A case for applying open science principles to bioassessment

Marcus W. Beck ([marcusb@sccwrp.org](mailto:marcusb@sccwrp.org)), Raphael D. Mazor ([raphaelm@sccwrp.org](mailto:raphaelm@sccwrp.org)), Susanna T. Theroux ([susannat@sccwrp.org](mailto:susannat@sccwrp.org))

Version Date: Fri Aug 31 15:03:49 2018 -0700

# Abstract

Progress in the field of applied bioassessment is impeded by a lack of methods that are accessible and reproducible for the management community. Open science principles that seek to democratize science can address these challenges, yet widespread adoption in research has yet to gain traction for the development and appplication bioassessment methods. At the core of this philosophy is the concept that research should be reproducible and transparent, in addition to having long-term provenance through effective modes of data preservation and sharing. This review will introduce core open science concepts that have been advocated more generally in the ecological sciences and will emphasize how adoption can benefit bioassessment for both prescriptive condition assessments and proactive applications that inform planning activities. Examples from the state of California will be used to demonstrate effective adoption of open science principles through data stewardship, reproducible research, and engagement of stakeholders with multimedia applications. Technical, sociocultural, and institutional challenges for adopting open science will also be discussed, including practical approaches for overcoming these hurdles in bioassessment applications.

# Introduction

Bioassessment is an essential element of environmental monitoring programs that informs decisions for managing aquatic resources. Decades of research have supported the development of methods that use a variety of assemblages with regional applications in streams, rivers, lakes, and marine environments. This body of applied tools represents significant achievements in developing accurate and interpretable assessment methods that rely on biological organisms as sentinels of environmental condition. Monitoring programs in the United States and internationally have collected millions of records of biological data spanning decades and hundreds of assessment methods have been developed from these data, yet the ability of these tools to reach the management community and to positively affect environmental change is imbalanced relative to the amount of information that is available through research and coordinated monitoring efforts. Morever, existing methods can lack transparency, require specialized training to implement, and are often not discoverable beyond specific research applications. Decision-makers require additional tools that synthesize information and bridge the gap between method and application.

The science of bioassessment has for decades focused on addressing technical challenges for developing indices that accurately describe environmental condition. Legal mandates to assess biological condition have set a precedent for developing bioassessment methods in the United States (Clean Water Act), Canada (Canada Waters Act), and Europe (Water Framework Directive). Basic research to address broad legislative objectives has historically focused on identifying which biological components respond predictably to environmental change, how these components can be measured with minimum uncerainty, and what basis of comparison is used to evaluate relative changes between communities. Most bioassessment indices developed at the assemblage-level are characterized as either multimetric, such as the index of biotic integrity, or multivariate where condition is assessed using ratios of observed and expected taxa. The reference-condition approach also establishes the foundation for many bioassessment methods whereby a set of reference sites are identified and used to evaluate levels of biological deviation to define potential impacts.

Methods for evaluating and comparing biological communities that have been developed by the research community have generally been accepted by environmental managers as robust for developing bioassessment tools. As a result, new assessment tools can be developed using readily available technical support documents that summarize the body of research and best practices to date on bioassessment. This has in part contributed to the proliferation of hundreds of assessment methods that have been developed for specific regional applcations. Although there are logistical and ecological rationale for why location- and taxa-specific methods are needed, concerns about redundancy, duplicated effort, and lack of coordinated monitoring have recently been discussed within the research community. Morever, the abundance of available methods can be a point of frustration for managers given a lack of guidance for choosing an appropriate method among alternatives. The process to charactize how an index could be used in practice to inform decisions and prioritize management actions is often opaque relative to why an index may have originally been developed.

What distinguishes bioassessment from others field of ecological study is the link to environmental management. Although bioassessment methods can and have been used to inform basic research, the intended use of these tools, often through legal definition, is in the applied sense. An argument could be made that an index is only as valuable as its level of integration with management and regulatory communities. In the United States, the CWA gives the power to states, tribes, and territories to develop their own methods, which in turn require federal approval to be implemented into a regulatory framework, e.g., TMDL reporting, permitting, etc. If federal approval is a rough assessment of index efficacy, a tremendous imbalance exists between the methods developed and those that are federally approved for regulatory use (see [here](https://www.epa.gov/wqc/information-bioassessment-and-biocriteria-programs-streams-and-wadeable-rivers) for EPA list). Of those that are approved, even more concerning is the manner of application within standard regulatory frameworks. Biological indices are typically used to develop post-hoc diagnoses that can trigger remediative or restoration actions. A growing concern is that these tools, although technically sound, are primarliy being used to document the long-term demise of environmental health. A much broader use for bioassessment to pro-actively guide planning decisions, such as identifying conservation priorities, could greatly extend the reach of tools that have already been developed.

Bioassessment currently suffers from an excess of information and forward progress will not be made unless this information meets the needs of the management community. A new mode of operation is needed whereby method development is open and transparent, existing methods are discoverable and reproducible, and information transfer to the management community is intuitive and purposeful. Open science principles that can democratize all aspects of the scientific method can meet these needs, yet bioassessment research and its application to better serve the environment has not fully embraced these principles. Others have advocated more broadly for inclusion of these principles in the ecological sciences (Hampton et al. 2015) and a growing wave of momentum has seen open science permeate how scientists conceptualize research in other disciplines (e.g., archeaology, behavioral ecology, vegetation sciences). Adopting an open science paradigm in biaossessment is particularly relevant compared to other fields given the explicit need to develop tools that are open and accessible to the management community. Legal and even ethical precedents in bioassessment may necessitate the open sharing of data given that environmental monitoring programs are often publicly funded.

This review will empower the research and management community to embrace open science as a new mode of thinking for bioassessment. These approaches are expected to benefit the research community by augmenting existing workflows for developing assessment tools, but more importantly, improve the ability of these methods to address environmental issues by bridging the gap between the scientific, management, and regulatory communities. An overview of the general principles of open science is provided, followed by a discussion of specific benefits and how these principles can be applied to bioassessment. We use examples from the state of California to demonstrate a real world scenario of how existing tools can be tailored to address legislative mandates for free and open sharing of data. We conclude with a discussion of technical, sociocultural, and instutional hurdles that have thus far prevented widespread adoption of open science and provide recommendations for the bioassessment community to address these challenges.

* What is the problem
  + Overall, management of water quality requires science that has been publicly funded and the application of the science is a public service that should be inherently open - but it typically is not.
  + Proliferation of methods - review of index coverage in US, internationally, Birk et al. 2012 describe nearly 300 methods that have been developed in Europe, Nichols et al. 2016 describe national down-scaling of bioassessment applications in Australia as a negative in absence of coordinated federal assessment networks, Kelly et al. 2016 describe issues of redundancy in ecological assessment of lakes regarding use of multiple taxa, are they all necessary?
  + Report/manuscript as final product paradigm and issues related to transparency, reproducibility, data provenance
  + Application requires specialized knowledge of an index, responsibility usually falls on one or two individuals
  + Lack of access to index calibration/validation data, information that is often collected through public funds although often treated as proprietary, data are not always discoverable (Hering et al. 2010 describe this issue in a ten-year assessment of WFD)
  + Linking a method to management requires synthesis of information and intuitive tools for applying/interpreting results
* What is open science
  + A philosophy and set of tools that can democratize scientific analysis by making data and analyses more accessible
  + Emphasis on reproducibility, transparency, communication, and longevity, researchers as data stewards not owners
  + Embraces all aspects of a project from idea conception to delivery of final products, implications for bioassessment
  + Overall, encourages collaboration and access to/sharing of data
* Objectives of this paper: Promote use of open science as a philosophy and set of tools for bioassessment applications by describing benefits that facilitate collaboration and stakeholder engagement, focus is on both the why and how to effectively advocate for and empower others to adopt open science

# Principles of open science and what they mean for bioassessment

* Overview of the open science process – follow Hampton paradigm, distinguish between benefits for the researcher vs research institution vs stakeholder/managers
* Why is open science particularly relevant for bioassessment?
  + data are publicly funded, typically, so legal/moral mandates necessitate open analysis, see Molloy 2011 for UK example
  + Multiple methods have been developed, an important component of open science is data discovery. Making methods open and transparent can facilitate synthesis and meta-analysis
  + Data used for bioassessment methods are typically not the “long-tail” of the ecological sciences, i.e., the carefully collected observational data meant to address specific research questions. Scientists in the long-tail are potentially more relucant to adopt open-science because of the perception of less benefit to making the data open. This suggests that bioassessment datasets and associated methods are inherently more likely to benefit from openness because more widespread appeal. Conversely, the long-tail datasets individually may not have broad relevance but collectively could serve larger purposes, some countries have abandonated national-scale coordinated monitoring efforts in favor localized sampling (Nichols et al. 2016)
* Aspects of the process that can benefit bioassessment
  + Data provenance and open data
  + Method development – existing software packages to facilitate
  + Method delivery –
    - portable packages and data visualization, emphasis on interactive online tools: Zastrow 2015 describes power of interactive mapping, Kelling et al. 2009 describe data-intensive approaches for biodiveristy to identify patterns born from the data and data viz is one approach, Fox and Hendler 2011 emphasize viz as part of analysis process not as end-product
    - Communication within collaborative teadm and management community, Kelling et al. 2009
* Why is open science particularly important for bioassessment? Vs. general ecological research? Vs. other kinds of environmental monitoring? Vs. other publicly funded data collection?

# California examples

* Example approach
  + What is the legal/policy framework for supporting/impeding open science in CA? Are we living up to our aspirations?
    - On July 10, 2018 the The State Water Resources Control Board “adopted a resolution on open data principles committing it and the Regional Water Boards to providing broader access to the data used to make local, regional and statewide water management and regulatory decisions in California.” [press release](https://www.waterboards.ca.gov/press_room/press_releases/2018/pr_water_data_071018.pdf), [resolution](https://www.waterboards.ca.gov/board_info/agendas/2018/jul/071018_5_drft_reso.pdf)
  + AB 1755, Dodd. The Open and Transparent Water Data Act. Passed in 2016, requires state water quality institutions to “create, operate, and maintain a statewide integrated water data platform that, among other things, would integrate existing water and ecological data information from multiple databases and provide data on completed water transfers and exchanges” and “develop protocols for data sharing, documentation, quality control, public access, and promotion of open-source platforms and decision support tools related to water data”
  + The California vision – describe legal/policy demands for bioassessment, current methods developed, developing tools to link technical products with management
  + Existing applications – assessment methods packaged as standalone applications complete with documentation, vignettes, versioning
  + Bioassessment as proactive vs reactive – SCAPE for regulatory applications, SCAPE for conservation, other examples

# Challenges and recommendations for bioassessment

* Challenges for application
  + Technical hurdles – technical and constantly expanding skillset is required, immediate returns difficult to see (e.g., for data sharing Hampton et al. 2015, need to find citation for learnign tech skills)
  + Sociocultural hurdles – unwillingness to share hard-earned data (less so for bioassesssment than traditional ecology, but could be an issue), vulnerability to criticism (Lewandowsky and Bishop 2016 describe concerns of transparency leading to damage of scientific integrity)
  + Institutional barriers – entrenched modes of operation can discourage novelty and exploration, no incentive for adoption
* The way forward
  + The holy grail is widespread adoption of open science in bioassessment, but this will never be completely integrated, see challenges above
  + Teaching as an approach – let the trainee become the trainer, Hampton et al. 2017 describes training initiatives to close the skill-transfer gap, Touchon and McCoy 2016 describe mismatch between grad programs and tech skills used in contemporary ecological analysis
  + Who is likely to adopt? Cultivate adopters (researcher benefits, institution benefits, stakeholder benefits), work with non-adopters (institution benefits, stakeholder benefits)
  + Roles for adopters, roles for non-adopters
  + Development as an approach – roles for adopters, develop specialized software packages (require vetting, Borregard et al. 2016, could link in to new but existing pathways for review such as ROpenSci or peer review journals like the R Journal), Touchon and McCoy 2016 advocate for a role of adopters as specialists to facilitate collaboration with “less quantitatively trained or interested students” rather than the latter analysizing their data in potentially suboptimal ways.
* Call to implement now - field is transitioning to molecular approaches where information acquisition will be orders of magnitude greater than traditional taxonomic-based approaches. Data acquisition and management will require systematic methods for documenting, cataloging, and sharing information - start now. Use of online eDNA archives have been established. Baird and Hajibabaei 2012 describe the bioassessment paradigm with molecular approaches

# References