Environ Monit Assess (2023) 195:1141
https://doi.org/10.1007/s10661-023-11788-4

RESEARCH

Sharing FAIR monitoring program data improves discoverability and reuse

Jennifer M. Bayer · **Rebecca A. Scully** · **Erin K. Dlabola** · **Jennifer L. Courtwright** · **Christine L. Hirsch** · **David Hockman-Wert** · **Scott W. Miller** · **Brett B. Roper** · **W. Carl Saunders** · **Marcía N. Snyder**

Received: 14 October 2022 / Accepted: 24 August 2023

Abstract Data resulting from environmental monitoring programs are valuable assets for natural resource managers, decision-makers, and researchers. These data are often collected to inform specific reporting needs or decisions with a specific timeframe. While program-oriented data and related publications are effective for meeting program goals, sharing well-documented data and metadata allows users to research aspects outside initial program intentions. As part of an effort to integrate data from four long-term large-scale US aquatic monitoring programs, we evaluated the original datasets against the FAIR (Findable, Accessible, Interoperable, Reusable) data principles and offer recommendations and lessons learned. Differences in data governance across these programs resulted in considerable effort to access and reuse the original datasets. Requirements, guidance, and resources available to support data publishing and documentation are inconsistent across agencies and monitoring programs, resulting in various data formats and storage locations that are not easily found, accessed, or reused. Making monitoring data FAIR will reduce barriers to data discovery and reuse. Programs are continuously striving to improve data management, data products, and metadata; however, provision of related tools, consistent guidelines and standards, and more resources to do this work is needed. Given the value of these data and the significant effort required to access and reuse them, actions and steps intended on improving data documentation and accessibility are described.

Keywords Environmental monitoring · FAIR data · Data management · Metadata · Data sharing · Aquatic habitat

J. M. Bayer (✉) · R. A. Scully
U.S. Geological Survey, Pacific Northwest Aquatic Monitoring Partnership, Cook, WA 98605, USA
e-mail: jbayer@usgs.gov

E. K. Dlabola
U.S. Geological Survey, Forest and Rangeland Ecosystem Science Center, Corvallis, OR 97331, USA

J. L. Courtwright
Watershed Sciences Department, College of Natural Resources, Utah State University, Logan, UT 84322, USA

C. L. Hirsch · D. Hockman-Wert · M. N. Snyder
United States Forest Service, Pacific Northwest Research Station, Corvallis, OR 97331, USA

S. W. Miller
Bureau of Land Management, National Operations Center, Denver, CO 80225, USA

B. B. Roper
United States Forest Service, National Stream and Aquatic Ecology Center, Logan, UT 84332, USA

W. C. Saunders
PACFISH/INFISH Biological Opinion Monitoring Program, United States Forest Service, Logan, UT 84332, USA

Published online: 04 September 2023

Introduction

State, tribal, and federal governments all have mandates to monitor aquatic ecosystems. In the USA, this includes many aquatic monitoring programs designed to understand landscape changes and help decide management actions, such as the challenge of managing landscapes for multiple uses (e.g., grazing, timber harvest, mining, roads; Roper et al., 2019), restoration priorities (Miller et al., 2016), and identifying limiting factors to endangered or threatened fish populations (Nickelson & Lawson, 1998). Collectively, millions of dollars are spent annually in the USA on aquatic monitoring; however, these programs are often independently managed and data access, sharing, and reuse are not at the forefront of program design and maintenance. This siloed approach results in more effort to discover and reuse data by those outside the program.

Data sharing is becoming increasingly common and is typically required by journals and funding organizations (Boeckhout et al., 2018; Fox et al., 2021; Nature, 2016; Smith et al., 2020), including the US federal government (Public Law 115–435, 132 Stat. 5529). FAIR (Findability, Accessibility, Interoperability, Reusability) principles were introduced to facilitate and promote data sharing, offering criteria to improve data findability, accessibility, interoperability, and reusability (Wilkinson et al., 2016). FAIR data are accessible via defined parameters, seamless to reuse and integrate, and heavily reliant on well-described metadata. Cited thousands of times, the FAIR principles (Wilkinson et al., 2016) have been widely adopted and used by scientific fields from plant phenotyping (Pommier et al., 2019) to oncology (Vesteghem et al., 2020). Further, the Foundations for Evidence-Based Policymaking Act of 2018 (2019) aligns with FAIR principles, requiring data to be "machine-readable," in a format that can be processed by a computer, "available in an open format […] not encumbered by restrictions" with an "open license" and "at no cost to the public" (Public Law 115–435, 132 Stat. 5529). Given the adoption of FAIR principles by the wider scientific community, monitoring programs will benefit by striving for FAIR data, as conforming to these widely adopted constructs will increase discoverability, reuse, and citability of program data.

Data management frequently overwhelms monitoring programs and adequate financial support can be difficult to procure. For many monitoring programs, a large quantity of data is collected and maintained and budgetary support for data management is minimal. As a result, data are often released on program-specific websites, solely in reports, or not at all. Furthermore, information on who manages the data or information (provenance), how data were collected or derived (protocols and metadata), and data quality assurance may be incomplete, difficult to find, or dispersed among sources. This can make accessing, sharing, and reusing data an obstacle. Sharing data allows others to conduct new and novel analyses within and outside the intended program objectives and increases the dataset's value and program visibility. For example, combining stream habitat data with other aquatic (water chemistry and quality), biological (fish assemblage, macroinvertebrates), and climate data would help water management decision-making (Larson et al., 2021). Furthermore, integrating data between programs monitoring in different geographic areas or time periods could help provide additional information for management decision-making across jurisdictional boundaries (Heberling et al., 2021; Northwest Indian Fisheries Commission Member Tribes, 2020; U.S. Fish & Wildlife Service, 2009) and inform investigations at international scales (Virro et al., 2021).

We evaluated four aquatic monitoring programs' compliance with FAIR principles while accessing and reusing their data and metadata as part of a stream physical habitat data integration project. Based on this experience and our understanding of challenges related to aquatic monitoring data, we offer lessons learned and highlight recommendations for monitoring programs to plan for, and achieve, FAIR data.

Participating aquatic monitoring programs background

We invited four long-term environmental monitoring programs that assess the condition and trend of wadeable streams and rivers (Table 1) to participate in creating a data exchange standard including controlled vocabularies and an integrated dataset of stream habitat metric data. These programs were selected based on their geographic ranges and long-term data (9–24 years) collected using consistent methods and statistically valid designs. These four programs have evolved alongside each other and

Table 1 Information about the aquatic monitoring programs participating in the data integration project

Program	Program description	Sponsoring entity	Spatio-temporal extent
Assessment, Inventory, and Monitoring (AIM)	The AIM Strategy provides a framework for the BLM to assess the condition and trend of terrestrial, riparian, and wetland and lotic (streams and rivers) resources on BLM-managed lands. The information is used to guide land use planning decisions, policy formulations, and adaptive management decisions for individual field offices to national level reporting (Kachergis et al., 2022; Toevs et al., 2011)	US Bureau of Land Management (BLM)	Western USA including Alaska since 2013
National Aquatic Resource Surveys (NARS)—National Rivers and Streams Assessment (NRSA)	The NARS are collaborative aquatic resource surveys conducted by EPA, states, and tribes. NARS is designed to assess the quality of the nation's coastal waters, lakes and reservoirs, rivers and streams, and wetlands using a statistical probability survey design and standardized field data collection and analysis. The NRSA uses NARS data to provide information on the ecological conditions of the nation's rivers and streams and the key stressors on a national and ecoregional scale (Shapiro et al., 2008; U.S. Environmental Protection Agency, 2016b, 2020b)	US Environmental Protection Agency (EPA)	Continental USA since early 2000s
Aquatic and Riparian Effectiveness Monitoring Program (AREMP)	AREMP focuses on assessing the effectiveness of federal land management under the aquatic conservation strategy (ACS) of the Northwest Forest Plan (NWFP) in California, Oregon, and Washington on maintaining and improving watershed conditions (U.S. Forest Service & U.S. Bureau of Land Management, 1994a, b)	USDA Forest Service (USFS) and US Bureau of Land Management (BLM)	California, Oregon, and Washington since 2002
PacFish/InFish Biological Opinion Monitoring Program (PIBO MP)	The PIBO MP goal is to monitor stream and riparian habitats within the PIBO MP study area to determine if the PacFish (Pacific Anadromous Fish) and InFish (Inland Fish) aquatic conservation strategies can effectively maintain or restore the structure and function of riparian and aquatic systems (Archer et al., 2012)	USDA Forest Service (USFS) and US Bureau of Land Management (BLM)	Continental Western USA since 1998

with collaboration across agencies, including many studies assessing the comparability of collected field data and resulting stream condition metrics (Houston et al., 2002; Lenz & Miller, 1996; Roper et al., 2010; Whitacre et al., 2007). Stream physical habitat data includes channel dimensions (bankfull width and height, stream gradient and sinuosity), pools (residual pool depth, percent pools), substrates (channel fine sediment and pool tail fines), and bank measurements (angle), which are used to assess changing stream conditions over time, which can indicate shifting watershed and ecosystem health.

We assessed each program's data and metadata based on FAIR principles as described by the Global Open FAIR Initiative (GOFAIR, n.d.) to facilitate the implementation of the FAIR data principles. Only a subset of the programs' data products was assessed here; some programs collect and maintain additional data (such as terrestrial or macroinvertebrate surveys), which were not assessed. When referring to program acronyms (AIM, AREMP, NRSA, and PIBO MP) throughout this paper, we are only referencing the specific subset of each program's stream data that we assessed (Bureau of Land Management [BLM] Assessment, Inventory, and Monitoring Project Team, 2021; Miller et al., 2017; U.S. Environmental Protection Agency [EPA], 2006, 2016a, 2020a, 2021; W. Carl Saunders, U.S. Forest Service PacFish/InFish Biological Opinion Monitoring Program, unpublished data, 2021[1]). The lessons learned and recommendations are based on our assessment of these specific datasets and metadata against the FAIR principles and our experience integrating a subset of the program's data.

Monitoring programs and FAIR assessment

These four programs were established before the introduction of FAIR data principles yet still managed to meet many of these principles to varying degrees (Table 2). Below, we describe the findings and summary of the FAIR assessment from Table 2.

[1] At the time of publication, data were not available from the US Forest Service PacFish/InFish Biological Opinion Monitoring Program.

Findability

Associating data and metadata with a citable persistent identifier, such as a digital object identifier (DOI), improves findability by providing access to the object regardless of its location. Also DOIs improve reproducibility of analyses in that they allow for documentation of versioning. Only AIM assigned a persistent identifier to the dataset and metadata on Data.gov, a metadata aggregator which facilitates discovery of open data at all levels of government. NRSA has a DOI registered at Data.gov for the NARS webpage where multiple surveys' datasets and metadata are posted, which obscures identification and versioning of individual datasets and metadata records. There is no persistent identifier for the individual NRSA datasets and metadata. AIM and NRSA data and metadata were found via a searchable resource (Data.gov). Without persistent identifiers, finding the other datasets required some prior knowledge. AREMP and PIBO MP data were findable only through Google, published reports, or the program website, which are not considered registered or indexed searchable resources according to FAIR principles.

Accessibility

Use of identifiers and avoidance of specialized or proprietary tools for data and metadata retrieval are best practice to make data and metadata accessible. Three programs made data and metadata retrievable using standard communications protocol, hypertext transfer protocol (https). While PIBO MP data and metadata required a data request to obtain, contact information was not clearly available on the program's website to do so (Table 2; A1). If contact information (such as a phone number or email address) was posted, it would qualify as a standard communications protocol according to FAIR.

Over time, data sharing methods and file formats may change, causing issues with future data accessibility. The use of a data repository with a long-term management plan ensures the metadata is available for users to find even if the data is no longer available at some point in time. While each program indicated data and metadata would be available in the future via data request (Table 2; A2), this is insufficient for FAIR principles. None of the programs used a data

Table 2 FAIR assessment of AIM, NRSA, AREMP, and PIBO MP programs' data and metadata based on individual FAIR criteria (GOFAIR, n.d.)

FAIR criteria	AIM	NRSA	AREMP	PIBO MP
Findable				
F1. (Meta)data are assigned a globally unique and persistent identifier	No	Yes	No	No
F2. Data are described with rich metadata (defined by R1)	Yes	Somewhat	Somewhat	Somewhat
F3. Metadata clearly and explicitly include the identifier of the data it describes	No	No	No	No
F4. (Meta)data are registered or indexed in a searchable resource	Yes	Yes	No	No
Accessible				
A1. (Meta)data are retrievable by their identifier using a standardized communications protocol	Yes	Yes	Yes	No
A1.1. The protocol is open, free, and universally implementable	Yes	Yes	Somewhat	Somewhat
A1.2. The [communication] protocol allows for an authentication and authorization procedure, where necessary	Not applicable	Not applicable	Not applicable	Yes
A2. Metadata are accessible, even when the data are no longer available	No	Yes	No	No
Interoperable				
I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation	Yes	No	Yes	No
I2. (Meta)data use vocabularies that follow FAIR principles	Not applicable	Not applicable	Not applicable	Not applicable
I3. (Meta)data include qualified references to other (meta)data	Yes	No	No	No
Reusable				
R1. Meta(data) are richly described with a plurality of accurate and relevant attributes	Yes	No	Yes	No
R1.1. (Meta)data are released with a clear and accessible data usage license	Yes	Yes	No	No
R1.2. (Meta)data are associated with detailed provenance	Yes	No	No	No
R1.3. (Meta)data meet domain-relevant community standards	Not applicable	Not applicable	Not applicable	Not applicable

repository to house permanent metadata to ensure longevity and accessibility to the information.

Publishing an Application Programmer Interface (API), a mechanism to access data via a defined communication protocol, aligns with the FAIR data principles, and is an important tool for data systems to pass information back and forth (Klein et al., 2019) and more and more data is being shared this way (Gorelick et al., 2017; Recreation.gov, 2022; U.S. Census Bureau, 2014; U.S. Environmental Protection Agency, 2019; U.S. Geological Survey, n.d.; U.S. Geological Survey Earthquake Hazards Program, n.d.). Only AIM published data with a representational state transfer (REST) API service, allowing researchers and developers to reuse information directly in web portals, mapping tools, and open-source scientific programs (e.g., R and Python) (Klein et al., 2019).

Interoperability

Open and/or standard data formats (i.e., non-proprietary) are key to data accessibility and allow users to access data and metadata without specific software resources. NRSA and AIM data and metadata were available in a free and open format where NRSA used comma separated value (.csv) files for data and text files (.txt) for metadata and AIM used comma separated value (.csv) files for data and eXtensible Markup Language files (.xml) for metadata. In addition to the.csv file fulfilling the open format criterion, AIM data are also available in a variety of other formats, such as.kml, GeoJSON, and shapefile providing a wide range of options for users. AREMP and PIBO MP data and metadata required proprietary software to access (A1.1; Table 2). AREMP published data and metadata as an ESRI ArcGIS geodatabase and PIBO

MP data were available by request as an ArcGIS geodatabase or a Microsoft Excel (.xlsx) spreadsheet.

Geodatabases and Microsoft Excel files can be a barrier to some users without the necessary software licenses to open files in their native environment. While ArcGIS geodatabases can be accessed via open-source software (e.g., QGIS, Python, R), geodatabases are still considered proprietary by the Library of Congress (2022) and are not acceptable file formats according to Federal Records Management (National Archives, 2022). Microsoft Excel (.xlsx) files are also accessible via open-source software (e.g., OpenOffice, LibreOffice) but the.xls file format is more widely compatible and acceptable (National Archives, 2022).

Using standard vocabularies and consistent accessible language makes interoperability easier within a program and between different program datasets. There are data exchange standards containing standard vocabularies for water quality data, such as EPA Water Quality Exchange (U.S. Environmental Protection Agency [EPA], 2019) and Observation Data Model 2 (ODM2; Horsburgh et al., 2008). Pre-existing standards and vocabularies were insufficient for describing the programs' stream physical habitat data and we acknowledge the lack of a standard inhibits each program's ability to achieve the FAIR Interoperable principles.

Without a standard vocabulary for stream physical attributes, each program created an independent vocabulary. As a result, vocabularies differed between programs as well as within individual programs. To meet specific reporting requirements, NRSA data and metadata are published as individual releases after 2-years' data collection. Individual NRSA datasets spanning 2 years had new metrics or altered field headers compared to other 2-year spans. NRSA maintains a single comprehensive dataset covering all years surveyed but does not make this information public. Other programs append new data to a pre-existing dataset, which eliminates the need for a data user to interoperate across years within the same program's data. Furthermore, interoperability (and findability) is improved by publishing machine (and human) readable metadata, which may be readable by a machine without specific mappings or translators or if translations are needed, to provide formal, shared documentation to allow these translations to be understood and repeatable. Only AIM and AREMP shared metadata in a machine-readable format as eXtensible Markup Language (XML).

FAIR principle I3 suggests providing direct links to any supplemental data (such as protocol documentation or published reports) within the data or metadata. In long-term aquatic monitoring, data collection protocols often change over time due to advances in technology, changes in monitoring questions, or work to coordinate collection with other monitoring programs. Only the AIM dataset has cross references to the specific data collection protocol used. PIBO MP data has coded fields to indicate the protocol used to collect the data but does not provide direct cross reference to protocol documentation. AREMP and NRSA datasets and metadata do not have cross references to the data collection protocols. None of these programs has direct references to changes in analysis over time.

Reusability

Richly described metadata promotes reusability. The metadata and the thoroughness of descriptions varied by program (F2 and R1; Table 2). AIM metadata includes field definitions and ranges, which helps the user know what is in the dataset and whether it meets their needs, as well as the scope, purpose of work, limitations, and keywords. AREMP's overview metadata includes keywords, purpose, and limitations, providing overall context. However, the overview record does not include field definitions or ranges, which makes it difficult for users to know what is in the dataset without accessing the entire dataset and extracting each individual layer's metadata file. PIBO MP and NRSA metadata include field definitions and units, but lack other important metadata fields, such as contact information, citation, purpose, limitations, spatial projections, and keywords.

AIM and AREMP used domain-relevant community metadata standards (R1.3; Table 2) by following the Federal Geographic Data Committee-Content Standard for Digital Geospatial Metadata (FGDC-CSDGM) and had more richly described metadata as a result. AIM also created metadata following the International Organization for Standardization Geographic Information standard (ISO-19139). NRSA and PIBO MP did not follow a metadata standard. However, a community standard did not exist for wadeable stream habitat data until we began this work, and it does not include all possible stream habitat variables, so this criterion is inapplicable and unachievable for the programs in our FAIR assessment.

While not included in the metadata or data files, field and lab protocols documenting data provenance (a record trail that accounts for the origin of a piece of data (in a database, document, or repository) together with an explanation of how and why it got to the present place) were available on each program's website (R1.2; Table 2). Providing this information with stream habitat monitoring data is essential because data collection and analysis methods impact the metric values and can influence the interpretation of resulting data (Al-Chokhachy et al., 2010). Reuse of stream habitat data can only happen if data provenance is understood through well-documented protocols readily available to users.

Lessons learned

Lack of high-level infrastructure and agency mandates

Since these programs are managed by different federal agencies and departments, there are no consistent data management requirements or comprehensive process or platform for sharing and publishing aquatic monitoring data and metadata. Programs whose sponsoring agencies have requirements to publish data and metadata met more of the FAIR principles because they have some guidance, resources, and support to complete the work to make data and metadata FAIR. However, the resources available to support infrastructure and staff are uneven across agencies and even within agencies. For example, the USFS has a publication or research mandate in place only for research stations, not the entire National Forest Service, which leaves some programs (such as PIBO MP and AREMP) without the necessary support and resources to reach all FAIR standards.

Lack of community standard or centralized repository

Our FAIR assessment found wide disparities across these four programs' ease of discovery, access, interoperability, and reuse of data and metadata. Overall, programs with persistent identifiers and richly described metadata following a metadata standard were easier to find and reuse. Open or standard (non-proprietary) file formats were also easier to access. Program datasets using consistent vocabulary throughout the temporal extent of the data were easier to interoperate and reuse, though doing so required extensive time and effort, and close collaboration with the programs to yield an integrated dataset.

Without a community standard, interoperability principles are difficult for programs to accomplish. Efforts have been made to improve interoperability between programs by coordinating data collection and analysis methods (Houston et al., 2002; Lenz & Miller, 1996; Roper et al., 2010; Whitacre et al., 2007), and some programs share historical roots in their initiation. There is currently no single repository, data dashboard dedicated to aquatic monitoring data, or a single location for a user to search for such data. Program data lived in a variety of locations, which required significant time spent searching and digging for information. The EPA Water Quality Exchange (WQX) supports the documentation and publication of aquatic data and may provide a vehicle for future data sharing but does not currently have the necessary standards or controlled vocabularies in place to do so. To address the lessons learned, we offer recommendations to aquatic monitoring programs below.

Limitations of FAIR principles

While FAIR principles help improve data sharing, they do not specifically define and clearly address all aspects of data stewardship and may be difficult to interpret and apply. Various sampling designs, different analytical techniques, and data quality assurance are not directly addressed by FAIR principles (Boeckhout et al., 2018). If these aspects are not well-documented and available, they may impede data reusability and interoperability (McCord et al., 2022). Additionally, data reusability and interoperability are two aspects of FAIR which tend to be harder to understand and achieve (Devaraju et al., 2021; Kinkade & Shepherd, 2021). For example, complications with interoperability arose even when multiple programs collected the same attribute, such as stream temperature (Isaak et al., 2017). To address this issue, programs collaborated to compare collected field data and resulting stream condition metrics (Houston et al., 2002; Lenz & Miller, 1996; Roper et al., 2010; Whitacre et al., 2007). However, some stream attributes remain problematic to interoperate due to differences in methodologies, equipment used, and training of personnel (volume of wood debris) (Isaak et al., 2017; Whitacre et al., 2007) so documentation of field and analysis protocols is essential. Furthermore, FAIR data does not necessarily conform to a discipline-specific standard, which may make findability more challenging than if such disciplinary standards existed and were widely used (Musen, 2022).

Recommendations

Natural resource monitoring programs spend substantial effort and money collecting and processing data, which should be treated as valuable assets for natural resource managers and decision-makers (Volk et al., 2014) with longevity and reuse in mind. Additionally, as remote sensing becomes a more common technique for aquatic monitoring, FAIR long-term datasets will be important to support the cost-effective integration of existing monitoring data with new remote sensing datasets by providing ground-truth validation to calibrate models. Designating or adapting an existing repository or building a new data repository that meets the characteristics described by the National Science and Technology Council (2022) would address many of the challenges programs face in sharing FAIR data, allowing the data to reach a broader, more diverse user community. A centralized repository could support the use of a single metadata standard, data standard, vocabulary, and open file formats, addressing many of the criteria that are more difficult for individual monitoring programs to address themselves. In the wider community, this is not a new concept (Bricker & Ruggiero, 1998; Jaeger et al., 2021). In 1994, the Government Accounting Office (GAO) pointed out that "understanding the ecology of an ecosystem will require collecting and linking large volumes of scientific data," but that "available data are often not comparable, and large gaps in information exist" (GAO, 1994).

In the absence of national infrastructure or agency mandates, monitoring programs could independently move toward more consistent data sharing by following the guidelines set out by the FAIR principles for data management. Furthermore, if such a resource becomes available in the future, prior adoption of FAIR data principles may save time and effort when transitioning to use larger scale data management infrastructure. The FAIR principles provide a valuable framework to improve data stewardship and it is important to recognize data and metadata do not have to meet all FAIR principles to be shared or reused. However, it is also important to recognize FAIR data do not automatically lead to data sharing (Boeckhout et al., 2018), guarantee reproducibility (Peer et al., 2021), ensure the ability to integrate data across programs (Mons et al., 2017), or address different spatial designs or analytical techniques (Boeckhout et al., 2018).

We acknowledge that each recommendation may not be possible for all programs to implement based on resource limitations. The four programs described here are all moving toward implementing FAIR data principles and have each made changes to their practices throughout the duration of our work on this project that were informed by working together. Also, as a result of this project these four programs have collaborated to make substantial progress towards a community domain-specific ontology (controlled vocabulary), which will be published via a separate mechanism. Prioritizing these actions for monitoring programs may maximize the benefits of FAIR data:

1. Share metadata in a machine-readable format to allow searchable harvesting by an aggregator (such as www.Data.gov) so keywords and search terms lead a user to each program's data from a single searchable resource. (Refer to "How to Get Your Open Data on Data.gov" for guidance; Data.gov, n.d.)
2. Assign a persistent identifier to data (and metadata) upon release. Like a journal article, a persistent identifier helps data users find data via citations and maintain longevity of data access by avoiding defunct links if websites move or change in the future. Established data repositories, such as USGS ScienceBase (www.sciencebase.gov), Dryad (www.datadryad.org), Environmental Data Repository (www.edirepository.org), and Pangaea (www.pangaea.de) assign digital object identifiers (DOIs; a type of persistent identifier).
3. Use a metadata standard (such as Federal Geographic Data Committee's Content Standard for Digital Geospatial Metadata (FGDC-CSDGM), Ecological Metadata Language (EML), or International Organization for Standards (ISO)) to ensure machine-readable metadata are richly described and contain necessary information for users to understand the data content and definitions, provenance, limitations, and attribution. For data provenance, include citations (and ide-

ally links) to any data collection or analysis protocol documentation, how metrics were calculated (and the software or code used to do so), and supporting material that provides context to the dataset (e.g., a report specific to a survey year) within the metadata record.

4. Develop and maintain consistent vocabularies and standards for monitoring program data. If a community-wide vocabulary cannot be agreed upon between programs, use consistent vocabulary within a single program over time. If changes occur, document them, and make the information readily available to users. Wherever possible, spell out words that are easily understood in field headers rather than using abbreviations or codes, which require definitions. Continue to build on community domain-specific vocabulary, such as being developed in this project (in review).
5. Release data in a free open and/or standard format (non-proprietary) alongside proprietary file types to maximize accessibility to a wide range of users and to future-proof data. Furthermore, adopt APIs for data sharing to provide a suite of available file types to suit different user knowledge, experience, and preferences.
6. Assign a reviewer for metadata and data prior to publishing or release. Like the journal publication process, an outside party without direct involvement with the product is preferred, but not required for this task, because they can provide a true assessment of the data and metadata from the user's point of view without inside knowledge of the product.

Conclusion

Individual monitoring programs often struggle to implement the FAIR principles due to a lack of agency guidance or requirements for data publishing, agreed-on data standards, or funding to support data management. This results in considerable challenges for secondary data users such as those working across programmatic jurisdictions. Improving data documentation and accessibility by following FAIR principles can enhance the reuse of data. Adapting data and metadata to the FAIR principles and explicitly focusing on the six actions above will insulate data against future changes to monitoring technologies and data storage, help streamline data integration, and therefore, save managers, researchers, and decision-makers time and resources. Making data and metadata FAIR is one step toward improving data sharing and facilitating data integration across programs. If programs also build or use existing publicly available data repositories, and publish data using consistent vocabularies, environmental monitoring data can be more easily combined with other information, including biological and climate data, ultimately providing data users more information to answer research questions and make management decisions.

Acknowledgements The use of trade, product, or firm names in this publication does not imply endorsement by the US Government. An animal care and use protocol was not required for this research.

Author contribution Jennifer M. Bayer, Rebecca A. Scully, and Erin K. Dlabola wrote the main manuscript text. Jennifer L. Courtwright, Christine L. Hirsch, David Hockman-Wert, Scott W. Miller, Brett B. Roper, W. Carl Saunders, and Marcía N. Snyder participated in analysis of data, analysis of dataset metadata, and programmatic metadata across programs. All authors reviewed the manuscript.

Funding This study was funded by US Geological Survey's Core Science Systems Mission Area and Ecosystem Mission Area, US Bureau of Land Management, and US Department of Agriculture Forest Service.

Data availability All data and metadata analyzed during this study are available from each respective program upon request. Programs may be accessed via links provided in Table 1.

Declarations

Ethical responsibilities of authors All authors have read, understood, and have complied as applicable with the statement on "Ethical responsibilities of authors" as found in the Instructions for Authors, and are aware that with minor exceptions, no changes can be made to authorship once the paper is submitted.

Conflict of interest The authors declare no competing interests.

References

Al-Chokhachy, R., Roper, B. B., & Archer, E. K. (2010). Evaluating the status and trends of physical stream habitat in headwater streams within the interior Columbia River and upper Missouri River basins using an index approach. *Transactions of the American Fisheries Society, 139*(4), 1041–1059. https://doi.org/10.1577/T08-221.1

Archer, E., Hough-Snee, N., Van Wagenen, A., Lokteff, R., & Roper, B. B. (2012). *The PACFISH/INFISH Biological Opinion (PIBO) Effectiveness Monitoring Program and invasive plant species detection: A retrospective summary 2003–2011*. 26.

Boeckhout, M., Zielhuis, G. A., & Bredenoord, A. L. (2018). The FAIR guiding principles for data stewardship: Fair enough? *European Journal of Human Genetics, 26*(7), 931–936. https://doi.org/10.1038/s41431-018-0160-0

Bricker, O. P., & Ruggiero, M. A. (1998). Toward a national program for monitoring environmental resources. *Ecological Applications, 8*(2), 326–329. https://doi.org/10.1890/1051-0761(1998)008[0326:TANPFM]2.0.CO;2

Bureau of Land Management Assessment, Inventory, and Monitoring Project Team. (2021). *I_Indicators* [vector digital data]. BLM national AIM lotic indicators ArcGIS Hub. Retrieved September 1, 2022, from https://gbp-blm-egis.hub.arcgis.com/datasets/BLM-EGIS::blm-natl-aim-lotic-indicators-hub/about

Data.gov. (n.d.). *How to get your open data on Data.gov*. Retrieved September 1, 2022, from https://resources.data.gov/resources/data-gov-open-data-howto/#:~:text=Federal%20agencies%20that%20do%20not,on%20resources.data.gov

Devaraju, A., Mokrane, M., Cepinskas, L., Huber, R., Herterich, P., de Vries, J., Akerman, V., L'Hours, H., Davidson, J., & Diepenbroek, M. (2021). From conceptualization to implementation: FAIR assessment of research data objects. *Data Science Journal, 20*, 4. https://doi.org/10.5334/dsj-2021-004

Foundations for Evidence-Based Policymaking Act of 2018, Pub. L. No. 115–435, 115 Stat. 1425. (2019). H.R. 4174. Retrieved September 1, 2022, from https://www.congress.gov/115/plaws/publ435/PLAW-115publ435.pdf

Fox, P., Erdmann, C., Stall, S., Griffies, S. M., Beal, L. M., Pinardi, N., Hanson, B., Friedrichs, M. A. M., Feakins, S., Bracco, A., Pirenne, B., & Legg, S. (2021). Data and software sharing guidance for authors submitting to AGU journals. Zenodo. https://doi.org/10.5281/zenodo.5124741

GAO, U. S. (1994). *Ecosystem management: Additional actions needed to adequately test a promising approach*. U.S. GAO/RCED-94–111. Retrieved September 1, 2022, from https://www.gao.gov/assets/t-rced-94-308.pdf

Global Open FAIR. (n.d.). *FAIR principles*. GOFAIR. Retrieved September 1, 2022, from http://www.go-fair.org/fair-principles/

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment, 202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Heberling, J. M., Miller, J. T., Noesgaard, D., Weingart, S. B., & Schigel, D. (2021). Data integration enables global biodiversity synthesis. *Proceedings of the National Academy of Sciences, 118*(6), e2018093118. https://doi.org/10.1073/pnas.2018093118

Horsburgh, J. S., Tarboton, D. G., Maidment, D. R., & Zaslavsky, I. (2008). A relational model for environmental and water resources data. *Water Resources Research, 44*(5). https://doi.org/10.1029/2007WR006392

Houston, L., Barbour, M. T., Lenat, D., & Penrose, D. (2002). A multi-agency comparison of aquatic macroinvertebrate-based stream bioassessment methodologies. *Ecological Indicators, 1*(4), 279–292. https://doi.org/10.1016/S1470-160X(02)00024-9

Isaak, D. J., Wenger, S. J., Peterson, E. E., Ver Hoef, J. M., Nagel, D. E., Luce, C. H., Hostetler, S. W., Dunham, J. B., Roper, B. B., Wollrab, S. P., Chandler, G. L., Horan, D. L., & Parkes-Payne, S. (2017). The NorWeST summer stream temperature model and scenarios for the western U.S.: A crowd-sourced database and new geospatial tools foster a user community and predict broad climate warming of rivers and streams. *Water Resources Research, 53*(11), 9181–9205. https://doi.org/10.1002/2017WR020969

Jaeger, K. L., Hafen, K. C., Dunham, J. B., Fritz, K. M., Kampf, S. K., Barnhart, T. B., Kaiser, K. E., Sando, R., Johnson, S. L., McShane, R. R., & Dunn, S. B. (2021). Beyond streamflow: Call for a national data repository of streamflow presence for streams and rivers in the United States. *Water, 13*(12), 1627. https://doi.org/10.3390/w13121627

Kachergis, E., Miller, S. W., McCord, S. E., Dickard, M., Savage, S., Reynolds, L. V., Lepak, N., Dietrich, C., Green, A., Nafus, A., Prentice, K., & Davidson, Z. (2022). Adaptive monitoring for multiscale land management: Lessons learned from the Assessment, Inventory, and Monitoring (AIM) principles. *Rangelands, 44*(1), 50–63. https://doi.org/10.1016/j.rala.2021.08.006

Kinkade, D., & Shepherd, A. (2021). Geoscience data publication: Practices and perspectives on enabling the FAIR guiding principles. *Geoscience Data Journal*, *9*(1). https://doi.org/10.1002/gdj3.120

Klein, E., Appeltans, W., Provoost, P., Saeedi, H., Benson, A., Bajona, L., Peralta, A. C., & Bristol, R. S. (2019). OBIS infrastructure, lessons learned, and vision for the future. *Frontiers in Marine Science, 6*, 588. https://doi.org/10.3389/fmars.2019.00588

Larson, K. B., Wong, M. A. Y. L., Mitchel, K. C., Tagestad, J. D., Saulsbury, J. W., Bellgraph, B. J., & Reicher, D. W. (2021). *Improving discovery, sharing, and use of water data: Initial findings and suggested future work* (PNNL-29737). Pacific Northwest National Laboratory. Retrieved September 1, 2022, from https://www.pnnl.gov/main/publications/external/technical_reports/PNNL-29737.pdf

Lenz, B. N., & Miller, M. A. (1996). *Comparison of aquatic macroinvertebrate samples collected using different field methods* (Fact Sheet No. 216–96; Fact Sheet). U.S. Geological Survey. https://doi.org/10.3133/fs21696

Library of Congress. (2022). *Sustainability of digital formats: Planning for Library of Congress collections.* Retrieved September 1, 2022, from https://www.loc.gov/preservation/digital/formats/fdd/fdd000293.shtml#sign

McCord, S. E., Welty, J. L., Courtwright, J., Dillon, C., Traynor, A., Burnett, S. H., Courtright, E. M., Fults, G., Karl, J. W., Van Zee, J. W., Webb, N. P., & Tweedie, C. (2022). Ten practical questions to improve data quality. *Rangelands, 44*(1), 17–28. https://doi.org/10.1016/j.rala.2021.07.006

Miller, S., Eldred, P., Muldoon, A., Anlauf-Dunn, K., Stein, C., Hubler, S., Merrick, L., Haxton, N., Larson, C., Rehn, A., Ode, P., & Vander Laan, J. (2016). A large-scale, multiagency approach to defining a reference network for Pacific Northwest streams. *Environmental Management, 58*(6), 1091–1104. https://doi.org/10.1007/s00267-016-0739-6

Miller, S. A., Gordon, S. N., Eldred, P., Beloin, R. M., Wilcox, S., Raggon, M., Andersen, H., & Muldoon, A. (2017). *Northwest Forest Plan—The first 20 years (1994–2013): Watershed condition status and trends* (Gen Tech. Rep. PNW-GTR-932). U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station.

Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L. O. B., & Wilkinson, M. D. (2017). Cloudy, increasingly FAIR; revisiting the FAIR data guiding principles for the European Open Science Cloud. *Information Services & Use, 37*(1), 49–56. https://doi.org/10.3233/ISU-170824

Musen, M. A. (2022). Demand standards to sort FAIR data from foul. *Nature, 609*(7926), 222–222. https://doi.org/10.1038/d41586-022-02820-7

National Archives. (2022). *Federal records management appendix A: Tables of file formats*. Archives.Gov. Retrieved September 1, 2022, from https://www.archives.gov/records-mgmt/policy/transfer-guidance-tables.html

Nature. (2016). Announcement: Where are the data? *Nature*, *537*(138). https://doi.org/10.1038/537138a

Nickelson, T. E., & Lawson, P. W. (1998). Population viability of coho salmon, *Oncorhynchus kisutch*, in Oregon coastal basins: Application of a habitat-based life cycle model. *Canadian Journal of Fisheries and Aquatic Sciences, 55*(11), 2383–2392. https://doi.org/10.1139/f98-123

Northwest Indian Fisheries Commission Member Tribes. (2020). *2020 state of our watersheds report: A report by the treaty tribes in western Washington*. Retrieved September 1, 2022, from https://files.nwifc.org/sow/2020/state-of-our-watersheds-sow-2020-final-web.pdf

Peer, L., Arguillas, F., Honeyman, T., Miljković, N., Gehlen, K. P.-V., & CURE-FAIR WG Subgroup 3. (2021). *Challenges of curating for reproducible and FAIR research output (2.1)*. https://doi.org/10.15497/RDA00063

Pommier, C., Michotey, C., Cornut, G., Roumet, P., Duchêne, E., Flores, R., Lebreton, A., Alaux, M., Durand, S., Kimmel, E., Letellier, T., Merceron, G., Laine, M., Guerche, C., Loaec, M., Steinbach, D., Laporte, M. A., Arnaud, E., Quesneville, H., & Adam-Blondon, A. F. (2019). Applying FAIR principles to plant phenotypic data management in GnpIS. *Plant Phenomics*, *2019*, 1–15. https://doi.org/10.34133/2019/1671403

Recreation.gov. (2022). *Recreation Information Database API*. Retrieved September 1, 2022, from https://ridb.recreation.gov/landing

Roper, B. B., Buffington, J. M., Bennett, S., Lanigan, S. H., Archer, E., Downie, S. T., Faustini, J., Hillman, T. W., Hubler, S., Jones, K., Jordan, C., Kaufmann, P. R., Merritt, G., Moyer, C., & Pleus, A. (2010). A comparison of the performance and compatibility of protocols used by seven monitoring groups to measure stream habitat in the Pacific Northwest. *North American Journal of Fisheries Management, 30*(2), 565–587. https://doi.org/10.1577/M09-061.1

Roper, B. B., Saunders, W. C., & Ojala, J. V. (2019). Did changes in western federal land management policies improve salmonid habitat in streams on public lands within the Interior Columbia River Basin? *Environmental Monitoring and Assessment, 191*(9), 574. https://doi.org/10.1007/s10661-019-7716-5

Shapiro, M. H., Holdsworth, S. M., & Paulsen, S. G. (2008). The need to assess the condition of aquatic resources in the US. *Journal of the North American Benthological Society, 27*(4), 808–811. https://doi.org/10.1899/08-116.1

Smith, J. A., Cessi, P., Fer, I., Foltz, G., Fox-Kemper, B., Heywood, K., Jones, N., Klymak, J., & LaCasce, J. (2020). Data availability principles and practice. *Journal of Physical Oceanography, 50*(12), 3377–3378. https://doi.org/10.1175/JPO-D-20-0266.1

The National Science and Technology Council. (2022). Desirable characteristics of data repositories for federally funded research. *Executive Office of the President of the United States*. https://doi.org/10.5479/10088/113528

Toevs, G. R., Taylor, J. J., Spurrier, C. S., MacKinnon, W. C., & Bobo, M. R. (2011). *Bureau of Land Management Assessment, Inventory, and Monitoring Strategy: For integrated renewable resources management*. U.S. Department of Interior, Bureau of Land Management, National Operations Center. Retrieved September 1, 2022, from https://www.blm.gov/sites/blm.gov/files/uploads/IB2012-080_att1.pdf

U.S. Census Bureau. (2014). *Census Data API Discovery Tool*. Retrieved September 1, 2022, from https://www.census.gov/data/developers/updates/new-discovery-tool.html

U.S. Environmental Protection Agency. (2006). *National Aquatic Resource Surveys, Wadeable Streams Assessment 2004* [tabular files]. Retrieved September 1, 2022, from https://www.epa.gov/national-aquatic-resource-surveys/data-national-aquatic-resource-surveys

U.S. Environmental Protection Agency. (2016a). *National Aquatic Resource Surveys, National Rivers and Streams Assessment 2008–2009* [tabular files]. Retrieved September 1, 2022, from https://www.epa.gov/national-aquatic-resource-surveys/data-national-aquatic-resource-surveys

U.S. Environmental Protection Agency. (2016b). *National Rivers and Streams Assessment 2008–2009: A collaborative survey* (EPA/841/R-16/007). Office of Water and Office of Research and Development. Retrieved September 1, 2022, from https://www.epa.gov/sites/default/files/2016-03/documents/nrsa_0809_march_2_final.pdf

U.S. Environmental Protection Agency. (2019, May 16). *Water Quality eXchange (WQX) web application programming interface*. Retrieved September 1, 2022, from www.epa.gov/waterdata/wqx-web-application-programming-interface

U.S. Environmental Protection Agency. (2020a). *National Aquatic Resource Surveys, National Rivers and Streams Assessment 2013–2014* [tabular files]. Retrieved September 1, 2022, from https://www.epa.gov/national-aquatic-resource-surveys/data-national-aquatic-resource-surveys

U.S. Environmental Protection Agency. (2020b). *National Rivers and Streams Assessment 2013–2014: A collaborative survey* (EPA 841-R-19–001). Retrieved September 1, 2022, from https://www.epa.gov/system/files/documents/2021-10/nrsa_13-14_report_508_ci_2021-10-15.pdf

U.S. Environmental Protection Agency. (2021). *National Aquatic Resource Surveys, National Rivers and Streams Assessment 2018–2019* [tabular files]. Retrieved September 1, 2022, from https://www.epa.gov/national-aquatic-resource-surveys/data-national-aquatic-resource-surveys

U.S. Fish & Wildlife Service. (2009). *National Fish Habitat Action Plan* (Fisheries Part 717). Division of Fisheries and Aquatic Resource Conservation. Retrieved September 1, 2022, from https://www.fws.gov/policy-library/717fw1

U.S. Forest Service, & U.S. Bureau of Land Management. (1994a). Appendix B6 – Aquatic conservation strategy. In *Final supplemental environmental impact statement on management of habitat for late-successional and old-growth species within the range of the northern spotted owl* (p. B-81-B-128). Retrieved September 1, 2022, from https://www.fs.usda.gov/r6/reo/library/downloads/documents/NWFP-FSEIS-1994-II.pdf

U.S. Forest Service, & U.S. Bureau of Land Management. (1994b). *Record of decision for amendments to Forest Service and Bureau of Land Management planning documents within the range of the northern spotted owl* (p. 74).

U.S. Forest Service PacFish/InFish Biological Opinion Monitoring Program. (2021). Unpublished data received from and used with permission of U.S. Forest Service.

U.S. Geological Survey. (n.d.). *USGS Water Services API*. Retrieved September 1, 2022, from https://waterservices.usgs.gov

U.S. Geological Survey Earthquake Hazards Program. (n.d.). *API documentation: Earthquake catalog*. Retrieved September 1, 2022, from https://earthquake.usgs.gov/fdsnws/event/1/

Vesteghem, C., Brøndum, R. F., Sønderkær, M., Sommer, M., Schmitz, A., Bødker, J. S., Dybkær, K., El-Galaly, T. C., & Bøgsted, M. (2020). Implementing the FAIR data principles in precision oncology: Review of supporting initiatives. *Briefings in Bioinformatics, 21*(3), 936–945. https://doi.org/10.1093/bib/bbz044

Virro, H., Amatulli, G., Kmoch, A., Shen, L., & Uuemaa, E. (2021). GRQA: Global river water quality archive. *Earth System Science Data, 13*(12), 5483–5507. https://doi.org/10.5194/essd-13-5483-2021

Volk, C. J., Lucero, Y., & Barnas, K. (2014). Why is data sharing in collaborative natural resource efforts so hard and what can we do to improve it? *Environmental Management, 53*(5), 883–893. https://doi.org/10.1007/s00267-014-0258-2

Whitacre, H. W., Roper, B. B., & Kershner, J. L. (2007). A comparison of protocols and observer precision for measuring physical stream attributes. *Journal of the American Water Resources Association, 43*(4), 923–937. https://doi.org/10.1111/j.1752-1688.2007.00074.x

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., & Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data, 3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.