Improving the IEA approach using principles of open data science

Kimberly Bastille[1]   Sean Hardison[2]   Lynn DeWitt[3]
Jennifer Brown[4]   Jameal Samhouri[5]   Sarah Gaichas[6]
Sean Lucey[6]   Kelly Kearney[7]   Ben Best[8]   Scott Cross[9]
Scott Large[6]   Ellen Spooner[10]

Abstract

The text of your abstract. 150 − 250 words.

[1] Integrated Statistics, Woods Hole, MA, USA
[2] University of Virginia, Charlottesville, VA, USA
[3] Three
[4] Four
[5] Five
[6] Woods Hole Laboratory, NOAA NMFS Northeast Fisheries Science Center, Woods Hole, MA, USA
[7] University of Washington, Joint Institute for the Study of the Atmosphere and Ocean, Seattle, WA, USA
[8] EcoQuants LLC, Santa Barbara, CA, USA
[9] nine
[10] ten

Introduction

As implemented so far, integrated ecosystem assessments are more often than not bespoke products that are specific to regional policy needs at varying spatial scales. While unique, these assessments are typically built around the flow of the IEA framework outlined in (Levin et al. 2009). This process consists of a scoping step to outline management objectives, the identification of ecosystem indicators to monitor key processes, the compilation of ecosystem status reports highlighting status and trends of indicators, a risk assessment step identifying where ecosystem considerations might threaten management objectives, and finally management strategy evaluation, in which potential management actions are tested using simulation models. The elements of the IEA framework are central to its implementation, but its flow is not prescriptive. Instead, the

IEA process is malleable to the needs of stakeholders, meaning that IEAs may manifest in many different forms.

However, data acquisition, management, communication, and dissemination are universal challenges that apply across the disparate applications of IEAs. Solving these challenges is not trivial, but over the past decade several examples have emerged suggesting the use of open data science tools as potential solutions (Rocchini and Neteler 2012; Lowndes et al. 2015; Lowndes et al. 2017; Ma et al. 2018). For example, Lowndes et al. 2017 discussed how embracing open science methods allowed for increased efficiency, transparency, and reproducibility in the development of the Ocean Health Index; a modular framework developed to assess the benefits of marine ecosystems to humans for sustainable management. Case studies for applying open science processes to IEAs have also been developed for ecosystem reporting in the Northeast Large Marine Ecosystem. Specifically, Ma et al. 2017 show how machine-readable provenance may be incorporated into automated analytical workflows using IPython and the Semantic Web.

Embracing these tools and strategies speaks to a broader philosophy of open science that has yet to fully catch on in the IEA community. Beyond making data publicly available, open science advocates for the "free and unfettered access to all aspects of the scientific endeavor" (Hampton et al. 2015), including methods, data and scientific products. These values lend themselves to the broad applicability of IEAs as decision-making tools for ecosystem-based management (EBM).

There are several entry and exit points for data, methodological, and scientific products throughout the execution of an IEA. Data entry points are formed during the exploratory phases of the IEA process, where objectives are scoped and representative indicators are developed. Exit points for data products can be found in the derived indicator data informing ecosystem status reports and risk assessments. Simulated products and model parameters, such as those resulting from MSE model runs, are also important to disseminate. Any data that leaves or enters an IEA brings with it metadata and code used to analyze, process, and visualize the data (in various stages of completion). Through the process these datasets, accompanying code, and metadata documentation become their own general, scientific and technical products feeding into subsequent products and adding transparency and efficiency (Fig. @ref(fig:iealoop)).

Without an underlying data management protocol, the vast quantities of data and documentation required to successfully execute an IEA can be overwhelming for practitioners and clients both. For example, the development of ecosystem status reports and risk assessments requires considering the management relevance of hundreds of data sets from many fields. In practice this meant sifting through over 400 unique spatial and temporal data sets that were submitted for consideration during the development of the New England State of the Ecosystem Report, an IEA product. However, only indicators of interest to managers and those necessary for storytelling were included in the final product; just over 90 (Gaichas, Sarah and Hardison, Sean and Large, Scott and Lucey, Sean 2019). Depending

Evaluate Strategies
Management Strategy Evaluation, Summary products, Web-based visualizations, Model inputs, outputs and parameters, Analysis/Visualization code, R packages*

Analyze Uncertainty & Risk
Risk Assessments, Methods documentation, Analysis/Visualization code

Develop Indicators
Indicator data sets, Web-based visualization apps, Metadata documentation, R packages/Shared R scripts

Assess Ecosystem
Ecosystem Status Reports, Report templates, Methods documentation, Analysis/Visualization code

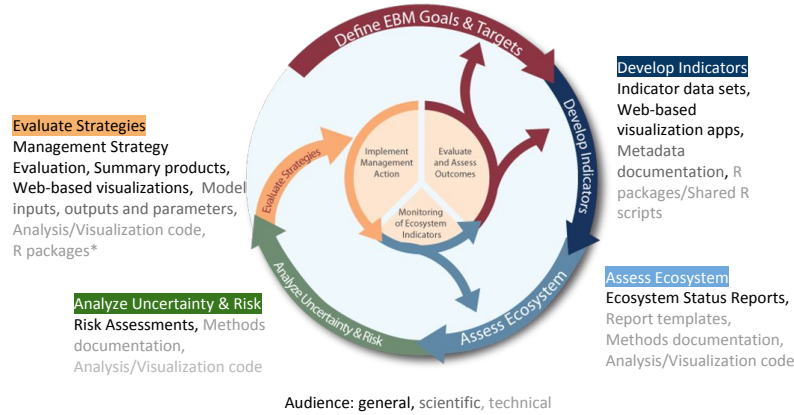Audience: general, scientific, technical

Figure 1: Products as part of the IEA loop according to audience: general, scientific and technical.

on the governing body implementing the IEA, each data set must be fully documented with metadata and made publicly available. Versioning data sets and associated metadata relative to specific IEA products also introduces challenges for scientists operating outside the bounds of their training. These hurdles result in data processing and documentation becoming a full time job, and indeed, it is becoming increasingly common for IEA teams to hire Data Analysts and Data Scientists to deal with the data deluge.

The implementation of IEA is an amalgam of the ecosystem and data sciences. Understanding this, we suggest that IEA teams embrace a philosophy of open science. This approach has gained momentum over the past decade within the ecological sciences as a way to deal with large amounts of data produced from a variety of sources (Michener and Jones 2012), as well as to improve transparency, reproducibility, repeatability, and ease of communication of data products (Rocchini and Neteler 2012; Lowndes et al. 2015; Lowndes et al. 2017; Ma et al. 2018). This can be accomplished using openly available tools, many of which are highlighted by Hampton et al. (2015) and Lowndes et al. (2017).

The benefits of implementing software development strategies range from the ethical to the practical. For example, Lowndes et al. (2017) suggests that changing ecosystem conditions due to anthropogenic impacts necessitate openness among environmental scientists whose data (and code to recreate their data) provide "snapshots" of their study systems amidst the changes. Further, openly-sourcing scientific data with accompanying code provides a record from which previous results may be built upon as more data are collected (Lowndes et al. 2017). From this perspective, iterative IEA products such as ecosystem reports

(a) Process
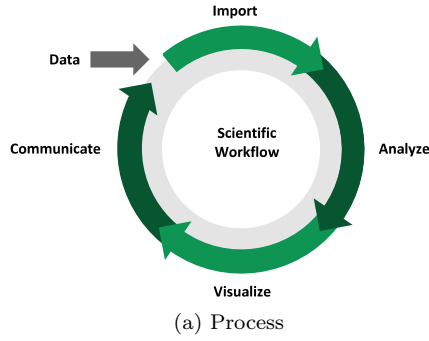
Function	Implementation	Tool
Import	ecodata	R package
	Data portals	ERDDAP
Analyze	lm to ecosystem modelling	Shared R scripts
	ecotrend	R package
	Ecosystem modelling	Ecopath with Ecosim
Visualize	Standardized plotting	Shared R scripts
	Web-based visuals	Shiny
	Simple plots	Excel
Communicate	Report writing, article writing, methods documentation	Rmarkdown
	Continuous integration	Travis
	Versioning software	Git
	Collaborative package/document development and data sharing	Github
	Data portals	ERDDAP

(b) Elements

Figure 2: Scientific workflow.

and risk assessments are themselves data points; forming ever-extending arrays of compiled information through time. This approach has been embraced by the Ocean Health Index, which is designed to monitor human-ocean relationships through time by repeated compilation and analysis of ecosystem data (Halpern et al. 2012; Halpern et al. 2015; Lowndes et al. 2015).

We suggest that the incorporation of version control and standardized methods for the collection, aggregation, and dissemination of IEA-oriented data and products can facilitate flexible and efficient use of IEAs, broadening their applicability and hastening their uptake by resource managers. These methods also confront the challenging spatial aspects of IEAs, which are inherently unique to the systems being addressed. As we discuss below, embracing standardized methods and open data science principles/tools would facilitate the implementation of future IEAs in new systems.

References

Gaichas, Sarah and Hardison, Sean and Large, Scott and Lucey, Sean. 2019. State of the ecosystem 2019: New england. 166 Water Street, Woods Hole, MA 02543-1026: National Marine Fisheries Service, NEFSC.

Halpern, Benjamin S., Catherine Longo, Darren Hardy, Karen L. McLeod,

Jameal F. Samhouri, Steven K. Katona, Kristin Kleisner, et al. 2012. An index to assess the health and benefits of the global ocean. *Nature.* https://doi.org/10.1038/nature11397.

Halpern, Benjamin S., Catherine Longo, Julia S. Stewart Lowndes, Benjamin D. Best, Melanie Frazier, Steven K. Katona, Kristin M. Kleisner, Andrew A. Rosenberg, Courtney Scarborough, and Elizabeth R. Selig. 2015. Patterns and Emerging Trends in Global Ocean Health. *PLoS ONE* 10: e0117863. https://doi.org/10.1371/journal.pone.0117863.

Hampton, Stephanie E, Sean S Anderson, Sarah C Bagby, Corinna Gries, Xueying Han, Edmund M Hart, Matthew B Jones, et al. 2015. The tao of open science for ecology. *Ecosphere* 6. Wiley Online Library: 1–13.

Levin, Phillip S, Michael J Fogarty, Steven A Murawski, and David Fluharty. 2009. Integrated ecosystem assessments: Developing the scientific basis for ecosystem-based management of the ocean. *PLOS Biology* 7. Public Library of Science: 1–6. https://doi.org/10.1371/journal.pbio.1000014.

Lowndes, Julia S Stewart, Benjamin D Best, Courtney Scarborough, Jamie C Afflerbach, Melanie R Frazier, Casey C O'Hara, Ning Jiang, and Benjamin S Halpern. 2017. Our path to better science in less time using open data science tools. *Nature ecology & evolution* 1. Nature Publishing Group: 0160.

Lowndes, Julia S Stewart, Erich J Pacheco, Benjamin D Best, Courtney Scarborough, Catherine Longo, Steven K Katona, and Benjamin S Halpern. 2015. Best practices for assessing ocean health in multiple contexts using tailorable frameworks. *PeerJ* 3. PeerJ Inc.: e1503.

Ma, Xiaogang, Stace E Beaulieu, Linyun Fu, Peter Fox, Massimo Di Stefano, and Patrick West. 2018. Documenting provenance for reproducible marine ecosystem assessment in open science. In *Information retrieval and management: Concepts, methodologies, tools, and applications*, 1051–1077. IGI Global.

Michener, William K, and Matthew B Jones. 2012. Ecoinformatics: Supporting ecology as a data-intensive science. *Trends in ecology & evolution* 27. Elsevier: 85–93.

Rocchini, Duccio, and Markus Neteler. 2012. Let the four freedoms paradigm apply to ecology. *Trends in ecology & evolution* 27. Elsevier: 310–311.