

This work was written as part of one of the author's official duties as an Employee of the United States Government and is therefore a work of the United States Government. In accordance with 17 U.S.C. 105, no copyright protection is available for such works under U.S. Law. Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it's important to you. Thank you.

Call to Action for Global Access to and Harmonization of Quality Information of Individual Earth Science Datasets

Ge Peng¹, Robert R. Downs², Carlo Lacagnina³, Hampapuram Ramapriyan⁴, Ivana Ivánová⁵, David Moroni⁶, Yaxing Wei⁷, Gilles Larnicol⁸, Lesley Wyborn⁹, Mitch Goldberg¹⁰, Jörg Schulz¹¹, Irina Bastrakova¹², Anette Ganske¹³, Lucy Bastin¹⁴, Siri Jodha S. Khalsa¹⁵, Mingfang Wu¹⁶, Chung-Lin Shie¹⁷, Nancy Ritchey¹⁸, Dave Jones¹⁹, Ted Habermann²⁰, Christina Lief²¹, Iolanda Maggio²², Mirko Albani²³, Shelley Stall²⁴, Lihang Zhou²⁵, Marie Drévillon²⁶, Sarah Champion²⁷, C. Sophie Hou²⁸, Francisco Doblas-Reyes²⁹, Kerstin Lehnert³⁰, Erin Robinson³¹, and Kaylin Bugbee³²

¹ Ge Peng, North Carolina Institute for Climate Studies, North Carolina State University, Asheville, NC, USA; gpeng@ncsu.edu; ORCID: 0000-0002-1986-9115

² Robert R. Downs, Center for International Earth Science Information Network (CIESIN), Columbia University, Palisades, NY, USA; rdowns@ciesin.columbia.edu; ORCID: 0000-0002-8595-5134

³ Carlo Lacagnina, Barcelona Supercomputing Center (BSC), Spain; carlo.lacagnina@bsc.es

⁴ Hampapuram K. Ramapriyan, Science Systems and Applications, Inc., Lanham, MD, USA and NASA Goddard Space Flight Center, Greenbelt, MD, USA; hampapuram.ramapriyan@ssaihq.com; ORCID: 0000-0002-8425-8943

⁵ Ivana Ivánová, Curtin University, Australia, ivana.ivanova@curtin.edu.au; ORCID: 0000-0001-6836-3463

⁶ David F. Moroni, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA; david.f.moroni@jpl.nasa.gov; ORCID: 0000-0003-2994-557X.

⁷ Yaxing Wei, Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA; weiy@ornl.gov; ORCID: 0000-0001-6924-0078.

⁸ Larnicol Gilles, Barcelona Supercomputing Center (BSC)/Magellium, Spain/France; gilles.larnicol@magellium.fr

⁹ Lesley Wyborn, National Computational Infrastructure, Australian National University, ACT, AUS; lesley.wyborn@anu.edu.au; ORCID: 0000-0001-5976-4943

¹⁰ Mitchell Goldberg, National Environmental Satellite, Data, and Information Service (NESDIS), Silver Spring, MD, USA; mitch.goldberg@noaa.gov

¹¹ Jörg Schulz, European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), Darmstadt, Germany; joerg.schulz@eumetsat.int

¹² Irina Bastrakova, Geoscience Australia, ACT, AUS, irina.bastrakova@ga.gov.au; ORCID: 0000-0002-4643-7289

¹³ Anette Ganske, Technische Informationsbibliothek (TIB), Hannover, Germany, Anette.Ganske@tib.eu; ORCID: 0000-0003-1043-4964

¹⁴ Lucy Bastin, Aston University, UK. l.bastin@aston.ac.uk; ORCID: 0000-0003-1321-0800

¹⁵ Siri Jodha Singh Khalsa; National Snow and Ice Data Center, University of Colorado, Boulder, CO, USA; sjskhalsa@gmail.com; ORCID: 0000-0001-9217-5550

- ¹⁶ Mingfang Wu, Australian Research Data Commons, Melbourne, Australia.
mingfang.wu@ardc.edu.au, ORCID: 0000-0003-1206-3431
- ¹⁷ Chung-Lin Shie, University of Maryland at Baltimore County, Baltimore, MD, USA and
NASA Goddard Space Flight Center, Greenbelt, MD, USA; clshie@umbc.edu and chung-lin.shie-1@nasa.gov, ORCID: 0000-0002-1115-1029
- ¹⁸ Nancy Ritchey, NOAA's National Centers for Environmental Information (NCEI), Asheville,
NC, USA; nancy.ritchey@noaa.gov; ORCID: 0000-0003-3939-6287
- ¹⁹ Dave Jones, StormCenter Communications | GeoCollaborate, Halethorpe, MD, USA,
dave@stormcenter.com
- ²⁰ Ted Habermann, Metadata Game Changers, Boulder, CO, USA; ted.habermann@gmail.com,
ORCID: 0000-0003-3585-6733
- ²¹ Christina Lief, ret. NOAA's National Centers for Environmental Information (NCEI),
Asheville, NC, USA; christina.lief@gmail.com
- ²² Iolanda Maggio, Rhea GROUP, La Piramide, Via di Grotte Portella 6/8, 00044 Frascati, Italy,
i.maggio@rheagroup.com, ORCID: 0000-0001-9409-1289
- ²³ Mirko Albani, European Space Agency, Frascati, Italy, mirko.albani@esa.int
- ²⁴ Shelley Stall, American Geophysical Union, Washington DC. USA; ssall@agu.org, ORCID:
0000-0003-2926-8353
- ²⁵ Lihang Zhou, National Environmental Satellite, Data, and Information Service (NESDIS),
Silver Spring, MD, USA; lihang.zhou@noaa.gov
- ²⁶ Marie Dré villon, Mercator Ocean International, France, mdrevillon@mercator-ocean.fr,
ORCID: 0000-0003-4586-8957
- ²⁷ Sarah Champion, North Carolina Institute for Climate Studies, North Carolina State
University, Asheville, NC, USA; smchampi@ncsu.edu; ORCID: 0000-0002-5080-6286
- ²⁸ C. Sophie Hou, Ronin Institute, USA, sophie.hou@ronininstitute.org, ORCID: [0000-0002-8087-1775](https://orcid.org/0000-0002-8087-1775)
- ²⁹ Francisco Doblas-Reyes, ICREA and Barcelona Supercomputing Center (BSC), Spain,
francisco.doblas-reyes@bsc.es, ORCID: 0000-0002-6622-4280
- ³⁰ Kerstin Lehnert, Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY,
USA; lehnert@ldeo.columbia.edu; ORCID: 0000-0001-7036-1977
- ³¹ Erin Robinson, Metadata Game Changers, Boulder, CO, USA; erinmr@gmail.com; ORCID:
0000-0001-9998-0114
- ³² Kaylin Bugbee, NASA Marshall Space Flight Center, Huntsville, AL, USA;
Kaylin.m.bugbee@nasa.gov; ORCID: 0000-0001-6733-5698

Corresponding author's email address: Ge Peng; gpeng@ncsu.edu

ABSTRACT

Knowledge about the quality of data and metadata is important to support informed decisions on the (re)use of individual datasets and is an essential part of the ecosystem that supports open science. Quality assessments reflect the reliability and usability of data and need to be consistently curated, fully traceable, and adequately documented, as these are crucial for sound decision- and policy-making efforts that rely on data. Quality assessments also need to be consistently represented and readily integrated across systems and tools to allow for improved sharing of information on quality at the dataset level for individual quality attribute or dimension. Although the need for assessing the quality of data and associated information is well recognized, methodologies for an evaluation framework and presentation of resultant quality information to end users may not have been comprehensively addressed within and across disciplines. Global interdisciplinary domain experts have come together to systematically explore needs, challenges and impacts of consistently curating and representing quality information through the entire lifecycle of a dataset. This paper describes the findings, calls for community action to develop practical guidelines, and outlines community recommendations for developing such guidelines. Community practical guidelines will allow for global access and harmonization of quality information at the level of individual Earth science datasets and support open science.

Keywords: Data Quality, Quality Dimension, Earth Science Information, Interoperability, FAIR, Stewardship

1. Needs for Curating and Sharing Dataset Quality Information

In this paper, data are representations of observations, objects, or other entities and can refer to anything that is collected, observed, generated or derived, and used as a basis for hypothesis testing, reasoning, discussion, or calculation. Observed data include in-situ and remotely sensed measurements. In-situ measurements can be from weather stations, rain gauges, buoys, or autonomous vehicles/vessels, while remotely sensed data can be from satellites or unmanned aircraft systems. Generated data can be results from a numerical model (e.g., a climate model) or a statistical model (e.g., a linear regression model). Model data can be analyses, predictions, or projections. Derived data can be produced from raw measurements or other products.

Dataset refers to an identifiable collection of physical records, a digital rendition of factual materials, or a product of a given version of an algorithm/model or experiment, often referred to as a collection. A dataset may contain one or many physical samples or data files or records in a database in an identical format, having the same geophysical variable(s) and product specification(s), such as the geospatial location or spatial grid. If more than one sample, file or records are included in a dataset, each may be referred to as a granule.

Dataset quality information consists of information about the quality of data, metadata and documentation. Documentation can include descriptions of measurement methods and instruments, software, provenance, as well as that on the state of practices, workflows, frameworks, tools, and systems associated with the dataset and production, data and quality management, data services and usage, customer support and user engagement. To enable decision- and policy-making processes, dataset quality information should be captured in the

metadata and should be a part of the ecosystem that supports open science. Although the importance of access to quality information is well recognized, methodologies for an evaluation framework and presentation of resultant quality information to end users may not have been comprehensively addressed, especially in an operational environment. Access to this information is especially important for enabling data to be findable, accessible, interoperable, and reusable (FAIR) (Wilkinson et al. 2016).

Considering that data quality can be affected by activities that are conducted throughout the data lifecycle, a data lifecycle approach to data quality assessment is necessary. Similarly, a data lifecycle approach to data quality assessment can facilitate effective recording of data quality information during various data lifecycle activities. The details of such information could be lost during later stages in the data lifecycle if they are not recorded when the data quality events or assessments occurred. Moreover, managing quality through the entire dataset lifecycle is imperative for ensuring that the information and knowledge gained are not contaminated by inaccurate or corrupt data, as well as for facilitating accurate uncertainty estimates in the derived analyses. The value of lifecycle approaches to data quality has been recognized for various kinds of data, including remote sensing observations (Barsi et al. 2019), health services (Kahn et al. 2015), and health and biomedical citizen science (Borda et al. 2020). Data lifecycle approaches to quality assessment also could be informed by lifecycle approaches to software quality (Lenhardt et al. 2014).

Technologies such as machine learning (ML) and artificial intelligence (AI) are becoming increasingly useful and effective as tools to uncover or gain new knowledge from various domains of Earth science data (See an overview by Maskey et al. 2020). However, any sound analysis needs to build on reliable data. Thus, it is becoming increasingly important to gain access to verifiable and consistently quality-controlled data as well as information on dataset quality.

Therefore, it is crucial to consistently record, curate, and represent quality information of individual datasets, and make it readily available and integratable. However, dataset quality information is not routinely curated and much less represented in a human- and machine-readable manner, despite the fact that international standards for describing the quality of geographic data have been in place since 2003 (e.g., ISO 19157: 2013; ISO 19115:2014).

2. Multi-Dimensionality of Dataset Quality

A dataset is associated with a number of distinct quality attributes or characteristics. For example, from a data consumer perspective, over 179 individual data quality attributes were identified by a survey analyzed in Wang and Strong (1996), such as accuracy, correctness, freedom from bias, etc.

Dataset quality attributes can be categorized into different perspectives or dimensions with emphasis on certain quality attributes (e.g., Wang and Strong 1996; Lee et al. 2002; Wilkinson et al. 2016; Redman 1996; Ramapriyan et al. 2017). For instance, Wang and Strong (1996) prioritized the 179 quality attributes down to 15, and categorized them into four perspectives:

- i) Intrinsic (accuracy, objectivity, believability, reputation);
- ii) Contextual (relevance, value-added, timeliness, completeness, appropriate amount of data);
- iii) Representational (interpretability, ease of understanding, concise representation and representational consistency); and
- iv) Accessibility (accessibility, access security).

While Wilkinson et al. (2016) focused on findability, accessibility, interoperability and reusability as four quality aspects of machine-based data sharing principles, Redman (1996) defined accuracy, completeness, consistency, and currency as four quality dimensions of data values. These quality attributes are explicitly listed to demonstrate that the quality attributes can differ greatly depending on the different perspectives. On the other hand, based on the full dataset lifecycle, Ramapriyan et al. (2017) categorized quality attributes into four quality dimensions: science, product, stewardship, and service.

Assessment models are developed in Earth science to measure the maturity of different quality perspectives and dimensions at the dataset or collection level (see an overview by Peng 2018). However, there are very limited and sparse actionable guidelines on how to curate and represent dataset quality information in a way that is consistent with FAIR principles for improved sharing.

Currently dataset quality information, when available, tends to focus on scientific quality and is published in science journals that are text-based and cannot be readily integrated into data management and stewardship processes or across different systems. For example, while uncertainty information is considered valuable, there are many approaches toward quantifying, characterizing, disseminating and interpreting that information (Moroni et al. 2019). In addition, dataset quality information needs to be readily understandable by both machine and human end users, including those who plan to use the described data as well as by those who are trying to determine whether the data are appropriate for their intended use. Therefore, we need to converge towards more harmonized approaches for curating dataset quality information in a way that is consistent with FAIR guiding principles to enable global access of this information.

3. Potential Benefits of FAIR Dataset Quality Information

The FAIR data guiding principles emphasize the importance of data sharing by ensuring that data and data descriptions (metadata) are Findable, Accessible, Interoperable, and Reusable (Wilkinson et al., 2016). Findable data is discovered and understood by a search agent (e.g., search engine or human user). Accessible data is rendered and used by machine and human end users via standard protocols within use and access constraints. Interoperable data can be readily used in conjunction with other data products or services and can also be integrated with other data to create new data products or services. Reusable data can be used, under the proper usage license, by diverse audiences beyond those who were initially envisioned as potential users by the original data producer(s). Since the inception, the FAIR data principles have been adopted by global entities and have had a major impact in prompting data sharing and reuse globally (e.g.,

G20 Leaders 2016; Australia FAIR Access Working Group 2017; European Commission 2018; Mons 2018; U.S. Public Law 115-435 2019; CODATA 2019).

To consider and apply the FAIR guiding principles when curating and representing dataset quality information can help ensure the information is optimal for sharing and enabling global access. It also contributes to or improves the FAIRness of a dataset. For example, when data can be discovered based on information about certain quality attributes, the findability of the data is improved for users who need data that contain such attributes, and further the quality information supports users to assess the relevance of a discovered dataset to a research or operational need.

Consistently curating and representing dataset quality information by following the FAIR principles could eventually lead to standardization and therefore harmonization of the information. In addition, describing quality information using standard formats, schemas, and terminology improves the interoperability and reusability of the data.

4. Call for Global Community Guidelines

Since 1991, Earth science community, as a whole, has been sharing essential and high value data. For example, Member States of the World Meteorological Organization (WMO) have been exchanging basic weather, hydrological, and climate data through a series of WMO resolutions (WMO 1991; 1999; 2015). Guidelines on acquisition, quality assurance and control of meteorological station data were developed and improved for WMO Member States (e.g., WMO 1986; 2004; 2019a). For the first time, WMO (2019b) issued a regulatory technical recommendation on managing climate data including evaluating the stewardship maturity of global datasets of climate data. Recognizing the need to address ever increasing data volume and variety of data types across disciplines, WMO (2019c) called for one unified data policy to support global environmental data sharing and open science. United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM) developed a strategic framework with recommendations for a geospatial community with a strong focus on data, data quality and standards. These recommendations will help countries make their geospatial data and information reliable, accessible, and easy to use (UN-GGIM 2018; 2019)

Sharing and reusing data is a big challenge but significant progress has been made collectively by the Earth science community over the last few decades. Success stories include greatly enhanced accessibility, usability, and interoperability of observational and climate model data through coordinated community efforts such as the Observations for Model Intercomparisons Project (Obs4MIP; Ferraro 2015) and the Coupled Model Intercomparison Projects (CMIP; Eyring 2016). Because CMIP data are distributed over several local data repositories, the concept of a distributed and federated quality assessment procedure was developed. Since CMIP5 (Taylor 2012) the data are quality checked (Stockhause et al. 2012) and CMIP6 data are published with persistent identifiers such as Digital Object Identifiers (DOIs) (Stockhause and Lautenschlager 2017).

Sharing and reusing dataset quality information is an even bigger challenge that requires community effort. Issuing a call to the community for such an effort is the first step towards achieving that goal.

To explore the needs for dataset quality information, approaches and challenges for consistently evaluating and representing the quality information, a one-day virtual workshop was held on Monday July 13, 2020, followed by a report-out session on Wednesday July 22, 2020, during the virtual Earth Science Information Partners (ESIP) 2020 Summer Meeting (SM20), July 14–24, 2020. The pre-ESIP workshop was sponsored by ESIP and co-organized by the ESIP Information Quality Cluster (IQC) and the Barcelona Supercomputing Center (BSC) Evaluation and Quality Control (EQC) team, in collaboration with the Australia/New Zealand Data Quality Interest Group (AU/NZ DQIG). An additional community engagement event was carried out by AU/NZ DQIG prior to the pre-ESIP workshop (Ivánová 2020). Additional information can be found in the workshop report by Peng et al. (2020a).

A total of 14 presentations from organizations across 9 countries were given during the two live sessions of the workshop and the ESIP SM20 report-out session. Presenters summarized the data quality assessment approaches that have been developed and/or adopted by organizations representing the scope of national and international Earth science data producers, data management stewardship programs, data and service centers as well as data and information providers. Participants also included data users from private sectors.

The needs, challenges and approaches of consistently curating digital Earth science data and products were discussed during the live sessions as well as online during the following weeks. The scope and path forward for developing community guidelines for Earth science dataset quality information were also communicated and discussed. Key takeaways from the discussions are listed and described in the following subsections.

4.1. Built by the Global Community

While needs are strong for practical community guidelines for curating data quality information, currently such information is very limited and sparse. Therefore, to ensure the relevance of such guidelines, it is crucial for the guidelines to be developed through a coordinated effort via an iterative process, leveraging the experiences and expertise of an international team of interdisciplinary subject matter experts, and community best practices. Domain experts from 9 countries across America, Europe and Oceania have participated in the workshop. An ESIP working group has been formed to develop practical community guidelines.

4.2. Quality-Attribute Agnostic Guidelines

As characterized in Section 2, assessing dataset quality is a multi-dimensional problem. The selection of the relevant attributes is context-dependent and it leads to different categorizations and practical dimensions (Redman 1996). The complexity exists even for a single quality attribute within one discipline in terms of its definition and how to measure and represent it consistently. An example of this is data uncertainty as explored by Moroni et al. (2019). Lessons

learned from various data curation initiatives indicate that an alternative approach is to equip data consumers with readily available information that is consistently captured, in a way that can be easily understood by end-users and integrated across tools and systems, regardless of the quality attribute and the assessment approach.

4.3. Community Consented Terminology for Enhanced Interoperability

For a given quality dimension, consistency in various components and attributes across entities in each component, namely, semantic and structured consistency as defined by Redman (1996), is important to generating machine-actionable quality information. Common terminology is necessary for integrating data and information across workflows, tools and systems, as well as for curating and representing dataset quality information. Moreover, terms should be defined for all stages of a dataset lifecycle.

4.4. Continuous Engagements with Stakeholders

The guidelines are being developed through an iterative process to allow for feedback from the community and all stakeholders, including those who contribute to the acquisition, curation, dissemination, and application of data. Continuous community engagements are planned by means of informal updates to various working groups and formal presentations to the targeted stakeholders, including those to the American Geophysical Union (AGU) community at the 2020 AGU Fall Meeting (Peng et al. 2020c), the ESIP community at its winter meeting in January 2021 (Peng et al. 2021), and the European Geosciences Union community at its general assembly in April 2021 (Lacagnina et al. 2021). Additionally, reviews of the guidelines draft are planned prior to its being baselined. The guidelines document will be a living document to allow for evolving community requirements and best practices.

4.5. Long-Term Sustainability

It has been suggested by the participants that long-term sustainability should be planned for such community guidelines. Once baselined, the guidelines will be publicly accessible by several means including at the ESIP WiKi website (https://wiki.esipfed.org/FAIR_Dataset_Quality_Information) and via an open science platform such as ESIP Figshare (esip.figshare.com) or Open Science Framework (osf.io) that provides a globally unique and persistent identifier with searching and sharing capability for the document. To ensure currency and timeliness of the guidelines, curation and revision planning is essential. These approaches will help with maintaining relevancy and long-term access to the guidelines.

5. Summary

Recognizing the needs, challenges, and impacts of sharing information on quality at a dataset/collection level, global interdisciplinary domain experts have come together and called for community guidelines towards global access and harmonization of information on quality of individual Earth science datasets.

The quality-attribute agnostic guidelines will be developed under community effort via an iterative process, leveraging the experiences and expertise of an international team of interdisciplinary subject matter experts and best practices, to address community needs and challenges summarized in Peng et al. (2020a). Description of what quality attributes are, how the attributes are assessed, and what assessment approaches are utilized, should be included in relevant metadata or a document, preferably in a consistent way for transparency and enhanced usability. The guidelines will call for a machine-actionable mechanism to represent assessed results for enhanced interoperability across systems and disciplines.

By adopting the FAIR principles, the guidelines will help to ensure global access of the dataset quality information. Effective sharing of structured dataset quality information will help to move towards its global harmonization, which in turn will support (re)use of the data by both human and machine end users and therefore further enhance the value of the data.

An ESIP working group has been formed and its membership is open to any domain expert who is willing to contribute to the development effort. Development of the guidelines has begun and the outcomes will be reported in a follow-up paper. The guidelines will be primarily developed for the Earth Science community. They will, however, be general enough so that other disciplines can readily adapt them, which will further promote global access and harmonization of dataset quality information.

Acknowledgement

The virtual pre-ESIP workshop held on July 13, 2020 was sponsored by ESIP with support from the National Aeronautics and Space Administration (NASA), the National Oceanic and Atmospheric Administration (NOAA) and the United States Geological Survey (USGS). The technological and infrastructural support during the preparation and conduct of the workshop was invaluable. In particular, we thank Megan Carter for supporting us throughout the workshop and providing helpful advice during the planning stage of the virtual workshop, and ESIP community fellow Alexis Garretson for supporting the ESIP SM20 report-out session. We thank all participants for attending the pre-ESIP workshop and the ESIP SM20 session and contributing to productive discussions during the live sessions and the two weeks of the ESIP SM20 period. Portions of this work have been extracted from Peng et al. (2020a), which reported on the workshop and the ESIP SM20 report-out session. The Australian participants would like to acknowledge the support of the Australian Research Data Commons.

References

- Australia FAIR Access Working Group 2017 Policy Statement on FAIR Access to Australia's Research Outputs. Version: Jan 2017. Available at: <https://www.fair-access.net.au/fair-statement>
- Barsi, Á, Kugler, Z, Juhász, A, Szabó, G, Batini, C, Abdulmuttalib, H, Huang, G, and Shen, H 2019 Remote sensing data quality model: from data sources to lifecycle phases. *International*

356 *Journal of Image and Data Fusion*, 10(4):280-99.
 357 DOI:<https://doi.org/10.1080/19479832.2019.1625977>
 358 Borda A, Gray K, and Fu Y 2020 Research data management in health and biomedical citizen
 359 science: practices and prospects. *JAMIA Open*, 3(1):113-25.
 360 DOI:<https://doi.org/10.1093/jamiaopen/ooz052>
 361 CODATA 2019 The Beijing Declaration on Research Data. Version: 7 November 2019.
 362 Available at: <http://www.codata.org/uploads/Beijing%20Declaration-19-11-07-FINAL.pdf>
 363 European Commission 2018 Turning FAIR into reality - Final Report and Action Plan from the
 364 European Commission Expert Group on FAIR data, European Commission: Brussels,
 365 DOI:<https://doi.org/10.2777/1524>
 366 Eyring, V, Bony S, Meehl G A, Senior C A, Stevens B, Stouffer R J and Taylor, K E 2016
 367 Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) Experimental
 368 Design and Organization. *Geoscientific Model Development* 9(5), 1937–1958.
 369 DOI:<https://doi.org/10.5194/gmd-9-1937-2016>
 370 Ferraro, R, Waliser, D E, Gleckler, P, Taylor, K E, and Eyring, V 2015 Evolving Obs4MIPs to
 371 Support Phase 6 of the Coupled Model Intercomparison Project (CMIP6). *Bull. Amer.*
 372 *Meteor. Soc.*, 96, ES131–ES133, DOI:<https://doi.org/10.1175/BAMS-D-14-00216.1>.
 373 G20 Leaders 2016 G20 Leaders’ Communique Hangzhou Summit. Version: 5 September 2016.
 374 Available at: https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_16_2967
 375 ISO 19115: 2014 Geographic Information—Metadata. Version: 2014-04. *International*
 376 *Organization for Standardization*. Geneva, Switzerland. Available at:
 377 <https://www.iso.org/standard/53798.html>
 378 ISO 19157: 2013 Geographic information—Data quality. Version: 2013-1. *International*
 379 *Organization for Standardization*. Geneva, Switzerland. Available at:
 380 <https://www.iso.org/standard/32575.html>
 381 Ivánová, I and AU/NZ DQIG 2020 Australia/New Zealand Data Quality Interest Group. *Pre-*
 382 *ESIP Workshop*, 13 July 2020, Virtual. DOI:<https://doi.org/10.6084/m9.figshare.12711944>
 383 Kahn, M G, Brown, J S, Chun, A T, Davidson, B N, Meeker, D, Ryan, P B, Schilling, L M,
 384 Weiskopf, N G, Williams, A E, and Zozus, M N 2015 Transparent reporting of data quality
 385 in distributed data networks. *Egems*, 3(1). DOI:<https://doi.org/10.13063/2327-9214.1052>
 386 Lacagnina, C, Peng, G, Downs, R R, Ramapriyan, H, Ivanova, I, Moroni, D F, Larnicol, G., Wei,
 387 Y, Bastin, L, Ritchey, N A, Wyborn, L A, Shie, C-L, Habermann, T, Ganske, A, Champion,
 388 S M, Wu, M, Bastrakova, I, Jones, D, and Berg-Cross, G 2021 Towards Developing
 389 Community Guidelines for Sharing and Reusing Quality Information of Earth Science
 390 Datasets. Abstract. EGU 2021 General Assembly.
 391 Lee, Y W, Strong, D M, Khan, B K, and Wang, R Y 2002 AIMQ: a methodology for
 392 information quality assessment, *Information & Management*, 40, 133-146.
 393 Lenhardt W, Ahalt S, Blanton B, Christopherson L, and Idaszak R 2014 Data management
 394 lifecycle and software lifecycle management in the context of conducting science. *Journal of*
 395 *Open Research Software*, 2(1). DOI:<http://doi.org/10.5334/jors.ax>

- Maskey, M, Alemohammad, H, Murphy, K J, and Ramachandran, R 2020 Advancing AI for Earth Science: A Data Systems Perspective. *EOS*, 101.
DOI:<https://doi.org/10.1029/2020EO151245>
- Mons, B 2018: Data Stewardship for open science: implementing FAIR principles. 1st Edition. Chapman and Hall/CRC Press, *Taylor & Francis*, New York. 244 pp. Available at: <https://www.taylorfrancis.com/books/9781315380711>
- Moroni, D F, Ramapriyan, H, Peng, G, Hobbs, J, Goldstein, J C, Downs, R R, Wolfe, R, Shie, C-L, Merchant, C J, Bourassa, M, Matthews, J L, Cornillon, P, Bastin, L, Kehoe, K, Smith, B, Privette, J L, Subramanian, A C, Brown, O, and Ivánová, I 2019 Understanding the Various Perspectives of Earth Science Observational Data Uncertainty. *Figshare*.
DOI:<https://doi.org/10.6084/m9.figshare.10271450>
- Peng, G 2018 The state of assessing data stewardship maturity – an overview. *Data Science Journal*, 17(0), DOI:<https://doi.org/10.5334/dsj-2018-007>
- Peng, G, Lacagnina, C, Downs, R R, Ivanova, I, Moroni, D F, Ramapriyan, H, Wei, Y, and Larnicol, G 2020a Laying the Groundwork for Developing International Community Guidelines to Effectively Share and Reuse Digital Data Quality Information – Case Statement, Workshop Summary Report, and Path Forward. *Open Science Framework*, DOI:<https://doi.org/10.31219/osf.io/75b92>
- Peng, G, Lacagnina, C, Downs, R R, Ramapriyan, H, Ivanova, I, Moroni, D F, Larnicol, G., Wei, Y, Bastin, L, Ritchey, N A, Wyborn, L A, Shie, C-L, Habermann, T, Ganske, A, Champion, S M, Wu, M, Bastrakova, I, Jones, D, and Berg-Cross, G 2020b Towards Developing Community Guidelines for Sharing and Reuse of Digital Data Quality Information. *AGU 2020 Fall Meeting*. Abstract 674372. Available at: <https://agu.confex.com/agu/fm20/meetingapp.cgi/Paper/674372>
- Peng, G, Lacagnina, C, Downs, R R, Ramapriyan, H, Ivanova, I, Moroni, D F, Larnicol, G., Wei, Y, Bastin, L, Ritchey, N A, Wyborn, L A, Shie, C-L, Habermann, T, Ganske, A, Champion, S M, Wu, M, Bastrakova, I, Jones, D, Hou, C-Y, and Berg-Cross, G 2021 An update on a community effort to prompt global sharing of dataset quality information. *ESIP 2021 Winter Meeting*. Virtual.
- Redman, C T 1996 Data quality of the information age. *Artech House*, Boston. 303 pp.
- Ramapriyan, H, Peng, G, Moroni, D, and Shie, C-L 2017 Ensuring and Improving Information Quality for Earth Science Data and Products. *D-Lib Magazine*, 23,
DOI:<https://doi.org/10.1045/july2017-ramapriyan>
- Stockhause, M, Höck, H, Toussaint, F, and Lautenschlager, M 2012 Quality assessment concept of the World Data Center for Climate and its application to CMIP5 data. *Geosci. Model Dev.* 5, DOI:<https://doi.org/10.5194/gmd-5-1023-2012>
- Stockhause, M and Lautenschlager, M 2017 CMIP6 Data Citation of Evolving Data. *Data Science Journal*, 16, DOI: <http://doi.org/10.5334/dsj-2017-030>

- Taylor, K E, Stouffer, R J, and Meehl, G A 2012 An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society* 93(4), 485–498.
DOI:<https://doi.org/10.1175/BAMS-D-11-00094.1>
- UN-GGIM 2018 Integrated Geospatial Information Framework Part 1. United Nations Committee of Experts for Global Geospatial Information Management,
<https://ggim.un.org/IGIF/part1.cshtml>
- UN-GGIM 2019 Integrated Geospatial Information Framework Part 2. United Nations Committee of Experts for Global Geospatial Information Management,
<https://ggim.un.org/IGIF/part2.cshtml>
- U.S. Public Law 115-435 2019 Foundations for Evidence-Based Policymaking Act of 2018. Title II OPEN Government Data Act. Version: 14 January 2019. 115th U.S. Congress. Available at: <https://www.govinfo.gov/content/pkg/PLAW-115publ435/pdf/PLAW-115publ435.pdf>
- Wang, R Y and Strong, D M 1996 Beyond Accuracy: What Data Quality Means to Consumers. *Journal of Management Information Systems*, 12(4),
DOI:<https://doi.org/10.1080/07421222.1996.11518099>
- Wilkinson, M D, Dumontier, M, Aalbersberg, I J, Appleton, G, Axton, M, Baak, A, and others 2016 The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018, DOI:<https://doi.org/10.1038/sdata.2016.18>
- WMO 1986 Guidelines On The Quality Control Of Surface Climatological Data. WMO/TD-No. 111. *World Meteorological Organization*, Geneva, Switzerland. Available at: https://library.wmo.int/doc_num.php?explnum_id=9205
- WMO 1991 Resolution 40 (Cg-XII) – WMO policy and practice for the exchange of meteorological and related data and products including guidelines on relationships in commercial meteorological activities. WMO-No. 827. *World Meteorological Organization*, Geneva, Switzerland. Available at: https://www.wmo.int/pages/prog/hwrr/documents/wmo_827_enCG-XII-Res40.pdf
- WMO 1999 Resolution 25 (Cg-XIII) – Exchange of Hydrological Data and Products. *World Meteorological Organization*, Geneva, Switzerland. Available at: https://www.wmo.int/pages/prog/hwrr/documents/Resolution_25.pdf
- WMO 2004 Guidelines on Quality Control Procedures for Data from Automatic Weather Stations. Expert Team on Surface Technology and Measurement Techniques, *World Meteorological Organization*, Geneva, Switzerland. Available at: [https://www.wmo.int/pages/prog/www/IMOP/meetings/Surface/ET-STMT1_Geneva2004/Doc6.1\(2\).pdf](https://www.wmo.int/pages/prog/www/IMOP/meetings/Surface/ET-STMT1_Geneva2004/Doc6.1(2).pdf)
- WMO 2015 Resolution 60 (Cg-17) – WMO Policy for the International Exchange of Climate Data and Products to Support the Implementation of the Global Framework for Climate Services. *World Meteorological Organization*, Geneva, Switzerland. Available at: https://library.wmo.int/doc_num.php?explnum_id=4192

473 WMO 2019a WMO Guidelines on Surface Station Data Quality Assurance for Climate
474 Applications. Draft: April 5, 2019. World Meteorological Organization, Geneva,
475 Switzerland. Available at: [https://www.wmo.int/pages/prog/wcp/wcdmp/hq-](https://www.wmo.int/pages/prog/wcp/wcdmp/hq-gdmfc/documents/QC_QAguidelines-April2019.pdf)
476 [gdmfc/documents/QC_QAguidelines-April2019.pdf](https://www.wmo.int/pages/prog/wcp/wcdmp/hq-gdmfc/documents/QC_QAguidelines-April2019.pdf)
477 WMO 2019b Manual on the high-quality global data management framework for climate.
478 WMO-No. 1238. *World Meteorological Organization*, Geneva, Switzerland. 43 pp.
479 Available at: https://library.wmo.int/doc_num.php?explnum_id=10197
480 WMO 2019c WMO data policy statement. Draft 1.0. *Study Group on Data Issues and Policies.*
481 *WMO Data Conference*. 16–19 November 2020, Virtual. Available at:
482 [https://meetings.wmo.int/WMO-Data-](https://meetings.wmo.int/WMO-Data-Conference/Documents/Flyer%20for%20Res.42%2011%2015.pdf)
483 [Conference/Documents/Flyer%20for%20Res.42%2011%2015.pdf](https://meetings.wmo.int/WMO-Data-Conference/Documents/Flyer%20for%20Res.42%2011%2015.pdf)