



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^{1*}, Benjamin D. Best², Courtney Scarborough¹, Jamie C. Afflerbach¹,
Melanie R. Frazier¹, Casey C. O'Hara¹, Ning Jiang¹ and Benjamin S. Halpern^{1,3,4}

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

Julia S. Stewart Lowndes^{1*}, Benjamin D. Best², Courtney Scarborough¹, Jamie C. Afflerbach¹,
Melanie R. Frazier¹, Casey C. O'Hara¹, Ning Jiang¹ and Benjamin S. Halpern^{1,3,4}

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^{1*}, Benjamin D. Best², Courtney Scarborough¹, Jamie C. Afflerbach¹,
Melanie R. Frazier¹, Casey C. O'Hara¹, Ning Jiang¹ and Benjamin S. Halpern^{1,3,4}

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)

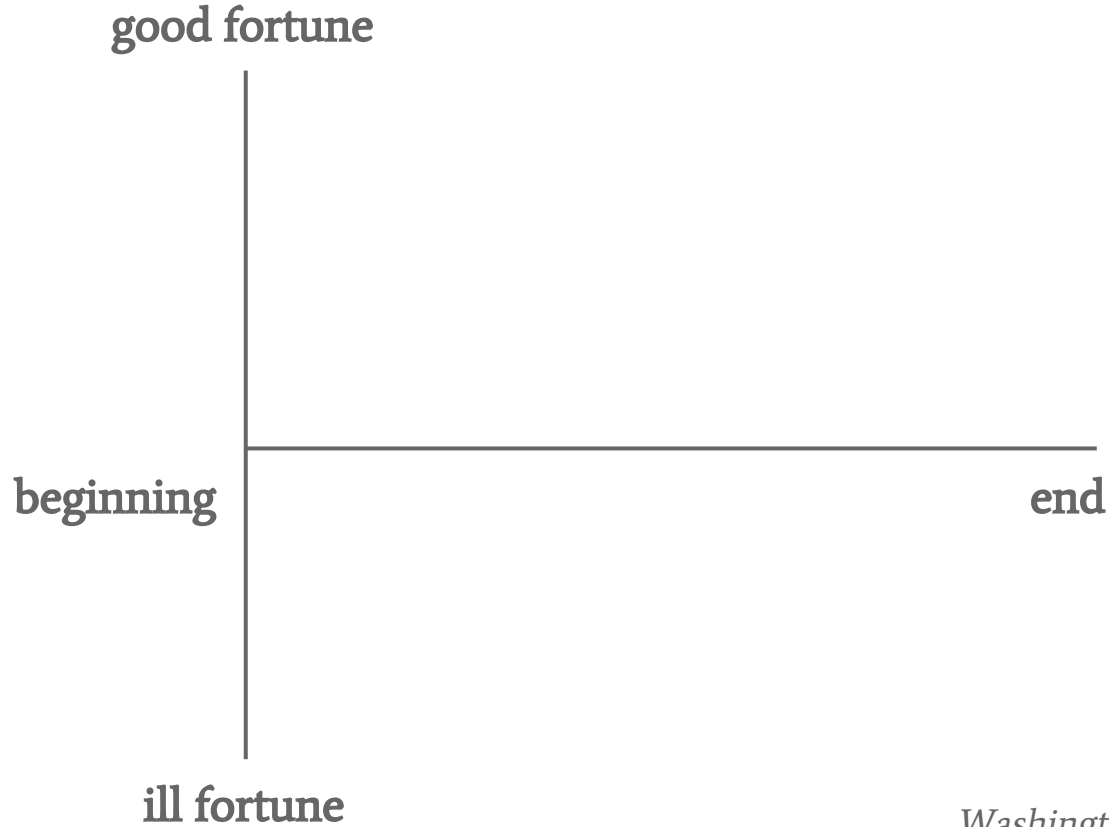
Our path to better science in less time using open data science tools

Julia S. Stewart Lowndes^{1*}, Benjamin D. Best², Courtney Scarborough¹, Jamie C. Afflerbach¹,
Melanie R. Frazier¹, Casey C. O'Hara¹, Ning Jiang¹ and Benjamin S. Halpern^{1,3,4}

Lessons learned:

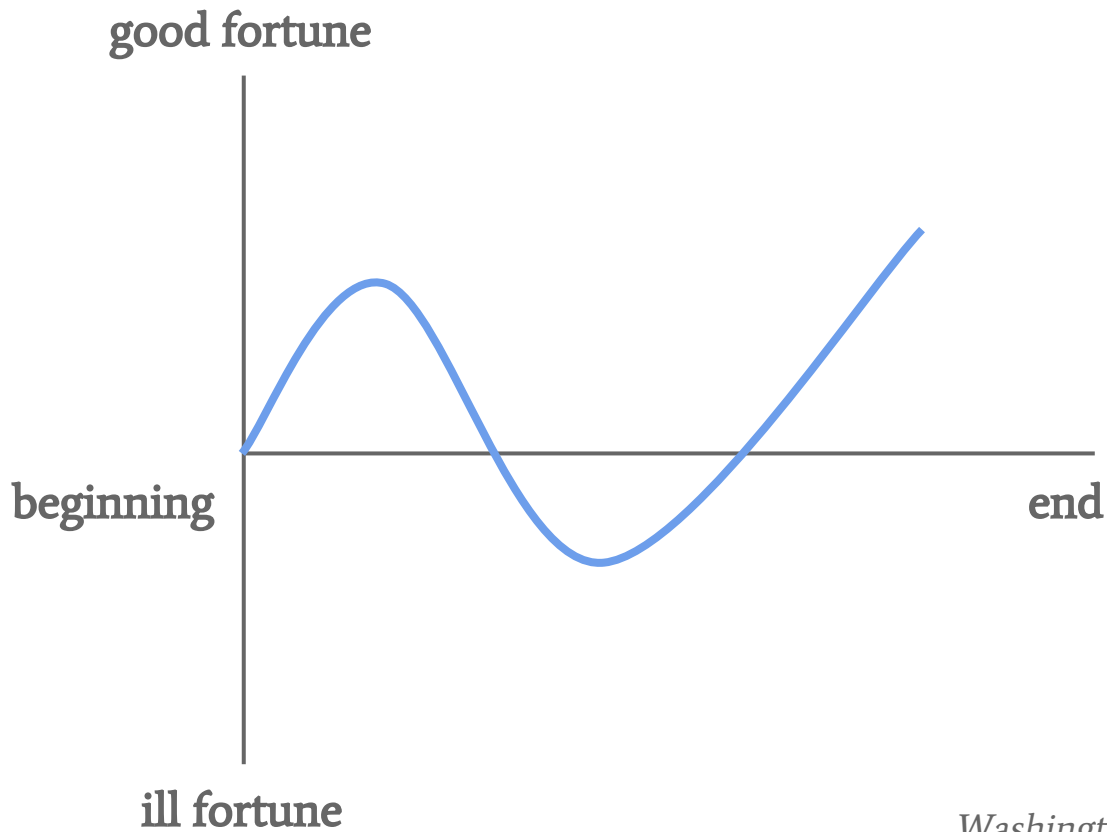
- barriers: exposure & change in mindset
- incorporate incrementally into existing practices (spectrum)

Kurt Vonnegut on the shapes of stories



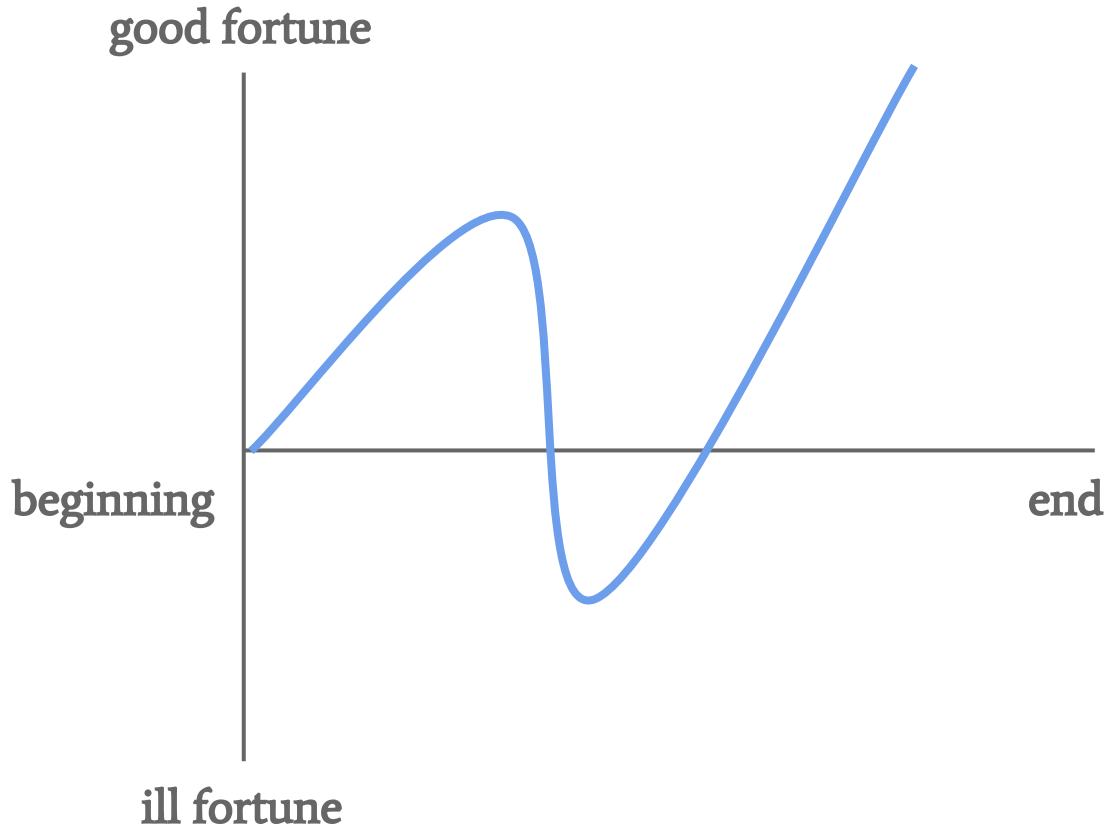
Washington Post 2015; *The Atlantic* 2016; YouTube

Story arc: 'boy meets girl'

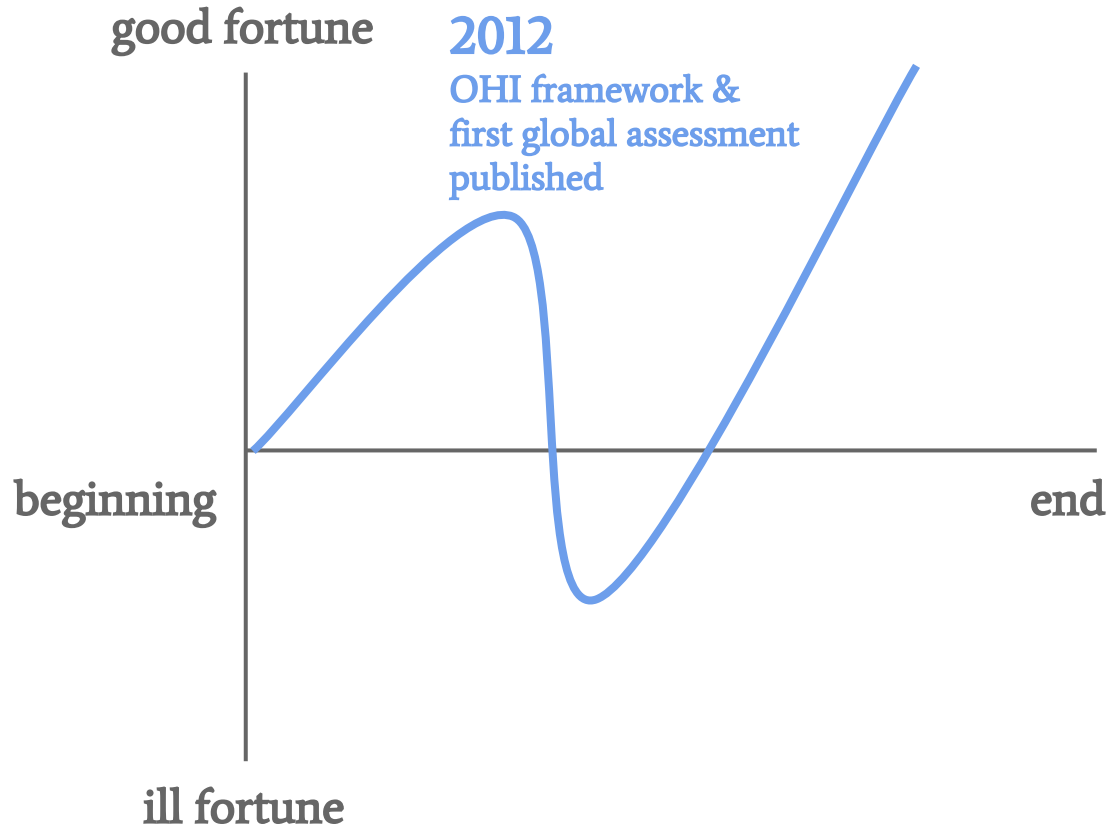


Washington Post 2015; The Atlantic 2016; YouTube

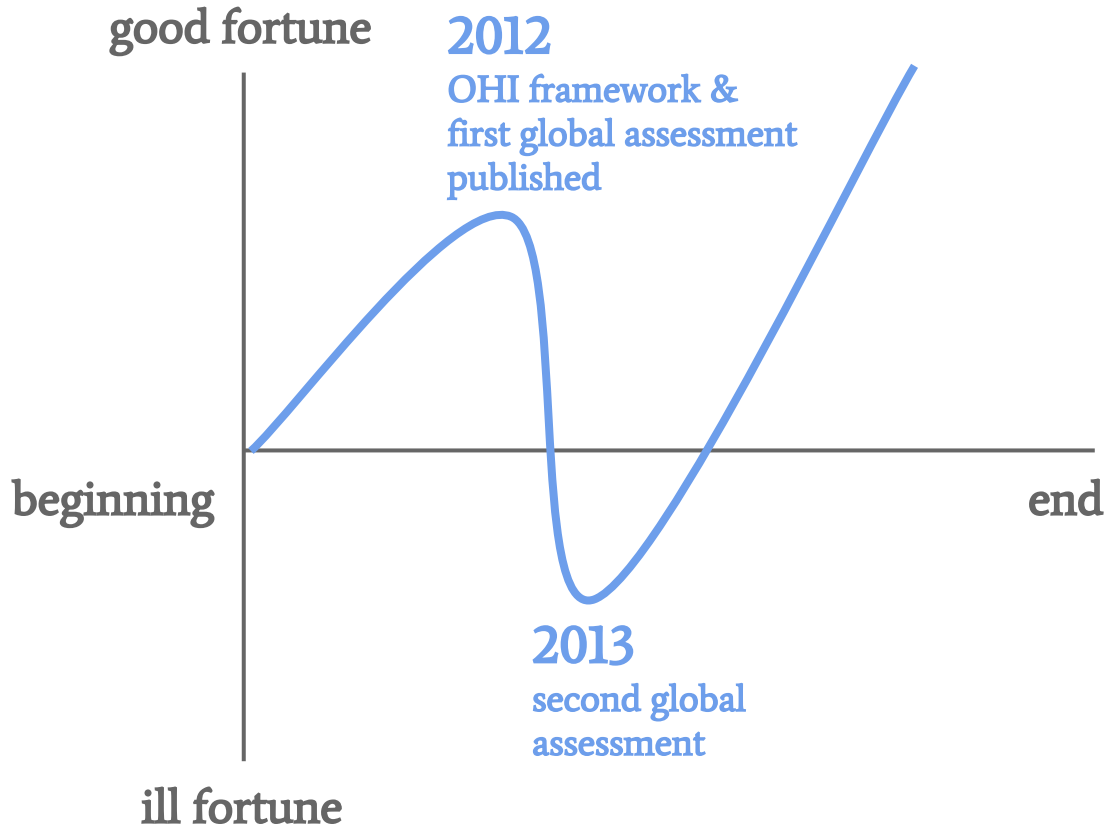
Story arc: The Ocean Health Index



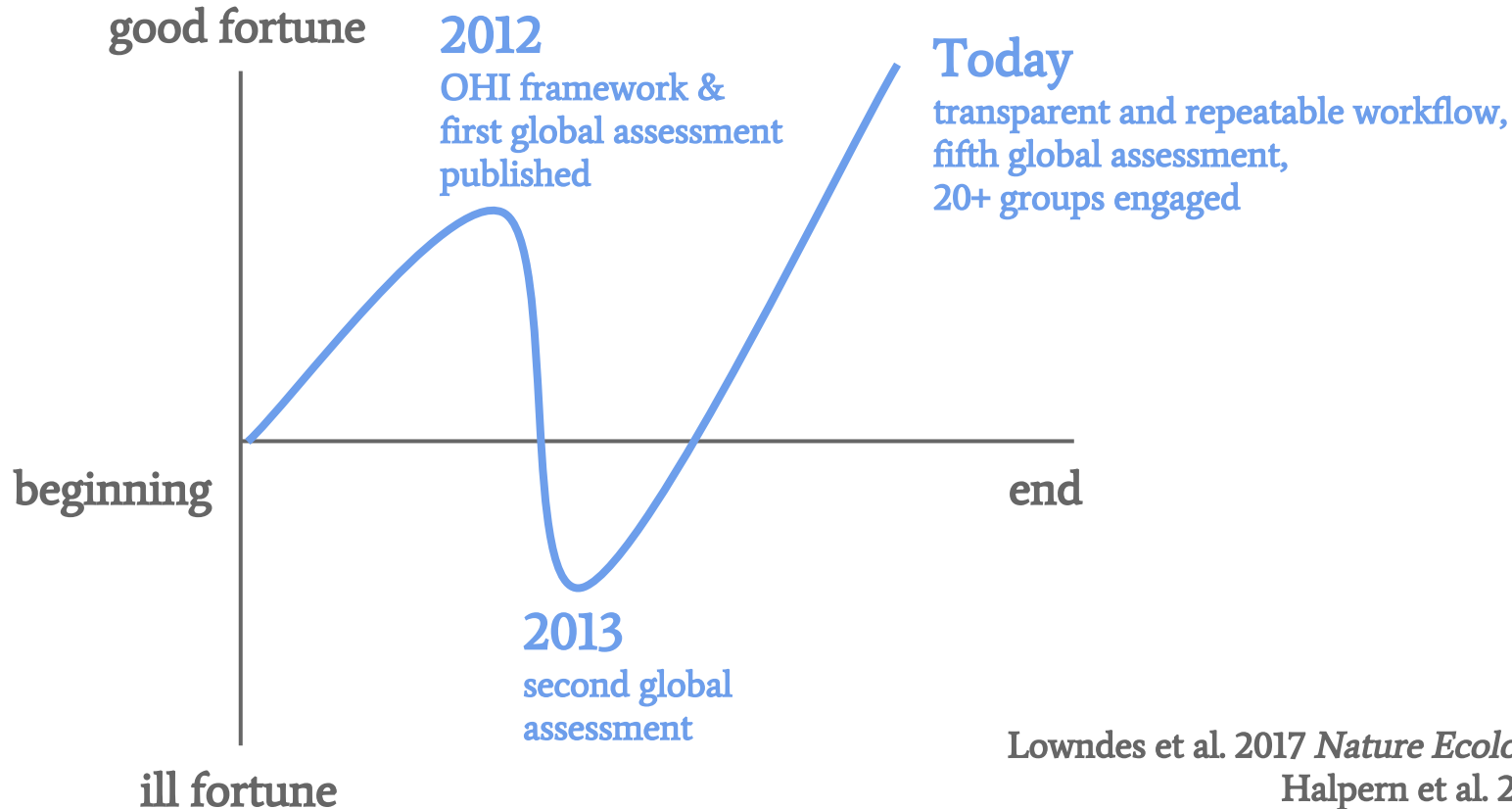
Story arc: The Ocean Health Index



Story arc: The Ocean Health Index



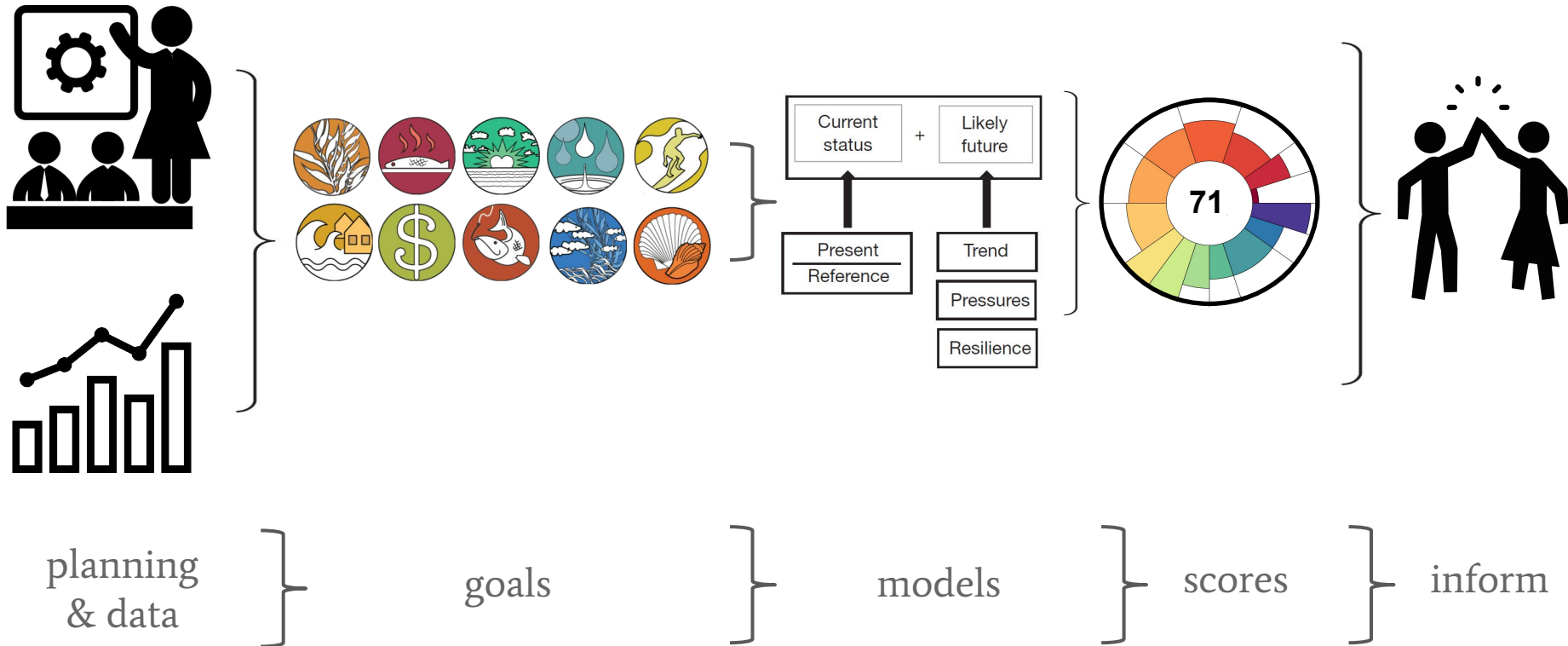
Story arc: The Ocean Health Index



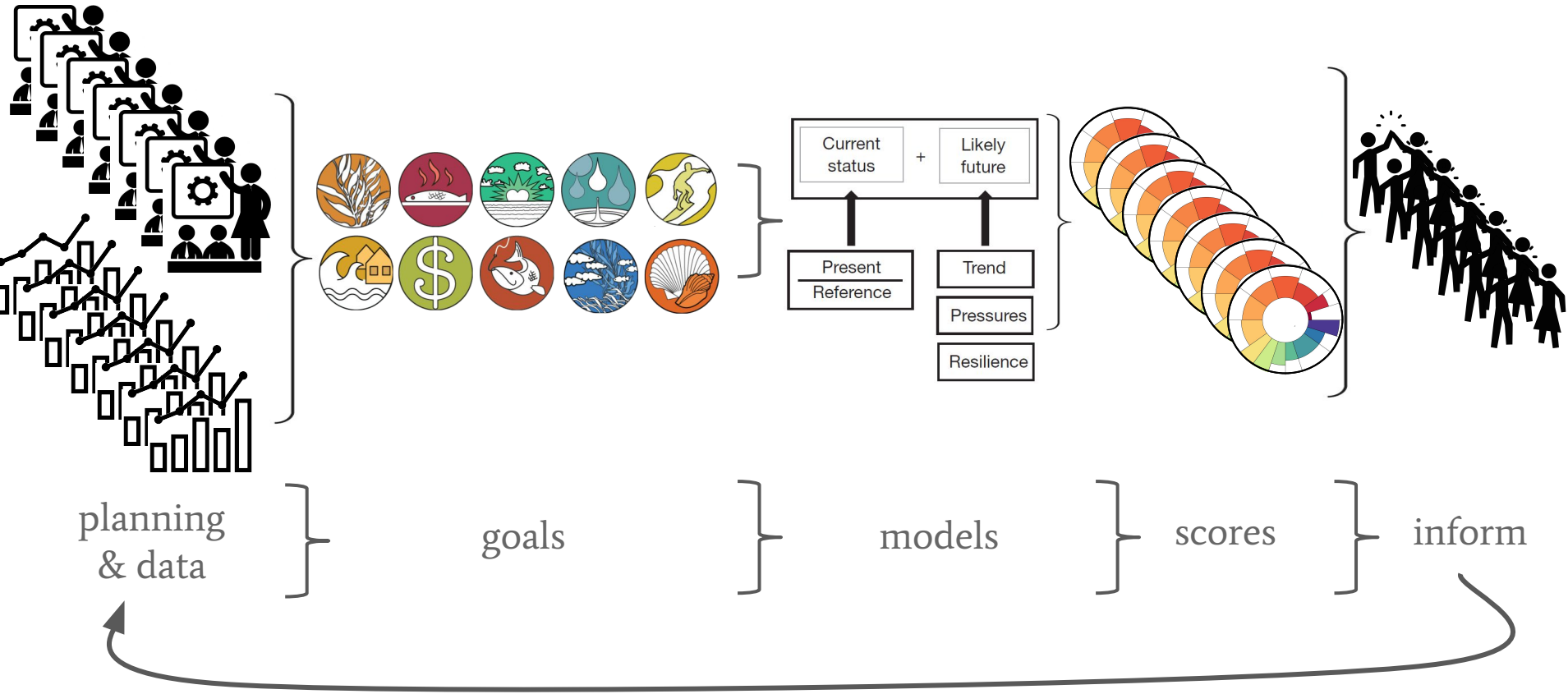
Lowndes et al. 2017 *Nature Ecology & Evolution*
Halpern et al. 2017, *PLOS One*

OHI framework

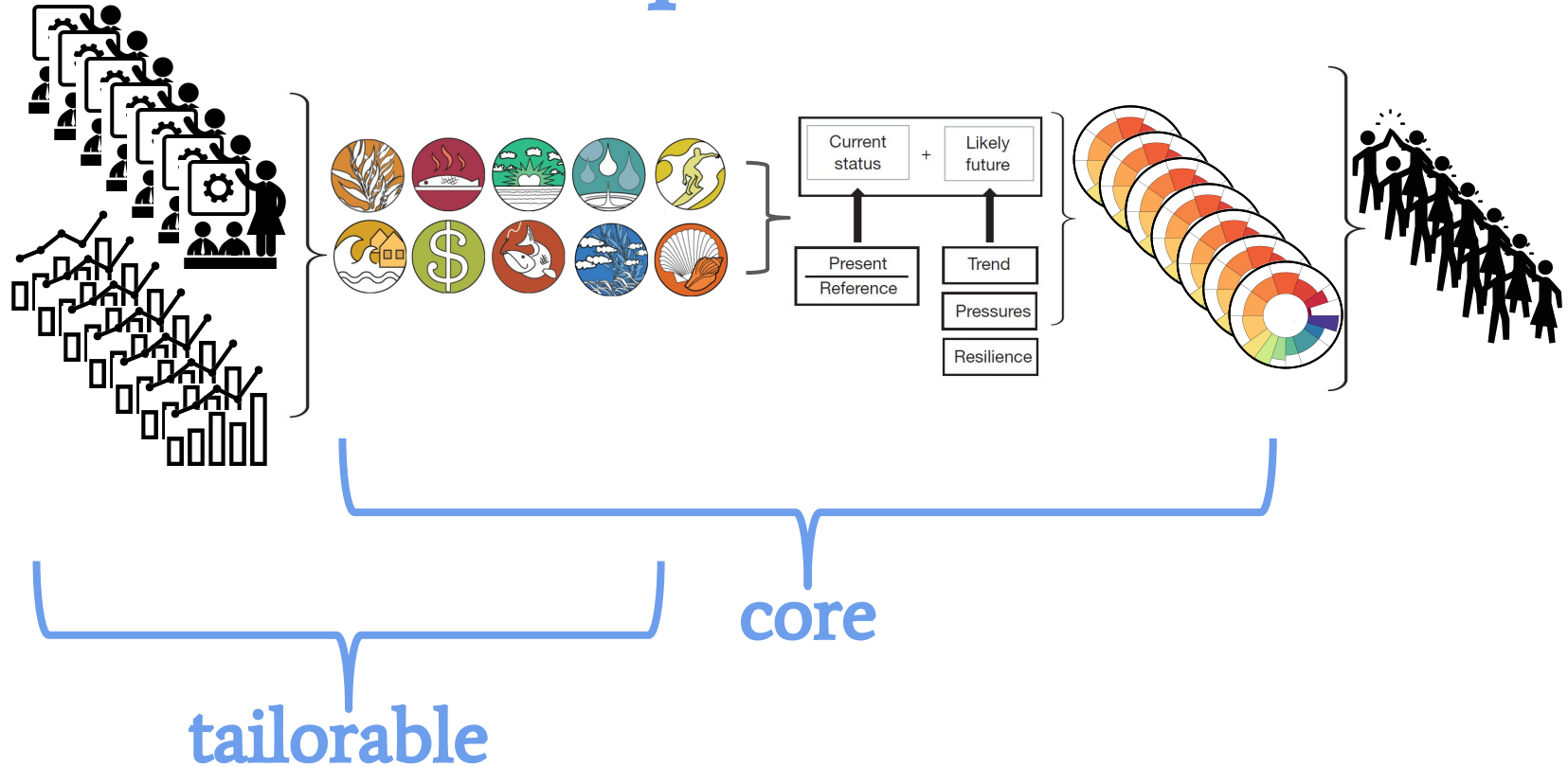
A healthy ocean sustainably delivers a range of benefits to people now and in the future.



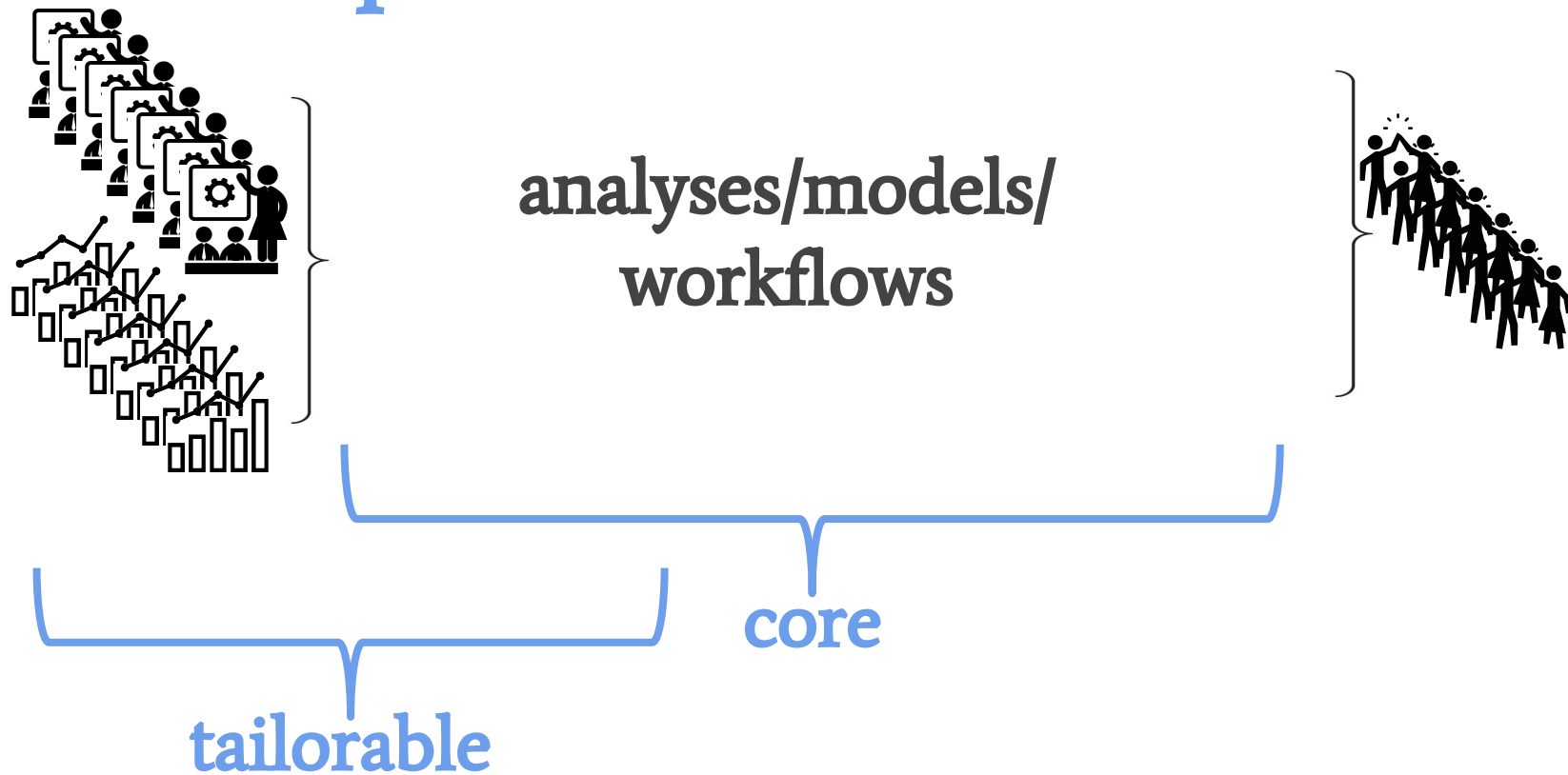
Repeatable OHI assessment process



OHI assessment process



Scientific process





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

Benjamin S. Halpern^{1,2}, Catherine Longo¹, Darren Hardy¹, Karen L. McLeod³, Jameal F. Samhoury⁴, Steven K. Katona⁵, Kristin Kleisner⁶, Sarah E. Lester^{7,8}, Jennifer O'Leary¹, Marla Ranelletti¹, Andrew A. Rosenberg⁵, Courtney Scarborough¹, Elizabeth R. Selig⁵, Benjamin D. Best⁹, Daniel R. Brumbaugh¹⁰, F. Stuart Chapin¹¹, Larry B. Crowder¹², Kendra L. Daly¹³, Scott C. Doney¹⁴, Cristiane Elfes^{15,16}, Michael J. Fogarty¹⁷, Steven D. Gaines⁸, Kelsey I. Jacobsen⁸, Leah Bunce Karrer⁵, Heather M. Leslie¹⁸, Elizabeth Neeley¹⁹, Daniel Pauly⁶, Stephen Polasky²⁰, Bud Ris²¹, Kevin St Martin²², Gregory S. Stone⁵, U. Rashid Sumaila⁶ & Dirk Zeller⁶



data processing



data_v1.xls



data_v1b.xls



data_v2_2012_02_26.xls

collaboration



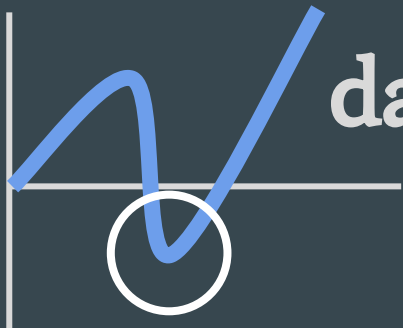
data question



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

tion

question

data_v2_2012_02_26.xls

FWD:Re:data question



data processing

collaboration



Studio®



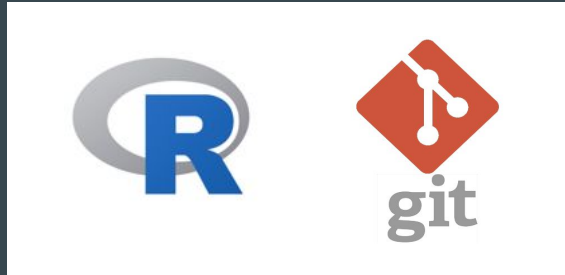
coding – reusable record
version control – bookkeeping

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



data processing

collaboration

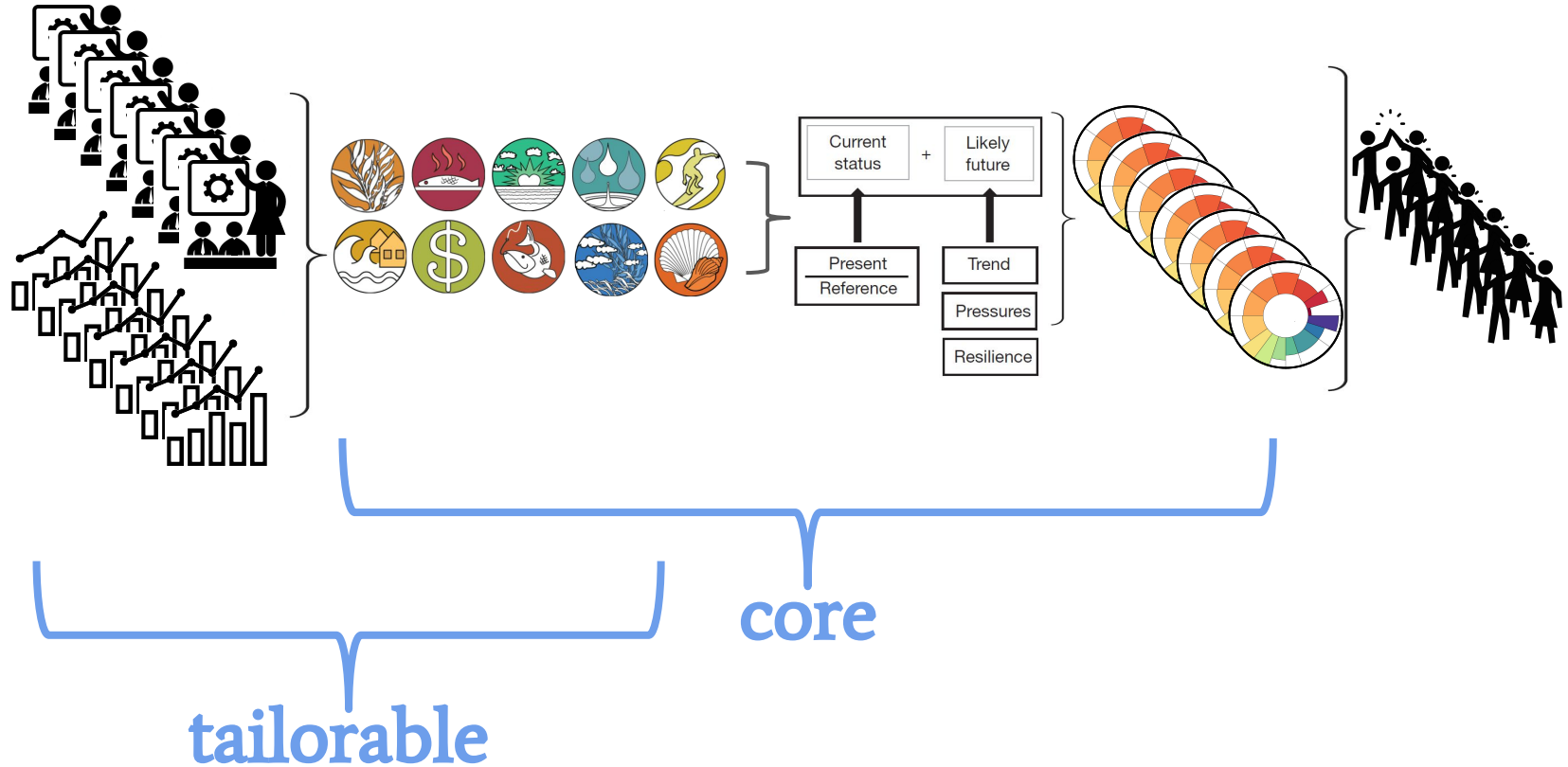


coding – reusable record
version control – bookkeeping

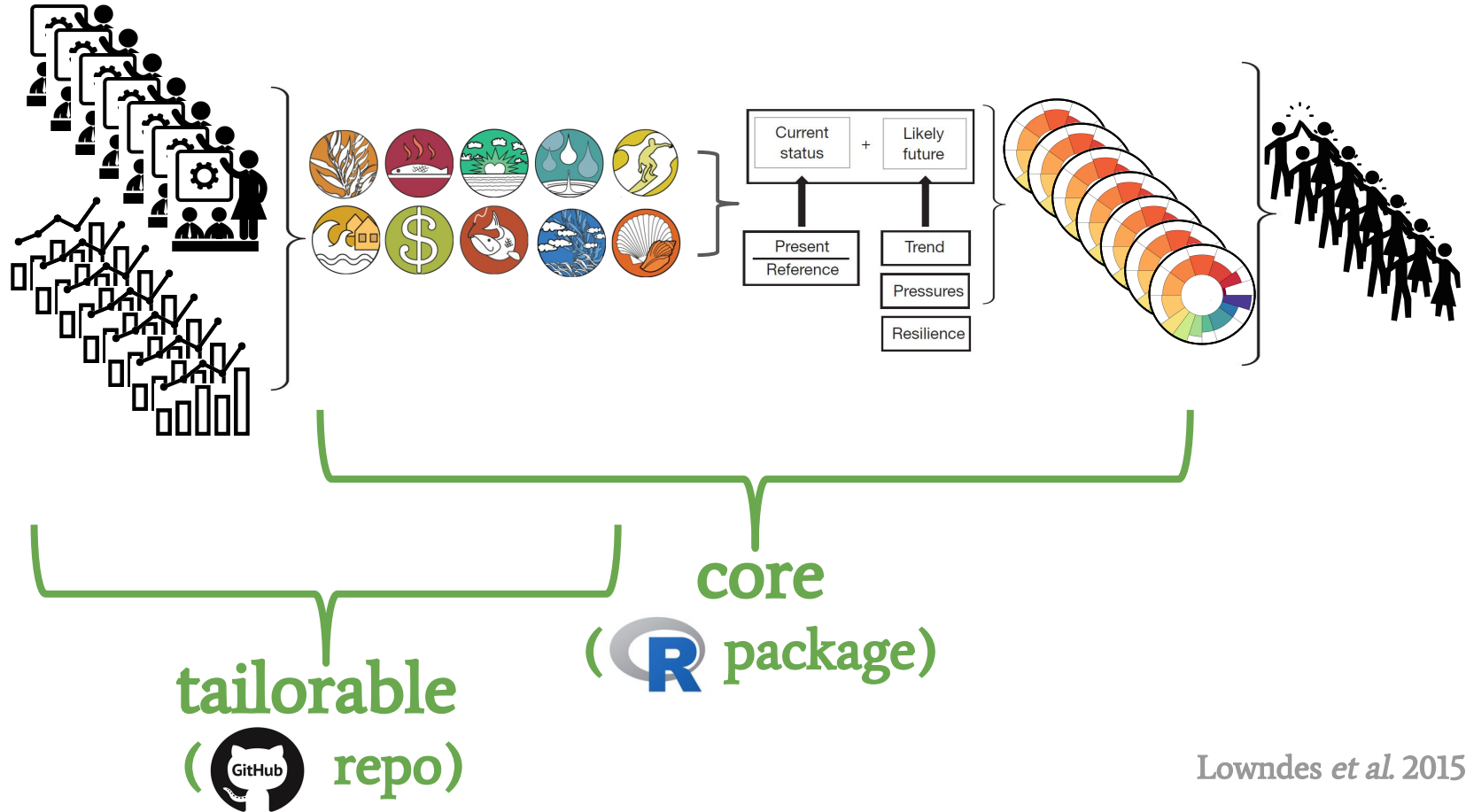
best practices – now + future
shared – websites, online docs

Communication. Open data science.

OHI framework



OHI Toolbox



How we work



Reproducibility

Code
Version control



local computer

How we work

Collaboration

Record keeping
File sharing



Communication

Sharing data,
code, methods



+



online cloud

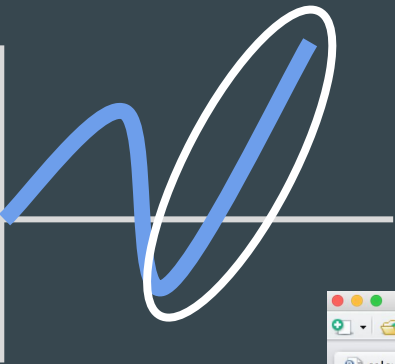


Reproducibility

Code
Version control



local computer



How we work

RStudio – GitHub workflow

The screenshot displays the RStudio environment with the following components:

- Source Editor:** Contains R code for loading and processing data from a CSV file. The code includes comments and uses functions like `bind_rows`, `read_csv`, `dplyr::select`, and `write_csv`.
- Console:** Shows the output of the code execution, including a column specification for the loaded data and a warning message: "There were 50 or more warnings (use warnings() to see the first 50)".
- Environment/History/Build/Git Panel:** Displays a list of files and folders, including `jamie_sf.r`, `circle2016/`, and `circle2016/`. A red callout box points to this panel with the text "GitHub connection, env, build".
- File Browser:** Shows the file structure and provides documentation for the `read_delim` function. A red callout box points to this panel with the text "File nav, help, plots, packages".

R code (scripts and console)

```
28
29 ## regions info
30 regions <- bind_rows(
31   data_frame(
32     region_id = 0,
33     region_name = 'Arctic'),
34   read_csv('spatial/regions_list.csv') %>%
35   dplyr::select(region_id = rgn_id,
36                 region_name = rgn_name))
37
38 ## set figure name
39 regions <- regions %>%
40   mutate(flower_png = sprintf('reports/figures/
41                               str_replac
42   readr::write_csv(regions, 'report
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60:2 Flower plots for each region
```

Console:

```
~/github/arc
name_fl
descrip
weight =
preindex_f
col_character(),
postindex_function = col_character()
)
Parsed with column specification:
cols(
  rgn_id = col_integer(),
  w_fis = col_double()
)
There were 50 or more warnings (use warnings() to see the first 50)
>|
```

Environment/History/Build/Git:

Staged	Status	Path
<input checked="" type="checkbox"/>	A	jamie_sf.r
<input checked="" type="checkbox"/>	M	circle2016/layers
<input checked="" type="checkbox"/>	M	circle2016/
<input checked="" type="checkbox"/>	M	circle2016/

File Browser:

Read a delimited file (including csv & tsv) into a tibble

`read_delim` (readr) R Documentation

Read a delimited file (including csv & tsv) into a tibble

Des

readr::read_delim() reads a delimited file into a tibble. It handles special cases like empty files, files with no header, and files with different line endings. They're the most common types of files. Files with comma separated values are read as delimited values.

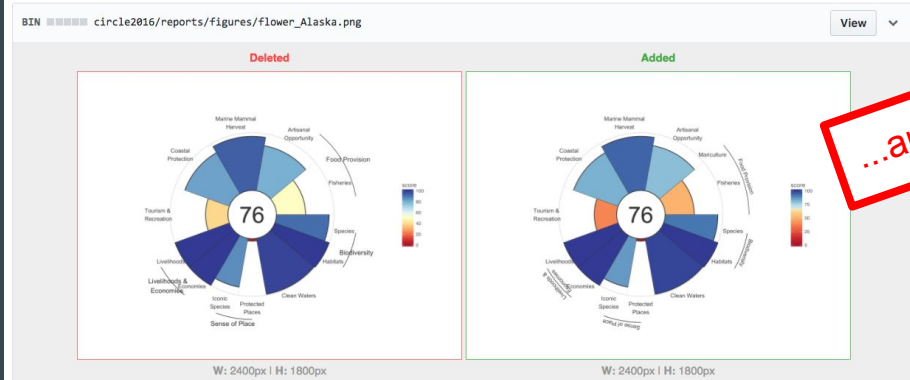


How we work

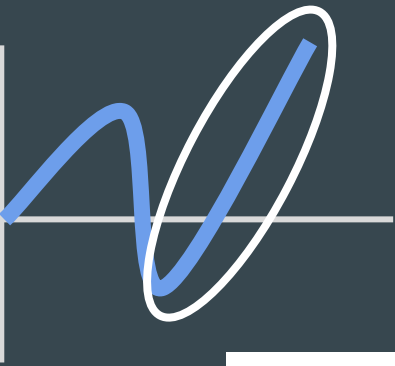
RStudio – GitHub workflow

```
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 %>%
  @@ -201,10 +205,10 @@ PlotFlower <- function(score_df,
201 205 mutate(supra_rad = supra_rad) %>%
202 206 filter(!is.na(name_supra0))
203 207
204 - plot_obj +
208 + plot_obj <- plot_obj +
205 209 geom_text(data = st2,
206 210 inherit.aes = FALSE,
207 - aes(label = st2$name_supra0, x = st2$pos_supra0, y = supra_rad, angle = st2$myAng),
211 + aes(label = name_supra0, x = pos_supra0, y = supra_rad, angle = myAng),
208 212 hjust = .5, vjust = .5,
209 213 size = 3,
210 214 color = dark_line)
  @@
```

See what changed
line-by-line

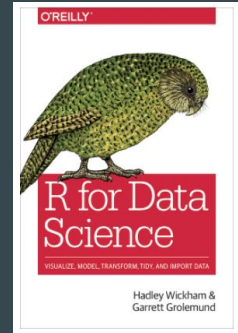
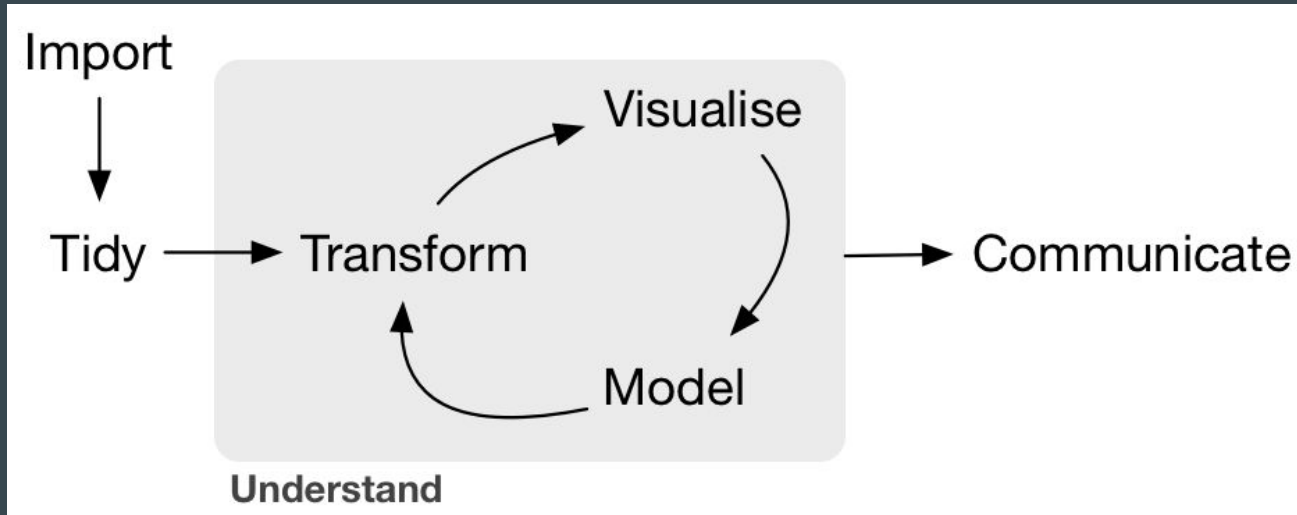


...and plot by plot



How we work

Data science theory & tidy data

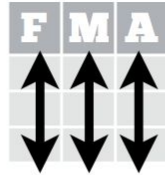




How we work

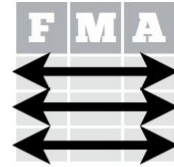
Data science theory & tidy data

In a tidy data set:

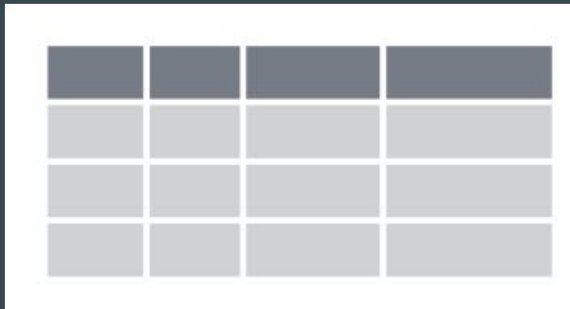


Each **variable** is saved in its own **column**

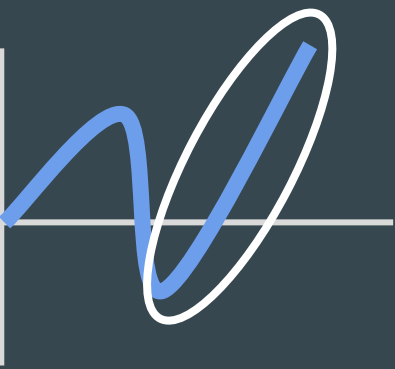
&



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



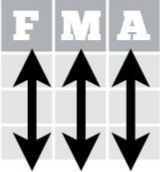
	country	continent	year	lifeExp	pop	gdpPercap
1	Afghanistan	Asia	1952	28.801	8425333	779.4453
2	Afghanistan	Asia	1957	30.332	9240934	820.8530
3	Afghanistan	Asia	1962	31.997	10267083	853.1007
4	Afghanistan	Asia	1967	34.020	11537966	836.1971
5	Afghanistan	Asia	1972	36.088	13079460	739.9811
6	Afghanistan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



How we work

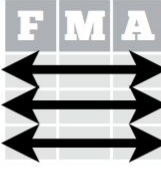
Data science theory & tidy data

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 the RStudio interface. On the left, the R Markdown source file is open, displaying the following 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](https://cran.r-project.org/web/packages/marmap/index.html)
15 package provides tools and data for visualizing the ocean
16 floor. Here is an example contour plot of marmap's "r"
17 params$data dataset.
18
19 ## [r echo = FALSE]
20 data(list = params$data)
21 autoplot(get(params$data))
22 ---
```

On the right, the rendered HTML output is shown, titled "Visualizing the ocean floor". It includes a contour plot of the ocean floor, which is a map of the Hawaiian Islands showing depth contours.

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

The screenshot shows the GitHub repository page for 'Ocean Health Index - Science'. The page header includes the Ocean Health Index logo, the repository name, and the tagline 'Open science with the Ocean Health Index'. Below the header, there are navigation tabs for 'Repositories', 'People 41', 'Teams 18', 'Projects 0', and 'Settings'. A search bar and filters for 'Type: All' and 'Language: All' are present. The main content area displays a list of repositories, each with a title, description, language, star count, and update time. A 'Top languages' section shows 'R', 'HTML', 'JavaScript', 'CSS', and 'Python'. A 'People' section shows a grid of 12 user avatars and an 'Invite someone' button.

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

Global assessment!



Manuscript website!



Training e-book!



OHI+ assessment!



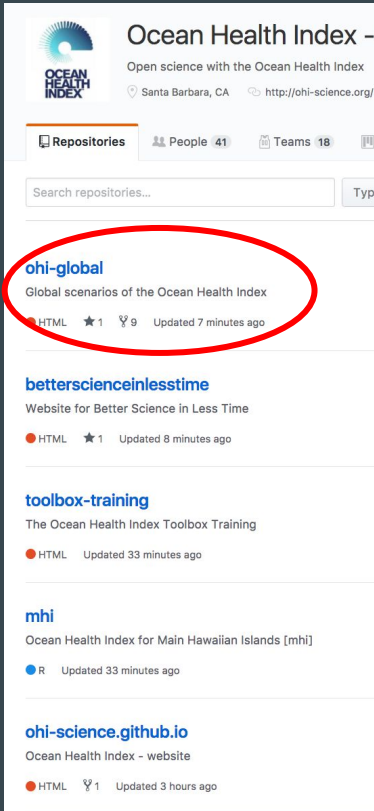
Science website!



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

beterscienceinlesstime
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 Ω_{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 Ω aragonite saturation state.

Methods

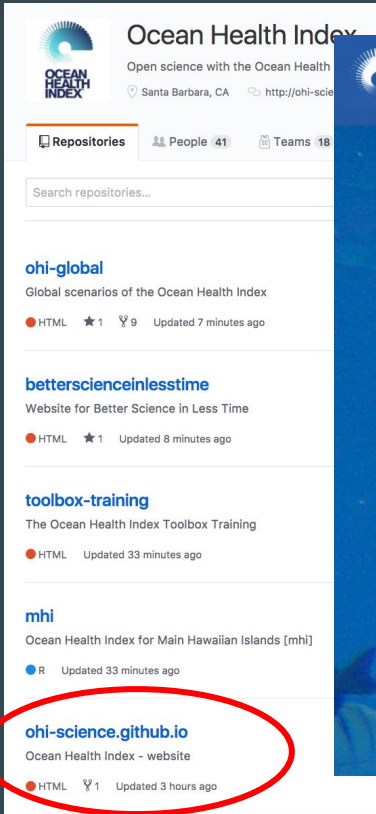
Setup

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

Hide

```
#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")
```

Science website: ohi-science.org

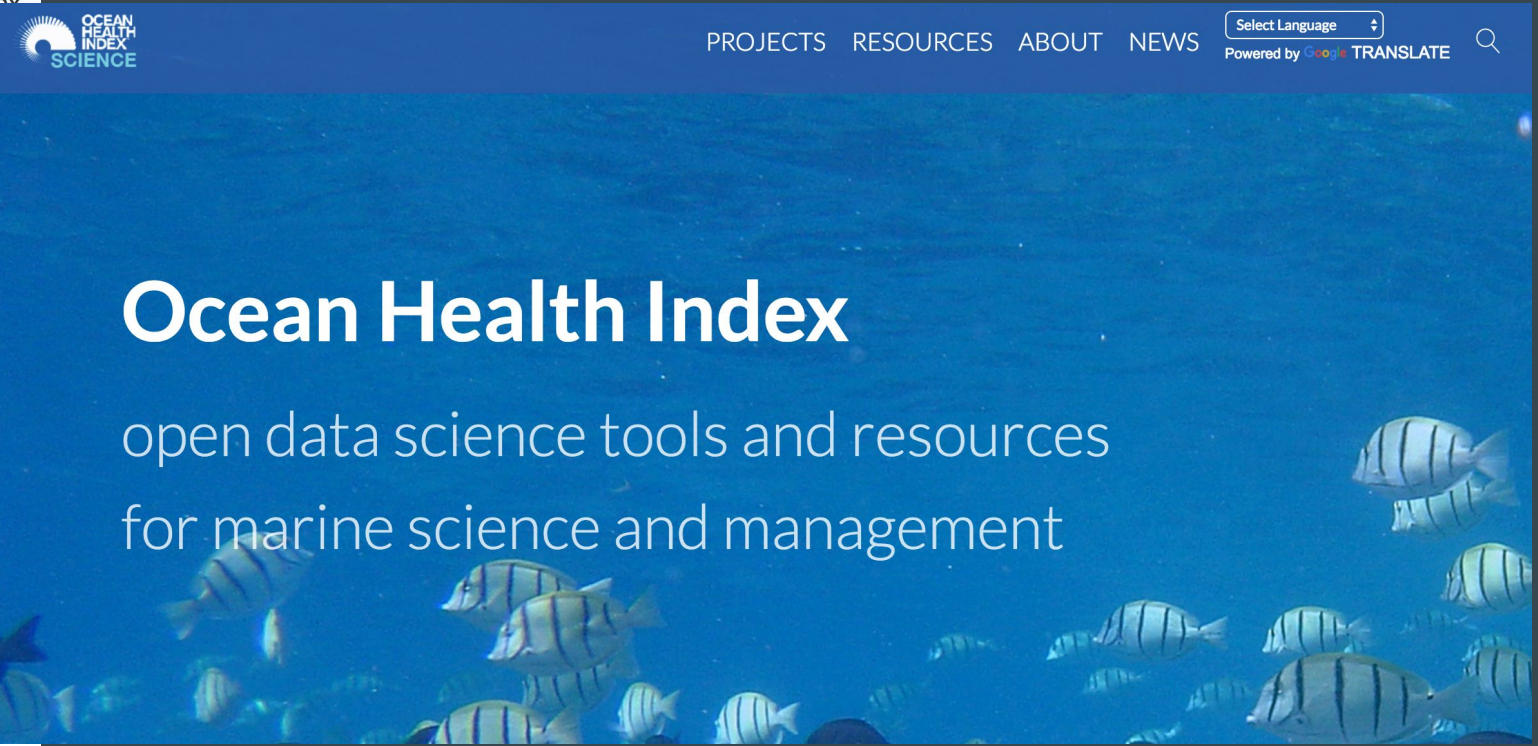


Ocean Health Index
Open science with the Ocean Health Index
Santa Barbara, CA | http://ohi-science.org

Repositories | People 41 | Teams 18

Search repositories...

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



OCEAN HEALTH INDEX SCIENCE

PROJECTS RESOURCES ABOUT NEWS

Select Language [v]
Powered by Google TRANSLATE

Ocean Health Index

open data science tools and resources
for marine science and management

Interactive websites for published articles: 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^{1,2,3*} • Melanie Frazier¹ • Jamie Afflerbach¹ • Casey O'Hara¹ • Steven Katona⁴ • Julia S. Stewart Lowndes¹ • Ning Jiang¹ • Erich Pacheco⁴ • Courtney Scarborough¹ • Johanna Polsenberg⁴

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, Afflerbach J, O'Hara C, Katona S, Lowndes JSS, Jiang N, Pacheco E, Scarborough C, Polsenberg 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 founding 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

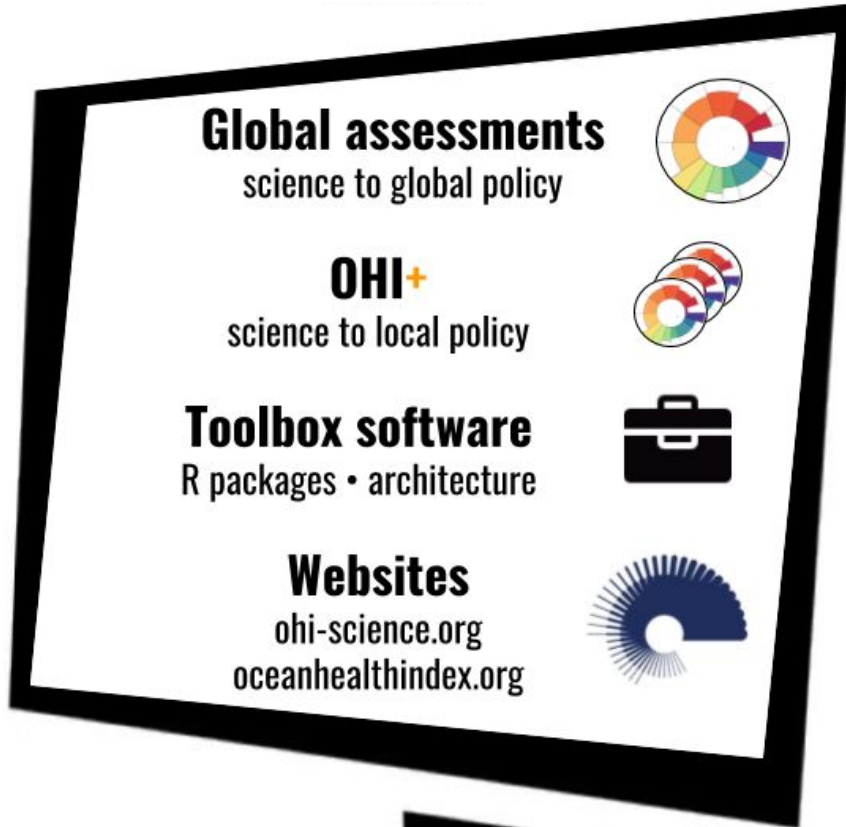
Drivers and implications of change in global ocean health over the past five years

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.

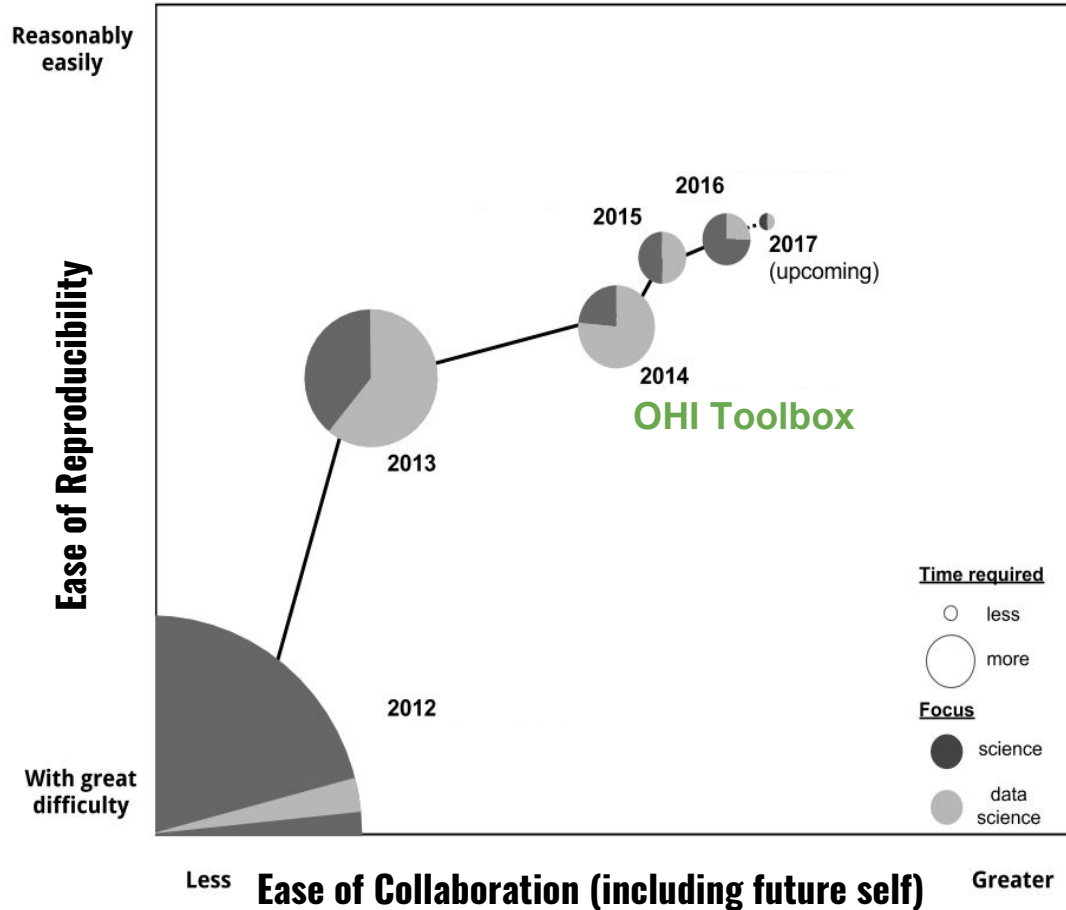
Globally scores have shown little change, as would be expected. However, individual countries have seen notable increases or declines due in particular to improvements in the harvest and management of wild-caught fisheries, the creation of marine protected areas (MPAs), and decreases in natural product 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.

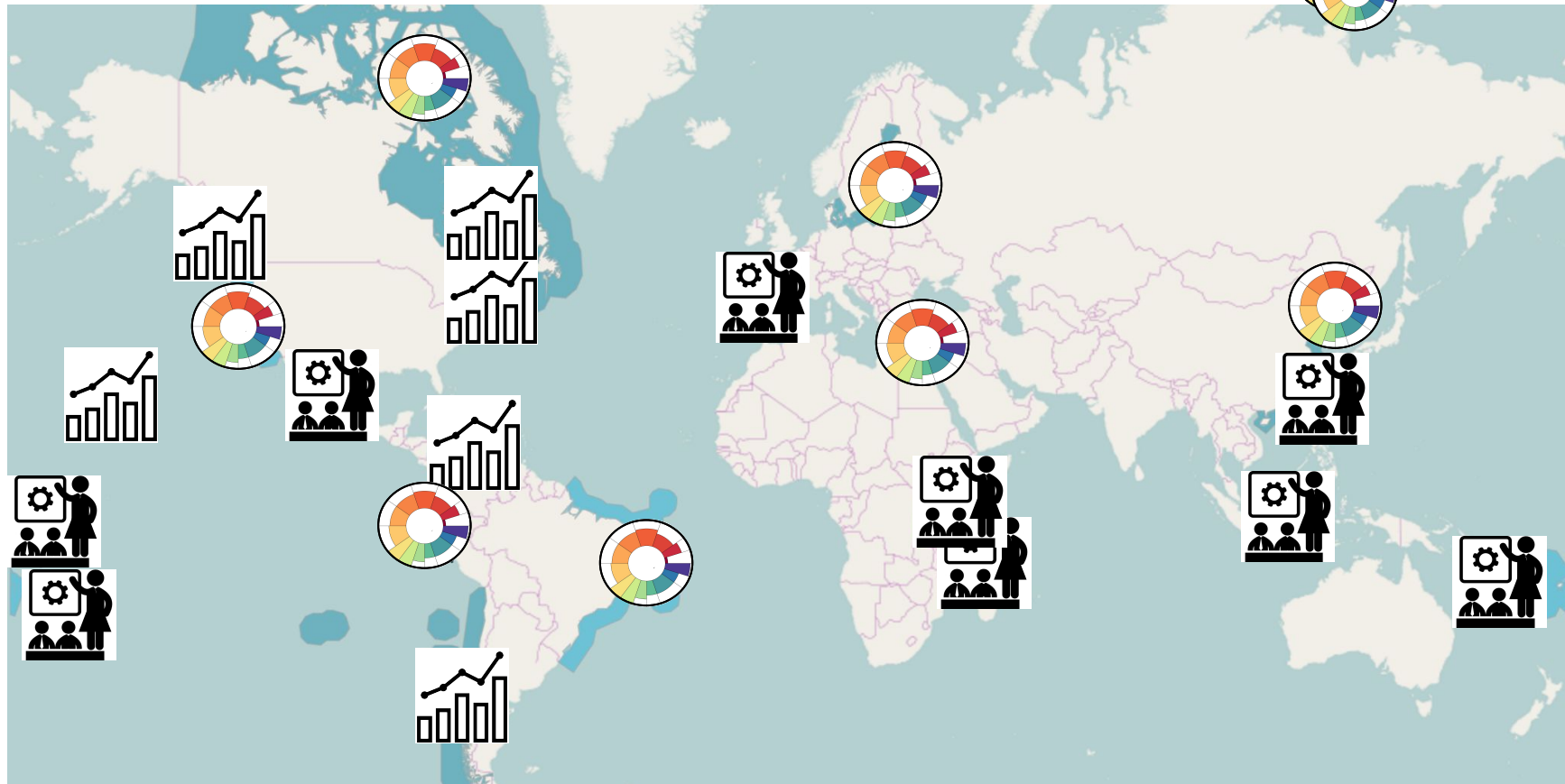
The OHI and open data science



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^{1*}, Benjamin D. Best², Courtney Scarborough¹, Jamie C. Afflerbach¹,
Melanie R. Frazier¹, Casey C. O'Hara¹, Ning Jiang¹ and Benjamin S. Halpern^{1,3,4}

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 // @OHIscience

oceanhealthindex.org // @OceanHealthIndx

Julia Stewart Lowndes

lowndes@nceas.ucsb.edu

twitter: @juliesquid



DataONE Webinar
October 9, 2017



TIDE MILL

Some of the most useful advice

The internet will make those bad words go away



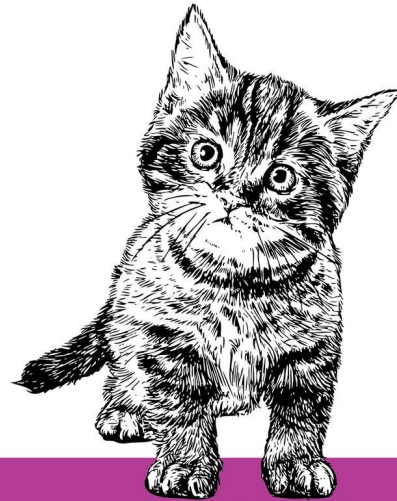
Essential

Googling the
Error Message

○ RLY?

The Practical Developer
@ThePracticalDev

How to actually learn any new programming concept



Essential

Changing Stuff and
Seeing What Happens

○ 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

The image shows a GitHub repository page for 'OHI-Science / mhi' and an RStudio window. The GitHub page has a red circle around the URL 'https://github.com/OHI-Science/mhi' in the browser address bar. The RStudio window has a red circle around the project path '~/.github/mhi - master - RStudio' in the title bar.