
REVIEW

The State of Assessing Data Stewardship Maturity – An Overview

Ge Peng

North Carolina State University, Cooperative Institute for Climate and Satellites – North Carolina (CICS-NC) at NOAA's National Centers for Environmental Information (NCEI), Asheville, North Carolina 28801, US
gpeng@ncsu.edu

Data stewardship encompasses all activities that preserve and improve the information content, accessibility, and usability of data and metadata. Recent regulations, mandates, policies, and guidelines set forth by the U.S. government, federal other, and funding agencies, scientific societies and scholarly publishers, have levied stewardship requirements on digital scientific data. This elevated level of requirements has increased the need for a formal approach to stewardship activities that supports compliance verification and reporting. Meeting or verifying compliance with stewardship requirements requires assessing the current state, identifying gaps, and, if necessary, defining a roadmap for improvement. This, however, touches on standards and best practices in multiple knowledge domains. Therefore, data stewardship practitioners, especially these at data repositories or data service centers or associated with data stewardship programs, can benefit from knowledge of existing maturity assessment models. This article provides an overview of the current state of assessing stewardship maturity for federally funded digital scientific data. A brief description of existing maturity assessment models and related application(s) is provided. This helps stewardship practitioners to readily obtain basic information about these models. It allows them to evaluate each model's suitability for their unique verification and improvement needs.

Keywords: Data Stewardship; Maturity Matrix; Data Preservation; Data Accessibility; Scientific Data; Information Quality; Metadata

1. Introduction

Data stewardship “encompasses all activities that preserve and improve the information content, accessibility, and usability of data and metadata” (National Research Council 2007). Scientific or research data is defined as: “the recorded factual material commonly accepted in the scientific community as necessary to validate research findings.” (OMB 1999). Federally funded scientific data are generally cared for by individual organizations such as repositories, data centers, and data stewardship programs or services.

U.S. laws and federal government mandates, policies and guidelines set forth in the last two decades have greatly expanded the scope of stewardship for federally funded digital scientific data. These directives include the following:

- U.S. Information Quality Act (US Public Law 106-554 2001, Section 515), also known as Data Quality Act;
- U.S. Federal Information Security Management Act (US Public Law 107-347 2002);
- Policies on Open Data and Data Sharing (OMB 2013; OSTP 2013);
- Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information (OMB 2002a);
- Guidelines on Ensuring Scientific Integrity (OSTP 2010).

In response to these governmental directives, federal and other funding agencies have issued their own policies and guidelines (e.g., NOAA 2010; see Valen and Blanchat (2015) for an overview of each federal agency's compliance with the OSTP open data policies).

Recognizing the impact and challenges of changing digital environment, the National Academy of Sciences (NAS), together with the National Academy of Engineering and the Institute of Medicine, have prompted the good stewardship of research data and called for transparency and data sharing to ensure data integrity and utility (NAS 2009). The Group on Earth Observations (GEO) has called for “full and open exchange of data, metadata and products” in its defined data sharing principles for its data collections to ensure the data are available and shared in a timely fashion (GEO Data Sharing Working Group 2014). Scientific societies and scholarly publishers, such as those involved in the Coalition on Publishing Data in the Earth and Space Sciences (COPDESS), have issued a position statement calling for data used in publications to be “available, open, discoverable, and usable” (COPDESS 2015). The World Data Service (WDS) of the interdisciplinary Body of the International Council for Science (ICSU) requires as a condition of membership that its members demonstrate their compliance with the WDS strong commitment to “open data sharing, data and service quality, and data preservation” (<https://www.icsu-wds.org/organization>; see also WDS Scientific Committee 2015). Stakeholders from academia, industry, funding agencies, and scholarly publishers have formally defined and endorsed a set of FAIR (namely, Findable, Accessible, Interoperable, Reusable) data principles for scientific data management and stewardship (Wilkinson et al. 2016).

These governmental regulations and mandates, along with principles and guidelines set forth by funding agencies, scientific organizations and societies, and scholarly publishers, have levied stewardship requirements on federally funded digital scientific data. As a result, stewardship activities are extremely critical for ensuring that data are: *scientifically sound and utilized, fully documented and transparent, well-preserved and integrated, and readily obtainable and usable*.

This elevated level of well-defined requirements has increased the need for a more formal approach to stewardship activities—one that supports rigorous compliance verification and reporting. Meeting or verifying compliance with stewardship requirements requires assessing the current state, identifying gaps, and, if necessary, defining a roadmap for improvement. However, such an effort requires comprehensive cross-disciplinary knowledge and expertise, which is extremely challenging for any single individual. Therefore, data stewardship practitioners can benefit from the existence of maturity models and basic information about them.

A maturity model is considered as a desired or anticipated evolution from a more ad hoc approach to a more managed process. It is usually defined in discrete stages for evaluating maturity of organizations or process (Becker, Knackstedt & Pöppelbuß 2009). A maturity model can also be developed to evaluate practices applied to individual data products (e.g., Bates and Privette 2012; Peng et al. 2015). A number of maturity models have been developed and utilized to quantifiably evaluate both stewardship processes and practices.

This article provides an overview of the current state of assessing the maturity of stewardship of digital scientific data. A list of existing or developing maturity models from various perspectives of scientific data stewardship is provided in **Table 1** with a high-level description of each model and its application(s) in Section 3. This allows stewardship practitioners to further evaluate the utility of these models for their unique stewardship maturity verification and improvement needs.

2. Perspectives of Scientific Data Stewardship

Figures 1 and **2** display different perspectives of maturity within the context of managing scientific data stewardship activities. They highlight the interconnectivity and interdependency of different levels of stewardship activities within individual organizations and different types of maturity for scientific data products through the entire data product lifecycle.

The diagram in **Figure 1** shows tiered maturity assessments from a top-down view. The top level represents an organization's processes while the lowest level represents the practices applied to individual data products of its data holdings. As indicated by the arrows in **Figure 1**, the maturity of organizational process capability can influence the maturity of portfolio management and individual data products, while the maturity of individual data products may reflect—and potentially impact—the maturity of portfolio management and organizational process capability.

The quality of a data product and its associated practices throughout its lifecycle can impact its overall quality. As the overall quality of a data product, unfortunately, tends to be dictated by the lowest quality from any stage of its entire life cycle, it is important to take a holistic approach to ensure and improve the

Table 1: A list of maturity assessment models based on various perspectives of data stewardship activities.

Maturity Perspective	Maturity Assessment Model and Reference Citation
Organizational data management maturity	CMMI Institute's Data Management Maturity Model (CMMI 2014) Enterprise Data Management Council (EDMC) Data Management Capability Assessment Model (EDMC 2015)
Repository data management procedure maturity	ISO standard for audit and certification of trustworthy digital repository (ISO 16363 2012) WDS-DSA-RDA core trustworthy data repository requirements (Edmunds et al. 2016)
Portfolio management maturity	Federal Geographic Data Committee (FGDC) lifecycle maturity assessment model (Peltz-Lewis et al. 2014; FGDC 2016)
Dataset science maturity	Gap Analysis for Integrated Atmospheric ECV CLimate Monitoring (GAIA-CLIM) measurement system maturity matrix (Thorne et al. 2015) NOAA's Center for Satellite Applications and Research (STAR) data product algorithm maturity matrix (Reed 2013; Zhou, Divakarla & Liu 2016) COordinating Earth observation data validation for RE-analysis for CLIMate ServiceS (CORE-CLIMAX) production system maturity matrix (EUMETSAT 2013)
Dataset product maturity	NOAA satellite-based climate data records (CDR) product maturity matrix (Bates et al. 2015) CORE-CLIMAX production system maturity matrix (EUMETSAT 2013)
Dataset stewardship maturity	NCEI/CICS-NC scientific data stewardship maturity matrix (Peng et al. 2015) CEOS Working Group on Information Systems and Services (WGISS) data management and stewardship maturity matrix (WGISS DSIG 2017)
Dataset use/service maturity	National Snow and Ice Data Center (NSIDC) level of services (Duerr et al. (2009) NCEI tiered scientific data stewardship services (Peng et al. 2016a) Global Climate Observing System (GCOS) ECV Data and Information Access Matrix Global Ocean Observing System (GOOS) framework NCEI data monitoring and user engagement maturity matrix (Arndt and Brewer 2016)

quality of a digital scientific data product. **Figure 2** depicts the maturity assessment from this horizontal view, adopting four dimensions of information quality defined by Ramapriyan et al. (2017): science, product, stewardship, and service. They correspond to the activities involved in the four different phases of the dataset life cycle: "1. define, develop, and validate; 2. produce, assess, and deliver (to an archive or data distributor); 3. maintain, preserve and disseminate; and 4. enable data use, provide data services and user support." (Ramapriyan et al. 2017). We adopt this categorization of information quality dimensions for general scientific data products because it better reflects the differing roles and responsibilities of entities involved in the different stages of dataset lifecycle. These distinct roles often require different domain knowledge and expertise.

Table 1 provides a list of existing maturity assessment models, including those highlighted in **Figures 1** and **2**. Brief descriptions of these models and, where available, their applications are provided in the next section.

3. Maturity Models Overview

a) Organizational data management maturity

Data management includes all activities for "planning, execution and oversight of policies, practices and projects that acquire, control, protect, deliver and enhance the value of data and information assets." (Mosely et al. 2009).

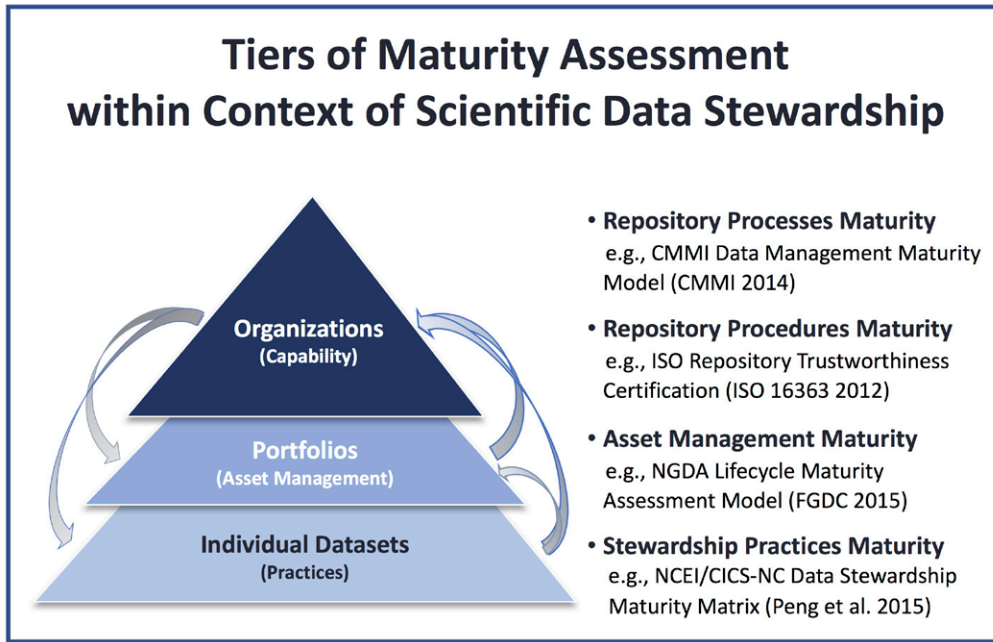


Figure 1: Category of tiered maturity assessment within the context of scientific data stewardship and examples of existing maturity assessment models. The arrows indicate that the maturity at the initiation point can impact that at the ending point. See Section 3 for a high-level description of each maturity assessment model listed in the diagram.

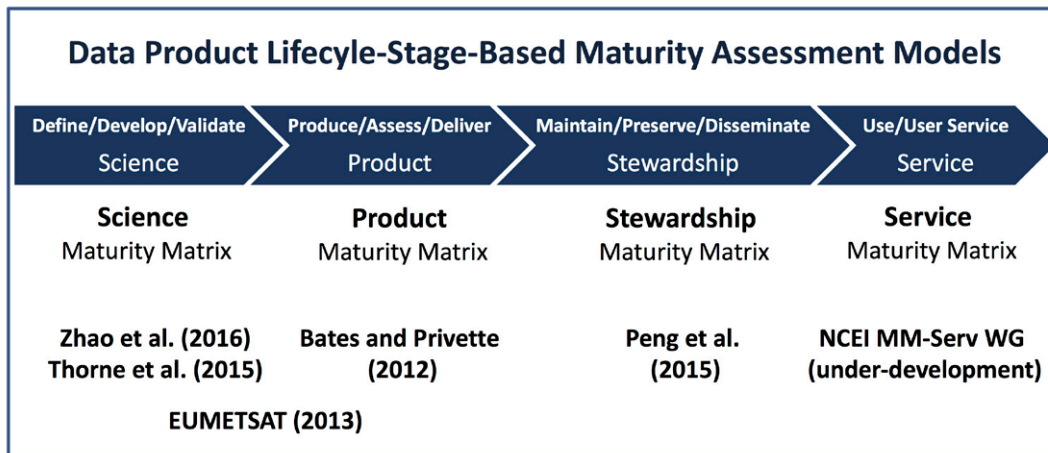


Figure 2: Category of data product lifecycle-stage-based maturity type and examples of existing assessment models in the form of a matrix. See Section 3 for a high-level description of each maturity assessment model listed in the diagram.

McSweeney (2013) reviewed four leading business data management maturity assessment models and concluded that there is lack of consensus about what comprises information management maturity and a lack of rigor and detailed validation to justify organization process structures. He called for a consistent approach, linked to an information lifecycle (McSweeney 2013).

Following the CMMI principles and approaches, the CMMI Institute’s Data Management Maturity (DMM) Model was released in August 2014. The DMM model is designed to cover all facets of data management and provides a reference framework for organizations to evaluate capability maturity, identify gaps, and provide guidelines for improvements across a project or an entire organization (CMMI 2014). The CMMI DMM model assesses 25 data management process areas organized around the following six categories: data management strategy, data governance, data quality, data platform & architecture, data operations, and supporting processes (CMMI 2014; Mecca 2015). It has been utilized by seven different businesses (Mecca 2015) and adopted by the AGU (American Geophysical Union) data management assessment program (Stall 2016).

The Enterprise Data Management Council (EDMC) Data Management Capability Assessment Model (DCMM) was released in July 2015 (EDMC 2015). DCMM defines a standard set of evaluation criteria for measuring data management capability and is designed to guide organizations to establish and maintain a mature data management program (EDMC 2015; Gorbali 2016). A detailed description and comparison of CMMI DMM and EDMC DCMM can be found in Gorbali (2016).

b) Repository data management procedure maturity

The trustworthiness of individual repositories has been the topic of study for the data management and preservation community for many years. Based on the Open Archival Information System (OAIS) reference model, ISO 16363 (2012) establishes comprehensive audit metrics for what a repository must do to be certified as a trustworthy digital repository (see also CCSDS 2012a). Three important qualities of trustworthiness are integrity, sustainability, and support for the entire range of digital repositories in three different aspects: organizational infrastructure, digital object management, and infrastructure and security risk management (ISO 16363 2012; CCSDS 2012b; Witt et al. 2012). A detailed justification for transparency is now recommended in the ISO 16363 repository trustworthiness assessment template.

Working with the Data Seal of Approval (DSA) and the Research Data Alliance (RDA), the WDS-DSA-RDA working Group developed a set of core trustworthy data repository requirements that can be utilized for certification of repositories at the core level (Edmunds et al. 2016) as a solid step towards meeting the ISO 16363 standards.

On the individual agency level, the United States Geological Survey (USGS) has adopted the WDS-DSA-RDA core trustworthy data repositories requirements and begun to evaluate and issue the “USGS Trusted Data Repository” certificate to its data centers (Faundeen, Kirk & Brown 2017). For organizations that do not wish to go through a formal audit process, utilizing this ISO assessment template and the WDS-DSA-RDA core requirements will still help them evaluate where they are and identify potential areas of improvement in their current data management and stewardship procedures.

c) Portfolio management maturity

An organization may identify and centrally manage a set of core data products because of the significance of those products in supporting the strategy or mission of the organization. For example, under OMB Circular A-16 (OMB 2002b) “Coordination of Geographic Information and Related Spatial Data Activities,” the Federal Geographic Data Committee (FGDC) designed a portfolio management process for 193 geospatial datasets contained within the 16 topical National Spatial Data Infrastructure themes (FGDC 2016). These 193 datasets “are designated as National Geospatial Data Assets (NGDA) because of their significance in implementing to the missions of multiple levels of government, partners and stakeholders” (Peltz-Lewis et al. 2014). The first NGDA lifecycle maturity assessment (LMA) model was developed and utilized to baseline the maturity of the NGDA datasets (FGDC 2015). The LMA model assesses the maturity in seven lifecycle stages of data portfolio management: define, inventory & evaluate, obtain, access, maintain, use & evaluate, and archive (FGDC 2016). The assessments were mostly carried out by data managers and summarized with improvement recommendations for future LMA assessment to support portfolio management process in FGDC (2016). LMA assessment reports of NGDA dataset are available online at: https://www.fgdc.gov/ngda-reports/NGDA_Datasets.html and maturity levels can be reviewed via an online tool at: <https://dashboard.geoplatform.gov> (user login may be required).

A similar approach can be adapted to product portfolio management. For example, by focusing on the user requirements and impacts, NOAA’s National Centers for Environmental Information (NCEI) developed product prioritization process and associated metrics to support an organization-wide product portfolio management process (Privette et al. 2017).

This portfolio management should be a part of an organizational data strategy that “ensures that the organization gets the most value from data and has a plan to prioritize data feeds and adapt the strategy to meet unanticipated needs in the future.” (Nelson, 2017). Data strategy needs to be complimentary to and aligned with organizational strategy. Nelson (2017) provides a set of key components for defining data strategy.

d) Data product lifecycle-stages-based maturity assessment models

Ensuring and improving data and metadata quality is an end-to-end process through the entire lifecycle of data products. They are shared responsibilities of all product key players and stakeholders (Peng et al. 2016a). As mentioned above, information quality is multi-dimensional and can be defined based on data product

lifecycle stages (e.g., Ramapriyan et al. 2017). Therefore, defining maturity assessment models for different phases of data products may reflect better the different roles and knowledge required for assessments.

i) Science Maturity Matrix (Define/Develop/Validate)

The scientific quality of data products is closely tied to the maturity of observing systems, product algorithms and production systems. Under the Gap Analysis for Integrated Atmospheric ECV CLimate Monitoring (GAIA-CLIM) project, a measurement system maturity matrix has been developed (Thorne et al. 2015). The measurement systems are categorized as: comprehensive observing networks, baseline networks, and reference networks, based on the observing quality and spatial density (Thorne et al. 2015). This matrix aims to assess the capability maturity of the measurement systems in the following areas: metadata, documentation, uncertainty characterization, public access, feedback, and update, usage, sustainability, and software (optional). The GAIA-CLIM Measurements System Maturity Matrix has been utilized by the GAIA-CLIM project to assess the geographical capabilities in the areas of data and metadata for Essential Climate Variables (ECVs) (Madonna et al. 2016a, 2016b).

The maturity metric of algorithms measures the scientific quality of developing data products and helps establish the credibility of the data products. A data product algorithm maturity matrix (referred to as MM-Algo) has been developed by NOAA's Center for Satellite Applications and Research (STAR) and applied to 68 products from S-NPP (National Polar-orbiting Partnership)/JPSS (Joint Polar Satellite System) as a measure of the readiness of the data product for operational use (Zhou, Divakarla & Liu 2016). The MM-Algo defines five stages of maturity levels for a data product in the areas of validation, documentation, and utility of the product: beta, provisional, validated (Stages 1, 2, and 3) (Reed 2013).

The S-NPP/JPSS Cal/Val program has developed a readiness review process. Information on S-NPP/JPSS data product algorithm maturity including the timeline and associated calibration/validation findings is available at: <https://www.star.nesdis.noaa.gov/jpss/AlgorithmMaturity.php>.

ii) Product Maturity Matrix (Produce/Assess/Deliver)

The use of a maturity matrix approach for individual Earth Science data products was pioneered by Bates and Privette (2012). Bates and Privette (2012) described a product maturity matrix (referred to as MM-Prod) developed by NOAA for satellite-based climate data records (CDRs). CDR MM-Prod provides a framework for evaluating the readiness and completeness of CDR products in six levels in the following six categories: software readiness, metadata, documentation, product validation, public access, and utility. It has been applied to about 35 NOAA CDRs (Bates et al. 2015). The assessments are performed mostly by the CDR data producers during the research-to-operations (R2O) transition process, and are reviewed by NCEI CDR R2O transition managers. (The MM-Prod scoreboard for each CDR can be found at <https://www.ncdc.noaa.gov/cdr>.)

Because the standards defined in CDR MM-Prod, such as data format, are mostly defined for and implemented by NOAA's satellite climate data records program, CDR MM-Prod may need to be generalized for a broader application to digital environmental data products.

A CDR production system maturity matrix, which originated from the CDR MM-Prod but was adapted for the CDR production system, has been developed under the COordinating Earth observation data validation for RE-analysis for CLIMate Services (CORE-CLIMAX) project (EUMETSAT 2013). The CORE-CLIMAX production system maturity matrix assesses whether the CDR can be sustainable in the following six categories: software readiness, metadata, user documentation, uncertainty characterization, public access/feedback/update, and usage. It has been applied to about 40 EU data records of ECV, including satellite, in situ and re-analysis data products (EUMETSAT 2015). It is utilized by the Sustained, Coordinated Processing of Environmental Satellite Data for Climate Monitoring (SCOPE-CM) project of the World Meteorological Organization (WMO) to monitor development processes (Schulz et al. 2015). The CORE-CLIMAX production maturity matrix is the first maturity model that dedicates an entire category to data uncertainty. To some extent, it measures production process quality control capability, as well some aspects of science and product maturity.

iii) Data Stewardship Maturity Matrix (Maintain/Preserve/Disseminate)

A scientific Data Stewardship Maturity Matrix (DSMM, referred to as MM-Stew) was developed jointly by NCEI and the Cooperative Institute for Climate and Satellites – North Carolina (CICS-NC), leveraging institutional knowledge and community best practices and standards (Peng et al. 2015). MM-Stew takes an approach similar to Bates and Privette (2012), but with a different scale structure. MM-Stew defines measurable, five-level progressive stewardship practices for nine key components: preservability, accessibility,

usability, production sustainability, data quality assurance, data quality control/monitoring, data quality assessment, transparency/traceability, and data integrity (Peng et al. 2015).

Over 700 + Earth Science data products were assessed utilizing MM-Stew, with most of the work done manually by NOAA OneStop Metadata Content Editors as a part of the OneStop-ready process (e.g., Ritchey et al. 2016, Hou et al. 2015; Peng et al. 2016b). The OneStop Metadata team, working with the NOAA Metadata Working Group, has developed a workflow and best practices to implement MM-Stew assessment ratings into ISO 19115 collection-level metadata records (Ritchey et al. 2016; Peng et al. 2017). These MM-Stew ratings are integrated into the OneStop data portal and used for discovery and search relevancy ranking. The ISO collection-level metadata records are integrated into NOAA and other catalog services. The detailed justifications for each of the OneStop-ready data products are captured in a data stewardship maturity report (DSMR). A tool has been developed by the OneStop project to systematically generate draft DSMRs with a consistent layout for both figures and DSMR (Zinn et al. 2017). An example of a citable DSMR can be found in Lemieux, Peng & Scott (2017). Both persistent and citable DSMRs and ISO collection-level metadata records can then be readily integrated into or linked by other systems and tools, for example, to be used for improved transparency and enhanced data discoverability. (Links to the MM-Stew related resources including examples of use case studies and ISO metadata record can be found in Peng (2017)). This quantitative and content-rich quality information may be used in decision-making process to support data asset management (e.g., Austin and Peng 2015).

The MM-Stew is designed for digital environmental data products that are extendable and publicly available. This may limit its usage for certain types of data products in specific key components. For example, the Production Sustainability component of MM-Stew may not be useful for data products from one-off research cruises because they are not extendable.

Utilizing the MM-Stew, the Data Stewardship Interest Group (DSIG) under the Committee on Earth Observation Satellites (CEOS) Working Group on Information Systems and Services, led by the European Space Agency, has started to develop a harmonized DSMM (Albani 2016). This harmonized DSMM is based on CEOS data management principles and preservation principles and is intended to be utilized in the Earth observation domain (Maggio 2017). Version 1 of the WGISS data management and stewardship maturity matrix has just been released to the global Earth observation community (WGISS DSIG 2018).

iv) Data Use/Services Maturity Matrix (Use/Services)

Unlike product services in business sector, which are quite mature, the scope of digital scientific data use/services is still evolving. In its levels of services, the National Snow and Ice Data Center (NSIDC) defined a variety of services that can be provided for a data product and described factors that could affect the amount of work required for each individual dataset in six categories: Archival, Metadata, Documentation, Distribution, USO (User Support Office) Infrastructure, and USO Support (Duerr et al. 2009). This list of services levels is designed to provide mechanisms for assessing and providing quantitative guidance on how much effort is required to provide the level of services that is needed for the data product in its current state.

In its tiered scientific data stewardship services, NCEI defined six levels of stewardship services that could be provided for individual data products (see Figure 4 in Peng et al. 2016a). Level 1 stewardship service preserves datasets and provides basic search and access capability for the datasets. With increased levels of stewardship services, more capabilities in scientific quality improvement, enhanced access, and reprocessing are provided, resulting in authoritative records status in Level 5 and providing national and international leadership in Level 6.

The Global Climate Observing System (GCOS) ECV Data and Information Access Matrix provides data and information access to the WMO GCOS ECVs (e.g., <https://www.ncdc.noaa.gov/gosic/gcos-essential-climate-variable-ecv-data-access-matrix>). The Global Ocean Observing System (GOOS) defines three maturity levels of the observing systems in the following three aspects: requirements processes, coordination of observations elements, and data management and information products. The three maturity levels are: concept, pilot, and mature. System performance is evaluated and tracked based on a series of metrics, including implementation, performance, data delivery, and impact metrics. More information on the GOOS framework can be found at: http://goosocean.org/index.php?option=com_content&view=article&id=125&Itemid=113.

A use/service maturity matrix for individual digital environmental data products is under development by the NCEI Service Maturity Matrix Working Group, in collaboration with the Data Stewardship Committee of Earth Science Information Partners (ESIP). The preliminary maturity levels for data monitoring and user engagement have been shared with the ESIP community for community-wide feedback (Arndt and Brewer 2016).

4. Summary

Recent U.S. laws and federal government mandates and policies, along with recommendations and guidelines set forth by federal and other funding agencies, scientific organizations and societies, and scholarly publishers, have levied stewardship requirements on federally funded digital scientific data. This elevated level of requirements has increased the need for a more formal approach to stewardship activities in order to support rigorous compliance verification and reporting. To help data stewardship practitioners, especially these at data centers or associated with data stewardship programs, this article provides an overview of the current state of assessing the maturity of stewardship of digital scientific data. A brief description of the existing maturity assessment models and their applications is provided. It aims at enabling data stewardship practitioners to further evaluate the utility of these models for their unique verification and improvement needs.

Generally speaking, one could utilize:

- the CMMI DMM model if the focus is on assessing and improving organizational processes or capabilities;
- the ISO 16363 model if the focus is on assessing and improving organizational infrastructure or procedures;
- the data product lifecycle-stage-based maturity models if the focus is on assessing and improving practices applied to individual data products.

Any organization can benefit from using a holistic approach to assess and improve the effectiveness of managing its data stewardship activities. Doing so will help ensure that processes are well-defined and procedures are well-implemented using the community best practices.

Given the multi-dimensional and incremental stages of these maturity models, they are not only practical in assessing the current state, identifying potential gaps, and defining a roadmap forward to a desired level of maturity from a certain stewardship perspective. They also offer the flexibility of allowing organizations or stewardship practitioners to define their own process capability requirements or data product maturity levels with a progressive, iterated improvement process or to tailor these models to a particular organizational process area, such as data quality management; or a particular practice applied to individual data products, such as verifying data integrity.

Increased application of these maturity assessment models will help demonstrate their utility and improve their own maturity. This is also beneficial in establishing the need for, and developing a community consensus on, best capturing and integrating quality descriptive information consistently in metadata records or in citable documents for both machine and human end-users. Doing so helps ensure that quality information about federally funded digital data products are findable and integrable, which in turn helps ensure that the data products are preserved for long-term use.

Disclaimer

The description of maturity assessment models in this article is for information only. It does not claim to be comprehensive. Any entity (person, project, program, or institution) is encouraged to do its due diligence and make the use decision based on its own unique needs. Any opinions or recommendations expressed in this article are those of the author and do not necessarily reflect the views of NOAA, NCEI, or CICS-NC.

Acknowledgements

Ge Peng is partially supported by NOAA's National Centers for Environmental Information (NCEI) under Cooperative Agreement NA14NES432003. Travel support for community engagement and feedback related to this work was provided by the NOAA OneStop Program and the Reference Environmental Data Records Program. Lorri Peltz-Lewis, Jörg Schulz, John J. Bates, Lihang Zhou, John Faundeen, Ruth Duerr, Richard Kauffold, Brian Newport, and M. Scott Koger have provided useful information. Communications with them, Hampapuram Ramapriyan, Jeffrey L. Privette, Curt Tilmes, and Sky Bristol are beneficial. Tom Maycock provided beneficial comment on the layout of Figure 1 and reviewed the paper. Brian Newport has painstakingly verified and updated all URLs in the references. Ge Peng thanks the management of CICS-NC and NCEI's Center for Weather and Climate and Data Stewardship Division for their continuing encouragement and support. She also thanks the Data Stewardship Committee and Information Quality Cluster of the Earth Science Information Partners (ESIP) for their continuing interest. Comments and suggestions from Data Science Journal anonymous reviewers are beneficial in improve the readability of the paper.

Competing Interests

The author has no competing interests to declare.

Author Information

Dr. Ge Peng is a Research Scholar at the Cooperative Institute for Climate and Satellite-North Carolina (CICS-NC) of North Carolina State University, which is co-located with the NOAA's National Centers for Environmental Information (NCEI). Dr. Peng holds a Ph. D. in meteorology and is experienced in assessing and monitoring the quality of Earth Science data products. She has extensive knowledge of digital data management and experience in working with metadata specialists and software developers. She is currently leading the effort on development of NOAA sea ice climate normal products and application of the NCEI/CICS-NC Scientific Data Stewardship Maturity Matrix. Dr. Peng has also been coordinating the development of an NCEI data use/service maturity matrix under the NCEI Use/Service Maturity Matrix Working Group. She is an active member of Earth Science Information Partners (ESIP) – a member of its Data Stewardship Committee and co-chair of Information Quality Cluster, where she leads the effort in defining roles and formalizing responsibilities of major product key players and stakeholders for ensuring data quality and improving usability, in collaboration with NCEI.

References

- Albani, M** 2016 ESA – EO data stewardship maturity matrix. *41th CEOS Working Group on Information Systems and Services (WGISS) Meeting*. Canberra, AUS. 14–18 March 2016. [Available online at: <http://ceos.org/meetings/wgiss-41>].
- Arndt, D S** and **Brewer, M** 2016 Assessing service maturity through end user engagement and climate monitoring. *2016 ESIP summer meeting*. 19–22 July 2016. Durham, NC, USA. [Available online at: http://commons.esipfed.org/sites/default/files/Arndt_Brewer_ESIP_Service_Maturity.pdf].
- Austin, M** and **Peng, G** 2015 A prototype for content-rich decision-making support in NOAA using data as an asset. Abstract #IN21A-1676. *AGU Fall Meeting 2015*. 14–18 December 2015. San Francisco, CA, USA.
- Bates, J J** and **Privette, J L** 2012 A maturity model for assessing the completeness of climate data records. *EOS*, 93(44): 441. *Trans. American Geophysical Union*. DOI: <https://doi.org/10.1029/2012EO440006>
- Bates, J J**, **Privette, J L**, **Kearns, E**, **Glance, W G** and **Zhao, X P** 2015 Sustained production of multidecadal climate data records – Lessons from the NOAA Climate Data Record Program. *Bull. Meteor. Soc.* DOI: <https://doi.org/10.1175/BAMS-D-15-00015.1>
- Becker, J**, **Knackstedt, R** and **Pöppelbuß, J** 2009 Developing maturity models for IT management – A procedure model and its application. *Business & Information Systems Engineering*, 3: 213–222. DOI: <https://doi.org/10.1007/s12599-009-0044-5>
- CCSDS** 2012a Reference Model for an Open Archival Information System (OAIS), Recommended Practices, Issue 2. Version: CCSDS 650.0-M-2. 135 [Available online at: <https://public.ccsds.org/pubs/650x0m2.pdf>].
- CCSDS** 2012b Audit and certification of trustworthy digital repositories – Recommended Practices. Version: CCSDS 652.0-M-1. 77. [Available online at: <https://public.ccsds.org/pubs/652x0m1.pdf>].
- CMMI** 2014 Data Management Maturity Model. *CMMI Institute*, 248. Version: 1.0 August 2014. Accessed: 7 July 2017.
- COPDESS** 2015 Statement of commitment from Earth and Space Science publishers and data facilities. Version. January 14. [Available online at: <http://www.copdess.org/statement-of-commitment/>].
- Duerr, R**, **Leon, A**, **Miller, D** and **Scott, D J** 2009 Level of services. *National Snow and Ice Data Center*. Version: v2 7/31/2009. [Available online at: [https://nsidc.org/sites/nsidc.org/files/files/NSIDCLevelsofService-V2_0a\(2\).pdf](https://nsidc.org/sites/nsidc.org/files/files/NSIDCLevelsofService-V2_0a(2).pdf)].
- EDMC (The Enterprise Data Management Council)** 2015 Data Management Capability Assessment Model. Version: v1.0. July 2015.
- Edmunds, R**, **L'Hours, H**, **Rickards, L**, **Trilsbeek, P** and **Vardigan, M** 2016 Core Trustworthy Data Repository Requirements. [Available online at: https://zenodo.org/record/168411-WV_NQBPYuSN]. DOI: <https://doi.org/10.5281/zenodo.168411>
- EUMETSAT** 2013 CORE-CLIMAX Climate Data Record Assessment Instruction Manual. Version: 2. 25 November 2013. [The latest version is available online at: http://www.sat.ltu.se/members/viju/publication/core-clim/Core-Climax_System_Maturity_Matrix_Instruction_Manual.pdf].
- EUMETSAT** 2015 CORE-CLIMAX European ECV CDR Capacity Assessment Report. Version: v1. 26 July 2015.
- Faundeen, J**, **Kirk, K** and **Brown, C** 2017 Certifying Trusted Digital Repositories: USGS Use Case. *2017 ESIP Summer Meeting*. 25–28 July 2017. Bloomington, IN, USA.

- FGDC** 2015 National Geospatial Dataset Asset Management Plan – Lifecycle Maturity Assessment Tool. [Available online at: https://cms.geoplatform.gov/sites/default/files/a16themeleads/ActiveDocuments/1_NGDA_BaselineAssessment_01_IntroAndAssessment_FINAL.pdf].
- FGDC** 2016 National Geospatial Data Asset Lifecycle Maturity Assessment 2015 Report – Analysis and Recommendations. Version. 8 December 2016. 93.
- GEO Data Sharing Working Group** 2014 GEOS Data Sharing Principles Post 2015. Version. March 10. [Available online at: http://www.earthobservations.org/documents/dswg/10_GEOS%20Data%20Sharing%20Principles%20post%202015.pdf].
- Gorball, J** 2016 Introduction to Data Management Maturity Models. *Slideshare*. Version. 28 July 2016. [Available online at: https://www.slideshare.net/Kingland_Systems/introduction-to-data-management-maturity-models].
- Hou, C-Y, Mayernik, M, Peng, G, Duerr, R and Rosati, A** 2015 Assessing information quality: Use cases for the data stewardship maturity matrix. Abstract #IN21A-1675. *AGU 2015 Fall Meeting*. 14–18 December 2015. San Francisco, CA, USA.
- ISO 16363** 2012 Space data and information transfer systems – Audit and certification of trustworthy digital repositories. Version. ISO 16363:2012. Geneva, Switzerland.
- Lemieux, P, Peng, G and Scott, D J** 2017 Data Stewardship Maturity Report for NOAA Climate Data Record (CDR) of Passive Microwave Sea Ice Concentration. Version 2. *figshare*.
- Madonna, F, Thorne, P, Rosoldi, M, Tramutola, E, Buschmann, M and DeMaziere, M** 2016a Report on data capabilities by ECV and by system of systems layer for ECVs measurable from space. *GAIA-CLIM Deliverable D1.6*. Version. 7 September 2016. [Available online at: http://www.gaia-clim.eu/system/files/workpkg_files/D1.6%20Report%20on%20data%20capabilities%20by%20ECV%20and%20by%20systems%20of%20systems%20layer.pdf].
- Madonna, F, Tramutola, E, Rosoldi, M, Thorne, P, Meier, A and Rannat, K** 2016b Report on the collection of metadata from existing network and on the proposed protocol for a common metadata format. *GAIA-CLIM Deliverable D1.7*. Version. 14 September 2016. [Available online at: http://www.gaia-clim.eu/system/files/workpkg_files/D1.7%20Report%20on%20the%20collection%20of%20metadata%20from%20existing%20network.pdf].
- Maggio, I** 2017 DMP IG as a Maturity Matrix. *44th CEOS WGISS meeting*. 03–06 April. Annapolis, MD, USA. [Available online at: <http://ceos.org/meetings/wgiss-43>].
- McSweeney, A** 2013 Review of data management maturity models. Version. 23 October 2013. *Slideshare*. [Available online at: <https://www.slideshare.net/alanmcsweeney/review-of-data-management-maturity-models>].
- Mecca, M** 2015 CMMI Data management maturity model ecosystem and deep dive. Version. April 21. [Available online at: <https://damapdx.org/docs/dama/Apr2015MelanieMecca.pdf>].
- Mosely, M, Brackett, M, Early, S and Henderson, D** (Eds.) 2009 *The Data Management Body of Knowledge (DAMA-DMBOK Guide)*, 406. Bradley Beach, NJ, USA: Technics Publications, LLC. 2nd Print Edition.
- NAS (National Academy of Sciences)** 2009 Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age. *National Academies Press*, 178. DOI: <https://doi.org/10.17226/12615>
- National Research Council** 2007 Environmental data management at NOAA: Archiving, stewardship, and access. 130. *The National Academies Press*. Washington, D.C. DOI: <https://doi.org/10.17226/12017> [Available online at: <https://www.nap.edu/catalog/12017.html>].
- Nelson, G S** 2017 Developing Your Data Strategy: A practical guide. Paper 0830-2017. ThotWave Technologies, Chapel Hill, NC. [Available online at: <http://support.sas.com/resources/papers/proceedings17/0830-2017.pdf>].
- NOAA (National Oceanic and Atmospheric Administration)** 2010 NOAA Administrative Order 212-15 – Management of environmental and geospatial data. 4. [Available online at: http://www.corporateservices.noaa.gov/ames/administrative_orders/chapter_212/212-15.pdf].
- OMB** 1999 Uniform Administrative Requirements for Grants and Agreements with Institutions of Higher Education, Hospitals, and Other Non-Profit Organizations. *OMB Circular A-110*. [Available online at: <https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/circulars/A110/2cfr215-0.pdf>].
- OMB** 2002a Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies. *Federal Register*, 67(36). OMB Notice February 22. [Available online at: https://obamawhitehouse.archives.gov/omb/fedreg_final_information_quality_guidelines].
- OMB** 2002b Coordination of Geographic Information and Related Spatial Data Activities. *OMB Circular A-16*. Revised. [Available online at: <https://www.whitehouse.gov/omb/circulars>].


- OMB** 2013 Open Data Policy – Managing Information as an Asset. Version: OMB Memorandum May 9. [Available online at: <https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>].
- OSTP** 2010 Scientific integrity. [Available online at: <https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/scientific-integrity-memo-12172010.pdf>].
- OSTP** 2013 Increasing access to the results of federally funded scientific research. Version: OSTP Memorandum. February 22. [Available online at: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf].
- Peltz-Lewis, L A, Blake-Coleman, W, Johnston, J and DeLoatch, I B** 2014 National Geospatial Data Asset Lifecycle Baseline Maturity Assessment for the Federal Geographic Data Committee. Abstract IN13A-3635, *AGU 2014 Fall Meeting*, 15–19. December 2014. San Francisco, CA, USA.
- Peng, G** 2017 Getting to know and to use the NCEI/CICS-NC data stewardship maturity matrix (DSMM). *Figshare*. Version. August 25. DOI: <https://doi.org/10.6084/m9.figshare.5346343>
- Peng, G, Lawrimore, J, Toner, V, Lief, C, Baldwin, R, Ritchey, N A, Brinegar, D and Delgreco, S A** 2016b Assessing Stewardship Maturity of the Global Historical Climatology Network-Monthly (GHCN-M) Dataset: Use Case Study and Lessons Learned. *D-Lib Magazine*, 22. DOI: <https://doi.org/10.1045/november2016-peng>
- Peng, G, Privette, J L, Kearns, E J, Ritchey, N A and Ansari, A** 2015 A unified framework for measuring stewardship practices applied to digital environmental datasets. *Data Science Journal*, 13. DOI: <https://doi.org/10.2481/dsj.14-049>
- Peng, G, Ritchey, N A, Casey, K S, Kearns, E J, Privette, J L, Saunders, D, Jones, P, Maycock, T and Ansari, S** 2016a Scientific stewardship in the Open Data and Big Data era – Roles and responsibilities of stewards and other major product stakeholders. *D-Lib Magazine*, 22. DOI: <https://doi.org/10.1045/may2016-gepeng>
- Peng, G, Ritchey, N, Milan, A, Zinn, S, Casey, K S, Neufeld, D, Lemieux, P, Ionin, R, Partee, R, Collins, D, Shapiro, J, Rosenberg, A, Jaensch, T and Jones, P** 2017 Towards Consistent and Citable Data Quality Descriptive Information for End-Users. 2017 DataOne Annual Meeting. *figshare*. DOI: <https://doi.org/10.6084/m9.figshare.5336191>
- Privette, J L, Cramer, B, Ellingson, G, Ellingson, L, Hutchins, C, McPherson, T and Wunder, D** 2017 Evolving NOAA's Climate Data Portfolio in Response to Validated Sectoral Needs. *AMS 97th Annual Meeting*, 22–26 January 2017. Seattle, WA, USA.
- Ramapriyan, H K, 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>
- Reed, B** 2013 Status of Operational Suomi NPP Algorithms. *The first Community Satellite Processing Package Use Group Meeting*. 21–23 May 2013. Madison, WI, USA. [Available online at: <http://www.ssec.wisc.edu/meetings/cspp/2013/presentations/Reed-OperationalSNPPAlgorithmsStatus.pdf>].
- Ritchey, N A, Peng, G, Jones, P, Milan, A, Lemieux, P, Partee, R, Lonin, R and Casey, K A** 2016 Practical Application of the Data Stewardship Maturity Model for NOAA's OneStop Project. Abstract #IN43D-08. *2016 AGU Fall Meeting*. 12–16 December 2016. San Francisco, CA, USA.
- Schulz, J and 14 others** 2015 System maturity assessment. Copernicus Workshop on Climate Observation Requirements. *ECMWF*. Reading, 29 June–2 July 2015.
- Stall, S** 2016 AGU's Data Management Maturity Model. *SciDataCon 2016*. 11–13 September. Denver, CO, USA.
- Thorne, P, Schulz, J, Tan, D, Ingleby, B, Madaoona, F, Pappalardo, G and Oakley, T** 2015 GAIA-CLIM Measurement Maturity Matrix Guidance: Gap Analysis for Integrated Atmospheric ECV Climate Monitoring: Report on system of systems approach adopted and rationale. Version. 27 Nov 2015. [Available online at: http://www.gaia-clim.eu/system/files/workpkg_files/640276_Report%20on%20system%20of%20systems%20approach%20adopted%20and%20rationale.pdf].
- US Public Law 106-554** 2001 Information Quality Act. *Publ. L. 106-554*, 101. [Available online at: <http://www.gpo.gov/fdsys/pkg/PLAW-106publ554/html/PLAW-106publ554.htm>].
- US Public Law 107-347** 2002 Federal Information Security Management Act. *Pub.L. 107-347*. [Available online at: <http://www.gpo.gov/fdsys/pkg/PLAW-107publ347/html/PLAW-107publ347.htm>].
- Valen, D and Blanchat, K** 2015 Overview of OSTP Responses. *figshare*. Access date: March 5. DOI: <https://doi.org/10.6084/m9.figshare.1367165>

- WDS Scientific Committee** 2015 WDS Data Sharing Principles. Version: v1. November 2015. DOI: <https://doi.org/10.5281/zenodo.34354>
- WGISS DSIG** 2018 WGISS Data stewardship maturity matrix. Document ID: CEOS.WGISS.DSIG.DSMM. Version: 1.0. 26 September 2017.
- Wilkinson, M D** and **51 others** 2016 The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3. DOI: <https://doi.org/10.1038/sdata.2016.18>
- Witt, M, Kroll, M, Minor, D** and **Relly, B** 2012 ISO 16363: Trustworthy Digital Repository Certification in Practice. *7th International Conference on Open Repositories*. 9–13 July 2012. Edinburgh, Scotland, UK. [Available online at: http://docs.lib.purdue.edu/lib_fspres/4].
- Zhou, L H, Divakarla, M** and **Liu, X P** 2016 An Overview of the Joint Polar Satellite System (JPSS) Science Data Product Calibration and Validation. *Remote Sensing*, 8(2). DOI: <https://doi.org/10.3390/rs8020139>
- Zinn, S, Relph, J, Peng, G, Milan, A** and **Rosenberg, A** 2017 Design and implementation of automation tools for DSMM diagrams and reports. 2017 ESIP winter meeting. January 11–13, in Bethesda, MD. [Available online at: http://commons.esipfed.org/sites/default/files/Zinn_etal_OneStop_DSMM_ESIP%2020170113_0.pdf].

How to cite this article: Peng, G 2018 The State of Assessing Data Stewardship Maturity – An Overview. *Data Science Journal*, 17: 7, pp. 1–12, DOI: <https://doi.org/10.5334/dsj-2018-007>

Submitted: 05 December 2017 **Accepted:** 05 March 2018 **Published:** 26 March 2018

Copyright: © 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 