

The Ocean Health Index Assessment Manual

Contents

1	What is OHI+?	6
1.0.1	The Process of OHI+	6
1.0.1.1	What to expect when conducting an assessment	6
1.0.1.2	Where to start when conducting an assessment	7
1.0.2	Expected outcomes	7
2	Introduction to OHI Assessments	7
3	Overview of the OHI WebApp	8
3.1	Background	9
3.2	Using the WebApp	9
3.3	The App Page	11
3.4	The App's Data tab	11
3.4.1	Overview of display options	11
3.4.2	Overview of variable options	11
3.5	The App's Compare tab	14
3.6	Discovering and Gathering Appropriate Data and Indicators	14
3.6.1	Data sources	15
3.6.2	Gathering responsibilities	15
3.6.3	The process of discovery	15
3.6.4	Requirements for data and indicators	16
3.6.4.1	Relevance to ocean health	16
3.6.4.2	Accessibility	17
3.6.4.3	Quality	17
3.6.4.4	Reference point	17
3.6.4.5	Appropriate spatial scale	17
3.6.4.6	Appropriate temporal scale	17
3.6.5	Example: US West Coast data discovery	17
3.6.5.1	Reasons data were excluded	18
3.7	Best approaches for Starting an Assessment	18
3.7.0.2	Strategies for Assessment	19
3.7.1	Using this Guide	19
3.7.2	Where to start?	19

3.7.2.1	To consider as you get started	19
3.7.2.2	Which data are available?	19
3.7.2.3	What scale is appropriate?	20
3.7.3	What's important spatially?	20
3.7.3.1	Reference Points	20
3.7.4	Practical Guidance	20
3.7.4.1	Synergies and Goals to approach together	20
3.7.4.2	Do Habitat-Based Goals Together	21
3.7.4.3	Do Fisheries goals and pressures together	21
3.7.4.4	Linkages: Note the connections between pressures and goals	21
3.7.5	Habitat-based Goals	22
3.7.5.1	Practical Considerations	22
3.7.5.2	<i>Ideal Approach</i>	22
3.7.5.3	Ready to start on Habitat-based Goals?	22
3.7.6	Ready to start the Carbon Storage goal?	23
3.7.6.1	<i>Ideal Approach</i>	23
3.7.6.2	Practical Approach	23
3.7.6.3	Ready to start on Coastal Protection?	24
3.7.6.4	<i>Ideal Approach</i>	25
3.7.6.5	Practical Approach	25
3.7.6.6	Linking Habitat and Biodiversity	25
3.7.6.7	Practical Considerations	26
3.7.6.8	<i>Ideal Approach</i>	26
3.7.6.9	Ready to start on Natural Products?	26
3.8	Pressures and resilience	28
3.8.1	<i>Ideal Approach:</i>	28
3.8.1.1	Pressure and resilience categories	28
3.8.2	Pressures and resilience matrices	28
3.8.3	Incorporating local pressures in your assessment	31
3.8.3.1	Including pressures from global assessments	31
3.9	Guide to searching for resilience metrics	32
3.9.1	<i>Ideal Approach</i>	32
3.9.1.1	Practical Considerations	32
3.9.1.2	Scoring: Turning Qualitative into Quantitative	33
3.9.1.3	Data sources	34
3.9.2	Incorporating local resilience measures in your assessment	34
3.9.2.1	Including resilience measures from global assessments	34

4 The Ocean Health Index Toolbox	35
4.1 File System for Assessment Repositories	36
4.1.1 Assessments and scenarios	36
4.1.2 <i>layers.csv</i>	36
4.1.3 <i>layers</i> folder	39
4.1.4 <i>conf</i> folder	40
4.1.4.1 <i>config.R</i>	40
4.1.4.2 <i>functions.R</i>	40
4.1.4.3 <i>goals.csv</i>	40
4.1.4.4 <i>pressures_matrix.csv</i>	41
4.1.4.5 <i>resilience_matrix.csv</i>	41
4.1.4.6 <i>resilience_weights.csv</i>	41
4.1.5 <i>install_ohicore.R</i>	41
4.1.6 <i>launch_app_code.R</i>	41
4.1.7 <i>calculate_scores.R</i>	41
4.1.8 <i>scores.csv</i>	41
4.1.9 <i>spatial</i> folder	43
4.1.10 <i>layers-empty_swapping-global-mean.csv</i>	43
4.1.11 Relaunching the Toolbox	43
4.2 Formatting Data for the Toolbox	43
4.2.1 Introduction	43
4.2.2 Gapfilling	44
4.2.2.1 Temporal gapfilling	44
4.2.2.2 Spatial gapfilling	46
4.2.3 Long formatting	47
4.3 Rescaling your data	48
4.3.0.1 Example: Global Data Approach	49
5 Installing the Toolbox	49
5.1 Overview	49
5.2 GitHub	50
5.2.1 Learning GitHub	50
5.3 Accessing GitHub Repositories	50
5.3.1 Create a GitHub account	50
5.3.2 Install <i>git</i> software	50
5.3.3 Set up your Git Identity	53
5.3.4 Install the GitHub application	53

5.3.5	Create a folder called <i>github</i> on your computer	53
5.3.6	Clone your repository to your computer	54
5.3.7	Update permissions	55
5.3.8	Work locally	55
5.3.9	Syncing	56
5.3.10	Using the GitHub App to synchronize your repository	56
5.3.11	Working with R and RStudio	56
5.3.12	Using RStudio to synchronize your repository	58
5.3.13	Install the latest version of R and RStudio	59
5.4	GitHub repository architecture	60
6	Using the Toolbox	60
6.0.1	File Preparation Workflow	62
6.0.1.1	Overview of the Process	62
6.1	Modifying and creating data layers	63
6.1.1	Create data layers with proper formatting	63
6.1.2	Save data layers in the <i>layers</i> folder	63
6.1.3	Register data layers in <i>layers.csv</i>	65
6.1.4	Check pressures and resilience matrices	66
6.2	Modifying pressures matrices	66
6.2.1	Create the new pressure layers and save in the <i>layers</i> folder	66
6.2.2	Register the new pressure layers in <i>layers.csv</i>	67
6.2.3	Register the new layers in <i>pressure_matrix.csv</i>	67
6.2.3.1	Set the pressure category	67
6.2.3.2	Identify the goals affected and set the weighting	67
6.2.4	Modify the resilience matrix (if necessary)	67
6.3	Modifying resilience matrices	68
6.3.1	Updating resilience matrix with local habitat information	68
6.3.1.1	Global resilience layers	68
6.3.1.2	Determining how to modify these resilience layers	69
6.4	Modifying goal models	71
6.4.1	Update <i>functions.R</i>	71
6.4.2	Check and possibly update <i>goals.csv</i>	71
6.4.3	Example modification:	74
6.4.4	What's the code trying to do?	74
6.4.4.1	Natural Products	74
6.4.4.2	Data Sources	75

6.4.4.3	Gap-filling	75
6.4.5	Appendix - source materials	76
6.4.5.1	Global Data Approach (Technical Notes)	76
6.4.5.2	Notes: Preparing the Data	76
6.5	76
6.5.0.3	Notes: Tech Specs	76
6.6	Removing goals	76
6.7	Modifying goals with categories	78
6.7.1	Example 1: Pressures	78
6.7.2	Example 2: Pressures	79
6.7.3	Example 3: Resilience	79
6.7.3.1	Resources	82
6.8	Notes about R	82
7	Frequently Asked Questions (FAQs)	82
7.1	Overall	83
7.2	Conceptual	83
7.3	Timing and Resources	83
7.4	Structure	84
7.5	Reference points	84
7.6	Appropriate data layers	85
7.7	Food Provision	86
7.8	Livelihoods & Economies	87
7.9	Tourism & Recreation	87
7.10	Natural Products	87
7.11	Species	87
7.12	Sense of Place	88
7.13	Pressures	88
8	Toolbox Troubleshooting	88
8.1	Error: RStudio won't push to GitHub	88
8.2	Loading RWorkspace on Restart	91
8.3	Calculating Pressures.	92
8.3.1	'The following components for [goal] are not in the aggregation layer [layer]...'	92
8.3.2	'Error in matrix...'	94
8.4	Calculating Resilience	94
8.4.1	'Error in match(x, table, nomatch = OL) : object id_num not found'	94

1 What is OHI+?

Your assessment will be an example of independent Ocean Health Index+ assessment (OHI+). At this stage, you should have **Learned** and **Planned** for your assessment, and are now ready to **Conduct** it.

The Ocean Health Index framework presents a highly tailorable approach ready to meet the needs and priorities of those who wish to use the ocean sustainably. Here you will learn how to conduct an Ocean Health Index+ (OHI+) assessment. The OHI+ approach adds relevant inputs and information to the overarching OHI framework and consists of data and indicators that are specific to the local context. Because the methods of the framework are repeatable, transparent, quantitative, and goal-driven, repeated assessments are valuable because they can be used to track and monitor the progress of ocean health through time. Even if yours is a first effort, your OHI+ is valuable because it establishes a monitoring baseline and calls attention to the state of the data quality and availability in your area. Such a flexible approach allows OHI+ assessments to adapt to local conditions while still being useful for management.

Remember that when conducting an assessment, the process is just as valuable as the results.

The Index is an assessment tool that provides a common platform for scientifically combining and comparing key elements from all dimensions of the oceans health — biological, physical, economic, and social – to measure how sustainably people are using the ocean. By allowing for a comprehensive and integrated view of marine systems as a way to understand the trade-offs and synergies among these goals, the Index represents a significant advance over conventional single-sector approaches to assessing sustainability as communicated through other kinds of index efforts. Because of this, preparing the conducting phase will require careful thought, consideration, and documentation, throughout this process.

1.0.1 The Process of OHI+

The OHI+ process consists of four distinct phases. In the first phase, you **learned** about the OHI and understood the philosophy behind it and the motivation for conducting your study. In the second phase, you actively **planned** to conduct an OHI+ Assessment. Now, in the third phase, you will **conduct** the assessment by engaging with the science of finding the data, preparing the goal models, and taking the necessary steps to learn the software and produce the results. In the final phase, you will communicate the results of your OHI+ Assessment in order to inform stakeholders who will use its findings.

- **Phase 1: Learn** about the OHI
- **Phase 2: Plan** an OHI+ Assessment
- **Phase 3: Conduct** the OH+ Assessment
- **Phase 4: Communicate** and Inform

1.0.1.1 What to expect when conducting an assessment There are key considerations and processes that will be a part of every assessment, however, the process for conducting each assessment will be unique depending on the local context. For example: what data and indicators are available will determine how goal models can be tailored to the region, and what skillsets and resources are involved will affect the time it takes to complete the assessment.

In OHI+, it is important for your team doing the study to make the decisions and modifications that are most important for the region, and then defend it clearly what was done and why. At the same, your team should be as creative and insightful as you can be.

1.0.1.2 Where to start when conducting an assessment You should start by understanding the structure of the global assessment and the data involved will help you think about what should be done differently in your local context.

The best way to do this is to begin with the WebApps. As described in the section, “**Overview of the OHI WebApp**,” most coastal countries have a WebApp that was created to facilitate planning and communication during your assessment. The WebApp presents data, goal models and calculated scores for each region (global administrative area identified by <http://gadm.org>) visually through maps, histograms, and tables. All data presented were extracted from the global analysis, and scores were calculated using global goal models. For a finer-scale assessment of ocean health in your region, these data files provided will need to be updated with available data and indicators for each region in your assessment. However, if better data are not available, you can use the data provided. Then, to dive deeper into data layers and goal models, you can explore your assessment’s GitHub repository, which stores all the information presented through the WebApp.

You should also be familiar with the approaches taken by other assessments adapted from the global context, including Brazil, the US West Coast, and Fiji. You can find these studies at <http://ohi-science.org>.

While our team of scientists and managers is prepared to provide guidance for assessments, you should follow the steps in this training program to complete your assessment as autonomously as possible.

1.0.2 Expected outcomes

The process of conducting an OHI assessment is as valuable as the final results. This is because while conducting an OHI assessment you will identify gaps in knowledge and data, produce decision-relevant information, and create an ocean alliance that combines knowledge and cultural values across disciplines. Conducting an OHI assessment requires engagement from as many different groups as possible, including research institutions, government agencies, policy groups, non-governmental organizations, and the civil and private sectors.

Finding the best data and indicators available is crucial for obtaining meaningful findings that can help inform decision-making. Assessments can incorporate higher-resolution data and indicators, local priorities and preferences, and develop tailored goal models and reference points, which produce scores that better reflect local realities. If a goal is not relevant in the local context, it can be excluded entirely. When you change goal models, though, it is important to capture the process in order to justify decisions that will inform the results. Similarly, pressures and resilience measures can be refined using local data and indicators. Index scores are only as good as the data on which they are based.

2 Introduction to OHI Assessments

Section Summary:

This section provides an overview of conducting an OHI assessment. In each assessment, goals are scored from 0 to 100 and then aggregated to produce an overall Index score. An assessment requires data, indicators, and priorities that are appropriate for each context.

TIP: The process of conducting an OHI assessment is as valuable as the results for policy and management decisions.

Assessments using the OHI framework can be most relevant to policy and management decisions when they are conducted at smaller spatial scales. These spatial scales can include political subdivisions like countries, states, provinces, and eco-regions such as bays. Using ten criteria (called goals), the Index scores on a scale of 0 to 100 how well coastal regions optimize their potential ocean benefits and services in a sustainable way relative to self-established reference points (targets). Relevant goal models are developed based on the best

available information, which can be either raw data or composite data, or previously developed indicators. **The process of conducting an OHI assessment is as valuable as the final results.** This is because while conducting an OHI assessment you will identify gaps in knowledge and data, produce decision-relevant information, and create an ocean alliance that combines knowledge and cultural values across disciplines. Conducting an OHI assessment requires engagement from as many different groups as possible, including research institutions, government agencies, policy groups, non-governmental organizations, and the civil and private sectors.

Assessments can incorporate higher-resolution data and indicators, local priorities and preferences, and develop tailored goal models and reference points, which produce scores that better reflect local realities. If a goal is not relevant in the local context, it can be excluded entirely. When you change goal models, though, it is important to capture the process in order to justify decisions that will inform the results. Similarly, pressures and resilience measures can be refined using local data and indicators. Index scores are only as good as the data on which they are based. **Finding the best data and indicators available is crucial for obtaining meaningful findings that can help inform decision-making.**

OHI assessments most often involve several **regions** within an overall **study area**. For example, in the global assessment, there were 221 regions (nations and territories) within the study area (global coastal oceans).

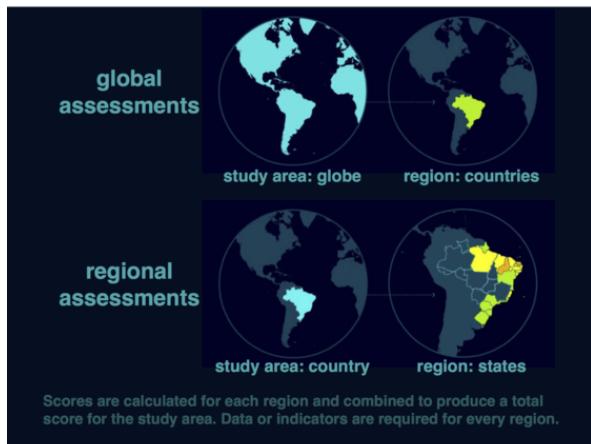


Figure 1: Figure. Global vs. regional assessments. Scores are calculated for each region and combined to produce a total score for the study area, whether at the regional, country, or global level.

Goal scores are calculated for each region separately and then combined to produce an overall Index score for the region. Index scores are combined using an area-weighted average to produce the Index score for the overall study area.

3 Overview of the OHI WebApp

Section Summary:

In this section, you will get an introduction to the OHI WebApps. When using the WebApp, you can conduct a preliminary assessment and use the built-in functions to compare input layers, output scores, and change data display options. You will return to the WebApp at the end of this phase.

OHI WebApps serve several purposes because they:

- allow for exploration of how the Index works: what data look like and which data layers are used in each goal

- are a communication platform for an assessment team, since information is presented in a manner that is accessible to group members of different disciplines and technical capacities
- can be used to set data gathering or goal model development priorities for the assessment
- display your assessment's data and calculated scores once you have finalized and formatted your data and modified goal models.

3.1 Background

OHI WebApps are websites created to facilitate independent assessments, and one is available for nearly every coastal nation or territory. The WebApps are meant to be a ‘Starter Kit’ and are available through <http://ohi-science.org> using a three-letter identifier in the URL. For example, Ecuador’s WebApp (“ECU”) is found at <http://ohi-science.org/ecu>. Each WebApp displays data layers*, which are raw data in this case, as well as the calculated OHI scores based on information extracted from global assessments. As such, they do not provide fine-scale resolution of data for each coastal nation or territory: the scores and data on which they are based are a starting point for an assessment to be conducted by an independent group. These data can be used as a default if better data for the region do not exist, but we encourage you to replace them wherever possible. (*Note: each data component that is included in the OHI is called a **data layer** because it will be combined with others to calculate the goal scores. Many data layers are rescaled from 0-1 to be combined with other data layers on the same unitless scale.)

Boundaries for exclusive economic zones (EEZs) were identified by <http://www.marineregions.org> and the largest subcountry regions (i.e., provinces, states, districts) were identified by <http://gadm.org>. Subcountry region boundaries were extended offshore to divide the EEZ of each study area into offshore regions. These subcountry regions have been provided as a starting point, and are typically coastal states or provinces, which, in our experience, is consistent with the scale at which most policy decisions are made. However, it is possible to change the boundaries for the regions and the study area depending on your preferences.

Note that the OHI+ team doesn’t take a stance on disputed territories. For independent assessments, we defer to the map-providers and the best judgment of the technical team.

The information displayed on the website is stored online, in a **GitHub repository**. GitHub is an open-source development platform allows for multiple users to collaborate, track changes, and document work such as data files and code. Therefore, any changes made to the files contained within the GitHub repository will be displayed on the WebApp for all team members to view. (See the section on **GitHub** for how to modify files using that platform).

3.2 Using the WebApp

When your team has finalized data layers and updated goal models, these data and scores will be visualized through the WebApp.

When first exploring a WebApp (for example, <http://ohi-science.org/ecu>), first note that it is possible to **translate** the site into any language that Google provides using the pull-down menu at the top. This will help with communication among your team.

The WebApp homepage provides several tabs for you to explore. The interactive **App** sub-page allows you to explore input data layers and output calculated scores for each region (See **Overview of Variable Options**). More detailed information is about the default **regions** and **data layers**, **goal models**, and **calculated scores** based on global data can be viewed in separate tabs, as well as through the App page. A quick reference about navigating the WebApp is available through the **Docs** link at the bottom of the page.

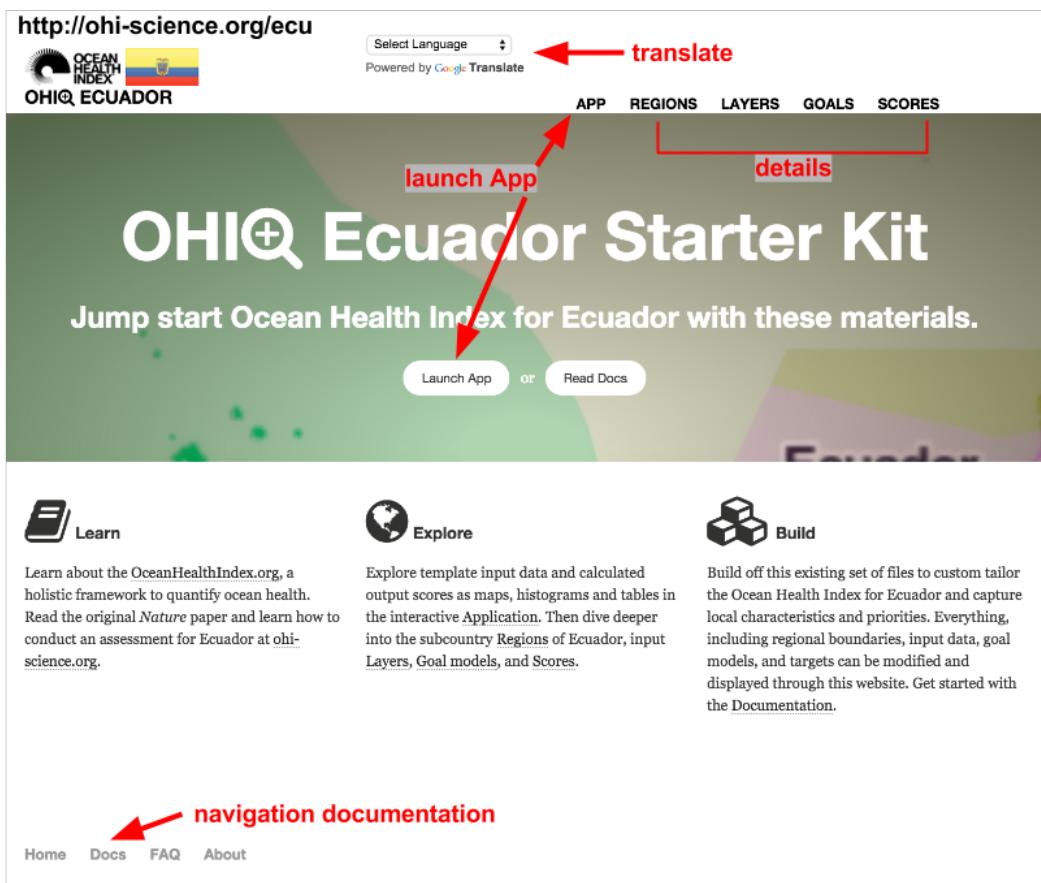


Figure 2: The WebApp start page. Note that it's possible to translate the page into your language of choice.

3.3 The App Page

The App page allows you to explore and visualize input data layers and calculated output scores for each region in the study area. By default, global data are presented for each subcountry region in the study area, and scores are calculated for each region using those data.

The App page displays this information through two tabs: Data and Compare. The **Data** tab provides several subtabs for viewing data (*Map, Histogram, Table*), and is the default tab when the Toolbox is launched. The **Compare** tab is most useful for comparing output scores when modifications are made to the underlying data or models (this provides a way to error check) once you have begun the process of calculating your own assessment.

The App provides two Branch/Scenario options to view, identified in the upper-left corner of the Data tab. The **Branch** options refer to the versions of the GitHub repository where data are stored. Branches start off as copies of the same repository, but can be modified independently of each other, enabling progress to be made on one ('draft' branch) while not altering the vetted original ('published' branch). These branches can be merged back together at any time. The App page will display the 'published branch' by default; we recommend working on the 'draft' branch until your assessment is finalized, at which point you would merge the draft branch with the published branch.

Scenario folders contain all the files needed to calculate scores. Scenario folders can differ from each other based on the years included (i.e., 2014 would be a different scenario from 2015), or they can be used to explore outcomes of policy alternatives, such as implementation of a proposed Marine Protected Area network or fisheries regulations. Running these different scenarios can be very useful for an OHI+ assessment in which multiple outcomes are to be evaluated.

3.4 The App's Data tab

3.4.1 Overview of display options

The Data tab displays input data layer or calculated scores for each goal parameter, and presents the information as a map, histogram, or table. These options (*Map, Histogram, Table*) are presented as subtabs located the map. The Map subtab is the default display option for the Data tab, and all data presented are drawn from data from the Global Assessments. This means they are either directly duplicated, or down-scaled using regional areas or population weightings.

Data displayed in the Map subtab:

The map displays data for every region as reported in the scenario. A color legend is displayed in the lower right corner of the map. The range of values will change as different variables are selected, and the colors will automatically change to create a visual scale of reference.

Data displayed in the Histogram subtab:

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

Data displayed in the Table subtab:

The table displays the target value for each region and the overall study area. It provides an identifying code (*rgn_id*), name (*rgn_name*), and value (*value*) for each.

3.4.2 Overview of variable options

The Data tab has drop-down menus from which you choose the data to be displayed. Data selected from the pull-down menus can be viewed in Map, Histogram, or Table form as described in the section above. Descriptions, statistics and metadata for the chosen fields are also displayed below the drop-down menus on the left side of the tab.

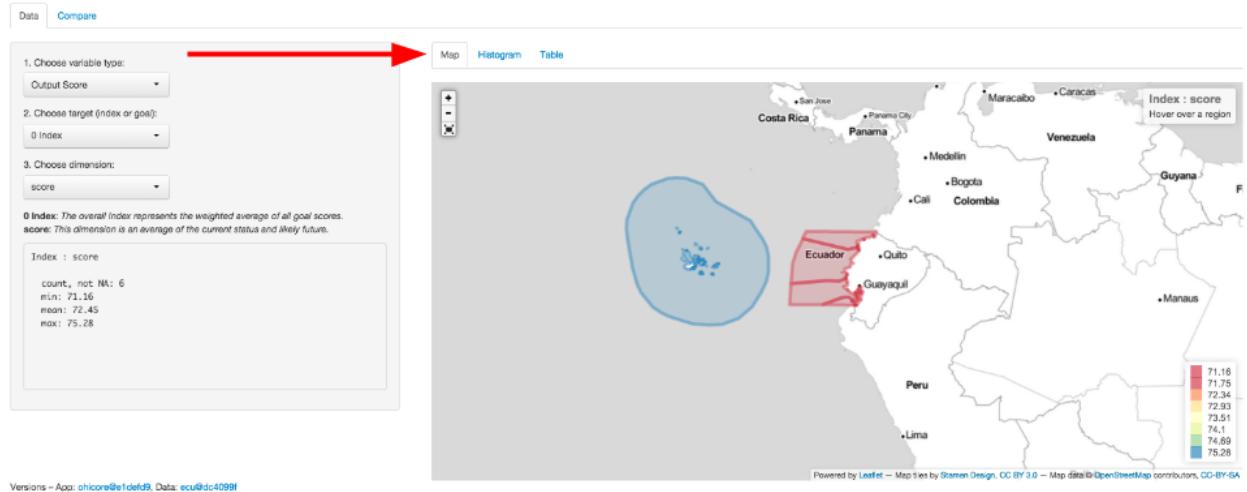


Figure 3: The Map subtab. Click on ‘Map’ to see a geographic view of your assessment region. Colors indicate scores or values for your input layers or output scores.

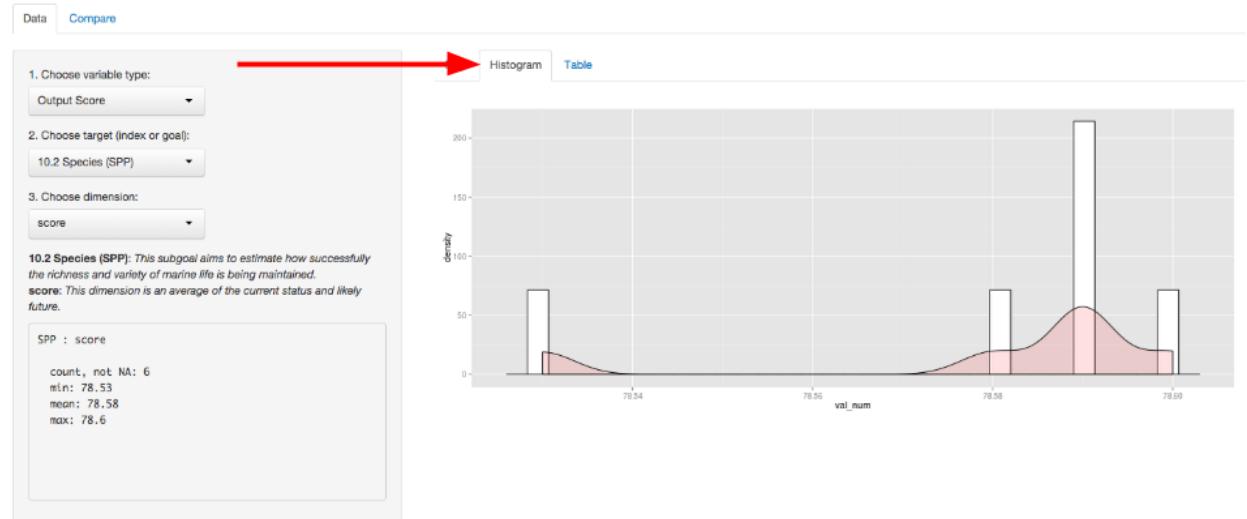


Figure 4: Click on ‘Histogram’ to see the distribution of your data or scores, after selecting a variable layer on the left. This example shows the Species sub-goal scores for the study regions of Ecuador.

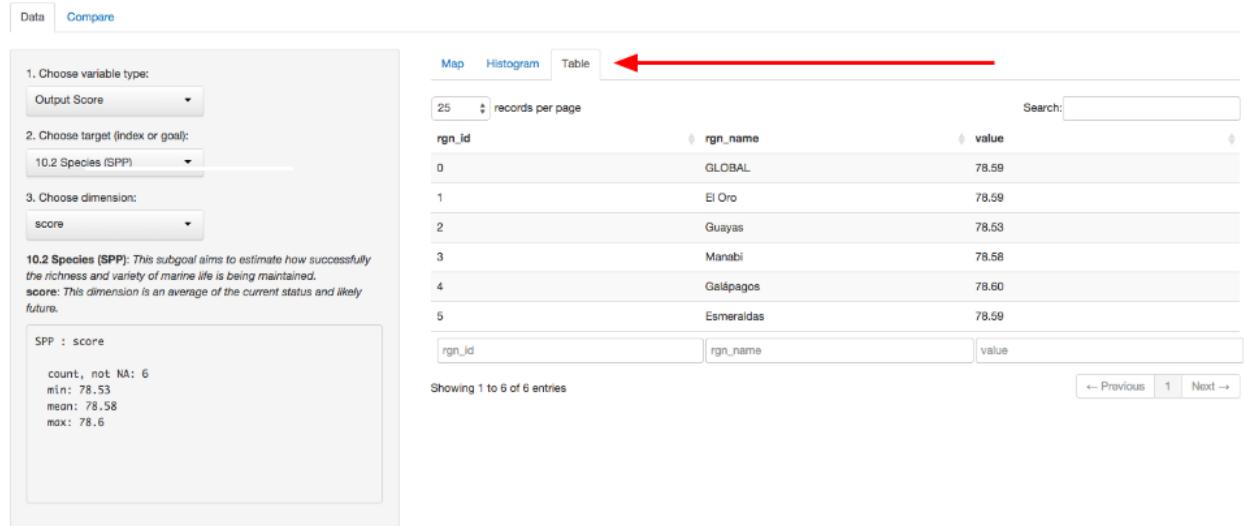


Figure 5: Click on ‘Table’ to see a table of your data or scores, after selecting a variable layer on the left. This example shows the Species sub-goal scores for the study regions of Ecuador.

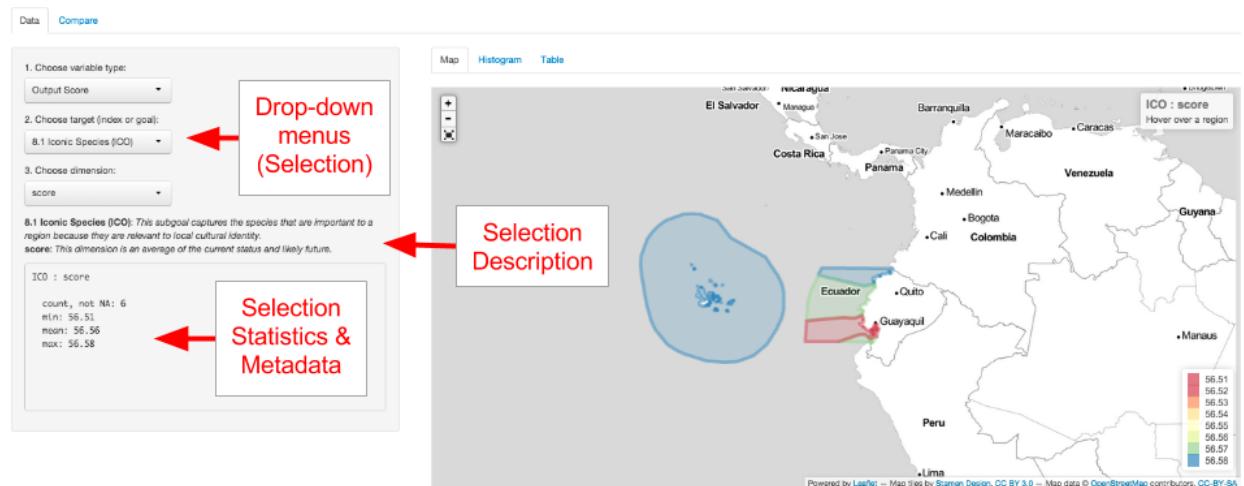


Figure 6: Overview of the Data tab. Choose the variable you would like to explore through the drop-down menus on the left-hand side of the page. Once you select either raw data or a score, you can view a description and statistical summary below.

TIP: Remember that your descriptions and values in `layers.csv` will appear here.

The first selection to be made from the drop-down menus is variable type, in which you can choose the **Output Score** that will show a calculated score (for a particular target chosen subsequently), or **Input Layer** that will show the data layer used to calculate the score of a particular target. To reiterate, Output Scores are the scores calculated using the Input Layers (data layers).

For example, if you select ‘Output Score’ as the variable type (which is the default), you will then be able to choose a target (goal or sub-goal), and the OHI dimension to be reported.

As another example, if you select ‘Input Layer’ as the variable type, you will be able to choose a target and a specific data layer associated with that target. If that layer has multiple categories or years available, you will be able to select a preference. Without selection, the default setting is the first category alphabetically and the most recent year.

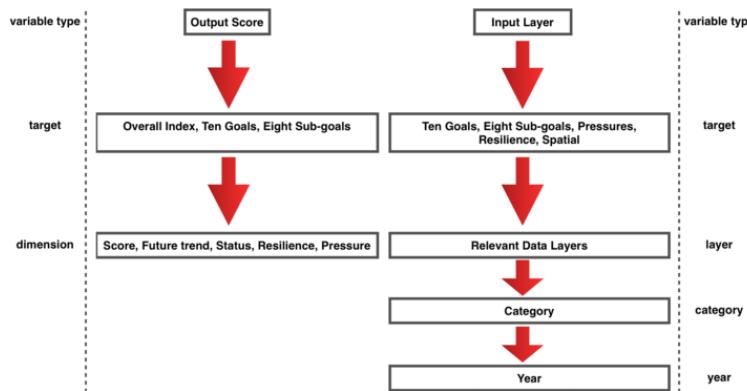


Figure 7:

3.5 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 data layers. Visualizing these differences is extremely helpful for confirming results and error checking. More context on the use of this function can be found in the section, “**The Ocean Health Index Toolbox**.”

3.6 Discovering and Gathering Appropriate Data and Indicators

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

** The accuracy of Index scores is a reflection of input data quality and the degree of understanding of the study area, and thus including the best quality and appropriate data and indicators available is of highest importance.**

Because the data and indicators you use will come from different sources and available from online databases, reports, spreadsheets and text files, 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, we call it a **data layer**. Because creating data 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.

There are many data layers included in the OHI framework. There are about eighty individual data layers from the global assessment that should be replaced with higher-resolution data in your study area where possible. You will need to search for data used to calculate status models as well as pressures and resilience layers.

3.6.1 Data sources

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

- government reports and project websites
- peer-reviewed literature
- masters and PhD theses
- university websites
- non-profit organizations

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 interpret the philosophy of each Index goal and sub-goal using the best available data and indicators. Some indicators already are scaled (e.g., from 0-1 or 0-10), and can easily be incorporated into your assessment since the reference points have already been identified.

3.6.2 Gathering responsibilities

Gathering appropriate data requires searching for and accessing existing data. You do not have to go collect the data itself, but you do need to discover and acquire existing data. It is important that team members responsible for data discovery make thoughtful decisions about whether data are appropriate for the assessment, and that they get feedback from the full team to discuss the merits of different data sources. Data discovery and acquisition are typically an iterative process, as there are both practical and philosophical reasons for including or excluding data.

When you begin exploring data possibilities, you can seek local data sources that could directly replace data from the global assessment provided in your repository. Such data would be better quality, i.e., higher accuracy and spatio-temporal resolution, than the data from the global assessment, and models may not need to change. However, we recommend first exploring other data possibilities that could capture specific characteristics to your study area. Assessments conducted at smaller scales are an opportunity to include characteristics specific to your study area that were not captured in the global assessment.

3.6.3 The process of 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.

Begin by understanding and comparing the best approaches used in assessments that have been completed, including global assessments, Brazil, Fiji, and the US West Coast. For the smaller-scale assessments (i.e., anything but the global 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. This should be a goal-by-goal process:

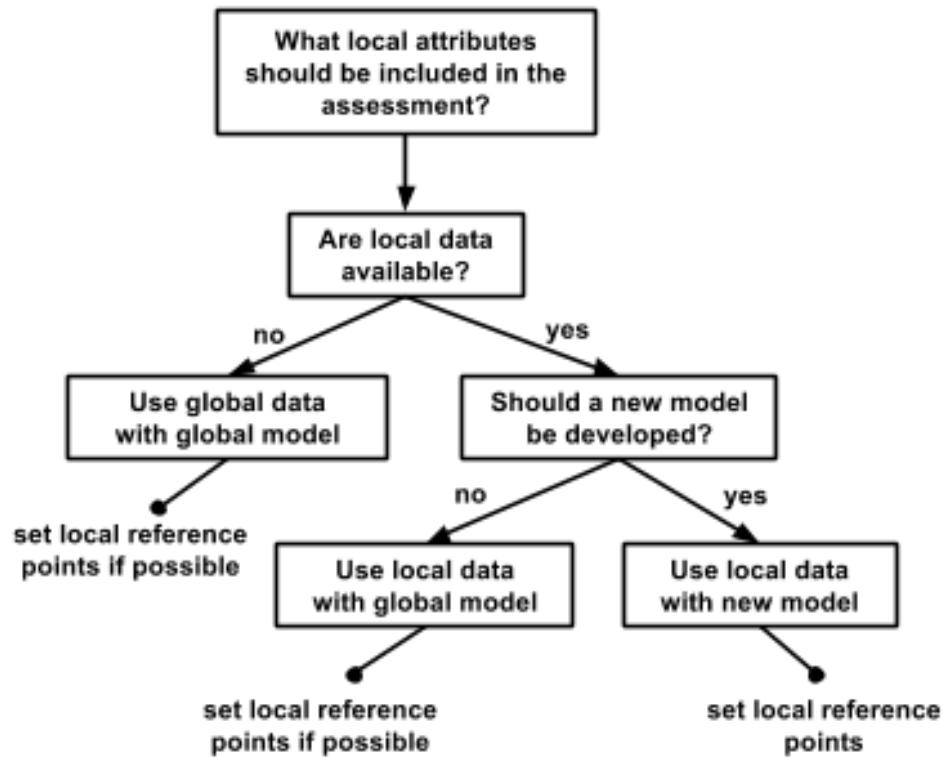


Figure 8:

3.6.4 Requirements for data and indicators

There are six requirements to remember when investigating (or ‘scoping’) potential data and indicators. It is important that data satisfy as many of these requirements as possible, at times requiring gap-filling solutions. If requirements are not met and gap-filling solutions are not possible, you will likely need to exclude a dataset from the analyses. If data cannot be included, you may elect to use the global data layers or identify other data and a different modeling approach.

1. relevance to ocean health
2. accessibility
3. quality
4. how to set the reference point
5. spatial scale
6. temporal scale

3.6.4.1 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.

3.6.4.2 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 data sources 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.

3.6.4.3 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, e.g., were there changes in the collection protocol to be aware of when interpreting temporal trends?

3.6.4.4 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?
- example: maximum sustainable yield in fisheries
- Have policy targets been set regarding these data?
- example: maximum levels of pollutants before beach closures
- Would a historic reference point be an appropriate target?
- example: percent of habitat coverage before coastal development
- Could a region within the study area be set as a spatial reference point?
- example: a certain region a leader in creating protected areas

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

3.6.4.6 Appropriate temporal scale Data must be available for at least three to five years to calculate the trend. For some goals, where temporal reference points are desirable, longer time series are preferable.*

* 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).

3.6.5 Example: US West Coast data discovery

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

3.6.5.1 Reasons data were excluded 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., resolution, coverage, or other requirements that have been listed) 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 US West Coast, kelp beds are a very important habitat because of their contribution to biodiversity and coastal protection. However, kelp coverage variation and is driven primarily by abiotic natural forcing (wave/storm disturbance and temperature) and thus it 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 US 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 US Endangered Species Act identifies a species list focused on species of concern within the US. As such, these data are biased in the context of ocean health since they 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 US 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.

3.7 Best approaches for Starting an Assessment

The Ocean Health Index team has a wealth of knowledge from years of developing the Index. However, with each new independent assessment, the team is learning even more, and we hope to share that information with you and encourage outside dialogue on the best approaches to conducting an Independent Assessment. Your OHI+ assessment will also be invaluable in learning how to develop better and more appropriate techniques to changing goal models.

3.7.0.2 Strategies for Assessment

For more information on Strategies, see the OHI Conceptual Guide sections, such as **Assembling a Team** and **Strategy**.

3.7.1 Using this Guide

For goals that have been conducted commonly across OHI+ assessments and the Global Assessment, we provide recommendations for how to approach them broadly here in the OHI Manual. For more descriptions of the models, evolution of the approach, and reference point, see the OHI **Conceptual Guide**>

3.7.2 Where to start?

Any assessment will depend on the available data, which requires creative thinking, particularly when ideal data are unavailable. Determining the appropriate scale and defining the study area and any regions within the study area will also depend on available data. The scale and resolution of available data can help inform the scale of the study area to be analyzed and how to divide it into regions.

3.7.2.1 To consider as you get started Data included to calculate scores are partly based on the philosophical framework of the Ocean Health Index (see Part I), and partly based on data available. Several goals that combine across categories have many potential data sources that could be included (or excluded): in particular the natural products, livelihoods and economies, and habitat based goals. The existence of the data itself will drive a lot of these decisions, but here are some guidelines to help determine if data are appropriate philosophically:

- Do the data represent something truly ocean-based, and natural (i.e. not man-made)?
- How would the reference point be calculated?
- Do data cover the appropriate spatial scales?
- Do data cover the appropriate temporal scales?

3.7.2.2 Which data are available? Humans interact with and depend upon the oceans in complex ways, some of which are familiar and easy to measure, such as providing seafood, transporting goods, or disposing of waste. Other benefits are more difficult to measure, such as the way marine-related jobs indirectly affect coastal communities, how different habitats mediate storm damage or the benefits people receive or perceive simply from living near the ocean. Thinking creatively and exploring the data available can make the Index more representative of reality.

It is important to think creatively and beyond the interests of a specific institution or primary field of study. Google or other internet search engines are great starting places: investigate what kinds of information are available from government and public records, scientific literature, academic studies, surveys and reports or other sources.

Data used in the Ocean Health Index spans a wide array of disciplines, both within and outside of oceanography and marine ecology. Therefore, it is necessary to look beyond the most known or obvious data sources to find data relevant for the goals in the region. 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.

3.7.2.3 What scale is appropriate? Here it is important to think about why is there interest in completing an OHI assessment. If managers or policy makers are interested, at what scale do they work? Where are the political boundaries? These questions are important to keep the OHI assessment relevant, but ultimately data availability will be most important in how to define regional boundaries for the OHI.

3.7.3 What's important spatially?

There are several spatial considerations you should have prepared before moving forward with the Toolbox. One is your map definitions for use in the assessment and in the WebApps. You must define your map regions and boundary. This can be politically drawn or, if appropriate, biogeographically drawn. Biographic considerations may be important when there are disputed territories or boundary lines in your area. *Note that OHI+ does not take a position on disputed territories, so it is up to you to vet the quality of the maps.*

Once you have the maps, you may also need to define buffer zones. Buffers are distances from shore, both inland and offland, that can be used later in the Toolbox to assess many impacts more accurately.

Finally, you will need to prepare your layers folder with spatial data. Only goals certain goals will require data layers that are spatial (see Section, **Habitat-Based Goals**). When considering spatial data, you should be conscientious of how it is presented: is it by square kilometers, or by a kind of region such as an EEZ or other identifier? Do you have better local data than global satellite or modeled data? Once you find the data, consider how far back in time it goes. Can you set a reference point with this data, or do you have to find another dataset or other source of information to find a historical reference point?

TIP: The only thing to do is remember that you are trying to capture ocean health—too far inland may no longer relate to ocean health.

3.7.3.1 Reference Points The decisions on choosing a reference point will be a theme in each of the goal models you develop. The choice of a reference point will affect how the final scores are calculated, and will have to be balanced between limitations of the data and expert judgment to assess the conditions of the various dimensions of ocean health.

In addition to set values, such as maximum or minimum observed value, it's possible to use spatial reference points and temporal reference points. In spatial reference points, you find the highest-scoring region and say that it is perfect. It's also possible to set a different reference point for each region of your study area, as was done in the U.S. West Coast study (2014). Using temporal reference points, a historical benchmark is used as a the “ideal” point in the past. A third type of reference point is a policy-set 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. 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—they should be “Specific,” “Measurable,” “Ambitious,” “Realistic,” and “Time-bound.”

After selecting the type of eref

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

For more, see the **Conceptual Guide**

3.7.4 Practical Guidance

3.7.4.1 Synergies and Goals to approach together

TIP: See section, **Assembling a Team**

Whether you are working goal-by-goal, or data layer by data layer, it is important to consider where you can find synergies. This will save you time, and effort, and will help build a picture of how your goals are interlocking in a way that is reflective of the actual linkages that exist in the coupled human and natural systems you are studying. Some examples include:

3.7.4.2 Do Habitat-Based Goals Together You should do the habitat-based goals together. These include **Carbon Storage**, **Coastal Protection**, **Biodiversity: Habitats** in the Global Assessment. This will be more efficient because the spatial analyst can do the data manipulation to create the spatial layers that get used for these goals, based off the same source material. This will greatly expedite your data search, and layer preparation. 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 these three goals.

If you look at `functions.R`, you will see that the reason this is the case is because the data layer, `Hab_extent`, is used in multiple places in the software:

- NP
- FIS
- CS
- CP

```

750 CP = function(layers){
751
752   # sum mangrove_offshore1km + mangrove_inland1km = mangrove to match with extent and trend
753   m = layers$data[["hab_extent"]] %>%
754     filter(habitat %in% c('mangrove_inland1km', 'mangrove_offshore1km')) %>%
755     select(rgn_id, habitat, km2)
756
757 if (nrow(m) > 0){
758   m = m %>%
759     group_by(rgn_id) %>%
760     summarize(km2 = sum(km2, na.rm=T)) %>%
761     mutate(habitat='mangrove') %>%
762     ungroup()
763 }

```

habitat layer

Figure 9: Note that Habitat Extent appears in several goal model functions.

3.7.4.3 Do Fisheries goals and pressures together Do the **Wild-Caught Fisheries** model and the fisheries-related pressures together. This is because you will already be searching for data on commercial high bycatch and low-bycatch, which are also used as pressures layers. While you are working on **Mariculture**, you should also consider how your

3.7.4.4 Linkages: Note the connections between pressures and goals Climate change appears in several places in the global index, and will certainly influence numerous goals in your assessment. You will see that climate pressures include UV radiation, Sea Surface Temperature (SST), sea-level rise (SLR), ocean acidification, and influence in the Global model the goals for Natural Products, Carbon Storage, Coastal Protection, Sense of Place, Livelihoods and Economies, and Biodiversity.

In the case of Global 2012, the **Clean Waters** goal is very much linked to the pressures because the same data layers for pressures are used as the input layers for the status. Note that because trash is also a pressure, it affects other goals as well, including **Tourism and Recreation**, **Lasting Special Places**, **Livelihoods and Economies**, and **Species**.

3.7.5 Habitat-based Goals

3.7.5.1 Practical Considerations

Habitat-based goals can be approached together in the data gathering process. Habitats appear in such goals as **Carbon Storage**, **Coastal Protection**, and **Biodiversity**. Your approach will largely depend on the type of data you have.

3.7.5.2 Ideal Approach Ideally, information on the extent and condition of every single habitat type would be available. The reference point for habitat-based goals will likely be temporal. This means that historic data are needed such that current habitat and value data can be compared to historic data. This could also be achieved by using or a proportion of historic data.

3.7.5.3 Ready to start on Habitat-based Goals? The first thing to consider with habitats is **what habitats are in your area**. You should consider what habitats are in your coastal regions as well as your offshore regions throughout your EEZ. Are there mangroves, coral reefs, seagrass beds, salt marshes, sea ice, or subtidal soft-bottom habitats? Are there kelp forests in your waters?

Once you determine what habitats are in your area, you should also consider **their range extent within your reporting regions**. You should consider whether and how far they go offshore and inland. You will likely use a combination of clipped spatial data and spatial data for whole regions, depending on the goal being assessed. For instance, the **Habitat** sub-goal of **Biodiversity** uses the entire region for habitats such as coral reefs in the calculation, but the **Coastal Protection** only uses the extent to 1 kilometer offshore and inland to 1 kilometer. This will affect the preparation of the spatial data layers later in your assessment.

For instance, do mangroves occur in river deltas in your area, and if so, how far upriver do they go? In the 2013 Global Assessment, for example, mangrove was assessed from 1 kilometer offshore to 1 kilometer inland, but this distance is variable and could change for your OHI+ assessments. It depends on your geography.

Several factors for habitats will have to be considered. In score calculations, habitats are weighted in different ways based on the amount of protection they provide and the amount carbon they sequester, and they are also derived from estimations of habitat health and condition (See **Carbon Storage**, **Coastal Protection**, and **Biodiversity**). The weights of the habitats relative to each other comes from the literature. If you have access to regional-specific studies, you should try to use that information. Otherwise, the Global data layers can be used as a secondary option.

When thinking about condition, for instance, coral health is based on the percentage of “living cover” on a reef relative to the potential range of the reef.

In the Global Assessment, coral reference points used the mean predicted values for a country from 1985-1987. When this data was not available for a country, we used the value of neighboring countries. There was also some analyses to control for the fact that the % live coral cover naturally varies over time, hence the use of “predicted values”.

In the 2012 Global Assessment, soft-bottom habitat had an estimated status based on the “intensity of trawl fishing” as a proxy (pages 51 and 52 of SOM 2012).

It is also important to think about the temporal aspect of your habitats. This will affect the **trend** and the choice of the **reference point**. Trend is the change in health or condition over time. However, this might be constrained by data availability due to the heterogeneity of habitat types and data sources. Ideally, there will be enough years of data to directly calculate the recent change in health of the habitat by using a

linear model to calculate 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.

For example, for sea ice in the Global Assessment, a linear model was set to a metric of sea ice cover for data from 2006-2011, after some smoothing to account for natural variation.

Overall, the historical **reference point** is particularly important for the habitat goals. What's the ideal reference point, however? You could consider setting the reference point further back in history if you have the data for it. This would assume there was a more pristine condition in the past, and would be based on the assumption that habitat destruction has been and still is occurring. Here it is useful to use the S.M.A.R.T. principles (See **Conceptual Guide**). It is highly encouraged to set an ambitious reference point, and deviate from those which were set in the Global.

3.7.6 Ready to start the Carbon Storage goal?

3.7.6.1 Ideal Approach Ideally, information would be available to allow you to assess the amount of carbon stored in every coastal habitat in your area. The best information would show area covered and some measure of the quality of the habitat for this goal. For example, a dense mangrove forest would be better quality than a sparse mangrove forest. Additionally, different weights would be assigned to the habitats based on their relative ability to store carbon, and this factor could be used to penalize areas where there is greater loss of carbon and reward areas where this is a greater storage of carbon. The carbon storage model could then incorporate such weights in a similar way to the **Coastal Protection** goal. The reference point for habitat-based goals will likely be temporal in most cases. This means that historic data are needed so that current habitat values can be compared to a past state.

TIP: 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.

3.7.6.2 Practical Approach The model and approach for **Carbon Storage** developed for the Global assessment will likely be appropriate for independent assessments. Knowing the area and the condition of carbon-storing habitats are the two components that are most important. It is important to spend time looking for a way to quantify condition, but it may not be possible to find.

One of the first decisions for your team to make is **whether to use the given data** in the Global Assessments or **to find and use a better dataset**. Even though your goal model shouldn't change entirely from the Global, you should find more localized data than the global data, because the global data layers have already been processed and include clipped regions that reduce the specificity of the information for your area. This is because in the default WebApp methods, the habitat area was calculated by using the global areal coverage of the habitat weighted by the size of the offshore area of the region. This is not consistent currently in the WebApps; for example, if you look at [the given layers a WebApp](#), you will see that the same values are applied to all regions. You should think here, however, of how your regions' size compare to any spatial dataset's scale, and return to the question of spatial coverage once you are thinking about the **reference points** farther on.

Questions to consider: Do you have maps, for example, 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. You could also use summarized habitat data that exists in tables or are already compiled in another source.

The types of habitat data you search for also matter. It is recommended to search for mangroves, saltmarshes, and sea grasses because these are viewed as carbon-sequestering habitats that are both ecologically threatened and sensitive to policy responses. 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. This would also affect other kinds of indicator considerations not discussed in the Global Assessments, such as chlorophyll abundance.

TIP: if you look at your default data on the WebApp, you will see **Carbon Storage** information presented for **mangroves**, **salt marshes**, and **sea grass** even if there are no mangroves in your area. You will not be scored on these if they are not in your area.

The second group of items to consider are the **condition** and **carbon contribution**. **Condition** is more essential than **contribution** in this case because it is a more direct measure of ecosystem health and would give a more accurate score for this goal. **Condition** means . It can be measured by a number of ways, often indirectly through studies on density of the habitats, disease affecting the particular habitats, or other impacts such as change in species composition or growth rates from impacts such as overgrazing. Has there been a study assessing habitat integrity or health specifically in your area? **Contribution** is how much each habitat stores relatively to the others—such as the rates of carbon uptake as measured by empirical data. For this you would have to go to the literature and find ratios of organic nutrient uptake between habitats, and you would have to make sure these studies are done correctly to represent your area. For example, were the studies done with a young mangrove forest, or an older one, which might have different growth rates?

In the 2012 Global Assessment, the condition of mangroves was assessed as, “Current square kilometers of mangrove coverage divided by the reference hectares, calculated from 1980 to 2005. Seagrass condition was calculated as,”Current percent cover, or hectares, of habitat divided by the reference percentage cover, or hectares from 1975 to 2010.”

A vital consideration for this goal is **the reference point**. It is an important decision to be made based on good science given the limitations of available data. The reference point is the extent of the habitat at some time in the past, in other words the ideal condition. You will need historical data for this—either from satellites, published papers, or even hand-drawn maps. For questions of sustainability, you must consider, what is a good year to set as your historical past extent? The reference year should be identified based on SMART principles, particularly that it is an ambitious goal, but also realistic. If you do not have trend data, you will have to use a proxy for the trend.

In the U.S. West Coast assessment (2014), researchers went to the local public library to find hand-drawn maps of historical salt marsh extents in California.

In this case, perhaps your choice of historical reference point may be guided by a policy target as well. 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?

3.7.6.3 Ready to start on Coastal Protection?

Coastal Protection aims to assess the amount of protection provided by marine and coastal habitats against flooding and erosion to coastal areas that people value.

3.7.6.4 Ideal Approach Ideally, data for all habitats within a region would be available, as well as information on the value of the land and the vulnerability of inhabitants being protected by these habitats. This requires data for habitat type at high spatial resolution as well as a measure of the value of what is protected by the habitats. The reference point for habitat-based goals will likely be a temporal baseline; this means that historic data are needed such that current habitat and value data can be compared to them.

3.7.6.5 Practical Approach The areas relevant to this goal include both inhabited places, such as homes and other structures, and uninhabited places such as parks and special places. However, no credit is given for sea-walls, because they are not regarded as sustainable. At local and regional scales data may exist on all these variables at a high enough resolution to map and calculate exactly which habitats are providing how much protection to which coastal areas. You must first decide which data sources to use for your maps. Do you have spatially-explicit satellite data? If so, how far back in time does it go? The data consideration will be a consideration when you set the reference points.

In the U.S. West Coast assessment (2014), researchers went to the local public library to find hand-drawn maps of historical sand-dunes in California, for example, in order to find information for sand dune extent in the 1850s.

In addition, one might want to know the level of vulnerability of the different coastal communities. Vulnerability can be quantified as the ability to evacuate, fragility of constructions, economic ability to reconstruct in case of damage. The reason this is done is to prioritize the protection of certain locations. Physical properties may be available in regional studies, allowing for more a detailed understanding of the protective ability, and likelihood of exposure for each habitat type in different portions of the coastline.

As with the other habitat-based goals, you will need to make careful decisions about your reference point. It's an important decision to be made based on good science given the limitations of available data. The reference point is the extent of the habitat at some time in the past, in other words the ideal condition. You will need historical data for this—either from satellites, published papers, or even hand-drawn maps. For questions of sustainability, you must consider, what is a good year to set as your historical past extent? Do you want to be ambitious and set it far in the past, so it gets closer to a more “pristine” condition? Or do you want to make it represent human use? Will it make your scores look better or worse if you choose a more recent or more distant example?

3.7.6.6 Linking Habitat and Biodiversity It is important to think about how habitat relate to the **Biodiversity** goal. The current approach tracks **Species** and **Habitats** separately because of imperfect data for **Species**. In other words, habitats are an estimation of the health of the species in an area. Therefore if you have comprehensive species assessments in your area you wouldn't need the **Habitat** sub-goal. However, the inventory of marine biodiversity may not often be sufficiently complete in your area.

You should also think about the linkages between the habitat-based goals, such as **Carbon Storage** and **Coastal Protection**, and the pressures and resilience. Does the encroachment or reduction of these habitats have an effect on carbon sequestration, or does it change ecosystem composition sufficiently to impact species? Could habitats have an effect on **Clean Waters**, or are they impacted more by local pollution than other impacts? If habitats change for natural reasons, you might not need to worry about their pressures on other goals, but if they consist of introduced species, you should think about the effects on other goals.

For example, the invasive marsh plants such as *Spartina* may afford some Carbon Storage, but it's a non-native habitat to an area such as the San Francisco Bay. It could still count towards the **Carbon Storage** goal, but would also be a pressure on **Biodiversity**.

3.7.6.7 Practical Considerations

The **Natural Products** goal describes how sustainably people harvest non-food products from the sea.

3.7.6.8 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 products depending on what is harvested in the study area, including corals, shells, seaweeds, aquarium fish, mangrove wood, or any non-food marine product that is harvested within a region. 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.

3.7.6.9 Ready to start on Natural Products? 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.

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 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 (See **Best Approaches**).

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 (See **Best Approaches**).

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.

$$S_p = 1 - \left(\frac{E + R}{N} \right)$$

Figure 10: Natural Products goal model from OHI Global Assessment 2013

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.

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.

[insert an image of time-series of harvest by product in region X?]

Example: The Global assessment used the following information in the Natural Products equations:

product	relative tonnes (1)	weighting (2)	Exposure (3)	Risk (4)
coral	FAO	FAO	coral habitat	all 1
sponges	FAO	FAO	coral + rocky reef habitat	all 0
ornamentals	FAO	FAO	coral + rocky reef habitat	1 if blast/cyanide fishing, otherwise 0
fish oil	FAO	FAO	fish score/100	-
shells	FAO	FAO	coral + rocky reef habitat	all 0
seaweeds	FAO	FAO	rocky reef habitat	-

3.8 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 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.

3.8.1 *Ideal Approach:*

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.

3.8.1.1 Pressure and resilience categories The Toolbox calculates pressures in five ecological pressure categories (pollution, habitat destruction, fishing pressure, species pollution, and climate change) and one social pressure category. 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.

3.8.2 Pressures and resilience matrices

After you identify the pressures and resilience measures for your study area and gather available data for each region, you will use matrix tables to determine how each of the measures affects each goal and sub-goal (for some goals you will also need to do this for habitat type or natural product categories). The pressures matrix establishes the relationships between stressors and goals are determined, and 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.

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*). In the pressures matrix ranks are categorized as follows:

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

Stressors that have no impact are not included 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.

	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 11: 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.

While pressures and resilience are usually displayed as two-dimensional matrices, they are actually three-dimensional matrices: each stressor should have data for each region in the study area, which is the third dimension of the matrix, as shown in the figure. The Toolbox will combine the data with the rank weights to calculate the pressures scores. Therefore, pressure weights should not be applied to the regions, but only to the goals. The Toolbox will multiply the stressor score for each region by the weight (1, 2, 3) assigned to that pressure for a specific goal and subgoal, and then it will combined that score within its appropriate pressure category as previously discussed.

Each pressure with a rank weight of 2 or 3 should have a corresponding resilience measure, which is meant to ‘balance’ the pressures since these have the greatest effect on ocean health. The Ocean Health Index considers resilience in two categories: **ecological resilience** to address ecological pressures, and **social resilience**,

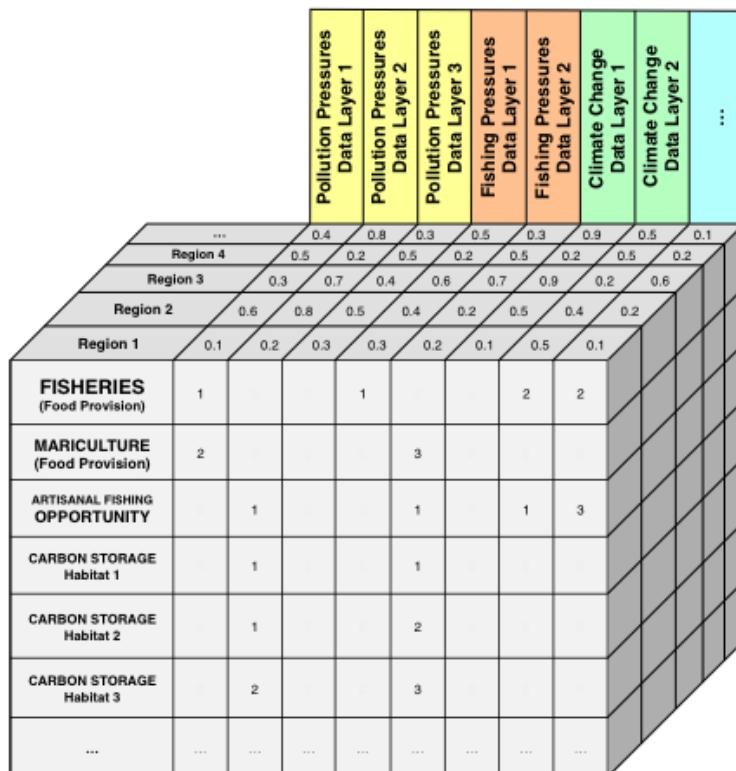


Figure 12: The pressures matrix is three-dimensional: each pressure layer has data per region, which is multiplied by the ranking weights of the pressures matrix.

which may not be strictly marine-related, but they can help estimate how a region may be able to respond to or prevent new 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?
- **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**.

3.8.3 Incorporating local pressures in your assessment

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 (See **Including pressures from global assessments**). 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 include 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.

3.8.3.1 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.

3.9 Guide to searching for resilience metrics

3.9.1 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.

3.9.1.1 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.

3.9.1.2 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.

3.9.1.3 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.

In the Brazil study (2014), the marine trash resilience was calculated by counting up whether localities had one of four garbage management services, including access to beach clean-up services, household garbage collection, household recycling collection, and garbage collection in public streets.

3.9.2 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.

3.9.2.1 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.

4 The Ocean Health Index Toolbox

Section Summary:

In this section, you will learn the basics of how to use OHI tools for conducting an assessment. You will be introduced to the files you will be working with, how to prepare them, and you will learn how concepts such as status, trend, pressures, and resilience are used together to create the final score. You will also learn what to do in cases of missing data.

TIP: Knowing where your data gaps are will make gapfilling easier in this process.

The **OHI Toolbox** is an ecosystem of data, scripts, and structure required to calculate OHI scores at any scale. Toolbox scripts are open source, written in the software language R, and data inputted into the Toolbox are **comma-separated-value**, or **.csv** files, which can be created or edited using text editors or Microsoft Excel. Files are stored within two folders called **repositories**, or **(repos)**, such that

OHI Toolbox = your assessment repo + `ohi core functions` repo.

We access and interact with the Toolbox ecosystem through an online collaborative platform called **GitHub**. GitHub stores the **R** scripts and **.csv** files in a folder called a repository, which is found online and can also be downloaded on your computer and synced with the online version. GitHub tracks changes by all collaborators working on the project through time, and saves all versions for comparison. The section, “**Installing the Toolbox**,” provides instruction on how to download GitHub repositories to your computer, but everything is also available online.

The Toolbox is used to calculate final scores. But, perhaps more importantly, it can also be used to organize an assessment, including data identification and management. The Toolbox can additionally be used to compare how different management scenarios could affect overall ocean health, which can inform effective strategies for ocean resource management at a local scale.

4.1 File System for Assessment Repositories

This section is an orientation to the files within your assessment repository. The file system structure is the same whether you view your assessment repository online or after downloading or cloning to your computer (see section, “[Installing the Toolbox](#)”).

Throughout this example, we will use Ecuador’s assessment repository as a guide. It’s available at <https://github.com/OHI-Science/ecu>.

TIP: Once you know the layout of the repository and the purpose of the files within it, you can plan an appropriate workflow with your team. See the [Using the Toolbox](#) introduction for more.

4.1.1 Assessments and scenarios

Your *assessment repository* contains a *scenario folder*, which by default is named `subcountry2014`. This scenario folder contains all the files needed to calculate scores, and they are described in detail below.

The scenario folder is named `subcountry2014` because it contains data for your country used in the 2014 global assessment. These data in most cases were attributed equally to all regions within your study area (for example, data used for Ecuador in the global assessment was attributed to all coastal states in the files within `subcountry2014`).

You will be able to rename your scenario folder to better reflect the spatial and temporal scale of your scenario after you have set up your GitHub account. We recommend that the name defines the scale of the regions and the year. Eventually, you will likely have multiple scenario folders that contain data for subsequent years or modifications to explore policy alternatives.

In the above figure, `ecu` is the **assessment repository** and `subcountry2014` is the **scenario folder**. Note that files with names preceded by a ‘?’ do not appear when not viewing from github.com; this is because these files are specific to GitHub.

Within the `subcountry2014` folder area all the inputs required by the Toolbox. See [Modifying and Creating Data Layers](#) for more information on the files you will commonly modify.

4.1.2 `layers.csv`

TIP: Keep `layers.csv` handy. It’s a very useful reference throughout the assessment process.

`layers.csv` is the registry that manages all data required for the assessment. All relevant data are prepared as a ‘data layer’ and registered in this file. The Toolbox will rely on information from this file to use the data layers and display information on the WebApp. You will update some of the layers in `layers.csv`, and some of them will be auto-generated by the Toolbox code when it’s running.

When you open `layers.csv`, you’ll see that each row of information represents a specific data layer that has been prepared for the Toolbox. The first columns (`targets`, `layer`, `name`, `description`, `fld_value`, `units`, `filename`) contain information that will be updated by your team as you incorporate your own data and edits; all other columns are generated later by the Toolbox as it confirms data formatting and content. The first columns have the following information:

- **targets** indicates which goal or dimension uses the data layer. Goals are indicated with two-letter codes and sub-goals are indicated with three-letter codes, with pressures, resilience, and spatial layers indicated separately.
- Food Provision (FP): Fisheries (FIS) and Mariculture (MAR)

<https://github.com/OHI-Science/ecu>

The screenshot shows the GitHub repository page for 'OHI-Science / ecu'. At the top, there's a navigation bar with 'GitHub', 'This repository', 'Search', 'Explore', 'Features', 'Enterprise', and 'Blog'. Below the navigation, the repository name 'OHI-Science / ecu' is displayed with a dropdown menu icon. A red arrow points from this area to the URL 'http://ohi-science.org/ecu' located below the repository name. Another red arrow points to a commit in the list, specifically to the file 'subcountry2014'. The commit list includes the following entries:

File	Commit Message	Time Ago
.travis.yml	Update .travis.yml	2 months ago
subcountry2014	Update goals.Rmd	2 months ago
.Rbuildignore	auto-calculate from commit c449dfc	2 months ago
.gitignore	adding debug files to .gitignore	2 months ago
.travis.yml	Update .travis.yml	2 months ago
README.md	Update README.md	2 months ago
ecu.Rproj	install_github git2r	2 months ago

Figure 13:

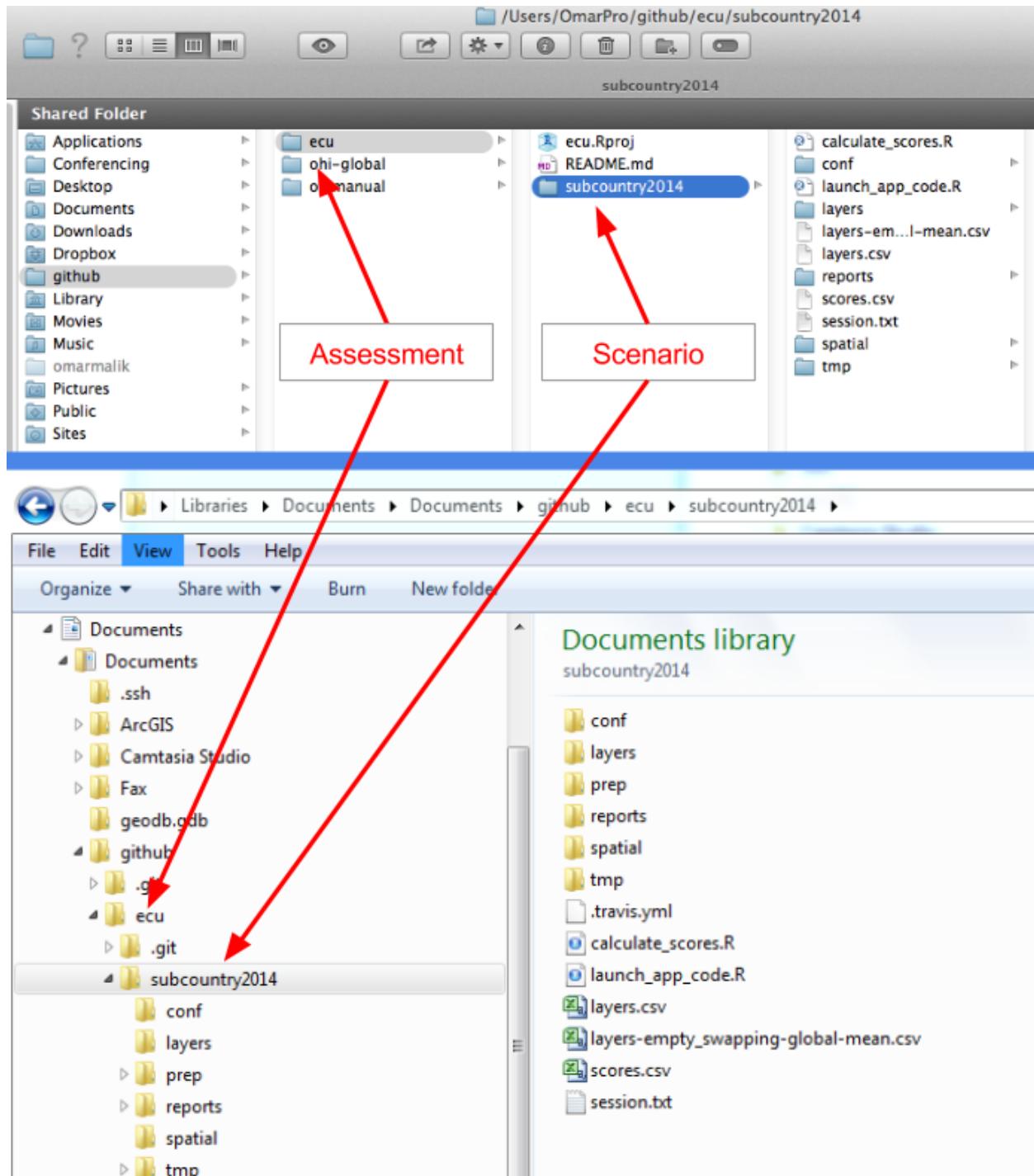


Figure 14: Navigating the assessment repository. The figure shows Mac folder navigation above and Windows navigation below.

1	targets	layer	name	description	fld_value	units	filename
2	AO	ao_access	Fisheries management	The opportunity for value	value		ao_access.csv
3	AO	ao_need	Purchasing power	The per capita pu value	value		ao_need.csv
4	CW	cw_coastalpopn_trend	Coastal human population	Coastal population trend	trend score		cw_coastalpopn_trend
5	CW	cw_fertilizer_trend	Fertilizer consumption	Statistics on fertilizer trend	trend score		cw_fertilizer_trend
6	CW	cw_pathogen_trend	Trends in access	Trends in percent trend	trend score		cw_pathogen_trend
7	CW	cw_pesticide_trend	Pesticide consumption	Statistics on pesticide trend	trend score		cw_pesticide_trend
8	FIS	fis_b_bmsy	B/Bmsy estimates obtained using the R script	B / B _{MSY}	B / B _{MSY}		fis_b_bmsy.csv
9	FIS	fis_meancatch	Catch data for each species	Reported data in mean catch	metric tons		fis_meancatch.csv

Figure 15:

- Artisanal Fishing Opportunity (AO)
- Natural Products (NP)
- Coastal Protection (CP)
- Carbon Storage (CS)
- Livelihoods and Economies (LE): Livelihoods (LIV) and Economies (ECO)
- Tourism and Recreation (TR)
- Sense of Place: Lasting Special Places (LSP) and Iconic Species (ICO)
- Clean Waters (CW)
- Biodiversity (BD): Habitats (HAB) and Species (SPP)
- **layer** is the identifying name of the data layer, which will be used in R scripts like `functions.R` and `.csv` files like `pressures_matrix.csv` and `resilience_matrix.csv`. This is also displayed on the WebApp under the drop-down menu when the variable type is ‘input layer’.
- **name** is a longer title of the data layer; this is displayed on the WebApp under the drop-down menu when the variable type is ‘input layer’.
- **description** is further description of the data layer; this is also displayed on the WebApp under the drop-down menu when the variable type is ‘input layer’.
- **fld_value** indicates the units along with the units column.
- **units** unit of measure in which the data are reported.
- **filename** is the `.csv` filename that holds the data layer information, and is located in the folder `subcountry2014/layers`.

4.1.3 *layers* folder

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

Note that each `.csv` file within the `layers` folder has a specific format that the Toolbox expects and requires. Comma separated value files (`.csv` files) can be opened with text editor software, or will open by default by Microsoft Excel or similar software.

Now, open the `layers/alien_species.csv` file: note the unique region identifier (`rgn_id`) with a single associated `score` or `value`, and that the data are presented in ‘long format’ with minimal columns. See the section on *Formatting Data for the Toolbox* for further details and instructions. Scores can be viewed through the WebApp using the ‘Input Layer’ pulldown menu on the App page.

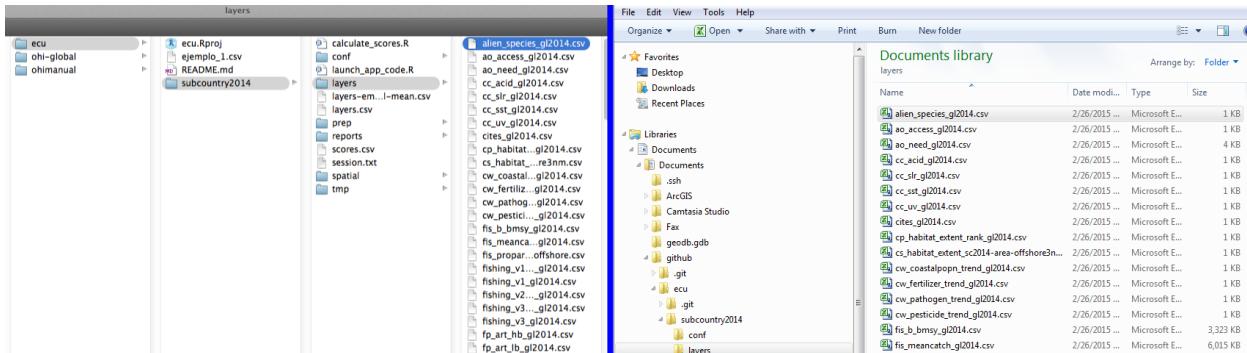


Figure 16: The `layers` folder contains every data layer as an individual `.csv` file. Mac navigation is shown on the left and Windows navigation is shown on the right.

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

4.1.4 `conf` folder

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

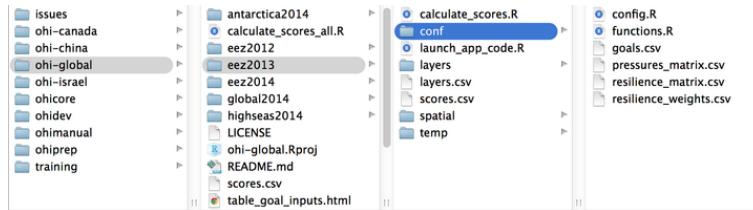


Figure 17: The `conf` folder contains important R functions and `.csv` files. Mac navigation is shown on the left and Windows is shown on the right.

4.1.4.1 `config.R`

`config.R` is an R script that configures labeling and constants appropriately.

4.1.4.2 `functions.R` `functions.R` contains functions for each goal and sub-goal model, which calculate the status and trend using data layers identified as ‘layers’ in `layers.csv`. When you modify or develop new goal models, you will modify `functions.R`.

TIP: It’s useful to skip to different sections of `functions.R` to see how key calculations are being done. See section, **Update `Functions.R`**.

4.1.4.3 `goals.csv` `goals.csv` is a list of goals and sub-goals and their weights used to calculate the final score for each goal. Other information includes the goal description that is also presented in the 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 the weightings and preindex functions if you’re planning to change the goal or sub-goal models.

4.1.4.4 *pressures_matrix.csv* *pressures_matrix.csv* defines the different types of ocean pressures and the goals they affect.

Each column in the pressures matrix identifies a data layer that is also registered in *layers.csv*: and has a prefix (for example: *po_* for the pollution category). The pressure data layers are also required to have a value for every region in the study area, with the region scores ranging from 0-1.

4.1.4.5 *resilience_matrix.csv* *resilience_matrix.csv* defines the different types of resilience with the goals that they affect.

Like the pressures matrix, the resilience matrix also has weights depending on the level of protection. However, these weights are in a separate file: *resilience_weights.csv*.

Each column in the resilience matrix is a data layer that is also registered in *layers.csv*. Resilience layers, like the pressure layers, are also required to have a value for every region in the study area. Resilience layers each have a score between 0-1.

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

4.1.5 *install_ohicore.R*

This script will install *ohicore*, the engine behind all Toolbox calculations. You will only need to run this script only one time.

4.1.6 *launch_app_code.R*

The Toolbox can be launched on your computer so that you can visualize any edits you make while you are offline. To do this, you will run the code in *launch_app_code.R*. Make sure you are in the *subcountry2014* directory at that time: `setwd("~/github/ecu/subcountry2014")`

4.1.7 *calculate_scores.R*

calculate_scores.R is a script that tells the Toolbox to calculate scores using the *.csv* files in the *layers* folder that are registered in *layers.csv* and the configurations identified in *config.R*. Scores will be saved in *scores.csv*.

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'.

4.1.8 *scores.csv*

scores.csv contains the calculated scores for the assessment. Currently, these scores were calculated using data for your country from the global 2014 assessment. Scores are reported for each dimension (future, pressures, resilience, score, status, trend) for each region in the study area (with region identifier), and are presented in 'long' format. Scores can be viewed through the WebApp using the 'Output Score' pulldown menu on the 'App' page.

~/github/chn - draft - RStudio

Excel File Edit View Insert Format Tools Data Window Help

~/github/chn - draft - RStudio

config.R calculate_scores.R data_prep.r cp_extent_chn2015_zb.csv

Environment History Build Git

Calibri (Body) 12 B I U

Home Layout Tables Charts Data

Global Environment

	D	E	F
1	description	fid_value	units
2	The opportunity	value	ao_access_gj2014.csv
3	The per capita	value	ao_need_gj2014.csv
4	Coastal popu	trend score	cw_coastalpoop_trend_gj2014.csv
5	Statistics on	trend.score	cw_fertilizer_trend_gj2014.csv
6	Trends in net trend	trend.score	cw_pathogen_trend_gj2014.csv
			cw_pesticide_trend_gj2014.csv
			fs_b_bmy_gj2014.csv
			fs_meancatch_gj2014.csv
			fs_propara_saup2rgn_s
			fs_ut_chn2015_LZH.csv
			fs_ft_chn2015_LZH.csv
			fs_mmyy_chn2015_LZH.csv
			fp_wildcaught_weight_gj2014.csv
			hab_extent_gj2014.csv
			hab_health_gj2014.csv
			hab_trend_gj2014.csv
			cs_contribution_chn2015.csv
			cs_condition_chn2015_H.csv
			cs_extreme_chn2015_HHM.csv
			cs_extreme_chn2015_HHM.csv
			cp_condition_chn2015_zb.csv
			cp_extent_chn2015_zb.csv

These errors mean your data layers have not been registered in 'layers.csv' correctly.

```

load required libraries
suppressWarnings(require(ohicore))

# set working directory to the scenario directory, ie containing conf and layers
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_file)

# 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)

```

1:1 (Top Level) Script

Console ~/github/chn/province2015/

- fi_ut: fi_ut_chn2015_name.csv
- fi_ft: fi_ft_chn2015_name.csv
- fi_tc: fi_tc_chn2015_name.csv
- fi_mmy: fi_mmy_chn2015_name.csv
- cp_condition: cp_condition_chn2015_zb
- cp_extent: cp_extent_chn2015_zb
- le_livjob: le_livjob_chn2015_zb
- le_livwage: le_livwage_chn2015_zb
- le_eco: le_eco_chn2015_zb
- mar_yk: mar_yk_chn2015_name.csv
- mar_smk: mar_smk_chn2015_name.csv
- mar_ac: mar_ac_chn2015_name.csv

index_descr total area of hectare Vietname

25	International category	category	lco_spp_extinction_status
26	International popn trend	trend score	lco_spp_popn_trend_gj2014.csv
27	Number of a value	value	lco_species_chn2015_YM.csv
28	Gross Domestic usd	2010 USD	le_gdp_sc2014.popn-infl
29	gapfilled	value	le_jobs_sector_year_gj2014.csv
30	gapfilled	percent	percent unerlo_unemployment_gj2014.csv
31	gapfilled	value	2010 USD
32	adjusted wot jobs	jobs	le_worforceadj_size_gj2014.csv
33	Jobs weight/weight	value	le_sector_weight_gj2014.csv
34	Population d count	count	le_popn_gj2014.csv
35	number of el 10k people	10,000 peop	le_livjob_chn2015_zb.csv

cc_uv_gj2014.csv 240 B 2013, 0:1 AM

cites_gj2014.csv 74 B May 27, 2015, 8:01 AM

cp_condition_chn2015_zb.csv 440 B May 27, 2015, 8:01 AM

cp_extent_chn2015_zb.csv 1.5 KB May 27, 2015, 8:01 AM

Figure 18: You can error-check your data layer registration if you see a ‘missing files’ warning when running `calculate_scores.R`.

4.1.9 spatial folder

The spatial folder contains a single file, `regions_gcs.js`. This is a spatial file in the JSON format; it spatially identifies the study area and regions for the assessment. If you plan to modify your study area or regions, you will need to upload a `.js` file with appropriate offshore boundaries. You will need a GIS analyst to do this: see http://ohi-science.org/pages/create_regions.html for some instruction.

4.1.10 layers-empty_swapping-global-mean.csv

This file contains a list of data layers that were used in the global assessment while not for your country. Without these data for your country, global averages are included in your `subcountry2014` scenario folder so the Toolbox can calculate scores until you replace these data with appropriate data for your study area. This file is not used anywhere by the Toolbox but is a registry of data layers that should prioritize to be replaced with your own local data layers.

4.1.11 Relaunching the Toolbox

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

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).

Example of data in the appropriate format:

rgn_id	year	count	rgn_id	product	year	tonnes
1	2005	177.14	1	ornamentals	2005	10327
1	2006	201.39	1	ornamentals	2006	10389
1	2007	199.81	1	ornamentals	2007	10897
1	2008	212.99	1	ornamentals	2008	9985
1	2009	228.81	1	ornamentals	2009	9001
2	2005	580.98	2	shells	2005	6179
2	2006	730.18	2	shells	2006	6823
2	2007	717.00	2	shells	2007	8239
2	2008	851.44	2	shells	2008	8819
2	2009	836.68	2	shells	2009	9205
3	2005	129.69	3	coral	2005	22079
3	2006	173.45	3	coral	2006	25297
3	2007	229.86	3	coral	2007	25361
3	2008	231.44	3	coral	2008	23817
3	2009	221.95	3	coral	2009	23623
4	2005	397.00	4	shells	2005	7500
4	2006	566.00	4	shells	2006	9700
4	2007	591.00	4	shells	2007	8600
4	2008	1154.00	4	shells	2008	9400
4	2009	1570.00	4	shells	2009	9300

Figure 19:

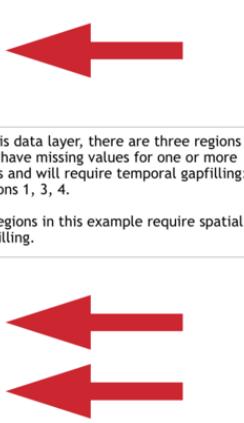
4.2.2 Gapfilling

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.

4.2.2.1 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.



rgn_id	year	count	Annotations			
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.

Figure 20:

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).

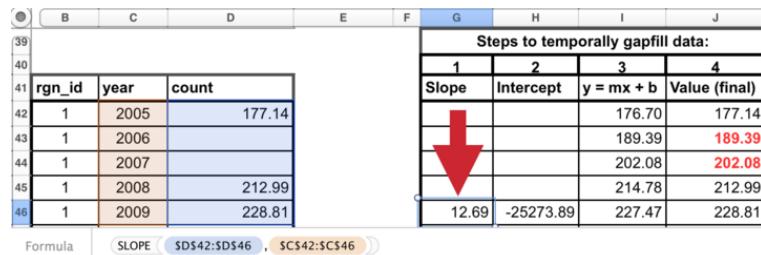


Figure 21:

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.

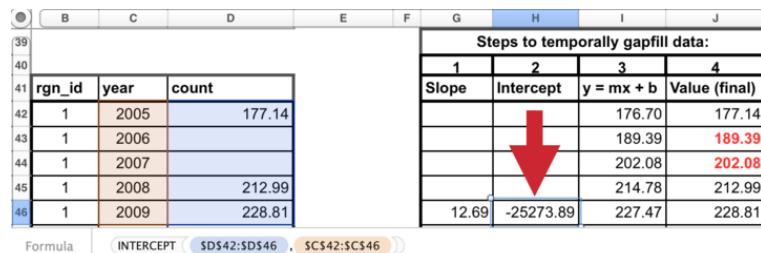


Figure 22:

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).

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.*

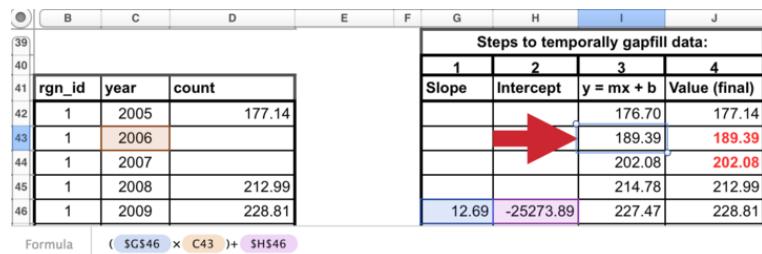


Figure 23:

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

The figure shows a Microsoft Excel spreadsheet with a large table containing three columns: 'rgn_id', 'year', and 'count'. The data spans multiple years (2005-2009) and regions (1-4). A callout box with a red arrow pointing to it contains the following text: 'In this data layer, Region 2 is missing from this dataset and requires spatial gapfilling.' Below this, another box states 'No temporal gapfilling is required in this example.'

Figure 24:

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:

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.

	A	B	C	D	E	F	G	H
42	rgn_id	year	count			rgn_id	year	count
43	1	2005	177.14			1	2005	177.14
44	1	2006	201.39			1	2006	201.39
45	1	2007	199.81			1	2007	199.81
46	1	2008	212.99			1	2008	212.99
47	1	2009	228.81			1	2009	228.81
48	2	2005				2	2005	153.42
49	2	2006				2	2006	187.42
50	2	2007				2	2007	214.84
51	2	2008				2	2008	222.22
52	2	2009				2	2009	225.38
53	3	2005	129.69			3	2005	129.69
54	3	2006	173.45			3	2006	173.45
55	3	2007	229.86			3	2007	229.86
56	3	2008	231.44			3	2008	231.44
57	3	2009	221.95			3	2009	221.95
58	4	2005	397.00			4	2005	397.00
59	4	2006	566.00			4	2006	566.00
60	4	2007	591.00			4	2007	591.00
61	4	2008	1154.00			4	2008	1154.00
62	4	2009	1570.00			4	2009	1570.00

Figure 25:

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

4.2.3 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:

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	0.7062	1	1	
B	governance_indicator					0.8564	0.7794	0.8615	1	1	
C	governance_indicator					0.8779	1	1	0.8986	1	
D	governance_indicator					0.8537	0.5373	0.7044	1	1	

Figure 26:

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

Transforming data into ‘narrow’ format:

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

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

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

Example code using the `melt` command in the `reshape2` library. Assume the data above is in a variable called `data_wide`:

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

Figure 27:

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 28:

4.3 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.

4.3.0.1 Example: Global Data Approach You should base your decision on whether you 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

Section Summary:

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

The **OHI Toolbox** is essentially several folders containing all the files required for an OHI assessment. These folders are stored online on www.github.com, and are called **GitHub repositories**. At this point, you should already be familiar with your assessment's repository, and all of the files it contains (if not, read the section, “**File System for Assessment Repositories**”).

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. Then, it explains how to after modifying files on your own computer, you can use GitHub to upload any modifications you make so that you can work collaboratively with your team.

Required software:

1. **Github App**
2. **git**
3. **R**
4. **RStudio**



Figure 29:

5.2 GitHub

GitHub 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: <http://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: <https://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*. In addition to cloning your GitHub repository to your computer, you will need to download and install *git* software and the GitHub App (application), both of which 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.

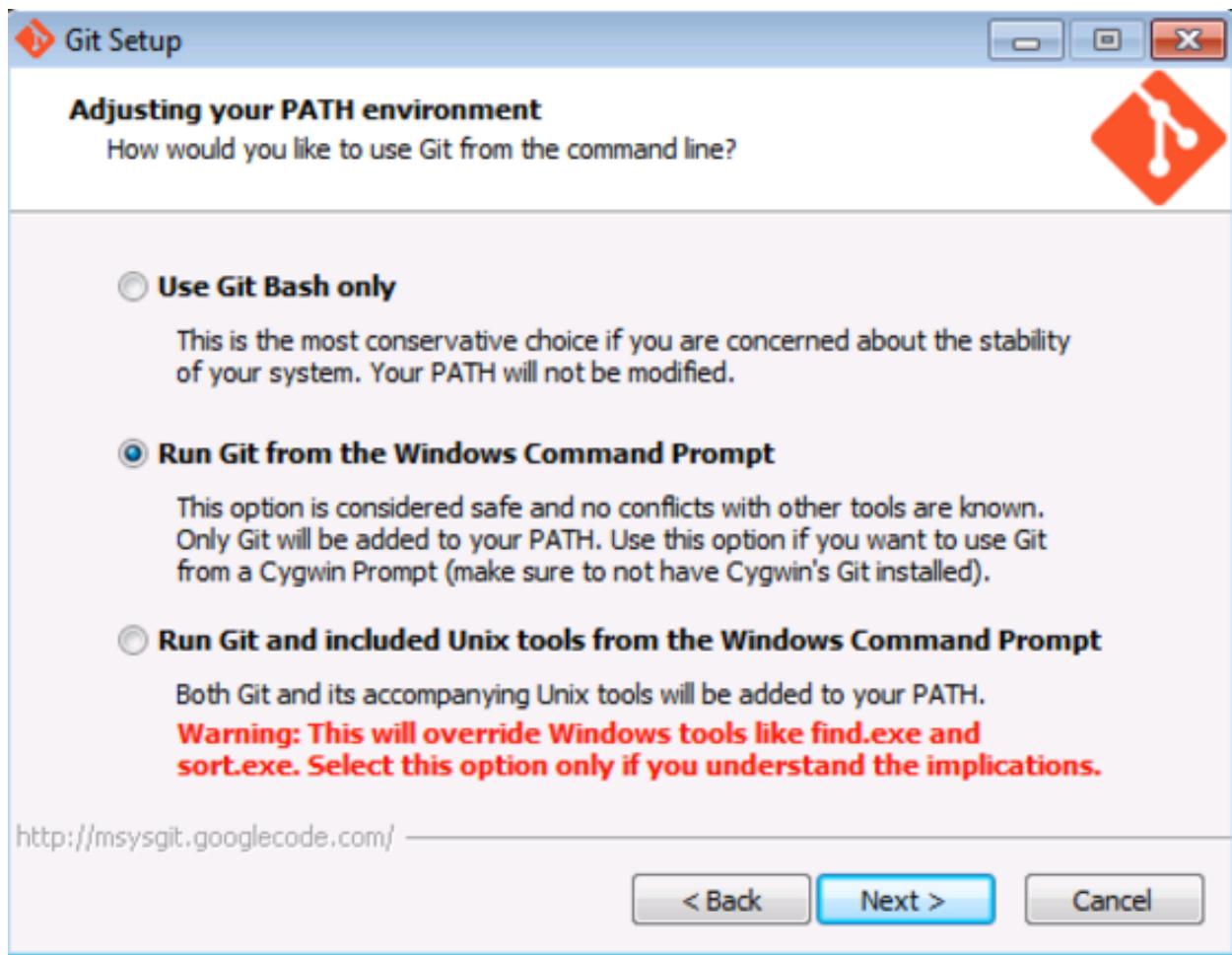


Figure 30:

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'. Type:

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

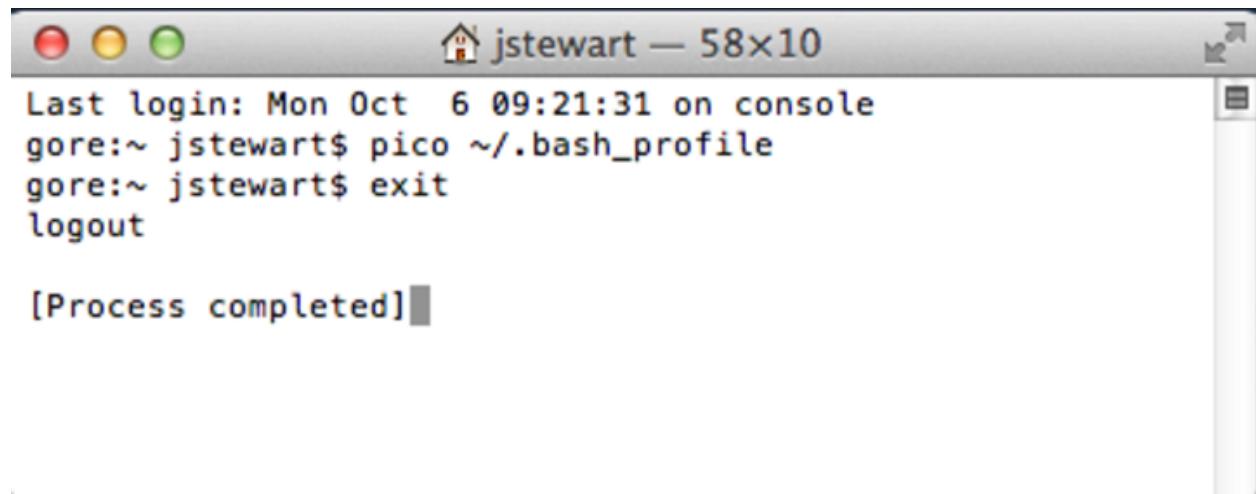
Exit pico by typing:

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

Exit Terminal by typing:

```
exit
```

Finally, quit Terminal.



```
Last login: Mon Oct  6 09:21:31 on console
gore:~ jstewart$ pico ~/.bash_profile
gore:~ jstewart$ exit
logout

[Process completed]
```

Figure 31:

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 Install the GitHub application

There are several options to clone your repository to your local machine. When getting started, we recommend using the GitHub application. This is freely available for download. Follow the default instructions for downloading and installing from the following:

- **Windows:** <https://windows.github.com/>.
- **Mac:** <https://mac.github.com/>.

5.3.5 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\[assessment]`
- Mac: `/Users/[User]/github/[assessment]`

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

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.

5.3.6 Clone your repository to your computer

Clone a repository by clicking the ‘Clone in Desktop’ button on your online repository’s homepage ([https://github.com/OHI-Science/{\[assessment\]}](https://github.com/OHI-Science/{[assessment]})):

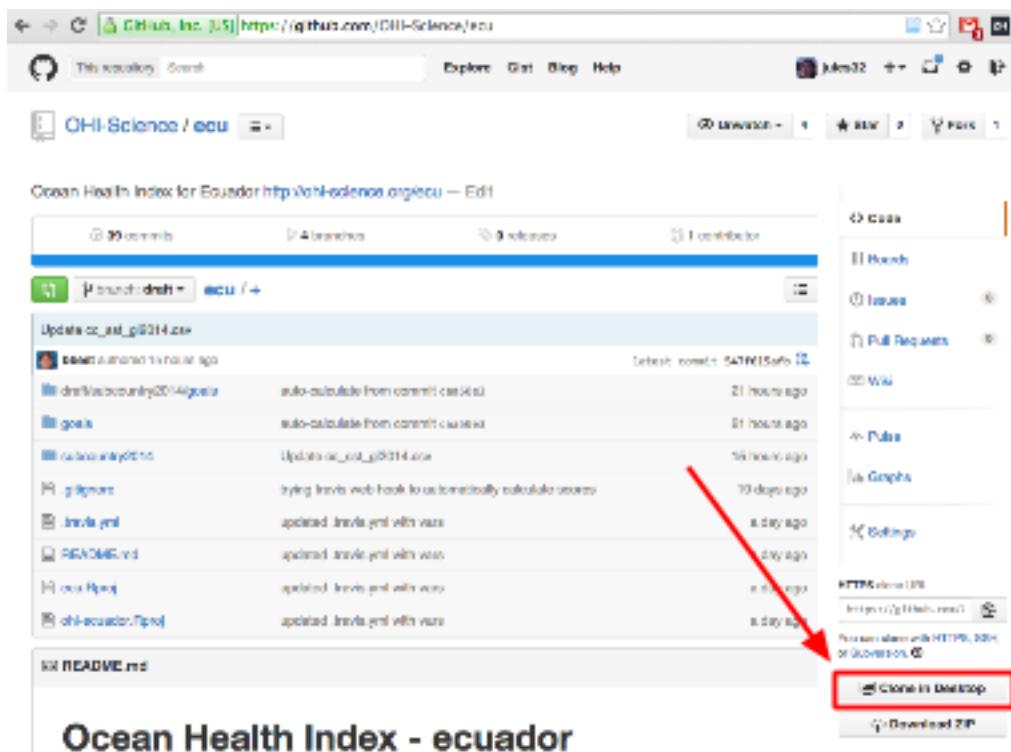


Figure 32:

You will be asked where to save this repository: save it into the `github` folder you created. The file path for your assessment will therefore be:

- Windows: `C:\Users\[User]\Documents\github\[assessment]` (example: `C:\Users\johndoe\Documents\github\ohi-ecuador`)
- on a Mac: `/Users/[User]/github/[assessment]` (example: `/Users/johndoe/github/ecu`)

The assessment can be identified by any computer as `~/github/[assessment]`.

The entire folder will now be saved on your computer.

5.3.7 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.

5.3.8 Work locally

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 with the `ohi-israel` repo owned by OHI-Science:

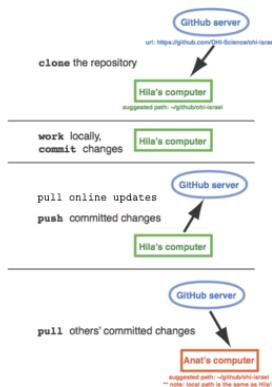


Figure 33:

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. When modifying R scripts such as `functions.R`, you will need to work in R.

We recommend syncing with either the GitHub App or with RStudio. Both methods require you to commit your changes, before pulling any updates and pushing your modifications. The GitHub App combines the pulling and pushing into one step, called syncing. The following sections show you how to synchronize the repository on your computer with the repository online.

5.3.9 Syncing

When you work on your computer, any edits you make to any files in your repo, using any program, will be tracked by *git*. You can use any of the above to commit and sync your changes back to GitHub. There are many options you can use to sync your edits on a repo with the online version.

- [GitHub App for Mac](#) and [for Windows](#)
- [RStudio](#)
- [Command line](#)

If you are just modifying data *.csv* files, you probably only need to use the GitHub App. RStudio is convenient if you are working with *.R* files. Also, the command line can be used by those interested, and there are resources available online.

TIP: Once you sync your repository, the updated information will be automatically available to the WebApps.

5.3.10 Using the GitHub App to synchronize your repository

The GitHub App will track your modifications and can be used to commit and sync any changes made locally to your repository. Once you are done working on the pertinent files and wish to commit and sync the changes to the online server on the Github server, open the GitHub App. The following example is with the *ecu* repository:

1. Make sure you select the correct repository, located on the left column of the GitHub App window (Step 1 in the figure).
2. Select the different files to which changes have been made (2a), and preview those changes on the right column of the GitHub App window (2b).
3. Once all the changes have been reviewed, write a summary/description in the respective message bars in the GitHub App window (3), then click on ‘Commit’ (3a) and then ‘Sync’ (3b) located on the top-right corner of the GitHub App window (Note: If a **Commit** button appears instead of **Commit & Sync**, you can either click **Commit** and then click the **Sync** in this way, or you can alternatively select *Edit > Automatically Sync After Committing* which will then allow you to click on ‘Commit and Sync’)

Go online and check that your changes are now visible on GitHub online.

5.3.11 Working with R and RStudio

RStudio is a program that can be used to synchronize any modifications you make to files in your assessment’s repository, and if you are working in R, it is convenient since you do not need to open the GitHub App. If you do not already have this installed, install the latest version of R and RStudio (and if you do have these installed, check for updates: there are frequent updates to the R software, and the current version is identified on the website). Both R and RStudio are freely available to download.

R: Download the current version of R appropriate for your operating system at <http://cran.r-project.org/> and follow the instructions to install it on your computer. If updating, compare the available version on their website with what you already have on your computer by typing `sessionInfo()` into your R console.

RStudio: Download the current version of RStudio software at www.rstudio.com. RStudio is not updated as often as R, but it is good to check for updates regularly. Note that in this case, you should follow the default install instructions.

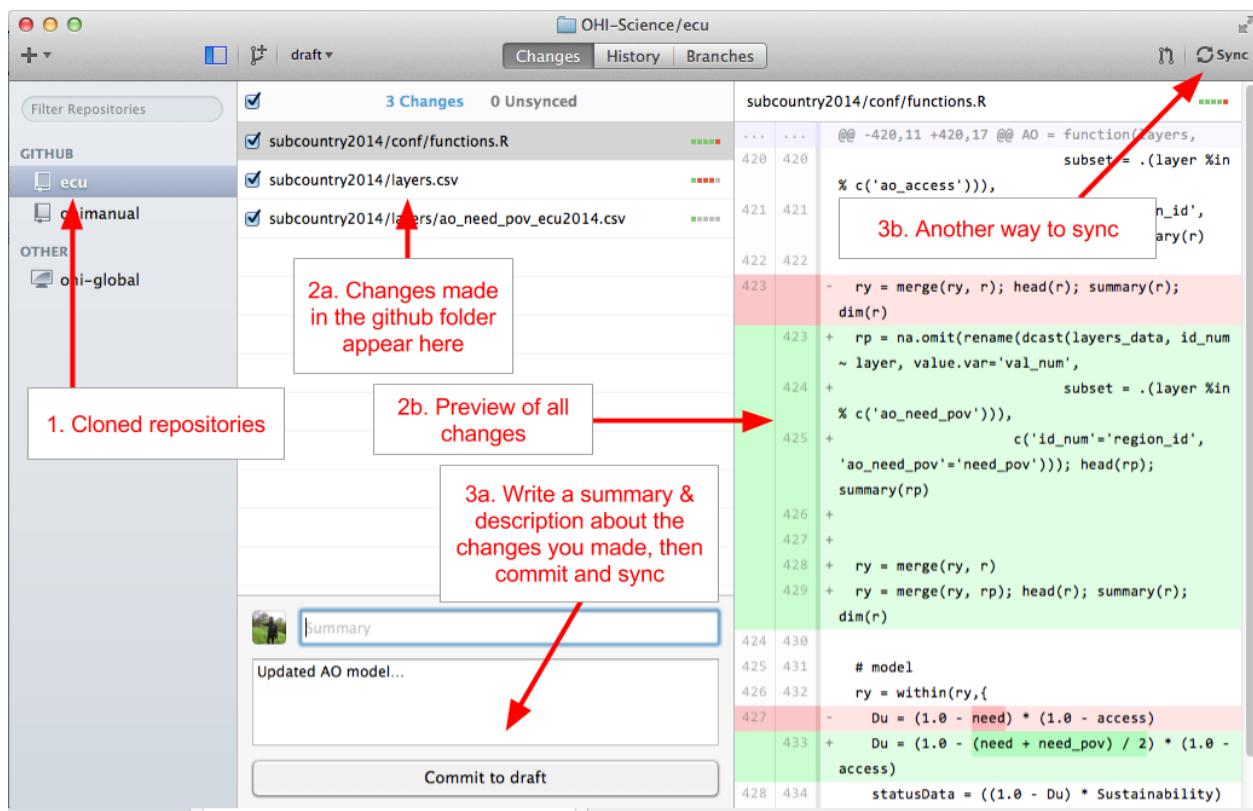


Figure 34: Figure showing the layout of the GitHub App when syncing. Click on ‘Commit’ and then ‘Sync’ to push changes to your repository.

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

5.3.12 Using RStudio to synchronize your repository

RStudio can sync files with GitHub directly, and can be used instead of the GitHub App. Like the GitHub App, it will capture the changes made to any files within the repository, no matter which software was used to modify them. The advantage for using RStudio to sync instead of the GitHub App is if you are working with R scripts already. 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.

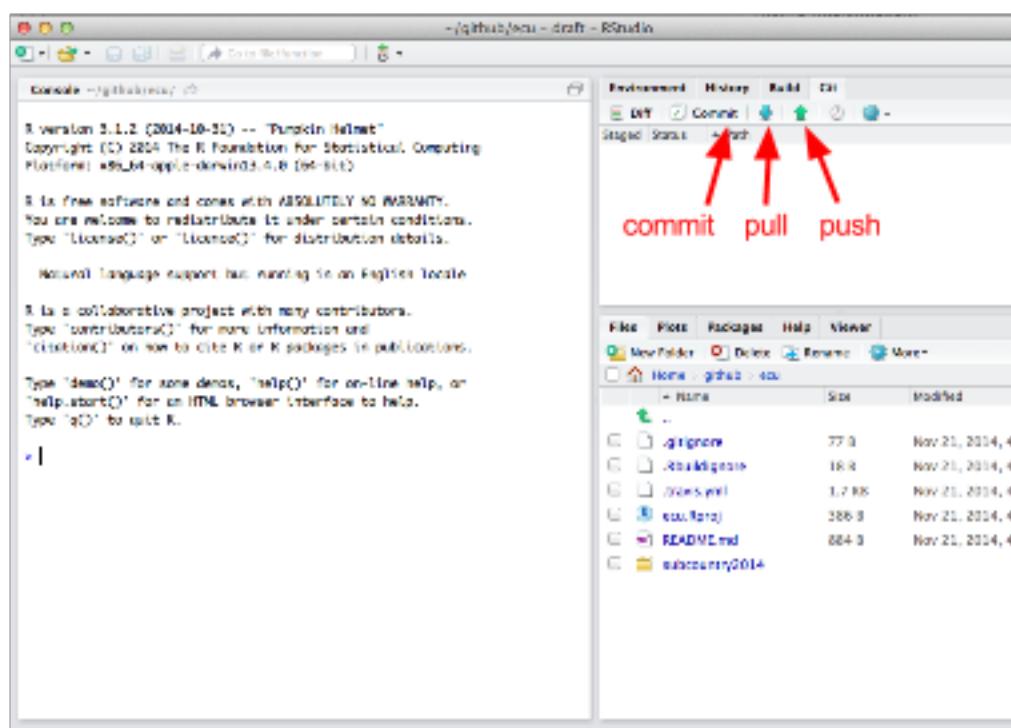


Figure 35:

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’ 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.
- 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.

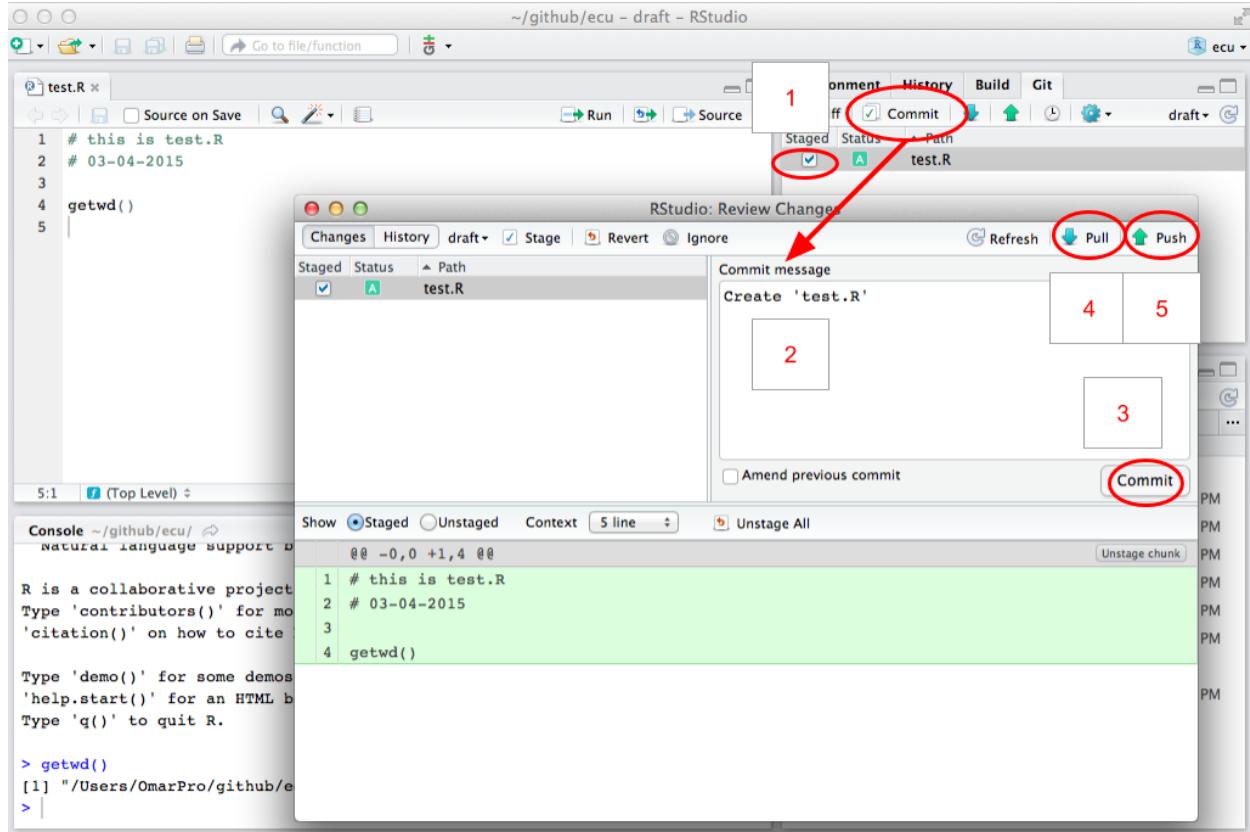


Figure 36: Figure showing RStudio when syncing. After first staging your changes, click the ‘commit’ button to open a new window where you can enter a ‘commit message’ and then pull and push new changes.

TIP: Another way to sync and open the project is to click on ‘New Project’ in the upper-right-hand corner of Rstudio, then choose ‘Version Control’, and then you can paste the URL of the desired repository. This URL can be found on your online repository’s homepage.

5.3.13 Install the latest version of R and RStudio

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

While not required, we highly recommend working with **RStudio**, which is an interface that makes working with R much easier, and it also interfaces with GitHub so you are able to synchronize without using the GitHub App. RStudio does not get updated as often as R does, but it is good to check for updates regularly.

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

Section Summary:

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.

TIP: You should now have your assessment repository opened and be familiar with the files in the folder.

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. Data layer files can be created and updated with any software that handles *.csv* files, but goal models must be updated in R. With any modifications you sync to the online repository, the Toolbox will automatically recalculate goal scores. Calculations can also be done locally and offline by running `subcountry2014/calculate_scores.R`.

This section gives instruction and examples for the most common modifications you will make to your repository:

- **modifying pressures and resilience matrices**
- **modifying and creating data layers for status, trend, pressures and resilience**
- **modifying goal models**
- **removing goals**

The files you will modify are:

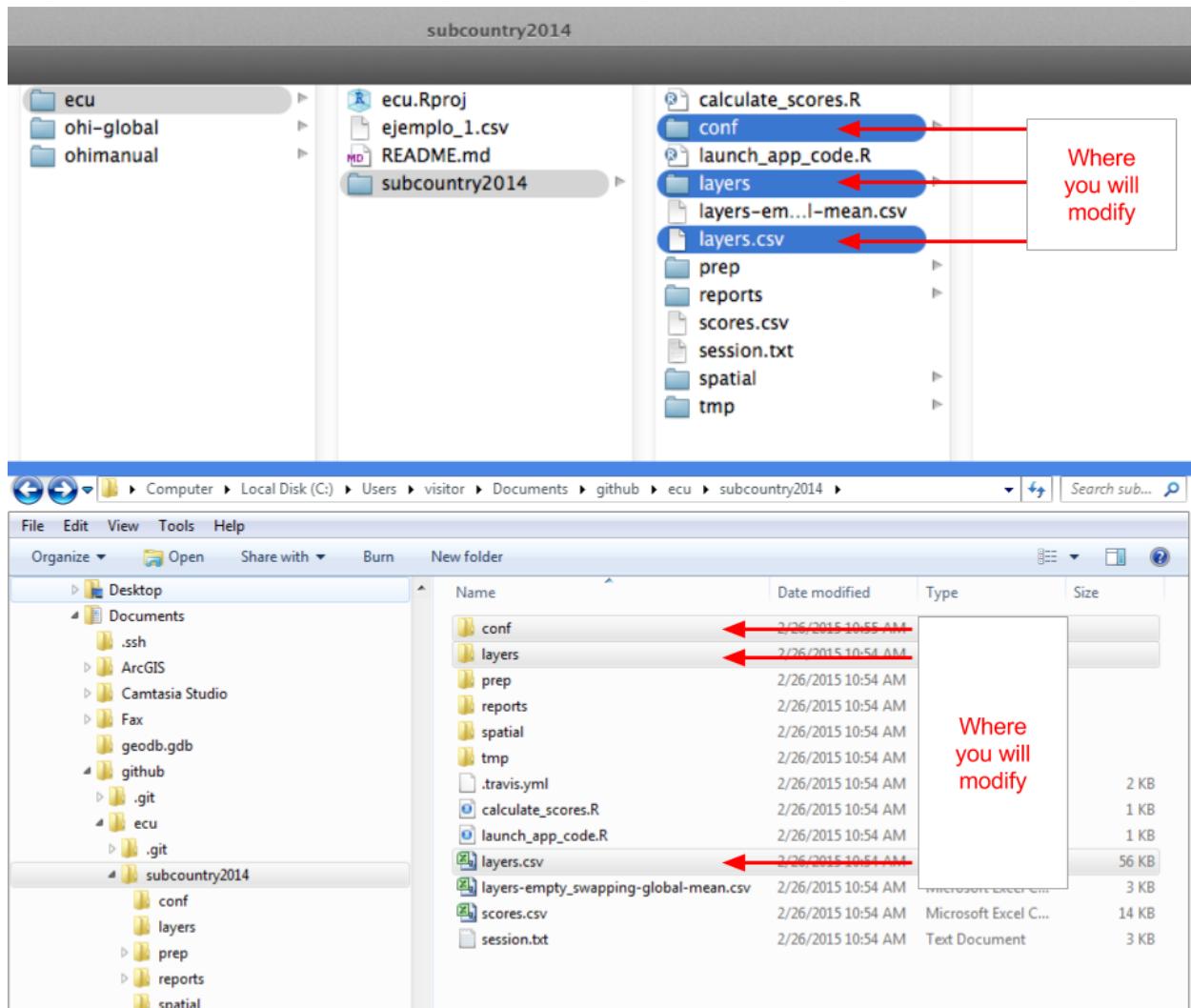


Figure 37: Files you will commonly modify are shown in the figure (Mac navigation is shown above and Windows is shown below). These include the `conf` folder, `layers` folder, and `layers.csv`.

6.0.1 File Preparation Workflow

It is generally recommended that you construct a useful workflow with your team when updating data layers for the Toolbox. This process is one that can be done by one person, or by several who are working through GitHub to sync the work. There are overall two main steps to preparing to input your layers into the Toolbox, starting with data layer preparation, and then going into data layer registration. The first step involves placing files into `layers` folder, and the second is registering those files in `layers.csv`. This file preparation process can occur in tandem with the model modification process. However, it must occur in order for you to run modified goal code using your new input data.

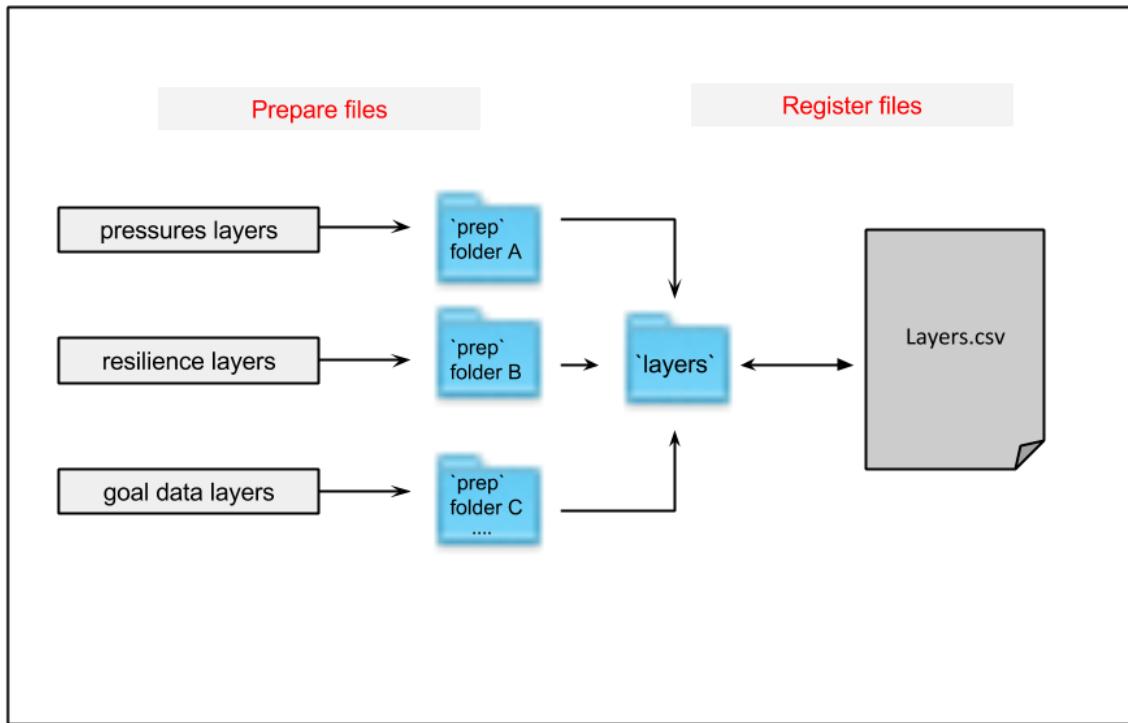


Figure 38: Diagram of OHI Toolbox data preparation workflow. You should start by prepping the files, loading them into the `layers` folder when they're ready for the Toolbox, and then registering them in `layers.csv`

6.0.1.1 Overview of the Process The following sections will describe the files included in the Toolbox. You will learn how to start preparing data layers for your assessment and how to start changing goal models. You will learn how the files interact to produce the calculated scores, starting from registering your layers to receiving the *scores* spreadsheet.

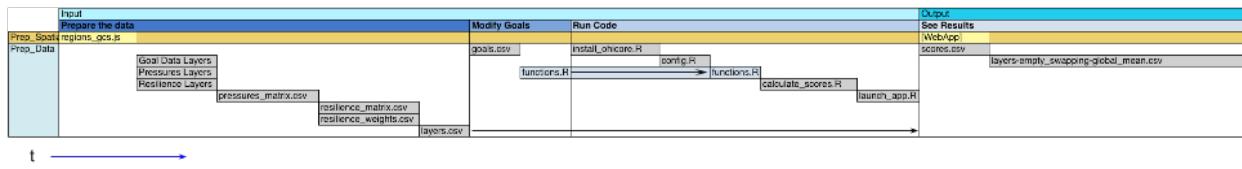


Figure 39: Recommended steps in which to engage with files in the OHI Toolbox.

6.1 Modifying and creating data layers

Data layers are `.csv` files and are located in the `[assessment]/subcountry2014/layers` folder. 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.

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`)

6.1.1 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.

6.1.2 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.

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

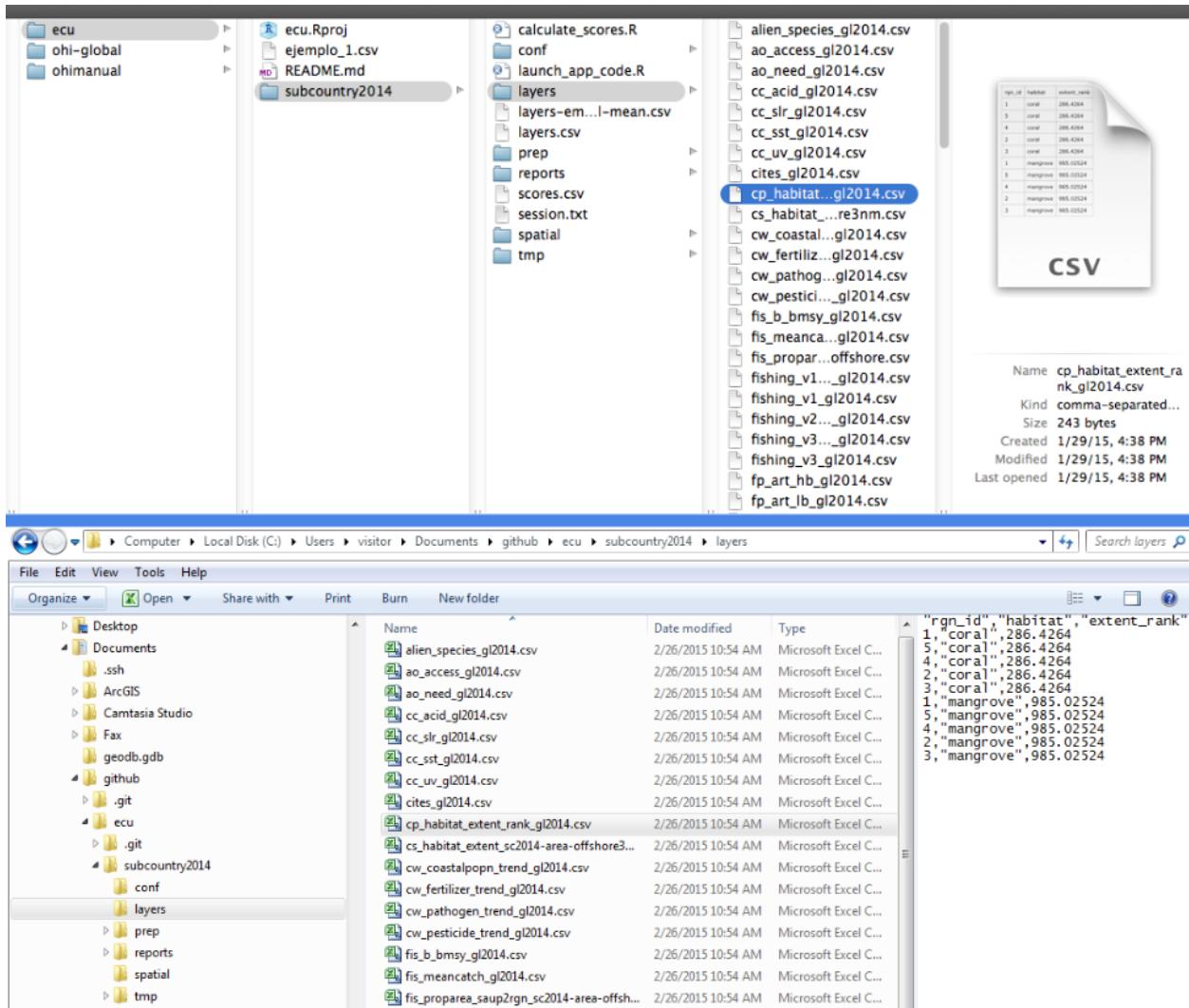
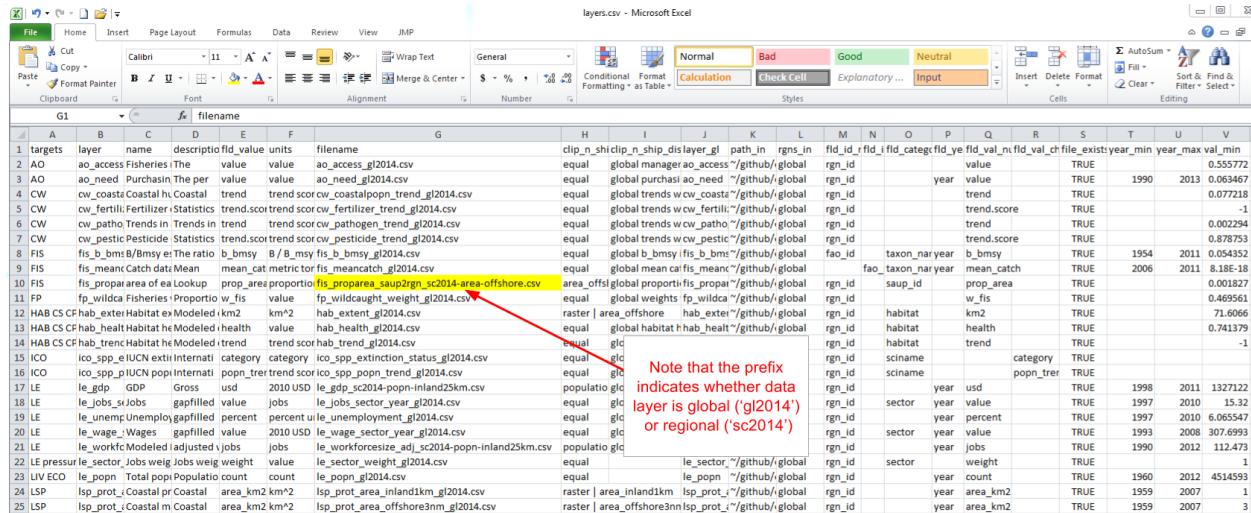


Figure 40: This figure shows the location of your data layers. Mac navigation is shown above and Windows is shown below.

6.1.3 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:



1	targets	layer	name	description	fld_value	units	filename	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	
2	AO	ao_access_Fisheries	The	value	value	ao_access	ao_access_gl2014.csv	equal	global	manager	ao_access	/github/global	rgn_id	value	TRUE	0.555772							
3	AO	ao_need_Purchasing	The per	value	value	ao_need	gl2014.csv	equal	global	purchasi	ao_need	/github/global	rgn_id	year	value	TRUE	1990	2013	0.063467				
4	CW	cw_coasta_Coastal	trend	trend	scor	cw_coastalpopn_trend	gl2014.csv	equal	global	trends	cw_coasta	/github/global	rgn_id	trend	TRUE	0.077218							
5	CW	cw_fertilizer_Statistics	trend	trend	scor	cw_fertilizer_trend	gl2014.csv	equal	global	trends	cw_fertiliz	/github/global	rgn_id	trend.score	TRUE	-1							
6	CW	cw_patho_Trends_in_Trends	trend	trend	scor	cw_pathogen_trend	gl2014.csv	equal	global	trends	cw_patho	/github/global	rgn_id	trend	TRUE	0.002294							
7	CW	cw_peptic_Pesticide_Statistics	trend	trend	scor	cw_peptic_trend	gl2014.csv	equal	global	trends	cw_peptic	/github/global	rgn_id	trend.score	TRUE	0.878753							
8	FIS	fs_b_bmsy_B/Bmsy	ratio	b_bmsy	B / B_msy	fs_b_bmsy	gl2014.csv	equal	global	b_bmsy	fs_b_bmsy	/github/global	fao_id	taxon_nar_year	b_bmsy	TRUE	1954	2011	0.054352				
9	FIS	fs_meanCatCatch_MeanCat	MeanCat	mean_cat	metric	fs_meanCat	meancatch	gl2014.csv	equal	global	meanCat	fs_meanCat	/github/global	fao_id	taxon_nar_year	mean_cat	TRUE	2006	2011	8.18E-18			
10	FIS	fs_meanCatArea_aLookup	prop_area	prop_area	prop_area	prop_area	saup2014-area-offshore.csv	equal	global	propArea	prop_area	/github/global	rgn_id	saup_id	prop_area	TRUE	0.001827						
11	FP	fp_wildca_Fisheries	Proprio	w_fis	value	fp_wildca	fp_wildca_weight_gl2014.csv	equal	global	weights	fp_wildca	/github/global	rgn_id	w_fis	TRUE	0.469561							
12	HAB CS CF hab_exten	Habitat_Ex	Modelled	km2	km2	hab_extent	gl2014.csv	equal	raster	area_offshore	hab_extent	/github/global	rgn_id	habitat	km2	TRUE	71.6066						
13	HAB CS CF hab_healt	Habitat_Health	Modelled	health	value	hab_health	gl2014.csv	equal	global	habitat	hab_healt	/github/global	rgn_id	habitat	health	TRUE	0.741379						
14	HAB CS CF hab_trend	Habitat_Trend	trend	trend	scor	hab_trend	gl2014.csv	equal	glc	trend	hab_trend	/github/global	rgn_id	habitat	trend	TRUE	-1						
15	ICO	ico_spp_e_IUCN_extInInternat	category	category	category	ico_spp_extinction_status	gl2014.csv	equal	glc	category	category	/github/global	rgn_id	sciname	category	TRUE							
16	ICO	ico_spp_p_IUCN_poplnternat	popn_trend	trend	scor	ico_spp_poppn_trend	gl2014.csv	equal	glc	popn_trend	popn_trend	/github/global	rgn_id	sciname	category	TRUE							
17	LE	le_gdp_GDP	Gross	usd	2010	USD	le_gdp_sc2014-popn-inland25km.csv	populatio	populatio	populatio	populatio	/github/global	rgn_id	year	usd	TRUE	1998	2011	1327122				
18	LE	le_jobs_sJobs_gapfilled	gapfilled	value	jobs	le_jobs_sector_year	gl2014.csv	equal	glc	sector	year	value	rgn_id	year	value	TRUE	1997	2010	15.32				
19	LE	le_unemp_Unemplgapfilled	gapfilled	percent	percent	unemployment	gl2014.csv	equal	glc	year	percent	value	rgn_id	year	percent	TRUE	1997	2010	6.065547				
20	LE	le_wage_Wages_gapfilled	gapfilled	value	2010	USD	le_wage_sector_year	gl2014.csv	equal	glc	year	value	rgn_id	year	value	TRUE	1993	2006	307.6993				
21	LE	le_workfctModelled	adjusted	jobs	jobs	le_workforce_size	adj_sc2014-popn-inland25km.csv	populatio	populatio	populatio	populatio	populatio	/github/global	rgn_id	year	jobs	TRUE	1990	2012	112.473			
22	LE	le_pressure_seCTOR_Jobs_welg	welg	weight	value	le_sector_weight	gl2014.csv	equal	le_sector	le_sector	le_sector	/github/global	rgn_id	sector	weight	TRUE	1						
23	LIV ECO	le_popn_Total popn	Popula	count	count	le_popn	gl2014.csv	equal	le_popn	le_popn	le_popn	/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_l	gl2014.csv	equal	raster	area_inland1km	lsp_prot_i	/github/global	rgn_id	year	area_km2	TRUE	1959	2007	1				
25	LSP	lsp_prot_Coastal_mc_Coastal	area_km2	km2	lsp_prot_area_offshore3nm	lsp_prot_o	gl2014.csv	equal	raster	area_offshore3nm	lsp_prot_o	/github/global	rgn_id	year	area_km2	TRUE	1959	2007	3				

Figure 41: 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); other columns are generated later by the Toolbox as it confirms data formatting and content:

- 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 used in R scripts like `functions.R` and `.csv` files like `pressures_matrix.csv` and `resilience_matrix.csv`.
- name:** Add a longer title for the data layer–this will be displayed 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 (which will be referenced in subsequent calculations).
- units:** Add a description about the *units* chosen in the *fld_value* column above.
- filename:** Add a filename for the new data layer that matches the name of the `.csv` file that was created previously in the `layers` folder.
- fld_id_num:** Area designation that applies to the newly created data layer, such as: `rgn_id` and `fao_id`.

TIP: Think about what units you would like to be displayed on the WebApp when filling out “units.”

6.1.4 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.

6.2 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. 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.

6.2.1 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.

6.2.2 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 42:

6.2.3 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.

6.2.3.1 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.

6.2.3.2 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_susviral	hd_susviral	hd_susviral	hd_intertidal
2	FIS			3	2	1			1		3		2	1
3	AD			3	2	2			1		1		3	1
4	AO						1							
5	NP					1	2	1			2			
6	NP					1	2	2			1			
7	NP					1	2	2			1			
8						1	2	2			2			
9						1	2	2			2			
10						1	2	2			2			
11						1	2	2			3			
12						1	2	2			2			
13						1	2	2			3			
14						1	2	2			2			
15						1	2	2			1			
16						1	2	2			2			
17						1	2	2			3			
18						1	2	2			2			
19						1	2	2			3			
20						1	2	2			2			
21						1	2	2			3			
22						1	2	2			2			
23						1	2	2			3			
24						1	2	2			2			
25						1	2	2			3			
26						1	2	2			2			
27						1	2	2			3			
28						1	2	2			2			
29						1	2	2			3			
30						1	2	2			2			
31						1	2	2			3			
32						1	2	2			2			
33						1	2	2			3			
34						1	2	2			2			
35						1	2	2			3			
36						1	2	2			2			
37						1	2	2			3			
38						1	2	2			2			
39						1	2	2			3			
40						1	2	2			2			
41						1	2	2			3			
42						1	2	2			2			
43						1	2	2			3			
44						1	2	2			2			
45						1	2	2			3			

Figure 43:

6.2.4 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.

6.3 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	habitat	habitat	habitat_li_gci	li_sector_even	mariculture	msi_gov	species_diver	species_diver_3nm	water	wgi_all
HAB	coral only	alien_species							habitat	habitat	habitat	habitat_li_gci	li_sector_even	mariculture	msi_gov	species_diver	tourism	water	wgi_all
HAB		alien_species							habitat	habitat	habitat	habitat_li_gci	li_sector_even	mariculture	msi_gov	species_diver	tourism	water	wgi_all
HAB	soft_bottom	alien_species							habitat	habitat	habitat	habitat_li_gci	li_sector_even	mariculture	msi_gov	species_diver	tourism	water	wgi_all
HAB	soft_bottom	alien_species							habitat	habitat	habitat	habitat_li_gci	li_sector_even	mariculture	msi_gov	species_diver	tourism	water	wgi_all
SPP									habitat	habitat	habitat	habitat_li_gci	li_sector_even	mariculture	msi_gov	species_diver	tourism	water	wgi_all
CS									habitat	habitat	habitat	habitat_li_gci	li_sector_even	mariculture	msi_gov	species_diver	tourism	water	wgi_all
CW									habitat	habitat	habitat	habitat_li_gci	li_sector_even	mariculture	msi_gov	species_diver	tourism	water	wgi_all
FIS									habitat	habitat	habitat	habitat_li_gci	li_sector_even	mariculture	msi_gov	species_diver	tourism	water	wgi_all
MAR									habitat	habitat	habitat	habitat_li_gci	li_sector_even	mariculture	msi_gov	species_diver	tourism	water	wgi_all
ECO									habitat	habitat	habitat	habitat_li_gci	li_sector_even	mariculture	msi_gov	species_diver	tourism	water	wgi_all
LIV									habitat	habitat	habitat	habitat_li_gci	li_sector_even	mariculture	msi_gov	species_diver	tourism	water	wgi_all
NP									habitat	habitat	habitat	habitat_li_gci	li_sector_even	mariculture	msi_gov	species_diver	tourism	water	wgi_all
CP									habitat	habitat	habitat	habitat_li_gci	li_sector_even	mariculture	msi_gov	species_diver	tourism	water	wgi_all
ICO									habitat	habitat	habitat	habitat_li_gci	li_sector_even	mariculture	msi_gov	species_diver	tourism	water	wgi_all
LSP									habitat	habitat	habitat	habitat_li_gci	li_sector_even	mariculture	msi_gov	species_diver	tourism	water	wgi_all
TR									habitat	habitat	habitat	habitat_li_gci	li_sector_even	mariculture	msi_gov	species_diver	tourism	water	wgi_all

Figure 44:

6.3.1 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)

6.3.1.1 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_combi, 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`

6.3.1.2 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.

6.4 Modifying goal models

When an existing layer is updated with new data, the Toolbox will automatically incorporate it into the goal calculations after the updated filenames are registered in `layers.csv`. However, if a new layer has been added to the layers folder and registered in `layers.csv`, the Toolbox will not use it unless it is called in a goal model. To integrate any new data layers registered in `layers.csv` you will need to modify the goal model to incorporate the data. Furthermore, in many cases, it will make sense to modify goal models based on data availability and/or local context. For example, the models for regional analyses can often be simplified because of improved data.

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

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

6.4.1 Update `functions.R`

To incorporate a new data layer into a goal model, open `functions.R` in RStudio: this script contains all the models for each goal and sub-goal. A member of your team with the ability to write R code will need to translate the updated goal model into the Toolbox format. Follow the structure of existing goal models in order to incorporate the new data layers, noting the use of certain R packages for data manipulation.

The image below shows the navigation pane in RStudio that can be used to easily navigate between goal models.

6.4.2 Check and possibly update `goals.csv`

`goals.csv` provides input information for `functions.R`, particularly about goal weighting and function calls. It also includes descriptions about goals and sub-goals, which is presented on the WebApp.

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. `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).

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

- check the years
- etc...

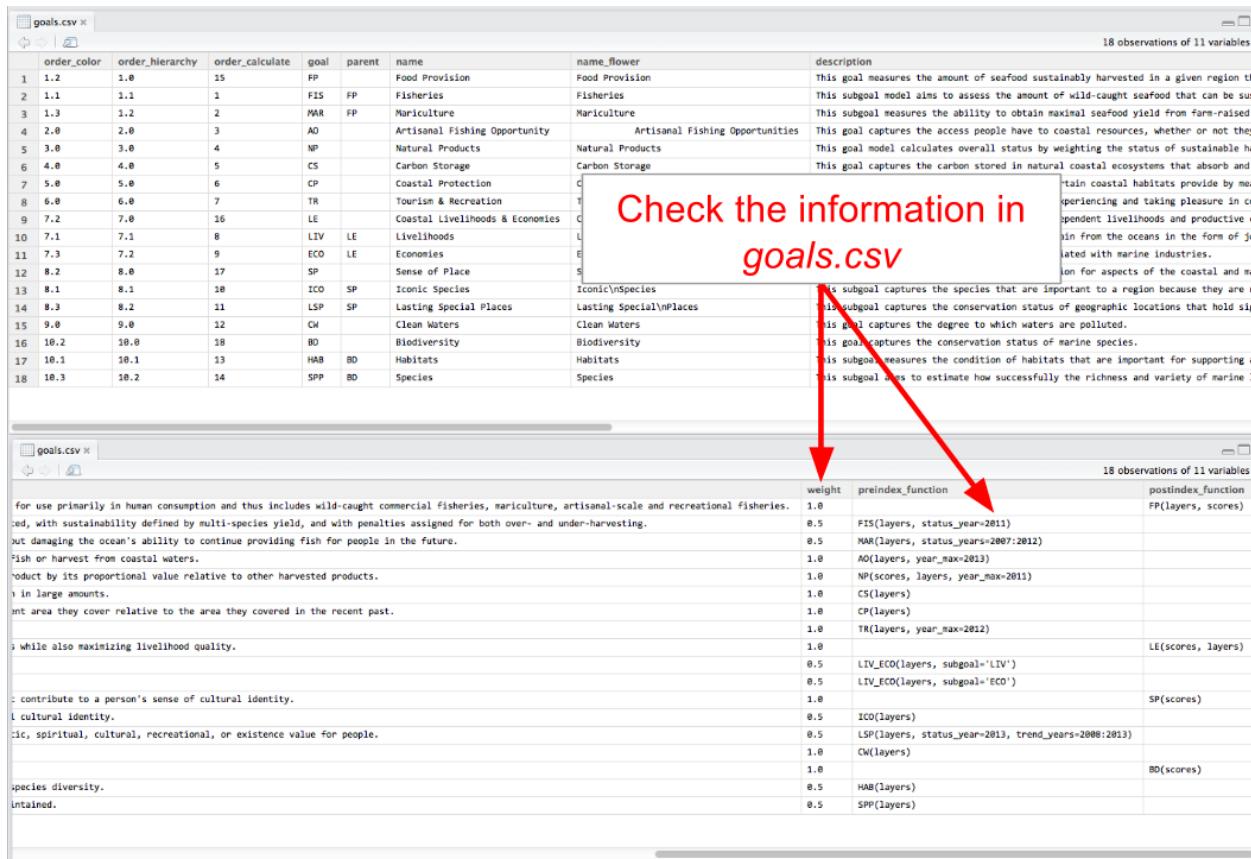
```
61
62 # separate out the region ids:
63 c$fao_id      <- as.numeric(sapply(as.character(c$fao_saup_id), "_"), function(x)x[1]))
64 c$saup_id     <- as.numeric(sapply(as.character(c$fao_saup_id), "_"), function(x)x[2]))
65 c$TaxonName   <- sapply(strsplit(as.character(c$taxon_name_key), "_"), function(x)x[1]))
66 c$TaxonKey    <- as.numeric(sapply(strsplit(as.character(c$taxon_name_key), "_"), function(x)x[2]))
67 c$catch       <- as.numeric(c$catch)
68 c$year        <- as.numeric(as.character(c$year))
69 #Create Identifier for linking assessed stocks with country-level catches
70 c$stock_id <- paste(as.character(c$TaxonName),
71                      as.character(c$fao_id), sep="_")
72
73 # b_bmsy data
74 b = SelectLayersData(layers, layer='fis_b_bmsy', narrow=T) %>%
75   select(
76     fao_id      = id_num,
77     TaxonName   = category,
78     year,
79     bmsy        = val_num)
80 # Identifier taxa/fao region:
81 b$stock_id <- paste(b$TaxonName, b$fao_id, sep="_")
82 b$bmsy      <- as.numeric(b$bmsy)
83 NP          <- as.numeric(as.character(b$fao_id))
84 CS          <- as.numeric(as.character(b$year))
85 CP          <-
86 TR          <-
87 LIV_ECO    <-
88 r_saup to rgn conversion
89 a[['fis_proparea_saup2rgn']] %>%
90   id, rgn_id, prop_area)
91   as.numeric(a$prop_area)
92   as.numeric(as.character(a$saup_id))
93   as.numeric(as.character(a$rgn_id))
94 -----
95 # the species status data with catch data
96 Catches: only taxa with catch status data
97
```

Modify *functions.R* when changing goal or sub-goal models

Shortcut to all goal sections

Figure 45: The navigation pane in RStudio can be used to easily navigate between goal models.

Check the information in
goals.csv



goals.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	FP	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	FP	Mariculture	Mariculture	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 me
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	dependent livelihoods and productive c
10	7.1	7.1	8	LIV	LE	Livelihoods	L	ain from the oceans in the form of jo
11	7.3	7.2	9	ECO	LE	Economies	E	ated with marine industries.
12	8.2	8.0	17	SP		Sense of Place	S	ion for aspects of the coastal and ma
13	8.1	8.1	18	ICO	SP	Iconic Species	I	is subgoal captures the species that are important to a region because they are r
14	8.3	8.2	11	LSP	SP	Lasting Special Places	L	is subgoal captures the conservation status of geographic locations that hold sig
15	9.0	9.0	12	CW		Clean Waters	C	nificance.
16	10.2	10.0	18	BD		Biodiversity	B	This goal captures the degree to which waters are polluted.
17	10.1	10.1	13	HAB	BD	Habitats	H	This goal captures the conservation status of marine species.
18	10.3	10.2	14	SPP	BD	Species	S	This subgoal measures the condition of habitats that are important for supporting a
								This subgoal tries to estimate how successfully the richness and variety of marine l

goals.csv

	weight	preindex_function	postindex_function
for use primarily in human consumption and thus includes wild-caught commercial fisheries, mariculture, artisanal-scale and recreational fisheries. sed, with sustainability defined by multi-species yield, and with penalties assigned for both over- and under-harvesting. but damaging the ocean's ability to continue providing fish for people in the future. fish or harvest from coastal waters. product by its proportional value relative to other harvested products. in large amounts. nt area they cover relative to the area they covered in the recent past. while also maximizing livelihood quality.	1.0	FIS(layers, status_year>2011)	
: contribute to a person's sense of cultural identity. l cultural identity. sic, spiritual, cultural, recreational, or existence value for people. species diversity.	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)	
			LE(scores, layers)
	1.0	LIV_ECO(layers, subgoal='LIV')	
	0.5	LIV_ECO(layers, subgoal='ECO')	
	1.0	SP(scores)	
	0.5	ICO(layers)	
	0.5	LSP(layers, status_year=2013, trend_years=2008:2013)	
	1.0	CW(layers)	
	1.0	BD(scores)	
	0.5	HAB(layers)	
	0.5	SPP(layers)	

Figure 46: Check the information in `goals.csv`. It provides input information for `functions.R`.

6.4.3 Example modification:

Suppose your team has decided to add an ‘artisanal access’ component to the Artisanal Fishing Opportunity goal because of locally available data. Once the data are obtained and properly formatted, the data layer is saved as `ao_access_art`. To include this new information in the goal model, you will need to do the following:

1. register the layer in `layers.csv`
2. update the goal model in `functions.R`
3. update the goal call in `goals.csv`

1. register in `layers.csv`

	A	B	C	D	E	F	G	H
1	targets	layer	name	description	fid_value	units	filename	fid_id_num
2	AO	ao_access	Fisheries management	The opportunity fit value	value		ao_access_china2014.csv	rgn_id
3	AO	ao_access_art	Example data	Made-up data	value	value	ao_access_art_china2014.csv	rgn_id
4	AO	ao_need	Purchasing power p	The per capita pu value	value		ao_need_global2013.csv	rgn_id
5	CW	cw_coastalpopn_trend	Coastal human popul.	Coastal population trend	trend score		cw_coastalpopn_trend_global2013.csv	rgn_id
6	CW	cw_fertiliser_trend	Earthline environmental	Chlorophyll a Fertiliser trends	trend score		cw_fertiliser_trend_global2013.csv	rgn_id

Figure 47:

2. update goal model

```

325 # cost
326 layers_data = SelectLayersData(layers, targets='AO')
327 ry = rename(dcast(layers_data, id_num ~ year - layer, value.var='val_num',
328   subset = .(layer %in% c('ao.access'))),
329   c('id_num'~'region_id', 'ao_need'~'need'));
330 Sustainability=1.0);
331
332 # cost data
333 layers_data = SelectLayersData(layers, targets='AO')
334 ry = rename(dcast(layers_data, id_num ~ year - layer, value.var='val_num',
335   subset = .(layer %in% c('ao.access'))),
336   c('id_num'~'region_id', 'ao_need'~'need'));
337 head(ry); summary(ry)
338 r = na.omit(rename(dcast(layers_data, id_num ~ layer, value.var='val_num',
339   subset = .(layer %in% c('ao.access'))),
340   c('id_num'~'region_id', 'ao.access'~'access')));
341 head(r); summary(r)
342 r0 = na.omit(rename(dcast(layers_data, id_num ~ layer, value.var='val_num',
343   subset = .(layer %in% c('ao.access_art'))),
344   c('id_num'~'region_id', 'ao.access_art'~'access_art')));
345 head(r0); summary(r0)
346
347 ry = merge(ry, r0)
348 ry = merge(ry, r0); head(ry); dim(ry)
349
350 # model
351 ry = within(ry,{
352   Du = (1.0 - need) * (1.0 - (access + access_art)/2 )
353   status = ((1.0 - Du) * Sustainability) * 100
354 })
355
356 # status
357 r.status = subset(ry, year==year_max, c(region_id, status)); summary(r.status); dim(r.status)
358
359 # trend

```

Figure 48:

3. update goal call in `goals.csv`

[develop]

6.4.4 What’s the code trying to do?

6.4.4.1 Natural Products You may have already looked at the **NP** section of `functions.R`. In simple terms, here is what the code is doing:

- It pulls out the appropriate data layers to find out the amount of each product per unit area. It does gap-filling as necessary for the Global data.

It should be noted that in the Global Assessments, the harvested amounts are derived from the information from the Food and Agriculture Organization of the United Nations (FAO), and these are combined with habitat values used elsewhere in the assessment. You should be conscious of this as you go through the model and change it, because you may be able to simplify aspects of the code such as gap-filling.

- It calculates Exposure by finding how intensely each identified product is being harvested (amount of product per km²), and then transforms this from a scale from 0 to 1.

TIP: The data layer, `hab_extent` is used here and in other goal models in the default code for the Global Assessment.

- In parallel to this, it finds the Risk of each product based on a scoring system that becomes binary: 0, or 1.
- It then averages the two factors, Exposure and Risk, to reveal where risk and intensity are highest. This value is then inverted to become Sustainability and to reward lower intensity and lower risk.
- The amount of each kind of product, relative to the peak historical yield of that product across all regions assessed, and Sustainability are both used to create sustainability-weighted scores for all regions and all years available.
- The latest year value is used in the status, and the past few years' values are used in the trend to produce the final score.

Figure needed? Diagram of sustainability calculation, NP; or perhaps a time-series of the available products in country X in order to show where the peak yield would be.]

6.4.4.2 Data Sources If the case is that corals, sponges, and then you might be able to use FAO data, which is the data source of the Global Assessments. Otherwise, you will have to find comparable data in your area or consult local statistical offices and local fisheries managers to get harvest values similar to landing values and any other kinds of stock assessments. The IUCN offers quantified assessments of risk to species, but that is more appropriate for biodiversity; CITES signatory data may be more appropriate for the trade products. Exposure can be calculated spatially, and for this you should be able to find or produce your own maps if possible. Your maps might have finer resolution than those in global resolution.

6.4.4.3 Gap-filling

TIP: When checking your data, check cases where country-product pair has 0 for sustainability score, but relatively high harvest ratio (curr harvest/peak harvest) – it may be a flag that the sustainability score is off (eg because the habitat area is off)

TIP: Explore simplifying gap-filling: use correlation model of dollar value vs. harvested tonnage, while discarding the part of script using dollar ratio (current dollar value)/(peak dollar value) as a gap-filler for harvest ratio.

TIP: Switch the gap-filling order: using the dollar value correlation model first (in cases where the most recent year has no harvest reported but has dollar value reported, that's a better estimate than using the harvest from the previous year), then gap-fill any remaining cases of missing harvest for the current year with harvest from the previous year

6.4.5 Appendix - source materials

6.4.5.1 Global Data Approach (Technical Notes) Data Overview

product	relative tonnes (1)	weighting (2)	Exposure (3)	Risk (4)
coral	FAO	FAO	coral habitat	all 1
sponges	FAO	FAO	coral + rocky reef habitat	all 0
ornamentals	FAO	FAO	coral + rocky reef habitat	1 if blast/cyanide fishing, otherwise 0
fish oil	FAO	FAO	fish score/100	-
shells	FAO	FAO	coral + rocky reef habitat	all 0
seaweeds	FAO	FAO	rocky reef habitat	-

- (1) relative tonnes: tonnes relative to max tonnes for region with 35% buffer. The maximum corresponds to the year with the highest \$ value - but it would probably be better to just base this off tonnes. When we redo these data lets evaluate this approach.
- (2) weighting: This weights the contribution of each product according to USD at max year for a region. It makes sense to use \$, because comparing extraction weight of sponges vs. ornamentals doesn't make sense.
- (3) Exposure: For fish oil this value is the FIS score (which is a bit different than what is described in the paper because FIS score can have penalties for underfishing). The other values are determined by: $\log(\text{harvest}/\text{habitat area} + 1) / \log[\max(\text{harvest}/\text{habitat area}) + 1]$.

The habitat area used for seaweeds: rocky reef The habitat area used for coral: coral The habitat area used for shells, ornamentals, sponges: coral plus rocky reef

6.4.5.2 Notes: Preparing the Data

6.5

6.5.0.3 Notes: Tech Specs

<!--Note Ecuador's approach to Natural Products:

Very low data for species, so it has been hard to advance this goal, unknown for each global product. but they need to be looking for data for things they actually produce, like madera, leña de manglares, etc.

<!--SPP status takes into account the IUCN score and the area that the species occupies (although the ICO status is calculated by taking the average of the IUCN ratings for all the iconic species in the EEZ).

<!--SPECIES The original logic was to represent the species present relative to the proportion of their range.

A disadvantage is that rare species (and those with contracting ranges) will have a relatively small influence.

An advantage is (actually a disadvantage of averaging species scores at the EEZ level): "The bigger an area, the more influence it has."

<!--SPECIES. For the global assessment, we started with a list of iconic species, and then found which ones were present in each EEZ. Are you using IUCN categories? If so, you can use Butchart's method. You'll see in the Toolbox that there are four places where you will need to remove the reference to this goal. Failing to delete all referenced

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**

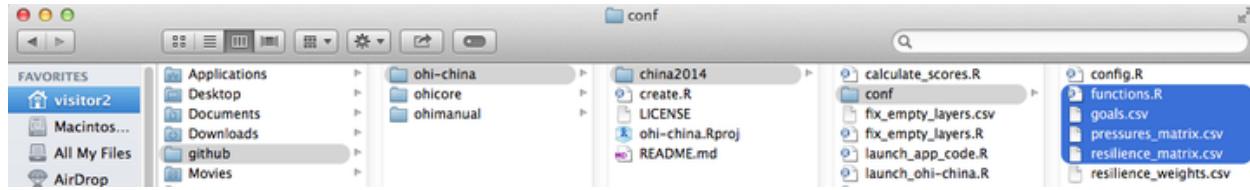


Figure 49:

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.

```

627 return(scores_AP)
628 }
629
630 CS = function(layers){
631
632 # layers
633 lyrss = llist("rk" = c("null_health", "health",
634 "null_exten", "sub_exten", "exten",
635 "null_trend", "trend"))
636 lyrss$names = sub("\\\\w\\\\w\\\\w\\\\.", "", names(unlist(lyrss)))
637
638 # cost data
639 D = SelectLayersData(layers, layers=lyrss$names,
640 rk = rename(D[,1:(dnum + category - layer, value.var="val_mn")], subset = .(layer %in% names(lyrss)[["rk"]])))
641
642 # filter to CS habitat
643 rk = subset(rk, habitat %in% c("mangrove", "saltmarsh", "seagrass"))
644
645 # sign switch if d < 0
646 rk$sign = ifelse(d < 0, -1, 1)
647 rk$exten = rk$exten * d
648
649 # status
650 r.status = ddply(no.onc(rk), c("region_id", "habitat", "extent", "health")), .(region_id), summarize,
651   goal = "CS",
652   dimension = "status",
653   score = min(1, sum(extent * health) / sum(extent)) * 100
654
655 # trend
656 r.trend = ddply(no.onc(rk), c("region_id", "habitat", "extent", "trend")), .(region_id), summarize,
657   goal = "CS",
658   dimension = "trend",
659   score = sum(extent * trend) / sum(extent)
660
661 # return scores
662 scores = cbind(rbind(r.status, r.trend))
663 return(scores)
664
665
666 CP = function(layers){
667
668 }
```

Figure 50:

- 2) Remove the CS row from **goals.csv**. Delete the highlighted row in the figure below that contains the CS goal.
- 3) Remove all CS rows from **pressures_matrix.csv**. Delete the highlighted rows in the figure below that contain CS pressures.
- 4) Remove all CS rows from **resilience_matrix.csv**. Delete the highlighted rows in the figure below that contain CS resilience.

| | A | B | C | D | E | F | G | H | I | J | K |
|----|-------------|----------------|---------------------|--------|------------------|----------------|------------------|--------|-----------------|---|---|
| 1 | order_color | order_hierarch | order_calculat_goal | parent | name | name_flower | description | weight | preindex_func | | |
| 2 | 1.2 | 1 | 15 FP | | Food Provision | Food Provision | This goal meas: | 1 | | | |
| 3 | 1.1 | 1.1 | 1 FIS | FP | Fisheries | Fisheries | This subgoal n | 0.5 | FIS(layers, sta | | |
| 4 | 1.3 | 1.2 | 2 MAR | FP | Marculture | Marculture | This subgoal n | 0.5 | MAR(layers, st | | |
| 5 | 2 | 2 | 3 AO | | Artisanal Oceani | Ocean | This goal capi: | 1 | AO(layers, yea | | |
| 6 | 3 | 3 | 4 NP | | Natural Produt | Natural Produt | This goal mod: | 1 | NP(layers, syst | | |
| 7 | 4 | 4 | 5 CS | | Carbon Storag | Carbon Storage | This goal capti: | 1 | CS(layers) | | |
| 8 | 5 | 5 | 6 CP | | Coastal Protec | Coastal\Prot | This goal meas: | 1 | CP(layers) | | |
| 9 | 6 | 6 | 7 TR | | Tourism & Rec | Tourism & VNR | This goal capti: | 1 | TR(layers, yea | | |
| 10 | 7.2 | 7 | 16 LE | | Coastal Liveli | Coastal Liveli | This goal aims: | 1 | | | |

Figure 51:

| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|----|------|-----------|--------------|-------------|--------------|--------------|--------------|--------------|--------------|----------|--------------|-----------------|---|
| 1 | goal | component | component_nt | po_desal_in | po_desal_out | po_chemicals | po_pathogens | po_nutrients | po_nutrients | po_trash | nd_auditable | sl_nd_auditable | h |
| 9 | NP | shells | | 1 | 2 | | | | 1 | | 2 | | |
| 10 | NP | sponges | | 1 | 2 | | | | 1 | | 3 | | |
| 11 | CS | medicinal | | 2 | 2 | | | | 1 | | 2 | | |
| 12 | CS | saltmarsh | | 2 | 2 | | | | 2 | | 3 | | |
| 13 | CS | seagrass | | 2 | 2 | | | | 2 | | | | |
| 14 | CP | coral | | 2 | 2 | | | 1 | | 2 | | | 3 |

Figure 52:

6.7 Modifying goals with categories

Several goals have categories that are assessed within them. In global assessments, these goals are:

| Goal | Category |
|------|------------------|
| NP | product types |
| CS | habitat types |
| CP | habitat types |
| HAB | habitat types |
| LIV | industry sectors |
| ECO | industry sectors |

These goal categories are identified in the file `config.r` located in the `conf` folder. The Toolbox uses `config.r` to identify the appropriate categories when pressures and resilience dimensions are calculated.

When you open `config.r`, you'll see that the `pressures_components` variable identifies which data layer indicates the category types. For global assessments, the following data layers are used:

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

If you have modified any of the categories for any goal, you will likely need to update the layer indicated in `config.r` in addition to updating the categories as unique rows in `pressures_matrix.csv`.

6.7.1 Example 1: Pressures

For example, there are three product types for Natural Products in the China assessment (seasalt, sea chemicals, and sea medicine), differ from those assessed in the global assessments (corals, fish_oil, ornamentals, seaweeds, sponges). To modify this goal in the Toolbox, input layers were updated, saved in the `layers` folder, and registered in `layers.csv`; the `NP` function in `functions.r` was updated, and new categories replaced the old in `pressures_matrix.csv`. The only remaining step to finalize the new NP goal is to update `config.r` to ensure it looks for the correct data layer and therefore categories. If you do not update `config.r` before running `calculate_scores.r`, you will get the following warning:

Calculating Pressures...
The following components for NP are not in the aggregation layer np_harvest_product_weight categories (corals, fish_oil, ornamentals, seaweeds, sponges):

6.7.2 Example 2: Pressures

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 CS layers and goal in the China 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 3: Resilience

For resilience, the proper categories also need to be identified both in `resilience_matrix.csv` and `config.r`. If neither are updated, 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 = lyrs),
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. NOT

```
Other example modifications

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_proarea_saup2rgn`
* `fp_wildcaught_weight`
```

#### 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
* *for genus/family/broader taxa*: the toolbox will use median B/Bmsy from species in that region + a pool of
* *data source (for CMSY)*: catch time-series (at least 10 years of catch >0), species resilience (if available)
```

\*\*Example data:\*\*

| fao_id | taxon_name     | year | b_bmsy   |
|--------|----------------|------|----------|
| 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 use multiple fao\_ids. 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

\*\*Example data:\*\*

| fao_saup_id | taxon_name_key             | year | mean_catch  |
|-------------|----------------------------|------|-------------|
| 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 'taxonomic key'.

`fis\_proarea\_saup2rgn`:

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

\*\*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 avoided  
 \*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:

`fp\_wildcaught\_weight`:

\*only needed if there is mariculture\*: for each region, this represents the relative proportion of catch (NOTE that, before all mariculture harvest from all species gets summed, the mariculture harvest for each

##### Running CMSY model

**\*\*Sample data to run CMSY:\*\***

| id   stock_id                    | res    | ct          | yr   |  |
|----------------------------------|--------|-------------|------|--|
| 6   Acanthistius brasiliensis_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_bmsyiQ25 | b_bmsyiQ75 | 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\*** is the

**\*\*How to:\*\***

**\*\*1. Include resilience in the CMSY code:\*\***

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

```
>
```

**\*\*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 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’).

**6.7.3.1 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](#)

## **6.8 Notes about R**

The Toolbox is written in R, and relies heavily on a few R packages created to facilitate data handling and manipulation. The primary R package used is called `dplyr` by Hadley Wickham. The `dplyr` package allows for ‘chaining’ between functions, which is represented with a `%>%`. See: <https://github.com/hadley/dplyr#dplyr> for documentation.

## **7 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).

## 7.1 Overall

### 7.2 Conceptual

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

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

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

#### **Q: 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.

## 7.3 Timing and Resources

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

A: Regional assessments can be completed at(varying costs depending on the local context.(Funds are needed for a management and scientific team, workshops and meetings (including travel), communications, policy

engagement, and operating costs. Therefore, securing funding is an important component to satisfactorily complete the assessment. We encourage the development of a local proposal or strategic action plan that details a timeline of activities and the resources needed to accomplish them.

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

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

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

A: 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

## 7.4 Structure

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

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

## 7.5 Reference points

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

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

**Q: What is sector evenness?**

A: Sector evenness (also called a diversity index) is an economic concept that is included in OHI to enable comparison across many different sectors included in the Livelihoods & Economies goal. This goal evaluates jobs, wages and revenues for nine marine employment sectors. The distribution of employment across these

nine sectors is an effective indicator of resilience. If total employment within a community is primarily based in one or two sectors, the overall economic system will be excessively vulnerable to downturns in those sectors. Conversely, if employment is spread relatively evenly throughout all nine sectors, the overall system will be more robust and resistant to such disturbances. Overall revenue within the community will remain more stable during such downturns, and workers displaced by a downturn in their sector may be able to find employment in another sector without leaving the community.

## 7.6 Appropriate data layers

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

A: Shipping and port activity are included as pressures only

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

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

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

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

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

A: Because they store carbon for less than 100 years, seaweeds and corals are not included in the carbon storage goal. While the pelagic oceanic carbon sink (phytoplankton) plays a large role in the sequestration of anthropogenic carbon, the pelagic ocean mechanisms are not amenable to local or regional management intervention. Phytoplankton contribute to carbon fixation when they die and sink to the sea bottom at sufficient depth, because it is effectively out of circulation. However, if those phytoplankton are eaten, the carbon is cycled back into the system and not sequestered. Something that could potentially be included in the carbon storage goal is mollusc shells, if they are added to a landfill and not recycled in the sea. So if information on mariculture production and waste disposal are available, this could be an interesting addition to carbon storage at a regional scale.

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

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

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

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

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

A: Desal would be incorporated into OHI in several places. The benefit is that there is freshwater produced, which could be incorporated into the Natural Products goal (or potentially into its own goal). Data required would be the volume of freshwater created based on the volume of seawater involved and spatial extent. Setting the reference point would not be based on how much can be produced, but some other targets perhaps set by government (percentage of the population served). Similar to the mariculture sub-goal and tourism

goals, any negative effects caused by desal that affect other goals (example: species) do not influence the ability to obtain desalination targets now and in the future. Therefore, the sustainability coefficient only measures the ability to sustain that goal, but not the impacts on other goals: instead, they are taken into account as pressures when calculating the other goals. Desal should be included as a pressure similar to cooling on-shore power plants since the discharge brine is dense, doesn't plume very well and there are chemicals involved.

**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?

## 7.7 Food Provision

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

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

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

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

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

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

\*\*Q: I have fish that are used as feed for other fish (e.g., sprat) in my country. Can I include them in this goal?

A: It would be more appropriate to include them in Natural Products rather than Food Provision. 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.

(Source: *OHI Baltic Workshop, February 2015*)

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

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

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

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

## **7.8 Livelihoods & Economies**

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

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

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

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

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

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

## **7.9 Tourism & Recreation**

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

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

## **7.10 Natural Products**

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

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

## **7.11 Species**

**Q: Can species and iconic species model scores be penalized if there are local flagship species that have not been evaluated?**

Global data are based on IUCN assessments. For these evaluations, IUCN chooses 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 IUCN 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 reports 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 unassessed 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.

It might be reasonable to penalize unassessed iconic species. It is a smaller list of species that are specifically identified as being of interest, for one reason or other. This would work for species that have some form of assessment - unless that information already exists, it might be unrealistic to try to produce the data layer required to develop a new model.

## 7.12 Sense of Place

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

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

**Q:** Should we calculate each category used in our assessment (e.g., antiquities, MPAs, beaches of special interest) independently, and then give the same weight (e.g., a third of the goal score) to the three categories, or should we instead pool the actual areas of the 3 categories?

A: 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.

(Source: OHI Israel assessment discussions, 2014-2015)

## 7.13 Pressures

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

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

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

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

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

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

# 8 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.

## 8.1 Error: RStudio won't push to GitHub

When pushing committed changes within RStudio, would return the error \* error: unable to read askpass response from 'rpostback-askpass' \* fatal: could not read Username for

RStudio: Review Changes

Git Push  
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:
  - `$ 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

```
GNU nano 2.0.6 File: /etc/paths
/usr/local/git/bin
/usr/local/bin
/usr/bin
/bin
/usr/sbin
/sbin

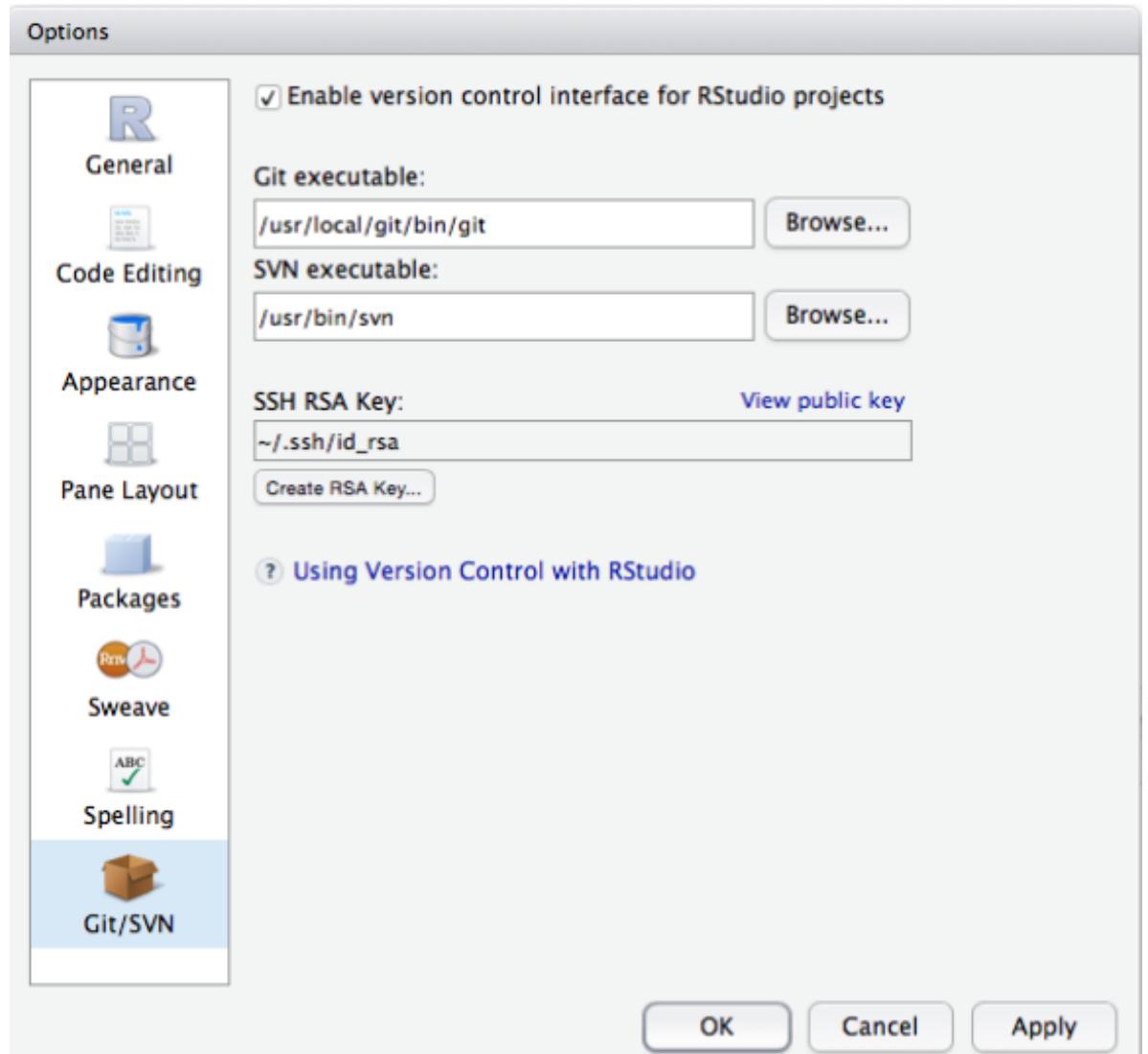
^G Get Help ^O WriteOut ^R Read File ^Y Prev Page ^K Cut Text
^X Exit ^J Justify ^W Where Is ^V Next Page ^U Uncut
```

- 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](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.

## 8.2 Loading RWorkspace on Restart

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

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

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

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

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

Figure 54:

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

### 8.3 Calculating Pressures...

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

Example:

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

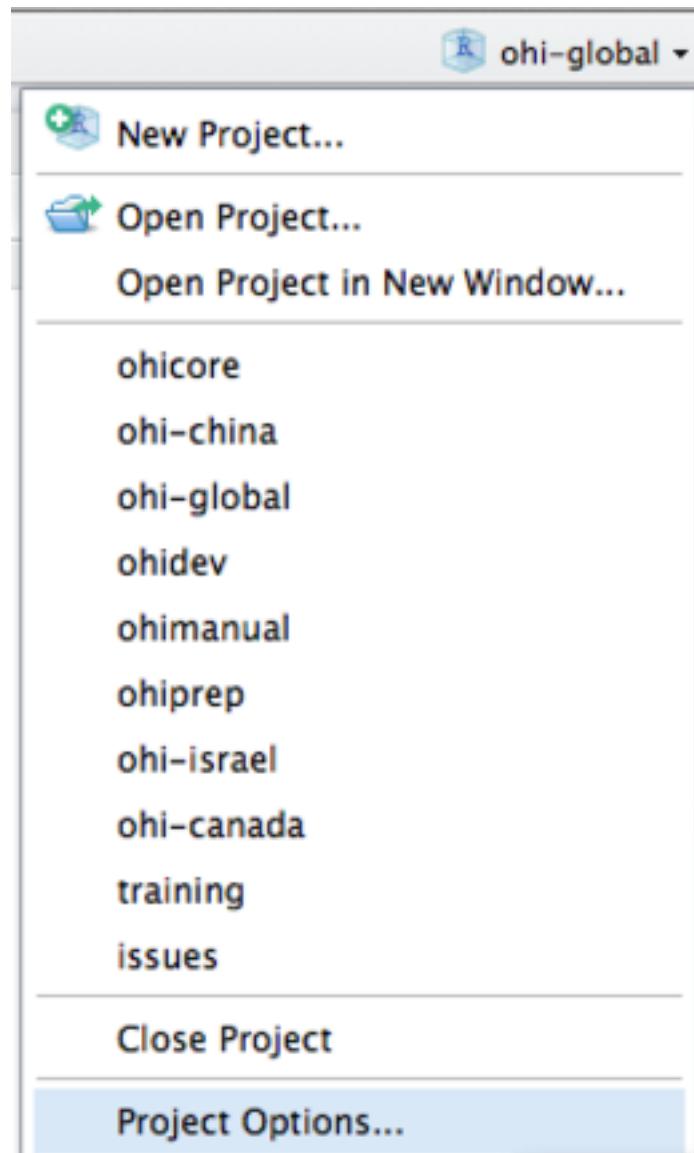


Figure 55:

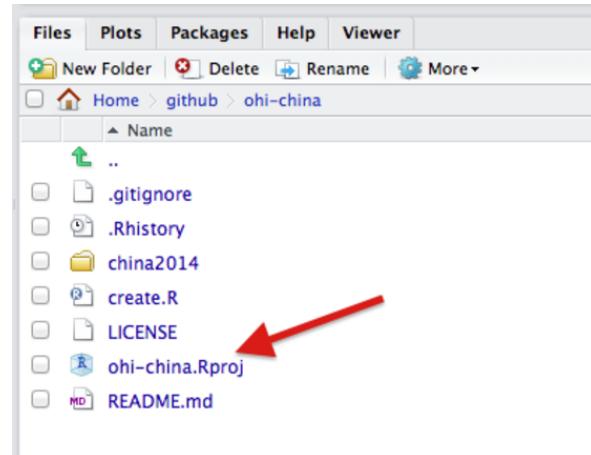


Figure 56:

### 8.3.2 ‘Error in matrix...’

The screenshot shows an Excel spreadsheet with the following data:

|    | 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            |

Example: >

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

## 8.4 Calculating Resilience ...

### 8.4.1 ‘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`.

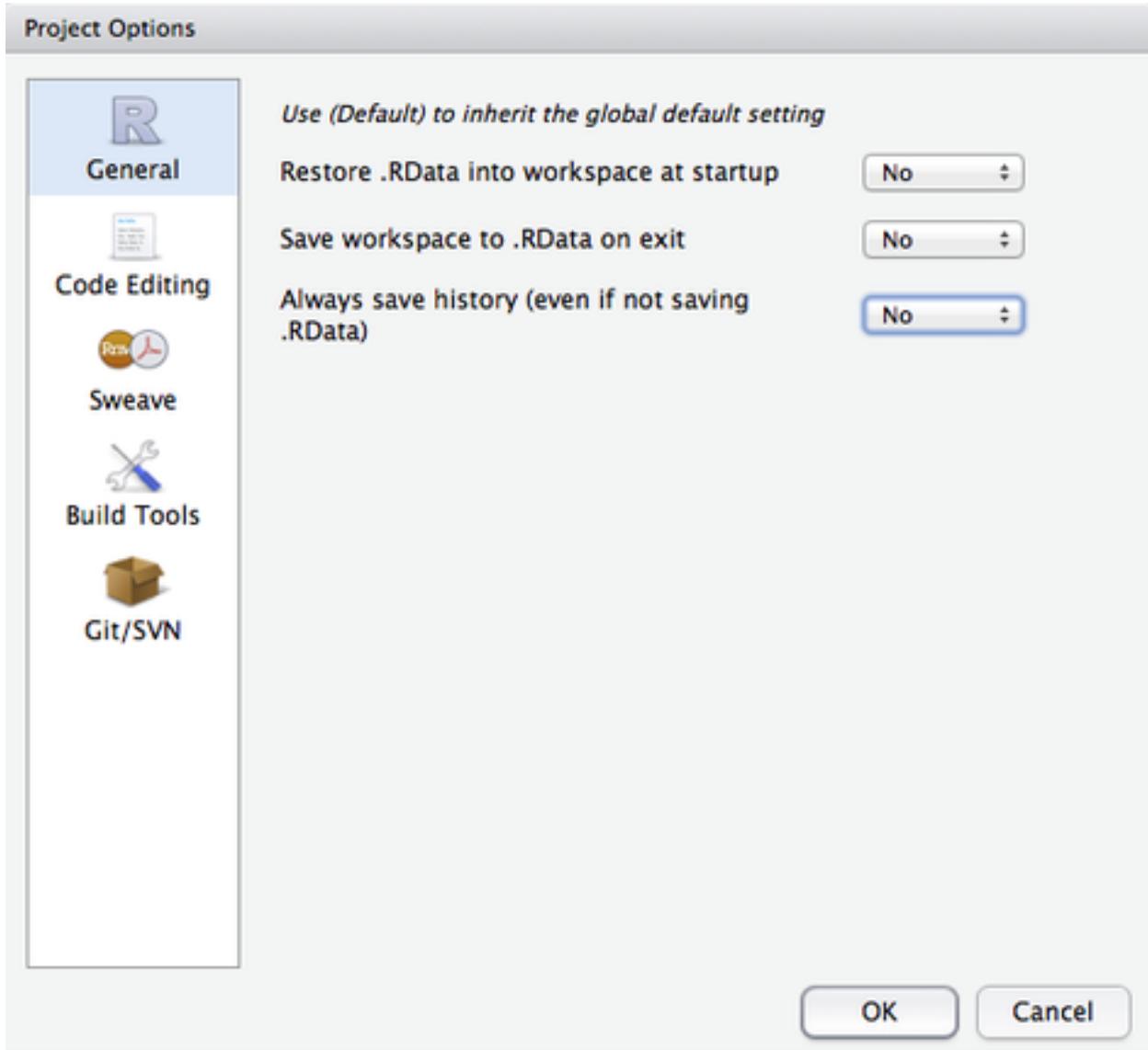


Figure 57:

```
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 58:

|    | 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 59:

```

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)
> |

```

Figure 60: