save scores as .csv file, tables and figures
write.csv(scores, 'scores.csv', na='', row.names=F)

```

After the calculation is done, you should be able to see the compiled score sheet for all goals in all regions in `sub-country/scores.csv`.

It is very likely that during the `CalculateAll` process you'll encounter problems and see error messages. In most cases, the error messages can specify what the error is and in which step it occurs, which should be helpful for trouble shooting. Some commonly occurring errors and how to fix them can be found in the Troubleshooting section of the manual.

6.9 Other example modifications

6.9.1 Preparing the fisheries sub-goal

Here is some background information about how to prepare fisheries data layers for the Toolbox.

Data layers used by the Toolbox:

- `fis_b_bmsy`
- `fis_meanCatch`
- `fis_proparea_saup2rgn`
- `fp_wildcaught_weight`

6.9.1.1 Description of data layers `fis_b_bmsy`

- *for species*: B/Bmsy estimate (either from formal stock assessment, or from a data-poor method such as CMSY)
- *for genus/family/broader taxa*: the toolbox will use median B/Bmsy from species in that region + a penalty for not reporting at species level. In order for the code to assign the correct penalty, the taxa need to include a numerical code of 6 digits, where the first digit behaves like an ISSCAAP code (the standardized species codes used by FAO): 6 means species, 5 means genus, 4 to 1 are increasingly broad taxonomic groups
- *data source (for CMSY)*: catch time-series (at least 10 years of catch >0), species resilience (if available)

Example data:

| <code>fao_id</code> | <code>taxon_name</code> | <code>year</code> | <code>b_bmsy</code> |
|---------------------|-------------------------|-------------------|---------------------|
| 51 | Ablennes hians | 1985 | 1.112412 |
| 51 | Ablennes hians | 1986 | 1.222996 |
| 51 | Ablennes hians | 1987 | 1.371058 |

NOTE: if a species that is caught in different sub-regions belongs to the same population, you don't want to split the catch among sub-regions, instead, you want to sum catch across all sub-regions, so you can calculate B/Bmsy for the whole population. For the global analysis we grouped all species catch by FAO major fishing area (www.fao.org/fishery/area/search/en), indicated in the column `fao_id`, assuming that all species caught within the same FAO area belonged to the same stock, while we assumed that the same species, if caught in a different fishing area, belonged to a separate stock.

Use `fao_id` as an identifier that separates different fisheries 'stocks' belonging to the same species. If you don't have multiple stocks in your study area, set all `fao_id = 1`.

fis_meancatch:

- average catch across all years, per species, per region
- *data source*: catch time-series (at least 10 years of catch >0), with a unique identifier for each population that you want to assess separately

Example data:

| <code>fao_saup_id</code> | <code>taxon_name_key</code> | <code>year</code> | <code>mean_catch</code> |
|--------------------------|-----------------------------|-------------------|-------------------------|
| 37_8 | Aristeus antennatus_690051 | 2014 | 14.24398116 |
| 37_8 | Atherinidae_400218 | 2014 | 27.30120156 |
| 37_8 | Balistes capriscus_607327 | 2014 | 3.247883895 |

The `taxon_name_key` column indicates the name of the species (e.g. *Aristeus antennatus*) and its 'taxonkey'. The taxonkey is a 6 digit numeric code used by the Sea Around Us Project, modified from FAO codes. The important element of this code is the first digit, because it reflects the taxonomic level (6=species, 5=genus, 4=family, etc.) of the reported catch. The toolbox uses this first digit to assign a score to all catch that was not reported at species level, taking the median of the B/Bmsy of assessed species, and adding a penalty that is increasingly strong for coarser taxa.

fis_proparea_saup2rgn:

- a conversion file that, for each region for which catch is reported, tells us what proportion of that region

falls within each of the final OHI reporting regions.

Example data:

| saup_id | rgn_id | prop_area |
|---------|--------|-----------|
| 166 | 1 | 1.0 |
| 162 | 2 | 1.0 |
| 574 | 3 | 0.7 |
| 37 | 4 | 0.8 |

Specific instances:

only if catch is reported for different regions than the ones used for the OHI assessment: this should be calculated using spatial analyses of overlap of the spatial units at which catch is reported with the spatial units at which the OHI assessment will be reported. The global data was reported by subregions (*saup_id*) and in some cases multiple subregions were part of the same, larger EEZ. Since for OHI we wanted results by EEZ (*rgn_id*), in those cases we needed to combine results from the subregions to get the final score, based on their size relative to the total EEZ size (*prop_area*). *If catch is reported for the same areas for which OHI is calculated:* then all the *prop_area* are = 1. *If catch is reported for the whole area of the assessment, but you want to calculate a separate OHI score for different sub-regions:* for each OHI reporting region (*rgn_id*) you'll repeat the same region in the *saup_id* column, and *prop_area* will be =1. This effectively means all the reporting regions will get assigned 100% of the catch and will have the same final stastus and trend score for the fisheries goal (but may have different pressures and resilience scores, if those layers are different in each sub-region).

fp_wildcaught_weight:

only needed if there is mariculture: for each region, this represents the relative proportion of catch coming from wild caught fisheries versus mariculture. The layer is used to weight how much the fisheries score influences the final food provision score, the higher the fisheries catch, the more the food provision score will reflect the fisheries score, and vice-versa if mariculture has a higher catch. (NOTE that, before all mariculture harvest from all species gets summed, the mariculture harvest for each species is smoothed and then multiplied by the resilience score).

6.9.1.2 Running CMSY model Sample data to run CMSY:

| id | stock_id | res | ct | yr |
|----|-----------------------------|--------|-------------|------|
| 6 | Acanthistius brasilianus_41 | Medium | 100 | 1950 |
| 23 | Acanthurus dussumieri_61 | | 0.059250269 | 1950 |
| 24 | Acanthurus dussumieri_71 | | 0.190749971 | 1950 |
| 25 | Acanthurus lineatus_61 | Low | 12.74821966 | 1950 |

The current CMSY script produces an output that looks something like this (split into 2 tables):

| stock_id | convergence | effective_sample_size | yr | b_bmsy | b_bmsyUpper |
|-------------------|-------------|-----------------------|------|----------|-------------|
| Ablennes hians_51 | SC | 30974 | 1985 | 1.112412 | 1.8 |
| Ablennes hians_51 | SC | 30974 | 1986 | 1.222996 | 1.768895 |

| stock_id | yr | b_bmsyLower | b_bmsyi25 | b_bmsyi75 | b_bmsyGM | b_bmsyMed |
|-------------------|------|-------------|-----------|-----------|----------|-----------|
| Ablennes hians_51 | 1985 | 1 | 1 | 1 | 1.093932 | 1 |
| Ablennes hians_51 | 1986 | 1.014688 | 1.075699 | 1.298437 | 1.209005 | 1.160329 |

where *stock_id* is the unique identifier for each stock that was used in the input file, *convergence* indicates whether the model converged and how strongly ('SC' = strong convergence), *effective_sample_size* reports the number of iterations used, *yr* = year, *b_bmsy* = B/Bmsy for the corresponding year (based on the median of all the estimated values: recommended), *b_bmsyUpper* = B/Bmsy at the upper 95% bootstrapped confidence bound, *b_bmsyLower* = B/Bmsy at the lower 95% bootstrapped confidence bound, *b_bmsyi25* = B/Bmsy at the first quartile, *b_bmsyi75* = B/Bmsy at the third quartile, *b_bmsyGM* = B/Bmsy based on the geometric mean of estimates, *b_bmsyMed* = B/Bmsy based on the median of estimates.

How to:

1. Include resilience in the CMSY code:

In the CMSY R script, in the PARAMETERS section, replace the following:

```
start_r      <- c(0.01,10) ## disable this line if you use resilience
with

if(res == "Very low"){
  start_r  <- c(0.015, 0.1)
} else {
  if(res == "Low"){
    start_r  <- c(0.05,0.5)
  } else {
    if(res == "High"){
      start_r  <- c(0.6,1.5)
    } else {
      start_r  <- c(0.1,1)
    }
  }
}
```

Figure 43:

2. Make assumptions about fisheries regulations:

If you assume that fisheries are depleted and there isn't very much fisheries regulation, and you are using the CMSY method to assess B/Bmsy, the original model may work well. If, however, the catch of a species declined because fisheries regulations have closed or limited the fishery, or if a fishery was abandoned for economic reasons (e.g., change in consumer preferences, market price dynamics, etc.), the model may be too pessimistic and underestimate B/Bmsy. In that case it may be best to use a version with a uniform prior on final biomass, instead of the constrained prior.

The original constrained prior on final biomass is set by this line within the code:

```
finalbio      <- if(ct[nyr]/max(ct) > 0.5) {c(0.3,0.7)} else {c(0.01,0.4)}
```

The model uses a uniform prior if that line is replaced with:

```
finalbio      <- c(0.01,0.7)
```

3. Use data at a different spatial resolution than the final assessment:

See notes above for *fis_proparea_saup2rgn*

4. Calculate B, or Bmsy:

The CMSY model calculates B/Bmsy as a ratio, it does not estimate the two variables separately.

5. Use catch per unit of effort (CPUE):

The CMSY model requires total biomass removed by fisheries, and uses catch as a proxy for that. It cannot use CPUE. Other more sophisticated stock assessment models use CPUE and may be employed. We do not provide documentation for the use of these other models.

6. Use other life-history characteristics, in addition to resilience:

The CMSY model does not use more detailed information. Other more sophisticated stock assessment models use other life-history traits such as fecundity, larval dispersal, r, K, Lmax, etc., and may be employed. We do not provide documentation for the use of these other models.

7. Create a ‘taxonkey’ to assign to each species:

When replacing the SAUP_FAO data with your own data, assign a key of 600000 to all species. For all catch that is reported at genus or coarser taxonomic level, you will have to choose an appropriate taxonkey. You can create your own key, from 100000 to 500000, based on your own judgment of how many species may be reported under that same denomination, and how different they may be (all that matters for the toolbox code is whether the number starts with a 1,2,3,4,5 or 6 with 1 being the coarsest, such as ‘miscellaneous marine animals’, or ‘crustaceans nei’).

6.9.1.3 Resources Martell, S & Froese, R (2013) “A simple method for estimating MSY from catch and resilience”. *Fish and Fisheries*, DOI: 10.1111/j.1467-2979.2012.00485.x. [Downloadable here](#)

Rosenberg, A.A., Fogarty, M.J., Cooper, A.B., Dickey-Collas, M., Fulton, E.A., Gutiérrez, N.L., Hyde, K.J.W., Kleisner, K.M., Kristiansen, T., Longo, C., Minte-Vera, C., Minto, C., Mosqueira, I., Chato Osio, G., Ovando, D., Selig, E.R., Thorson, J.T. & Ye, Y. (2014) Developing new approaches to global stock status assessment and fishery production potential of the seas. *FAO Fisheries and Aquaculture Circular No. 1086*. Rome, FAO. 175 pp. [Downloadable here](#)

7 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.

7.1 General Software Errors

7.1.1 rpostback-askpass error

Sometimes when RStudio won’t push committed changes to GitHub, RStudio displays an `rpostback-askpass` error:

```
error: unable to read askpass response from 'rpostback-askpass'  
fatal: could not read Username for 'https://github.com': Device not configured
```

Here’s how we fixed it: we updated `git.exe` to the latest version, 2.2.1, edited the search path to point to the new version, made sure the `git credential.helper` was configured to be able to access the OS X keychain, and pushed a test commit from terminal to store the username and password in the keychain, where it can be accessed from other apps like RStudio. Easy peasy!

1. To check your current version of `git.exe`, type this at the terminal command line:

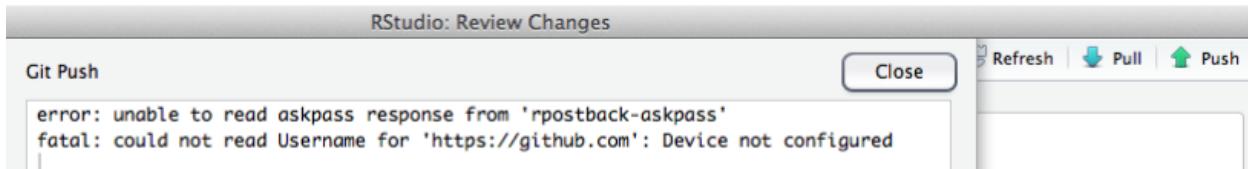
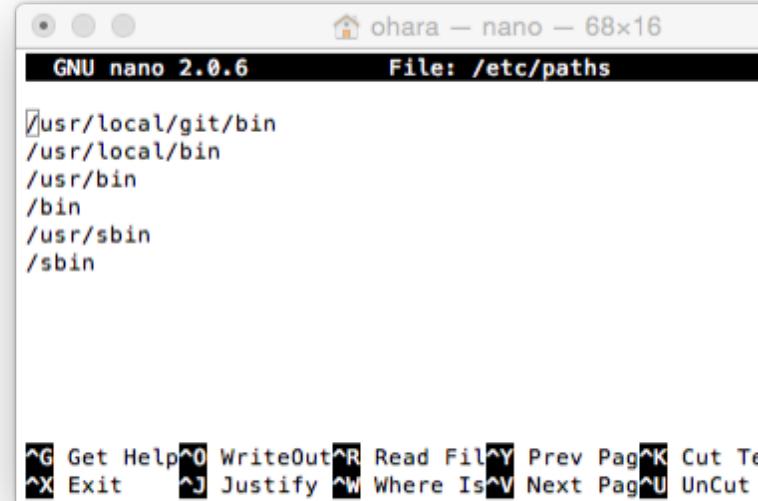


Figure 44: Error screen window: ‘error: unable to read askpass response’.

- \$ git --version should return something like:
 - git version 2.2.1 (check online to see if this is the latest version)
2. To update, go to <http://git-scm.com/download/mac>, download the latest *git* for OS X, install it.
 3. In terminal, type `git --version` and verify that it reports the new version. If it shows the new version, great! Skip to Step 5.
 - Don’t be sad if it doesn’t! If you still see the old version, the installer put the new version into a different directory, which has a lower priority in the search path, so now to update the search path. The default Apple *git* seems to install the *git.exe* into `/usr/bin/` directory, this particular updater seems to install into `/usr/local/git/bin/` directory. The search path needs to be updated to look for *git.exe* in the new directory first.
 4. To change the search path, open up the paths file in `nano` editor using `sudo`:
 - \$ sudo nano /etc/paths
 - At the top line of the paths file, add the directory for the updated *git*: `'/usr/local/git/bin'` (without



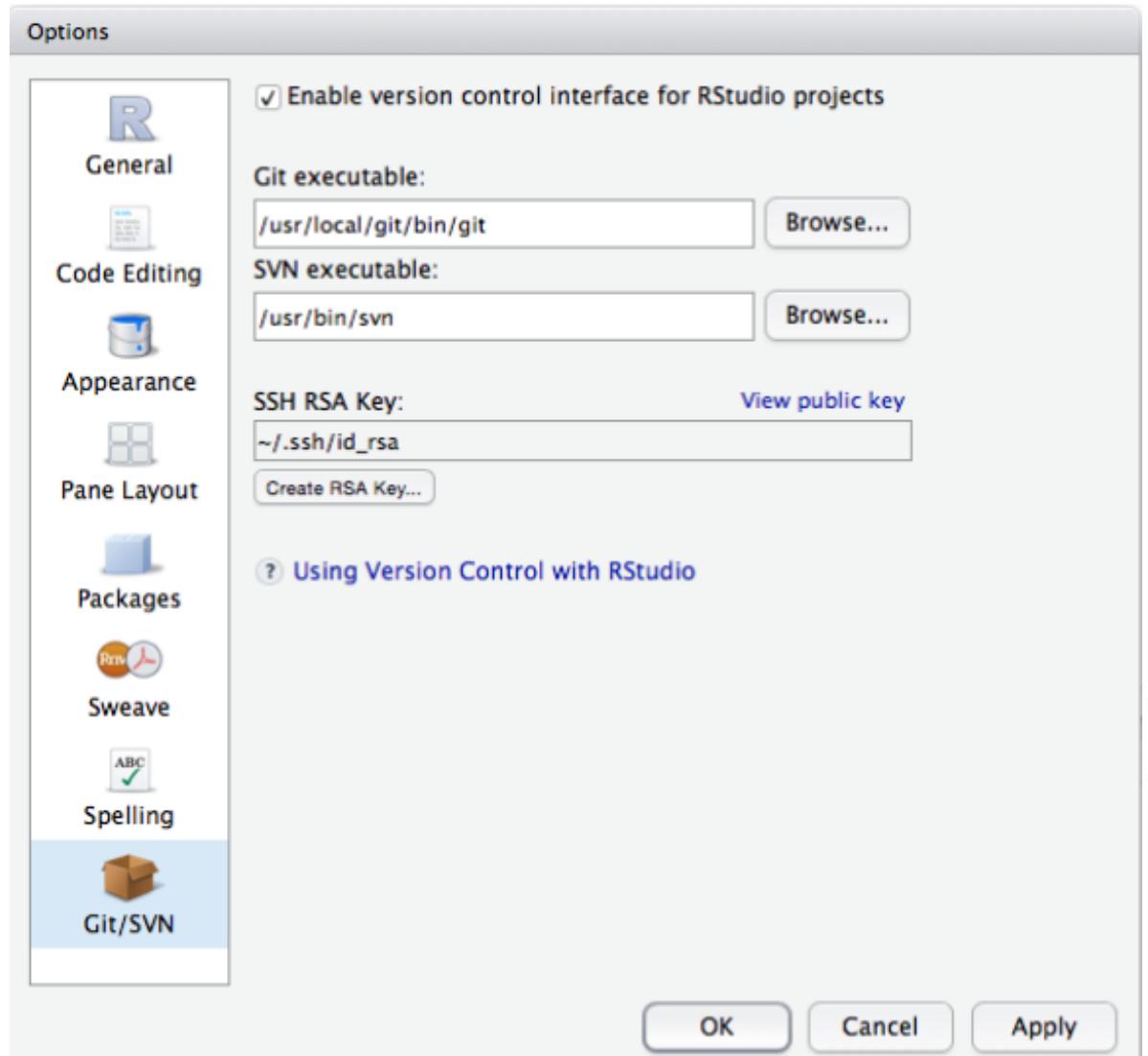
- the quotes) so it looks like the top line here:
- Then hit control-X to exit, then Y in response to the save prompt.
5. Make sure your `git config` is up to date, including `credential.helper`:
 - \$ `git config --global -l` should return something like:
 - `user.name="Casey O'Hara"`
 - `user.email=ohara@nceas.ucsb.edu`
 - see https://github.com/OHI-Science/ohiprep/wiki/Setup#git_identity for help on updating user.name and user.email

- `credential.helper=osxkeychain`
 - (if you need to configure the credential helper: <https://help.github.com/articles/caching-your-github-password-in-git/>)

6. Now while you are in Terminal, it is important to sync with a repository to establish your security credentials. You must clone a repository and push a ‘test’ commit, and then once you are prompted for your username and password your information will get stored in the keychain. Here are the steps:

- Change your working directory to your local github directory: `$ cd github`
- (Tip: you can check if you’re in the right folder by entering `pwd`, short for “print working directory”; or you could look at the line of code preceding the “\$”.)
- Clone into a repository with a URL *for which you have permissions*. As an example, the following steps use a repository called ‘ZAF’ but you should use your own URL with a three-letter country code in place of ‘ZAF’:
 - `$ git clone https://github.com/omalik/zaf.git`
 - Change your working directory to the folder you just created (here, ‘ZAF’): `$ cd zaf`
 - Push a test commit to repository ‘ZAF’:
 - `$ touch test.md`
 - `$ git add test.md`
 - `$ git commit -m "testing"`
 - `$ git status`
 - `$ git push`
 - Check your status again: `$ git status`
 - (TIP: You can check your status with `$ git status` and you can use ‘ls’ to see if your new changes have registered in this repository.)

7. Now that *git* is updated and your username and password are set, make sure RStudio knows the location of the new `git.exe`. In RStudio, select **Tools > Global Options...**, select the **Git/SVN**, and browse to the new **Git executable** (it should appear as `/usr/local/git/bin/git` if you updated your *git* version



as above).

Next time you push a commit from RStudio, it should remember the username and password from your test commit in Step 6, and you should be good to go.

7.1.2 Git not detected on system path

After downloading and installing git, sometimes RStudio is unable to locate git and will show the following error:

It is important to make sure you do in fact have git installed on your computer first. You can do this by opening up terminal (Mac OSX) or the cmd line (Windows) and typing:

- `$ which git` for Mac/Linux,
- `$ where git` for Windows

These commands will tell you where the `git.exe` file is located. Typically it will look something like this: `/usr/local/git/bin/git` or `/usr/bin/git` or some variation of those.

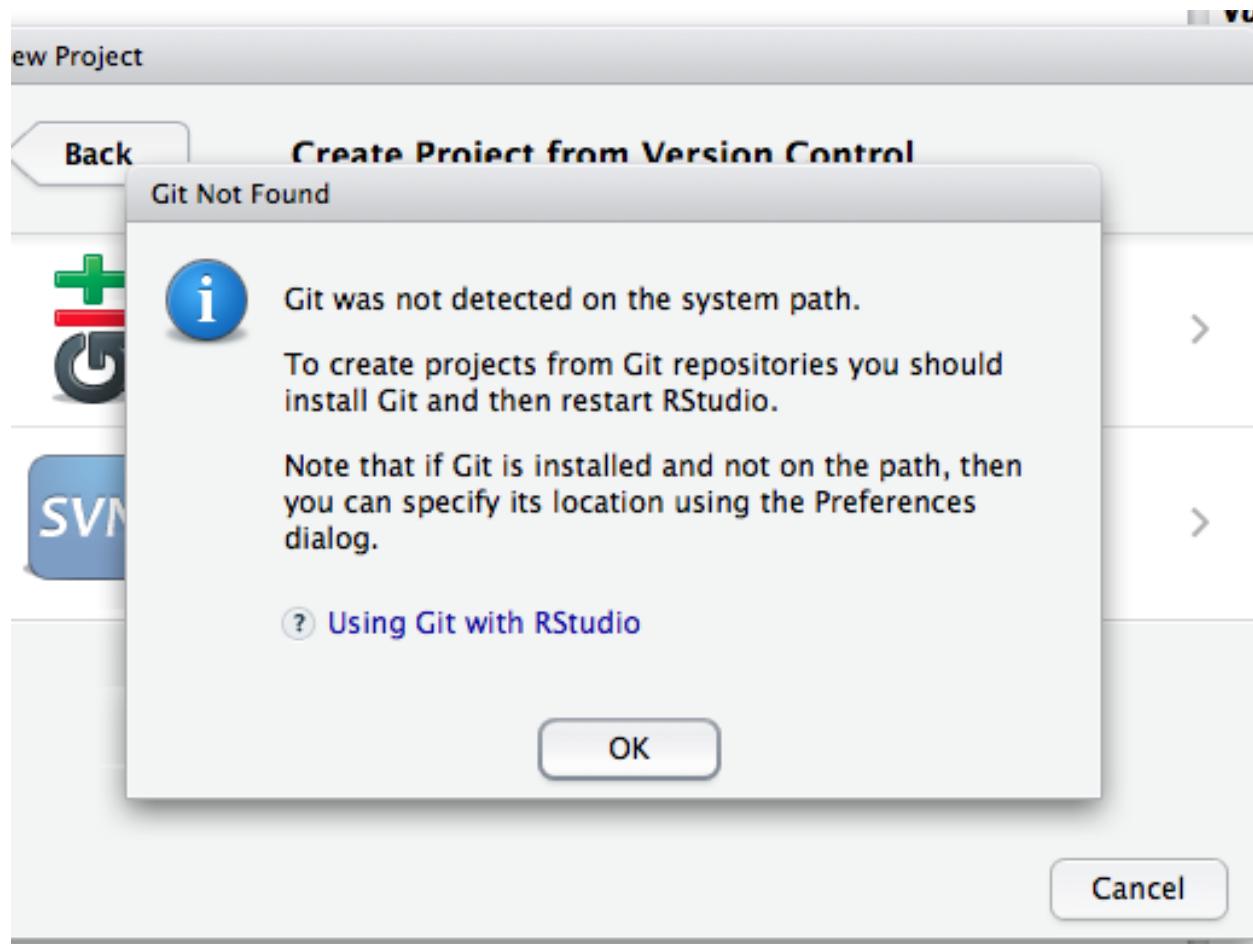


Figure 45: Error screen window:‘Git was not detected on the system path’

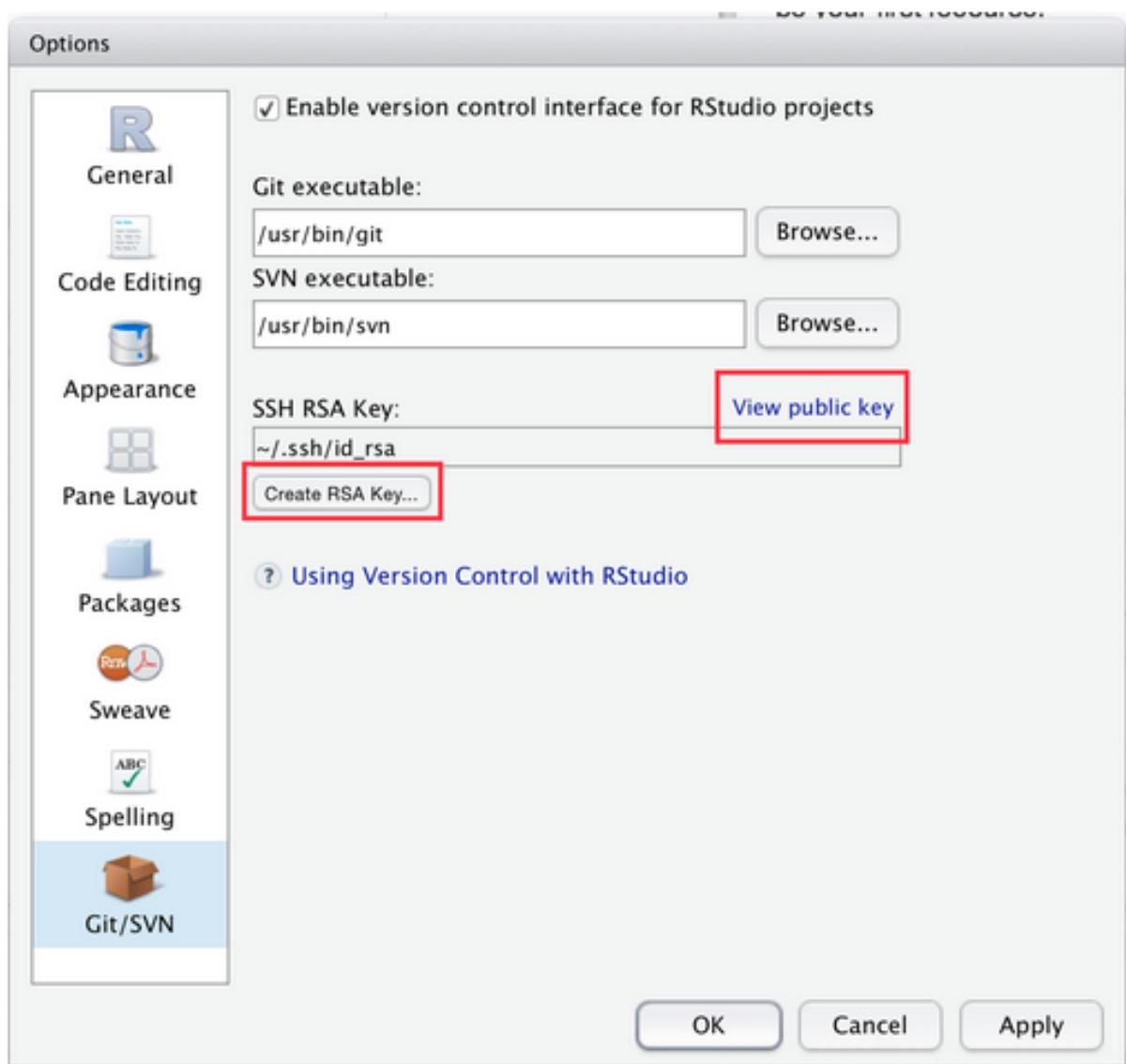


Figure 46:

Once confirming the location of `git.exe` you need to tell RStudio where it is. Open up RStudio, got to Preferences and select the Git/SVN option:

In the Git executable area, fill in the path to your `git.exe`. If RStudio does not let you manually enter your path, select Browse... and navigate to the `git.exe` file. If you are not able to navigate to the file it is likely a hidden file.

On a Mac, to make hidden files visible, close RStudio and do the following:

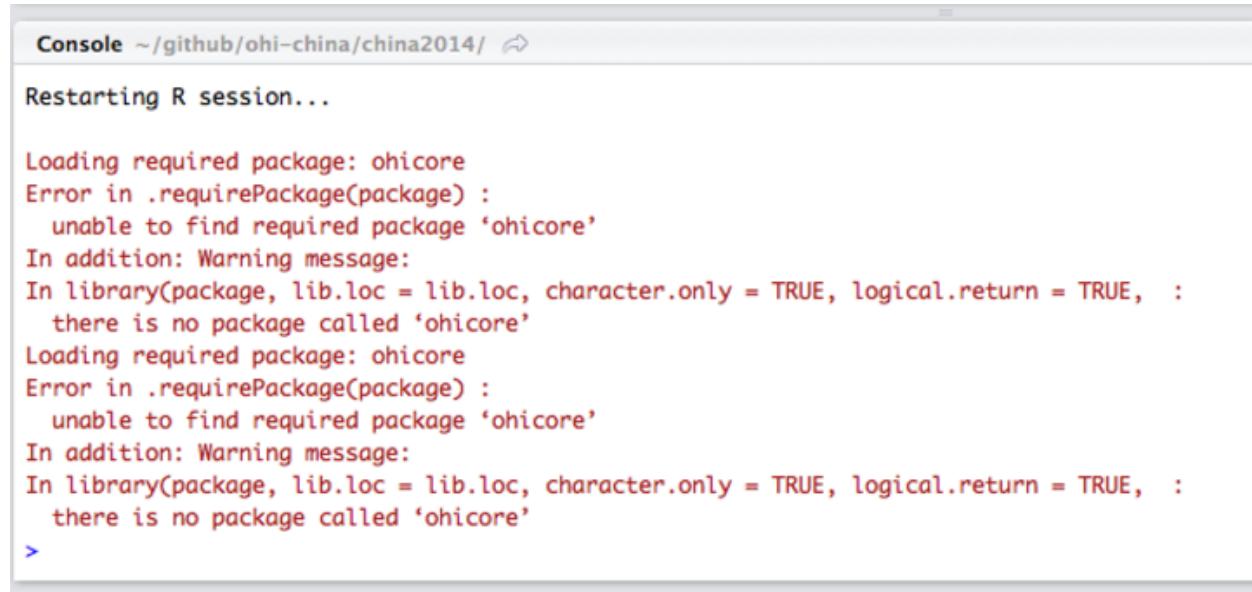
- Open Terminal found in Finder > Applications > Utilities.
- In Terminal, paste the following: defaults write com.apple.finder AppleShowAllFiles YES.
- Press return.
- Hold the ‘Option/alt’ key, then right click on the Finder icon in the dock and click Relaunch.

And then reopen RStudio, go to Preferences -> Git/SVN -> Browse... and you should be able to navigate to the `git.exe`

You will then need to create an RSA Key. You can do this by clicking on ‘Create RSA Key...’ at the bottom of the Git/SVN panel, then ‘View public key’. Copy the key you see, and add it to your GitHub account by using the instructions provided [here](#).

7.1.3 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:



The screenshot shows an RStudio console window with the title bar "Console ~/github/ohi-china/china2014/". The console output is as follows:

```
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'
>
```

Figure 47:

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

1. Go to Project Options, either in the pull-down menu or by double-clicking the `.Rproj` file:

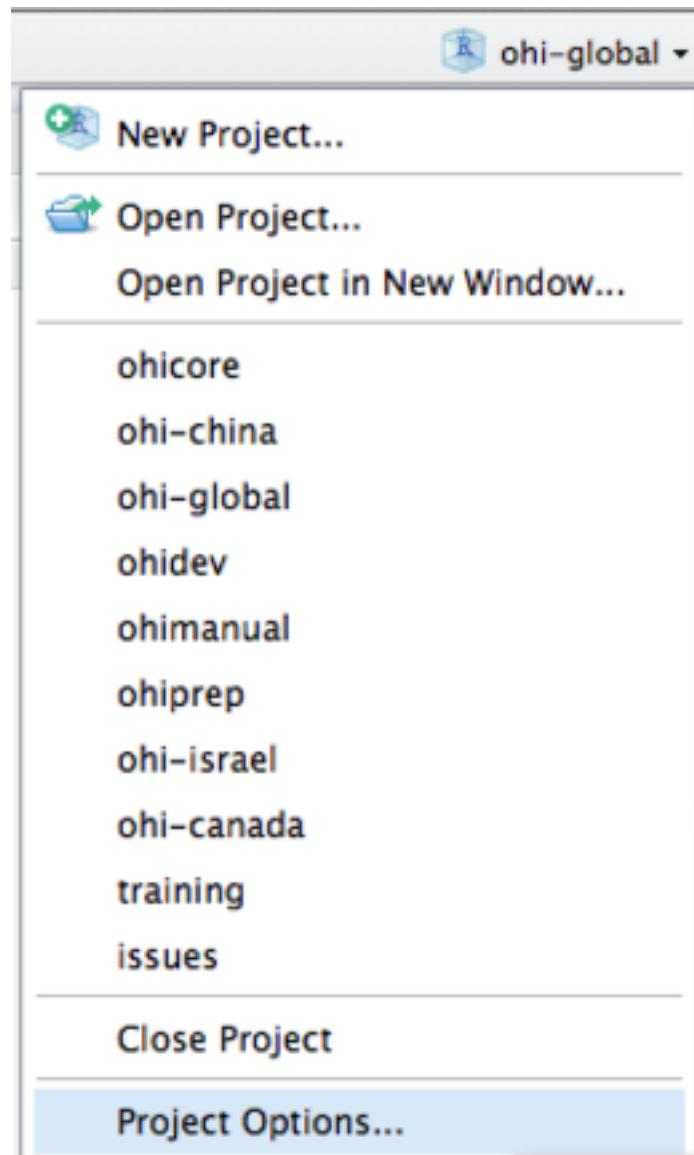


Figure 48:

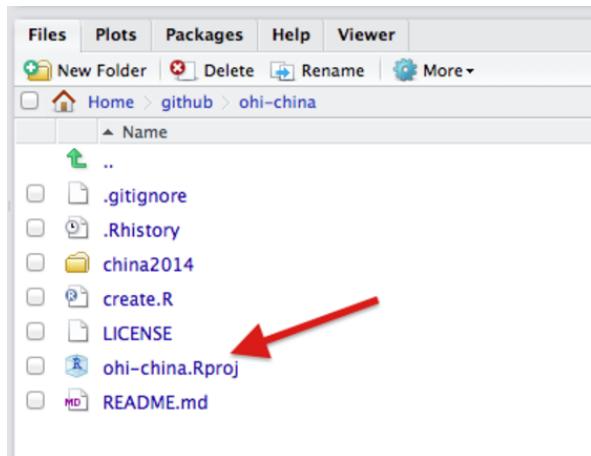


Figure 49:

2. Change all options to **No**:

7.2 Errors when Using the Toolbox

7.2.1 Useful Errors when Calculating Scores

TIP: You can use the `layers` function in `calculate_scores.R` to error-check whether you have registered your files in `layers.csv` correctly or not. If you haven't, you will get an error message regarding 'missing files'.

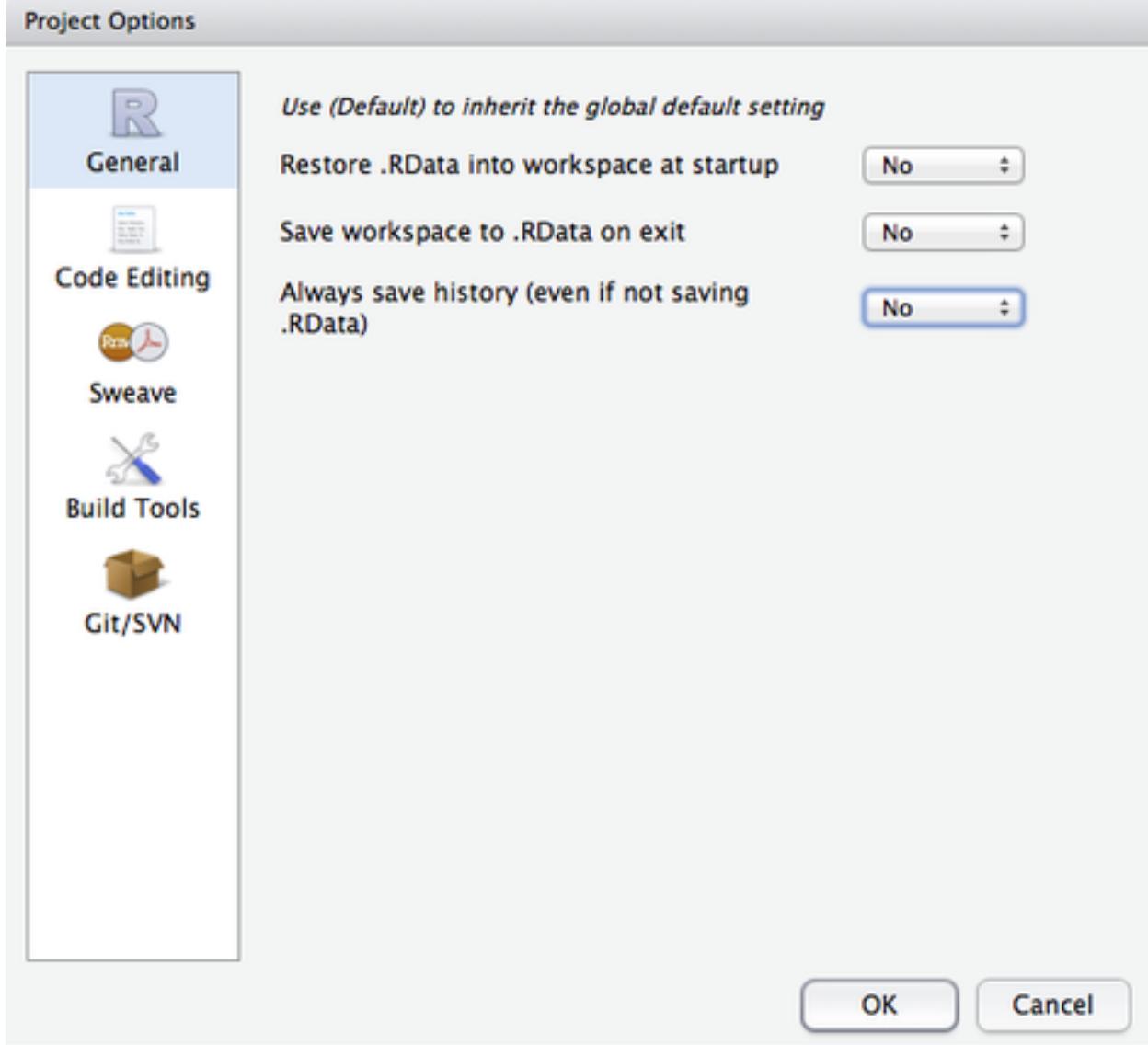


Figure 50:

The screenshot shows an RStudio interface with several tabs open: config.R, calculate_scores.R, data_prep.r, and cp_extent_chn2015_zb.csv. The Environment pane displays a table of data frames, including 'layers' which has 6 rows. A red arrow points from the 'layers' table in the Environment pane to the 'layers.csv' file in the File Explorer pane. A large red box highlights the error message: "These errors mean your data layers have not been registered in 'layers.csv' correctly." The message is overlaid on the RStudio interface.

```

# load required libraries
suppressWarnings(require(ohicore))

# set working directory to the scenario directory, ie containing conf and la
setwd('~/github/chn/province2015')

# load scenario configuration
conf = Conf('conf')

# run checks on scenario layers
CheckLayers('layers.csv', 'layers', flds_id=conf$config$layers_id_fie

# load scenario layers
layers = Layers('layers.csv', 'layers')

# calculate scenario scores
scores = CalculateAll(conf, layers, debug=F)
write.csv(scores, 'scores.csv', na='', row.names=F)

```

7.2.2 Calculating Pressures...

7.2.2.1 ‘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 (fish_oil,
Error in data.frame(names(P), P) :
  arguments imply differing number of rows: 0, 1

```

Figure 51:

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 |

Figure 52:

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

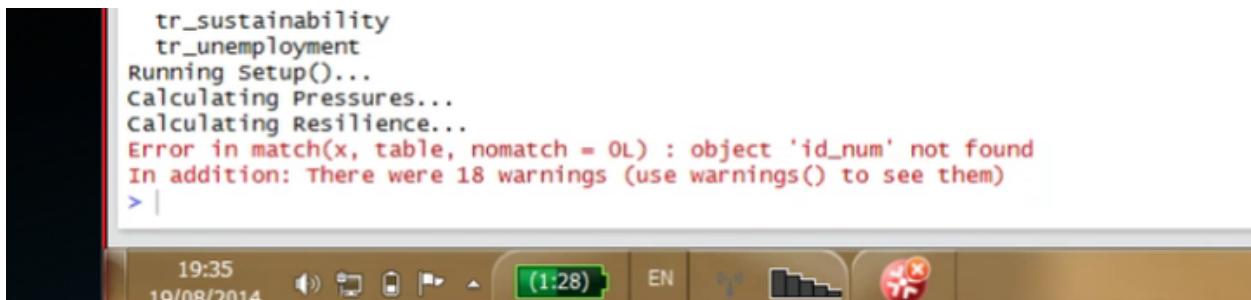
7.2.2.2 ‘Error in matrix...’ Example: >

This error means there is an empty column in `pressures_matrix.csv`, and the Toolbox cannot handle empty columns.

7.2.3 Calculating Resilience ...

7.2.3.1 ‘Error in match(x, table, nomatch = OL) : object id_num not found’

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



```
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)
> |
```

The screenshot shows a terminal window with R code being run. The code includes functions for sustainability and unemployment, followed by a series of calculations. An error message is displayed in red text, indicating that the object 'id_num' was not found. Below the terminal window, a taskbar is visible with various icons and a system tray.

Figure 53:

8 Overview of the OHI WebApp

OHI+ WebApps are websites created to facilitate independent assessments. You can easily navigate how information is organized and displayed when conducting an assessment. It is also meant to be used to visualize and communicate results.

The WebApp displays input information (data and indicators) as well as final OHI scores. When prepared and formatted for the OHI, inputs are called **layers** and are used in all OHI calculations, including goal models, pressures and resilience. By default, the WebApp only displays layers and score information that have been extracted from the latest global assessment and allocated to subcountry regions with the study area. The default display therefore does not provide fine resolution nor does it guarantee accurate data for each study area. You will substitute these layers with higher-quality information at the local scale in your assessment. However, the default layers can be used as inputs into your assessment in cases where no better information exists. **Incorporating the best information possible will generate results that best represent your study area.**

The WebApp is powered by the **OHI Toolbox**, which organizes all of the layers and calculates Index scores. The Toolbox is where you will actively work to prepare and format layers and develop goal models, which can then be displayed with the WebApp.

A default WebApp is available for most coastal nations. For example, Ecuador's WebApp (ECU) is found at <http://ohi-science.org/ecu>. Note that it is possible to translate the page into your language of choice.

Remember that this information is publicly available when stored on free GitHub accounts. If you would like your assessment to be private, you can choose a premium option.

The WebApp homepage provides tabs for you to explore your data, regions, and calculated scores. The main pages are **App**, **Regions**, **Layers**, **Goals**, and **Scores**, which were described in the **Conceptual Guide**. The App page is described more below.

The interactive **App** page allows you to explore input and output variables. The inputs are the layers and the outputs are calculated scores for each goal and dimension of the Index for each region in the study area. This page is where you should start your exploration. By choosing input layers you can see the range of values for a given variable and you can also see information about data sources. More detailed information about the default goal models and the calculation methods is found in the other pages of the WebApp.

The information displayed on the website is stored online in a **GitHub repository**. GitHub is an open-source development platform that allows multiple users to collaborate, track changes, and share their work to prepare data files and write code. Some members of your team will use GitHub to track layer preparation and view the history of changes made in this process. It also a way for your team to document the decisions made during your assessment. Any changes made to files contained within the GitHub repository will be automatically displayed on the WebApp for other team members and collaborators to view. The history of these changes is also stored on this platform as an archive, and it can be used to display changes made over time (See the section on **GitHub** for how to modify files using this platform).

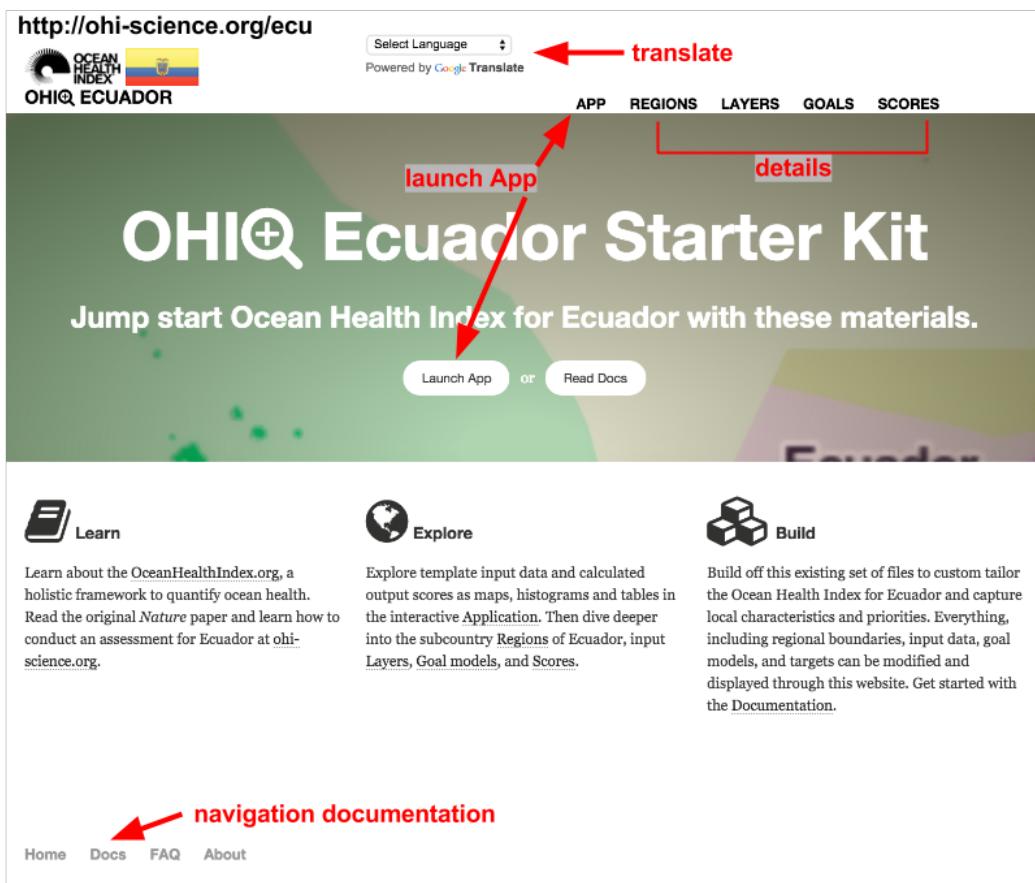


Figure 54: The WebApp start page. Note that it is possible to translate the page into your language of choice.

8.1 Defining and displaying regions

WebApps display subcountry regions within each study area. The boundaries for these subcountry regions are usually states, provinces, or districts reported to Global Administrative Areas (GADM: www.gadm.org). These land-based regions are extended offshore to divide the Exclusive Economic Zones (EEZs) into offshore regions of the study area. Offshore regions are important for Index calculations, in part because scores for each region are combined using the offshore area to weight the average of the final Index score. You can redefine these regional boundaries; these subcountry regions have been provided as a starting point. To redefine the boundaries you will need a spatial analyst; details are below.

It is important to note that the provided WebApps do not claim to take a stance on disputed territories. The boundaries for all EEZs were identified by MarineRegions.org (<http://www.marineregions.org>), and subcountry regions were identified by the Global Administrative Regions database (<http://gadm.org>).

8.2 Exploring inputs and outputs with the WebApp's App page

The App page allows you to explore the input layers and calculated output scores for each region in the study area. The page presents data and scores from the global assessment that are applied to each subcountry region in the study area by default. In order to explore data and scores, you can select them on the left to view their attributes and you can also visualize them in a number of ways on this page.

The App page allows you to view displays through the **Data** and **Compare** tabs. The **Data tab** provides summary information on each layer and metadata descriptions that accompany them. The **Compare tab** is mainly used for comparing output scores when modifications are made to the underlying data or models.

8.2.1 The App's Data tab

8.2.1.1 Overview of display options The Data tab displays input layer or calculated scores for each goal parameter. It presents the information through a *Map*, *Histogram*, or *Table*. These options are available as sub-tabs on the Data tab page. The *Map* sub-tab is the default display option for the Data tab, and all data presented are drawn from the global assessments by default. This means they are either directly duplicated across regions, or the raw values are down-scaled using offshore area- or population-weightings. The *Histogram* sub-tab likewise draws from the same data source, but it displays a histogram of observed values with a smoothed line added. The *Table* sub-tab also draws from the same data but offers information in a table.

Data displayed in the Map sub-tab:

The *Map* displays data for every region. A legend is displayed in the lower right-hand corner of the map to explain the meaning of the colors presented. The range of values will change when variables are selected, and the colors will automatically change to match that range.

Data displayed in the Histogram sub-tab:

The *Histogram* shows the distribution of values of the selected variable as the number of observations for each value bin (shown as white bars) and it also automatically creates a smoothed density function (shown as pink shading).

Data displayed in the Table sub-tab:

The *Table* displays the variable's value for each region in the study area. It provides an identifying code (*rgn_id*), name (*rgn_name*), and value (*value*) for each observation. The variables are also searchable since you can use the textbox at the bottom of the page to filter the results displayed.

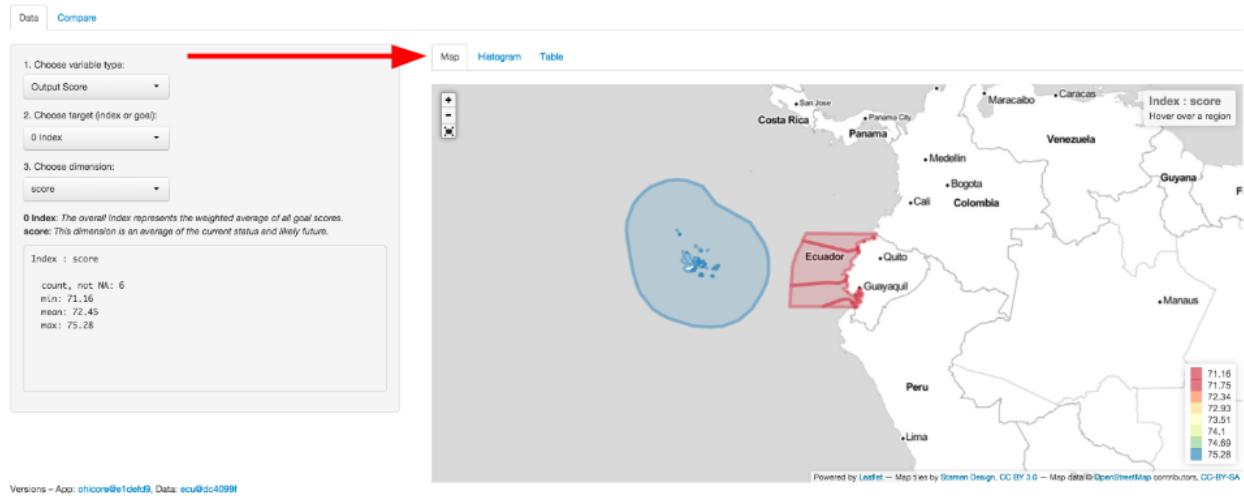


Figure 55: The Map sub-tab. Click on ‘Map’ to see a geographic view of your assessment region. Colors indicate scores or values for your input layers or output scores. This example shows Index scores for each region in Ecuador.

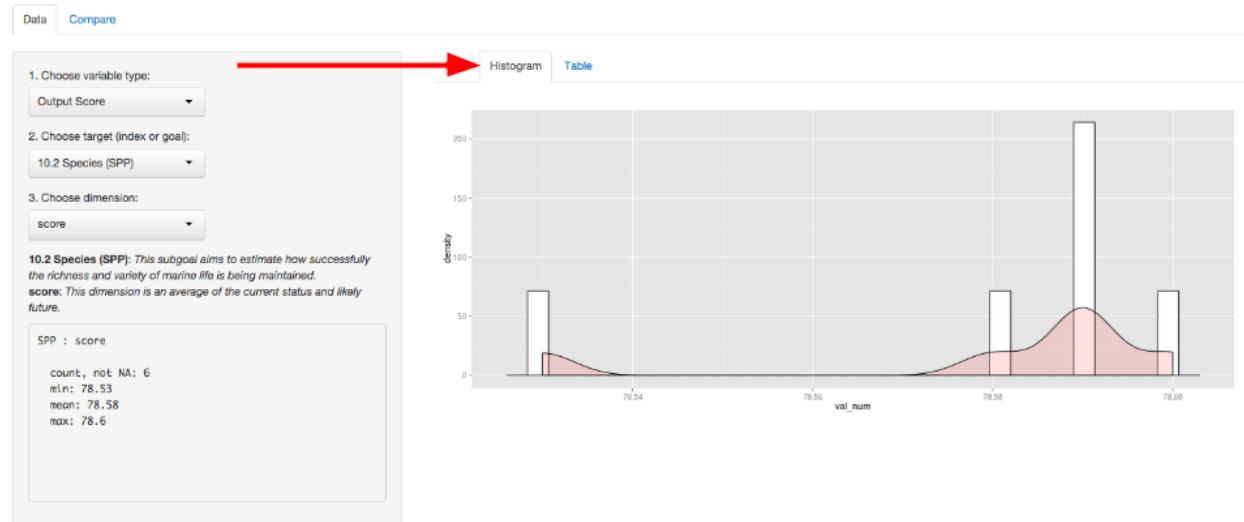


Figure 56: The Histogram sub-tab. Click on ‘Histogram’ to see the distribution of layers or scores after selecting a variable layer on the left. This example shows the Species sub-goal scores for the study regions of Ecuador.

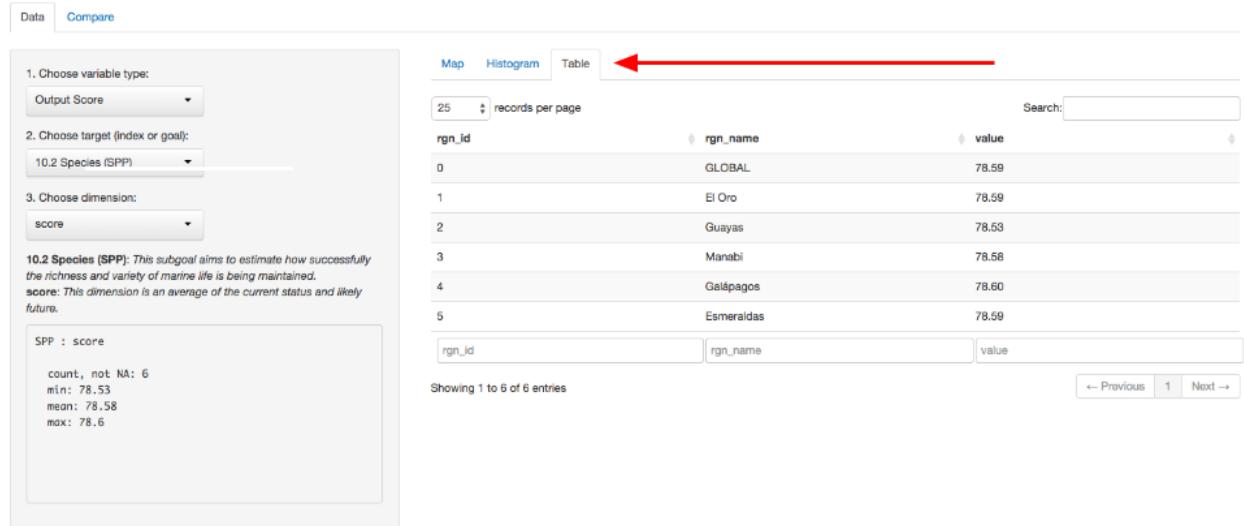


Figure 57: The Table sub-tab. Click on ‘Table’ to see a table of your data or scores, after selecting a variable layer on the left. This example shows the Species sub-goal scores for the regions of Ecuador.

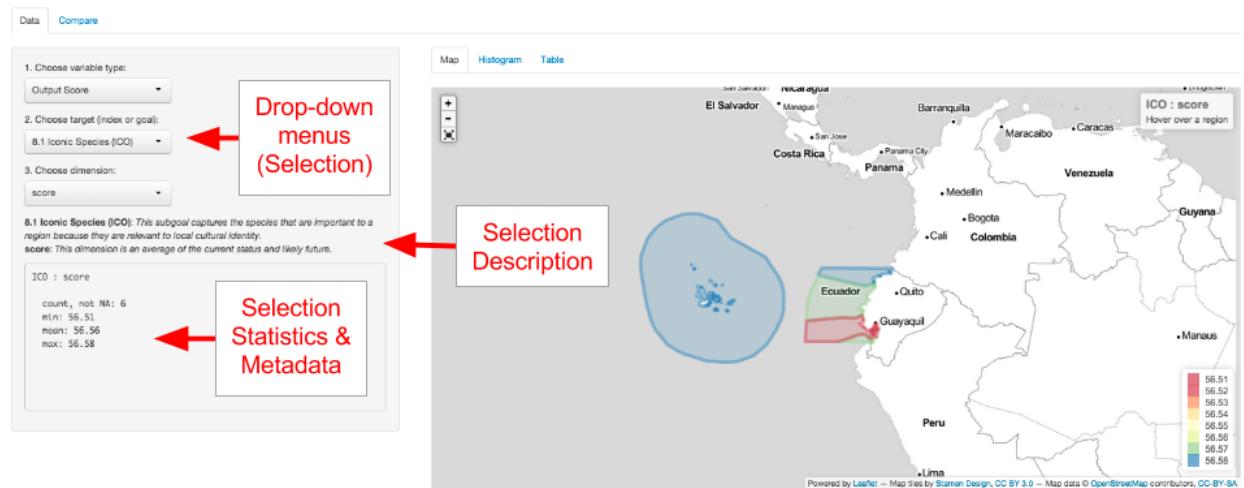


Figure 58: Overview of the Data tab. Choose the variable you would like to explore through the drop-down menus on the left-hand side of the page. Once you select either input data or an output score, you can view a description and summary of values below.

8.2.1.2 Overview of variable options When you choose the variables to be displayed, you will also see summary descriptions for the layer. These descriptions, statistics, and metadata for the chosen fields are displayed below the drop-down menus on the left side of the page.

TIP: As you prepare new layers, your updated descriptions will appear here.

The first selection you should make from the drop-down menus is the variable type. This means you can choose either **Input Layer** or **Output Score**. The **Input Layer** will show the layer used for a particular target you select. The targets in this case are either goals, pressures, resilience, or spatial information. The **Output Score** will show calculated scores for the alternative target selections you will make. In this case, the targets are Index or goal scores. In either case, you then have the option to further refine your search by either going into a specific layer or a specific dimension that is used in the overall Index calculations. If you do not make a selection, the **Output Score** is displayed by default.

For example, if you select **Output Score** as the variable type, you will then be able to choose a target goal or sub-goal, and then you will be able to choose the OHI dimension to be displayed. Remember that the dimensions are status, trend, pressures, resilience, future state, and score. In this way you can investigate the components that combine to create the goal scores.

As another example, if you select **Input Layer** as the variable type, you will be able to choose a target such as a goal and a specific layer associated with that goal. If that input layer has multiple categories of input types, or if it has multiple years of information available, you will be able to select more specific information. If you do not make a selection, the default setting is the first alphabetical category and the most recent year available.

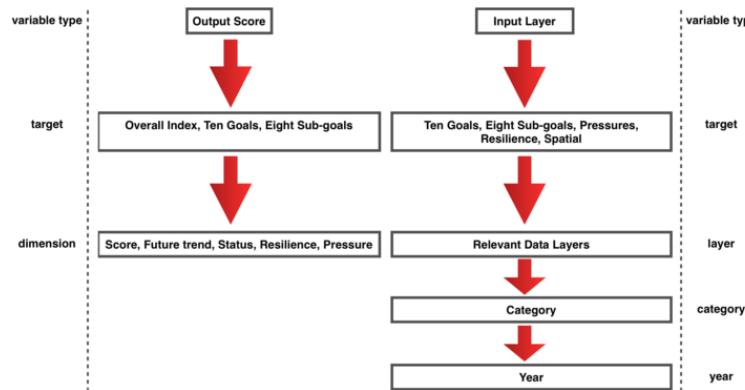


Figure 59: Overview of the variable options. You can choose to select and view either `input layers` or `output scores`.

8.2.2 The App's Compare tab

The **Compare** tab allows you to compare differences in calculated scores based on changes you have made to the underlying layers. These changes can be the values of the layers themselves, or they can be from changing the goal models. Any component you change is tracked through the archiving system of **GitHub** (See the section on **GitHub**) and each version of the changes be visualized here. You can take advantage of this ability to compare one saved version of your calculated Index output to another version of your calculated output. This is done to compare the how changes made to your data or indicators for goals, pressures, or resilience would affect the resulting scores.

You can use this in two ways. One way is use this is to visualize updates as you make them: viewing differences is extremely helpful for error checking and for sharing tests to the data among your technical team. Another way to use the Compare tab is to compare different management scenario to how changes in

your assumptions would impact score results. These changes could occur in the goal models themselves, such as through changes to targets or reference points, or they could be changes made to the values of pressures and resilience layers, for example.

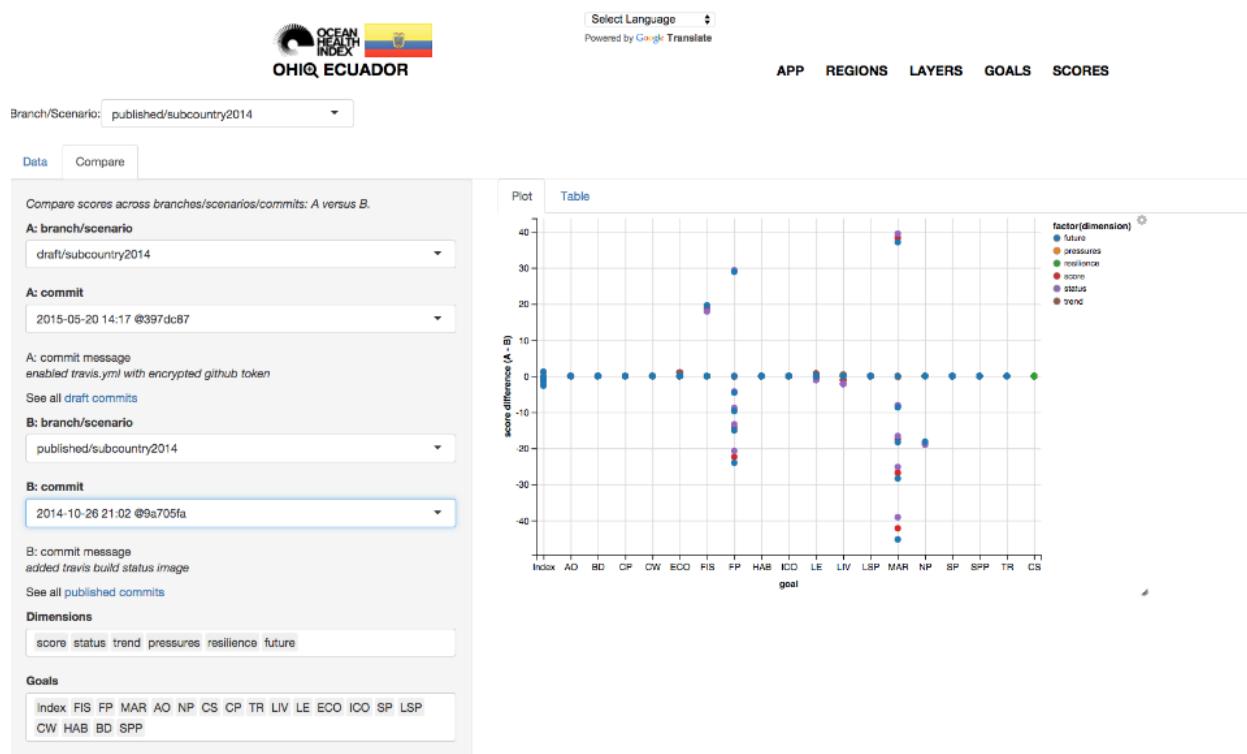


Figure 60: You can use the OHI+ WebApp’s ‘Compare’ tab to error check, and check the outcomes of alternate scenarios of your assessment.

The App page also offers the ability to view different **branches** or **scenarios** in the upper left-hand corner of the page. The **branches** refer to unique copies of a GitHub repository where information is stored. Each branch is a copy of a repository that is meant to be modified independently of other branches. This is done to ensure that changes made to one branch will not affect the information in another branch. This allows for active collaboration and offers a way to archive different outputs to Index calculations. For example, the *published* branch shows information that has been vetted for sharing, while the *draft* branch can be used for experimentation. These branches can be merged together at any time, and that is typically done when important milestones in the assessment process are reached. The *subcountry* folders displayed also offer another way to compartmentalize outcomes by allowing you to compare different **scenarios** within the same branch of your repository.

The App displays a *published* branch by default. It is recommended work on the *draft* branch until your assessment is finalized. When it is finalized, you can then merge the *draft* branch with the *published* branch.

These options for displaying and comparing information will be useful for understanding the multiple objectives in your OHI+ assessment.

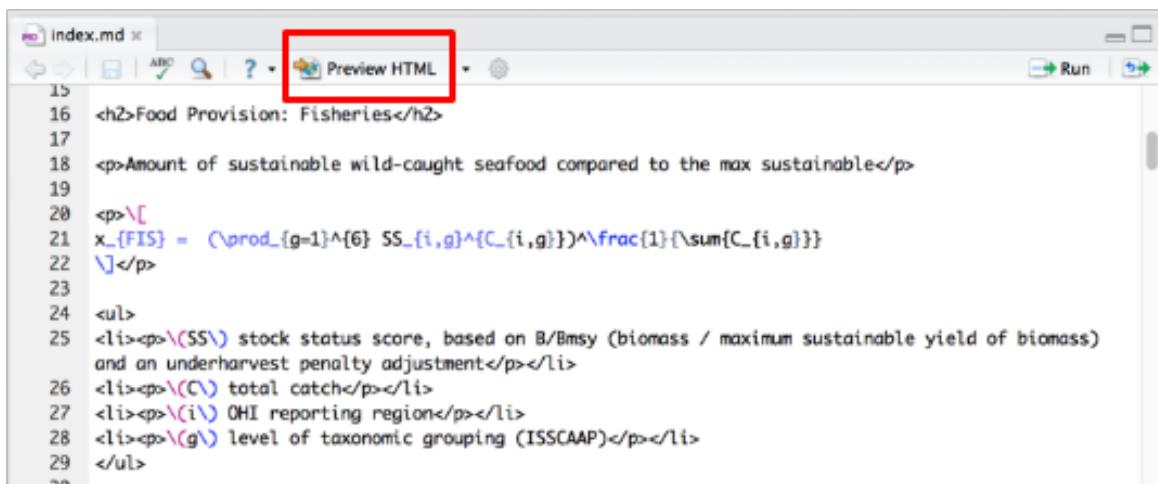
8.3 Updating the WebApp’s pages

The WebApp displays input layers on several pages: on the App page, Layers page, and Scores page. These input layers are displayed from `layers.csv` and the layers within the `layers` folder and the scores are

displayed from `scores.csv`. While the input layers and scores will be automatically displayed on the WebApp, there is other content on WebApps pages that can be edited by your team and displayed. You will likely spend the most time updating the equations displayed on the Goals page to be consistent with the updated methods you have used in your assessment.

For the WebApp to display the pages properly, not everything on each page can be edited as it is written in a language to create the website. But it is possible to explore the files and update much of the text that is displayed while maintaining the required formatting. To do this, run `copy_webapps_templates.r` once, before further modifications on the WebApp contents. This actions creates a folder called `webapps_templates` and copies template files there.

It is best to edit the files in RStudio: you will be able to view your work as it will be displayed on the WebApp by clicking the ‘Preview HTML’.



The screenshot shows the RStudio interface with the 'index.md' file open. The top bar includes standard icons for file operations, a search function, and a 'Preview HTML' button, which is highlighted with a red rectangle. Below the preview button are other icons for 'Run' and 'Edit'. The main area displays the content of the 'index.md' file, which is a Markdown document containing R code for generating HTML output. The code includes sections for 'Food Provision: Fisheries', calculations for 'FIS' scores, and lists of stock status, total catch, reporting regions, and taxonomic grouping levels.

```
15<h2>Food Provision: Fisheries</h2>
16<p>Amount of sustainable wild-caught seafood compared to the max sustainable</p>
17<20<p>\[<21x_{FIS} = (\prod_{g=1}^6 SS_{i,g})^{C_{i,g}}\] / \sum C_{i,g}</p>
22\]</p>
23<ul>
24<li><p>\(CSS\)</p> stock status score, based on B/Bmsy (biomass / maximum sustainable yield of biomass)
and an underharvest penalty adjustment</li>
26<li><p>\(CV\)</p> total catch</li>
27<li><p>\(CI\)</p> OHI reporting region</li>
28<li><p>\(g\)</p> level of taxonomic grouping (ISSCAAP)</li>
29</ul>
```

Figure 61:

8.3.1 Regions

You may have redefined the spatial boundaries of the regions used in your assessment, or you may want to update the information provided about them. This can be done with the file called `webapps_templates/regions.brew.md`.

8.3.2 Layers

Most of the information displayed on the Layers page of the WebApp is taken from the `layers.csv`, and therefore to modify any information about specific data layers, you will need to modify the `layers.csv` file within the draft branch of your repository. However, you are able to edit the header text information at the top of the Layers page if you wish. This can be done with the file called `webapps_templates/layers.brew.md`.

8.3.3 Goals

You will likely spend the most time modifying the information displayed on the Goals page, as these show and describe the models used in the assessment. Text can be modified with the file called

`webapps_templates/goals.brew.md`.

To edit the goal equations themselves, you will edit the `goals.Rmd` found in the `conf` folder (example: `ecu/subcountry2014/conf/goals.Rmd`). This is an RMarkdown file, with equations written in LaTex. When rendered by RStudio or the WebApp, it displays nicely formatted. To update model equations, you will need to use the LaTex format. You can learn the syntax by studying how the equations from the global assessments are displayed, and from many resources online. One resource is <https://en.wikibooks.org/wiki/LaTeX/Mathematics>. Learn more about .Rmd formatting at <http://shiny.rstudio.com/articles/rmarkdown.html>.

8.3.4 Scores

The scores displayed on the Scores page of the WebApp are the calculated scores from the `scores.csv` file in the draft branch, and therefore cannot be modified. However, you are able to edit the header text information at the top of the Scores page if you wish. This can be done with the file called `webapps_templates/scores.brew.md`.

9 Appendix 1: Goal Models, Reference Points, and Pressure and Resilience

Once you have determined which goals are assessed and have begun searching for data and indicators, you can start to develop goal models and set reference points. The decision tree of the data discovery process also applies here: first consider how goals can be tailored to your local context before you consider replicating what was done in the Global Assessments. It is always better to use local goal model and reference point approaches where possible. This section aims to provide you with goal-by-goal guidance on how to find data, pick indicators, set reference point, and develop the model, as well as guideline on how to think about pressure and resilience. But first, let's see some general tips before diving into the details of each goal model.

9.1 Developing multiple goal models at the same time

Goals can be assessed independent of one another. As each goal model is developed and data gathered, it can be assessed without affecting other goals.

However, you can develop some goal models simultaneously and streamline the data search. For example, the habitat-based goals **Carbon Storage**, **Coastal Protection**, and the **Habitats sub-goal of Biodiversity** all rely on the same underlying data, and their models can be developed together. A spatial analyst can create the spatial layers used for these goals with the same source material. This will greatly expedite your data layer preparation. Species data for **Iconic Species sub-goal of Sense of Place** is often a subset of data from **Species sub-goal of Biodiversity**. Data for non-food marine products for **Natural Products** and food products for **Fisheries sub-goal of Food Provision** are often recorded in similar data sources and may need partitioning.

If you wish to further coordinate these activities on a higher level, you could have the same team member coordinate activities for the development of certain goals. That is a consideration when assembling your team and planning your workflow. For more details, please see the goal-specific sections.

9.2 Keeping Reference Points in Mind

Setting a reference point is required for every goal model you develop. It is an “ideal” condition, or target, where the goal is considered to be achieved to its full potential. Achieving or exceeding the reference point will result in a score of 100. The choice of a reference point will thus affect how the final scores are calculated, and must be balanced between knowledge of the system, expert judgment, and limitations of the data. You

may set an universal reference point for all regions in your study area, or you may set a unique reference point for each region.

Generally there are four types of reference points: + *Functional*: Scientifically sound target set based on the known responses of variables measured, such as Maximum Sustainable Yield. + *Temporal*: A historical benchmark is used as a the “ideal” point in the past, such as mangrove coverage in the 1980’s. + *Spatial*: A region within the study area with the highest input values, and all others are scaled to it. + *Established target*: Such as a sustainable catch yield by a certain year, or the number of people employed in a marine sector by a certain year.

Which type of reference point to use depends on the goal and available data. How many years of data are available? Can you set a temporal reference point with these data, or do you have to find another dataset or other source of information? In any case, you must balance being realistic with being ambitious. We suggest following the S.M.A.R.T. criteria when choosing a reference point: “*Specific*,” “*Measurable*,” “*Ambitious*,” “*Realistic*,” and “*Time-bound*.”

You will learn more, and think more critically about reference points, as you develop the data layers for your assessment.

How to use the reference point in a model

It’s best to explicitly include the reference point in the model equation whenever possible. For example, the Carbon Storage goal model in the global assessment is written like this:

$$x_{CS} = \sum_1^k \frac{C_s}{C_r} * \frac{A_k}{A_T},$$

where Cr is the reference condition of each habitat. See goal-specific sections for more examples.

9.3 Identifying pressures and resilience

While you are developing goal models, you should note the links between your goals and pressures and resilience: both the pressures and resilience that affect them and whether the goal acts as pressure or resilience on other goals. It is recommended to begin gathering data of pressure and resilience from the start of the assessment. The team members who are developing specific goals should think about the pressures that act upon those goals as they are data-gathering, and they should think about the data sources that could provide pressures information. However, it may be most useful when one team member gathers all of the data for pressures, since the same pressures often affect multiple goals. See **Pressure and Resilience** section of this chapter for more information.

Some pressure data are the same or closely-related to data for goals. For example, the **Wild-Caught Fisheries** goal model requires catch data, which may be the same data source for information on commercial high- and low-bycatch data, which are used as pressures layers that affect **Livelihoods and Economies** and **Biodiversity**. In global assessments, the **Clean Waters** goal is very much linked to pressures layers because the input layers for its status are used as pressure layers. Trash pollution is a pressure that affects **Tourism and Recreation**, **Lasting Special Places**, **Livelihoods and Economies**, and **Species**. It is important to remember these linkages as you go through the data discovery process.

You should also start searching for pressures data independent from data for goals. An example would be how climate change impacts will appear in various places in your assessment. Climate change pressures layers can include UV radiation, sea surface temperature (SST), sea-level rise (SLR), and ocean acidification, and these impacts might affect such goals as **Natural Products**, **Carbon Storage**, **Coastal Protection**, **Sense of Place**, **Livelihoods and Economies**, and **Biodiversity**. These linkages will become more clear as you go through the OHI+ assessment process.

->

9.3.1 Artisanal Fishing Opportunities

Artisanal fishing, often also called small-scale fishing, provides a critical source of food, nutrition, poverty alleviation and livelihood opportunities for many people around the world, in particular in developing nations. As opposed to large, commercial fisheries that usually involves industrial, energy-intensive vessels and long trips, artisanal fisheries refer to households, cooperatives or small firms that use relatively small amounts of capital, energy and small fishing vessels (if any), make relatively short fishing trips, and use fish mainly for local consumption or trade.

In addition, this goal is not about recreational fishing often done in developed countries, which is captured in *Food Provision* (if it provides food) and Tourism and Recreation goals. Nor is it about the actual amount of fish caught or revenue generated, which are captured by *Coastal Livelihood and Economies*. Rather, **artisanal fishing opportunities measure whether people who need to fish on a small, local scale have the opportunity to do so.** A score of 100 means the country or region is meeting the needs of artisanal fishermen or communities by implementing institutional supports, providing access to near-shore water, and maintaining the health of targeted species.

Ideal Approach

Ideally, this goal would include some measure of how easy or hard it is for fishermen to *access* ocean resources when they *need* them and the *sustainability* of harvest of all nearshore stocks used by artisanal fishers. The need for artisanal fishing is most directly tied to the percent of population below the poverty line. Access, or ability to meet this need, are institutional permits or encouragements for small fisheries. Sustainability of artisanal fisheries refers to the sustainability of fishing gears. Sustainable gears are traditional hook and line, as opposed to unsustainable practices such as dynamite and gill net fishing. Sustainability can also be estimated by assessing the health of targeted species. The type of reference point you use will depend on the data available.

Practical Guidance

Ideal indicators mentioned above are often scarce, especially percent poverty or the sustainability of fishing gears. You will want to find proxy data for *access* in whatever way best suits your areas. This could be drawn from physical, economic, regulatory, or stock condition data as an indication of availability. A combination of all of these would be best to more accurately speak to the philosophy, but is usually limited by data.

In the U.S. West Coast Assessment (2014), three metrics were used to define **Artisanal Opportunity** that you can use to study:

| Type of Access | Data Used | Reference Point |
|----------------|--|-------------------------------------|
| Physical | Number of coastal access points per mile | 1 coastal access point per mile |
| Economic | Change in gas price over time | No change in gas price |
| Resource | Condition of fish stocks through NOAA Fish Stock Index | Perfect sustainability score on FSI |

You should include data that are distinct from the catch data used in other areas of **Food Provision**. So once you find catch data and access data, you should think about which goals to use it in. For instance, if you find tonnage of artisanally-caught fish, you should include that elsewhere. In any case, you will likely modify the default Global Assessment model using different and better-resolved data.

For example, in the Baltic Sea region, **Artisanal Opportunities** is very closely connected to **Tourism and Recreation** since there are a lot of locals and tourists using the shared sea for enjoyment. You will have to decide how to apportion the data.

The **reference point for this goal** would likely be a metric of having full access to the resource (see

examples in the table above). Keep in mind that the access is for people, and therefore a marine protected area may not count towards full access.

Examples of the Approach

| Assessment | Developing the Model |
|-------------------------------|--|
| Global 2012 | The status was the demand as estimated by poverty levels. The data were measured by the government. |
| Global 2013 | The approach was the same as Global 2012 |
| Brazil (2014) | The model was simplified to reflect the primary driver of opportunity as the availability of fish. |
| U.S. West Coast (2014) | This study developed a model using three key variables of physical and economic access to coastlines. |
| China (2015) | Status model is rewritten and is similar to the 2014 U.S. West Coast assessment. It is based on the same principles. |

9.3.2 Biodiversity

People value biodiversity in particular for its existence value. The risk of species extinction generates great emotional and moral concern for many people. As such, this goal assesses **the conservation status of species based on the best available global data through two sub-goals: Species and Habitats**. Species were assessed because they are what one typically thinks of in relation to biodiversity. Because only a small proportion of marine species worldwide have been mapped and assessed, we also assessed Habitats as part of this goal, and considered them a proxy for condition of the broad suite of species that depend on them. We calculate each of these sub-goals separately and treat them equally when calculating the overall goal score. A score of 100 means all species are at very low risk of extinction, and all habitats are conserved.

9.3.2.1 Sub-goal: Species This sub-goal assesses the health of all marine species present in a region, including endangered species and species in relatively good conditions. The presence of higher-risk species leads to a higher score.

Data for this goal is also used in Sense of Places sub-goal: Iconic Species. It will be effective for goal keepers of Biodiversity and Sense of Place to work together on data gathering.

Ideal Approach

Ideally, you would find data for all species present in your region including information on their *habitat ranges* along with scientific studies that speak to the *health of their populations*.

Practical Guidance

You should start by trying to *find spatial information for species* that occur in your area and determine *whether or not they have been scientifically assessed and given a conservation status*. It is best if you only use species for which there are both spatial data and an assessment. In the global assessment, we combined data from recent assessments of the International Union for the Conservation of Nature ([IUCN](#)) (www.iucnredlist.org/technical-documents/spatial-data) and [AquaMaps](#) (aquamaps.org). These data sets provide a geographic snapshot of how total marine biodiversity is faring, even though it is a very small sub-sample of overall species diversity.

IUCN provides global species assessments that indicate the distribution and the conservation status of species, ranging from Least Concern to Critically Endangered to Extinct. These risk categories were turned into weights (between 0 and 1) for calculations. AquaMaps offers additional spatial data for species not covered by the IUCN distribution maps. Both evaluations choose a taxon (e.g. sharks) and a group of world experts assess it comprehensively. Locally identified species identified in a regional assessment may not be in the global database because they do not belong to one of the taxa that have been selected for assessment, or because the experts that did the assessment did not know that information existed. In either case, there is no connection between what IUCN or AquaMaps report and what assessments are done locally. Therefore, it

might not be fair to penalize a study area for missing species. For biodiversity, it is unrealistic to expect that all species are assessed, so it seems unfair to penalize for un-assessed species. In the fisheries goal, there are penalties for species that are exploited but not assessed, because if there are landings data, it means they are somewhat measurable, and so it is reasonable to expect they should be at least monitored.

For regional assessments, local studies of marine species status and local datasets are best here. The spatial information can be a range map with simple presence or absence information, or it can have more detailed data. You can complement the species list search with a scientific literature search to see if anyone has scored the species status in a way that you can use. One thing to keep in mind when using local data sets is that *local lists might focus on endangered species and leave species in relatively good conditions un-assessed. This could result in poorer scores.*

If spatial distribution information for individual species is not available, the goal model could be simplified to be based on threat-status-weighted average in your region.

You should also think about the **reference point** for scores that signal poor Biodiversity status. You can use the same threshold as the Global Assessments which say that *places with extinction risk scores greater than seventy-five percent will get scores of zero*. This is an estimation based on the literature of mass extinctions (e.g., Barnosky *et al.*, 2011) and could be applied across scales. You don't need all species extinct for there to be a zero, so you will have to choose how to rescale it, and whether the risk effects are linear or nonlinear.

Additional Note: Defining spatial regions and map considerations

Once you have gathered the data, the treatment of it will matter for the model and goal score calculation. Do you know how the data were collected? Do you have information on sampling effort? If you don't know, you may not be sure whether changes in condition are due to monitoring efforts or biodiversity change, and you therefore may want to consider the uncertainty of your model.

In any case, the original logic of the **Species** sub-goal of the **Biodiversity** goal is to represent the species present relative to the proportion of their range within a given region. The goal is to summarize extinction risk for an area across assessed species, and assign it appropriately so that the loss of species scores poorly. You should consider whether the impacts to local species status are linear or non-linear. Will drawing borders affect how your scores are assigned?

When considering how to change the model, you should think about the outcome of the score on your decisions. For instance, will weighting a “Critically Endangered” species higher on the scale result in the inclusion or exclusion of more rare species? Will the way you aggregate spatial data to summarize extinction risk for your area take into account the influence of species with smaller ranges size, or will that information be lost in the averaging process? An inherent disadvantage for conservation may occur when rare species get rarer in the future, and will therefore have a relatively small influence on the score while common species drive the results.

Range size has an impact on score results. For example, if you use the current model, none of the cone snail species listed below will have a big impact on the **Species** sub-goal score because to their small range size that covers one cell of map area. However, the 0.8 score for *Conus roeckeli* shows that it is a rare species as assessed by the IUCN. On the other hand, the coral *Acropora palmata* is also rare and yet covers a large range.

| Scientific name | IUCN Category | Trend | Map Cells |
|--------------------------|-----------------------|------------|-----------|
| <i>Conus salreiensis</i> | Critically Endangered | Decreasing | 1 |
| <i>Conus trochulus</i> | Near Threatened | Unknown | 1 |
| <i>Conus roeckeli</i> | Least Concern | Unknown | 1 |
| <i>Acropora palmata</i> | Critically Endangered | Stable | 1158 |

Examples of the Approach

| Assessment | Developing the Model |
|-------------------------------|---|
| Global 2012 | The status of assessed species was calculated as the area- and threat status-weighted average |
| Global 2013 | The goal model was the same as Global 2012. There were data updates available for 15 out of |
| Brazil (2014) | The status of 504 assessed species was calculated as the threat status-weighted average of spec |
| U.S. West Coast (2014) | The model description and reference point were the same as Global 2012, with regional data a |

9.3.2.2 Sub-goal: Habitats The Habitats sub-goal includes all habitats in the study area, and assess their health condition and coverage area.

Habitat is included in the Biodiversity goal to provide a more complete picture of diversity in the system. This is because in global assessments assessed species data were limited and the diversity of habitats can be included with the assumption that healthier habitats mean healthier species. Therefore if you have comprehensive species assessments in your area you may not need to include all habitats as a sub-goal, instead only including living habitats (algae, corals) along with species in the Species sub-goal.

Ideal Approach

Ideally, information on the area extent (square kilometers) and condition of every habitat type would be available over a long period of time.

Practical Guidance

You will first find what habitats are in your study area, both in the *coastal regions* and *offshore*. In the global assessments, data were available for mangroves, coral reefs, seagrass beds, salt marshes, sea ice, or subtidal soft-bottom habitats, but there are likely other important habitats in your region.

Once you determine what habitats are in your area, you will need information about *the area extent of each habitat* within each region of your study area. You should consider whether and how far they go offshore and inland. Spatial data are preferred: you will be able to calculate the total square kilometers of each habitat within each region. For example, in global assessments mangrove area within 1 kilometer of the coastline were included, but this distance could change in OHI+ assessments.

You will need to include area and condition data specific to your study area and not rely on the global data provided. This is because the habitat data provided for your assessment are either over-representative by allocating the study area's data equally to each subcountry region, or can be misrepresentative by allocating a proportion of those data to each subcountry region (based on the offshore area of each region to the total area). Either case assumes that if a habitat was present in the country, it could be found anywhere (which is not the case with study areas that span many degrees of latitude), and the first case inflates the habitat coverage in the study area.

TIP 1: Do you have maps that show current habitat distributions and maps that show historical habitat distributions? If so, you could extract that data for each of your regions to get a current and reference area. In the U.S. West Coast assessment (2014), researchers went to the local public library to find hand-drawn maps of historical salt marsh and sand dune extents in California. You could also use summarized habitat data that exists in tables or are already compiled in another source.

TIP 2: If local data is not available, you obtain certain habitat extent data from the global assessment, which used high resolution spatial extend data from this study: *Hamilton, S. & Casey, D. (2014). Creation of a high spatiotemporal resolution global database of continuous mangrove forest cover for the 21st Century (CGMFC-21): a big-data fusion approach. arXiv:1412.0722*. Good data to obtain from this study include shoreline, mangrove, soft bottom, and sea ice.

The *condition* of the habitats can be measured in different ways, depending on the data available. Indicators of condition could include habitat density, susceptibility to pathogens, or change in species composition or growth rates from impacts such as overgrazing. Look for studies assessing habitat integrity or condition specifically in your area.

In global assessments, direct information about coral condition was not available so it was based on the percentage of “living cover” on a coral reef relative to the potential range of the reef. 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. Mean predicted values for each region from 1985-1987 were set as reference points.

If possible, you will want to gather habitat area and condition data through time so that you can calculate the **trend**. Ideally, there will be enough years of data to directly calculate the recent change in habitat condition as the trend. This isn’t always the case, and proxies or estimates might need to be used. For example, due to spotty salt marsh data we created trend categories of increasing (0.5), stable (0), and decreasing (-0.5) based on available data.

The **reference point** will likely be based on historic habitat area coverage and condition, with the assumption that habitat destruction has been and still is occurring and the target is to return to some point in the past. You will need historical data for this, from satellites, published papers, or even hand-drawn maps. The challenge is to find a reference point that is both *ambitious* and *realistic* (based on the S.M.A.R.T. principles: see *Conceptual Guide*), using the data available. If data allow, it will be possible to set a reference point that is more ambitious than that used in global assessments.

Alternatively, the reference point could be guided by a policy target. For example, are there any climate change policies in your area, with defined targets and objectives? Are there any restoration or carbon storage projects in your area? Do any organizations offer guidance on the amount of carbon storage your management policies should be aiming for?

Examples of the Approaches

| Assessment | Developing the Model |
|-------------------------------|--|
| Global 2012 | The status was assessed for all habitats for mangroves, coral reefs, seagrass beds, salt marshes |
| Global 2013 | The goal model was the same as 2012. |
| Brazil (2014) | The goal model was the same as as Global 2012 for mangroves, coral reefs, seagrass beds, salt |
| U.S. West Coast (2014) | Salt marshes and seagrass beds were considered. Extent was used and habitat health was not |

9.3.3 Coastal protection

This goal aims to assess **the amount of protection provided by marine and coastal habitats against flooding and erosion to coastal areas that people value, both inhabited (homes and other structures) and uninhabited (parks, special places, etc.).** A score of 100 would indicate that these habitats are all still intact or have been restored to their reference conditions.

Habitat-based goals should be considered together during the data gathering process because the same data underly three goals: **Carbon Storage, Coastal Protection, and Biodiversity.** Goal models to date have depended on the area (square kilometers) of each habitat type in each region, the condition of each habitat, and a weighting to distinguish how different habitats contribute to each goal.

Ideal Approach

Ideally you will have information on the relative potential of habitats to provide coastal protection, and extent and condition of each habitat. The reference point would likely be a historic reference before destruction of coastal habitats.

Practical Guidance

To see how to obtain data on habitat *extent* and *condition*, read the section on **Biodiversity sub-goal: Habitats**.

There can be many *habitat types* included in Coastal Protection goal. In global assessments, coral, mangroves, saltmarshes, seagrasses, and seaice were included. Habitats were *weighted* based on their protective ability identified by [Natural Capital Project](#). Depending on the habitats you include, you will need to find additional weights.

In the global assessment, we did not include an assessment of the protection afforded by man-made structures, such as jetties and seawalls. These structures cannot be preserved without maintenance and often have negative side effects (e.g. alter sedimentation rates causing erosion in new locations), thus they do not constitute long-term sustainable service and are mostly seen as a pressure. 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.

The goal status model for Coastal Protection developed for global assessments will likely be appropriate for independent assessments. However, it could be possible to incorporate the *differences in vulnerability* between subcountry regions. Vulnerability can be quantified as the ability to evacuate, economic ability to reconstruct in case of damage. Physical properties may be available in OHI+ assessments, allowing for more a detailed understanding of the protective ability, and likelihood of exposure for each habitat type in different portions of the coastline.

Examples of the Approaches

| Assessment | Model Description and Reference Point |
|------------------------|---|
| Global 2012 | The habitats included mangroves, coral reefs, seagrasses, salt marshes, and sea ice. The status model was the same as in Global 2012. |
| Global 2013 | The goal model was the same as in Global 2012. |
| Brazil (2014) | The 12 nmi boundary was used for each habitat type for mangroves, seagrasses, coral reefs, and sand dunes. |
| U.S. West Coast (2014) | Salt marshes, seagrasses, and sand dunes were included. |

9.3.4 Carbon storage

Highly productive coastal wetland ecosystems store substantially larger amount of carbon than terrestrial forests and have the highest sequestration rates of any habitats on earth. They are also threatened by under-regulated coastal development but are amenable to restoration and conservation efforts. This goal intends to capture the **ability of the coastal habitats to remove carbon given their carbon uptake rate and health conditions**. A score of 100 means all habitats that contribute to carbon removal are still intact or have been restored and they can function to their full carbon burial potential.

Habitat-based goals should be considered together during the data gathering process because the same data underly three goals: **Carbon Storage**, **Coastal Protection**, and **Biodiversity**. Goal models to date have depended on the area (square kilometers) of each habitat type in each region, the condition of each habitat, and a weighting to distinguish how different habitats contribute to each goal.

Ideal Approach

Ideally, you will know which habitats contribute to carbon *sequestration and storage*, and their *capacity* to do so. You will also have information on *habitat extent* and *health* conditions. The reference point for habitat-based goals will likely be temporal, meaning you will compare the current area of these habitats to some area in the past, and thus historic data are needed. Reference points could also be set by with a proportion increase (or decrease) of known conditions.

Practical Guidance

In terms of habitat types for this goal, it is recommended to search for *mangroves*, *saltmarshes*, and *seagrasses* because these are viewed as carbon-sequestering habitats that are both ecologically threatened and sensitive to policy responses. It is possible that there are other carbon-sequestering habitats present in your study area. But you should remember that we recommend using habitats that can store carbon on the order of 100 years, thereby limiting the types of habitat types you will need. *For details on how to obtain data on habitat extent and condition, read section on Biodiversity sub-goal: Habitats.*

Seaweeds and coral store carbon for less than 100 years, and 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.

Contribution is relative contribution of each habitat to total carbon storage, measured by the per area, empirical rates of carbon uptake and coverage area. For this you would have to search the literature and find ratios of organic nutrient uptake between habitats, and you would have to make sure these studies represent your study area. For example, were the studies done with a young mangrove forest, or an older one, which might have different growth rates?

Understanding habitat carbon storage rates is an area of ongoing research. The capacity for habitats to store carbon varies, and depends on the morphology of plants in the system. In the global assessment, we assumed equal potential for carbon storage by each habitat, and thus contribution to carbon storage is measured by habitat extents directly.

Examples of the Approach

| Assessment | Developing the Model |
|------------------------|---|
| Global 2012 | Seagrasses, tidal marshes and mangroves, were assessed. The whole extent of mangroves was included. |
| Global 2013 | The goal model was the same as in Global 2012. Mangrove data included 1km inland in addition to the coastal area. |
| Brazil (2014) | The goal model was the same as in Global 2012. The greatest data gaps were for sea grasses. |
| U.S. West Coast (2014) | Salt marshes and seagrass beds were considered. Extent was used and habitat health was not considered. |

9.3.5 Clean Waters

People value marine waters that are free of pollution and debris for aesthetic and health reasons. Contamination of waters comes from oil spills, chemicals, eutrophication, algal blooms, disease pathogens (e.g., fecal coliform, viruses, and parasites from sewage outflow), floating trash, and mass kills of organisms due to pollution. People are sensitive to these phenomena occurring in areas that they access for recreation or other purposes as well as for simply knowing that clean waters exist. The Clean Water goal captures **the degree to which**

local waters are unpolluted by natural and human-made causes. This goal scores highest when the contamination level is zero.

Ideal Approach

Ideally, data would be available and combined from different categories of marine pollution that can directly cause waters to become unsuitable for recreation, enjoyment, and other purposes. These factors typically include eutrophication (nutrients), chemicals, pathogens, oil pollution, and marine debris. The Status of these components is the inverse of their intensity (i.e., high input is a bad status).

Practical Guidance

First think about what kind of point and non-point sources of pollution are in your area. Are there known sources of trash and marine debris? Is there a population that does not have access to sanitation? Does your wastewater get effectively treated before it is discharged into the environment? How does urban runoff contribute to your local coastal waters? Where possible, you should think about categories of inputs that are not used in the Global Assessments. These include information on toxic algal blooms, oil spills, turbidity or sediment input, and floating trash, and think about how they can fit in to the model areas.

Once you identify the sources, you should try to find *in situ* measurements of contamination. This could include monitoring data for pathogen levels, chemical contaminants, or data on the frequency and location of anoxic conditions or eutrophication. If direct measurements of water pollution are unavailable, indirect indicators, or proxy data, could be used. For example, land-based nutrients pollution is a proxy measure to nutrient input. The number of people who do not have improved access to sanitation could indicate the level of pathogen pollution.

If you don't have such information, or have partial information on one of these inputs, you could combine it with population data or model the data to estimate how much of an impact it has on coastal areas. Do you have population density information over time? You could then use this later to calculate the trend.

In the 2012 Global Assessment, these data layers are scores spatially for 3 nm out from shore. This is because it concentrates the effect of the inputs to coastal regions and makes them visible to the shoreline. You can change this value in your spatial analysis.

You should use more refined data than the Global Assessment data, because they rely heavily on proxy data for water quality. However, you should follow the same approach of the *geometric mean* to sum up the data layers that you find. This guarantees that if any one of the components scores poorly, the higher scores from other components will not hide the effect.

The Clean Waters goal is connected to the **Pressures** layers. You should approach the both of them at the same time where possible. (See the sections on **Pressures and Resilience** to understand more.)

In the Global Assessment, marine debris from plastic pollution is one of the pressures layers. The same data are also used in the **Clean Waters** goal, but they are inverted such that a low value of debris is a high score in the goal. This is commonly done for these interchangeable data layers.

For each contamination category, a reference point is set and all data in this category is scaled to 1. The type of **reference point** used will depend on the data available. You may decide that an ocean completely rid of pollution is ideal, or you may find that beach closure of less than 10 days per year due to *E.coli* contamination is acceptable.

Examples of the Approach

| Reference point is when the contamination level is zero. | The lack of direct measurements meant that modeled and proxy data were used. The status of this goal was also used in the pressures layers.

Global 2013 | The model was same as Global 2012, with a few simplifications; revenue data were adjusted by dividing by GDP per region, reported in 2013 USD. | The reference point was the same as Global 2012. | The approach was the same as Global 2012, with simplifications.

Brazil (2014) | The goal model and reference point were the same as Global 2012. Data used to model the components for eutrophication (nutrients) and chemicals was the same as in Global 2012, while pathogens and debris were localized to state level data. | The reference point approach was the same as Global 2012. | The study used better, or more local, data than the Global.

U.S. West Coast (2014) | The model was the same as Global 2012, with regional instead of global data. | The reference point was set as the number of days when beaches were closed to bathers because pathogen counts were higher than state standards. | The study used more local data than the Global.

China (2015) | Status model is similar to global assessments. Pollution is based on nitrogen, phosphate, chemical oxygen demand, and oil pollution. | The same reference point as Global 2012 was used that waters are free from all pollution. | The study used all local data. Data on pathogens and marine debris are poor or unavailable and thus were ignored in the model.

9.3.6 Food Provision

One of the most fundamental services the ocean provides people is the provision of seafood, whether it is helping meet the basic nutritional needs of over half of the world's population to high-end sushi. This goal, then, **measures the amount of seafood sustainably harvested in a given EEZ or region through any means for use primarily in human consumption or export** and thus includes wild-caught commercial fisheries, mariculture, artisanal-scale and recreational fisheries.

The Food Provision goals aim to maximize the amount of sustainably produced seafood from wild or ocean-cultured stocks; any unsustainable extractive practices is penalized, and so is over- or under- harvesting.

Because we do not track where the fish go after being caught or produced, this goal does not aim to measure food security for the population of a given country, but instead measures the food provided from its waters.

It would be ideal to include catch and effort information for commercially, recreationally, and artisanally fished species in your area, although most completed assessments have only been able to include catch information from commercial fisheries due to data availability limitations. When data become available for artisanal and/or recreational catch, they could be included as part of the fisheries sub-goal or as a separate sub-goal depending on the context.

The sub-goals of Food Provision (Fisheries and Mariculture) measure the amount of goods sustainably harvested from the sea for human consumption, while the Natural Products goal measures the amounts of *non-food* goods for trade (eg. coral, fish for aquarium, etc) in your study area. Data for both goals are often recorded in the same sources. It may be time-saving for the goal keepers for these two goals to join efforts to gather data.

General Approach

Ideally, you would have information about the quantity of species caught or harvested (tonnes), the effort involved (particularly important for the Fisheries sub-goal), the practices used (fishing gear, mariculture inputs, extraction methods), and the spatial extent where the practices occur (fishing locations, mariculture farms). Regarding the species caught or harvested in your area, in Global Assessments, information by species for fisheries and mariculture are processed separately before being combined to calculate status scores.

Goal models should also incorporate a *indicator of the sustainability of each practice*. In the case of *Fisheries* sub-goal, Maximum Sustainable Yield (MSY) is used as a benchmark, calculated from available harvest (and efforts) information.

For *Mariculture*, Mariculture Sustainability Index (MSI) is used in the model. In the global assessment, MSI is calculated based on three sub-indices that directly measured long-term renewability of a given mariculture

practice: the wastewater treatment index, the origin of feed index (i.e. fishmeal or other) and the origin of seed (i.e. hatchery or wild caught).

The *overall Food Provision* model should not change; it is a combination of fisheries and mariculture scores, with the contribution of each type of practice to the overall score is weighted by its relative contribution to the total seafood yield. This assumes that one tonne of seafood by any method is the same as any other tonne of seafood. It would be possible to work with different assumptions and apportion weighting differently.

Pressures and Resilience measures must be considered as well. What pressures *act on the harvesting of these species?* If information allows, pressures and resilience measures can act on different species, group of species, or practices separately, as is done with Natural Products. Fishing or harvesting practices can also act as pressures for other goals, for example destructive fishing practices can impact habitat-based goals and genetic escapes from mariculture practices can affect Fisheries and Biodiversity.

9.3.6.1 Fisheries sub-goal The Fisheries sub-goal describes the amount of wild-caught seafood harvested and its sustainability for human consumption. The model generally compares landings with Maximum Sustainable Yield. A score of 100 means the country or region is harvesting seafood to the ecosystem's production potential in an sustainable manner.

Practical Guidance

Fisheries science is a discipline that in part aims to estimate the amount of fish that can sustainably be extracted from the sea. For this reason, it is important to consult with fisheries experts in your study area. Fisheries experts will be able to advise how to best estimate the maximum amount of catch that can be sustainably fished, and the information available will determine what type of modeling to take. There are many different modeling approaches, and most are based on either catch alone, or catch-per-unit-effort.

The Ocean Health Index relies on freely accessible, open-source data that is consistently updated over time. Here are two datasets used in the OHI-Global assessments, which could be also be useful for OHI+ assessments: Fisheries catch over time from the [Sea Around Us Project](#) (seaaroundus.org) and Global fisheries landings over time from [FAO](#) (fao.org/fishery/statistics/en).

If only catch data are available, it is highly recommended to follow the approach in the *2013 global assessment* (Halpern *et al.* 2015) rather than the 2012 global assessment (Halpern *et al.* 2012). All global assessments use national fisheries catches reported to the Food and Agricultural Organization (FAO), and the 2013 global assessment used the fisheries modeling method for data-poor sources developed by *Martell & Froese* (2013). With this method, fisheries catch information would be used to calculate the population biomass (B), and its maximum sustainable yield (BMSY). The reference functional relationship between fisheries catch and effort information would be used to calculate the present biomass against BMSY would be used to set the reference point. The current status would be calculated using the present state of every individual species and combining each species together as the weighted proportion of the total catch.

You can find data for catch-per-unit effort data, and then create a functional relationship to determine the reference point.

At a global scale, catch, effort, and MSY estimates are not available for either commercial, artisanal or recreational fishing: only landings data for commercial fisheries are available through the United Nations Food and Agriculture program (UN FAO). You will hopefully be able to find more localized data when conducting your assessment.

When collecting data on fish landings, it's important to consider how you will divide the data among regions. You should try to assess each fish species by its entire population across all regions in your study area. The status in the global assessment model (2013) was calculated based on estimating population biomass relative to the biomass that can deliver maximum sustainable yield for each landed stock (B/BMSY). This ratio is conventionally used to inform fisheries management. This approach adopts the population biomass at MSY (BMSY) as a single-species reference point.

If you are replicating models from the global assessment (2013-2015), do not split the catch among regions; instead, you want to sum catch across all regions so you can calculate B/BMSY for the whole population.

The principle of the **reference point** should not change. You should be creating models that penalize scores for harvesting above the maximum sustainable yield, as defined in your assessment, and scores that penalize for harvesting below the sustainable yield. The penalties vary for models developed in the global assessments, where overfished species negatively influence scores more than under-fished species do.

It is important to also consider *buffering around the reference point* (eg. 75% of BMSY) because of imperfect knowledge about the data. Part of this depends on the type of assumptions you want to make about the ecology of fish species in your area and the impacts upon them from fishing practices. For instance, when all species are exploited simultaneously, fishing pressure on each population might be lower due to changes in interactions between species that occur when a predator population is reduced.

9.3.6.1.1 A note about methods used in the fisheries goal Since Global 2012, several new data-poor approaches have been developed to assess fisheries that leverage globally-available information (Costello *et al.*, 2012; Martell & Froese, 2013; Thorson *et al.*, 2013). The estimates of B/BMSY used in Global 2013 were obtained by applying a model developed by Martell & Froese, (2013), and referred to as the “catch-MSY” method. This approach adopts the population biomass at MSY (BMSY) as a single-species reference point.

The catch-MSY approach improves upon the method used in Global 2012 in that it leverages a mechanistic understanding of the connection between harvest dynamics and population dynamics and uses this to infer stock depletion levels as a function of both historical patterns in catch and of species-specific resilience traits (Thorson *et al.* 2013). In addition, this model is more informative in the case of developing fisheries, whereas the Global 2012 approach assumed a perfect score in cases where a peak with successive decline had yet to be observed.

Although it is a data-limited method, the more complex approach better takes into account species-specific fishery dynamics. The scores for each population were also combined using a geometric mean, which ensures that smaller, rarer populations have more weight so that the biodiversity of the catch is taken into account as well.

9.3.6.1.2 Calculating the fisheries goal Calculating the fisheries goal using methods from global assessments (2013-2015 as it is not recommended to use methods from the 2012 global assessment) relies on catch or catch-per-unit-effort data. The fisheries goal model uses mean catch and B/BMSY for all species. These are two separate steps: the mean catch is not used to calculate B/BMSY. Once you have B/BMSY for all the species, you will combine them together to get a single fisheries score for each of the regions in your study area. This is where the mean catch is used: it is a weighting factor so that species with higher mean catch will contribute more to the final score. You will not need to calculate this because the Toolbox calculates the weighted mean when you provide mean catch as an input layer to the Toolbox. The input layer should simply be the mean catch for each species.

When calculating B/BMSY and mean catch, use as much information as possible: all the years available. Species that have few years of available data will likely be less accurate and it is important to document this. It is admissible to have different ranges of years for different species, since too much information would be lost if all species required the same species range. However, interpreting why different species have different years of time series is important. There may be data for 30 years of history for some species, 10 for others and 5 or 6 for others. It is important to know whether there is only 6 years of data because it is a recent fishery that developed in the last 6 years; in this case 6 years is enough. But if only 6 years of data are available because they stopped collecting the data in recent years, you might get a completely misleading assessment. And of course if you know there was no catch at all in recent years, it is important to include those recent years as 0's. 0's must also be included when calculating B/BMSY or the results will be nonsensical.

How B/BMSY is calculated for global assessments is a bit unsatisfactory and is in the process of being better developed. This is because the global model is very imprecise, which also affects how to interpret B/BMSY

results. It is important to consult with a fisheries scientist in your study area, as they will have expertise with the information and knowledge available in the local context. If possible, we suggest calculating the scores using a more precise model as well, so fisheries experts can assess whether results look reasonable. This is important so that the scores produced have credibility.

| Assessment | Developing the Model |
|------------------------|--|
| Global 2012 | The status was calculated as a function of the absolute difference between a region's total land-based biomass and its maximum sustainable yield (BMSY). |
| Global 2013 | The status was calculated based on estimating population biomass relative to the biomass that would be required to reach BMSY. |
| Brazil (2014) | The status was calculated in the same manner as Global 2012, with a modified sustainability threshold. |
| U.S. West Coast (2014) | The status was based on B/BMSY for each landed stock and fishing mortality that can deliver BMSY. |

9.3.6.2 Examples of the Approach

9.3.6.3 Mariculture sub-goal Mariculture measures the ability to reach the highest levels of seafood gained from farm-raised facilities without damaging the ocean's ability to provide fish sustainably now and in the future.

Higher scores reflect high food provisioning in a sustainable manner, while not compromising the water quality in the farmed area and not relying on wild populations to feed or replenish the cultivated species. A score of 100 means that a region is sustainably harvesting the greatest amount of farmed seafood possible based on its own potential (where its maximum potential is estimated in different ways depending on the assessment).

Ideal Approach

Ideally, you would find information about the *harvest and practices*, on the *total area* available for mariculture. This could be in terms of a physical area or area based on siting priorities. You would also find information on the *sustainability* of the mariculture practices. This is important because mariculture competes for space with many other ocean uses, including fishing, tourism, and other activities.

This approach would not penalize regions that have less geographic area available for mariculture, though places with fewer sheltered bays or lower primary production could be at a disadvantage.

Culture of marine fish in closed pools on-shore should not be included because onshore aquaculture does not require a marine environment. However, if the aquaculture farms receive seawater supply and return seawater back to the sea, they should be included in the Mariculture sub-goal. Finer-scale, environmental damages due to the intake pipes and the processed brine back into the marine system could be incorporated as a pressure layer as well.

Fish that are used as feed for other fish (e.g., sprat) would be more appropriate to include in Natural Products. This is because they are not being consumed directly. Fish such as sprat, for example, may be used to feed pigs in addition to other fish, and therefore you would need to know how much (tonnage) is being produced, and where it is going to be able to accurately distinguish these categories to avoid double-counting.

Practical Guidance

Besides harvest information, either by species or total amount, this goal requires *spatial information*. You would have to find maps determining where coastal activities are allowed, or find other kind of data that tells you the usage restrictions for activities in your waters. For example, you can look for assessments that have been made to identify the coastal and offshore habitats that are appropriate for each intended type of mariculture species.

The 2012 Global Assessment included the entire area of the coastline assuming that mariculture could be developed everywhere. This was done because there was no information about limitations for how much coastal area could be allotted to mariculture.

The **reference point** will depend on your context. You should try to account for the full potential extent of mariculture in your marine areas. This could be based on a number of variables such as the habitat suitability for each cultured species, the distance from the coast, information on how local preferences favor the allotment of marine space to mariculture versus sports, hotels, beaches, tourism, or other uses. This definition of mariculture potential therefore can be physical or social. Keep in mind that in the Global approach, a low score can indicate one of two things – that species are being farmed in an unsustainable manner or that regions are not maximizing the potential to farm in their marine territorial waters.

It is important to think about the meaning of your reference point. The spatial reference point in the Global approach assumes that because one country is able to produce such a high proportion of mariculture to its coastal area, all regions should be able to do well, given current ecological and technological conditions. This also assumes that it is socially desirable for all regions to produce farmed species at this level, which is likely not true for all regions. The Global approach compared observed production density of all areas to the highest observed production density after the sustainability coefficient was applied.

There was a significant change between Global 2012 and Global 2013 to improve the reference point for this sub-goal. The coastal population was factored in to the reference point in 2013, where the harvested tonnes per inhabitants within 25 km of the shore was considered. This change was done under the assumption that production depends on the presence of coastal communities who can provide the labor force, infrastructures, and economic demand to support the development and economic viability of mariculture facilities. Thus, two regions with an equal number of coastal inhabitants harvesting an equal tonnage of cultured seafood should score the same, even if one is larger than the other, as the productivity is commensurate to each region's socio-economic potential to develop mariculture. This is an example of how improvements in methods occur over time.

| Assessment | Developing the Model |
|-------------------------------|--|
| Global 2012 | Mariculture was calculated as the yield reported multiplied by the sustainability for each species. |
| Global 2013 | This study used a model similar to the Global 2012. |
| Brazil (2014) | The status was calculated using harvest data reported by the Brazilian Institute of the Environment and Renewable Natural Resources. |
| U.S. West Coast (2014) | The status was calculated as the sustainable production density of shellfish biomass from mariculture. |
| China (2015) | Status model is similar to the 2012 global assessment. But the status model was based on tonnes harvested. |

9.3.6.3.1 Examples of the Approach

9.3.7 Livelihoods and Economies

The jobs and revenue produced from marine-related industries are clearly of huge value to many people, even for those people who do not directly participate in marine-related industries. People value community identity, tax revenue, and indirect economic and social impacts of a stable coastal economy. This goal tracks the **number and quality of jobs and the amount of revenue produced across as many marine-related industries/sectors as possible** through two sub-goals, Livelihoods and Economies. A score of 100 reflects productive coastal economies that avoid the loss of ocean-dependent livelihoods while maximizing livelihood quality.

In the global study, this goal does not attempt to capture any aspects of *job identity* (i.e. the reputation, desirability or other social or cultural perspectives associated with different jobs). We make the assumption

that all marine-related jobs are equivalent, such that, for example, a fisherman could transition to a job in mariculture or ship-building without affecting the score of this goal. There are also not adequate data to assess workers' *job satisfaction* or *ecological sustainability* of sectors. Future, finer scale applications of the Index could incorporate these key considerations.

Oil and gas industries typically are not included in this goal. 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. However, oil and gas industries are considered a large employer and often generate large revenue for a region. Similarly, industries such as shipping, boat building, ports and harbors do not directly depend on a healthy ocean, but the demand for some of those boats (fishing boats, sailboats, yachts) is dependent on a healthy ocean and are also important providers of jobs and revenue. Ultimately, whether and how to include those data is a decision that should be made on a case-by-case basis by the individual OHI+ group.

9.3.7.1 Sub-goal: Livelihoods This sub-goal describes livelihood quantity and quality for people living on the coast. Livelihoods includes two equally important sub-components, the number of jobs, which is a proxy for livelihood quantity, and the per capita average annual wages, which is a proxy for job quality.

Ideal Approach

Ideally, this sub-goal would speak to the quality and quantity of marine jobs in an area. It would encompass all the marine sectors that supply jobs and wages to coastal communities, incorporating information on the *sustainability* of different sectors while also telling about the *working conditions and job satisfaction*. The jobs and revenue produced from marine-related industries directly benefit those who are employed, and also those who gain indirect value from related economic and social impacts of a stable coastal economy, such as community identity and tax revenue. You should capture the indirect as well as direct benefits from jobs, wages and revenue from coastal communities.

In your assessment, there is the opportunity to study the behavior of economic trends in your area. You can examine time-series with greater detail and, for example, establish a different time-periods that reflect economic cycles in your area, or even process the data to eliminate the “noise” from fluctuations and capture more persistent trends.

You will most likely simplify the models from the global assessments. Models in the global assessments adjust for currency differences using the Consumer Price Index and Purchasing Power Parity, which you likely will not need to do.

Practical Guidance

The first step of this goal is to **identify the marine-related sectors in your area**. These sectors are important because the *Livelihoods* goal includes information about **jobs** and **wages** for those sectors. There are jobs that are directly connected to the marine environment, such as shipping, fishing, longshore workers, but also some that are connected indirectly, such as suppliers and supporting industries. For example, the sectors for which data were found in global assessments included tourism, commercial fishing, marine mammal watching, aquarium fishing, water and tidal energy jobs, mariculture, transportation and shipping, ports and harbors, ship and boatbuilding.

After you have identified which jobs are in your area, you will want to find some **measure of their direct and indirect benefits**. Direct benefits of jobs include the number of jobs in each area, and the wages or income for such jobs. You could find such information from your local national statistical office, or economics bureaus, for example. Indirect benefits of these jobs to the local communities are calculated through the use

of *economic multipliers*, with which you can attempt to estimate the revenue generated by jobs more broadly associated with marine sectors. It's encouraged to use economic multipliers from the literature.

One example to use economic multipliers is to multiply the number of fishermen by a multiplier to estimate larger economic effects, ranging from gear manufacturing companies to restaurants and movie theaters where the fishermen spend their income.

Next you must think about how to use the data to infer **quality and quantity of jobs**. Do you have data through time so that you can see how the number of jobs or the wages per sector have changed over time? If all of the sectors change in the same way, for instance, this might show broader economic trends that you should account for.

It is highly recommended that you set the **reference point** for *jobs* as a *temporal comparison* (eg. the number of jobs five years before), and for wages as a *spatial comparison* (eg. highest observed value across reporting units). Comparing the number of jobs across different places, for instance, would require at the very least adjusting values by the size of the workforce in each location.

Examples of the Approaches

| Assessment | Developing the Model |
|-------------------------------|---|
| Global (2012) | This was measured as the number of direct and indirect jobs across sectors within a region plus |
| Global (2013-2015) | The model was similar to Global 2012, with some simplifications. |
| Brazil (2014) | The method was the same as Global 2012. |
| U.S. West Coast (2014) | This goal follows the same model as in Global 2012, using local data for the sectors of living resources. |
| China (2015) | Status model is based on the number of direct jobs across marine sectors and the average |

9.3.7.2 Sub-goal: Economies Economies captures the economic value associated with marine industries using revenue from marine sectors. It is composed of a single component, revenue.

Ideal Approach

Ideally, revenue data would be collected for all coastal regions, and traced from sectors both directly and indirectly related to marine industries. When these data are not available it is possible to use revenue data at a larger scale and adapt them to a coastal area based on the population distribution. The reference point in this sub-goal will likely be set with a moving-window temporal approach.

Practical Guidance

A number of sectors were not included in the Global Assessments because sufficient data did not exist. However, it might be desirable to include sectors such as ecotourism, sailing, kayaking, surfing, and offshore wind energy production, and scientific research, among others. It may be practical to include the same sectors in the *Livelihoods* and the *Economies* subgoals.

The quantity of Wild-caught fisheries, Mariculture, and participation of Tourism & Recreation are considered separately in specific goals, whereas the monetary component is captured in L&E.

In the global assessment, revenue has a moving-target temporal **reference point** by comparing the value in the current year to values in previous years. Global assessments used a five-year moving-window because it is intended to capture short-term changes in the trajectory. But then you must consider if that amount of time would represent economic trends. If there is a not economic downturn, do you want to reward an increasing number of jobs, or reward maintaining the same number of jobs?

We highly recommend that this remains a temporal comparison so that a specific place is compared to its performance in the past and not to anywhere else.

You may find resilience metrics that can be used throughout your assessment. There are many economic indices and some might be appropriate for your area. For example, the *Global Competitiveness Index* was used in global assessments as a resilience for this goal because it means that marine jobs are more likely to be maintained if an area gets a good score on the Competitiveness Index.

Examples of the Approaches

| Assessment | Developing the Model |
|-------------------------------|---|
| Global (2012) | The status used the total adjusted revenue generated directly and indirectly from each sector |
| Global (2013-2015) | The model was same as Global 2012, with a few simplifications; revenue data were adjusted by |
| Brazil (2014) | The method was the same as Global 2012. |
| U.S. West Coast (2014) | The method was the same as Global 2012, but with local sectors represented. |
| China (2015) | Status model is the same as in global assessments |

9.3.7.3 Natural Products In many countries the harvest of non-food natural products is important for local economies and can also be traded internationally. The sustainable harvest of these products is therefore an important component of a healthy ocean. **This goal assesses the ability of countries to maximize the sustainable harvest of living marine resources, such as corals, shells, seaweeds, and fish for the aquarium trade.**

It does not include bioprospecting which focuses on potential (and largely unknowable and potentially infinite) value rather than current realized value, or non-living products such as oil and gas or mining products which by definition are not sustainable.

Similar to Food Provisions, this goal is trying to capture the value of goods sustainably harvested by the ocean. While Food Provision focuses on goods for human consumption, Natural Products assesses non-food goods. As these two goals often share similar data sources, it may be time-saving for the goal keepers of these two goals to work together and gather data.

Ideal Approach

Ideally, *quantity, value, and a sustainability rating of the harvest method* would be available for every marine and coastally-derived natural product within the regions of a study area. This could include a wide range of non-food products depending on what is harvested in the study area, including corals, shells, seaweeds, aquarium fish, and mangrove wood. The ideal reference point would be derived from a functional relationship of the sustainability of the harvest for each product relative to the amount of product available in the ecosystem, informed by scientific studies. Without such information, assumptions and expert judgment will need to be made to set the reference point.

Practical Guidance

Whether you use the approach from the global assessment or are developing your own new model entirely, there are a few tasks that will remain the same because are key to the philosophy of this goal.

The first is to identify **identify which products are in your study area**. For example, does your study area have corals, ornamental fishes, sponges? Does your area yield medicines from the sea, or other products that are not used for nutrition under *Food Provision*? Does your area harvest drinking water from the ocean through desalination plants? Is there a kelp or seaweed industry in your area? If there are multiple uses of the product, you must also consider what proportion of the product is used for food, and what proportion is used for other purposes. As another example, oil from marine mammals was considered but excluded from

the global models, but if a region has a considerable amount of mammal oil harvest, they should include it in the calculation, keeping in mind that the sustainability of this type of harvest is likely to be low and should be reflected in the score.

In the global assessments, Natural Products data come from the [UN's Food and Agriculture Administration] (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.

If natural products are all produced through on-land aquaculture, you would probably have good reason to exclude the natural product goal. This would also depend on the origin of these natural products—are they from the region's waters? You should consult with experts and any reports/papers on the topic. In addition, this 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.

The second task is to think about **where these products are harvested and how much of them are harvested** in these areas through a period of time. You should find spatial representation of these products, which can be done by knowing where they are derived from. Do they come from certain habitats (in the case of coral) or animals (in the case of fish oil)? This information will help calculate the sustainability of the harvest of the natural product. Harvest amounts and the spatial data are used to calculate *exposure* further on, and can also be used to set the *relative weighting* between the products. These spatial data may have already been used in other goals, or they may lead you to find useful data that can be used in other parts of the assessment.

The second task is to think about **where these products are harvested and how much of them are harvested** in these areas through a period of time. You will have to assign geographic representation of these products, which can be done by knowing where they are derived from, ideally, or by assigning relative weightings. Do they come from certain habitats (in the case of coral) or animals (in the case of fish oil)? This information will help calculate the sustainability of the harvest of each natural product. Harvest amounts and the spatial data are used to calculate *exposure* further on, and can also be used to set the *relative weighting* between the products. These spatial data may have already been used in other goals, or they may lead you to find useful data that can be used in other parts of the assessment.

The third component is to try to find the **sustainability** coefficients of the identified products. It is possible to measure sustainability in a number of different ways. Quantitative information can be used, or expert judgment, perhaps based on information or rough estimates of how sustainable the harvest method is, which is what was done in Global 2012. We based the sustainability component on the historical maximum harvest recorded, the maximum harvesting density recorded, and risk status assessments by the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). In the absence of these, we borrowed general principles from fisheries models to provide rough estimates. If these are given values you could simplify the model, or they could be derived from two factors, *exposure and risk*. The *exposure* will come from the spatio-temporal harvest amount data already prepared, and the *risk* will come from the scientific literature or a developed indicator. For both of those cases, the values can be calculated in separate equations as part of your data preparation process.

Global assessments borrow principles from fisheries science to make estimates of product sustainability. In the Global 2013 assessment the sustainability component was derived from the historical maximum harvest recorded, the maximum harvesting density recorded, and risk status assessments by the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES).

One very important thing to consider at this point is your **reference point for the relative harvest amount**. The relative harvest of your data is multiplied by the sustainability coefficient in the last step.

Setting the reference point is a decision your team must make based on the available data and an inferred functional relationship between the harvest of the product and the amount in the system. Understanding the patterns in harvest can help inform how to set the reference point. For example, knowing whether harvesting effort was constant or whether product yields changed due to the market demand and not the availability. This information could help inform whether it is more appropriate to set the reference point as the peak yield of the time-series, or some percentage above or below, or some other approach that is both ambitious and realistic (*SMART* principles). The decision you make for the reference point should be based on the trend of the data; for instance, if your harvests have only increased over time, which may be indicative of an emerging economy, you will have to account for that.

Examples of the Approach

| Assessment | Developing the Model |
|-------------------------------|---|
| Global 2012 | The products used were coral, ornamental fish, fish oil, seaweeds and marine plants, shells, an |
| Global 2013 | The goal model had the same approach as Global 2012, with updated data processing. |
| Brazil (2014) | The method was the same as Global 2012. |
| U.S. West Coast (2014) | This goal was not included in this assessment due to lack of data availability. |
| China (2015) | Status model is the same as in global assessments. Three natural products were assessed: sea |

9.3.8 Tourism and Recreation

Tourism and recreation in coastal areas is a major component of thriving coastal communities and a measure of how much people value ocean systems, i.e. by traveling to coastal and ocean areas. This goal is not about the revenue or livelihoods that are generated by tourism and recreation (that is captured in the livelihoods goal, section 6F above) but instead captures **the value that people have for experiencing and enjoying coastal areas**. A score of 100 means a region utilizes its full recreational potential without harming the ecosystem.

Ideal Approach

Ideally, you would find information for how the ocean in your area is used and enjoyed by local residents and tourists alike. How many international or domestic visitors? When are they visiting? How long are they staying? What activities are they enjoying? Your model should capture the full range of values and touristic and recreational activities. A measure of sustainability should also be included in the model. The type of reference point used will depend on the data available.

Practical Guidance

First look for information that directly reflect how people enjoy the ocean, such as boat rentals, resort registries, whale watching etc. If you can't find information on visitors, look for proxy data that are indirectly related. For example, do people have access to boating areas, or to surfing spots? Are tourism-related industries (eg. hotel employment, restaurant) increasing or decreasing?

This goal demonstrates the flexibility of the OHI+ approach and will necessarily draw from different data sources than the Global Assessment. It is encouraged to think creatively and to look at what other OHI+ assessments have done when developing this goal model. For example, in the Brazil Assessment the density of hotel employees per state was used as a metric to determine how well touristied coastal areas were. This was better than using international travel information, as was used in the Global Assessment, because for a large country like Brazil, internal travel would not have been accounted for.

The best way to represent the sustainability term for this goal is to use a local indicator or a measure of tourism sustainability or competitiveness. Otherwise use the TCCI value from the Global 2013 assessment for the study area (applied evenly across all regions).

The **reference point** used will depend upon the types of data. Does your country have industry growth rate targets? Are there measures on sustainable levels of tourism and recreational activities based on ecological factors? Do you want to increase tourism, or instead ensure it does not decline?

Examples of the Approach

| Assessment | Developing the Model |
|-------------------------------|---|
| Global 2012 | This goal measured the number of international tourists arriving by airline to coastal regions, and the number of international tourists arriving by land to coastal regions. |
| Global 2013 | The study used the direct employment in the tourism industry relative to total labor force and the number of international tourists arriving by land to coastal regions. |
| Brazil (2014) | The model developed for Global 2012 was changed to use information on hotel employees for coastal areas. |
| U.S. West Coast (2014) | There were data available for changes in participation in 19 different marine and coastal species. |
| China (2015) | Status model is based on the ratio of visitors and coastal area. |

9.3.9 Sense of Place

This goal tries to capture **the aspects of the coastal and marine system that people value as part of their cultural identity**. This definition includes people living near the ocean and those who live far from it but still derive a sense of identity or value from knowing particular places or species exist. We divided this goal into two sub-goals: Iconic Species and Lasting Special Places, and weighted them equally when combining to create a single goal score. A score of 100 means the species and places important for cultural identity are protected and conserved.

9.3.9.1 Sub-goal: Iconic Species Iconic species are those that are relevant to local cultural identity through a species' relationship to one or more of the following: 1) traditional activities such as fishing, hunting or commerce; 2) local ethnic or religious practices; 3) existence value; and 4) locally-recognized aesthetic value (e.g., touristic attractions/common subjects for art such as whales). Habitat-forming species are not included in this definition of iconic species, nor are species that are harvested solely for economic or utilitarian purposes (even though they may be iconic to a sector or individual). This sub-goal assesses how well those species are conserved.

Data for this goal will likely have considerable overlap with *Biodiversity sub-goal: Species*. It will be effective for the goal keepers of Biodiversity and Sense of Place to share the data searching efforts.

Ideal Approach

Ideally, you would have a list of species that are valued as iconic and information on their *habitat ranges* along with scientific studies that speak to the *health* of their populations. A reference point is where all iconic species are at a healthy level of existence.

Practical Guidance

First you need to *identify the Iconic Species*. In practice, Iconic Species are usually a subset of the broader list of species in an area, and so you should be able to find Iconic Species after having found assessed species data for the *Species sub-goal of the Biodiversity goal*. (See *Biodiversity* for more detailed instructions.) Once you have the full list of assessed species, you can determine a subset for *Iconic Species*. For instance, are there known “indicator species” in your area? Are there species that are culturally held as valuable? Do any species appear on the currency or postage stamps?

Since different species are be iconic to different groups, defining which species are iconic can be challenging when it's a cultural question. You might have to find information from experts on local customs and tradition.

The choice of inclusion of iconic species in your list can be subjective. You could also come up with specific inclusion criteria, for instance, that would filter a list of species or filter a subset of the gathered data for

Biodiversity. This would be a more rigorous approach because then it could be documented and you could replicate the study in future assessments.

Ultimately, almost any species can be iconic to someone, and so the intent with this goal was to focus on those species widely seen as iconic within a country, and iconic from a cultural or existence value (rather than for a livelihoods or extractive reason). Many lists exist for globally important, threatened, endemic, etc. species, but in all cases it is not clear if or to what extent these species represent culturally iconic species. In the global assessment, species were drawn from the World Wildlife Fund's global and regional lists for Priority Species (especially important to people for their health, livelihoods, and/or culture) and Flagship Species ('charismatic' and/or well-known). These lists are note the only source that included cultural reasons for listing iconic species and you should try to acquire more information on locally relevant species.

After identifying the list of iconic species, you need to find a matrix to reflect the *health* of their population. In the global study, we used IUCN threat categories and weights. The reference point is to have the risk status of all assessed species as Least Concern. Alternatively, you can use a different kind of assessment approach to see if the populations are healthy, which could be indicated, for instance, by the stability of their populations.

Examples of the Approach

| Assessment | Developing the Model |
|-------------------------------|--|
| Global 2012 | The status was the average extinction risk of iconic species, calculated as the weighted sum of |
| Global 2013 | The method was the same as Global 2012. |
| Brazil (2014) | The method was the same as Global 2012. |
| U.S. West Coast (2014) | This study replaced the global IUCN risk assessments with regionally-specific species assessments. |
| China (2015) | Status model is similar to global assessments. However, average extinction risk of all assessed |

9.3.9.2 Sub-goal: Lasting Special Places This sub-goal focuses on those geographic locations that hold particular value for aesthetic, spiritual, cultural, recreational or existence reasons, and assesses how well they are protected.

This sub-goal makes use of protected areas, and some of the same information gathered on projected areas can be used to create **resilience** data layers like Marine Protected Areas.

Ideal Approach

Ideally, you would be able to survey everyone in your region and produce a list of all the places they consider special, and then assess how well they are protected relative to a desired state. How well they are protected could be the percentage of area protected, and you could also find how well they are protected using other data. This sub-goal could also be based on the extent to which people participate in spiritual or religious activities in an area.

Practical Guidance

The ideal list of special places don't exist in practice, and this is a difficult goal to express accurately, since it attempts to capture how people interact culturally with their coastal places. You can define "special" with specific criteria. A good example to look at is the *U.S. West Coast assessment (2014)* and *Brazil Assessment (2014)*.

More likely, you would follow the global assessment and use *lists of protected areas* as catalogues of special places, and the area of designated protected places relative to a **reference point** (eg. of thirty percent coastal area protected) is used as a measure. Coastal area could be based off a 1 square km buffer inland, as in the Global Assessment, or it could be based on what is reasonable to your area. In any case, you would want to consider how far out from shore you should include as well; would it be 3 nautical miles, or as far as your territorial waters up to 12 nautical miles?

Data sources should be specific to your region. International databases, like the [World Database of Protected Areas](http://www.protectedplanet.net/) (www.protectedplanet.net/), offer rich information, but they may not be as up-to-date as the list of national parks in your area, and may not have as much information on the quality of protection. If you have more information on quality, you could think about another approach than the thirty percent reference point target.

Global assessment incorporated both marine and terrestrial protected areas. However, if only data on marine protected areas is available, you can calculate only the marine component of the goal. This is not ideal but OHI is flexible to work with the data available.

If your assessment involves multiple categories (e.g., antiquities, MPAs, beaches of special interest), whether you group them together or calculate each category separately depends on reference points. Maybe you want 10% of offshore water to be in MPAs, but only 5% of coastlines to be beaches and 3% Antiquities, for example; in this case, you would calculate them separately and then add them together. But if you want 10% of your country's coast to have any combination of these things, you would keep them together.

In the Global Assessment framework, the **Sense of Place** sub-goals were weighted equally and combined in an average to create a single goal score. The two sub-goals are averaged currently in the framework. But these could be combined with a weighted average.

Examples of the Approach

| Assessment | Developing the Model |
|------------------------|--|
| Global 2012 | The status was calculated by combining the percent of coastal waters that are coastal marine |
| Global 2013 | The method was the same as Global 2012. |
| Brazil (2014) | The method was the same as Global 2012. |
| U.S. West Coast (2014) | The model was the same as Global 2012. |
| China (2015) | Status model is similar to global assessments |

9.4 Pressures and resilience

Pressures and Resilience are two of the four dimensions used to evaluate each goal or sub-goal, along with **Status** and **Trend**.

It is important to identify the pressures that affect the ocean and coastal systems in your study area, and to search for additional pressures not included in the global assessments. Once you have identified pressures within your study area, you should identify what resilience measures could counteract or nullify those pressures. Alternatively, you can start with a list of known resiliences, such as the relevant environmental laws in your country, and then map them onto pressure layers.

TIP: The same considerations and requirements about data presented in the “**Gathering Appropriate Data**” section also apply to pressures and resilience. Every measure you include for pressures and resilience requires data for each region in your assessment.

Ideally, every stressor with an identified strong impact should have a corresponding resilience measure. The rationale is that as resiliences in the study area increase (for instance, by improving environmental regulations), they would balance out and eventually neutralize the pressures. This would lead to an increase in the overall goal or sub-goal score. By including regulatory responses in your assessment, you ensure that the actions taken in your country are relevant to ocean health.

In practice, however, the pressures and resilience measures you include in your assessment will be highly determined by data availability. It is best to *first consider what pressures are acting in your study area and*

then determine if data are available to measure them. You should also decide if the pressures data included in the global assessment are relevant for your assessment and determine if local data better capture pressures for all the regions in your study area. When considering resilience measures, look for regulations or indicators that could be encompassed in one of the pressures categories.

9.4.1 Pressure

The Toolbox calculates pressures in five **ecological pressure categories** (eg. pollution, habitat destruction, fishing pressure, species pollution, and climate change) and one **social pressure category** (eg. Weakness of Governance Indicator). Under each category, you could have multiple pressure data layers. For example, the “pollution” category could include pathogen, nutrients, and chemicals pollution.

The reason behind the ecological categories is to avoid hidden weighting (e.g., overrepresentation of pressures for which there is more data). For example, in the global assessment there were many pollution datasets available, but few distinct habitat destruction datasets. If we simply averaged the scores of each individual stressor, pollution scores would have a greater influence on the results (stronger weight) due to the relative higher availability of measurements of various pollutants. Instead, aggregating by pressure categories ensures that different stressor types influence the score based on ranks. Nonetheless, the scores are combined in a cumulative way within each category to account for the fact that multiple stressors within a category have a cumulative impact that is greater than if only one of the stressors were present. The resulting scores for the five ecological categories are averaged to produce a single ecological pressures score. This score is then averaged with the social pressures score to produce the final overall pressure score.

There are two types of pressures data you need to provide for the toolbox: *local data layers for each pressure* and *a pressure matrix*.

9.4.1.1 Pressures data layers The pressures you will include in your assessment will depend on what is important in your study area and what data are available. If local pressures data are not available, you may default to using data from the global assessment, but this means in most cases that you will not have different information for each region. You will determine the weight ranks required in the pressures matrix only after you have identified the data you will include.

The following steps outline the process of how to identify pressures in your assessment. The steps are iterative; return to previous steps to ensure you capture all important pressures in your study area:

1. Begin by exploring pressures important to your study area. What are big stressors acting along your coastlines?
2. Are data available to measure these stressors? If not, are other indirect measures or proxies available to represent these stressors?
3. Evaluate the pressures included in the global assessment. For example, if there is no mariculture in your study area, you could remove pressures data layers that only affect this goal (i.e. genetic escapes).
4. Are all of them relevant? Are there local data that can be substituted in the place of global data?
5. Determine the pressure category for any additional stressors in your study area, and add it to the pressures matrix.
6. When all stressors are included in the pressures matrix, determine which goals it affects. Then, determine the weight rankings of all stressors for each goal. Use literature and expert judgement to determine this.
7. Prepare each pressure data layer as described in this manual only after steps 1-6 are completed. In addition to the proper formatting for the Toolbox, *pressures data must be rescaled (normalized) on a unitless scale from 0 - 1, where 0 is no stressor at all and 1 is the highest possible value for the stressor, or the value at which the goal achievement is completely impaired. You will have to determine how to rescale the data, whether it is based on the highest value in the data range or other methods.*

9.4.1.1.1 Including pressures from global assessments If you are not able to find local data for stressors, you may use the data from the global assessments for your country. For most of the stressors, this means that there will not be differences between the regions within your study area. However, several stressors included in the global assessment are based on spatial data at high resolution from previous work by Halpern *et al.* (2008) in *Science: A global map of human impact on marine ecosystems*. These data are available at a resolution of 1 km² for the entire global ocean, and can be extracted for the regions in your study area. The stressors available at 1km² resolution are indicated below with ** ** **.

Table of pressures layers and descriptions

| layer | name |
|--------------------|---|
| cc_acid** | Ocean acidification |
| cc_slr** | Sea level rise |
| cc_sst** | Sea surface temperature (SST) anomalies |
| cc_uv** | UV radiation |
| fp_art_hb | High bycatch caused by artisanal fishing |
| fp_art_lb | Low bycatch caused by artisanal fishing |
| fp_com_hb | High bycatch caused by commercial fishing |
| fp_com_lb | Low bycatch caused by commercial fishing |
| fp_targetharvest | Targeted harvest of cetaceans and sea turtles |
| hd_intertidal | Coastal population density as a proxy for intertidal habitat destruction |
| hd_subtidal_hb | High bycatch artisanal fishing practices as a proxy for subtidal hard bottom habitat destruction |
| hd_subtidal_sb | High bycatch commercial fishing practices as a proxy for subtidal soft bottom habitat destruction |
| po_chemicals** | Ocean-based chemical pollution |
| po_chemicals_3nm** | Land-based chemical pollution |
| po_nutrients** | Ocean nutrient pollution |
| po_nutrients_3nm** | Coastal nutrient pollution |
| po_pathogens | Access to improved sanitation as a proxy for pathogen pollution |
| po_trash | Trash pollution |
| sp_alien | Alien species |
| sp_genetic | Introduced species as a proxy for genetic escapes |
| ss_wgi | Weakness of governance indicated with the WGI |

Note that chemical and nutrient pollution have both land-based (within 3 nautical miles) and ocean-based (within the entire 200 nautical mile EEZ) elements. This is because how pollution affects different goals will depend on the spatial scale of the goal's activity. Some goals occur far from shore, and nutrient and chemical pollution should be included for all offshore waters: FIS, MAR, ECO, and SPP. However, some goals are really only relevant nearshore, so nutrient and chemical pollution should only be included close to the shoreline (3nm in the global study): AO, CS, CP, TR, ICO, LSP, HAB.

These distinctions won't always apply for smaller-scale assessments. For example, in the US West Coast study (Halpern *et al.* 2014), we did not distinguish between offshore and 3nm and therefore only used the `po_nutrients` data layer.

Some pressure data are the same or closely-related to data for goals. For example, the **Wild-Caught Fisheries** goal model requires catch data, which may be the same data source for information on commercial high- and low-bycatch data, which are used as pressures layers that affect **Livelihoods and Economies** and **Biodiversity**. In global assessments, the **Clean Waters** goal is very much linked to pressures layers because the input layers for its status are used as pressure layers. Trash pollution is a pressure that affects **Tourism and Recreation**, **Lasting Special Places**, **Livelihoods and Economies**, and **Species**. It is important to remember these linkages as you go through the data discovery process.

You should also start searching for pressures data independent from data for goals. An example

would be how climate change impacts will appear in various places in your assessment. Climate change pressures layers can include UV radiation, sea surface temperature (SST), sea-level rise (SLR), and ocean acidification, and these impacts might affect such goals as **Natural Products**, **Carbon Storage**, **Coastal Protection**, **Sense of Place**, **Livelihoods and Economies**, and **Biodiversity**. These linkages will become more clear as you go through the OHI+ assessment process.

9.4.1.2 Pressures matrix Independent from local pressure layers, you will develop a Pressures Matrix table that establishes the relationships between stressors and goals, ie. how each pressure measure affects each goal, or an individual habitat type or natural product categories within a goal. It uses a *rank from 1-3 to weight how strongly a given pressure affects a goal or sub-goal relative to all the other pressures affecting it.*

It is easy to confuse the weights with pressure data layers. But pressure weights should not be applied to the regions, only to the goals.

The rank weights used in the pressures matrix were determined by Halpern *et al.* 2012 (*Nature*) based on scientific literature and expert opinion (see Supplemental Table S28 of *Halpern et al. 2012*). For example, 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. In the pressures matrix ranks are categorized as follows:

- 3 = high pressure
- 2 = medium pressure
- 1 = low pressure

Stressors that have no impact are left blank in the matrix rather than being assigned a rank of zero, which would affect the average score. Pressures are ranked rather than being represented as a binary (yes/no) measure because the range of consequence of different pressures on each goal can be quite large, and to classify all those pressures as a simple ‘yes’ would give too much influence to the weakest stressors. For example, food provision is most heavily impacted by unsustainable, high-bycatch fishing, but pollution does have some impact on fish populations. Without a weighting system, these stressors would be treated equally in their impact on the food provision goal.

9.4.1.3 Calculating Pressures Pressures are represented by three-dimensional matrix. The data of each stressor in each region and the rank weights are two dimensions of the 3-D matrix. The Toolbox will: 1. For each stressor, multiply the local data for each region (between 0 and 1) by the weight (NA, 1, 2, or 3) assigned to that pressure for a specific goal and subgoal 2. Within each category, combine all stressor data from step 1 to get a category score 3. Average scores of all categories to get an overall pressures score for that goal or sub-goal

9.4.2 Resilience

Ideally, each pressure should have a corresponding resilience measure, which is meant to ‘balance’ the pressures that negatively effects on ocean health. The Ocean Health Index considers resilience in two categories: **ecological resilience** to address ecological pressures, and **social resilience** to estimate how a region may be able to respond to or prevent environmental challenges. Additionally, **goal-specific regulations** are intended to address ecological pressures, and are measured as laws, regulations, and other institutional measures related to a specific goal. Ideally, for any resilience measure, you would have three tiers of information:

- **Existence of regulations:** Are regulations in place to appropriately address the ecological pressure?

| | Pollution Pressures
Data Layer 1 | Pollution Pressures
Data Layer 2 | Pollution Pressures
Data Layer 3 | Fishing Pressures
Data Layer 1 | Fishing Pressures
Data Layer 2 | Climate Change
Data Layer 1 | Climate Change
Data Layer 2 | ... |
|----------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-----------------------------------|-----------------------------------|--------------------------------|--------------------------------|-----|
| FISHERIES
(Food Provision) | 1 | | | 1 | | | 2 | 2 |
| MARICULTURE
(Food Provision) | 2 | | | | 3 | | | |
| ARTISANAL FISHING
OPPORTUNITY | | 1 | | | 1 | | 1 | 3 |
| CARBON STORAGE
Habitat 1 | | 1 | | | 1 | | | |
| CARBON STORAGE
Habitat 2 | | 1 | | | 2 | | | |
| CARBON STORAGE
Habitat 3 | | 2 | | | 3 | | | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |

Figure 62: Scores from 1-3 are given to rank the importance of each pressure. Only values of 2 or 3 require that a resilience layer be activated when calculating the goal scores.

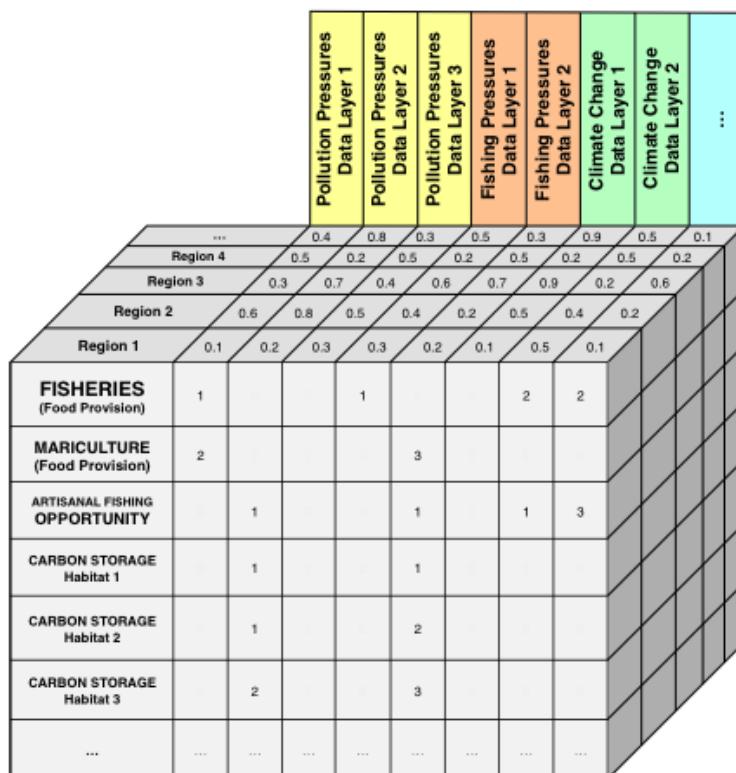


Figure 63: The pressures matrix is three-dimensional: each pressure layer has data per region, which is multiplied by the ranking weights of the pressures matrix.

- **Implementation and enforcement:** Have these regulations been appropriately implemented and are there enforcement mechanisms in place?
- **Effectiveness and compliance:** How effective have the regulations been at mitigating these pressures and is there compliance with these regulations?

Ideally, information would exist for these three tiers, and you would be able to weight the resilience measure based on the quality of the information as 1 (existence of regulation), 2 (implementation and enforcement), or 3 (effectiveness and compliance). This approach is different from the way ranks are assigned in pressures, which is based on impact. However, in most cases, information is not available for these three tiers: often, the existence of regulations is all that is available, and this does not always vary by region. In some cases, you may want to consider building your own set of indicators to determine **implementation and enforcement** and **effectiveness and compliance**.

9.4.2.1 Ideal Approach Ideally, assessments of social resilience would include national-level and as well as local rules and other relevant institutional mechanisms that are meant to safeguard ocean health. The global focus has been on international treaties and indices, so your region should have more localized information. There would also be information as to their effectiveness and enforcement. of more. Information on social norms and community (and other local-scale) institutions (such as tenure or use rights) that influence resource use and management would be useful too.

9.4.2.2 Practical Considerations In practical terms, resilience is hard to define and finding data can be difficult. It is often difficult to find regulations and indicators that would directly ‘balance’ individual stressors, but it is worth the effort to explore what information is available in the local context and how it could be included as resilience measures. You may be able to construct your own set of indicators for resilience (particularly social resilience) using proxy data. Your team may have to get creative to develop appropriate assessment measures here. It can be metrically be defined as presence-versus-absence (value of zero or one), or on a scale (value between zero and one) if the measure is an assessment or score. For instance in the global study, resilience measures that were counted in the socio-economic resilience class of data came from the World Governance Indicators (<http://govindicators.org>). In a regional context, however, a more appropriate data layer might be a local governance index of some kind, preferably developed by a reputable organization using credible methods.

When available, National-level data are preferable to global-level data for your assessment. These include national laws on the environment, or protection of the marine environment or rivers that lead to coastal waters. National laws include things like the Clean Water Act (CWA) and the Endangered Species Act (ESA) in the U.S., or the national implementations of the E.U. Water Framework Directive. National actions can also be broadened beyond just legislation to include administrative procedures such as those involving permits, licenses, court cases, administrative action, and compliance mechanisms. [Cultural items at the national scale, such as holidays, are also applied at this scale]

State or province-level laws provide more regionally-specific information and thus work well for assessments. This would involve looking at the same types of laws and policies that exist on the national level, but specifically incorporating those that have been tailored to fit the needs of a particular sub-national area. This includes things such as California’s state-level California Environmental Quality Act (CEQA), or the California Ocean Protection Act (COPA), which have laws designed specifically to protect California’s environment. This would tell you more relevant information than using data from a national or international law. Local level regulations will usually provide you with the most accurate information for your assessment in order to tailor it best to the local context.

9.4.2.3 Scoring: Turning Qualitative into Quantitative There are several ways to turn the qualitative information of regulations and social actions into quantitative metrics for analysis. A robust way is to give credit for different aspects of the resilience measures. In addition to a score for having the law, policy, or action, in place, it is possible to gauge the effectiveness of that activity.

The simplest way is to give credit for having a resilience measure in place. This means assigning a binary score of zero or one for “presence” versus “absence” of the resilience measure. For international conventions, this can be done by assigning a value of 1 for having signed a convention. A more rigorous score can be given for countries that have further *ratified* a convention in addition to signing it; this is one way to further differentiate scores. This can be done by seeing if a country has signed and ratified CITES, for example.

For example, if you were trying to find out if there are regulations in place that guide fishing pressure, you could look see if regulations exist for trawl-fishing limitations, or see if there are regulations for fish size, length, or if there are any seasonal restrictions. Another option would be to see if formal stock assessments exist for commercially-fished species.

A further step is to assess how well those measures are being complied with. This will give you more robust way is to assess how well a resilience mechanism is working to maintain the integrity of the regulation and thereby the ecosystem.

For example, once you have found out whether regulations for fishing pressure exist, you would then try to find values for compliance with these regulations. These could be raw data or calculated statistics such as rate of compliance or proportion of compliance. It should answer the question, “Are there indicators of compliance with fishing pressure guidelines”?

A subsequent, and final, step to creating a robust resilience assessment is to determine whether there are enforcement mechanisms in place to deal with non-compliance of the regulations. This is because a regulation is only as good as its implementation, and having both enforcement and compliance actions in place would reinforce the regulation and make it more effective.

For example, in the case of fishing pressures, a further look into available data could lead you learn whether there are reported values of inspector visits and enforcement coverage of permitted facilities. Or you could look at reported numbers of enforcement actions in response to non-compliance. Further, you could also see if there are fines that have been paid or exist in association with non-compliance.

9.4.2.4 Data sources Environmental laws and policies offer tangible information on resilience. The most common type of environmental regulations come from administrative law, such as pollution regulation of various kinds. Land-use law is also important to the integration of social and ecosystem issues, so finding zoning laws relevant for coastal areas could be useful, and so could finding whether or not a region requires environmental impact statements before allowing construction for either coastal land or for marine planning. Other kinds of law some countries include court cases settling disputes or requiring reparation of pollution damages, for example.

Resilience also goes beyond just the law, however. Insurance policies present another option, for instance.. Coastal areas are increasingly requiring climate-related insurance in some countries, and so the existence of such markets in a vulnerable area would be an example of a climate change resilience measure. Social initiatives also present another way to tackle resilience. There might be a beach clean-up day, a percentage of the refuse material that is recycled by the population, or some other social factor that reduces trash inputs into the ocean. A local law banning plastic bags is another way that local jurisdictions control plastic trash.

9.4.3 Incorporating local resilience measures in your assessment

1. Begin by exploring how resilience could be measured in your study area. What laws and regulations are in place that could provide resilience to ocean health?
2. Are there locally-developed indices that capture social or ecological resilience? Is there information about how each region in your study area are implementing or enforcing the laws?

3. Evaluate the resilience measures are included in the global assessment. Are all of them relevant? Are there local data that can be substituted in the place of global data? Are there resilience measures that should be excluded entirely?
4. Assign the resilience measure to the appropriate goal. Since resilience measures are in response to pressures that have a weight rank of 2 or 3 effect on a certain goal, determining which goals ecological and social resilience measures effect follows the same pattern as the pressures matrix. For goal-specific resilience measures, assign the resilience measure to the appropriate goal.
5. Prepare each resilience data layer only after steps 1-6 are completed. In addition to the proper formatting for the Toolbox, resilience data must be rescaled (normalized) on a unitless scale from 0 - 1. You will have to determine how to rescale the data, whether it is based on the highest value in the data range or other methods.

9.4.3.1 Including resilience measures from global assessments Remember that local measures are far more appropriate than those included in global assessments, which likely do not reflect local management targets. However, international data were used in the global assessments that are available to you if you cannot find better local data:

Table of resilience layers and descriptions

| layer | name |
|-----------------------|---|
| alien_species | Alien species |
| cites | Resilience from commitment to CITES |
| fishng_v1 | CBD survey: coastal fishing v1 |
| fishng_v1_eez | CBD survey: ocean fishing v1 |
| fishng_v2_eez | CBD survey: ocean fishing v2 |
| fishng_v3 | CBD survey: coastal fishing v3 |
| fishng_v3_eez | CBD survey: ocean fishing v3 |
| habitat | CBD survey: habitat |
| habitat_combo | CBD survey: coastal habitat |
| habitat_combo_eez | CBD survey: ocean habitat |
| li_gci | GCI: competitiveness in achieving sustained economic prosperity |
| li_sector_evenness | Sector evenness as a measure of economic diversity |
| mariculture | CBD survey: mariculture |
| msi_gov | MSI sustainability and regulations |
| species_diversity | Ocean ecological integrity |
| species_diversity_3nm | Coastal ecological integrity |
| tourism | CBD survey: tourism |
| water | CBD survey: water |
| wgi_all | Strength of governance indicated with the WGI |

* CBD = Convention on Biological Diversity; GCI = Global Competitiveness Index; MSI = Mariculture Sustainability Index; WGI = World Governance Indicators.

10 Appendix 2: Instructions for preparing data files for the OHI Toolbox

This document explains what to provide the toolbox master for your OHI assessment, using China 2015 assessment as an example.

1. For each goal, please prepare a document describing the goal model, including the equation, an explanation of the variables, and the data layer name for each variable

2. For each data layer,
 - 2a. please provide separate *.csv* files for data layer
 - 2b. please provide separate description files for each data layer
3. For pressures and resilience:
 - 3a. please format pressure and resilience layers as described in 2) above
 - 3b. please update the pressures matrix rank weights for any added pressures
 - 3c. please add columns to the matrices for any pressures or resilience measures in China

10.1 1. Description document for each goal

Please prepare a document with the following information. This can be used in your China OHI Evaluation Manual to help demonstrate and justify your approaches. This document can be written in Chinese.

1. Goal model equation
2. Explanation of each variable
3. Data layer name for each variable
4. Written description of each goal model. This should include some or all of the following:
 - why the goal models were changed (or not changed)
 - why some data were included and other data weren't
 - any data limitations that also limited the model
 - what data would ideally be available and how that could improve the model
 - describe the 'ideal approach' for the goal, which may or may not be possible due to data availability, etc
 - list of pressures that negatively affect the status of the goal and the corresponding weight of the incidence of the pressure on the goal/subgoal
 - list of resilience measures that positively affect the status of the goal

10.2 2a. *.csv* files for each data layer

Each file containing data in the proper format for the OHI Toolbox is called a **data layer**.

10.2.1 Contents of data layers

Finalized data layers have at least two columns: the `rgn_id` column and a column with data, that is best identified by its units (km2 or often score). There also may be a `year` column or a `category` column (for natural product categories or habitat types). An example of a two-column file is '[al_access_gl2014.csv](#)', and an example of a three-column file is '[hab_extent_gl2014.csv](#)'.

An example of a data file is `ao_access_gl2014.csv`:

```
"rgn_id","value"
1,0.555771907253878
2,0.555771907253878
3,0.555771907253878
4,0.555771907253878
```

An example of a data file is `hab_extent_gl2014.csv`:

```
"rgn_id","habitat","km2"
1,"coral",261.102
2,"coral",261.102
3,"coral",261.102
4,"coral",261.102
```

The ‘region identifier’ (`rgn_id`) column has numbers identifying each province in China, numbered from north to south. You can explore the identifier numbers at [<https://ohi-science.org/chn/regions>]. Note: the OHI China team may modify the which number is associated with which province, however, it is important that there is consistency across all data layers because the OHI Toolbox calculates scores based on `rgn_id`. You can work with Ning to change the regions in `rgn_labels.csv`.

10.2.2 Naming data layers

Please name each data layer with the following format so it is easy to keep all data organized:

`prefix_layername_scenario_suffix.extension`

There cannot be any white spaces in any part of the filename: instead, use underscores ('').

The **prefix** will be the letters identifying each goal (two letters) or sub-goal (three letters):

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

The **layername** should be made of words or abbreviations to identify what the layer is.

The **scenario** should identify the assessment scenario: **chn2015**. This will help separate updated data layers from global data layers ('gl2014').

The **suffix** of the filename should identify who prepared the data so any questions can easily be sent to the correct person. For the suffix, please identify yourself with your initials.

The **extension** identifies the filetype and is separated by a period (.). You must save your files as comma separated values (`.csv`) since this is the format used by the OHI Toolbox. You can work in Excel and after data are formatted properly, it is possible to save as a `.csv` instead of a `.xlsx` file (select File > Save as > `.csv`).

Here is an example of a properly named data layer for the tourism and recreation goal, where the data are the percent of unemployment prepared by Julia Lowndes.

`tr_unemployment_chn2015_JL.csv`

10.3 2b. Description files to accompany each data layer

Please write a document to accompany each data file. Please save the document as a *.txt* file or a *.pdf* file, but not a Word document (*.docx*).

Each document should have the following information:

1. Filename of the data layer
2. Identify the target for this data layer: goal, sub-goal, pressures, or resilience
3. Short title of the data layer (for the ‘name’ column in `layers.csv`)
4. One-sentence explanation of the data layer (for the ‘description’ column in `layers.csv`)
5. Long description of the data layer that includes:
 - any data manipulation that has occurred, how to interpret it, and why (e.g. rescaling from 0-1)
 - any further data manipulation required, how to interpret it, and why
6. The data source:
 - the institution that collected the original data
 - the institution providing the data (if different from the original)
 - the year it was accessed
 - the URL or publication reference
7. The years of data available
8. The units of data
9. Any other observations or explanations about the data

10.4 3. Pressures and resilience

Remember that for any pressures and resilience measures included in the OHI matrices, there must be data available for each region in your study area.

10.4.1 Pressures

Several pressures layers are based on 1km² resolution spatial files that we can extract for regional studies. These layers are:

- ocean acidification (cc_acid)
- sea level rise (cc_slr)
- sea surface temperature (cc_sst)
- ultra violet radiation (cc_uv)
- nutrient pollution (po_nutrients)
- chemical pollution (po_chemicals)

10.4.2 Resilience

Resilience measures are often regulatory measures that indicate how provinces are performing compared with each other. If there are any environmental or social performance indicators for China, these could be used here.

11 Appendix 3: Instructions for preparing model descriptions

This document explains what information are needed from the goal keepers in order to calculate the goal scores and update the WebApp. An example is provided for each step of the instructions.

1. Goal model equation:

$$x_{CW} = \sqrt[4]{a * u * l * d}$$

2. Explanation of each variable:

The Status of this goal (XCW) is calculated as the geometric mean of four components where a = the number of people without access to sanitation (i.e., coastal population times % without access to improved sanitation) rescaled to the global maximum, u = 1 – (nutrient input), l = 1 – (chemical input), and d = 1 – (marine debris input). The intensity or input values for each sub-component come from the data sources listed in the separate data description files.

Xcw = clean waters score.

a = access to sanitation, by country. Access to improved sanitation as a proxy for pathogen pollution.

u = 1 – (nutrient input, by country). Ocean nutrient pollution.

l = 1 – (chemical input, by country). Ocean-based chemical pollution.

d = 1 – (marine debris input, by country). Trash pollution.

3. Data layer name for each variable:

- access to sanitation = ‘po_pathogens’
- nutrient input = ‘po_nutrients’
- chemical input = ‘po_chemicals’
- marine debris input = ‘po_trash’

4. Written description of goal model. This should include some or all of the following: (a) why the goal models were changed (or not changed); (b) why some data were included and other data weren’t; (c) any data limitations that also limited the model; (d) what data would ideally be available and how that could improve the model; (e) describe the ‘ideal approach’ for the goal, which may or may not be possible due to data availability

This goal measures the degree to which coastal waters are free of contaminants. This is the final score that is the geometric mean of the four clean waters variables of pathogens (a), nutrients (u), chemicals (l), and trash (d). This is the same model used in the Halpern et al. (2013) global assessment because it was the most robust based on the available data. Data are not disaggregated data at the sub-country and instead here represent the national value applied across provinces.

We used a geometric mean, as is commonly done for water quality indices, because a very bad score for any one sub-component would pollute the waters sufficiently to make people feel the waters were ‘too dirty’ to enjoy for recreational or aesthetic purposes.

Currently the model relies on a number of proxies, such as trends in access to improved sanitation as a proxy for pathogen pollution trend. Ideal data would consist of direct measurements of pathogen levels from monitoring stations at a sub-country level in a time-series. In the future as data become available from new monitoring efforts we will include them in this model.

5. List of pressures that negatively affect the status of the goal and the corresponding weight of the incidence of the pressure on the goal/subgoal:

| Pressure | Weight |
|------------------|--------|
| po_chemicals_3nm | 3 |
| po_pathogens | 3 |
| po_nutrients_3nm | 3 |
| po_trash | 3 |
| ss_wgi | 1 |
| ... | ... |

6. List of resilience measures that positively affect the status of the goal:

| Resilience | Weight |
|-----------------------------|--------|
| CBD Water | 1 |
| World Governance Indicators | 1 |
| ... | ... |

12 Appendix 4: Develop a record-keeping spreadsheet

If you are a toolbox master responsible for indexing all the data and modifying all the goal models, it is helpful to keep a master record-keeping spreadsheet from the start. This will allow to note who to contact for each goal and whether the key items for calculations are received, as well as to track your own progress. Goal keepers might be at different stages of data gathering and model development, and might send you pieces of information at different times. When you are updating multiple goal models, it is easy to lose track where you are in the progress for each goal.

If you are a goal keeper, having a record-keeping place to remind you where you are in the process and record all the important decisions may also be helpful. Here are some suggestions on what to and how to keep track of the progress.

Key items for each goal: * Goal keeper and their contact information * Data * Data layer description (see template) * Model description (see template) * Pressure * Resilience * Reference point

To track your own progress on coding, you can also include columns such as: * Coded? * Questions * Additional notes >It is important to record all the decision makings regarding data and models along the way. You can record those questions, the date when you contact the responsible perso(s), and the responses. If there is too much information, you can also keep this section on a separate document.

Here is an example of the spreadsheet:

13 Appendix 5: R Tutorials for OHI

Ocean Health Index R code uses several packages and best practices to facilitate understanding and collaboration. These approaches are presented here, along with examples using data included in global OHI assessments.

This document describes several packages that are used extensively in OHI assessments and introduces you to typical coding practices commonly seen in OHI scripts and functions.

Also see the accompanying R script to test examples using these packages.

| Goal | Contact | Data Layer | | Model | | Reference point | Pressure | Resilience | Questions | | Additional notes |
|---|---------|------------|-------------|-------------|--------|-----------------|----------|------------|-----------|-----------|------------------|
| | | Data | Description | Description | Coded? | | | | Questions | Response? | |
| <i>Food Provision, Fisheries (FIS)</i> | | | | | | | | | | | |
| <i>Food Provision, Mariculture (MAR)</i> | | | | | | | | | | | |
| <i>Artisanal Fishing Opportunities (AO)</i> | | | | | | | | | | | |
| <i>Natural Products (NP)</i> | | | | | | | | | | | |
| <i>Carbon Storage (CS)</i> | | | | | | | | | | | |
| <i>Coastal Protection (CP)</i> | | | | | | | | | | | |
| <i>Livelihood and Economies (LIV+ECO)</i> | | | | | | | | | | | |
| <i>Tourism and Recreation (TR)</i> | | | | | | | | | | | |
| <i>Sense of Place, Iconic Species (ICO)</i> | | | | | | | | | | | |
| <i>Sense of Place, Lasting Special Places (LSP)</i> | | | | | | | | | | | |
| <i>Clean Water (CW)</i> | | | | | | | | | | | |
| <i>Biodiversity, Habitats (HAB)</i> | | | | | | | | | | | |
| <i>Biodiversity, Species (SPP)</i> | | | | | | | | | | | |

Figure 64: Example spreadsheet

13.0.3 R Very Basics:

- Have you already downloaded and installed [R](#)?
- Have you already downloaded and installed [RStudio](#)?
- Have you walked through the excellent interactive tutorials from [swirl](#)?

13.0.4 tidy functions

‘Tidy’ up your messy data using [tidyverse](#) to make it easier to work with. The ‘tidy tools’ functions in the [dplyr](#) package work best with tidy data.

From Hadley Wickham’s [Tidy Data paper](#): >It is often said that 80% of data analysis is spent on the cleaning and preparing data. And it’s not just a first step, but it must be repeated many over the course of analysis as new problems come to light or new data is collected. To get a handle on the problem, this paper focuses on a small, but important, aspect of data cleaning that I call data tidying: structuring datasets to facilitate analysis.

From [RStudio’s introduction to tidyverse](#):

The two most important properties of tidy data are: 1. Each column is a variable. 2. Each row is an observation.

Arranging your data in this way makes it easier to work with because you have a consistent way of referring to variables (as column names) and observations (as row indices). When you use tidy data and tidy tools, you spend less time worrying about how to feed the output from one function into the input of another, and more time answering your questions about the data.

`gather()` is arguably the most useful function in [tidyverse](#), and is explained in more detail below. `spread()` and `separate()` are other useful functions in [tidyverse](#).

Other ‘tidy’ references: * [Hadley Wickham’s Tidy Data paper](#): Download the pre-print version for the whys and hows of tidy data. * [Cran tidy data vignette](#): An informal and code heavy version of Hadley’s full *Tidy Data* paper. * [RStudio Blogs: Introducing tidyverse](#): Basics and philosophy of [tidyverse](#) * [swirl tutorial package](#): A tutorial package built directly into R. Section 2: ‘Getting and Cleaning Data’ runs you through [dplyr](#) and [tidyverse](#) basics * [R data wrangling cheat sheet](#): A quick reference guide to [tidyverse](#) and [dplyr](#) functions

13.0.4.1 `tidy::gather()` Description

`gather()` takes data organized in rows and collapses them into a column format (a key column and a value column), duplicating all other columns as needed. Use `gather()` when your data is organized in “wide” format, in which some of your variables are in row form, rather than column form. Another `tidy` function, `spread()`, is more or less the reverse of `gather()`, to reformat long data into wide data. It is more difficult to work with wide data, but may be more convenient for examining data in a table format.

Note: `gather()` essentially replaces `melt()` in `plyr` package.

Example

The sample data set (see intro) contains harvest data of a number of marine commodities, separated by country, commodity, and year. In its original form, the harvest data (in tonnes) is spread across five different harvest years. * Counter to ‘tidy data’ principles, we have multiple columns (X2007:X2011) representing a single variable (year), and multiple observations of harvest tonnage in each row. * To transform this into ‘tidy data’ we will gather the five annual harvests into a single column called ‘tonnes’ and note the year of harvest in a new column called ‘year’.

The example in the figure below shows how the original wide data is transformed into long data using the command `gather`. Here are two ways of achieving this:

1. Here, information from columns X2007 through X2011 are gathered into a single column called `year`, and the information in each column are put into a new column called `tonnes`.

```
data_long <- data_wide %>% gather(year, tonnes, X2007:X2011)
```

2. Here, the `-` unselects the named columns, so they will not be gathered; all other columns are gathered into columns named `year` and `tonnes`. This approach will yield the same result.

```
data_long <- data_wide %>% gather(year, tonnes, -Country, -Commodity, -Trade)
```

| Wide data: 15 rows of 8 columns. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--|--|---|-------|--------|--|--|--|---------|-----------|-------|------|--------|-------------|---------------|--------|-------|-----|-------------|--------------------|--------|-------|-----|-------------|---------------------|--------|-------|-----|-------------|----------------------|--------|-------|-----|-------------|------------------------|--------|-------|-----|-------------|---------------------|--------|-------|---|-------------|---------------------------|--------|-------|-----|-------------|--|--------|-------|------|-------------|------------|--------|----|----|------------|---------------|--------|-----|---|------------|--------------------|--------|------|------|------------|---------------------|--------|------|-------|------------|----------------------|--------|---|---|------------|---------------------|--------|----|----|------------|--|--------|-----|-----|---------|-----------|-------|------|--------|-------------|---------------|--------|-------|-----|-------------|--------------------|--------|-------|-----|-------------|---------------------|--------|-------|-----|-------------|----------------------|--------|-------|-----|-------------|------------------------|--------|-------|-----|-------------|---------------------|--------|-------|---|-------------|---------------------------|--------|-------|-----|-------------|--|--------|-------|------|-------------|------------|--------|-------|----|------------|---------------|--------|-------|-----|------------|--------------------|--------|-------|------|------------|---------------------|--------|-------|------|------------|----------------------|--------|-------|---|------------|---------------------|--------|-------|----|------------|--|--------|-------|-----|--------------|---------------|--------|-------|-----|--------------|--------------------|--------|-------|-----|--------------|---------------------|--------|-------|-----|--------------|----------------------|--------|-------|-----|--------------|------------------------|--------|-------|-----|
| The data (tonnes of harvest) is spread out across multiple columns, separated by year. Each row represents multiple observations. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| data_wide <- data_long %>% spread(year, tonnes) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <table border="1"> <thead> <tr> <th>Country</th> <th>Commodity</th> <th>Trade</th> <th>year</th> <th>tonnes</th> </tr> </thead> <tbody> <tr><td>1 Australia</td><td>Agar agar nei</td><td>Export</td><td>X2007</td><td>208</td></tr> <tr><td>2 Australia</td><td>Coral and the like</td><td>Export</td><td>X2007</td><td>616</td></tr> <tr><td>3 Australia</td><td>Fish body oils, nei</td><td>Export</td><td>X2007</td><td>448</td></tr> <tr><td>4 Australia</td><td>Fish liver oils, nei</td><td>Export</td><td>X2007</td><td>...</td></tr> <tr><td>5 Australia</td><td>Mother of pearl shells</td><td>Export</td><td>X2007</td><td>189</td></tr> <tr><td>6 Australia</td><td>Ornamental fish nei</td><td>Export</td><td>X2007</td><td>5</td></tr> <tr><td>7 Australia</td><td>Ornamental saltwater fish</td><td>Export</td><td>X2007</td><td>...</td></tr> <tr><td>8 Australia</td><td>Other seaweeds and aquatic plants and products thereof</td><td>Export</td><td>X2007</td><td>3348</td></tr> <tr><td>9 Australia</td><td>Shells nei</td><td>Export</td><td>NA</td><td>NA</td></tr> <tr><td>10 Vietnam</td><td>Agar agar nei</td><td>Export</td><td>0 0</td><td>2</td></tr> <tr><td>11 Vietnam</td><td>Coral and the like</td><td>Export</td><td>2480</td><td>2600</td></tr> <tr><td>12 Vietnam</td><td>Fish body oils, nei</td><td>Export</td><td>8200</td><td>29000</td></tr> <tr><td>13 Vietnam</td><td>Fish liver oils, nei</td><td>Export</td><td>7</td><td>5</td></tr> <tr><td>14 Vietnam</td><td>Ornamental fish nei</td><td>Export</td><td>74</td><td>84</td></tr> <tr><td>15 Vietnam</td><td>Other seaweeds and aquatic plants and products thereof</td><td>Export</td><td>298</td><td>608</td></tr> </tbody> </table>
<table border="1"> <thead> <tr> <th>Country</th> <th>Commodity</th> <th>Trade</th> <th>year</th> <th>tonnes</th> </tr> </thead> <tbody> <tr><td>1 Australia</td><td>Agar agar nei</td><td>Export</td><td>X2007</td><td>208</td></tr> <tr><td>2 Australia</td><td>Coral and the like</td><td>Export</td><td>X2007</td><td>616</td></tr> <tr><td>3 Australia</td><td>Fish body oils, nei</td><td>Export</td><td>X2007</td><td>448</td></tr> <tr><td>4 Australia</td><td>Fish liver oils, nei</td><td>Export</td><td>X2007</td><td>...</td></tr> <tr><td>5 Australia</td><td>Mother of pearl shells</td><td>Export</td><td>X2007</td><td>189</td></tr> <tr><td>6 Australia</td><td>Ornamental fish nei</td><td>Export</td><td>X2007</td><td>5</td></tr> <tr><td>7 Australia</td><td>Ornamental saltwater fish</td><td>Export</td><td>X2007</td><td>...</td></tr> <tr><td>8 Australia</td><td>Other seaweeds and aquatic plants and products thereof</td><td>Export</td><td>X2007</td><td>3348</td></tr> <tr><td>9 Australia</td><td>Shells nei</td><td>Export</td><td>X2007</td><td>NA</td></tr> <tr><td>10 Vietnam</td><td>Agar agar nei</td><td>Export</td><td>X2007</td><td>0 0</td></tr> <tr><td>11 Vietnam</td><td>Coral and the like</td><td>Export</td><td>X2007</td><td>2480</td></tr> <tr><td>12 Vietnam</td><td>Fish body oils, nei</td><td>Export</td><td>X2007</td><td>8200</td></tr> <tr><td>13 Vietnam</td><td>Fish liver oils, nei</td><td>Export</td><td>X2007</td><td>7</td></tr> <tr><td>14 Vietnam</td><td>Ornamental fish nei</td><td>Export</td><td>X2007</td><td>74</td></tr> <tr><td>15 Vietnam</td><td>Other seaweeds and aquatic plants and products thereof</td><td>Export</td><td>X2007</td><td>298</td></tr> <tr><td>16 Australia</td><td>Agar agar nei</td><td>Export</td><td>X2008</td><td>122</td></tr> <tr><td>17 Australia</td><td>Coral and the like</td><td>Export</td><td>X2008</td><td>446</td></tr> <tr><td>18 Australia</td><td>Fish body oils, nei</td><td>Export</td><td>X2008</td><td>302</td></tr> <tr><td>19 Australia</td><td>Fish liver oils, nei</td><td>Export</td><td>X2008</td><td>...</td></tr> <tr><td>20 Australia</td><td>Mother of pearl shells</td><td>Export</td><td>X2008</td><td>215</td></tr> </tbody> </table> | | | | | | | | Country | Commodity | Trade | year | tonnes | 1 Australia | Agar agar nei | Export | X2007 | 208 | 2 Australia | Coral and the like | Export | X2007 | 616 | 3 Australia | Fish body oils, nei | Export | X2007 | 448 | 4 Australia | Fish liver oils, nei | Export | X2007 | ... | 5 Australia | Mother of pearl shells | Export | X2007 | 189 | 6 Australia | Ornamental fish nei | Export | X2007 | 5 | 7 Australia | Ornamental saltwater fish | Export | X2007 | ... | 8 Australia | Other seaweeds and aquatic plants and products thereof | Export | X2007 | 3348 | 9 Australia | Shells nei | Export | NA | NA | 10 Vietnam | Agar agar nei | Export | 0 0 | 2 | 11 Vietnam | Coral and the like | Export | 2480 | 2600 | 12 Vietnam | Fish body oils, nei | Export | 8200 | 29000 | 13 Vietnam | Fish liver oils, nei | Export | 7 | 5 | 14 Vietnam | Ornamental fish nei | Export | 74 | 84 | 15 Vietnam | Other seaweeds and aquatic plants and products thereof | Export | 298 | 608 | Country | Commodity | Trade | year | tonnes | 1 Australia | Agar agar nei | Export | X2007 | 208 | 2 Australia | Coral and the like | Export | X2007 | 616 | 3 Australia | Fish body oils, nei | Export | X2007 | 448 | 4 Australia | Fish liver oils, nei | Export | X2007 | ... | 5 Australia | Mother of pearl shells | Export | X2007 | 189 | 6 Australia | Ornamental fish nei | Export | X2007 | 5 | 7 Australia | Ornamental saltwater fish | Export | X2007 | ... | 8 Australia | Other seaweeds and aquatic plants and products thereof | Export | X2007 | 3348 | 9 Australia | Shells nei | Export | X2007 | NA | 10 Vietnam | Agar agar nei | Export | X2007 | 0 0 | 11 Vietnam | Coral and the like | Export | X2007 | 2480 | 12 Vietnam | Fish body oils, nei | Export | X2007 | 8200 | 13 Vietnam | Fish liver oils, nei | Export | X2007 | 7 | 14 Vietnam | Ornamental fish nei | Export | X2007 | 74 | 15 Vietnam | Other seaweeds and aquatic plants and products thereof | Export | X2007 | 298 | 16 Australia | Agar agar nei | Export | X2008 | 122 | 17 Australia | Coral and the like | Export | X2008 | 446 | 18 Australia | Fish body oils, nei | Export | X2008 | 302 | 19 Australia | Fish liver oils, nei | Export | X2008 | ... | 20 Australia | Mother of pearl shells | Export | X2008 | 215 |
| Country | Commodity | Trade | year | tonnes | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 Australia | Agar agar nei | Export | X2007 | 208 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 Australia | Coral and the like | Export | X2007 | 616 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3 Australia | Fish body oils, nei | Export | X2007 | 448 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4 Australia | Fish liver oils, nei | Export | X2007 | ... | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 5 Australia | Mother of pearl shells | Export | X2007 | 189 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 6 Australia | Ornamental fish nei | Export | X2007 | 5 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 7 Australia | Ornamental saltwater fish | Export | X2007 | ... | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 8 Australia | Other seaweeds and aquatic plants and products thereof | Export | X2007 | 3348 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 9 Australia | Shells nei | Export | NA | NA | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 10 Vietnam | Agar agar nei | Export | 0 0 | 2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 11 Vietnam | Coral and the like | Export | 2480 | 2600 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 12 Vietnam | Fish body oils, nei | Export | 8200 | 29000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 13 Vietnam | Fish liver oils, nei | Export | 7 | 5 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 14 Vietnam | Ornamental fish nei | Export | 74 | 84 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 15 Vietnam | Other seaweeds and aquatic plants and products thereof | Export | 298 | 608 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Country | Commodity | Trade | year | tonnes | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 Australia | Agar agar nei | Export | X2007 | 208 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 Australia | Coral and the like | Export | X2007 | 616 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3 Australia | Fish body oils, nei | Export | X2007 | 448 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4 Australia | Fish liver oils, nei | Export | X2007 | ... | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 5 Australia | Mother of pearl shells | Export | X2007 | 189 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 6 Australia | Ornamental fish nei | Export | X2007 | 5 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 7 Australia | Ornamental saltwater fish | Export | X2007 | ... | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 8 Australia | Other seaweeds and aquatic plants and products thereof | Export | X2007 | 3348 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 9 Australia | Shells nei | Export | X2007 | NA | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 10 Vietnam | Agar agar nei | Export | X2007 | 0 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 11 Vietnam | Coral and the like | Export | X2007 | 2480 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 12 Vietnam | Fish body oils, nei | Export | X2007 | 8200 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 13 Vietnam | Fish liver oils, nei | Export | X2007 | 7 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 14 Vietnam | Ornamental fish nei | Export | X2007 | 74 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 15 Vietnam | Other seaweeds and aquatic plants and products thereof | Export | X2007 | 298 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 16 Australia | Agar agar nei | Export | X2008 | 122 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 17 Australia | Coral and the like | Export | X2008 | 446 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 18 Australia | Fish body oils, nei | Export | X2008 | 302 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 19 Australia | Fish liver oils, nei | Export | X2008 | ... | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 20 Australia | Mother of pearl shells | Export | X2008 | 215 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | data_long <- data_wide %>% gather(year, tonnes, -Country, -Commodity, -Trade) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Long data: 75 rows of five columns. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Each column represents a single variable. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Each row represents a single observation. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Figure 65: wide data to long data using `gather()` and `spread()`

13.0.5 dplyr functions

The `dplyr` package includes a number of functions to easily, quickly, and intuitively wrangle your data. Here is a quick introduction with examples from data used in the Ocean Health Index.

From [RStudio's introduction to dplyr](#):

The bottleneck in most data analyses is the time it takes for you to figure out what to do with your data, and `dplyr` makes this easier by having individual functions that correspond to the most common operations...

Each function does one only thing, but does it well.

The most important `dplyr` functions to understand for data processing will be `group_by()`, `mutate()`, and `summarize()`. Also important, `dplyr` introduces the ability to perform subsequent functions in a logical and intuitive manner, using the `%>%` chain operator.

- `%>%` (chaining operator): allows sequential chaining of functions for cleaner, easier-to-read code
- `dplyr::select()`: selects variables to be retained or dropped from dataset
- `dplyr::filter()`: filters data set by specified criteria
- `dplyr::arrange()`: sorts dataset by specified variables
- `dplyr::mutate()`: adds variables or modifies existing variables
- `dplyr::summarize()`: uses analysis functions (sum, mean, etc) to summarize/aggregate specified variables
- `dplyr::group_by()`: groups data by specified variables, allowing for group-level data processing.

Other `dplyr` references:

- [RStudio blogs: Introducing dplyr](#): philosophy, examples, and basics of `dplyr`
- [Cran dplyr vignette](#): Walkthrough of `dplyr` with examples
- [dplyr and pipes: the basics](#): More examples of `dplyr` functions, and more depth on `%>%`
- [swirl tutorial package](#): A tutorial package built directly into R. Section 2: ‘Getting and Cleaning Data’ runs you through `dplyr` and `tidyverse` basics
- [R data wrangling cheat sheet](#): a quick reference guide to `tidyverse` and `dplyr` functions

13.0.5.1 %>% operator Description

The `%>%` operator allows you to ‘pipe’ or ‘chain’ a number of function calls, in which the output dataframe of one function is fed directly into the next function as the input dataframe. This lets you avoid creating temporary variables to store intermediate values, and lets you avoid nesting multiple functions. Using `%>%` makes your code more elegant, streamlined, and easy to read since you are able to write your code on multiple indented lines. From [dplyr and pipes: the basics](#):

OK, here’s where it gets cool. We can chain `dplyr` functions in succession. This lets us write data manipulation steps in the order we think of them and avoid creating temporary variables in the middle to capture the output. This works because the output from every `dplyr` function is a data frame and the first argument of every `dplyr` function is a data frame.

Usage

```

data_out <- f(data_in, args)
# standard function call

data_out <- data_in %>% f(args)
# function call using %>% operator. data_in is passed as first argument
# of function().

data_out <- data_in %>%
  f1(args1) %>%
  f2(args2) %>%
  f3(args3) %>% ...
# Output of function can be passed to another function immediately,
# without need for temporary storage. Indented format for legibility,
# see how pretty it looks?

```

Example

```

### Bad! Nested functions: read from inside out - hard to decipher
h_recent_totals1 <- arrange(mutate(filter(group_by(harvest, country, commodity),
  year >= 2009), harvest_tot = sum(tonnes, na.rm = TRUE)), country, commodity)

### Better: Line by line. Easier to read, but have to wait for the end to see
### what it does. Temp variables add more places for errors and bugs.
h_temp <- group_by(harvest, country, commodity)
h_temp <- filter(h_temp, year >= 2009)
h_temp <- mutate(h_temp, harvest_tot = sum(tonnes, na.rm = TRUE))
h_recent_totals2 <- arrange(h_temp, country, commodity)

### Best! Chained format intuitively links together the functions. Saves
### typing, fewer opportunities for errors, easier to debug. The %>% operator
### automatically indents each following line for easy reading.
h_recent_totals3 <- harvest %>%
  group_by(country, commodity) %>%
  filter(year >= 2009) %>%
  mutate(harvest_tot = sum(tonnes, na.rm = TRUE)) %>%
  arrange(country, commodity)

```

13.0.5.2 dplyr::select() Description

`select()` allows you to choose specific columns/variables from your dataset, and drop all others. Alternately, you can select specific variables to drop, leaving others in place. `rename()` is a relative of `select()` that allows you to rename variables, while leaving all variables in place.

Examples

The sample dataset includes the annual harvest, in tonnes, of a number of commodities exported by two countries. Type of trade provides no information (it is all Export), so that variable can be dropped. The names of all the variables should be converted to lower-case, to match the OHI style guide. See the figure below.

```

### Example 1:
harvest <- harvest %>%
  select(Country, Commodity, year, tonnes)
### Selects the named variables, and drops all others. Useful to choose a

```

```

### subset of key variables from a complicated data set.

### Example 2 (same result as example 1):
harvest <- harvest %>%
  select(-Trade)
### Using the '-' drops 'Trade' column and leaves other variables intact.
### Useful if you would like to clear out temporary variables.

harvest <- harvest %>%
  rename(country = Country, commodity = Commodity)
### Drops no variables. Syntax: rename(new_var_name = old_var_name) w/o quotes.

```

Using the chain operator, we can string these two functions into one smooth, easy-to-read flow:

```

harvest1 <- harvest %>%
  select(-Trade) %>%
  rename(country = Country, commodity = Commodity)

```

The `harvest` data is fed into `select()`, and the output is fed into `rename()`. The final output of this complete

harvest1 <- harvest %>% select(-Trade) %>% rename(country = Country)

Eliminate unneeded column

| Country | Commodity | Trade | year | tonnes |
|-------------|--|------------------|--------|--------|
| 1 Australia | Agar agar nei | Export | X2007 | 200 |
| 2 Australia | Coral and the like | Export | X2007 | 616 |
| 3 Australia | Fish body oils, nei | Export | X2007 | 448 |
| 4 Australia | Fish liver oils, nei | Export | X2007 | ... |
| 5 Australia | Mother of pearl shells | Export | X2007 | 189 |
| 6 Australia | Ornamental fish nei | Export | X2007 | 5 |
| 7 Australia | Ornamental saltwater fish | Export | X2007 | ... |
| 8 Australia | Other seaweeds and aquatic plants and products thereof | Export | X2007 | 3340 |
| 9 Australia | Shells nei | Export | X2007 | NA |
| 10 Vietnam | Agar agar nei | harvest (before) | Export | 0 0 |
| 11 Vietnam | Coral and the like | Export | X2007 | 2400 |
| 12 Vietnam | Fish body oils, nei | Export | X2007 | 8200 |

flow is assigned to the new variable `harvest1`.

13.0.5.3 dplyr::filter() Description

`filter()` allows you to select observations (rows) that match search criteria, using values in specified variables (columns). Drops all observations that do not match the criteria. * Use logical operators & and | to filter on multiple criteria simultaneously

Example

```

harvest_vnm <- harvest %>%
  filter(country == 'Vietnam')
### Single criterion filter: keeps only data with country matching 'Vietnam'.

```

```

h_vnm_recent <- harvest %>%
  filter(country == 'Vietnam' & year >= 2009)
### filter with multiple criteria: selects 'Vietnam' data from 2009 or later.

```

13.0.5.4 dplyr::arrange() Description

`arrange()` sorts observations (rows) based upon a specified variable or list of variables. Does not actually change the data in any way, only the appearance. Useful for inspecting your data after each processing step.

Example

```

harvest_sorted <- harvest %>%
  arrange(country, commodity, year)
  ### Sorts commodity harvest values for each country, chronologically

harvest_sorted <- harvest %>%
  arrange(country, commodity, desc(year))
  ### Sorts harvest values by most recent year (descending order)

```

13.0.5.5 dplyr::mutate() Description

`mutate()` is a powerful and useful tool for processing data. You can add new variables or modify existing variables, using all variety of functions to perform operations on the dataset. `mutate()` works well with `group_by()` to perform calculations and analysis at a group level rather than dataset level.

Example

From the sample data set (see figure below), we would like to:

- Remove the ‘X’ from the ‘year’ values.
- Translate the text codes in ‘tonnes’ into numbers and NAs. These codes are specific to FAO’s data reporting format: ... is the same as NA, and 0 0 means greater than zero, but less than half a tonne.
- Convert these text fields into numeric fields so they can be analyzed properly.

| | country | commodity | year | tonnes |
|----|-----------|--|-------|--------|
| 1 | Australia | Agar agar nei | X2007 | 280 |
| 2 | Australia | Coral and the like | X2007 | 616 |
| 3 | Australia | Fish body oils, nei | X2007 | 448 |
| 4 | Australia | Fish liver oils, nei | X2007 | ... |
| 5 | Australia | Mother of pearl shells | X2007 | 189 |
| 6 | Australia | Ornamental fish nei | X2007 | 5 |
| 7 | Australia | Ornamental saltwater fish | X2007 | ... |
| 8 | Australia | Other seaweeds and aquatic plants and products thereof | X2007 | 3348 |
| 9 | Australia | Shells nei | X2007 | NA |
| 10 | Vietnam | Agar agar nei | X2007 | 0 0 |
| 11 | Vietnam | Coral and the like | X2007 | 2400 |
| 12 | Vietnam | Fish body oils, nei | X2007 | 8286 |
| 13 | Vietnam | Fish liver oils, nei | X2007 | 7 |

harvest (before)

| | country | commodity | year | tonnes |
|----|-----------|--|------|--------|
| 1 | Australia | Agar agar nei | 2007 | 280.0 |
| 2 | Australia | Coral and the like | 2007 | 616.0 |
| 3 | Australia | Fish body oils, nei | 2007 | 448.0 |
| 4 | Australia | Fish liver oils, nei | 2007 | NA |
| 5 | Australia | Mother of pearl shells | 2007 | 189.0 |
| 6 | Australia | Ornamental fish nei | 2007 | 5.0 |
| 7 | Australia | Ornamental saltwater fish | 2007 | NA |
| 8 | Australia | Other seaweeds and aquatic plants and products thereof | 2007 | 3348.0 |
| 9 | Australia | Shells nei | 2007 | NA |
| 10 | Vietnam | Agar agar nei | 2007 | 0.1 |
| 11 | Vietnam | Coral and the like | 2007 | 2400.0 |
| 12 | Vietnam | Fish body oils, nei | 2007 | 8286.0 |
| 13 | Vietnam | Fish liver oils, nei | 2007 | 7.0 |

harvest1 (after)

Figure 66: Using mutate to alter data in a dataframe

```
library(stringr)  ### to access 'str_replace()' string functions
```

```

harvest1 <- harvest %>%
  mutate(
    year = str_replace(year, fixed('X'), ''), # remove the 'X'
    tonnes = str_replace(tonnes, fixed('...'), NA), # replace '...' with 'NA'
    tonnes = str_replace(tonnes, fixed('0 0'), 0.1), # replace '0 0' with '0.1'
    tonnes = ifelse(tonnes == '', NA, tonnes)) %>%
  mutate(
    tonnes = as.numeric(as.character(tonnes)),
    year = as.integer(as.character(year)))

```

Notes: * In this example, no new variables were added. Multiple variables can be changed with one call to `mutate()`. Multiple modifications to ‘tonnes’ happen sequentially, so order is important. * The `as.numeric(as.character(...))` gets around the fact that these text variables are stored as ‘factor’ class, rather than ‘character’ class. `as.character()` forces them into character class, and then the `as.numeric()` can convert the character strings to numeric where applicable. Similar for `as.integer(...)`

13.0.5.6 dplyr::summarize() (or summarise()) Description

`summarize()` combines multiple values of a variable into a single summary value. `summarize()` works well with `group_by()` - for grouped data, each group will be summarized and reported separately. For ungrouped data, the summary covers the entire dataset.

- `summarize()` compresses the dataset and drops individual observations. To maintain individual observations, consider creating a summary variable using `mutate()` instead.
- NA values can be problematic - use `na.rm=TRUE` or similar methods.

Example

To determine the total harvest of each country, for each commodity:

```
h_summary <- harvest %>%
  group_by(country, commodity) %>%
  summarize(harvest_tot = sum(tonnes, na.rm = TRUE)) %>%
  ungroup()
```

13.0.5.7 dplyr::group_by() Description

`group_by()` allows you to easily group a dataset by one or more variables/columns.

By itself, it does nothing to change your data. But once your dataset has been sorted into useful groups, other `dplyr` functions will operate on each group separately, rather than operating on the entire dataset. *

The function `groups(data)` reports back the current grouping status of dataframe `data`.

- `group_by()` alters the grouping, but does not alter the sort order.
- `arrange()` does not alter the current grouping - it will sort by groups first, then sorts within each group.
- Multiple calls to `group_by()` will reset the groupings each time (by default), rather than adding additional layers of groups.
- Once you have finished with your operation at the group level, it is a good practice to use the `ungroup()` function to remove the groupings, to avoid unintended consequences due to forgotten `group_by()` calls.

Example

If you want to find the total tonnage harvested for each commodity for each country, you would want to group by country and by commodity, and then perform a `sum()` function on the grouped data. Two options presented here: `summarize()` to collapse data to just the summary, and `mutate()` to add a new column that includes the summary values.

```
h_tot_sum <- harvest %>%
  group_by(country, commodity) %>%
  summarize(harvest_tot = sum(tonnes, na.rm = TRUE))
  ### Summarize information by collapsing each group to a single summary value
  ### (total tonnage by commodity by country). Note ungroup() at end.
```

```
h_tot_mut <- harvest %>%
  group_by(country, commodity) %>%
  mutate(harvest_tot = sum(tonnes, na.rm = TRUE)) %>%
  arrange(country, commodity) %>%
  ungroup()
  ### Summarize information by creating a new variable to contain summary
  ### value; report value for every observation. Note ungroup() at end.
```

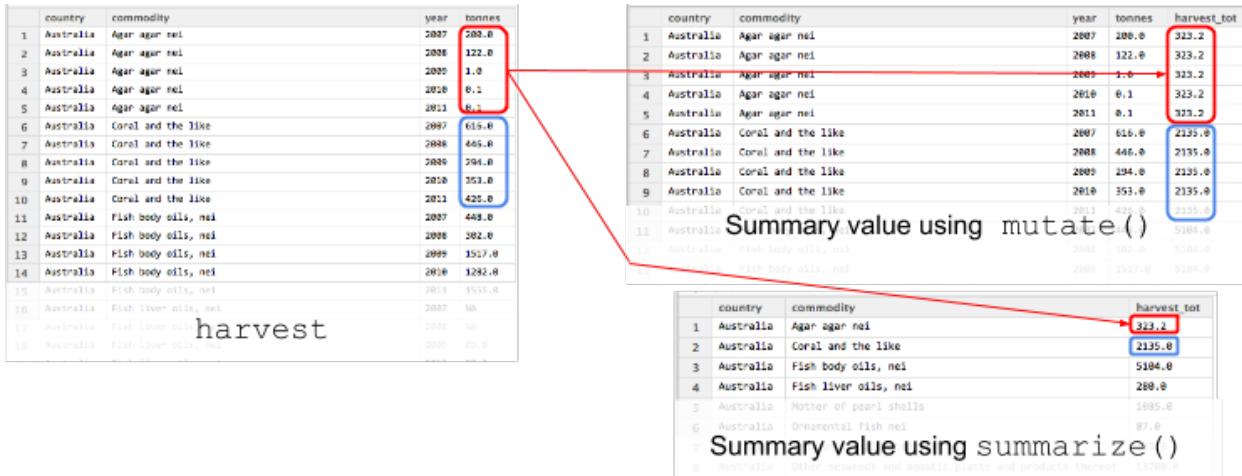


Figure 67: group_by to find group-level information

13.0.6 Coding style

Code unto others as you would have them code unto you.

Why style? ask Hadley Wickham, developer of many wonderful R packages:

Good style is important because while your code only has one author, it'll usually have multiple readers. This is especially true when you're writing code with others. In that case, it's a good idea to agree on a common style up-front. Since no style is strictly better than another, working with others may mean that you'll need to sacrifice some preferred aspects of your style.

The Ocean Health Index is founded upon principles of open-source science, so our code should be not just available, but legible to others. For OHI+, we expect people to modify code to implement new goal models, and we may need to provide support in developing and debugging their code.

Certain coding techniques are more efficient than others (e.g. in R, looping across elements in a vector is much slower than operating on the entire vector at once), but rarely does OHI code push any performance envelopes. Much more of our time is spent writing code, translating old code into new models, and debugging. Transparent, readable code will save more time in the future than a perfectly-optimized but opaque algorithm.

Readable code is:

- collaborative
- easier for others to understand and debug
- easier for others to update and modify
- easier for ‘future you’ to interpret what ‘past you’ meant when you wrote that chunk of code.

Check out Hadley Wickham’s [style guide](#):

- How many of these suggestions are second-nature to you? how many are you guilty of breaking?
- Note that these are guidelines, not rules; non-stylish code can still work.

13.0.6.1 Best practices for coding in OHI assessments:

- use a consistent format for variable names, filenames, function names, etc.
 - `lower_case_with_underscores` (preferred) or `camelCase` (ok I suppose)
 - * not `periods.in.between`
 - use names that are brief but intuitive
- Comment clearly for your own purposes, and for others.
 - Comment on the purpose of each important block of code.
 - Comment on the reasoning behind any unusual lines of code, for example an odd function call that gets around a problem.
- Take advantage of R Studio section labels functionality:
 - If a comment line ends with four or more -, =, or # signs, R Studio recognizes it as a new section.
 - Text within the comment becomes the section name, accessible in the drop-down menu in the bottom left of the RStudio script window.
- use <- to assign values to variables (not necessary, but preferred)
- use %>% to create intuitive chains of related functions
 - one function per line
 - break long function calls into separate lines (e.g. multiple mutated variables)
- use proper spacing and formatting for legibility
 - don't crowd the code - use spaces between math operators and after commas
 - use indents to indicate nested or sequential/chained code
 - break sequences or long function calls into separate lines logically - e.g. one function call per line
- use functions to add intuitive names to chunks of code
- Use ‘tidy data’ practices - take advantage of `tidyverse`
 - clean up unused columns using `select(-colname)`
- if you are working on an older script, spend a few extra minutes to update it according to these best practices
 - technical debt - you can do it quickly or you can do it right. Time saved now may cost you or someone else more time later.

13.0.6.2 Writing functions <http://nicercode.github.io/guides/functions/> Why write functions? * name a chunk of code for easier reading * easily reuse a chunk of code

What makes a good function: * It's short * Performs a single operation * Uses intuitive names

13.0.6.3 Directories and files

* Store files in a folder called 'github' in your home directory; access it with `~/github` so that user

14 Appendix 6: 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).

14.1 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: How does the Index account for ecosystem benefits?

A: The OHI is not an index of ecosystem services. The Index prefers to describe benefits from a healthy ocean and emphasize their relevance, but the ideas are closely related. The ten goals roughly fall into areas of ecosystem services such as food provisioning (**Food Provision**), regulatory services (**Carbon Storage**), cultural services (**Tourism and Recreation, Special Places**), supporting services (**Clean Waters, Biodiversity**), and other values (**Livelihoods and Economies**).

(Source: *OHI Baltic workshop*)

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.

Q: Can 10 goals really capture all the ocean's values? Shouldn't we talk about ocean's values in more detailed terms?

A: This question speaks of the interdisciplinary nature of OHI. It is a part of the design of the model. It takes experts from all fields many years to develop the 10 goals, and each goal includes different aspects. Ten goals can easily become 100. For example, the economy goal can be divided into individual goals of tourism, transportation, production, value added (eg. fish to burger) etc. However, getting solid scores for 100 goals could take more than a year or two, and trying to communicate all the information to stakeholders and managers will be cumbersome. But we can communicate with people across fields about 10 goals (and their pressures and resilience) without being overwhelming.

Ten goals may seem simplified. But they are useful for talking to people across fields without compromising scientific validity.

Q: How are the goals weighted to calculate the final score?

A: By default, we weighed them equally because we don't have solid, better information to do otherwise, even though we understand that each goal may/should have different weights depending on regional conditions

and needs. We didn't put emphasis on goal weighing because ten individual goal scores are more important than one total score, and can tell more about the status of ocean health. If you are interested in seeing how weighing affects the total index score, you can try the following:

- Scenario studies. The OHI framework is flexible so that you can change the weights and see how scores change
- Studies on people's preferences or management needs to weigh goals differently

Q: How should we compare regional scores, spatially or temporally?

A: Temporal comparison is good for repeated measurements in one location, so that managers can see how goal conditions change and whether their policies take effect. Spatial comparison is used to compare regional differences in your country.

For OHI+, the most helpful comparison is temporal comparison. Management goals are different spatially, and data are different. We can't compare spatially unless they use the same model and data. The goal of OHI is to help local agencies adapt and improve ocean resources management. Therefore it may be an unnecessary burden to compare spatially. Without worrying about comparisons, different regions can still learn from one another.

Q: We calculate trends based on 5 years of data. Wouldn't it be more scientifically valid to have a 20 or 30 year trend?

A: Yes you can use more years of data if available. But 5 years is a good compromise between scientific validity and meaningfulness for local stakeholders. On one hand, we have to do it scientifically vigorous with a solid reference. But on the other hand, a 5-yr trend is more helpful and actionable to stakeholders than a 30-yr trend.

Q: Is OHI an internationally recognized monitoring system? What other partnership does OHI have?

A: OHI is gaining recognition from world governance organizations, NGOs, and local governments. There is only limited amount of funds for conservation. If we work together, we can reach maximum return on investment. We partner with UCSB, who is on the scientific frontier. They innovate science and technology, develop toolbox, fill data gap, and provide technical support for local OHI assessments. Conservation International focuses on external partnerships. CI is always looking at multilateral agreements. Some of our active partnerships include:

- The United Nation. Many countries follow UN leads.
- World Bank and The Nature Conservancy, who has good research on economics, and is interested in incorporating OHI in their monitoring systems that are already in place.
- World Wildlife Fund, who is interested in using OHI in all of their 14 global sites.
- Clinton Climate Initiative.

How to draw spatial boundary: ecosystem or jurisdictional boundary?

A. We recommend jurisdictional boundary. It comes down to who will care, and who will take responsibility for the results and act on the results.

There seems to be overlapping between goals, for example, habitats data re used for CS, CP, and HAB. Is that double counting?

A. No. There is no complete overlap of data. Different habitat data are used for different goals. For example, habitats good for CS aren't necessarily appropriate for CP.

OHI is also human benefits-driven, not ecologically-driven. It happens that CS, CP, and HAB are affected by similar factors, but they provide non-overlapping benefits. It is the design of OHI to show the connectivity among goals. If certain habitats are good for multiple goals, there is strong incentive to focus energy and resources to protect those habitats.

14.2 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 skill sets. 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: The duration of an OHI assessment depends on a number of factors, such as the 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. For independent assessments (OHI+), we have found that the average time has ranged from 1.5 to 3 years (See **Task Timeline** in the **Conceptual Guide**).

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

14.3 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.

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. Sector evenness was used as a resilience measure for Livelihood & Economies goal.

14.4 Appropriate data layers

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: 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.

Q: Where do energy activities fit in to OHI?

It depends. Energy could be part of a **Natural Products** goal, for instance, such as wave energy – but then the question is, what is the reference point? It is partially accounted for in **Livelihoods & Economies** through sectoral jobs data. The infrastructure is also something to consider. It could also be a pressure or resilience factor if there is a measurable footprint of the activity. You may want to consider for resilience, do you have governance measures that promote more sustainable practices in the energy industry?