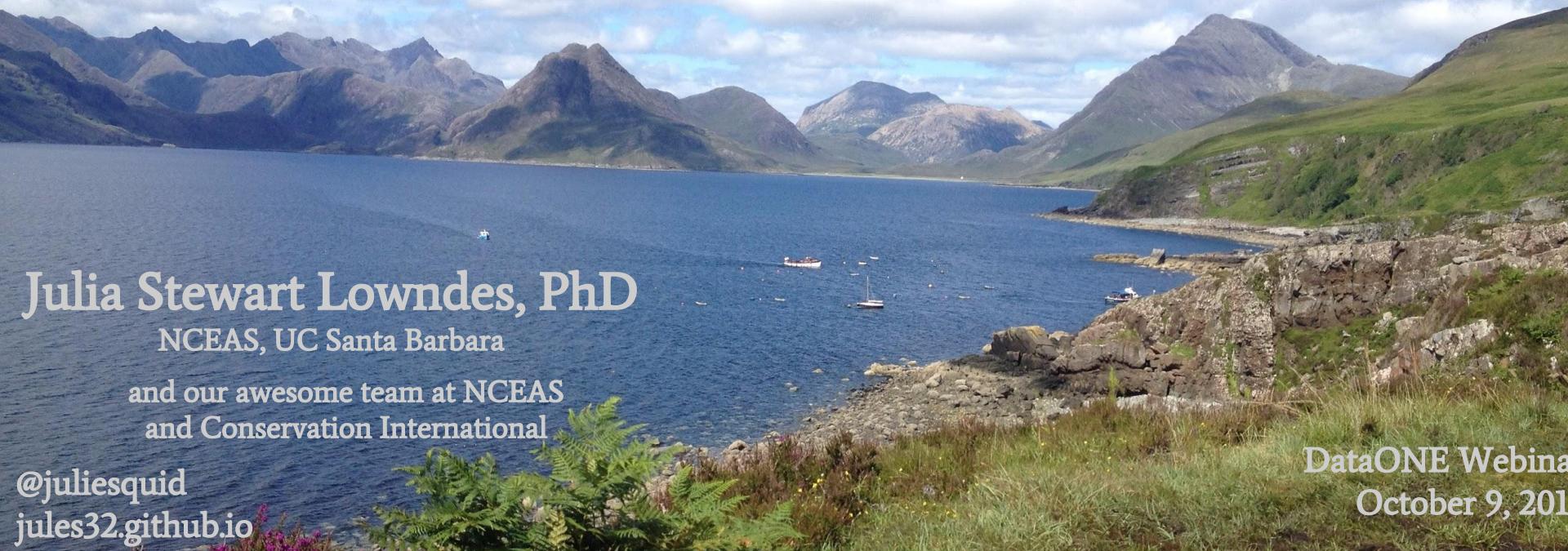




# Better science in less time

## How open data science tools have improved our science



**Julia Stewart Lowndes, PhD**

NCEAS, UC Santa Barbara

and our awesome team at NCEAS  
and Conservation International

@juliesquid  
jules32.github.io

DataONE Webina  
October 9, 201

# Our path to better science in less time using open data science tools

Julia S. Stewart Lowndes<sup>1\*</sup>, Benjamin D. Best<sup>2</sup>, Courtney Scarborough<sup>1</sup>, Jamie C. Afflerbach<sup>1</sup>,  
Melanie R. Frazier<sup>1</sup>, Casey C. O'Hara<sup>1</sup>, Ning Jiang<sup>1</sup> and Benjamin S. Halpern<sup>1,3,4</sup>

# Our path to better science in less time using open data science tools

Julia S. Stewart Lowndes<sup>1\*</sup>, Benjamin D. Best<sup>2</sup>, Courtney Scarborough<sup>1</sup>, Jamie C. Afflerbach<sup>1</sup>,  
Melanie R. Frazier<sup>1</sup>, Casey C. O'Hara<sup>1</sup>, Ning Jiang<sup>1</sup> and Benjamin S. Halpern<sup>1,3,4</sup>

## Better science is:

- Reproducible
- Collaborative
- Open

*Separate things, but  
deeply interwoven.  
Your most  
important  
collaborator is you!*

# Our path to better science in less time using open data science tools

Julia S. Stewart Lowndes<sup>1\*</sup>, Benjamin D. Best<sup>2</sup>, Courtney Scarborough<sup>1</sup>, Jamie C. Afflerbach<sup>1</sup>,  
Melanie R. Frazier<sup>1</sup>, Casey C. O'Hara<sup>1</sup>, Ning Jiang<sup>1</sup> and Benjamin S. Halpern<sup>1,3,4</sup>

Better science is:

- Reproducible
- Collaborative
- Open

Open data science tools:

- Align with data science practices
- Enable better science
- Exist! For your science! (meet scientists where we are)

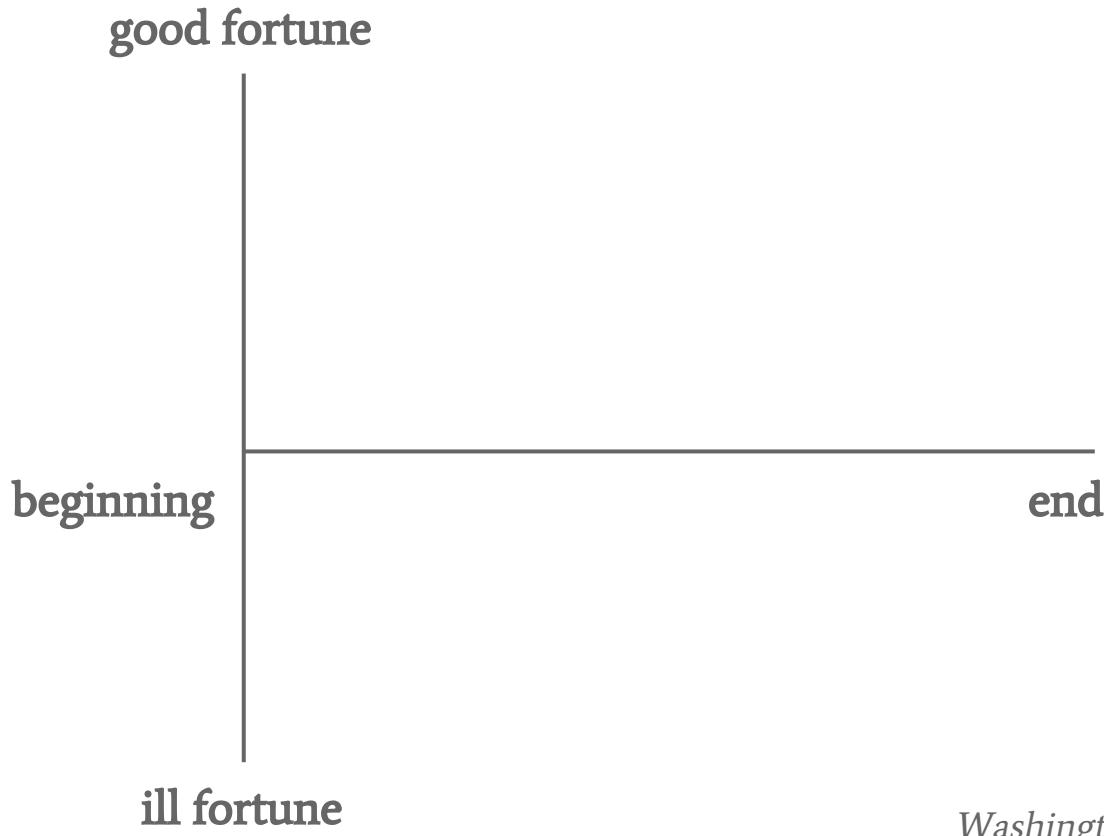
# Our path to better science in less time using open data science tools

Julia S. Stewart Lowndes<sup>1\*</sup>, Benjamin D. Best<sup>2</sup>, Courtney Scarborough<sup>1</sup>, Jamie C. Afflerbach<sup>1</sup>,  
Melanie R. Frazier<sup>1</sup>, Casey C. O'Hara<sup>1</sup>, Ning Jiang<sup>1</sup> and Benjamin S. Halpern<sup>1,3,4</sup>

## Lessons learned:

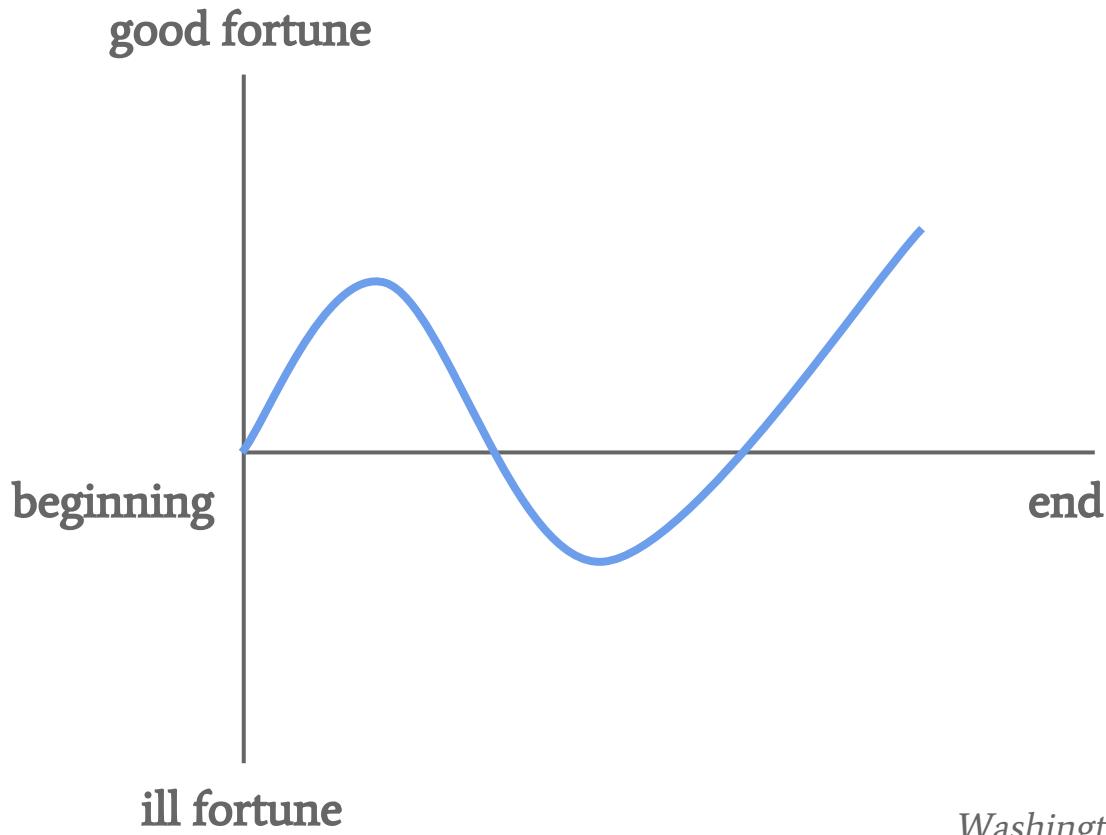
- barriers: exposure & change in mindset
- incorporate incrementally into existing practices (spectrum)

# Kurt Vonnegut on the shapes of stories



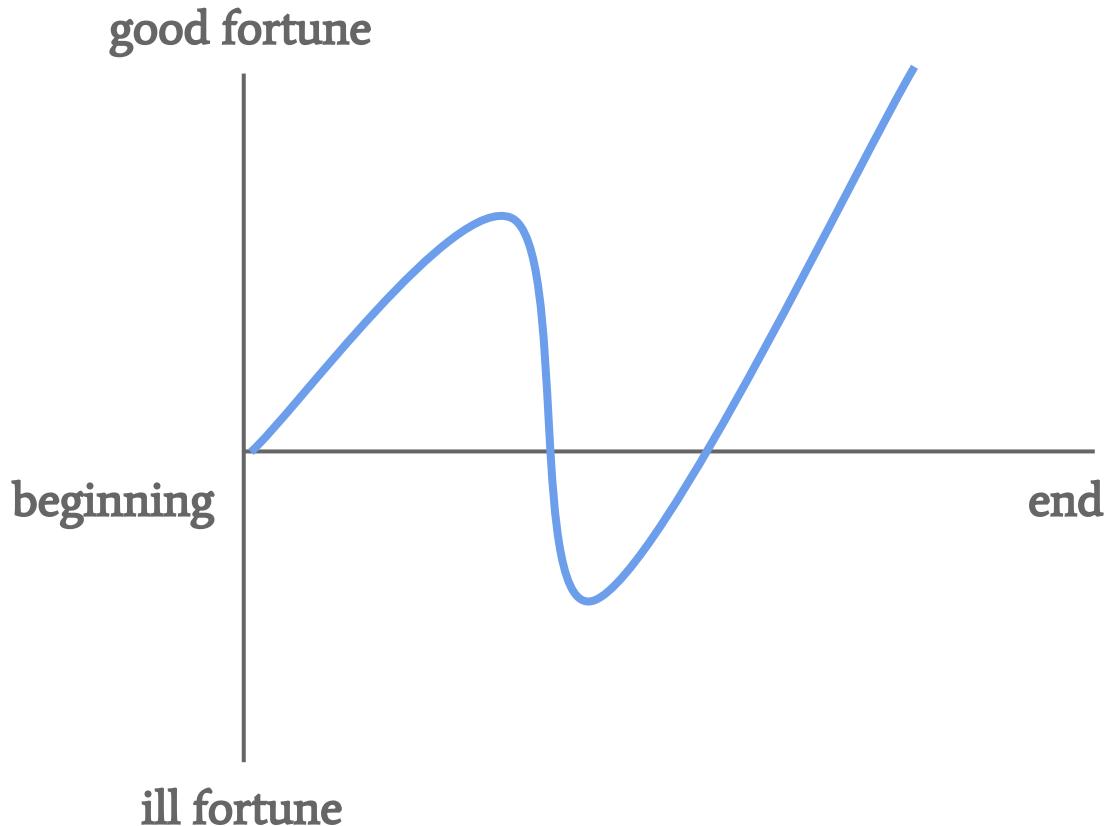
*Washington Post 2015; The Atlantic 2016; YouTube*

# Story arc: 'boy meets girl'

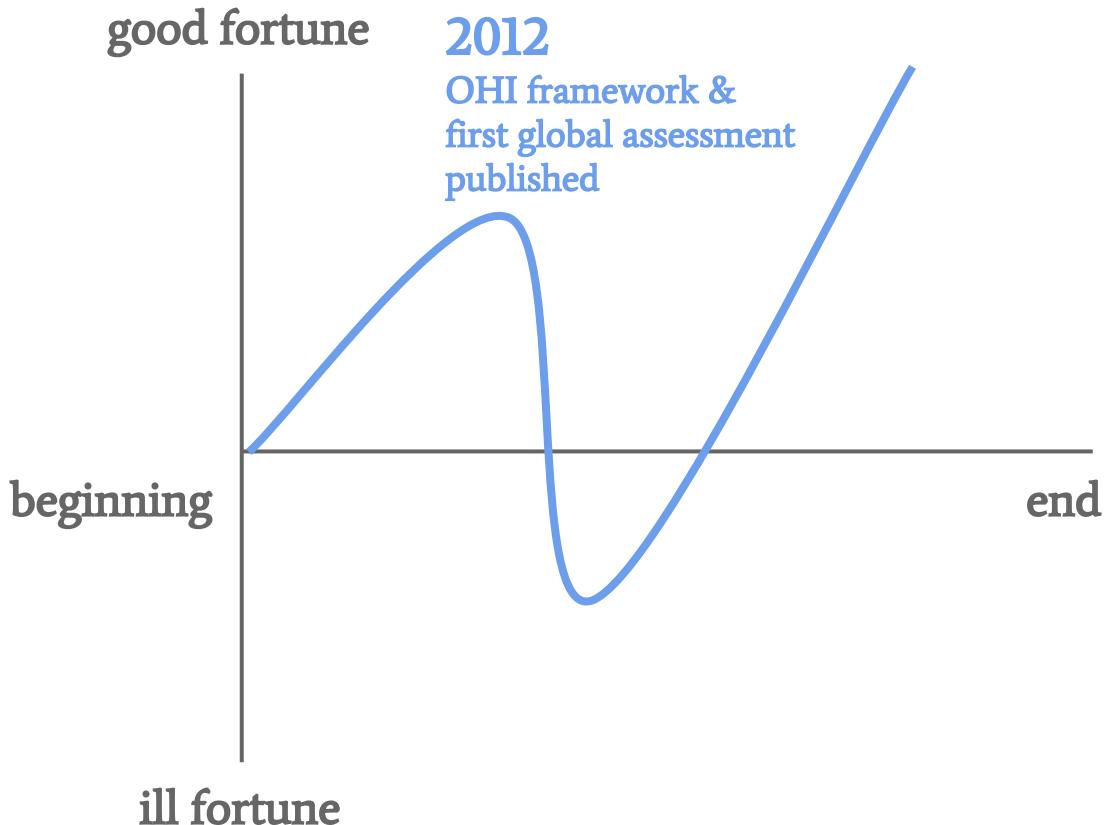


*Washington Post 2015; The Atlantic 2016; YouTube*

# Story arc: The Ocean Health Index

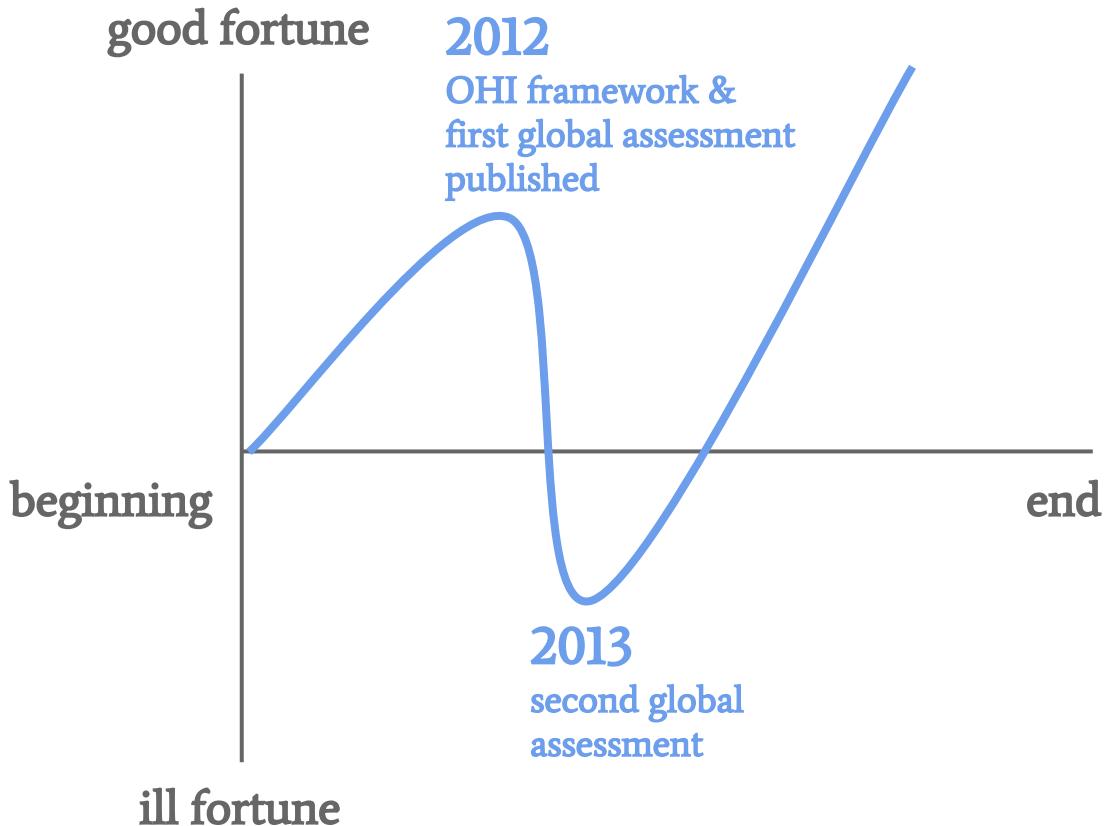


# Story arc: The Ocean Health Index

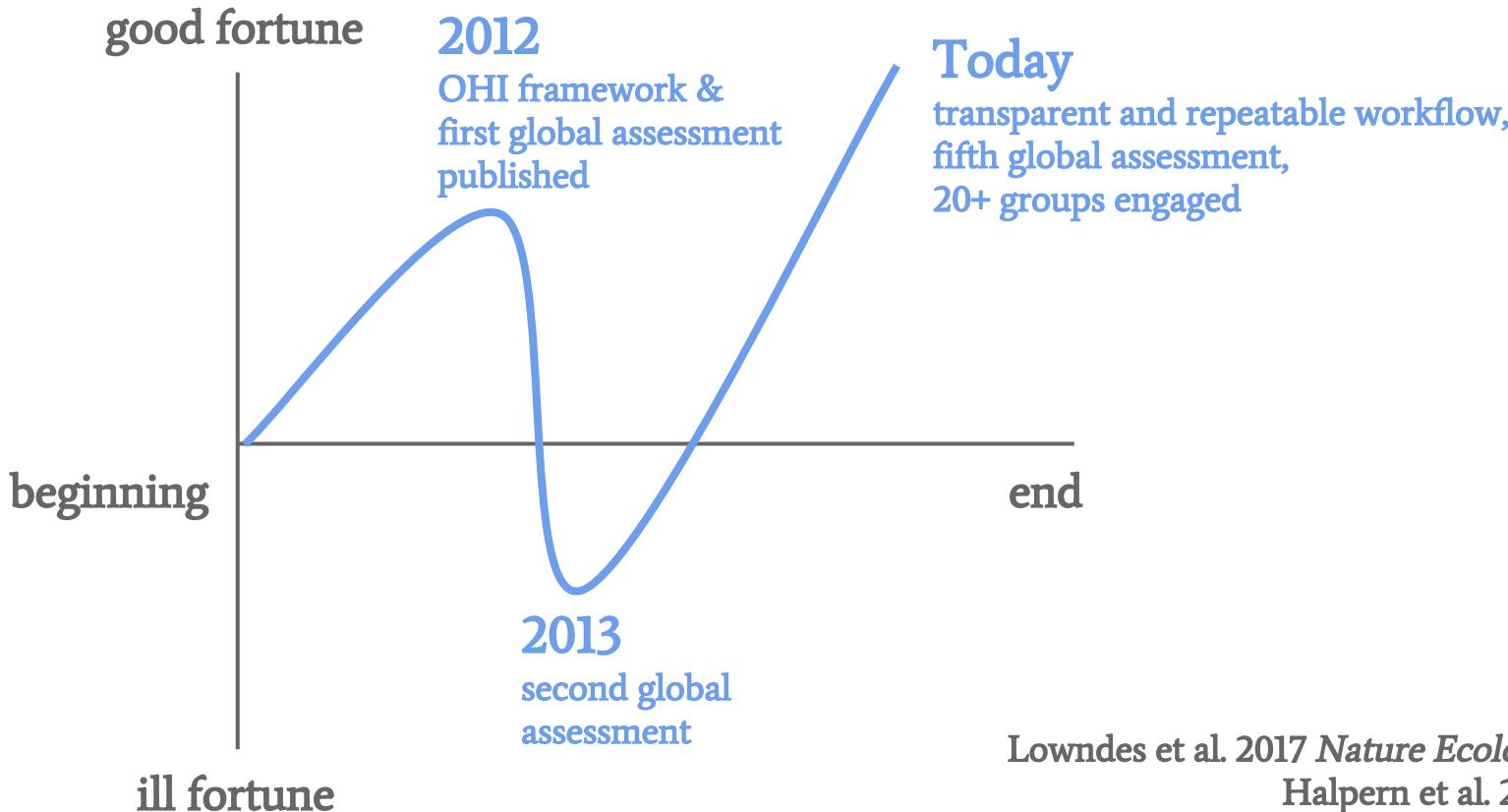


Halpern et al. 2017, *Nature*

# Story arc: The Ocean Health Index



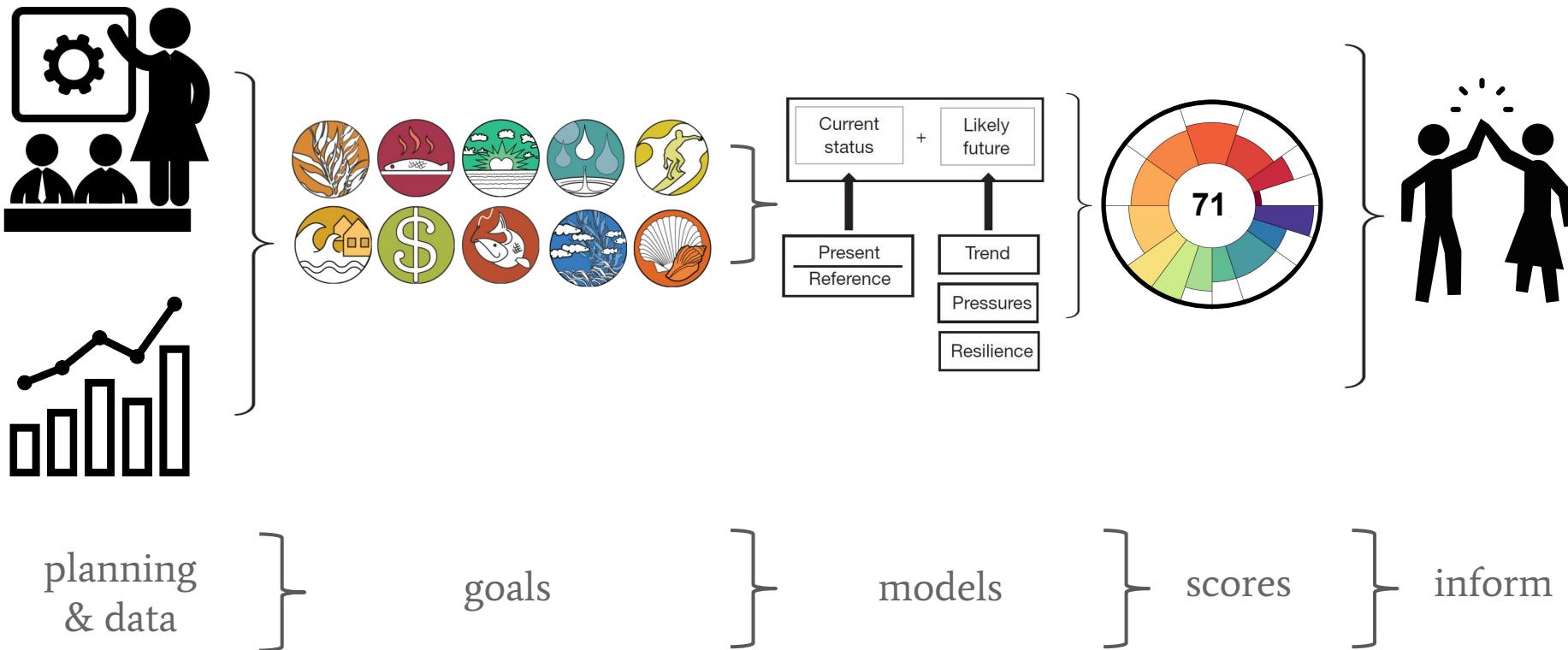
# Story arc: The Ocean Health Index



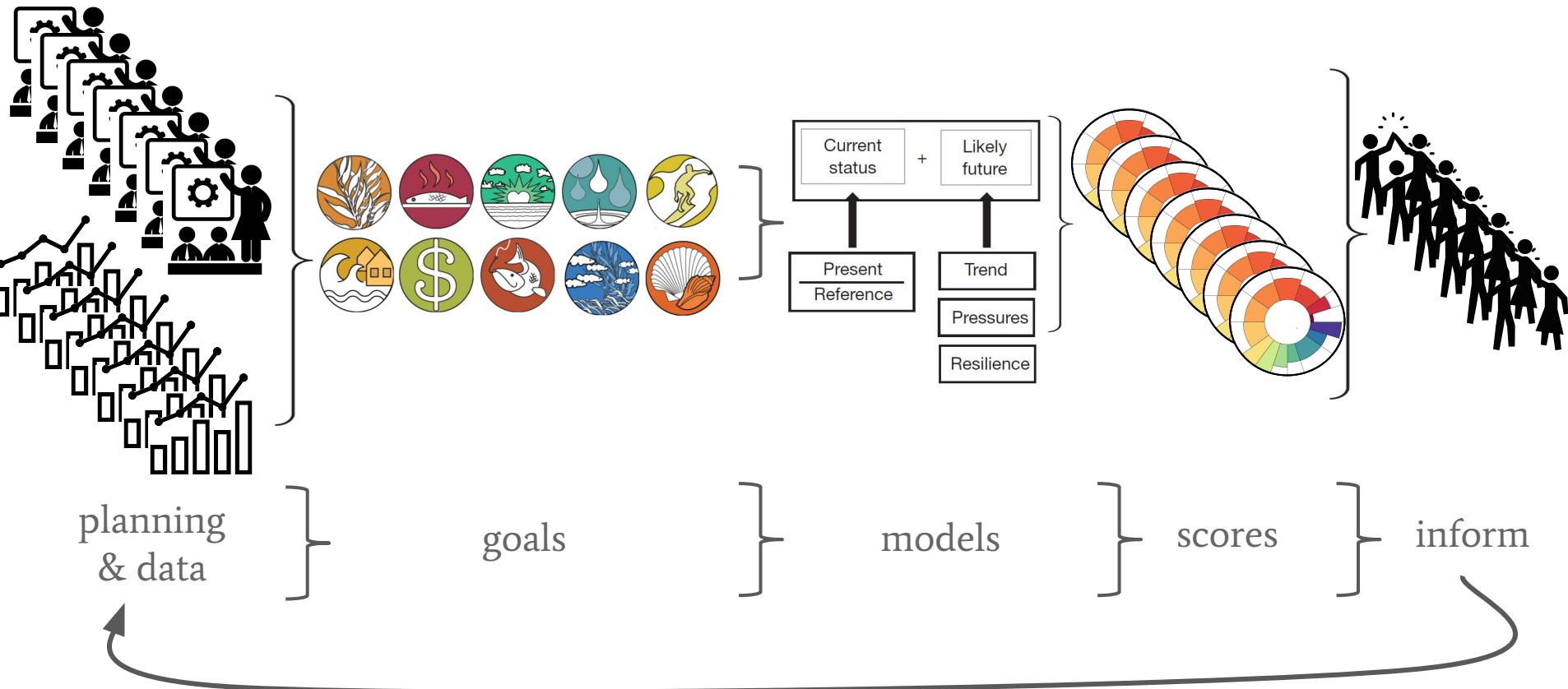
Lowndes et al. 2017 *Nature Ecology & Evolution*  
Halpern et al. 2017, *PLOS One*

# OHI framework

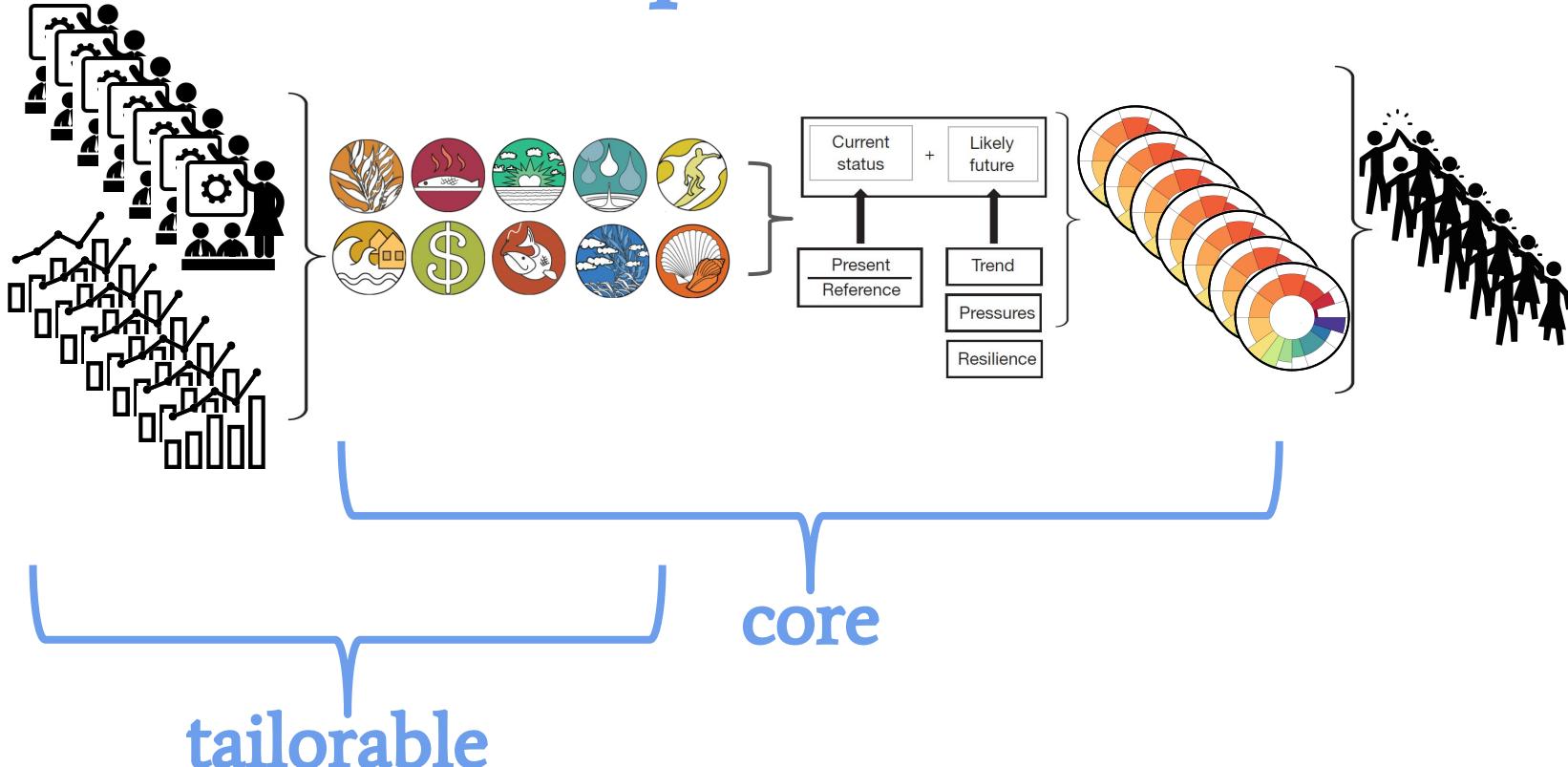
A healthy ocean sustainably delivers a range of benefits to people now and in the future.



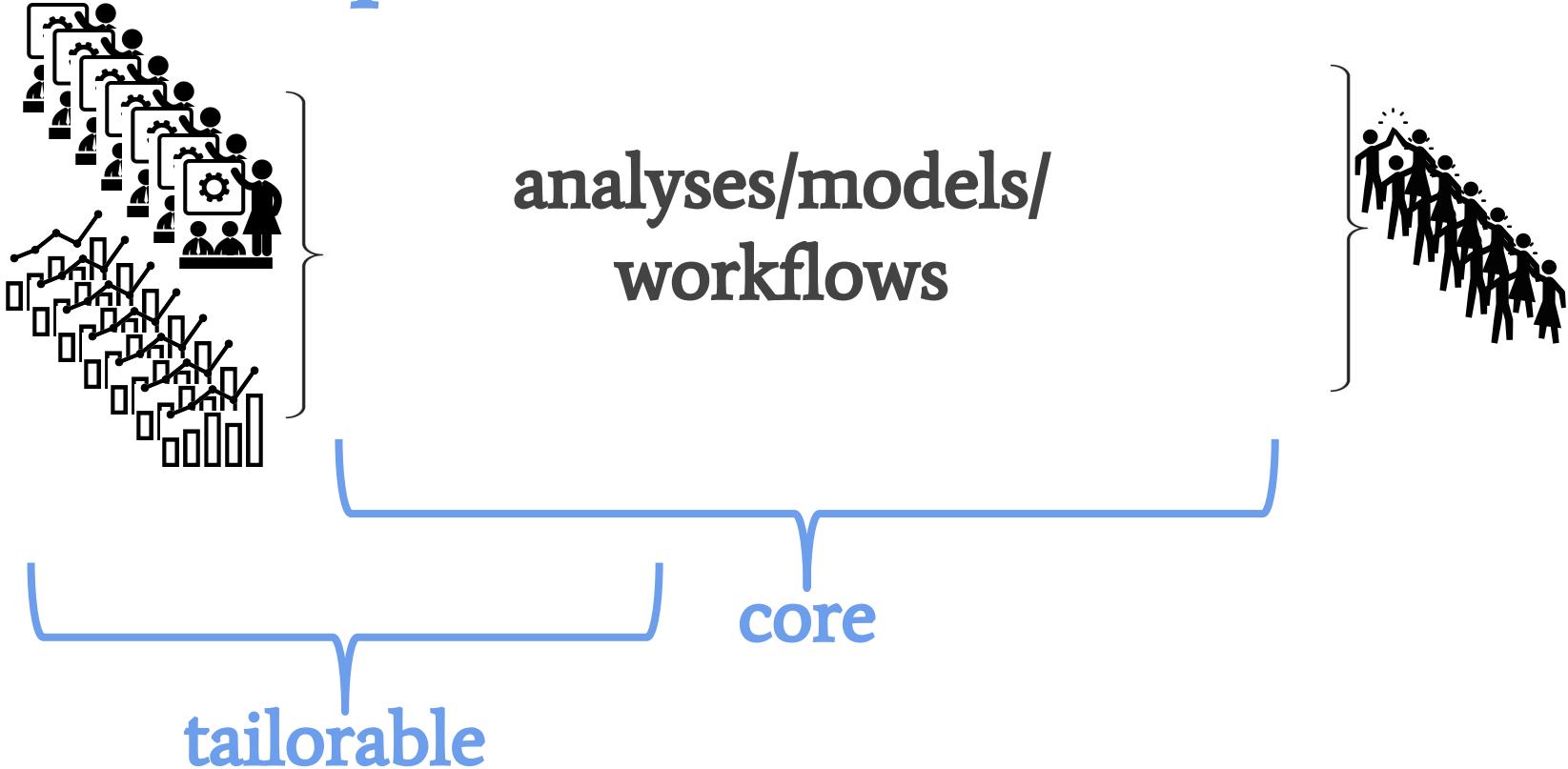
# Repeatable OHI assessment process



# OHI assessment process



# Scientific process





# An index to assess the health and benefits of the global ocean

Benjamin S. Halpern<sup>1,2</sup>, Catherine Longo<sup>1</sup>, Darren Hardy<sup>1</sup>, Karen L. McLeod<sup>3</sup>, Jameal F. Samhouri<sup>4</sup>, Steven K. Katona<sup>5</sup>, Kristin Kleisner<sup>6</sup>, Sarah E. Lester<sup>7,8</sup>, Jennifer O'Leary<sup>1</sup>, Marla Ranelletti<sup>1</sup>, Andrew A. Rosenberg<sup>5</sup>, Courtney Scarborough<sup>1</sup>, Elizabeth R. Selig<sup>5</sup>, Benjamin D. Best<sup>9</sup>, Daniel R. Brumbaugh<sup>10</sup>, F. Stuart Chapin<sup>11</sup>, Larry B. Crowder<sup>12</sup>, Kendra L. Daly<sup>13</sup>, Scott C. Doney<sup>14</sup>, Cristiane Elfes<sup>15,16</sup>, Michael J. Fogarty<sup>17</sup>, Steven D. Gaines<sup>8</sup>, Kelsey I. Jacobsen<sup>8</sup>, Leah Bunce Karrer<sup>5</sup>, Heather M. Leslie<sup>18</sup>, Elizabeth Neeley<sup>19</sup>, Daniel Pauly<sup>6</sup>, Stephen Polasky<sup>20</sup>, Bud Ris<sup>21</sup>, Kevin St Martin<sup>22</sup>, Gregory S. Stone<sup>5</sup>, U. Rashid Sumaila<sup>6</sup> & Dirk Zeller<sup>6</sup>



# data processing



data\_v1.xls



data\_v1b.xls



data\_v2\_2012\_02\_26.xls

# collaboration



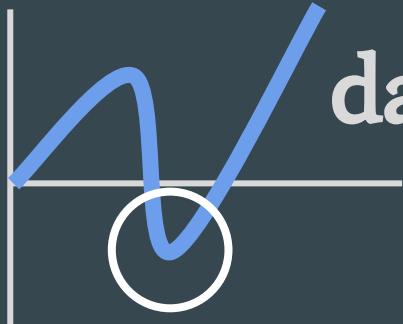
data question



Re:data question



FWD:Re:data question



# data processing



data\_v1.xls



data\_v1b.xls



data\_v2\_2012\_02\_26.xls

# collaboration



data question



Re:data question



FWD:Re:data question



# data processing

# collaboration



data



data\_v2\_2012\_02\_26.xls



tion



question



FWD : Re : data question



# data processing

# collaboration



coding – reusable record  
version control – bookkeeping

best practices – now + future  
shared – websites, online docs



# data processing

# collaboration



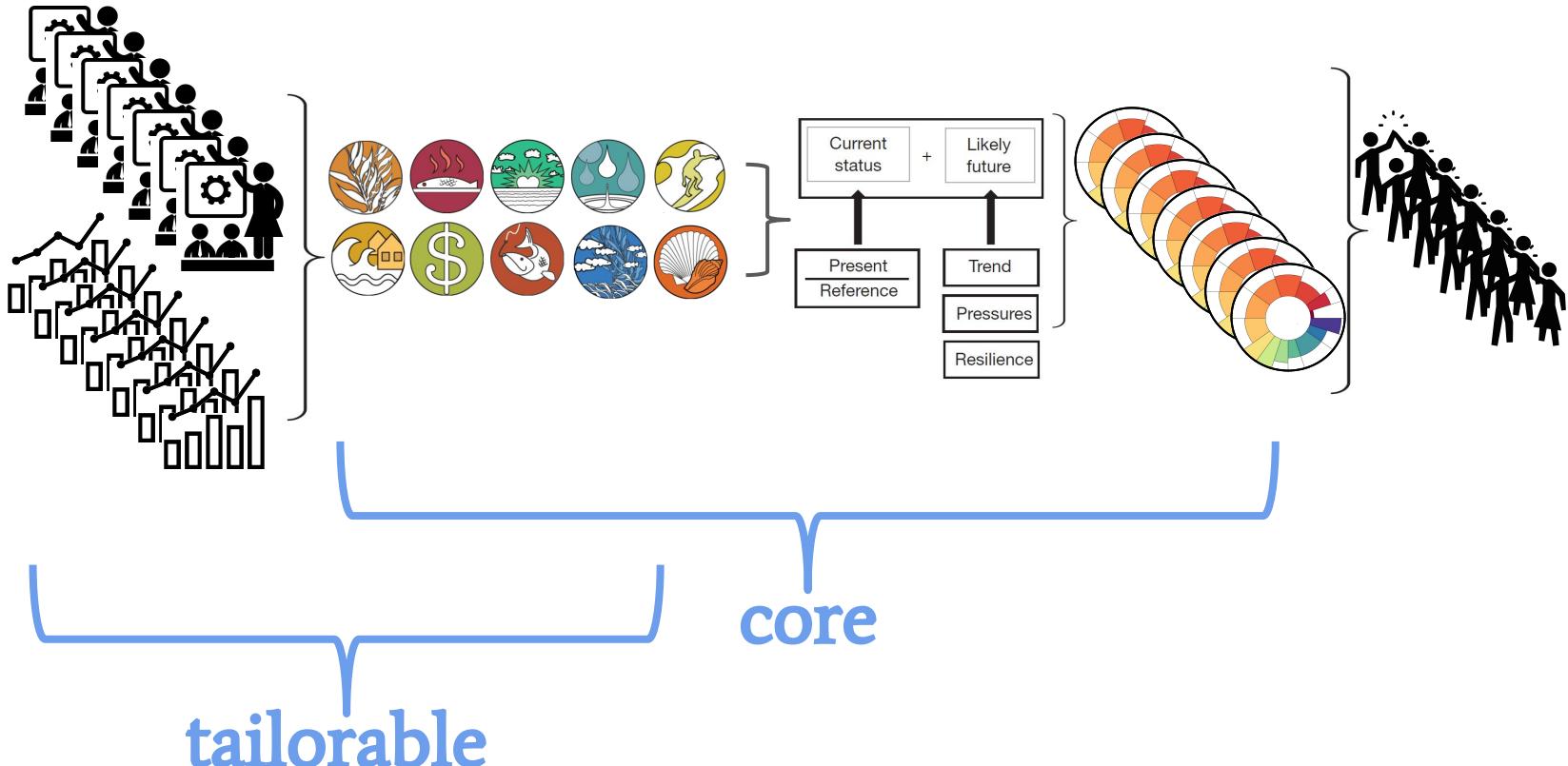
coding – reusable record  
version control – bookkeeping

best practices – now + future  
shared – websites, online docs

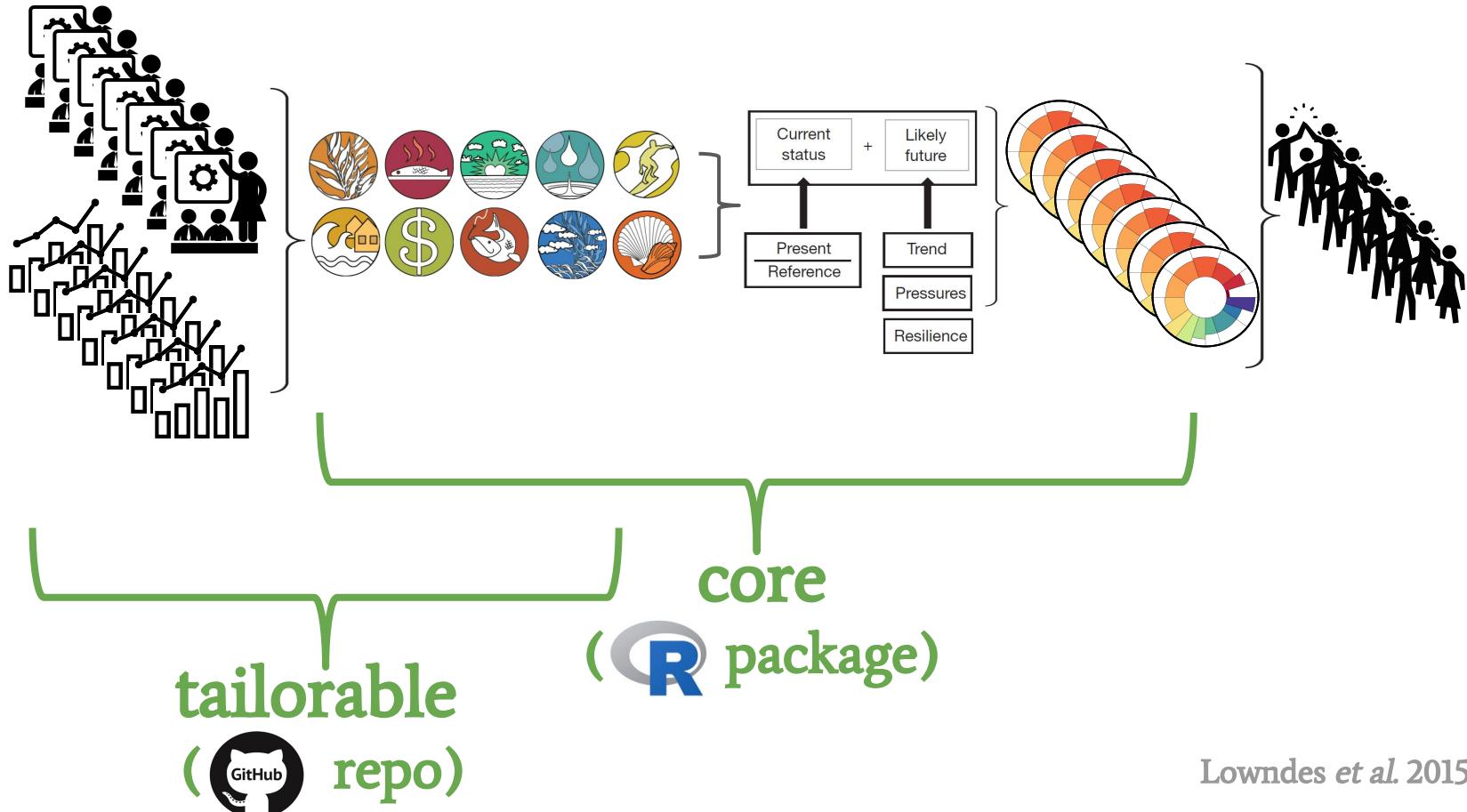


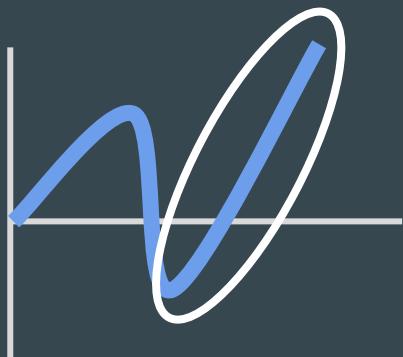
Communication. Open data science.

# OHI framework



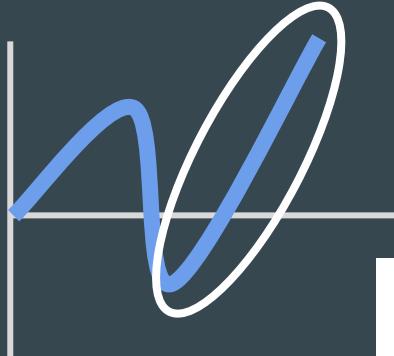
# OHI Toolbox





# How we work





# How we work

## Collaboration      Communication

Record keeping  
File sharing



Sharing data,  
code, methods



## Reproducibility

Code  
Version control

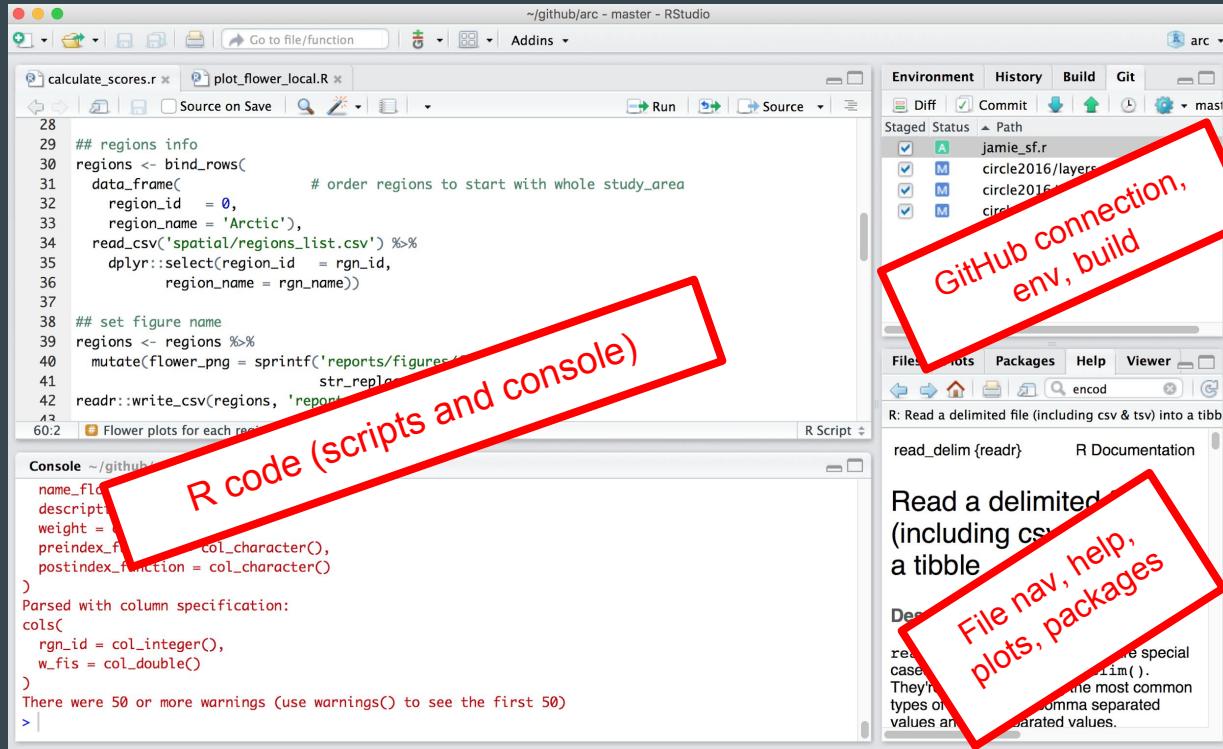


*local computer*



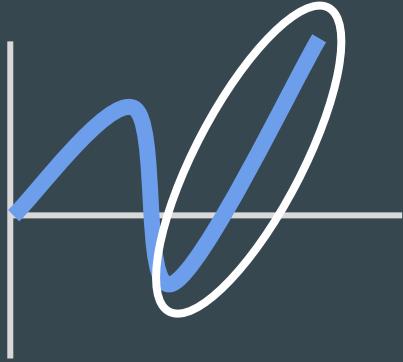
# How we work

## RStudio – GitHub workflow



The image shows a screenshot of the RStudio IDE. On the left, the code editor displays two files: `calculate_scores.r` and `plot_flower_local.R`. The `plot_flower_local.R` file contains R code for processing flower data. A red callout box points to the code in the editor with the text "R code (scripts and console)". The R console at the bottom shows the execution of the script, including the creation of a tibble and some warnings. A red callout box points to the console output with the text "R code (scripts and console)".

The right side of the screen shows the GitHub interface within RStudio. The sidebar displays the repository structure under "Staged Status". A red callout box points to the GitHub interface with the text "GitHub connection, env, build". The bottom right corner shows a partial view of the `read_delim` documentation page.

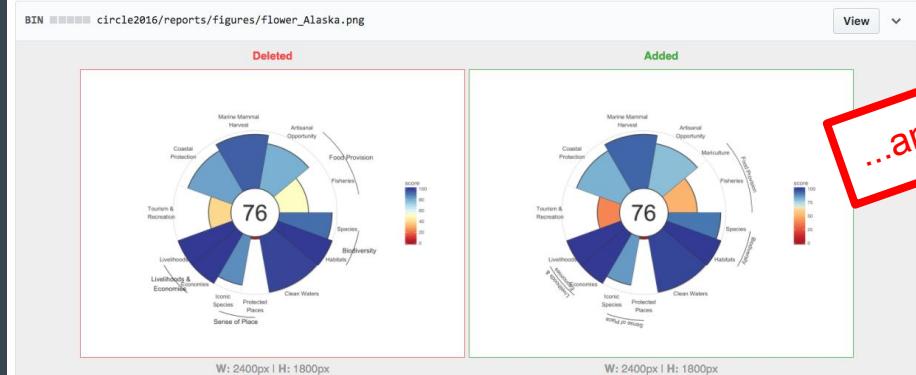


# How we work

## RStudio – GitHub workflow

```
64  +
61  65      ## set up positions for the bar centers:
62  66      ## cumulative sum of weights (incl current) minus half the current weight
63  67  score_df <- score_df %>
64  68  @@ -201,10 +205,10 @@ @ PlotFlower <- function(score_df,
65  69      mutate(supra_rad = supra_rad) %>
66  70      filter(!is.na(name_supra0))
67  71
68  72  - plot_obj +
69  73  + plot_obj <- plot_obj +
70  74      geom_text(data = st2,
71  75          inherit.aes = FALSE,
72  76          aes(label = st2$name_supra0, x = st2$pos_supra0, y = supra_rad, angle = st2$myAng),
73  77          aes(label = name_supra0, x = pos_supra0, y = supra_rad, angle = myAng),
74  78          hjust = .5, vjust = .5,
75  79          size = 3,
76  80          color = dark_line)
77  81
78  82  @@
```

See what changed  
line-by-line

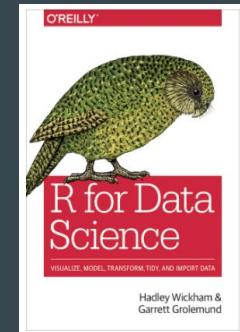
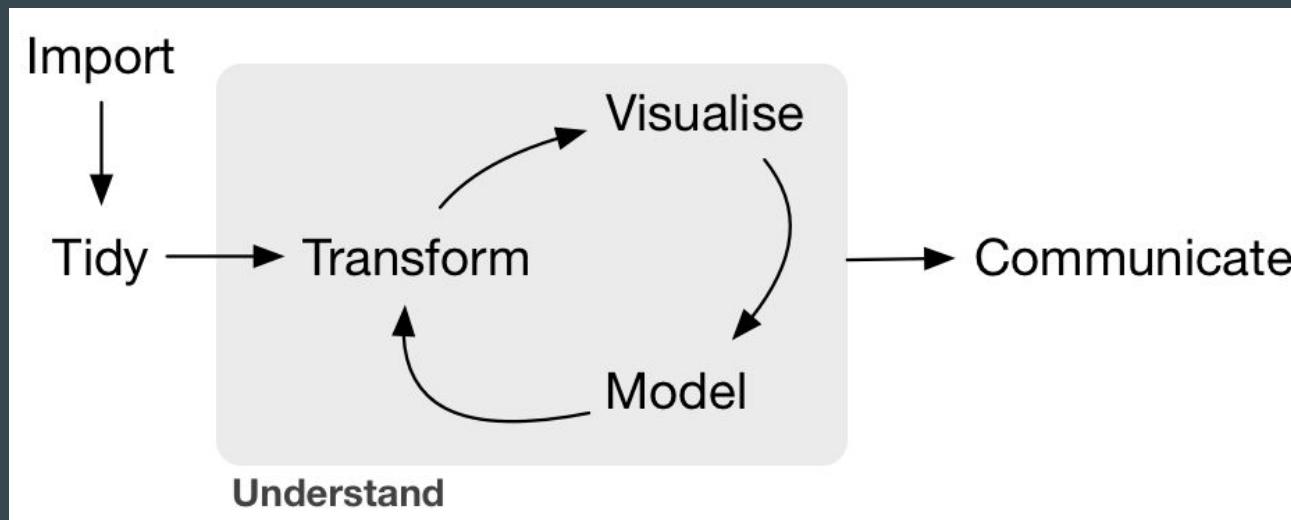


...and plot by plot



# How we work

## Data science theory & tidy data



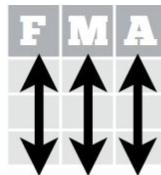
Wickham & Grolemund 2016: R for Data Science  
<http://r4ds.had.co.nz>



# How we work

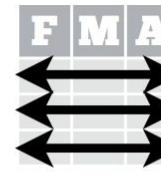
## Data science theory & tidy data

In a tidy data set:

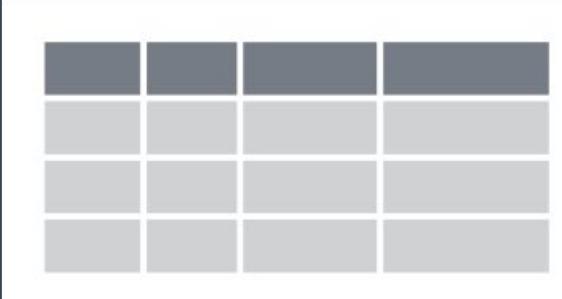


Each **variable** is saved  
in its own **column**

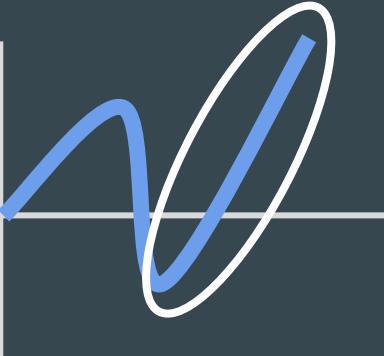
&



Each **observation** is  
saved in its own **row**



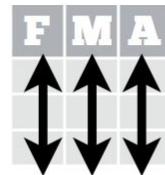
	country	continent	year	lifeExp	pop	gdpPercap
1	Afghanistan	Asia	1952	28.801	8425333	779.4453
2	Afghanistan	Asia	1957	30.332	9240934	820.8530
3	Afghanistan	Asia	1962	31.997	10267083	853.1007
4	Afghanistan	Asia	1967	34.020	11537966	836.1971
5	Afghanistan	Asia	1972	36.088	13079460	739.9811
6	Afghanistan	Asia	1977	38.438	14880372	786.1134
7	Afghanistan	Asia	1982	39.854	12881816	978.0114



# How we work

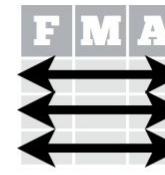
## Data science theory & tidy data

In a tidy data set:



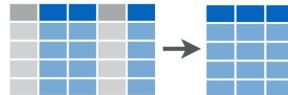
Each **variable** is saved  
in its own **column**

&



Each **observation** is  
saved in its own **row**

### Subset Variables (Columns)

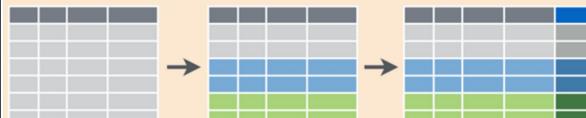


### Subset Observations (Rows)



### Group Data

Compute new variables by group.





# How we work

Best community practices  
for naming & organization

## Jenny Bryan: Naming files

machine readable, human readable,  
plays well with default ordering

NO

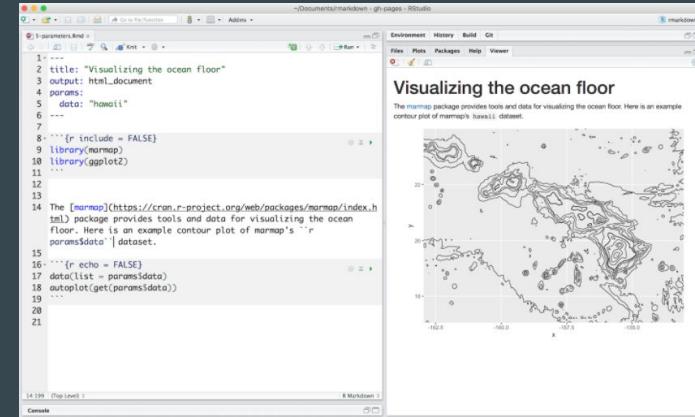
myabstract.docx  
Joe's Filenames Use Spaces and Punctuation.xlsx  
figure 1.png  
fig 2.png  
JW7d^(2sl@deletethisandyourcareerisoverWx2\*.txt

YES

2014-06-08\_abstract-for-sla.docx  
joes-filenames-are-getting-better.xlsx  
fig01\_scatterplot-talk-length-vs-interest.png  
fig02\_histogram-talk-attendance.png  
1986-01-28\_raw-data-from-challenger-o-rings.txt

## RStudio projects as GitHub repos

## RMarkdown combines code & text



The screenshot shows an RStudio interface with a code editor and a viewer pane. The code editor contains RMarkdown code:

```
1 ---  
2 title: "Visualizing the ocean floor"  
3 output: html_document  
4 params:  
5   data: "hawaii"  
6 ---  
7  
8 ---{r include = FALSE}  
9 library(marmap)  
10 library(ggplot2)  
11 ...  
12  
13  
14 The marmap package provides tools and data for visualizing the ocean floor. Here is an example contour plot of marmap's `hawaii` dataset.  
15  
16 ---{r echo = FALSE}  
17 data(list = params$data)  
18 autoplot(get(params$data))  
19 ...  
20  
21
```

The viewer pane displays a contour plot titled "Visualizing the ocean floor" showing the "hawaii" dataset. The plot shows depth contours in the Pacific Ocean around Hawaii.

# Working openly online - for science and communication: github.com/ohi-science

Global assessment!



Manuscript website!



Training e-book!



OHI+ assessment!



Science website!



**Ocean Health Index - Science**  
Open science with the Ocean Health Index  
Santa Barbara, CA http://ohi-science.org/

Repositories People 41 Teams 18 Projects 0 Settings

Search repositories... Type: All Language: All Customize pinned repositories New

**ohi-global**  
Global scenarios of the Ocean Health Index  
HTML ★ 1 9 Updated 7 minutes ago

**betterscienceinlesstime**  
Website for Better Science in Less Time  
HTML ★ 1 Updated 8 minutes ago

**toolbox-training**  
The Ocean Health Index Toolbox Training  
HTML Updated 33 minutes ago

**mhi**  
Ocean Health Index for Main Hawaiian Islands [mhi]  
R Updated 33 minutes ago

**ohi-science.github.io**  
Ocean Health Index - website  
HTML 1 Updated 3 hours ago

Top languages R HTML JavaScript CSS Python

People 41 >

Invite someone

Collaborators!



# Analyses – R code and text together (R Markdown)

Ocean Health Index -  
Open science with the Ocean Health Index  
Santa Barbara, CA http://ohi-science.org/

Repositories People 41 Teams 18

Search repositories... Type

**ohi-global**  
Global scenarios of the Ocean Health Index

HTML ★ 1 9 Updated 7 minutes ago

**betterscienceinlesslime**  
Website for Better Science in Less Time

HTML ★ 1 Updated 8 minutes ago

**toolbox-training**  
The Ocean Health Index Toolbox Training

HTML Updated 33 minutes ago

**mhi**  
Ocean Health Index for Main Hawaiian Islands [mhi]

R Updated 33 minutes ago

**ohi-science.github.io**  
Ocean Health Index - website

HTML 1 Updated 3 hours ago

## Data Source

Reference: Feely et al.(2009)

Downloaded: March 15, 2016

Description: Aragonite Saturation State  $\Omega_{\text{arg}}$

Native data resolution: 1 degree cells

Time range: 1880-1889 and 2005-2015, monthly data provided for each year

Format: NetCDF

### Notes about the data:

This data was shared with us by Ivan Lima from Woods Hole Institute for Oceanography in December 2014 and again February 2016. The data came as NetCDFs in an irregular grid format with a resolution of about 1 degree. The data values are monthly average surface  $\Omega$  aragonite saturation state.

## Methods

### Setup

The main R libraries needed for this analysis are the `raster`, and `ncdf4` packages.

```
#set options for all chunks in code
knitr::opts_chunk$set(warning=FALSE, message=FALSE,fig.width=6, fig.height=6)
source("~/github/ohiprep/src/R/common.R")
```

Hide

# Science website: ohi-science.org

The screenshot shows the homepage of the Ocean Health Index Science website. The header features a blue navigation bar with the "OCEAN HEALTH INDEX SCIENCE" logo, a search bar, and language selection options. Below the header is a large, scenic background image of a school of fish swimming in the ocean. Overlaid on this image is the main title "Ocean Health Index" in large white letters, followed by the subtitle "open data science tools and resources for marine science and management" in a slightly smaller white font.

**OCEAN HEALTH INDEX SCIENCE**

PROJECTS RESOURCES ABOUT NEWS

Select Language Powered by Google TRANSLATE

Search repositories...

**ohi-global**  
Global scenarios of the Ocean Health Index  
HTML ★ 1 Updated 7 minutes ago

**betterscienceinlesstime**  
Website for Better Science in Less Time  
HTML ★ 1 Updated 8 minutes ago

**toolbox-training**  
The Ocean Health Index Toolbox Training  
HTML Updated 33 minutes ago

**mhi**  
Ocean Health Index for Main Hawaiian Islands [mhi]  
R Updated 33 minutes ago

**ohi-science.github.io**  
Ocean Health Index - website  
HTML ★ 1 Updated 3 hours ago

# Interactive websites for published articles: [http://ohi-science.nceas.ucsb.edu/plos\\_change\\_in\\_global\\_ocean\\_health](http://ohi-science.nceas.ucsb.edu/plos_change_in_global_ocean_health)

Abstract

Score maps

Trends by goal

Trend v. score

Trend bars

Model eval

Rank change

Tables

Data

## Drivers and implications of change in global ocean health over the past five years

Benjamin S. Halpern<sup>1,2,3\*</sup> • Melanie Frazier<sup>1</sup> • Jamie Aflerbach<sup>1</sup>  
• Casey O'Hara<sup>1</sup> • Steven Katona<sup>4</sup> • Julia S. Stewart Lowndes<sup>1</sup>  
• Ning Jiang<sup>1</sup> • Erich Pacheco<sup>4</sup> • Courtney Scarborough<sup>1</sup> •  
Johanna Polensberg<sup>4</sup>

1. National Center for Ecological Analysis and Synthesis, University of California, Santa Barbara, California, USA

2. Bren School of Environmental Science & Management, University of California, Santa Barbara, California, USA

3. Imperial College London, Silwood Park Campus, Ascot, United Kingdom

4. Conservation International, Arlington, Virginia, USA

Please cite as: Halpern BS, Frazier M, Aflerbach J, O'Hara C, Katona S, Lowndes JSS, Jiang N, Pacheco E, Scarborough C, Polensberg P (2017)

Drivers and implications of change in global ocean health over the past five years. PLoS ONE 12(7): e0178267. doi:10.1371/journal.pone.0178267

Beau and Heather Wrigley generously provided the funding grant to the Ocean Health Index. The Pacific Life Foundation is the Founding Presenting Sponsor.

Financial support has also been provided by Jayne and Hans Hufschmid and Dan Sten Olsson, CEO of Stena AB.

## The Ocean Health Index

A healthy ocean sustainably delivers a range of benefits to people now and in the future. The Ocean Health Index is the comprehensive framework used to quantify ocean-derived benefits to humans and to help inform sustainable ocean management using the best available information. Assessments using the OHI framework require synthesizing existing data representing those benefits, using methods that are reproducible and repeatable. Repeated assessments using the same methods enable quantifiable comparison of changes in ocean health through time, which can be used to inform policy and track progress.

Visit (<http://www.ohi-science.org/>) for more about the science and methods behind the Ocean Health Index, or (<http://www.oceanhealthindex.org/>) for an overview of the Ocean Health Index project.

## Abstract of published paper

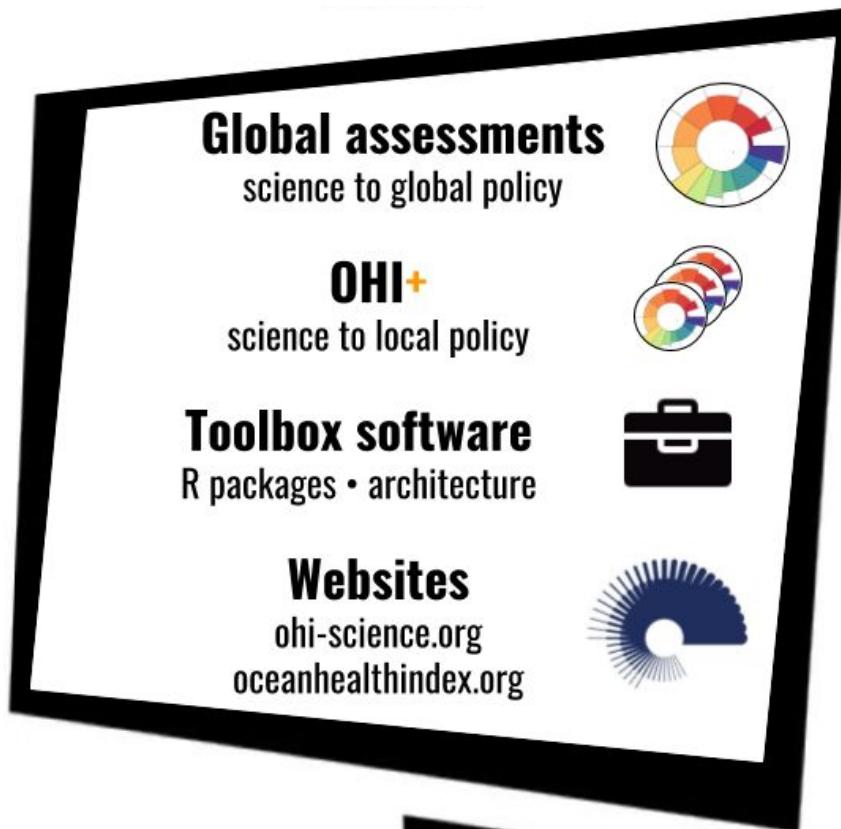
### Drivers and implications of change in global ocean health over the past five years

Growing international and national focus on quantitatively measuring and improving ocean health has increased the need for comprehensive, scientific, and repeated indicators to track progress towards achieving policy and societal goals. The Ocean Health Index (OHI) is one of the few indicators available for this purpose. Here we present results from five years of annual global assessment for 220 countries and territories, evaluating potential drivers and consequences of changes and presenting lessons learned about the challenges of using composite indicators to measure sustainability goals.

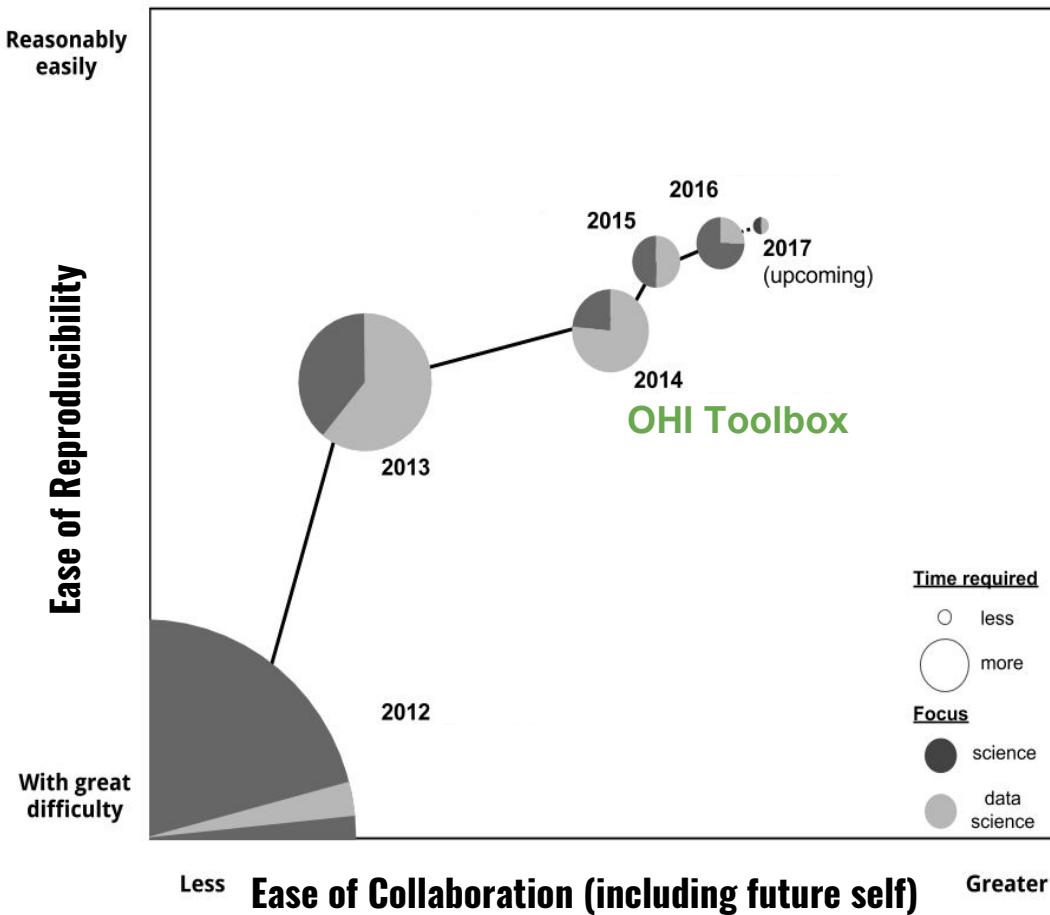
Globally scores have shown little change, as would be expected. However, individual countries have seen notable increases or declines due in particular to improvements in the harvest and management of wild-caught fisheries, the creation of marine protected areas (MPAs), and decreases in natural product harvest. Rapid loss of sea ice and the consequent reduction of coastal protection from that sea ice was also responsible for declines in overall ocean health in many Arctic and sub-Arctic countries. The OHI performed reasonably well at predicting near-term future scores for many of the ten goals measured, but data gaps and limitations hindered these predictions for many other goals.

Ultimately, all indicators face the substantial challenge of informing policy for progress toward broad goals and objectives with insufficient monitoring and assessment data. If countries and the global community hope to achieve and maintain healthy oceans, we will need to dedicate significant resources to measuring what we are trying to manage.

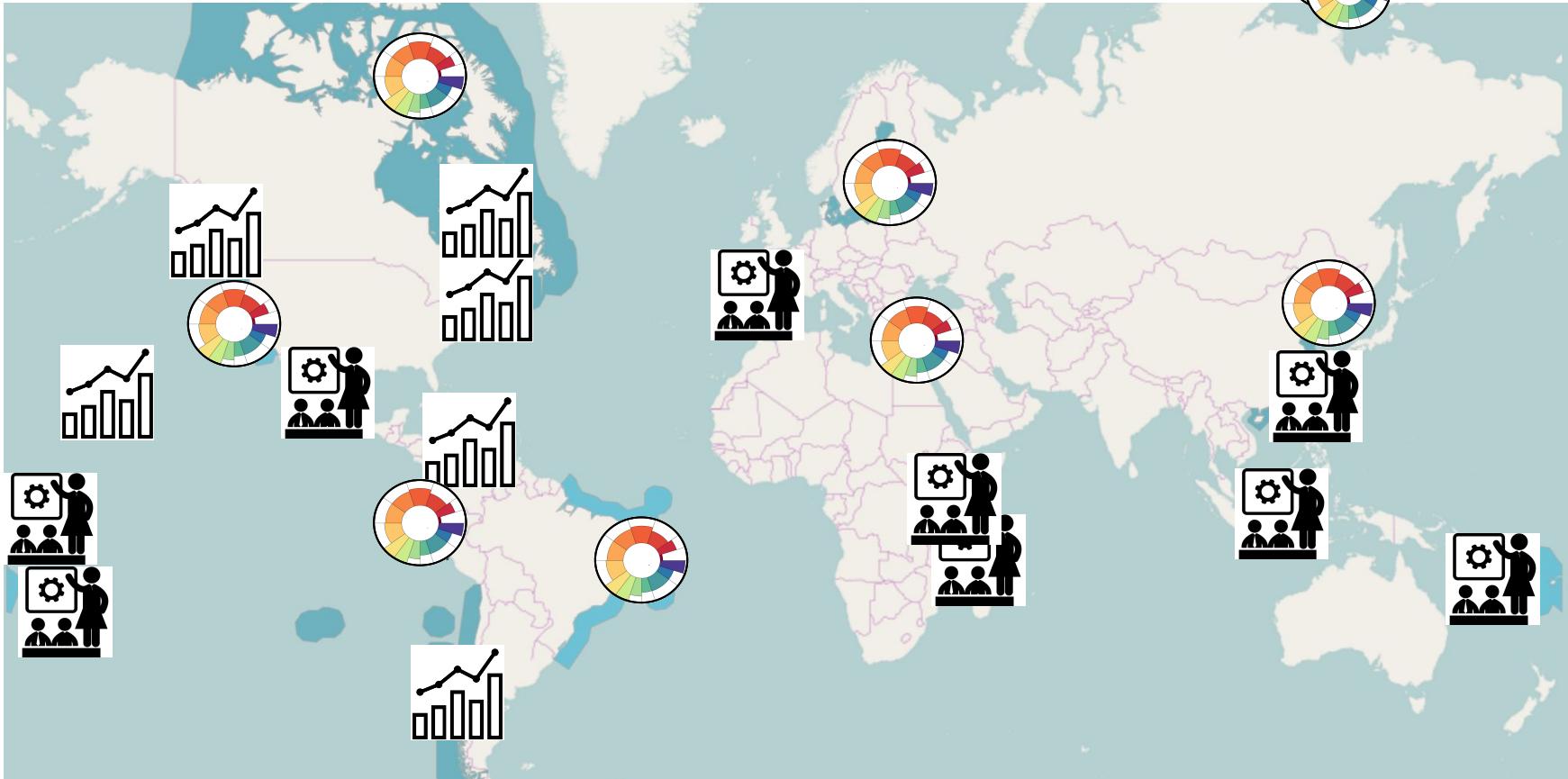
# The OHI and open data science



# With open data science, repeating analyses takes less time



# OHI Assessments



**So what can you do?**

# Resources

So many awesome resources...

academic pubs, webinars, books, blogs, trainings, tutorials, podcasts, etc

A few lists we're trying to keep updated:

<http://ohi-science.org/betterscienceinlesstime>

1. **Specific resources that helped us learn**

e.g. R for Data Science, RStudio webinars & cheatsheets, Software Carpentry

2. **Academic literature + media on the importance of open data science for science**

e.g. How open science helps researchers succeed (McKiernan 2016),

Git can facilitate greater reproducibility and increased transparency in science (Ram 2013)

## **Learn in an intentional way**

**Use online resources for self-paced learning**

books, tutorials, slide decks, archived webinars, presentations, blogs & more

## **Learn in an intentional way**

### **Use online resources for self-paced learning**

books, tutorials, slide decks, archived webinars, presentations, blogs & more

### **Learn together by joining and/or creating local communities**

academic articles or webinars during lab meetings, journal clubs, or happy hours.

PIs can encourage this culture even if they don't code themselves

## **Learn in an intentional way**

### **Use online resources for self-paced learning**

books, tutorials, slide decks, archived webinars, presentations, blogs & more

### **Learn together by joining and/or creating local communities**

academic articles or webinars during lab meetings, journal clubs, or happy hours.

PIs can encourage this culture even if they don't code themselves

### **Ask for help**

expect that someone has already had this question/this error (paste errors into Google!)

# **Learn in an intentional way**

## **Use online resources for self-paced learning**

books, tutorials, slide decks, archived webinars, presentations, blogs & more

## **Learn together by joining and/or creating local communities**

academic articles or webinars during lab meetings, journal clubs, or happy hours.

PIs can encourage this culture even if they don't code themselves

## **Ask for help**

expect that someone has already had this question/this error (paste errors into Google!)

## **Attend in-person workshops and conferences**

workshops to learn skills, conferences to see examples “in the wild”

# Learn in an intentional way

## Use online resources for self-paced learning

books, tutorials, slide decks, archived webinars, presentations, blogs & more

## Learn together by joining and/or creating local communities

academic articles or webinars during lab meetings, journal clubs, or happy hours.

PIs can encourage this culture even if they don't code themselves

## Ask for help

expect that someone has already had this question/this error (paste errors into Google!)

## Attend in-person workshops and conferences

workshops to learn skills, conferences to see examples “in the wild”

## Engage online

see what's new, how people are working:

#rstats on Twitter, rOpenSci, RStudio, Software Carpentry, etc

# Our path to better science in less time using open data science tools

Julia S. Stewart Lowndes<sup>1\*</sup>, Benjamin D. Best<sup>2</sup>, Courtney Scarborough<sup>1</sup>, Jamie C. Afflerbach<sup>1</sup>,  
Melanie R. Frazier<sup>1</sup>, Casey C. O'Hara<sup>1</sup>, Ning Jiang<sup>1</sup> and Benjamin S. Halpern<sup>1,3,4</sup>

## Lessons learned:

- barriers: exposure & change in mindset
- incorporate incrementally into existing practices (spectrum)



# Better science in less time

How open data science tools have improved our science

## Thank you

[ohi-science.org](http://ohi-science.org) // @OHlscience

[oceanhealthindex.org](http://oceanhealthindex.org) // @OceanHealthIdx

# Julia Stewart Lowndes

[lowndes@nceas.ucsb.edu](mailto:lowndes@nceas.ucsb.edu)

twitter: @juliesquid



DataONE Webinar  
October 9, 2017



# Some of the most useful advice

*The internet will make those bad words go away*



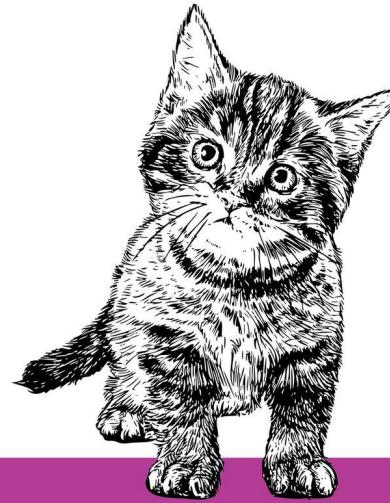
*Essential*

Googling the  
Error Message

O RLY?

*The Practical Developer  
@ThePracticalDev*

*How to actually learn any new programming concept*

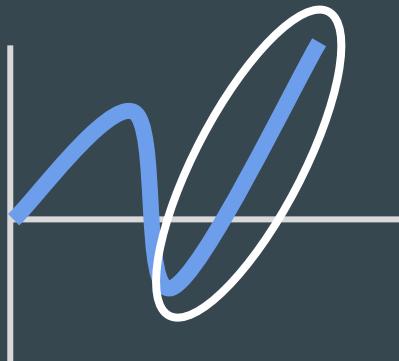


*Essential*

Changing Stuff and  
Seeing What Happens

O RLY?

*@ThePracticalDev*



# How we work

Best community practices  
for naming & organization

## Jenny Bryan: Naming files

machine readable, human readable,  
plays well with default ordering

NO

myabstract.docx  
Joe's Filenames Use Spaces and Punctuation.xlsx  
figure 1.png  
fig 2.png  
JW7d^(2sl@deletethisandyourcareerisoverWx2\*.txt

YES

2014-06-08\_abstract-for-sla.docx  
joes-filenames-are-getting-better.xlsx  
fig01\_scatterplot-talk-length-vs-interest.png  
fig02\_histogram-talk-attendance.png  
1986-01-28\_raw-data-from-challenger-o-rings.txt

## RStudio projects as GitHub repos

