

PRACTICE PAPER

Fitness for Use of Data Objects Described with Quality Maturity Matrix at Different Phases of Data Production

Heinke Höck^{1,2}, Frank Toussaint^{1,2} and Hannes Thiemann^{1,2}

¹ Data Management, World Data Center for Climate (WDCC), DE

² Deutsches Klimarechenzentrum (DKRZ), Hamburg, DE

Corresponding author: Heinke Höck (hoeck@dkrz.de)

Fitness for use information should be stored to enable easy identification of data objects that are suitable for re-use – a feature which can only be assessed by the data user. With the described Quality Maturity Matrix (QMM), we want to provide a metric for a discrete measurement of the fitness for use of data objects. We use the data maturity to describe the degree of formalization and standardization of the data with respect to the quality of data and metadata. The data objects mature as they pass through the different post-production steps where they undergo different curation measures. The higher the maturity and the level in the QMM, the easier is it for the user to judge the appropriateness of the data for a possible re-use.

For our development of the Quality Maturity Matrix we link the maturity levels to the five phases concept, production/processing, project collaboration/intended use, long-term archiving, and impact re-use. Each of the five levels is measured with regard to the four criteria consistency, completeness, accessibility, and accuracy. For the description we use the terms of the Open Archival Information System (OAIS).

We relate our data focused QMM to some existing maturity matrices which put the focus on the maturity of the curation process rather than of the data objects themselves. In addition, we make an attempt to establish a connection between the QMM criteria of data assessment and the FAIR Data principles.

Keywords: data management; Fitness for Use; data production steps; Quality Maturity Matrix; FAIR Data Principles

1 Introduction

1.1 Background

In recent years, the digital storage of data objects (data and metadata) in most scientific fields faced new challenges. Big projects mostly are global collaborations and require an increasing orientation to international standards. The data storage techniques in parallel shifted to federated, distributed storage systems like the Earth System Grid Federation (ESGF¹) for climate model data. For the long term archival (LTA) on the other hand, communities, funders, and data users make stronger demands for data and metadata quality to facilitate data use and re-use. Thus, for the efficient re-use of data objects, the metadata should contain the maximum possible information to judge on the data's fitness for use from the users view – the intended user as well as any not yet known re-users.

For the assessment of data objects, stakeholders from academia, industry, funding agencies, and scholarly publishers have formally defined and endorsed a set of FAIR Data Principles (Wilkinson et al. 2016).

At the same time, on the other hand, there was growing interest of scientists, journals, and other parts of the community, to assess the quality of different approaches to research data management (RDM). This led

¹ esgf.llnl.gov.

to a rising importance of RDM assessment systems, for which already several layouts of maturity matrices have been developed and some of them published (Crowston & Quin 2012, Peng et al. 2015). For example regarding repositories, the Core Trust Seal certification² is a standardised assessment. The World Data System³ (WDS) of the International Science Council (ISC) requires this certificate as a condition of membership. Additionally, applicants need to show their compliance with the WDS' strong commitment to "open data sharing, data and service quality, and data preservation".

These two assessment types, for data objects and for RDM techniques, have various overlaps, as the assessment of a repository is not independent of the quality of its data. The assessment of data, on the other hand, depends on the procedure of data curation which can be evaluated with regard to several criteria.

These curation criteria, in turn, well might differ for different data objects in the storage of one single repository dedicated to different scientific communities. This was already pointed out by Treloar & Harboe-Ree (2008) who described this situation systematically by identifying eight different curation criteria in which the data undergoes changes during its maturation process. As these changes mostly represent an evolutionary process, Treloar & Harboe-Ree refer to Curation Continua. Furthermore, they divided a data object's evolution into the three phases Private Research, Shared Research, and Public.

However, the term maturity is problematic in a different respect as was discussed by Cox et al. (2017): 'It might be taken to imply a single development path leading to a fixed mature finishing place. This is not normally the case. Also, terms like *immature* or *underdeveloped*, sometimes associated with maturity models might be seen as pejorative.' This issue can be adopted for the term *quality*, as well.

An interaction between Treloar's domain model and the maturity matrix (see below **Table 1**) may improve this situation because the workflow does not terminate at impact/re-use. Maturity and quality depend on the phase. Each phase has its own options for maturity and quality. For example, the persistence of data differs between the private production domain with local storage and the mostly public long-term phase in a long-term archive.

This shows that 'data products produced by the same organization are often in various levels of maturity in terms of their data quality, accessibility, and usability as well as the states of completeness of data quality metadata and documentation' (Peng et al. 2015).

1.2 The users' needs – early access and tools

Peng et al. (2016) describe the requirements and challenges for the Public Domain: 'as data are increasingly treated as valuable assets for decision-makers, decision support based on fast data analysis has made ensuring data quality a critical but challenging task. Therefore, having tools available is not just helpful but a necessity for effectively stewarding and serving digital scientific data. Those tools allow data and scientific stewards to effectively capture, describe, and convey data quality information'. Other helpful tools allow users to view data products before they request aggregated or subsetted data for their specific applications or they automatically select metadata from file headers.

However, any type of tools and other forms of automated processing require rich metadata. Partly for technical reasons, partly to bring a complete picture of the data into the user's view. So metadata is an important criterion for the data quality in terms of data re-use.

Additionally, during the project's runtime of shared research, scientists want to use the data product at the earliest stage possible and often settle themselves with poor information on the data, i.e. with poor

Table 1: Assignment of the DKRZ data dissemination system to the domains as described by Treloar & Harboe-Ree (2008).

Domain	Phase	DKRZ system
Research preparation phase	Concept generation	data management (DM) planning tool RDMO
Private Research	production/processing	DKRZ storage on hard disc and tape HPSS ⁴
Shared Research	project collaboration intended use	ESGF, globally distributed project repository
Public	long-term archiving impact re-use	Long-term Archive

² <https://www.coretrustseal.org>.

³ <https://www.worlddatasystem.org>.

⁴ <https://www.dkrz.de/up/systems/hpss/hpss>.

metadata quality. To cover these needs, some projects might want to publish the data in an early and poorly conceived stage.

For the Shared Research Domain and Public Domain, it is important to differentiate between the intended use and the actual re-use of the data because this reflects that with increases in maturity, the data curation processes become more refined, institutionalized, and standardized (Crowston & Qin 2012). The data producers/providers have the knowledge to deal with their raw data. So the need for more standardisation mainly arises from the requirements of the re-use.

At the German Climate Computing Centre (Deutsches Klimarechenzentrum, DKRZ⁵), this situation led to a twofold dissemination system for the data, which follow the internal storage on project hard disks. The ESGF data dissemination system focuses on the needs of users as partners in globally running projects which want to access the data during the project collaboration – this is the intended use phase. The DKRZ's digital long-term archive (DKRZ-LTA), in contrast, aims for long term data holding and data re-use far beyond the project run time. This requires high generic standards for the metadata quality.

In this paper we describe the development of a Maturity Matrix assessment for the quality of data and metadata (Chapter 2. and 3.). We present the different criteria used at DKRZ to rate the maturity of data during the data curation (Chapter 4.).

2 The Quality Maturity Matrix

2.1 Initial Considerations

In addition to using maturity matrices for RDM services and capabilities, this technique also has been applied to other areas as pointed out by Cox et al. (2017). Examples are software engineering (Paulk et al. 1993), digital preservation (Kenney & McGovern 2003), and data intensive research (Lyon et al. 2012). Maturity models have also been applied in the RDM space, within institutions (ANDS 2011), and within research projects (Crowston & Qin 2012). In Peng (2018) the reader finds a good overview of the current state of assessing the maturity of stewardship of digital scientific data.

In 2015 we took the initiative to apply the Quality Maturity Matrix (QMM) technique to implement quality assessment for the digital storage workflow of climate model data at DKRZ. The first application field for implementing the QMM was the DKRZ-LTA World Data Center for Climate (WDCC⁶) sector. We decided to use the **System Maturity Matrix** (SMM⁷) of the German Weather Service (DWD) and EUMETSAT, which collaborate on climate related data products. The DWD in turn cooperates with us in many climate modelling projects. So the SMM became the initial point of the QMM development at DKRZ.

For the DKRZ QMM, the quality with respect to the repository itself was not considered, e.g., persistency of access and physical reliability of long term storage. This should be done by the repository certifications as for instance the CoreTrustSeal. We took those aspects out of our QMM scheme that are contained in published stewardship maturity matrices, as for example described by Peng et al. (2015) and focus on the quality of stewardship. This reflects that the QMM is intended to be used purely for data objects. It can be used by anyone interested in the quality of data objects and is not limited to science and research.

2.2 Initial points for the development of a Maturity Matrix for data

Peng (2018) provides us with the motivation to develop a Maturity Matrix for data stewardship and other processes, using the maturity model description: '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'. This can be expanded to similar techniques for the assessment of data objects like the DKRZ QMM.

This motivation led us to the following points for the development of a Maturity Matrix for data objects:

- Definition of a level system for quality assessment

The levels correspond to the discrete stages Peng (2018) announced for the maturity model. This is the metric we have referred to in the background chapter.

⁵ www.dkrz.de.

⁶ wdc-climate.de.

⁷ <https://www.eumetsat.int/website/home/Data/TechnicalDocuments/index.html>.

- Description and definition of data object quality for data and metadata
In combination with the OAIS terms (CCSDS 2012) and the QMM we want to provide an assessment frame for data object quality.
- Support the estimation of data usability for the re-users' intended application
We are using the ISO19157 (2013) explanation of data usability: Usability is based on user requirements. Here ISO 19157 distinguishes between rationale – the intended application for creating a data set – and the user requirements for a particular application – the re-use.
Furthermore, 'non-quantitative information is illustrative for users and can help assessing the quality of a data set, especially in cases where it is used for a particular application that differs from the intended application' (ISO 19157:2013). We implemented most of the information on usability by the QMM criterion 'Accuracy', specifically by the topic 'references to evaluation results (data) and methods'.

3 Description of the Quality Maturity Matrix for Data Management at DKRZ

The two dimensions of the QMM are the levels and criteria/aspects. The levels and their characteristics are given in **Figure 1**. The QMM levels are numbered 1 to 5. The QMM criteria are consistency, completeness, accessibility, and accuracy; these are described fully in section 4 below.

The criteria are developed to support the phases of the data production: concept, production/processing, project collaboration/intended use, long-term archiving and impact/re-use.

In this context, support means description of experiential knowledge which makes sense in the context of data production steps and helps to reach the next level. This is the coarse outline of the levels. Data in the production phase are able to reach level 5 if the aspects' criteria are fulfilled.

In **Figure 1** the level colouring of the QMM levels changes from red for making a concept for the data to green for data of the highest degree of maturation.

3.1 The way to DKRZ's QMM

The initial point of the QMM development at DKRZ was the DWD/EUMETSAT **System Maturity Matrix (SMM⁷)** which is used for monitoring the process of generating Climate Data Records for satellite data (Bates & Privette 2012).

The monitoring aspect of SMM is missing in the QMM approach because software readiness (SMM criterion) does not change after the climate model data and associated data (observations) have been transferred to the long-term archive of WDCC. The documentation of the methods could make progress, but the software itself cannot (e.g. coding standards) (**Table 2**).

Other modification aspects from SMM to QMM definition are:

- One main matrix, no submatrices, but outsourcing of details in checklists (Höck 2019a) to adjust details into informal rules (level 2), project requirements (level 3), long-term archive requirements and discipline-specific standards (level 4), and interdisciplinary standards (level 5)
- Generic descriptions in the matrix cells
- Depiction of guidance on how to reach the next level (Höck 2019a) with connected common cells and transposed matrix

Quality Maturity Matrix Level	Level 1	Level 2	Level 3	Level 4	Level 5
Data Production Steps/Phases	concept	production processing	project collaboration intended use	long-term archiving	impact re-use
Characteristic	not evaluated	informal	data management plan	discipline-specific standards	interdisciplinary standard

Figure 1: Characteristics of data and metadata Quality Assurance Maturity Levels. QMM levels corresponding a) to different steps of the data production workflow and b) to the five data production phases with their standardisation characteristics and increasing degrees of formalisation.

- The quality with respect to the repository itself was not considered, e.g., persistency of access and physical reliability of long term storage. This should be done by repository certifications like CoreTrustSeal², the nestor Seal,⁸ or an ISO 16363 certification.
- The criterion accuracy is part of the level 3 evaluation process with the provision of associated documents and/or information about:
 - (i) procedure about methodological and technical sources of errors and deviation/inaccuracy,
 - (ii) procedure with validation against independent data (if feasible),
 - (iii) evaluation results (data) and methods of data production,
 - (iv) indication of missing values,
 - (v) procedure of statistical quality control of the data.

Most of these modifications were included in the first draft of the DKRZ QMM, as we reported in a presentation (Höck et al. 2015).

In addition, we adapted the relevant terms to the reference model of the Open Archival Information System (OAIS, **Figure 2**) and we implemented the *OAIS Preservation Description Information* (PDI,

Table 2: Shows a comparison of SMM and QMM.

SMM	QMM
Software Readiness	Omitted: the data object is considered as persistent. Software development would lead to new data objects except software documentation. That is part of the metadata provenance.
Metadata	Criterion: Completeness Aspect: Existence of Metadata
User Documentation	Criterion: Completeness Aspect: Existence of Metadata
Uncertainty Characterisation	Criterion: Accuracy
Public Access/Feedback/Update	Criterion: Accessibility/Criterion: Completeness Aspect: Existence of Metadata level 5: data provenance chain exists including internal and external objects e.g. software, articles, method and workflow description/Criterion: Consistency Aspect: Versioning and Controlled Vocabularies (CVs)
Usage	Omitted: we use the ISO19157 explanation of data usability. It depends on the 'particular application'. From this point of view, an evaluation of usage is not possible.

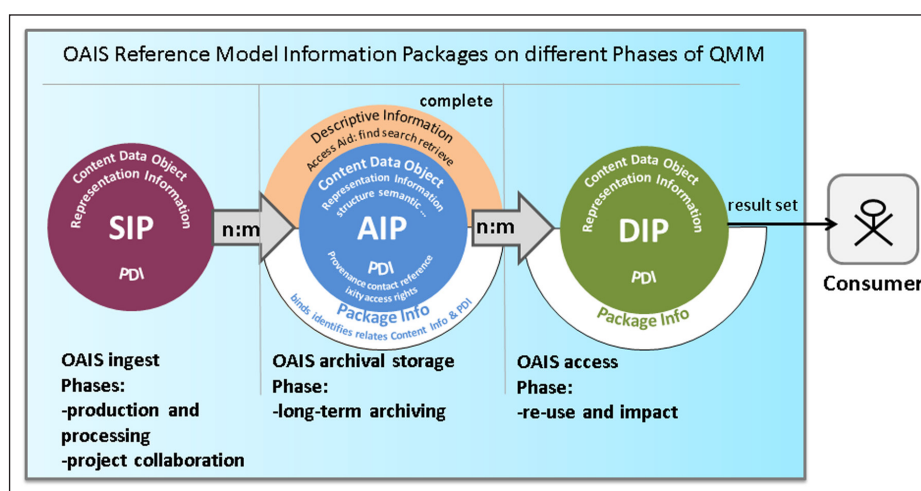


Figure 2: OAIS Reference Model Information Packages on different Phases of the QMM process, showing the submission (SIP), archival (AIP), and dissemination information packages (DIP).

⁸ https://www.langzeitarchivierung.de/Webs/nestor/EN/Zertifizierung/nestor_Siegel/siegel.html.

Figure 3) as obligate where applicable. The latter should be a minimum set of metadata in the long-term archive, which should accompany the Content Data Object (CDO).

4 The Criteria of the QMM levels

For the Quality Maturity Matrix we regard the four criteria consistency, completeness, accessibility, and accuracy. Each of these criteria is subdivided into aspects, for example, for Completeness the aspects are 'Existence of Data' and 'Existence of Metadata', as shown in **Table 3**.

One of the ways to obtain the best possible re-use (and impact) of data objects is to make data FAIR. In this respect we are guided by the interpretation of Mons et al. (2017) for the European Open Science Cloud (EOSC): '...as long as such data are clearly associated with FAIR metadata, we would consider them fully participating in the FAIR ecosystem.'

All in all, the FAIR mission statement consists of 15 aspects. With the QMM, one can assess to which degree FAIR Data Principles are fulfilled for a data object and data can therefore be marked as FAIR.

As the criteria and the levels of the QMM represent a matrix and also for space reasons, a presentation in tabular form was chosen for the following subsections. In the four tables (**Tables 4 to 7**), we give an overview of the different factors relevant for the four criteria and their aspects. Connections between the QMM and the FAIR Data Principles of Wilkinson (2016) are presently light faced.

Content Information Type	Reference	Provenance	Context	Fixity	Access Rights
DKRZ-LTA Data	Citation- Creators Publication Year Title DOI Publisher Persistent Identifier - DOI Journal reference	Data source Creators and contributors Description of data production steps and methods	Mission Funding history Pointer to related project description in original environment	Checksum and data object size protection against citation alteration	Access and use constraints Terms of use

Figure 3: DKRZ Long Term Archive – example of minimum metadata (PDI, following the OAIS reference model).

Table 3: Overview of the QMM quality criteria and sub-criteria (aspects).

Criterion	Aspect
Consistency	Data Organisation and Data Object
	Versioning and Controlled Vocabularies (CVs)
	Data-Metadata Consistency
Completeness	Existence of Metadata
	Existence of Data
Accessibility	Metadata Access by Identifier
	Data Access by Identifier
Accuracy	Plausibility
	Statistical Anomalies

4.1 Criterion: Consistency

Table 4: QMM criterion consistency.

Level 1	Level 2	Level 3	Level 4 R1.2	Level 5
Aspect: Data Organisation and Data Object				
conceptual development	<i>data organisation is structured/conform to</i> internal rules informal documented	project specification	well-defined rule e.g. discipline-specific standards and long-term archive requirements (OAIS Package Info -binds)	interdisciplinary standards
<i>data objects (OAIS) are</i>				
	SIPs consistent to internal rules	SIPs correspond to project requirements	I1, I2 AIPs conform to well-defined rules e.g. discipline-specific standards and long-term archive requirements	AIPs conform to interdisciplinary standards up-to-date and consistent to external scientific objects if feasible DIPs are fully machine-readable with references to sources
<i>data formats – Content Data Object (OAIS)</i>				
	file extensions are consistent	data sizes are consistent	I1 DIPs datasets are self-describing I1 conform to well-defined rules e.g. discipline-specific standards and long-term archive requirements	conform to interdisciplinary standards

(Contd.)

Level 1	Level 2	Level 3	Level 4 R1.2	Level 5
Aspect: Versioning and Controlled Vocabularies (CVs)				
conceptual development	<i>versioning follows/is</i> internal rules informal documented	systematic corresponds to project requirements	systematic collection including documentation of enhancement conform to well-defined rules old versions stored if feasible	In case new versions are published: documentation is consistent to previous versions
data labelled with CVs conform to				
	informal CVs if feasible	formal project defined CVs if feasible	11, 12 discipline-specific standards	interdisciplinary standards
Aspect: Data-Metadata Consistency				
not evaluated	OAIS metadata components are consistent PDI components: Provenance- unsystematically documented: Reference- creators	PDI components: Provenance – basically documented: Reference –creators contact Descriptive Information -naming conventions for discovery – find and search	Complete PDI * Provenance Context Reference – cross Fixity Access Rights and Representation Information Descriptive Information Package Info	13 external meta-data and data are consistent
*maintenance and storage policy are not affected, since they belong to the repository certification.				

4.2 Criterion: Completeness

Table 5: QMM criterion completeness.

Level 1	Level 2	Level 3	Level 4 R1.2	Level 5
Aspect: Existence of Data (Completeness and Persistence)				
not evaluated	data is in production and may be deleted or overwritten	datasets exist, not complete and may be deleted but not overwritten unless explicitly specified	data entities (conform to discipline-specific standards) are complete dynamic datasets – data stream are not affected number of datasets (aggregation) is consistent data are persistent, as long as expiration date requires	data entities (conform to interdisciplinary standards) are complete dynamic datasets – data stream are not affected number of datasets (aggregation) is consistent data are persistent, as long as expiration date requires
Aspect: Existence of Metadata				
not evaluated	OAIS metadata components exist PDI components: Provenance- unsystematically documented Reference- creators	PDI components: Provenance – basically documented: Reference –creators contact Descriptive Information: naming conventions for discovery – find and search	F2, R1 Complete PDI * R1.2 Provenance Context Reference Fixity Access Rights and Representation Information R1.1 Descriptive Information F4 Package Info	metadata is conform to interdisciplinary standards data provenance chain exists including internal and external objects e.g. software, articles, method and workflow description
*maintenance and storage policy are not affected, since they belong to the repository certification.				

4.3 Criterion: Accessibility

Table 6: QMM criterion accessibility.

Level 1	Level 2	Level 3	Level 4 R1.2	Level 5
Aspect: Data Access by Identifier				
not evaluated	file names data is accessible by	internal unique identifier correspond to project requirements	permanent identifier (expiration is documented) (OAIS Package Info – identifies) datasets have an expiration date and are accessible for at least 10 years (conform to rules of good scientific practice)	F1, A1 global resolvable identifier (PID-persistent identifier) registered with resolving to data access including backup where it is commonly accepted that the identifier is persistently resolvable at least to information about fate of the object data is accessible within other data infrastructures including cross references
checksups are accessible				
a bijective mapping between identifier and datasets is documented e.g. in data header (OAIS Package Info – binds, identifies)				
Aspect: Metadata Access by Identifier				
not evaluated	not specified metadata is accessible by	internal unique identifier correspond to project requirements	by permanent identifier expiration is documented (F4 OAIS Package Info – identifies) complete data citation is persistent	F1, A1 global resolvable identifier including backup complete data citation is persistent I3 external PID references are supported
a mapping between data access identifier and metadata access identifier is implemented (OAIS Package Info relates Content Info and PDI)				

4.4 Criterion: Accuracy

Table 7: QMM criterion accuracy.

Level 1	Level 2	Level 3	Level 4	Level 5
Aspect: Plausibility				
not evaluated	R1 documented procedure about technical sources of errors and deviation/inaccuracy exists (data header and content is consistent)			
		R1 documented procedure about methodological sources of errors and deviation/inaccuracy documented procedure with validation against independent data		
		R1 references to evaluation results (data) and methods exist		
Aspect: Statistical Anomalies				
not evaluated	R1 missing values are indicated e.g. with fill values			
		R1 documented procedure of statistical quality control is available		
			scientific consistency among multiple data sets and their relationships is documented if feasible	

5 Evaluation Process by the QMM

The evaluation planning and implementation involve the feasibility of the evaluation for the specific level.

At DKRZ, we first identify which QMM level the data object has reached when it is submitted to us. For a more detailed level evaluation, implementation check lists are provided (Höck 2019a) to assess whether or not the criteria are successfully obtained for the specific level.

Once the positive outcome of the level evaluation is confirmed, we offer to the user guidance to enhance the data object to the next level of maturity.

For the implementation of the evaluation process at the WDCC, the submission process has so far been analysed for model data. The WDCC has to check some points to those it normally carries out in the workflow of the data object submission process. We found out that the following check points are sufficient to reach at least QMM level 4 (Höck 2019b).

- 1) The WDCC LTA submission process of data and metadata must have been completed (completely archived) with DOI assignment.
- 2) The data format of the model data should be netCDF CF⁹ or WMO GRIB¹⁰
- 3) The accuracy has to be documented with the model description in case of model data.
- 4) The DIP data size must be appropriate and the use constraint (license data providers apply to their particular data) should be checked.
- 5) The documentation of submission agreement between DKRZ-LTA and the project should be available.

This corresponds to the FAIR presentation of the (meta)data at WDCC LTA (DKRZ-User Portal 2019). The (meta)data is FAIR with the exception of guiding topic 13 (Wilkinson 2016) under the sufficient but not necessary conditions: Meta(data) progress is 'completely archived' (long term archiving process of data and metadata is finished), DataCite DOI(s) are assigned and the data format is netCDF CF or GRIB.

The FAIR Data Principles do not contain the persistence of the data. However, this is included in the QMM under the aspect Existence of Data (Completeness and Persistence).

6 Granularity of QMM Level Assignment

It is recommended to assign levels on aspect granularity and not to use sub-aspects such as *data formats – Content Data Object (OAS)* because netCDF CF is a discipline-specific standard (level 4) and netCDF is an interdisciplinary standard (level 5). To rule out the need for less stringent requirements follow to a higher level, the entire aspect must be fulfilled. Several quality results for different data quality aspects can be aggregated to the associated criterion, if all aspects have reached the corresponding level (**Table 3**).

The evaluation process was carried out as an exemplar at the DKRZ-LTA. The corresponding protocol¹¹ is available online.

7 Conclusion

The DKRZ Long Term Archive stores Earth System data with a strong focus on climate model data. Especially for the latter, the described Quality Maturity Matrix for data has been developed. However, it can easily be adapted to other data types like satellite or in-situ data.

The aim of this data assessment by QMM is to give the data user the opportunity to appraise the Fitness for Use of the data objects. Metrics that represent this for the data records are useful for this purpose. The QMM described here should additionally provide clues for the improvement of the working method. With the QMM the data user is given an idea of how far the disseminated data follow the FAIR Data Principles and other standards and recommendations. The QMM goes beyond the FAIR Data Principles in the field of data persistence, which is of particular interest for archives and their users.

Abbreviations

AIP: Archival Information Package (CCSDS 2012)

CDO: Content Data Object (CCSDS 2012)

⁹ <https://cfconventions.org/>.

¹⁰ <https://www.wmo.int/pages/prog/www/WDM/Guides/Guide-binary-2.html>.

¹¹ https://cera-www.dkrz.de/WDCC/ui/cersearch/entry?acronym=QMM_MOMERGOMBSEMEP.

CV: Controlled Vocabulary
DIP: Dissemination Information Package (CCSDS 2012)
DKRZ: Deutsches Klimarechenzentrum (German Climate Computing Center)
DOI: Data Object Identifier for Data see <https://datacite.org/>
DM: data management
DWD/EUMETSAT: Deutscher Wetterdienst and European Organisation for the Exploitation of Meteorological Satellites collaboration
EOSC: European Open Science Cloud
ESGF: Earth System Grid Federation
FAIR: Findable, Accessible, Interoperable, Reusable
WMO GRIB: World Meteorological Organization GRIdded Binary
HPSS: High Performance Storage System
ISC: International Science Council
LTA: long term archival storage
NetCDF CF: Network Common Data Form Climate and Forecast
OAIS (CCSDS): Open Archival Information System (The Consultative Committee for Space Data Systems)
PID: Persistent Identifier
PDI: Preservation Description Information (CCSDS 2012)
QMM: Quality Maturity Matrix
RDM: Research Data Management
RDMO: Research Data Management Organiser
SIP: Submission Information Package (CCSDS 2012)
SMM: System Maturity Matrix
WDCC: World Data Center for Climate
WDS: World Data System

Acknowledgements

The authors wish to thank Michael Lautenschlager and Martina Stockhause for their assistance in developing the Quality Maturity Matrix.

Competing Interests

The authors have no competing interests to declare.

Author Information

The Authors have been providing data management services to the climate research community for over 20 years. With expertise in earth sciences, informatics, mathematics, astronomy, legal issues and database administration, they bring a broad range of expertise to solve the challenges of research data management.

References

- ANDS.** 2011. Research data management framework: Capability maturity guide. Melbourne: Australian National Data Service. Available at <https://docplayer.net/15343597-Research-data-management-framework-capability-maturity-guide.html>.
- Bates, JJ** and **Privette, JL.** 2012. A maturity model for assessing the completeness of climate data records, *EOS. Transactions of the AGU*, 93(44): 441. DOI: <https://doi.org/10.1029/2012EO440006>
- CCSDS.** 2012. (OAIS), Recommended Practice, CCSDS 650.0-M-2 (Magenta Book) Issue 2. Available at <https://public.ccsds.org/pubs/650x0m2.pdf> [Last accessed 23 Mai 2018].
- Cox, AM,** et al. 2017. Developments in research data management in academic libraries: Towards an understanding of research data service maturity. *Journal of the Association for Information Science and Technology*, 68(9): 2182–2200. DOI: <https://doi.org/10.1002/asi.23781>
- Crowston, K** and **Qin, J.** 2012. A capability maturity model for scientific data management: Evidence from the literature. *Proceedings of the American Society for Information Science and Technology*, 48(1): 1–9. DOI: <https://doi.org/10.1002/meet.2011.14504801036>
- DKRZ-User Portal.** 2019. FAIRness of DKRZ's LTA WDCC service. Hamburg, Germany: DKRZ. Available at <https://www.dkrz.de/up/services/data-management/LTA/fairness> [Last accessed 18 Aug 2020].
- Höck, H,** et al. 2015. Maturity Matrices for Quality of Model- and Observation-Based Data Records in Climate Science. <https://meetingorganizer.copernicus.org/EGU2015/EGU2015-10158-1.pdf>.

- Höck, H.** 2019a. Technical Report Quality Maturity Matrix (QMM) Checklist. Hamburg, Germany: WDCC. DOI: https://doi.org/10.2312/WDCC/TR_QMM_Checklist
- Höck, H.** 2019b. QC Checklist QMM Level 4 and 5 with Protocols at DKRZ-LTA. Hamburg, Germany: WDCC. DOI: http://doi.org/10.2312/WDCC/TR_QMM_Checkl_Levels_4-5_Prots
- ISO 19157:2013-12.** Geographic information – Data quality (ISO 19157:2013(E)).
- Kenney, AR and McGovern, NY.** 2003. The five organizational stages of digital preservation. In Hodges, P, Sandler, M, Bonn, M and Wilkin, JP. (eds.), *Digital libraries: A vision for the 21st century*. Ann Arbor, MI: University of Michigan Scholarly Publishing Office. Available at <http://hdl.handle.net/2027/spo.bbv9812.0001.001>.
- Lyon, L,** et al. 2012. Developing a community capability model framework for data-intensive research. In *iPres 2012. Proceedings of the Ninth International Conference on the Preservation of Digital Objects* (pp. 9–16). Available at: <https://ipres.ischool.utoronto.ca/sites/ipres.ischool.utoronto.ca/files/iPres%202012%20Conference%20Proceedings%20Final.pdf> [Last accessed 15 Aug 2020].
- Mons, B,** et al. 2017. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. DOI: <https://doi.org/10.3233/ISU-170824>
- Paulk, MC,** et al. 1993. Capability maturity model, Version 1.1. *IEEE Software*, 10(4): 18–27. DOI: <https://doi.org/10.1109/52.219617>
- Peng, G.** 2018. The state of assessing data stewardship maturity – An overview. *Data Science Journal*, 17: Article 7. DOI: <https://doi.org/10.5334/dsj-2018-007>
- Peng, G,** et al. 2015. A unified framework for measuring stewardship practices applied to digital environmental datasets. *Data Science Journal*, 13: 231–253. DOI: <https://doi.org/10.2481/dsj.14-049>
- Peng, G,** et al. 2016. Scientific stewardship in the open data and big data era – Roles and responsibilities of stewards and other major product stakeholders. *D-Lib Magazine*, 22(5/6). DOI: <https://doi.org/10.1045/may2016-peng>
- Treloar, AE and Harboe-Ree, C.** 2008. Data management and the curation continuum: how the Monash experience is informing repository relationships. Available at: https://bridges.monash.edu/articles/Data_management_and_the_curation_continuum_how_the_Monash_experience_is_informing_repository_relationships/5627773 [Last accessed 18 Aug 2020].
- Wilkinson, M,** et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*, 3: 160018. DOI: <https://doi.org/10.1038/sdata.2016.18>


How to cite this article: Höck, H, Toussaint, F and Thiemann, H. 2020. *Fitness for Use of Data Objects Described with Quality Maturity Matrix at Different Phases of Data Production*. *Data Science Journal*, 19: 45, pp. 1–14. DOI: <https://doi.org/10.5334/dsj-2020-045>

Submitted: 20 February 2020

Accepted: 03 November 2020

Published: 17 November 2020

Copyright: © 2020 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 Ubiqity Press.

OPEN ACCESS 