UC Davis
San Francisco Estuary and Watershed Science

Title
An Open Data Framework for the San Francisco Estuary

Permalink
https://escholarship.org/uc/item/02q969kt

Journal
San Francisco Estuary and Watershed Science, 18(2)

Authors
Baerwald, Melinda R.
Davis, Brittany E.
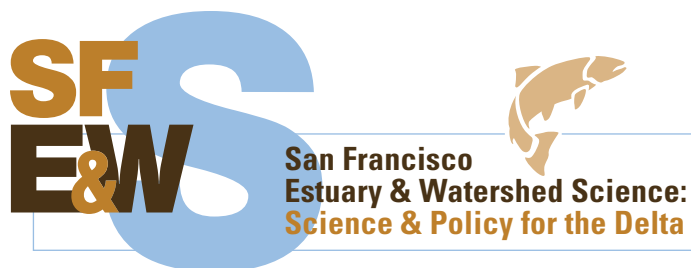Lesmeister, Sarah
et al.

Publication Date
2020

DOI
10.15447/sfews.2020v18iss2art1

Copyright Information

Peer reviewed

An Open Data Framework for the San Francisco Estuary

Melinda R. Baerwald,*[1] Brittany E. Davis,[1] Sarah Lesmeister,[1] Brian Mahardja,[2] Rachel Pisor,[1] Jenna Rinde,[1] Brian Schreier,[1] Vanessa Tobias[2]

INTRODUCTION

Transparency and reproducibility are cornerstones of the scientific method (Nosek et al. 2015), yet, in practice, scientific journal articles with readily accessible data and fully reproducible methods are uncommon (Hampton et al. 2013; Kidwell et al. 2016). To improve the dissemination of scientific information and data, there has been a push to change the way scientific information is shared (Reichman et al. 2011). Referred to as the "open science" movement, the objective is to encourage transparency throughout all stages of the scientific research process, including data availability; publication access; and detailed, transparent, and reproducible methods. The implementation of "open science" practices improves scientific rigor, encourages collaboration, and accelerates discoveries and innovations overall (Woelfle et al. 2011). Science is built on data, and accordingly, open data is a key aspect of the open science movement. Open data means that data are freely available to the public, and are described by sufficient documentation for appropriate reuse.

The benefits of open science and open data are not limited to altruistic improvements for the broad scientific community. Individual researchers and teams also directly benefit from using open practices. Open scientific publications have been shown to receive more citations and media coverage, and are associated with more job and funding opportunities (McKiernan et al. 2016). When data are reused regularly, as in the case of mandated monitoring programs, making data publicly available with supporting information can streamline data delivery (i.e., there is no need to wait for the data manager to send the data set) and reduce costs. Open access to data also helps data producers gain trust from both the scientific community and the public through transparency and reproducibility. In part to address the history of a lack of openness, the scientific community has responded by more frequently including open data requirements for peer-reviewed scientific journal publications, grant applications, and governmental entities.

The purpose of this article is to highlight the benefits of open science and to share tools for scientists within the San Francisco Estuary (estuary) region to help streamline the process. We are not attempting to provide an overview of the open science movement in the field of ecology, as others have already written recent thorough reviews and perspectives on this topic (Molloy 2011; Hampton et al. 2015; Culina et al. 2018; Powers and Hampton 2019). Although many individuals and organizations collect data of various types and sizes, we highlight the estuary in particular because of the quantity of data that exists. Our own efforts to make environmental data from long-term monitoring programs and research studies in the estuary more accessible daylighted the challenges that many researchers face on their path to open data. To help future data providers make their data fully open, we used our collective experience to develop a set of steps for producing open data and briefly discuss why each step is important.

INTERAGENCY ECOLOGICAL PROGRAM DATA UTILIZATION WORK GROUP

One of the primary groups involved with the generation of long-term data sets of the estuary's natural resources is the Interagency Ecological Program (IEP). The IEP is a consortium of state and federal agencies that has been conducting collaborative ecological studies in the estuary since the 1970s to help inform the potential environmental effects of state and federal pumping operations that divert water from the estuary to the southern part of California. Information gathered by the IEP has been used many times to guide management actions whose goals are to help the recovery of endangered species that reside within the estuary. The IEP currently has over 30 long-term monitoring programs conducted by government agencies and universities, some of which started decades ago between the 1950s and the 1990s. This combined effort amounts to more than 500,000 independent sampling events over the years. Numerous studies that have played a key role in improving our scientific understanding of the estuary relied on these data sets, and information from these efforts have led to follow-up studies that range in topics from genetics, to physiology, to stable isotopes (Sommer et al. 2019).

Given the large spatial and temporal scale of the data collected by IEP, the range of topics covered, and the increasing demand for information, the IEP established in the late 1990s a work group focused on data management. Dubbed the Data Utilization Work Group (DUWG), this group's goal is to ensure that all data and information generated by IEP are of high quality and are consistent with management and science priorities. It does so by setting internal procedures and guidelines, defining and implementing shared data standards across member agencies, facilitating data sharing in a timely manner, and coordinating with other data management teams in the local science community. The DUWG's recent effort to set guidelines and coordinate open data protocols has helped with preparing for implementation of The Open and Transparent Water Data Act (AB 1755, *https:// water.ca.gov/Programs/All-Programs/AB-1755*) in 2016, which requires that all water data collected by governmental state agencies be published in an open data repository.
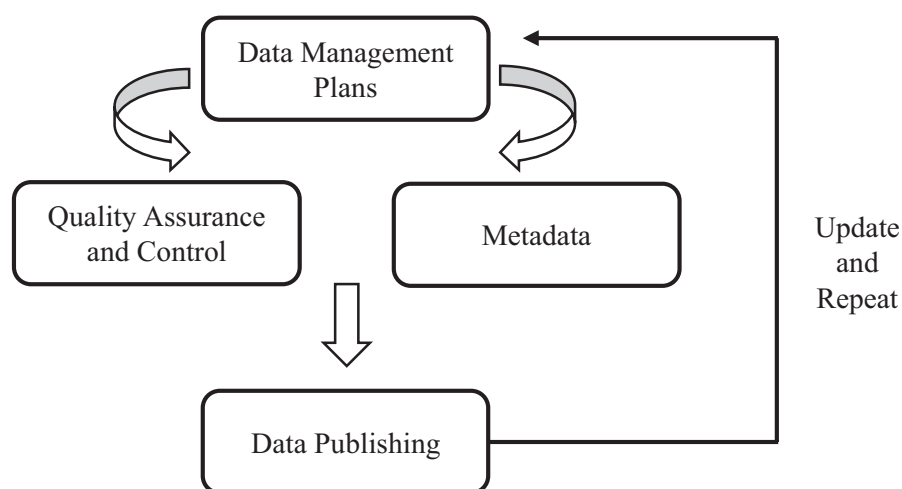
Figure 1 Steps in the open data framework

Data from monitoring programs provides information that ecosystem managers need to determine the status of a species and decide on management actions. However, for data analysts who turn monitoring data into information, locating such data can sometimes be an arduous task that involves finding the contact information of the data manager, formally requesting the data from the individual, and waiting to hear a reply. Even after successfully gathering data, those hoping to conduct analysis sometimes face the additional challenge of having little to no metadata to give context to the data. From the perspective of the data producer, concerns may arise about data requests, because data users who have little understanding of the design or nuances of the study or monitoring program may inadvertently misuse data. There is also concern that data requesters can selectively use the data to support personal or institutional goals or without crediting the agency that collected the data. To address these common concerns, the IEP DUWG has developed an open data framework to facilitate reuse and synthesis of IEP data and to aid other organizations or individuals that are interested in data sharing. This open data framework follows four simple steps:

1. Write a Data Management Plan (DMP) to ensure the future integrity of the data.

2. Develop a Quality Assurance/Quality Control (QA/QC) procedure to maintain the high quality of information.

3. Write a metadata document for data users and new members joining the study or monitoring program.

4. Publish the data set to allow replication of studies and to provide credit for the data producers.

Data Management Plans

Data Management Plans (DMP) are documents created to provide essential information on how data is administrated from raw collection to a final archived format. Formal DMPs are increasingly becoming required for federal and private competitive grants (Thoegersen 2015). DMPs are typically brief documents (a couple of pages) that describe the data life-cycle from data collection to storage, QA/QC procedures, where it will be archived, and steps for sharing (Michener 2015).

DMPs have five general components: program description, data description, data preservation, QA/QC, and data-sharing requirements. (1) The program description should contain data set name, points of contact, and types of data collected or their data sources. (2) The DMP's data description component briefly describes the metadata, any metadata standards used, and where data users can find the metadata document. (3) Data preservation documents the data format (data type and file extension) and how they are stored, backed up, and archived. Data storage and archiving are similar and often used interchangeably, but, for data-management purposes, are different. Data *storage* typically refers to procedures for storing data in the original format in which the data were gathered or entered. Data *archiving* refers to procedures for the final, QA/QC-ed data for long-term sharing and preservation. Both data storage and data archiving should include back-up copies in two to three locations, such as an agency's server and a cloud network. (4) The QA/QC components briefly describe the standards or procedures used to qualify the data, and where additional documents such as standard operating procedures or a QA Program Plan can be located. (5) The data-sharing requirements section describes how and where the public can access the data, as well as how the data should be cited and any requirements for using the data.

DMPs provide many benefits as a planning and communication tool (Jones 2011). For planning, DMPs ensure transparency between data producers and consumers by outlining the steps of how the data will be generated and QA/QC-ed, and where it will be archived. As a communication tool, DMPs can help with onboarding new staff and avoiding unnecessary duplication, and, for long-term maintenance, promote strong metadata on how the data were processed and stored. DMPs can be used for short-term studies or long-term monitoring programs. They are also applicable to synthesis efforts to document the various data sources used in an effort, even when no original data are produced. Table 1 shows a DMP template developed by the IEP DUWG and more information about developing a DMP.

Quality Assurance and Control

Quality Assurance is a system of management activities that include planning, implementing, assessing, reporting, and continuously improving (EPA 2002a). A component of QA is QC, which is the system of technical activities whose purpose is to quantify data errors and/or level of uncertainty, and to determine the effect of those errors (EPA 2006). The foundation of QA is systematic planning, which directs efficient sample collection, early detection of errors, and better results and effective decision-making based on the research (EPA 2001a). Planning

Table 1   Resources for each step of open data planning and publishing

Resource	Citation or Website
General Open Data	
Book: Open Science by Design	National Academies of Sciences Engineering and Medicine (2018)
Data Management Plans	
Article: Ten Simple Rules for Creating a Good Data Management Plan	Michener (2015)
Guide: How to Develop a Data Management and Sharing Plan	Jones (2011)
Template: Interagency Ecological Program Data Management Plan Template and Instructions	*https://github.com/InteragencyEcologicalProgram/An-Open-Data-Framework-for-the-San-Francisco-Estuary-/tree/master/Data%20 Management%20Plans*
Quality Assurance/Quality Control	
Report: EPA Requirements for Quality Assurance Project Plans	EPA (2001a)
Report: EPA Quality Manual for Environmental Programs	EPA (2000)
Report: Guidance for Preparing Standard Operating Procedures	EPA (2007)
Report: Guidance on Environmental Data Verification and Data Validation	EPA (2002c)
Metadata	
Guide: Metadata Standards and Schemas	*https://guides.lib.unc.edu/metadata/standards*
Schema: Ecological Metadata Language	Jones et al. (2019)
Schema: Federal Geographic Data Committee	*https://www.fgdc.gov/metadata/csdgm/organization.html*
Template: Interagency Ecological Program Ecological Metadata Language MS Word Document	*https://github.com/InteragencyEcologicalProgram/IEP-to-EDI-Publishing/blob/master/IEP_EDI_Metadata_Template_05222019.docx*
R Code: Interagency Ecological Program Ecological Metadata Language in machine readable format	*https://github.com/InteragencyEcologicalProgram/IEP-to-EDI-Publishing/blob/master/IEP-to-EDI-Publishing.Rproj*
Guide: EMLassemblyLine in R	*https://github.com/InteragencyEcologicalProgram/IEP-to-EDI-Publishing/blob/master/DIY%20Instructions%20 EMLassemblyLine%20in%20R.docx*
Publishing	
Repository: Environmental Data Initiative	*https://environmentaldatainitiative.org/*
Repository: Knowledge Network for Biocomplexity	*https://knb.ecoinformatics.org/*
Repository: Dataverse	*https://dataverse.org/*
Additional Resources	
Presentation Slides: 2019 Interagency Ecological Program Data Management Showcase	*https://github.com/InteragencyEcologicalProgram/Open-Data-Workshop/blob/master/resources/Data%20Management%20 Showcase%20Slides.pdf*
Data Workshop	
Presentation Video: 2019 Interagency Ecological Program Data Management Showcase	*https://youtu.be/rgGFogjhePc*
California Legislation: Open and Transparent Data Act	*https://water.ca.gov/Programs/All-Programs/AB-1755*

involves laying out the study objectives and identifying the data quality needed to meet those objectives, including appropriate data type, how the data will be used, performance criteria, QC activities, and how the data will be analyzed (EPA 2006). Quality Control activities systematically determine if the data are of the appropriate type, frequency, and quality needed to answer study questions and support resource-management decisions. This includes QC activities for each sampling, analysis, or measurement technique; for each method and procedure; for acceptance criteria; and for corrective actions (EPA 2001a; EPA 2002b).

Quality Assurance is an integral part of open data and involves every stage of a project's life-cycle. The US Environmental Protection Agency (EPA) has been at the forefront in developing quality management systems and directs local regulatory agencies such as the State Water Resources Control Board to follow those models; therefore, our recommendations are built on EPA guidance. Further, when the robust QA practices outlined in EPA guidance are supported at the institution through a Quality Management System (EPA 2001b), quality is better integrated into its culture. Key components of a quality management system include a QA policy, quality system documentation, annual assessments, training and education, systematic planning of projects, project-specific quality documentation, and project and data assessments (EPA 2001b). There are also tools that can be implemented on a project level to ensure data of known and documented quality. These include study plans and/or quality assurance project plans, training plans, and data assessments (EPA 2001a). The necessary QC activities may be identified during the planning process and may include documents such as checklists, instrument-calibration sheets, field sheets, laboratory chain-of-custodies and reports, and/or codes or scripts used in data-validation activities. All these QC activities, tools, and components are interwoven into each step of the data life-cycle, and are applicable to anyone involved with the data—from data collector to lab technician to data scientist. Table 1 shows additional resources for how to get started.

Having proper QA/QC processes allow data producers to communicate the quality, limitations, and appropriate use of the data. With this transparency, data users may be more coordinated in their methods of checking data quality, because they will know the methods originally used to check the data. Having this information at hand may significantly reduce the time spent in performing the data-cleansing activities required for data reuse (Vetrò et al. 2016).

Metadata

Metadata is information about data that describes the 'who, what, where, when, why, and how' of data sets. Metadata is a critical component of a data set because it enables users to conduct the most accurate data analysis and reduces the likelihood of misuse. There are several types of metadata (Zeng 2015; Riley 2017), ranging from: (1) information that describes how the data were collected and produced (process metadata); to (2) the actual content of the data, parameter ranges, and quality of the data (reference metadata); to (3) more general information such as the title, a set of key words, an abstract, and contacts that

help better identify the data and make it discoverable (descriptive metadata), to, finally, (4) how the data are archived over the long-term, permissions, and property rights (administrative metadata). In addition to the various types of metadata, different formats of metadata (also called schemas) exist. While numerous scientists still generate metadata using a text document format (e.g., Microsoft Word), many agencies are requiring the use of standardized science or geospatial metadata formats such as Ecological Metadata Language (EML) (Jones et al. 2019) and Federal Geographic Data Committee (FDGC) schemas. Standardized metadata formats allow for greater readability and interoperability amongst diverse data sets by both scientists and machines (Duval et al. 2002). Both EML and FDGC are used by state and federal agencies within the IEP, and applications are available for interoperability (i.e., cross-walking) between formats. Table 1 shows a complete list of available schemas. Using standardized schemas, IEP data and metadata can be harvested (machine-read) and accessed on a variety of web-based data platforms. Standardized and accessible metadata increases discoverability, contributing to big-data synthesis projects, meeting state AB1755 protocols, and helping decrease data misuse.

With participation from state, federal, and program partners, the IEP DUWG developed recommendations for best metadata practices. Recommended practices include using standardized EML with inclusion of a robust methodology section that better describes field and data collection practices for water quality and biological data for long-term monitoring programs. Detailed metadata for long-term data sets that span decades are particularly important to capture changes in methodologies over time, which may affect analysis and integration of similar data sets used in synthesis efforts. In addition to recommendations for metadata, the IEP DUWG provides resources for creating standardized metadata. These resources include metadata Microsoft Word templates, as well as a number of important metadata elements or types (descriptive, process, reference, and administrative); R-scripts for creating standardized machine-readable EML metadata (in .xml formats); and instructions for how to generate and publish usable metadata (Table 1).

Data Publishing

Data publishing is, essentially, making data sets available for reuse by others. There are several best practices to consider for data publication. One is that a data set and its associated metadata are considered equally valuable and create a complete data package only when combined. The data package can be made public via a public website, a data repository, a journal focused on data publication, or as a supplement to a research article. The latter three allow the data package to obtain a Digital Object Identifier (DOI), which is a universally recognized alpha-numeric sequence assigned by a registration agency (e.g., DataCite or Crossref) to provide a persistent link to the data package's location on the internet. Having a DOI assigned to a data package facilitates data discovery, usage tracking, documentation of data set versions, and retention of the relationship between a data set and its metadata. It is also recommended that machine-readable, stable, non-proprietary, and standardized data formats are used (e.g., .csv, .txt). Promoting

discovery and reuse of a high-quality data package can reap later rewards through increased exposure and citations of your scientific research. Some possibilities include posting the data package or a hyperlink to it on data portals, including a hyperlink on a professional website; requesting it be a featured contribution in a data repository; and calling attention to it on your preferred professional social media site (e.g., Twitter). Promoting a high-quality data package can reap later rewards through increased exposure and citations of your scientific research. Recently, IEP has made a concerted effort to publish and obtain DOIs for six of its monitoring programs and one discrete study (Table 2). This effort is ongoing, and in the next few years many more IEP data packages are anticipated to be added to IEP's data repository of choice (i.e., Environmental Data Initiative) or similar data repositories (Table 1).

Published data sets may range from short-term research studies to multi-decade monitoring programs. Regardless of the data type or scope, publishing is beneficial for data generators, data users, funders, decision-making managers, and the public. Data generators benefit by getting recognition for their scientific contribution outside of the typical peer-reviewed research article and can add the citation to their resume as a publication. Data generators and users both benefit from having data accessible to answer new questions and from becoming part of larger meta-analyses. Researchers participating in larger meta-analyses using accessible data can generate synthesis products that lead to broader discoveries (e.g., Stompe et al., this issue.) Maximizing the data's utility through publication and subsequent meta-analyses can promote increased knowledge sharing and collaborative opportunities, as well as reduce scientific redundancy to make funding dollars go further. The overall increase in efficiency and opportunities to broaden scientific scope can enable the more rapid and informed decision-making needed for a successful adaptive-management framework. The public can also benefit from data publication because the practice promotes increased transparency and data accessibility, and—in conjunction with quality metadata—increases scientific credibility. Given that the benefits of data publishing are numerous, we recommend that more effort be directed towards incentivizing it. For agencies, their executives, managers, and supervisors can strongly promote data publication by including it in job duty statements, by considering it a valued product during performance reviews, and by giving staff sufficient time to work on producing this type of publication.

OPEN DATA CHALLENGES

While the benefits of open data to researchers and the broader scientific community are substantial, scientists face roadblocks in making their data open. Here, we list some of the most common barriers identified by IEP scientists, as well as some insight into how to move past these barriers.

1. **Lack of time or funding.** Making data open and accessible often does not get prioritized, as it necessarily comes toward the end of a project, when researchers are busy writing reports/manuscripts and trying to secure funding

Table 2 Interagency Ecological Program Datasets Published on Environmental Data Initiative

Repository to Date		
Lead Agency	**Dataset Topic**	**Citation**
CA Dept. of Fish and Wildlife	Fish and invertebrate catch and water quality from Delta	CDFW et al. (2018)
CA Dept. of Water Resources	Dissolved oxygen in Stockton Deep Water Ship Channel	IEP et al. (2019a)
CA Dept. of Water Resources	Yolo Bypass Fish Monitoring Program	IEP et al. (2019b)
CA Dept. of Water Resources	Discrete water quality monitoring in Bay-Delta	IEP et al. (2020a)
CA Dept. of Water Resources	Zooplankton in Yolo Bypass and Sacramento River	IEP et al. (2020c)
US Fish and Wildlife Service	Delta electrofishing during drought	IEP et al. (2019c)
US Fish and Wildlife Service	Delta Juvenile Fish Monitoring Program	IEP et al. {2020b}
US Fish and Wildlife Service	Enhanced Delta Smelt Monitoring Program	USFWS et al. (2019)

for the next project. For long-term monitoring programs, there is no "end" to the project, so published data must constantly be updated. The solutions to this barrier are twofold. Agency scientists need to communicate to supervisors and managers the value of open data so that time and resources can be set aside for completing these tasks. Academic scientists need to include time in budgets to publish data. Likewise, to ensure open data is prioritized, funding agencies can include data set publication as a deliverable in contract language.

2. **Lack of knowledge.** Because open science and open data initiatives are still in their relative infancy, researchers often either aren't aware of the value or simply don't know how to efficiently make their data open. This current essay, along with the IEP DUWG and its GitHub page (Table 1), are excellent sources of information. One of the goals of the IEP DUWG is to help make IEP data more open by supporting and guiding researchers. (For additional resources, please refer to Table 1.)

3. **Fear of data misuse or theft.** Researchers can justifiably fear having their data misused by other scientists, or being scooped when multiple publications are planned from a data set.

   a. **Data misuse** concerns can be ameliorated by ensuring that robust metadata is included with any published data, and by ensuring that whatever online repository is used enables users to download all metadata with the data set.

   b. **Data plagiarism** concerns can be addressed by ensuring that published data contain clear instructions for how to cite the data, and that journals require authors to properly cite any third-party data in manuscripts.

   c. **Scooping** concerns can be addressed by carefully timing publications of both manuscripts and data, as well as the use of preprints. Preprints are complete draft manuscripts that are stored in a repository before journal submission or publication. Preprints are publicly available, and each one receives a DOI, allowing them to be cited. Having preprints for your work

ensures that your ideas, study design, and results remain yours. Any potential scooping of a study via open data would be minimized, because other researchers would be more likely to pursue study questions or methods with your data that are distinct from your original study.

STARTING (OR CONTINUING) YOUR OPEN SCIENCE JOURNEY

The path to open data and, ultimately, open science can be intimidating. Fortunately, plentiful resources are available (e.g., Table 1) to individuals, projects, or programs that want to start the journey. Most scientific projects and programs likely have the four key steps (data management plans, QA/QC, metadata, and publishing) of an open data framework. A natural first step to open science is to have open data, because many public agencies and funding sources require grantees to make data open and easily accessible. While there may be legal requirements to make data publicly available, being actively engaged in the culture shift to open science has many benefits. Increased information, transparency, communication, and accessibility will benefit the scientific community and the broader public by increasing efficiency, promoting informed decision-making, creating learning opportunities, and much more. The IEP DUWG supports and promotes a culture of open science and open data in the estuary, its watershed, and beyond.

ACKNOWLEDGEMENTS

REFERENCES

[CDFW et al.] California Department of Fish and Wildlife, Hartman R, Sherman S, Contreras D, Ellis D. 2018. Fish catch, invertebrate catch, and water quality data from the Sacramento–San Joaquin Delta collected by the Fish Restoration Monitoring Program, 2015–2017. version 2. Environmental Data Initiative. [accessed 2020 Jan 21]. *https://doi.org/10.6073/pasta/86810e72766ad19fccb1b9dd3955bdf8*

Culina A, Baglioni M, Crowther TW, Visser ME, Woutersen–Windhouwer S, Manghi P. 2018. Navigating the unfolding open data landscape in ecology and evolution. Nat Ecol Evol. [accessed 2020 Jan 21];2(3):420–426. *https://doi.org/10.1038/s41559-017-0458-2*

Duval E, Hodgins W, Sutton S, Weibel SL. 2002. Metadata principles and practicalities. D-lib Mag. [accessed 2020 Jan 21];8(4). Available from: *http://www.dlib.org/dlib/april02/weibel/04weibel.html*

[EPA] US Environmental Protection Agency. 2000. Quality manual for environmental programs. Washington (DC): EPA Office of Environmental Information. CIO 2105-P-01-0 [accessed 2020 Jan 21]. Available from: *https://www.epa.gov/sites/production/files/2013-10/documents/2105p010.pdf*

[EPA] US Environmental Protection Agency. 2001a. EPA requirements for quality assurance project plans (QA/R-5). Washington (DC): EPA Office of Environmental Information. EPA/240/B-01/003. [accessed 2020 Jan 21]. Available from: *https://www.epa.gov/sites/production/files/2016-06/documents/r5-final_0.pdf*

[EPA] US Environmental Protection Agency. 2001b. EPA requirements for quality management plans (QA/R-2). Washington (DC): EPA Office of Environmental Information. EPA/240/B-01/002. [accessed 2020 Jan 21]. Available from: *https://www.epa.gov/sites/production/files/2016-06/documents/r2-final.pdf*

[EPA] US Environmental Protection Agency. 2002a. Guidance for developing quality systems for environmental programs (QA/G-1). Washington (DC): EPA Office of Environmental Information. EPA/240/R-02/008. [accessed 2020 Jan 21]. Available from: *https://www.epa.gov/sites/production/files/2015-08/documents/g1-final.pdf*

[EPA] US Environmental Protection Agency. 2002b. Guidance for quality assurance project plans (QA-G-5). Washington (DC): EPA Office of Environmental Information. EPA/240/R-02/009. [accessed 2020 Jan 21]. Available from: *https://www.epa.gov/sites/production/files/2015-06/documents/g5-final.pdf*

[EPA] US Environmental Protection Agency. 2002c. Guidance on environmental data verification and data validation (QA/G-8). Washington (DC): EPA Office of Environmental Information. EPA/240/R-02/004. [accessed 2020 Jan 21]. Available from: *https://www.epa.gov/sites/production/files/2015-06/documents/g8-final.pdf*

[EPA] US Environmental Protection Agency. 2006. Guidance on systematic planning using the data quality objectives process (QA/G-4). Washington (DC): EPA Office of Environmental Information. EPA/240/B-06/001. [accessed 2020 Jan 21]. Available from: *https://www.epa.gov/sites/production/files/documents/guidance_systematic_planning_dqo_process.pdf*

[EPA] US Environmental Protection Agency. 2007. Guidance for preparing standard operating procedures (QA-G6). Washington (DC): EPA Office of Environmental Information. EPA/600/B-07/001. [accessed 2020 Jan 21]. Available from: *https://www.epa.gov/sites/production/files/2015-06/documents/g6-final.pdf*

Hampton SE, Anderson SS, Bagby SC, Gries C, Han X, Hart EM, Jones MB, Lenhardt WC, MacDonald A, Michener WK, et al. 2015. The Tao of open science for ecology. Ecosphere. [accessed 2020 Jan 21];6(7):120. *https://doi.org/10.1890/es14-00402.1*

Hampton SE, Strasser CA, Tewksbury JJ, Gram WK, Budden AE, Batcheller AL, Duke CS, Porter JH. 2013. Big data and the future of ecology. Front Ecol Environ. [accessed 2020 Jan 21];11(3):156–162. *https://doi.org/10.1890/120103*

 [IEP et al.] Interagency Ecological Program, Lesmeister S, Rinde J. 2019a. Interagency Ecological Program: discrete dissolved oxygen monitoring in the Stockton Deep Water Ship Channel, collected by the Environmental Monitoring Program, 1997–2018. version 1. Environmental Data Initiative. [accessed 2020 Jan 21].
https://doi.org/10.6073/pasta/4f254205afb605220a6453d933fb5d47

[IEP et al.] Interagency Ecological Program, Schreier B, Davis B, Ikemiyagi N. 2019b. Interagency Ecological Program: fish catch and water quality data from the Sacramento River floodplain and tidal slough, collected by the Yolo Bypass Fish Monitoring Program, 1998–2018. version. 2. Environmental Data Initiative. [accessed 2020 Jan 21].
https://doi.org/10.6073/pasta/b0b15aef7f3b52d2c5adc10004c05a6f

[IEP et al.] Interagency Ecological Program, Speegle J, Kirsch J, Mahardja B. 2019c. Interagency Ecological Program: Sacramento–San Joaquin Delta boat electrofishing survey during the critical drought of 2014. version 1. Environmental Data Initiative. [accessed 2020 Jan 21].
https://doi.org/10.6073/pasta/489ef5790e5e8448ba39d396923d876c

[IEP et al.] Interagency Ecological Program, Lesmeister S, Martinez M. 2020a. Interagency Ecological Program: discrete water quality monitoring in the Sacramento–San Joaquin Bay–Delta, collected by the Environmental Monitoring Program, 2000–2018. version 2. Environmental Data Initiative. [accessed 2020 May 6].
https://doi.org/10.6073/pasta/a215752cb9ac47f9ed9bb0fdb7fc7c19

[IEP et al.] Interagency Ecological Program, McKenzie R, Speegle J, Nanninga A, Cook JR, Hagen J, Mahardja B. 2020b. Interagency Ecological Program: over four decades of juvenile fish monitoring data from the San Francisco Estuary, collected by the Delta Juvenile Fish Monitoring Program, 1976–2019. version 4. Environmental Data Initiative. [accessed 2020 May 6]. *https://doi.org/10.6073/pasta/41b9eebed270c0463b41c5795537ca7c*

[IEP et al.] Interagency Ecological Program, Schreier BM, Pien CL, Adams JB. 2020c. Interagency Ecological Program: zooplankton catch and water quality data from the Sacramento River floodplain and tidal slough, collected by the Yolo Bypass Fish Monitoring Program, 1998–2018. version 1. Environmental Data Initiative. [accessed 2020 May 6]. *https://doi.org/10.6073/pasta/ea437db178d6f7b93213cc0e4a915885*

Jones MB, O'Brien M, Mecum B, Boettiger C, Schildhauer M, Maier M, Whiteaker T, Earl S, Chong S. 2019. Ecological Metadata Language version 2.2.0. KNB Data Repository. [accessed 2020 Jan 21]. *https://doi.org/10.5063/F11834T2*

Jones S. 2011. How to develop a data management and sharing plan. Edinburgh: DCC How-to Guides. [accessed 2020 Jan 21]. Available from:
http://www.dcc.ac.uk/resources/how-guides

Kidwell MC, Lazarević LB, Baranski E, Hardwicke TE, Piechowski S, Falkenberg LS, Kennett C, Slowik A, Sonnleitner C, Hess-Holden C, et al. 2016. Badges to acknowledge open practices: a simple, low-cost, effective method for increasing transparency. PLoS Biol. [accessed 2020 Jan 21];14(5):e1002456. *https://doi.org/10.1371/journal.pbio.1002456*

McKiernan EC, Bourne PE, Brown CT, Buck S, Kenall A, Lin J, McDougall D, Nosek BA, Ram K, Soderberg CK, et al. 2016. How open science helps researchers succeed. eLife. [accessed 2020 Jan 21];5:e16800. *https://doi.org/10.7554/eLife.16800*

Michener WK. 2015. Ten simple rules for creating a good data management plan. PLoS Comput Biol. [accessed 2020 Jan 21];11(10):e1004525. *https://doi.org/10.1371/journal.pcbi.1004525*

Molloy JC. 2011. The Open Knowledge Foundation: open data means better science. PLoS Biol. [accessed 2020 Jan 21];9(12):e1001195. *https://doi.org/10.1371/journal.pbio.1001195*

National Academies of Sciences Engineering and Medicine. 2018. Open science by design: realizing a vision for 21st century research. Washington (DC): The National Academies Press. [accessed 2020 Jan 21]. *https://doi.org/10.17226/25116*

Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, Buck S, Chambers CD, Chin G, Christensen G, et al. 2015. Promoting an open research culture. Science [accessed 2020 Jan 21];348(6242):1422–1425. *https://doi.org/10.1126/science.aab2374*

Powers SM, Hampton SE. 2019. Open science, reproducibility, and transparency in ecology. Ecol Appl. [accessed 2020 Jan 21];29(1):8. *https://doi.org/10.1002/eap.1822*

Reichman OJ, Jones MB, Schildhauer MP. 2011. Challenges and opportunities of open data in ecology. Science. [accessed 2020 Jan 21];331(6018):703–705. *https://doi.org/10.1126/science.1197962*

Riley J. 2017. Understanding metadata: what is metadata, and what is it for?: a primer. National Information Standards Organization (NISO). [accessed 2020 Jan 21]. Available from: *https://www.niso.org/publications/understanding-metadata-2017*

Sommer T, Conrad JL, Culberson S. 2019. Ten essential Bay–Delta articles. San Franc Estuary Watershed Sci. [accessed 2020 Jan 21];17(2). *https://doi.org/10.15447/sfews.2019v17iss2art1*

Thoegersen JL. 2015. Examination of federal data management plan guidelines. J eScience Librarianship. [accessed 2020 Jan 21];4(1):e1072. *https://doi.org/10.7191/jeslib.2015.1072*

[USFWS et al.] US Fish and Wildlife Service, Johnston C, Lee S, Mahardja B, Speegle J, Bernard D. 2019. USFWS: San Francisco Estuary enhanced Delta Smelt monitoring program data, 2016–2019. version 1. Environmental Data Initiative. *https://doi.org/10.6073/pasta/98bce400502fae3a6b77b3e96f6d51e7*

Vetrò A, Canova L, Torchiano M, Minotas CO, Iemma R, Morando F. 2016. Open data quality measurement framework: definition and application to open government data. Gov Inf Q. [accessed 2020 Jan 21];33(2):325–337. *https://doi.org/10.1016/j.giq.2016.02.001*

Woelfle M, Olliaro P, Todd MH. 2011. Open science is a research accelerator. Nat Chem. [accessed 2020 Jan 21];3(10):745–748. *https://doi.org/10.1038/nchem.1149*

Zeng M. 2015. Metadata types and functions. [accessed 2020 Jan 21]. Available from: *https://marciazeng.slis.kent.edu/metadatabasics/types.htm*