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

Ge Peng^{1,2*}, Carlo Lacagnina^{3*}, Robert R. Downs^{2,4}, Ivana Ivanova⁵, David F. Moroni^{2,6}, Hampapuram Ramapriyan^{2,7}, Yaxing Wei^{2,8}, and Gilles Larnicol³

¹ Cooperative Institute for Satellite Earth System Studies (CISESS) between NOAA and North Carolina State University, at National Centers for Environmental Information (NCEI), Asheville, NC, USA

² Information Quality Cluster (IQC) of Earth Science Information Partners (ESIP), USA

³ Barcelona Supercomputing Center (BSC), Barcelona, Spain

⁴ Columbia University, Center for International Earth Science Information Network (CIESIN) and NASA Socioeconomic Data and Applications Center (SEDAC), Palisades, NY, USA

⁵ Curtin University, Perth, Australia

⁶ NASA Jet Propulsion Laboratory (JPL), Pasadena, CA, USA

⁷ Science Systems and Applications, Inc. (SSAI) and NASA Goddard Space Flight Center, Greenbelt, MD, USA

⁸ NASA Oak Ridge National Laboratory (ORNL), Oak Ridge, TN, USA

* Corresponding Authors: gpeng@ncsu.edu; carlo.lacagnina@bsc.es

1. Case Statement

Knowledge about the quality of data and associated information is important to support decision-making for taking solid, informed actions. Assessment of the data quality is key for ensuring that the available information is creditable and for establishing relations of trust between the data provider and various downstream users. Quality assessments inform the reliability and usability of the data, and quality information needs to be consistently curated, fully traceable, adequately documented and able to support users to address their specific needs. This is especially important for data used to support policy-making and enabling data to be reusable, in addition to being findable, accessible, and interoperable, as defined by the FAIR data principles (Wilkinson et al. 2016). Although the importance of access to quality-assured data is well recognized, methodologies of an evaluation framework and presentation of resultant quality information to end users may not have been comprehensively addressed, especially in an operational environment.

The Cloud offers both opportunities and challenges. These can be observed when testing the ability to integrate data across different systems, i.e., interoperability. The Cloud allows users to discover, access, subset/integrate, and analyze large volumes and varieties of data in a more efficient, scalable manner compared with legacy, on-premises data storage and distribution

technologies. Determining how to establish trust and enable the correct use of the data is crucial to data sharing (Digital Science et al. 2019).

Technologies such as machine learning (ML) and artificial intelligence (AI) are becoming increasingly useful as tools to uncover or gain new knowledge from data. However, sound analysis needs to build on reliable data. Thus, it is becoming increasingly important to gain access to verifiable and consistently quality-controlled data when making data-driven decisions. Managing quality through the entire data life cycle is imperative for ensuring that the information and knowledge gained are not contaminated by inaccurate or corrupted data and for facilitating accurate uncertainty estimates in the derived analyses.

Therefore, it is crucial to consistently curate and represent quality information of data and associated information and make it readily available and integratable. However, dataset quality information is not routinely curated in a human- and machine-readable manner, despite the fact that the ISO 19157 standards for describing the quality of geographic data have been in place since 2013 (ISO 19157 2013). This is because there are many challenges, including but not limited to the following:

- Dataset quality information is multi-dimensional, i.e., science, product, stewardship, and services (Ramapriyan et al. 2017), with many quality attributes (Wang and Strong 1996) and many different maturity assessment models (see Peng 2018 for an overview; also see references in Table 1). Information about dataset quality traverses different knowledge domains and curating it requires cross-disciplinary collaborations.
- Currently, dataset quality information, when available, tends to focus on scientific quality and is published in science journals that are free-text-based and are not readily integratable 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).
- Dataset quality information needs to be understandable by 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.

Community guidelines for curating and representing dataset quality information in a consistent fashion and in line with the FAIR principles can help science data centers, repositories, data producers, publishers, data managers, and stewards establish trustworthiness and maximize reuse and value of their data. Developing such community guidelines requires a community-wide, cross-disciplinary effort to set a solid foundation for wide implementation. This paper formally introduces the case statement to provide the background for initiating the development effort.

To kick off the development, a workshop was convened with international subject matter experts (SMEs) sharing different experiences and approaches in curating and representing dataset quality information. The workshop is described in the next section along with outcomes and key takeaways, followed by the scope of the guidelines and a development timeline,

leveraging discussions among the participants. The intended use and users along with potential benefits of such guidelines are then described.

2. Workshop Summary Report

2.1. An Outline of the Workshop

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).

ESIP, which was founded in 1998, is mostly supported by USA Earth science governmental agencies, namely, the National Aeronautics and Space Administration (NASA), the National Oceanic and Atmospheric Administration (NOAA), and the United States Geological Survey (USGS), and has over 150 national and international partner organizations as members. ESIP IQC aims to facilitate cross-disciplinary collaborations to assess various aspects of Earth science data and information quality and provides best practices to the community. The BSC EQC team supports the EQC function of Copernicus Climate Change Service (C3S) Climate Data Store via the C3S_512 contract. C3S is one of six services of the European Union's Earth observation programme. The AU/NZ DQIG is a forum for AU/NZ data providers, repository operators and data consumers. It was founded in late 2019 by the Australian Research Data Commons (ARDC), Curtin University and the Australian National University (ANU) and facilitated by ARDC.

The purpose of this workshop, in summary, was to bring together national and international SMEs on the quality of datasets or data products to initiate the development of community guidelines for consistently curating and representing dataset quality information in a way that is in line with the FAIR principles as defined by Wilkinson et al. (2016).

About 80 ESIP and invited international SMEs, affiliated with over 40 private, academic and governmental institutions from 9 countries across America, Europe and Oceania (Figure 1) registered for the workshop. About 67% of the registrants were from the United States while the rest were from the European and AU/NZ regions (Figure 2; see Table A1 for a list of attendees who signed in and their affiliations). For many of the approximately 20 non-USA attendees, this was their first time to engage with the ESIP community.

There were two live 90-minute virtual workshop sessions to accommodate attendees from different time zones. More than 45 workshop registrants attended the first live session while approximately 25 attended the second. About 45 ESIP SM20 registrants attended the report-out session. Prior to this workshop, a mini-workshop was held by AU/NZ DQIG on July 6, 2020, where 57 had registered and 27 had participated actively.

Through presentations and discussions in and out of the live sessions during the workshop, invited speakers shared their knowledge to help attendees gauge the complexity and multi-dimensionality of dataset quality information. This knowledge exchange allowed attendees to

understand why Earth science organizations need to curate and represent data quality information throughout the entire life cycle of a dataset – from the data product design and production stage, through data and metadata curation stage for preservation and access, to data use stage by servicing data to consumers. It also helped attendees appreciate the challenges those organizations face, and learn the different approaches they take. (In this report, a dataset refers to a digital rendition of factual materials, a derived product, or model output. Dataset quality includes the quality of both the data product and the associated information. Associated information includes that captured in documents, e.g., scientific papers or reports, and/or as quality metadata in a dataset-level metadata record.)

In this section, we will describe briefly the context and goals of the workshop and SM20 session and summarize the outcomes and key takeaways.

2.2. The Context of the Workshop

Earth science produces a lot of digital data to support a variety of sectors of society and to facilitate informed decision-making. Different tools are made available to deal with big data, like cloud infrastructure and machine learning. All these need data to be interoperable, reliable and usable. This calls for dataset quality management processes that guarantee data reliability and accessibility to the acquired dataset quality information. Aspects of quality control, assurance and dissemination are the main pillars of dataset quality management processes. In the Earth sciences, dataset quality information has historically focused on scientific quality and is often published in scientific journals, which makes it difficult to make this information readily accessible and maintainable for data management in operational services. In recent years, different approaches are moving towards making dataset quality information available in a way that is both human and machine readable. These approaches build a solid foundation for us to develop a consensus for community guidelines.

2.3. The Goals of the Workshop

- Examine the needs for dataset quality information
- Explore approaches for evaluating dataset quality in an operational environment
- Discuss challenges for consistently curating and representing dataset quality information
- Define the needs and scope of community guidelines and path forward

2.4. The Goals of the ESIP SM20 Report-Out Session

- Communicate to the ESIP community about the Pre-ESIP Workshop and its outcomes
- Continue to explore needs, challenges, and approaches with invited presentations from workshop participants
- Facilitate open discussion among the ESIP community, workshop participants, and organizing committee members about the scope and desirable type of the community guidelines

2.5. Outline of Presentations

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 (additional details such as presentation titles, presenter name and affiliation(s) can be found in Table 1 and a short bio for presenters can be found in Appendix B). 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 data and information providers. Those organizations include:

- major international Earth science satellite programs and space agencies, such as the U.S. Joint Polar Satellite System (JPSS; Goldberg and Zhou 2020), the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT; Schulz 2020), and the European Space Agency (Albani and Maggio 2020)
- major international Earth science data centers and data and information stewardship programs, such as JPSS (Goldberg and Zhou 2020), U.S. NOAA National Centers for Environmental Information (NCEI; Ritchey 2020), U.S. NASA Distributed Active Archive Centers (DAACs; Wei et al. 2020), USGS (Hou et al. 2020), U.S. National Climate Assessment (Champion 2020), the Committee on Earth Observation Satellites (CEOS; Albani and Maggio 2020), and the Group on Earth Observations (GEOSS; Downs 2020)
- major organizations of international environmental data service providers and publishers, such as NCEI (Ritchey 2020), C3S (Lacagnina et al. 2020); Copernicus Marine Environment Monitoring Service (CMEMS; Dré villon et al. 2020), and the World Meteorological Organization (Lief 2020)
- international Earth science data partners such as ESIP (Peng, Ramapriyan & Moroni 2020) and Australian ARDC (Ivánová 2020)
- American Geophysical Union (AGU), which is a major international Earth science journal publisher and a key proponent for FAIR data (Stall 2020).

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 two weeks, within a collaborative Google Doc. The scope and path forward for developing community guidelines for Earth science dataset quality information were also communicated and discussed. The case statement, drafted by the workshop organizing committee (Peng et al. 2020), was reviewed by the participants during the workshop and during the ESIP SM20 period. The latest version is available and will be maintained at: bit.ly/FAIR-DQI-CaseStatement.

Figure 1: Countries and institutional logos of pre-ESIP workshop registrant affiliations. For demonstration only—logos are not necessarily the official version.

Figure 2: Country breakdowns of the pre-ESIP workshop registrants. The country codes follow that of ISO 3166 - ALPHA-2 (<https://www.iso.org/obp/ui/#search>), where US, UK, DE, ES, FR, AU, CH, IT, and NZ refers to the United State of America, United Kingdom, Germany, Spain, France, Australia, Switzerland, Italy, and New Zealand, respectively.

Table 1: A list of the presentation titles, affiliated organizations, and references

Presentation Title	Affiliated Organization(s)	Reference
ESIP Information Quality Cluster (IQC) & Multi-dimensions of data and information quality	ESIP IQC	Peng, Ramapriyan & Moroni (2020)
Evaluation and Quality Control Function of the C3S Climate Data Store (CDS)	Barcelona Supercomputing Center/Contractor to the Copernicus	Lacagnina, Doblas-Reyes, and C3S_512 Team (2020)
JPSS & Product Algorithm Maturity Matrix and Application	NOAA/NESDIS/JPSS (Joint Polar Satellite System)	Goldberg and Zhou (2020)
System Maturity and Application Performance for Climate Data Record	European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT)	Schulz (2020)
CEOS WGISS Data Management and Stewardship Maturity Matrix and Application at ESA	European Space Agency (ESA)	Albani and Maggio (2020)
NOAA/NCEI Data Stewardship Maturity Assessment	NOAA/NCEI (National Centers for Environmental Information)	Ritchey (2020)
Australia/New Zealand Data Quality Interest Group	Curtin University/Australian Research Data Commons (ARDC)	Ivánová (2020)
NASA ESDSWG Data Quality WG	National Aeronautics and Space Administration (NASA)	Wei et al. (2020)
GEOSS Data Management & Data Sharing Principles and TRUST - Implications for Information Quality	Columbia University and NASA SEDAC (Socioeconomic Data and Applications Center)	Downs (2020)
The High Quality Global Data Management Framework for Climate	World Meteorological Organization (WMO)	Lief et al. (2020)
Why FAIR?	American Geophysical Union (AGU)	Stall (2020)
Production of quality information at the Copernicus Marine Environment Monitoring Service (CMEMS)*	Mercator Ocean/Contractor to the Copernicus	Dré villon et al. (2020)
Information Quality and National Climate Assessment*	North Carolina State University (NCSU)	Champion (2020)

Interpreting and Applying the FAIR Principle Checks*	Apogee Engineering/Contractor to the United States Geological Survey (USGS)	Hou (2020)
--	---	------------

* Presentations at the workshop report-out session on July 22, 2020, during the 2020 ESIP Summer Meeting.

2.4. Online Dataset Quality Information Resources

A list of references on existing approaches for assessing the quality of data and associated information is captured in the website (<https://bit.ly/Pre-ESIP-Wkshp-Resources>), which may be updated to reflect the current state. Outcomes of the development of community guidelines will be also updated at this site.

2.5. Workshop Outcomes

Discussions among workshop participants during the two live sessions, leveraging knowledge from the presentations, have resulted in the following consensus among the majority of the participants:

- Dataset quality and associated information is multi-dimensional and cross-disciplinary.
- It is unlikely to have ONE dataset quality assessment model for all intended purposes and needs.
- Transparent and quantifiable quality assessments should be a part of a dataset quality management framework.
- Any approach that is used to evaluate certain dataset quality attribute(s) should be findable and accessible, and preferably be interoperable and reusable.
- The assessment results, i.e., dataset quality information, should be findable, accessible, interoperable and reusable, aka, FAIR

2.6. Workshop Key Takeaways

Discussions among workshop participants resulted in the following insights, which can inform the development of community guidelines. (A comprehensive list of workshop takeaways can be found in Appendix C.)

- The lack of organizational strategy for evaluating dataset quality causes inefficiency—guidelines can narrow this gap.
- There is a strong need for commonly defined terminology and community-vetted profiles for data quality information.
- Defined approaches and best practices should follow FAIR principles, be scalable and support automation with cross-domain support for Earth science data, such as accommodating observational and model data.
- Continued community collaboration, interoperability, machine-readable metadata, enabling both usability and understandability, and sufficient guidance documentation are

all critical foundations to the level of reach and support that can be provided through a common endeavor to advocate for and enhance data quality information.

2.7. Report-Out Session Key Takeaways

Discussions about the workshop and approaches for preparing the community guidelines identified the following additional issues that can inform efforts to develop the guidelines.

- Guidelines should be created via an iterative process and with a feedback loop between the funders, data producers, data managers, and data users.
- A document should be produced utilizing use cases to make guidelines more relevant and applicable than a peer-reviewed theoretical paper.
- The long-term sustainability of such community guidelines should be considered upfront.

In conclusion, there is a strong need for consistent dataset quality information and for standardizing quality information so that it can be readily accessible and integratable across different systems and tools. It has been emphasized that both data quality assessment methods and results should be FAIR. Actionable community guidelines with sustainable maintenance and update processes are viewed as necessary and beneficial to various stakeholders.

3. Path Forward

Building on the outcomes and momentum of the workshop, an ESIP FAIR dataset quality information community guidelines working group (FAIR DQI WG), consisting of the workshop organizing committee members and other workshop participants, has been formed and will be developing and improving the guidelines. In this section, the preliminary scope, future plans and timelines are described, along with intended use and users of such community guidelines.

3.1. The Preliminary Scope of the Guidelines

Based on the needs, challenges and existing approaches of assessing dataset quality among various organizations with various types of roles revealed by all the presentations, the discussions about the scope of the guidelines during the workshop and SM20 helped identify the following issues to consider when developing the guidelines.

- The guidelines will focus on providing guidance for capturing and representing dataset quality information that is consistent, findable, accessible, interoperable, and reusable (FAIR), but not on how to assess dataset quality, given the complexity of data and information quality attributes and different interpretations of fitness for purpose.
- The guidelines are intended to be applied to quality information of digital datasets of Earth Science measurements and derived products, analysis and reanalysis, numerical forecast, predictions, and projections. However, the guidelines will be general enough to be applied to digital datasets in other disciplines.
- The guidelines need to define the basic ingredients needed to perform dataset quality assessments, independently of the assessment approach used. For example, do not say

“use the CORE-CLIMAX system maturity matrix” but “perform maturity assessment”. Use cases may be utilized to demonstrate the utility of guidelines in establishing or maintaining dataset quality management functions, for example, in an operational environment.

- The guidelines should provide practical recommendations that are directly actionable, not just principles.
- The guidelines should be developed iteratively, leveraging community feedback for different experiences and from similar workshops.
- The long-term sustainability of such guidelines should be considered, supporting continual learning and improvement processes.

3.2. Future Plans and Timelines

In the coming months, the FAIR DQI WG will be drafting guidelines. Figure 3 shows a timeline of early planning and upcoming events that were identified to follow the workshop, culminating in a baseline guidelines document. Updates will be posted to the online resources site (<https://bit.ly/Pre-ESIP-Wkshp-Resources>).

In summary, the guidelines will aim to encourage provision of fully described assessment methods and results that are findable and accessible, machine- and human-readable, interoperable, and reusable with appropriate usage licenses.

To help with interoperability of dataset quality information, the guidelines will attempt to standardize terminology for relevant elements, for example, the stages of digital Earth science dataset lifecycles, quality dimensions, descriptive documents, and maturity metadata tags associated with different stages of the life cycle, as well as metadata entities. The common terminology will be refined and included in the guidelines. The guidelines may also include recommendations for the basic ingredients needed, such as those for an operational environment.

Figure 3: Proposed timeline of the development of the international community guidelines.

3.3. Intended Users

The audience for the community guidelines also was discussed among participants, leading to the identification of the following categories of potential users of the guidelines.

- Data producers, publishers, providers, and service providers for improved data sharing and reuse
- Data quality management professionals for improved data quality and usability
- Entities or organizations that manage and steward Earth science datasets during any stage of their full life cycle for improved enterprise data management and stewardship
- End users who integrate various datasets and associated quality information for improved interoperability and reusability
- Sponsors who fund and otherwise support efforts to collect, curate, manage, preserve and disseminate Earth science datasets

3.4. Potential Use and Possible Benefits

Opportunities that would be presented for the users of the guidelines were described in the case statement and were also discussed among participants, leading to the identification of the following affordances that the guidelines could provide for users.

- The guidelines should help users curate and represent dataset quality information that is in line with FAIR principles for improved quality information sharing and reuse.
- Persistent and readily findable and accessible dataset quality information minimizes potential redundancy in having to assess the quality of data and associated information prior to application of a data product.
- By applying and being compliant with the guidelines, organizations, including data centers and stewardship programs, will be able to demonstrate the trustworthiness of their datasets.
- By providing guidance about what elements need to be considered as basic ingredients required to perform dataset quality assessments, the guidelines can form the basis for structuring DQM practices, for example, in an operational environment.
- Guidance on the curation of dataset quality information can help improve the practices of various stakeholders who contribute to the collection, curation and dissemination of data and associated information.
- Clear documentation of the quality assessment protocols used can promote the re-use of quality assurance practices and thus support the generation of more easily comparable datasets and quality metrics.
- Machine-actionable dataset quality information allows for easy integration across systems and tools, enabling correct use and reuse of the data product and helping ensure the quality of analysis outcomes, especially those that use machine learning technologies.
- Consistently curated, traceable and adequately documented dataset quality information with sufficient evidence can guide users to address their specific needs.

- Following the guidelines helps to make the DQM processes more efficient by building on the accumulated experience from different communities.
- Consolidated guidelines can form the basis for developing in the future specific standards about dataset quality assessments to be followed by the organizations.

4. Acknowledgment

The workshop 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 Erin Robinson and Megan Carter for supporting us throughout the workshop and for 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 presenters, especially Mitch Goldberg, Jörg Schulz, Mirko Albani, Nancy Ritchey, Christina Lief, Shelley Stall, Lihang Zhou, Iolanda Maggio, Marie Drévilion, Sarah Champion, and Sophie Hou, for their excellent presentations and for engaging in the discussions and/or with the ESIP community. In addition to all the presenters and committee members, we thank the following attendees for their active participation in discussions and contribution to contents: Lesley Wyborn, Lucy Bastin, Anette Ganske, Ted Habermann, Dave Jones, Gary Berg-Cross, Brian Westra, Siri Jodha Khalsa, Irina Bastrakova, Chung-Lin Shie, Kerstin Lehnert, Paul Lemieux, Don Collins, Gastil Gastil-Buhl, and Danie Kinkade. We thank Tom Maycock for proofreading the report and Janet Fennema for formatting the references and contributing to the acronym list.

Last but not least, we thank all participants for this successful pre-ESIP workshop and the ESIP SM20 session and productive discussions during the live sessions and the two weeks of the ESIP SM20 period. Contents from all the presentations as well as discussions will be leveraged and the expertise of participants will be tapped into when developing and improving the international community guidelines.

5. References

- Albani, M. and I. Maggio, 2020: CEOS WGISS Data Management and Stewardship Maturity Matrix and Application at ESA. *Pre-ESIP Workshop*, 13 July 2020, Virtual. <https://doi.org/10.6084/m9.figshare.12711896>.
- Champion, S., 2020: Information Quality and National Climate Assessment. *ESIP Summer Meeting*, 22 July 2020, Virtual. <https://doi.org/10.6084/m9.figshare.12721610>.
- Digital Science, B. Fane, P. Ayriss, M. Hahnel, I. Hrynaszkiewicz, G. Baynes, et al. 2019: The State of Open Data Report 2019. Digital Science. Report. <https://doi.org/10.6084/m9.figshare.9980783>.
- Downs, R. R., 2020: GEOSS Data Management & Data Sharing Principles and TRUST - Implications for Information Quality. *Pre-ESIP Workshop*, 13 July 2020, Virtual. <https://doi.org/10.6084/m9.figshare.12711989>.

- Dré villon, M., I. García-Hermosa, M. G. Sotillo, C. Régnier, and the CMEMS Product Quality Working Group, 2020: Production of Quality Information at the Copernicus Marine Environment Monitoring Service (CMEMS). *ESIP Summer Meeting*, 22 July 2020, Virtual. <https://doi.org/10.6084/m9.figshare.12721592>.
- Goldberg, M. and L. Zhou, 2020: JPSS & Product Algorithm Maturity Matrix and Application. *Pre-ESIP Workshop*, 13 July 2020, Virtual. <https://doi.org/10.6084/m9.figshare.12711869>.
- Hou, Y., 2020: Interpreting and Applying the FAIR Principle Checks. *ESIP Summer Meeting*, 22 July 2020, Virtual. <https://doi.org/10.6084/m9.figshare.12721628>.
- ISO 19115-1, 2014: Geographic information—Metadata—Part 1: Fundamentals. Version: 2014-04-01. Geneva, Switzerland. Available at: <https://www.iso.org/standard/53798.html>
- ISO 19157, 2013: Geographic information—Data quality. Version: 2013-12. Geneva, Switzerland. Available at: <https://www.iso.org/standard/32575.html>
- Ivánová, I., and AU/NZ DQIG, 2020: Australia/New Zealand Data Quality Interest Group. *Pre-ESIP Workshop*, 13 July 2020, Virtual. <https://doi.org/10.6084/m9.figshare.12711944>.
- Lacagnina, C., F. Doblas-Reyes, and C3s_512 Team, 2020: Evaluation and Quality Control Function of the C3S Climate Data Store (CDS). *Pre-ESIP Workshop*, 13 July 2020, Virtual. <https://doi.org/10.6084/m9.figshare.12711656>.
- Lief, C., W. Wright, G. Peng, O. Baddour, P. Siegmund, D. Berod, R. Dunn, A. Cazenave, and M. brunet, 2020: The High Quality Global Data Management Framework for Climate – Improving the Quality of Climate Data Management for Better Climate Monitoring. *Pre-ESIP Workshop*, 13 July 2020, Virtual. <https://doi.org/10.6084/m9.figshare.12712001>.
- Moroni, D. F., H. Ramapriyan, G. Peng, J. Hobbs, J. C. Goldstein, R. R. Downs, R. Wolfe, C.-L. Shie, C. J. Merchant, M. Bourassa, J. L. Matthews, P. Cornillon, L. Bastin, K. Kehoe, B. Smith, J. L. Privette, A. C. Subramanian, O. Brown, & I. Ivánová, 2019: Understanding the Various Perspectives of Earth Science Observational Data Uncertainty. *Figshare*. <https://doi.org/10.6084/m9.figshare.10271450>
- Open Geospatial Consortium, 2016: OGC® Geospatial User Feedback Standard: Conceptual Model. Document ID: 15-097r1. Editors: Joan Masó and Lucy Bastin. Version: 1.0 2016-05-25. Available at: <http://docs.opengeospatial.org/is/15-097r1/15-097r1.html>
- Peng, G., 2018: The State of Assessing Data Stewardship Maturity – An Overview. *Data Science Journal*. 17, <http://doi.org/10.5334/dsj-2018-007>.
- Peng, G., H. Ramapriyan, and D. Moroni, 2020: ESIP Information Quality Cluster (ICQ) & Multi-Dimensions of Data and Information Quality. *Pre-ESIP Workshop*, 13 July 2020, Virtual. <https://doi.org/10.6084/m9.figshare.12710315>.
- Peng, G., C. Lacagnina, R. Downs, I. Ivánová, G. Larnicol, D. Moroni, H. Ramapriyan, and Y. Wei, 2020: Case Statement for Developing Community Guidelines for Consistently Curating and Representing Dataset Quality Information. Version: v01r00-20200702. <https://doi.org/10.6084/m9.figshare.12605438>.

- Ramapriyan, H., G. Peng, D. Moroni, and C.-L. Shie, 2017: Ensuring and Improving Information Quality for Earth Science Data and Products. *D-Lib Magazine*, 23, [10.1045/july2017-ramapriyan](https://doi.org/10.1045/july2017-ramapriyan).
- Ritchey, N., 2020: NOAA/NCEI Data Stewardship Maturity Assessment. Pre-ESIP Workshop, 13 July 2020, Virtual. <https://doi.org/10.6084/m9.figshare.12711929>.
- Schulz, J., 2020: System Maturity and Application Performance for Climate Data Record. Pre-ESIP Workshop, 13 July 2020, Virtual. <https://doi.org/10.6084/m9.figshare.12711875>
- Stall, S., 2020: Why FAIR? Pre-ESIP Workshop, 13 July 2020, Virtual. <https://doi.org/10.6084/m9.figshare.12712010>.
- Wang, R.Y. and D. M. Strong, 1996: Beyond Accuracy: What Data Quality Means to Consumers. *Journal of Management Information Systems*, 12(4):5, <https://doi.org/10.1080/07421222.1996.11518099>.
- Wei, Y., D. Moroni, H. Ramapriyan, R. Downs, Z. Liu, D. Scott, & NASA ESDSWG, 2020: NASA ESDSWG Data Quality Working Group. Pre-ESIP Workshop, 13 July 2020, Virtual. <https://doi.org/10.6084/m9.figshare.12711962>.
- Wilkinson, M., and Coauthors, 2016: The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data*, 3, <https://doi.org/10.1038/sdata.2016.18>.

Appendix A: List of Workshop and Report-Out Session Attendees

Table A1: List of the pre-ESIP workshop attendees alphabetically (names, affiliations, countries) who have signed in, with the presenters and organizing committee members identified.

Name	Affiliation(s)	Country Code	Presenter/ Committee Member
Mirko Albani	European Space Agency (ESA)	IT	Presenter
Edward Armstrong	NASA Jet Propulsion Laboratory (JPL)	US	
Lucy Bastin	Aston University	UK	
Irina Bastrakova	Geoscience Australia	AU	
Gary Berg-Cross	Ontolog Forum	US	
Fred Bingham	University of North Carolina Wilmington	US	
Anne Bowser	Wilson Center	US	
Sarah Champion	North Carolina State University (NCSU)	US	Presenter
Lisa Dallas	NASA GSFC - GLOBE Program	US	
Amanda Dean	CASE Consultants, Int. /Contractor to NCEI	US	
Candida Dewes	National Snow and Ice Data Center (NSIDC)	US	
Francisco Doblas-Reyes	BSC	ES	
Robert R. Downs	CIESIN, Columbia University/NASA SEDAC	US	Presenter/Committee
Marie Drévillon	Mercator Ocean International/Contractor to Copernicus	FR	Presenter
Anette Ganske	Technical Information Library (TIB)	DE	
Mitch Goldberg	NOAA	US	Presenter
Ted Habermann	Metadata Game Changers	US	
Daniel Heydebreck	German Climate Computing Center (DKRZ)	DE	

Keri Hornbuckle	University of Iowa	US	
Sophie Hou	Apogee Engineering /Contractor to USGS	US	Presenter
Rebecca Hudak	Woods Hole Oceanographic Institution (WHOI)	US	
Philip Jones	Riverside Technology, Inc/Contractor to NCEI	US	
Ivana Ivánová	Curtin University	AU	Presenter/Committee
Dave Jones	StormCenter Communications	US	
Amandine Kaiser	DKRZ	DE	
Siri Jodha Khalsa	NSIDC	US	
Carlo Lacagnina	BSC	ES	Presenter/Committee - Co-Lead
Kerstin Lehnert	Lamont-Doherty Earth Observatory of Columbia University	US	
Paul Lemieux	Riverside Technology, Inc.	US	
Shannon Leslie	NSIDC	US	
Christina Lief	World Meteorological Organization (WMO)	CH	Presenter
Carlee Loeser	ADNET Systems, Inc./Contractor to NASA	US	
Iolanda Maggio	RHEA Group/Contractor to ESA	IT	Presenter
David Moroni	NASA JPL	US	Committee
Rance Nault	Michigan State University	US	
Ge Peng	NCSU; CISESS/NCEI	US	Presenter/Committee-Lead
Abigail Pennington	UC San Diego	US	
Hampapuram Ramapriyan	SSAI and NASA GSFC	US	Presenter/Committee
Nancy Ritchey	NOAA NCEI	US	Presenter
Erin Robinson	ESIP	US	

Robert J. Rovetto	University of Maryland (UMD)	US	
Jörg Schulz	EUMETSAT	DE	Presenter
Chung-Lin Shie	UMD/GSFC	US	
Shelley Stall	American Geophysical Union (AGU)	US	Presenter
Laura Stolp	WHOI	US	
Terri Velliquette	NASA ORNL	US	
Yaxing Wei	NASA ORNL	US	Presenter/Committee
Brian Westra	University of Iowa	US	
Lesley Wyborn	National Computational Infrastructure	AU	
Mingfang Wu	Australian Research Data Commons (ARDC)	AU	
Qianjin Zhang	University of Iowa	US	
Lihang Zhou	NOAA/NESDIS/JPSS	US	Presenter

Table A2: List of the 2020 ESIP Summer Meeting session attendees who have signed in

Name	Affiliation(s)	Country
Joe Adams	USGS UAS	US
Frederick Bingham	UNC Wilmington	US
Josh Blumenfeld	NASA EOSDIS/ADNET Systems, Inc.	US
Sarah Champion	NCSU/NCICS, CISESS/NOAA NCEI	US
Don Collins	NCEI	US
Joan Damerow	Lawrence Berkeley National Lab	US
Lisa Dallas	NASA GSFC	US
Amanda Dean	CASE, Inc./NOAA NCEI	US
Robert R. Downs	ESIP IQC, CIESIN, Columbia University & NASA SEDAC	US
Marie Drévillon	Mercator Ocean International	FR
Alexis Garretson	ESIP/Jackson Laboratory	US

Hap Garritt	PIE LTER	US
Gastil Gastil-Buhl	MCR LTER	US
Rob Crystal-Ornelas	Lawrence Berkeley National Lab	US
Liz Harper	NOAA NCEI	US
Sophie Hou	Apogee Engineering/USGS	US
Ivana Ivánová	Curtin University	AU
Dave Jones	StormCenter Communications/GeoCollaborate	US
Danie Kinkade	BCO-DMO/WHOI	US
Kitty Kolb	USGS	US
Madison Langseth	USGS	US
Mary Martin	HBR LTER	US
David Moroni	NASA JPL, California Institute of Technology	US
Karen Moe	NASA GSFC Retired	US
Frank Nitsche	USAP-DC, LDEO, Columbia University	US
Ge Peng	ESIP IQC, NCSU/NCICS, CISESS/NOAA NCEI	US
Abby Pennington	UC San Diego	US
Carrie Pulliam	University of Kentucky	US
H. K. Ramapriyan	ESIP IQC, SSAI/NASA GSFC	US
Chung-Lin Shie	UMBC/JCET, NASA/GES DISC	US
Donna Scott	NSIDC	US
Adam Sapp	GCE LTER	US
Yaxing Wei	NASA ORNL DAAC	US
Brian Westra	University of Iowa	US
Qianjin Zhang	University of Iowa Engineering Library	US

Appendix B: Short Bio of the Presenters and Workshop Organizing Committee Members

Mirko Albani is responsible of the Heritage Space Programme (LTDP+) addressing the management and preservation of heritage space data in the Earth Observation Directorate at the European Space Agency. He is also responsible of internal coordination activities and representing the Agency in GEO (Group on Earth Observation) and CEOS (Committee on Earth Observation Satellites) where he is leading the Data Stewardship Interest Group. He has an advanced Degree in Electronic Engineering from the University of Rome Tor Vergata and a post-graduate Master in Space System Engineering from the Technological University of Delft. In his career, he developed sound and broad knowledge of Navigation, Telecommunication and Earth Observation systems working at ESA, ASI (Italian Space Agency) and industry.

Sarah Champion holds a Bachelor of Science degree in Meteorology from Penn State University and Master of Science degree in Meteorology from The Naval Postgraduate School, which she earned while serving on active duty with the U.S. Air Force. For the last 6 years, Sarah has served as the Data Architect for the National Climate Assessment (NCA), and is responsible for ensuring the NCA's data transparency, accessibility, and reproducibility in compliance with the Information Quality Act, Highly Influential Scientific Assessments and the Foundations for Evidence-Based Policymaking Act. This role is fulfilled through the Cooperative Institute for Satellite Earth Systems Studies (CISESS), the North Carolina Institute for Climate Studies (NCICS), and North Carolina State University, where Sarah holds the position of Research Scholar, focused on topics in atmospheric science and climate change.

Dr. Robert R. Downs serves as senior digital archivist and acting head of cyberinfrastructure and informatics R&D at CIESIN, the Center for International Earth Science Information Network, in the Earth Institute of Columbia University; vice-chair of the Columbia University Morningside Campus Institutional Review Board (IRB) and co-chair of the ESIP Information Quality Cluster, the Group on Earth Observation (GEO) Data Working Group, and the Research Data Alliance (RDA) Repository Platforms for Research Data (RPRD) IG. He is on the CoreTrustSeal Standards and Certification Board; the CODATA Data Science Journal Editorial Board, and the Executive Committee of the EarthCube Council of Data Facilities.

Dr. Marie Drévillon is an oceanographer at Mercator Ocean, Toulouse, France. After a PhD in climate physics, and a postdoctoral fellowship in ocean physics and biogeochemistry, she joined Mercator Ocean in 2005 to participate in the development of their first global ocean forecasting system. Since 2009, she has been managing the scientific qualification of Mercator Ocean reanalyses, analyses and forecasts, within the research and development department. Since 2015, she has been involved in the coordination of the European Copernicus Marine Environment Monitoring Service activities on multi-year products and scientific quality.

Dr. Mitch Goldberg of the NOAA National Environmental Satellite, Data, and Information Service (NESDIS). Dr. Goldberg serves as an expert and lead representative of the NOAA-NESDIS science and user communities and coordinates NESDIS applied science. He leads the NESDIS Satellite Proving Ground program to improve NOAA services by facilitating collaboration between users and product developers to optimize the use of satellite data. Dr. Goldberg is involved in a number of international coordinating groups including the World Meteorological Organization (WMO), Coordination Group for Meteorological Satellites (CGMS), and the Committee on Earth Observation

Satellites (CEOS). Dr. Goldberg is a Fellow of the American Meteorological Society (AMS), and the current chair of the AMS Satellite Meteorological, Oceanography, and Climatology Committee.

Sophie Hou is a Data and Usability Analyst from Apogee Engineering working with the USGS Science Data Management, Science Analytics and Synthesis, and the Western Geographic Science Center. Sophie is interested in designing and implementing best practices for scientific data, especially in the areas of user interface (UI)/user experience (UX) and trusted infrastructure. She is also interested in developing data management and usability training materials in order to meet the growing data workforce needs. Sophie has a Master of Science degree in Library and Information Science with a specialization in Data Curation from the University of Illinois, Urbana-Champaign. Sophie also has additional degrees and professional work experience in the field of Electrical Engineering.

Dr. Ivana Ivánová is a Senior Lecturer at Curtin University and Research Fellow at FrontierSI. Her research interests and expertise are in the quality of spatial resources and spatial information infrastructures. She has an engineering and doctoral degree in geodesy and cartography with specialization in Geoinformatics and spatial data quality. Ivana has more than 20 years of experience with national and international standardization; currently she is co-chairing the Open Geospatial Consortium (OGC)'s Domain Working Group on Quality, and, at ISO/TC211, is leading the revision of ISO 19157-1 standard on Geographic Information Quality.

Christina Lief is a Physical Scientist who specializes in global observing systems and environmental data management. Retired from NOAA/NCEI in 2017, Christina now leads the WMO International Expert Group on Climate Data Modernization focused on developing applications and tools to assess climate data quality and was elected chair of the WMO Expert Team on Data Development and Stewardship (2018–2020) and co-chair of the WMO Expert Team on Data Requirements for Climate Services (2020–2022). Christina earned a BS in Geological and Earth Sciences from the George Washington University and a MS in Geographic and Cartographic Sciences from the George Mason University.

Iolanda Maggio has a computer science background with specialization in Data Science, with extensive work experience in the Space and Defence Operations environment and space project development and operations service industries. Currently she is covering a Senior Project Manager/Engineering and Operation of Space Mission Support position on the ESA Establishment in Frascati (Rome) with the following responsibilities: coordination of a team acting as a Team Leader, in order to manage specific projects relating to Mission Data Operation, Preservation, Reprocessing and exploitation and giving support for space mission international cooperation and standardization committees activities and initiatives. She is acting as Service Manager with responsibility for service improvements, KPI management and Risk Management.

Dr. Carlo Lacagnina works at BSC as technical manager of the service contract in charge of developing and implementing the Evaluation and Quality Control (EQC) function of the CDS C3S. He has recognized scientific expertise both in climate model biases and remote-sensing of clouds and aerosols. He has been working in data quality maintenance for both business and climatological applications.

Dr. Gilles Larnicol has 25 years of experience in remote sensing applications and services and in leading of complex, international projects—in particular in the frame of the European Copernicus

Programme. He has well-recognized scientific expertise both in remote-sensing and operational and research oceanography and climate. He has participated in numerous EC and Space Agencies projects (CNES, ESA, EUMETSAT). He led projects aiming at implementing operational service related to “Sea Level satellite observations for the Copernicus Marine and Climate Service” and for the ESA Climate Change Initiative. More recently, he is co-lead of the Evaluation and Quality Control component of the Climate Service project (C3S_512 project lead by Barcelona Super Computer Center).

Mr. David Moroni is co-chair of the ESIP IQC and Data Publication Team manager at the Physical Oceanography Distributed Active Archive Center (PO.DAAC) located at the Jet Propulsion Laboratory (JPL), California Institute of Technology, Pasadena, CA. Since joining PO.DAAC in 2008, Mr. Moroni has coordinated and overseen the publication of hundreds of cross-disciplinary Earth science datasets, encompassing satellite remote sensing, data assimilation, and airborne and in situ data sources. Throughout his tenure at JPL, Mr. Moroni has also been a researcher and co-investigator on 5 funded NASA-ROSES proposals, contributing to fundamental physical oceanography and air–sea interaction studies, as well as data-centric informatics technology to improve data search relevancy.

Dr. Ge Peng, co-chair of ESIP IQC, is a scientist at the Cooperative Institute for Satellite Earth System Studies (CISESS), a partnership between North Carolina State University and NOAA National Centers for Environmental Information (NCEI). Dr. Peng is experienced in evaluating data product quality. She led and participated in the development and application of several data maturity assessment models. She serves as the NCEI Snow and Ice product area lead and is a chief editor of the Earth System Science Data journal. Dr. Peng will support the WMO Expert Team on Data Requirements for Climate Services and is an Information Management and Technology Expert Team co-chair (2020–2022).

Dr. Hampapuram K. Ramapriyan is a Research Scientist at Science Systems and Applications, Inc. (SSAI) supporting the ESDIS Project at NASA Goddard Space Flight Center. His primary areas of focus are data stewardship and preservation, and data and information quality. He chaired ESIP’s Information Quality Cluster during 2015–2019 and is a former vice-chair of ESIP’s Data Stewardship Committee. Recently, he led the development of an ISO Standard (ISO 19165-2:2020) specifying Earth science data and metadata content to be preserved to ensure dataset reusability for the long-term future. He chairs NASA’s Data Product Developers Guide Working Group, which recently published a guide for data producers. He is a member of the AGU and a Senior Member of the IEEE.

Dr. Jörg Schulz is Climate Service & Product Manager at the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), leading the climate activities at EUMETSAT that are delivering climate data records for climate services and science. He addresses the generation of climate data records from historical, currently operated, and new satellite instruments, elaborating on new methods for uncertainty quantification and improvements of satellite instrument calibration. He authored or co-authored more than 80 publications. Jörg has been engaged in international science and Earth Observation coordination, leading the WCRP GEWEX Data and Analysis Panel from 2014-2016, chairing the Joint Committee on Earth Observation Satellites/Coordination Group for Meteorological Satellite Working Group on Climate (2017–2020), and is also a member of the WCRP Data Advisory Council.

Shelley Stall is the Senior Director for the American Geophysical Union's Data Leadership Program. She works with AGU's members, their organizations, and the broader research community to improve data and digital object practices with the ultimate goal of elevating how research data is managed and valued. Better data management results in better science. Shelley's recent work includes the Enabling FAIR Data project (<https://copdess.org/enabling-fair-data-project/>) engaging over 300 stakeholders in the Earth, space, and environmental sciences to make data open and FAIR and targeting the publishing and repository communities to change practices by no longer archiving data in the supplemental information of a paper but instead depositing the data supporting the research into a trusted repository where it can be discovered, managed, and preserved. ORCID: 0000-0003-2926-8353

Dr. Yaxing Wei is the Lead Scientist for the ORNL Distributed Active Archive Center (DAAC) and a geoinformatics data scientist at the Remote Sensing and Environmental Informatics (RSEI) group in the Environmental Sciences Division. Since joining ORNL in 2008, Wei has been dedicated in data management, visualization, sharing, and analysis for a number of research projects, including the ORNL DAAC, the Multi-scale Synthesis and Terrestrial Model Intercomparison Project (MsTMIP), and the Data Observation Network for Earth (DataONE). Wei is an IEEE Senior Member. He received the International Council for Science's World Data System (WDS) Data Stewardship Award in 2015. He is currently chairing the ESIP Information Quality Cluster (IQC) and NASA Earth Science Data and Information System (ESDIS) Standards Office (ESO).

Lihang Zhou is the Data Products Management and Services (DPMS) Deputy and the Satellite Products Portfolio Manager in NOAA Joint Polar Satellite System (JPSS). Lihang has extensive experience on satellite algorithms, cal val, and program management. From 2006 to 2009, she worked on GOES-R as the quality assurance manager for the algorithm working group (AWG). From 2009 to 2011, she worked in NOAA's Integrated Program Office (IPO) as the Senior Data Product and Algorithm Scientist. Before joining JPSS, she worked at NOAA/STAR as the STAR JPSS Program Manager and the science deputy for the JPSS Algorithm Management Project. She has also worked on algorithms retrieving products from the hyperspectral infrared sounding instruments onboard U.S. and international satellites.

Appendix C: Main Takeaways of the Workshop

This appendix summarizes main takeaways of the workshop held virtually to kick off the development of community guidelines for consistently curating and representing dataset quality information that is in line with the FAIR principles.

Twelve presentations laid the groundwork for discussion, showing the complexity and multi-dimensionality of dataset quality information and existing approaches used by various types of organizations (see Section 2 of the main text for more details). Discussions about the challenges and approaches for preparing the community guidelines, along with questions inspired by the presentations, identified the following takeaways. These takeaways can be used to drive the development of the guidelines.

- To start with, a few firm points about terminology. *Dataset quality* is different from data quality. The former includes quality of both the data and the associated information, including that captured in documents, e.g., scientific papers or reports, and that captured in the metadata record. Hereafter, the terms data products and datasets are used interchangeably. *Dataset quality management (DQM)*: a data quality function is made of different groups: scientific quality, technical quality, and dissemination of the quality information to the users and possibly more groups. All these aspects together are considered as dataset quality management, which is part of data governance. What guarantees that data is not corrupted and is accurate are the quality control procedures plus a series of protocols to avoid new quality issues in the system. This latter element is part of the quality assurance. Both quality control and quality assurance can consider scientific (e.g., uncertainty) and technical (e.g., temporal completeness) aspects. The information acquired during the quality control/assurance processes has to be disseminated to improve “usability” of the data and “verifiability” of the quality procedures applied, with the ultimate goal of increasing trustworthiness in data and information disseminated by the operational services. How to disseminate this quality-related information is another aspect of quality management.
- Dataset quality is a multi-dimensional problem, and there is no one assessment model valid for all intended purposes. Instead, the guidelines can be very useful for setting the direction for the right approach to follow in each specific case.
- Any approach used for DQM and the related results should follow the FAIR principles and be as transparent and quantifiable as possible. Documentation that supports the assessments should be provided along with the assessment results.
- There is a need for commonly defined terminology for data quality information. Consistency across assessment groups allows wider application of the same tools.
- DQM approaches should be sustainable, which calls for efficiency and scalability. Scalability can be supported by considering automated processes for evaluation, while having clear guidelines for a consistent assessment framework would strongly help to reach efficient quality assessments. The lack of organizational strategy for evaluating dataset quality causes inefficiency—guidelines can narrow this gap.

- There is a strong need for producing consistent dataset quality information so that this information can be readily accessible and integratable across systems and tools. This calls for machine-readable assessment results, which need to be captured in a standardized way, while keeping them human-readable. The tradeoff between machine- and human-readable assessment results has to be a foundation of the DQM approaches. For reasons of scalability and machine readability, dataset quality assessments are recommended to be as much as possible automatized, particularly in an operational environment.
- Importance of user feedback: Involving users in the assessments increases trust. Best practices need to be vetted by users. The Usage section of ISO metadata standard (19115-1) provides a standard way to share usage information (e.g., who is the user, intended use, and limitations identified) (ISO 19115-1 2014). The standards developed by the Open Geospatial Consortium (2016) on geospatial user feedback could be used for completing the ‘usage’ metadata. The revision of ISO 19157-1 and -2 of the current ‘usability’ standard element aims to make it a type of quality report.
- Maturity assessments constitute an important ingredient for the dataset quality assessment tasks. Indeed, high maturity provides trust that the data records are produced following best practices of the community. It should be noted that a low maturity rating does not necessarily imply low scientific value for a dataset. It can happen especially for datasets managed by a single investigator that may be deemed to have low stewardship maturity due to poor quality in metadata, documentation and accessibility.
- Challenges of the Maturity Model (MM) assessments: detailed explanations of each MM question and examples of already assessed datasets and training are necessary for providing homogeneous and correctly interpreted assessments. Moreover, assessment approaches should be sustainable, which calls for scalability; an automated process for evaluation can also contribute to scalability. An example which tackles this issue of scalability is represented by the NCEI approach. NCEI involves data producers in the assessment by means of questionnaires. This also guarantees an independent assessment. An additional gap identified regarding MMs is that a reference model to be applied for the assessment of model-based datasets is currently missing.
- Quality indicators, for instance the ESIP Operational Readiness Levels. (<https://www.esipfed.org/orl>), are useful to inform quickly about the level of data quality
- Independent assessment: having peer-reviewed papers written by the data creators is not enough—there is a need for an independent assessment exercise by an end user, which can be a third party external to the data provider.
- Cost of quality: human resources and technological costs associated with capacity building efforts. This topic was slightly touched in one question, but it is an important ingredient to at least mention when an organization has to set the direction of the DQM practices to consider.
- Machine learning: ML can help with detecting potential issues with data, and availability of high-quality datasets can help ML to better utilize the data and information.

- Inspiring principles: FAIR. Facilitate discovery, access, usability, preservation, and curation of data quality information (GEOSS data management principles). Provide data quality information with open data, minimal restrictions and limited costs, in a timely manner (GEOSS data-sharing principles). Employ transparency, responsibility, user focus, sustainability and technology to provide data quality information with data (TRUST principles for digital repositories). The WMO Global Framework for Climate Services (GFCS) should also be considered as a foundation for these guidelines.
- Motivations for disseminating quality assessments: i) They tend to originate from the needs and requirements to ensure and establish the trustworthiness of practices or processes, in reference to data, services or organizational capabilities. Organizations can also leverage those assessment results to improve their practices and workflows. It also helps funders get a measure of how compliant the funded datasets or services are with their requirements. ii) Dataset quality information helps users compare different datasets. Clear documentation of the quality assessment protocols used can promote the re-use of quality assurance practices and thus support the generation of more easily comparable datasets and quality metrics. iii) It is good practice to connect the quality information with the dataset before release, and users should even be required to acknowledge having read the quality information before using the data. iv) Dataset quality information should help users identify whether the data fits their own purpose and prevent data product misuse, which in turn may result in damage to the reputation of the data provider. v) Persistent, readily findable and accessible dataset quality information minimizes potential redundancy in having to assess the quality of data and associated information prior to application of a data product.
- There is a lack of the identified minimum metadata information required to make the dataset reusable.

Appendix D: List of Acronyms

AGU	American Geophysical Union
AI	Artificial Intelligence
ARDC	Australian Research Data Commons
AU/NZ DQIG	Australia/New Zealand Data Quality Interest Group
ANU	Australian National University
BCO-DMO	Biological and Chemical Oceanography Data Management Office
BSC	Barcelona Supercomputing Center, Spain
C3S	Copernicus Climate Change Service
CDR	Climate Data Record
CDS	Copernicus Climate Data Store
CEOS	Committee on Earth Observation Satellites
CIESIN	Center for International Earth Science Information Network
CISESS	Cooperative Institute for Satellite Earth System Studies
CMEMS	Copernicus Marine Environment Monitoring Service
CNES	National Centre for Space Studies, France
CODATA	Committee on Data of the International Science Council (ISC)
DAAC	NASA Distributed Active Archive Centers
DataONE	Data Observation Network for Earth
DKRZ	German Climate Computing Center
DQD WG	OGC Data Quality Domain Working Group
DQI	Dataset Quality Information
DQ IG	ARDC Data Quality Interest Group
DQM	Data Quality Management
ECV	Essential Climate Variables
EOSDIS	NASA's Earth Observing System Data and Information System
ESA	European Space Agency
ESDSWG	NASA Earth Science Data System Working Groups
ESIP	Earth Science Information Partners
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FAIR	Guiding principles for Findable, Accessible, Interoperable, and Reusable
GCE LTER	Georgia Coastal Ecosystems Long Term Ecological Research
GEO	Group on Earth Observations
GEOSS	Global Earth Observation System of Systems
GES DISC	NASA's Goddard Earth Sciences Data and Information Services Center
GEWEX	Global Energy and Water Exchanges project, part of the World Climate Research Programme (WCRP)
GLOBE	NASA's Global Learning & Observations to Benefit the Environment Program
GFCS	WMO Global Framework for Climate Services
GSFC	NASA Goddard Space Flight Center
HBR LTER	Hubbard Brook Long Term Ecological Research
IEEE	Institute of Electrical and Electronics Engineers
IQA	Information Quality Act
IQC	Information Quality Cluster of ESIP

ISO	International Standards Organization
JCEST	Joint Center for Earth Systems Technology
JPL	NASA Jet Propulsion Laboratory
JPSS	Joint Polar Satellite System, a collaborative program between the National Oceanic and Atmospheric Administration (NOAA) and its acquisition agent, National Aeronautics and Space Administration (NASA)
MCR LTER	Moorea Coral Reef Long Term Ecological Research
ML	Machine Learning
MM	Maturity Model
MsTMIP	Multi-scale Synthesis and Terrestrial Model Intercomparison Project
NASA	National Aeronautics and Space Administration of the United States
NCA	National Climate Assessment
NCEI	National Centers for Environmental Information
NCI	National Computational Infrastructure, Australia
NCICS	North Carolina Institute for Climate Studies
NCSU	North Carolina State University
NESDIS	National Environmental Satellite, Data, and Information Service
NOAA	National Oceanic and Atmospheric Administration of the United States
NSIDC	National Snow and Ice Data Center
OGC	Open Geospatial Consortium, an international voluntary consensus standards organization
ORNL	NASA Oak Ridge National Laboratory
PIE LTER	Plum Island Ecosystems Long Term Ecological Research
RDA	Research Data Alliance
RPRD IG	Repository Platforms for Research Data Interest Group
RSEI	Remote Sensing and Environmental Informatics
SEDAC	NASA Socioeconomic Data and Applications Center
SSAI	Science Systems and Applications, Inc.
TRUST	Principles for Digital Repositories: Transparency, Responsibility, User Focus, Sustainability, Technology
UAS	USGS Unmanned Aircraft Systems
UC	University of California, San Diego
UMBC	University of Maryland, Baltimore County
UMD	University of Maryland
USAP-DC	United States Antarctic Program Data Center
USGS	United States Geological Survey
WCRP	World Climate Research Programme
WDS	World Data System
WGISS	Working Group on Information Systems and Services, a subsidiary body supporting CEOS
WHOI	Woods Hole Oceanographic Institution
WMO	World Meteorological Organization