Open
Geospatial
Consortium

ANALYSIS READY DATA (ARD) MATURITY REPORT

ENGINEERING REPORT

DRAFT

CONTENTS

LIST OF TABLES

I   OVERVIEW

The Analysis Ready Data (ARD) Maturity Report evaluates the maturity of crucial ARD sources for disaster risk response and climate assessments, focusing on NOAA data sources. It reviews three major data maturity assessment models—Data Stewardship Maturity Matrix (DSMM), CEOS (Committee on Earth Observation Satellites) Analysis Ready Data (CEOS-ARD), and WGISS (Working Group on Information Systems and Services) Data Management and Stewardship Maturity Matrix (DMSMM)—and develops an updated data maturity matrix combining their strengths. The report highlights the importance of data quality, accessibility, and interoperability, and evaluates datasets like NOAA's Climate Data Record of Passive Microwave Sea Ice Concentration and Landsat Collection 2 Surface Reflectance. It also identifies the need for tools to maximize automation in data maturity evaluation, emphasizing the development of a comprehensive suite for ARD maturity assessment.

II   EXECUTIVE SUMMARY

The Analysis Ready Data (ARD) Maturity Report provides a comprehensive evaluation of the maturity of key ARD sources, focusing on their role in supporting disaster risk response and climate assessments. Centered on NOAA datasets, the report assesses the readiness of data products to meet the growing demand for high-quality, accessible, and interoperable geospatial information. To establish a robust framework for ARD evaluation, the report reviews and synthesizes three leading data maturity models: (1) the Data Stewardship Maturity Matrix (DSMM), (2) the CEOS Analysis Ready Data (CEOS-ARD), and (3) the WGISS Data Management and Stewardship Maturity Matrix (DMSMM). By integrating the strengths of these models, the report introduces an enhanced ARD maturity matrix that supports the principles of FAIRness (Findable, Accessible, Interoperable, Reusable), AI readiness, and metadata standardization. Two key datasets evaluated using the enhanced maturity matrix are NOAA's Climate Data Record of Passive Microwave Sea Ice Concentration and Landsat Collection 2 Surface Reflectance, which represent typical datasets from two different communities—the climate and weather community and the Earth observation community. The report also reviews tools that support ARD maturity evaluation, highlighting the importance of automation to streamline assessments and reduce manual overhead. As a result, it advocates for the development of a comprehensive tool suite to enable self-service evaluations and improve consistency across data products.

III   KEYWORDS

The following are keywords to be used by search engines and document catalogues.

analysis ready data, data stewardhsip, data maturity, application-to-the-cloud, pilot, docker, web
service

IV CONTRIBUTORS

All questions regarding this document should be directed either to the editor or to the
contributors.

NAME	ORGANIZATION	ROLE
Eugene Yu	George Mason University.	Editor
Liping Di	George Mason University	Editor
Micah Brachman	OGC	Task Lead

V FUTURE OUTLOOK

The future outlook for Analysis Ready Data (ARD) maturity assessment emphasizes the need for
developing comprehensive tools to maximize automation in evaluating data maturity. Building on
the integration of existing data maturity matrices like DSMM, CEOS-ARD, and DMSMM, future
efforts should focus on enhancing productivity by creating a suite of tools that streamline the
assessment process across all components. This will support FAIRness, AI readiness, and ease of
integration, ultimately improving data accessibility, interoperability, and quality for disaster risk
response and climate assessments.

VI VALUE PROPOSITION

The Analysis Ready Data Maturity Report provides a comprehensive evaluation of crucial ARD
sources for disaster risk response and climate assessments, focusing on NOAA data sources. By
reviewing and integrating existing data maturity matrices like DSMM, CEOS-ARD, and DMSMM,
the report develops an updated, comprehensive ARD maturity assessment matrix. This matrix
supports FAIRness, AI readiness, and conformance to metadata standards, enhancing data
accessibility, interoperability, and integration. The report also identifies tools for self-service data

maturity assessments, recommending improvements and analyzing dataset component maturity to ensure high-quality, reliable, and accessible data for effective disaster and climate resilience.

1

INTRODUCTION

---

① INTRODUCTION

In an era of increasing environmental challenges, the maturity and accessibility of NOAA's Analysis Ready Data (ARD) are crucial for effective disaster risk response and climate assessments.

Analysis Ready Data (ARD) refers to datasets that have been pre-processed and standardized to a level where they can be immediately used for analysis without requiring additional preparation[1]. These datasets are typically geospatial in nature and are processed to ensure consistency, accuracy, and usability across various applications. ARD often includes corrections for geometric alignment, atmospheric conditions, and radiometric calibration, making them suitable for time-series analysis, machine learning, and other advanced data processing tasks. The goal of ARD is to minimize the time and effort required by users to prepare raw data, thereby enabling more efficient and effective analysis.

Building on this concept, the study of data maturity evaluation or levels of readiness involves assessing how well an organization or system is prepared to utilize data effectively. This evaluation typically includes frameworks or models that measure various aspects of data management, such as data quality, governance, integration, and analytics capabilities. By understanding the maturity levels, organizations can identify gaps, prioritize improvements, and develop strategies to enhance their data readiness. This progression often aligns with the principles of ARD, as higher maturity levels emphasize the availability of well-prepared, standardized datasets that can drive actionable insights.

1.1. Aims

The ARD Maturity Report for Climate and Disaster Resilience Pilot aims to evaluate the maturity of crucial NOAA Analysis Ready Data (ARD) sources integral to disaster risk response and climate assessments. This report is designed to facilitate the agency's uptake of NOAA data, both internally and externally, by providing an updated data maturity matrix and evaluating selected datasets against these criteria.

1.2. Objectives

In an era where climate change and natural disasters pose significant risks to communities and ecosystems, the availability and maturity of reliable data are paramount. This report focuses on two primary objectives:

- Evaluate the maturity of NOAA ARD sources
  - Primary Focus: Disaster risk response.

- Secondary Focus: Climate assessments.

- Develop Updated Criteria and Requirements

  - Build on existing data maturity matrices used by NOAA and allied government agencies.

  - Produce updated criteria and requirements for different ARD maturity levels, considering:

    - FAIRness and AI readiness.

    - Storage locations and access protocols (e.g., cloud-native storage, object storage, HTTP access, lazy access, intelligent subsetting, integration with high-level analysis libraries and distributed frameworks).

    - Conformance to indexed metadata standards.

    - Ease of integration into climate and disaster applications.

2

DATA MATURITY

---

2 DATA MATURITY

2.1. Updated Data Maturity Matrix

To ensure NOAA's Analysis Ready Data (ARD) sources are optimized for disaster resilience and climate assessments, an updated data maturity matrix that evaluates key aspects of data readiness and usability will be developed.

- **FAIRness and AI Readiness**: Ensuring that data is Findable, Accessible, Interoperable, and Reusable (FAIR) is essential for maximizing its utility. This involves:

  - Findable: Implementing robust metadata standards and indexing to ensure datasets can be easily discovered by users.

  - Accessible: Providing clear access protocols and ensuring data is available through multiple platforms and interfaces.

  - Interoperable: Ensuring data can be seamlessly integrated with other datasets and systems, supporting various formats and standards.

  - Reusable: Guaranteeing that data is well-documented and maintained, allowing for repeated use in different contexts.

  - Additionally, assessing AI readiness involves evaluating the data's suitability for advanced analytics and machine learning applications. This includes ensuring data quality, consistency, and the presence of necessary annotations and metadata.

- **Storage Locations and Access Protocols**: Evaluating storage solutions is critical for efficient data management and accessibility. This includes:

  - Cloud-Native and Cloud-Optimized Storage: Assessing the use of cloud-native storage solutions that offer scalability, reliability, and cost-effectiveness. HTTP Access: Ensuring data can be accessed via HTTP from multiple compute instances, facilitating ease of use and integration.

  - Lazy Access and Intelligent Subsetting: Supporting mechanisms that allow users to access only the data they need, reducing overhead and improving performance.

  - Integration with Analysis Libraries and Frameworks: Ensuring compatibility with high-level analysis libraries and distributed computing frameworks to support advanced data processing and analysis.

- **Conformance to Indexed Metadata Standards**: Conforming to standardized metadata indexing is crucial for enhancing data discoverability and usability. This involves:

- Standardized Metadata: Ensuring datasets adhere to recognized metadata standards, facilitating easier search and retrieval.

- Indexed Metadata: Implementing indexing mechanisms that improve the efficiency of data discovery and access.

- **Ease of Integration**: Assessing the ease with which datasets can be integrated into climate and disaster applications is vital for their practical use. This includes:

  - Integration with Applications: Evaluating how readily datasets can be incorporated into existing climate and disaster resilience applications.

  - User-Friendly Interfaces: Providing interfaces and tools that simplify the process of data integration for science users.

  - Documentation and Support: Ensuring comprehensive documentation and support resources are available to assist users in integrating and utilizing the data effectively.

The updated Data Maturity Matrix integrates the strengths of the CEOS Analysis Ready Data (ARD) specification, the NOAA Data Stewardship Maturity Matrix (DSMM), and the CEOS WGISS Data Management and Stewardship Maturity Matrix (DMSMM). It provides a comprehensive framework for assessing the maturity of data products and stewardship practices, with four levels of maturity: 0 (Not Managed), 1 (Partially Managed), 2 (Managed), and 3 (Fully Managed). Key components include general metadata and accessibility, per-pixel metadata and usability, radiometric, atmospheric, and geometric corrections, data quality assurance, product family specifications and data quality control/monitoring, product assessments and data quality assessment, preservability and production sustainability, transparency/traceability, and data integrity. The updated data maturity matrix is shown in the table below. Full details on the survey, the updated data maturity matrix, and the evaluation template can be found in Annex B.

Table 1 — The Updated Data Maturity Matrix

CATEGORY	LEVEL 0 (NOT MANAGED)	LEVEL 1 (PARTIALLY MANAGED)	LEVEL 2 (MANAGED)	LEVEL 3 (FULLY MANAGED)
GENERAL METADATA AND ACCESSIBILITY	- No sufficient metadata - Not registered and released in a recognized catalogue - Data and metadata are not accessible online	- No standard adopted for metadata - Registered and released in a specific catalogue - Catalogue search at product-specific level - Basic online service for data and metadata	- Detailed and comprehensive metadata - Product metadata considered international standard - International standard adopted for collection metadata - Registered and released in an online catalogue	- International or community agreed standard for detailed and comprehensive metadata - Both product and collection metadata adopted international standard - Registered and released in a well-known online catalogue for searching

CATEGORY	LEVEL 0 (NOT MANAGED)	LEVEL 1 (PARTIALLY MANAGED)	LEVEL 2 (MANAGED)	LEVEL 3 (FULLY MANAGED)
			for searching against data and metadata • Search interface partially adopted international standard	against data and metadata • Information for periodic updates of metadata available online • Search interface adopted international standard and accessible online • Advanced services available online, such as visualization, summary, and customizable retrieval • Dissemnination reports and data policy available online
PER-PIXEL METADATA AND USABILITY	• No per-pixel metadata • Data not structured • Non-standard or propriertary data format • Incomplete documentation • No document online	• Partial per-pixel metadata • Schema available for automated data use • Data in open or well-known formats • Data documentation avalable online • No link between mission documentation and data records	• Per-pixel metadata but not completely following international standard • International standards for data encoding • Basic interoperable capabilities (such as subsetting, aggregating, retrieval prtocol) • Data in internationally recognized standard formats • All documents are available online • Link between mission documentation and data records created and managed	• Per-pixel metadata in internationally recognized standards • Standard vocabularies are adopted • International standards for data semantic encoding • Advanced interoperable capabilities (such as semantic discovery and retrieval, interoperable visualization) • Standarad-based metadata for doucmentation and accessible online • Link between mission documentation and data records published online
RADIOMETRIC, GEOMETRIC, AND	• No radiometric and atmospheric corrections • No geometric correction	• Low level radiometric and atmospheric corrections	• Radiometric and atmospheric corrections • Basic geometric correction	• Radiometric and atmospheric corrections are complete. Calibrations adopted

CATEGORY	LEVEL 0 (NOT MANAGED)	LEVEL 1 (PARTIALLY MANAGED)	LEVEL 2 (MANAGED)	LEVEL 3 (FULLY MANAGED)
ATMOSPHERIC CORRECTIONS	• No algorithm/ processing documentation • No uncertainty characterization • No information about ancillary data	• No geometric correction, insufficient information for geometric calibration • Partial algorithm/ processing documentation • Limited uncertainty characterization, not in cross-comparison • Limited information about ancillary data	• Algorithm/ processing documentation is available online, major steps are documented. • Uncertainty characterization are available for processes and calibrations applied • Information about ancillary data is available, but ancillary data may not be available in open standards	international or community standard procedures. • Geometric correction is complete. • All algorithm/ processing documentation is available online and following international or community protocols. • Complete ncertainty characterization. Per-pixel uncertainty is available. • Ancillary data are available online in international standards. Meet the standard requirements of high level processed products.
DATA QUALITY ASSURANCE	• Data quality assurance procedure is unknown • No validation activity performed • No validation results	• Ad hoc data quality assurance procedure is applied. No ducumentation is available. • Simple comparison. Limited validation activity performed • Simple validation results without solid support	• Data quality assurance procedure is performed and documented. Fully available online with master reference data. Limited data quality assurance metadata. • Complete methodology is documented with characterized uncertainties. Reference measurements are well representative. • Independent validation results are available. Mission prefmance	• Cmplete data quality assurance procedure is performed and documented following community or international protocols. External review is applied. • Comprehensive validation activities are performed. Reference measurements independently assessed to be fully represenative. • All validation results are available online. Mission prefmance shows excellent

CATEGORY	LEVEL 0 (NOT MANAGED)	LEVEL 1 (PARTIALLY MANAGED)	LEVEL 2 (MANAGED)	LEVEL 3 (FULLY MANAGED)
			shows excellent agreement with validation results.	agreement with validation results with validated uncertainty characterization. Analysis performed independently.
PRODUCT FAMILY SPECIFICATIONS (PFS) AND DATA QUALITY CONTROL/ MONITORING	• No product family specificaion • Sampling unknown • Analysis unknown • No control and monitoring check • No quality indicator in metadata • No procedure documentation	• Product family specificaion in draft, not fully developed and approved • Sampling is frequent and available • Analysis is systematic but not automatic • Basic control and monitoring check • Community quality metrics are partially implemented • Procedure documentation is limited to implemented basic procedures	• Product family specificaion is developed and approved within the community • Sampling is extensive • Analysis is automatic following community metrics • Complete control and monitoring check • Quality indicator is available in metadata • Quality control rocedure is documented and available online	• Product family specificaion is fully developed, tested, and approved as a standard through an internationally recognized standard body • Sampling is cross-validated • Analysis is automatic, comprehensive, and independent • Full control and monitoring check compliant with an international standard • Quality indicator, both pre- and post-processing, are available in metadata • Full procedure documentation is available online. Quality metadata is accessible online.
PRODUCT ASSESSMENTS AND DATA QUALITY ASSESSMENT	• Algorithm / method / model theoretical basiss assessed (available online) • No calibration algorithm document • No Geometric processing algirhtm document • No Retrieval algorithm document	• Operational product assessed (available online) • Calibration algorithm is partially documented • Geometric processing algirhtm is documented, but missing all or part of the calibration parameters.	• Quality metadata assessed (available online), but limited assessments. • Calibration algorithm is documented to cover all expected use cases of the mission. • Geometric processing	• Assessment performed on a recurring basis conforming to community quality metadata and standards (available online) • Calibration algorithm is well-documented with state-of-the-art calibration algorithm

| | • No additional processing step document | Confidence in the calibration qulaity is minimal.
• Retrieval algorithm is documented, but the retrieval algorithm is with limited performance.
• Additional processing steps are documented, but they are limited. | algirhtm is documented with all input calibration parameters and meet the performance of the mission for all expected use cases. Quality flags indicate good gemetric accuracy with less than 5% excpeptional.
• Retrieval algorithm is documented and meet the mission's stated prefmance in all epxected use cases and validated perfmance against similar algoriithms or with empirical evidence
• Additional processing steps are documented and fit for the stated purpose of the mission. | applied and considered to support the mission's stated performance
• Geometric processing is well-documented with state-of-the-art methodology used to support the mission's stated performance. Quality flags indicate excellent geometric accuracy.
• Retrieval algorithm document is well-documente with state-of-the-art retrieval to support the mission's stated performance with full uncertainty derived and validated.
• Additional processing steps are well-documented for the mission's stated purpose. |
| PRESERVABILITY AND PRODUCTION SUSTAINABILITY | • Uncontrolled storage location
• Only data are stored
• No sustainability information is available
• No persistent and resolvable identifiers available | • Basic archiving repository
• Metadata preserved
• Midium term, institutional sustainaining plan (contractual deliverables with specs and schedule defined)
• Persistent identifier asssignment only available for particular data records collection | • Preservation repository certified internally
• Community-standard-conformed metadata preservation
• Long term, institutional commitment to sustainain data preservation
• Persistent identifier asssignment available for all desiminated data records | • Preservation repository certified officially
• International standard for metadata preservation
• Future archiving standard chnages planned
• Long term, national or international commitment to sustain the data preservation
• Persistent identifier for |

CATEGORY	LEVEL 0 (NOT MANAGED)	LEVEL 1 (PARTIALLY MANAGED)	LEVEL 2 (MANAGED)	LEVEL 3 (FULLY MANAGED)
		• Basic landing page management	collections and metadata • Automatic landing page generation, updating, and maintenance	accessible data and metadata • Data include the identifier in its metadata • Metadata follows international standards that can be harvested and indexed automatically
TRANSPARENCY, TRACEABILITY	• Limited product information available online • Detailed product information is scattered and difficult to obtain	• Theoretical basis documents on algorithms / method / model and code are available online • Overall data product citation trackable with unique identifier • Detailed product information is not accessible online • Detailed product information is available unpon inquiry	• Operational algorithms and documents are available online • Data are trackable through unique identifiers • Detailed product information is accessible online • Data tested for presence of correct provenance metadata	• All data product information is available online • Automatic metadata generation and updating for provenance documentation • Complete and updated data provenance available online
DATA INTEGRITY	• No data ingest integrity check • No data/ associated information integrity, authenticity and readibility check	• Data archive integrity verifiable • Data records/ associated information integrity basic check	• Data access integrity verifiable • Conforming to community data integrity technology standard • Data records/ associated information content integrity check and verification • Media readability and accessbility testing	• Data authenticity verifiable • Performance of data integrity check monitored and reported • Automatic data records/ associated information content integrity check and verification • Data authenticity verifiable • Automatic verification process, including monitoring and reporting

2.2. Evaluation of Selected NOAA Data Sources

To ensure the effectiveness and usability of NOAA's Analysis Ready Data (ARD) for climate and disaster resilience applications, it is essential to evaluate key aspects of data accessibility, interoperability, and assimilation barriers. This section provides a detailed assessment of these critical factors on selected datasets.

- **Data Accessibility**: Evaluating data accessibility involves assessing how easily NOAA's ARD can be accessed and utilized by users. Key considerations include:

  - Ease of Access: Determining whether data can be readily accessed through user-friendly interfaces and protocols. This includes evaluating the availability of data through web portals, APIs, and other access points.

  - Documentation and Support: Ensuring comprehensive documentation is available to guide users in accessing and using the data. This includes user manuals, tutorials, and support services. Access Speed and Reliability: Assessing the speed and reliability of data access, including the performance of data retrieval processes and the stability of access platforms.

- **Interoperability**: Interoperability is crucial for ensuring that NOAA's ARD can work seamlessly with other datasets and systems. This involves:

  - Standardization: Ensuring data conforms to common standards and formats that facilitate integration with other datasets. This includes adherence to metadata standards and data formats widely used in the climate and disaster resilience communities.

  - Compatibility: Evaluating the ability of data to be used in conjunction with various software tools and platforms. This includes compatibility with data analysis software, geographic information systems (GIS), and other relevant applications.

  - Data Linking and Integration: Assessing the ease with which data can be linked and integrated with other datasets. This includes evaluating the presence of unique identifiers and metadata that support data merging and cross-referencing.

- **Assimilation Barriers**: Identifying and addressing assimilation barriers is essential for integrating NOAA's ARD into specific applications. Key challenges include:

  - Technical Barriers: Evaluating technical challenges such as data format conversions, software compatibility issues, and the need for specialized tools or expertise to use the data effectively.

  - Data Quality and Consistency: Assessing the quality and consistency of data, including the presence of gaps, errors, or inconsistencies that may hinder its use in applications.

- User Training and Expertise: Identifying the need for user training and expertise to effectively assimilate and utilize the data. This includes evaluating the availability of training resources and the level of expertise required to work with the data.

- Resource Availability: Considering the availability of computational and storage resources necessary to handle large datasets and perform complex analyses.

In this pilot, several collections of datasets have been reviewed for disaster resilience and climate assessments, evaluating their data maturity and analysis readiness using the updated data maturity matrix developed by CEOS and CGMS. The evaluation focused on Essential Climate Variables (ECVs) and included datasets like the NOAA Climate Data Record of Passive Microwave Sea Ice Concentration and Landsat Collection 2 Surface Reflectance, assessing their readiness and maturity levels. The analysis highlighted the importance of well-documented, accessible, and high-quality data for effective disaster and climate assessments, with ongoing development in ARD datasets to enhance specifications, interoperability, and readiness for AI/ML/cloud applications. Full details on dataset collections and examples of evaluating data maturity with the updated matrix can be found in Annex C .

2.3. Tools for Self-Service Assessments

To empower domain scientists and data users in evaluating and improving the maturity of NOAA's Analysis Ready Data (ARD), it is essential to review existing tools and identify any gaps. These tools should support the following key functionalities:

- **Identification of Datasets**: Effective tools for identifying datasets that meet specified ARD maturity levels are crucial. Examples include:

  - CEOS ARD: The Committee on Earth Observation Satellites (CEOS) ARD tools help users identify datasets that conform to ARD standards, ensuring they meet the necessary criteria for disaster resilience and climate assessments.

  - USGS Landsat ARD: The United States Geological Survey (USGS) provides ARD for Landsat data, which includes pre-processed, analysis-ready datasets that are easy to use for various applications.

These tools facilitate the discovery of datasets that are already aligned with ARD maturity standards, streamlining the process for users.

- **Recommendations for Improvement**: Tools that provide actionable recommendations for improving the ARD maturity level of datasets are invaluable. These tools should:

  - Analyze Current Maturity Levels: Assess the current state of datasets against the updated data maturity matrix.

- Identify Areas for Enhancement: Highlight specific elements that need improvement, such as metadata completeness, data accessibility, or interoperability.

- Provide Clear Guidance: Offer step-by-step recommendations and best practices for enhancing the maturity of datasets, making them more suitable for advanced analytics and integration into climate and disaster resilience applications.

- **Categorization of Non-Pass Datasets**: It is also important to have tools that categorize datasets needing adjustments to meet specified levels in the data maturity model. These tools should:

  - Assess Conformance: Evaluate datasets against the criteria defined in the ARD maturity matrix.

  - Categorize Adjustments Needed: Classify datasets based on the extent of adjustments required, ranging from minor updates to substantial modifications.

  - Prioritize Actions: Help users prioritize the necessary actions to bring datasets up to the desired maturity level, ensuring efficient use of resources and time.

In this pilot, existing tools and libraries were reviewed for evaluating data maturity and analysis readiness for disaster resilience and climate assessments, particularly in the context of artificial intelligence and machine learning. Key tools include data maturity assessment templates like the Scientific Data Stewardship Maturity Assessment (DSMM) Model Template, WGISS Data Management and Stewardship Maturity Matrix (DMSMM) Schema, and CEOS Analysis Ready Data (ARD) Self-Assessment User Guide. Compliance test tools such as the OGC Compliance Test Suites, CF-Checker, Geospatial Metadata Validation Service (ISO 19115), and ESDSWG DIWG Compliance Test ensure datasets adhere to specific standards and conventions, supporting high-quality, reliable, and accessible data management practices. Full details on survey and review of data maturity assessment templates and compliance test tools for effective data stewardship and maturity evaluation can be found in Annex D .

3

OUTLOOK

—

3   OUTLOOK

The Analysis Ready Data (ARD) Maturity Report highlights significant advancements in data accessibility and quality, particularly for NOAA data sources. Moving forward, the primary focus will be on evaluating the maturity of crucial ARD sources integral to disaster risk response and climate assessments. This involves developing updated criteria and requirements for ARD maturity evaluation by reviewing existing data maturity matrices and creating a comprehensive ARD maturity assessment matrix. The updated matrix aims to support FAIRness and AI readiness, facilitate storage location and access protocols, enable conformance to metadata standards, and enhance ease of integration.

Future research should address the gaps identified in the current data maturity assessment processes. One major gap is the absence of a single suite that maximizes automation for data maturity evaluation. Developing a comprehensive suite of tools that streamline the assessment process across all components will be crucial. This suite should integrate templates for manual processes and programmatic checkers for specific standards, enhancing productivity and ensuring high-quality, reliable, and accessible data.

Additionally, the ongoing development of ARD datasets should focus on specifications, interoperability, and readiness for AI/ML/cloud applications. By leveraging the strengths of existing data maturity matrices like DSMM, CEOS-ARD, and DMSMM, and integrating them into a unified framework, future efforts can significantly improve data management practices. This will empower researchers and practitioners to tackle complex geospatial challenges with increased efficiency and effectiveness, ultimately advancing the field of disaster risk response and climate assessments.

4

SECURITY, PRIVACY AND ETHICAL CONSIDERATIONS

4 SECURITY, PRIVACY AND ETHICAL CONSIDERATIONS

The Analysis Ready Data (ARD) Maturity Report emphasizes the importance of security, privacy, and ethical considerations in the evaluation and management of ARD sources. As data accessibility and quality improve, it is crucial to ensure that data handling practices adhere to stringent security protocols to protect sensitive information from unauthorized access and breaches. Privacy concerns must be addressed by implementing robust data anonymization techniques and ensuring compliance with relevant data protection regulations, such as GDPR (General Data Protection Regulation).

Ethical considerations are equally important, particularly in the context of disaster risk response and climate assessments. The use of ARD must be guided by principles of transparency, accountability, and fairness to avoid biases and ensure equitable access to data. This includes providing clear documentation of data provenance, processing steps, and any potential limitations or uncertainties associated with the data.

Furthermore, the development and deployment of tools for data maturity assessment should prioritize ethical AI practices, ensuring that algorithms used for data evaluation are transparent, explainable, and free from biases. Continuous monitoring and auditing of these tools are necessary to maintain their integrity and trustworthiness.

The ARD Maturity Report advocates for a comprehensive approach to data management that integrates security, privacy, and ethical considerations, thereby fostering trust and reliability in the use of ARD for disaster resilience and climate assessments.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1]   Liping Di, David J. Meyere, and Eugene Yu (eds.), 2023, OGC Testbed-19 Engineering Report. "Analysis Ready Data". https://docs.ogc.org/per/23-043.html.

[2]   G. Peng, "The State of Assessing Data Stewardship Maturity – An Overview," Data Sci. J., vol. 17, p. 7, Mar. 2018, doi: 10.5334/dsj-2018-007.

[3]   R. Dunn et al., "Stewardship Maturity Assessment Tools for Modernization of Climate Data Management," Data Sci. J., vol. 20, p. 7, Feb. 2021, doi: 10.5334/dsj-2021-007.

[4]   G. Peng, W. S. Gross, and R. Edmunds, "Crosswalks among stewardship maturity assessment approaches promoting trustworthy FAIR data and repositories," Sci. Data, vol. 9, no. 1, p. 576, Sep. 2022, doi: 10.1038/s41597-022-01683-x.

[5]   "CEOS Analysis Ready Data." Accessed: Dec. 12, 2024. [Online]. Available: https://ceos.org/ard/

[6]   CEOS-ARD Product Self-Assessment User Guide, 2023. URL: https://ceos.org/ard/files/User%20Guide/CEOS_ARD%20User%20Guide%20v1_3_Final_06152023.pdf

[7]   G. Peng et al., "Practical Application of a Data Stewardship Maturity Matrix for the NOAA OneStop Project," Data Sci. J., vol. 18, no. 1, p. 41, Aug. 2019, doi: 10.5334/dsj-2019-041.

[8]   P. Harrison and M. Thankappan, "CEOS-ARD Product Self-Assessment User Guide," 2023. [Online]. Available: https://ceos.org/ard/files/User%20Guide/CEOS_ARD%20User%20Guide%20v1_3_Final_06152023.pdf

[9]   CEOS, "Repository for CEOS Analysis Ready Data (CEOS-ARD), including Product Family Specifications (PFSs)." [Online]. Available: https://github.com/ceos-org/ceos-ard

[10]   CDOS WGISS Data Stewardship Interest Group, "WGISS Data Management and Stewardship Maturity Matrix", CEOS.WGISS. DSIG.DSMM Issue 1.1 Oct 2024. [Online]. Available: https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/WGISS%20Data%20Management%20and%20Stewardship%20Maturity%20Matrix.pdf

[11]   Research Data Alliance FAIR Data Maturity Model Working Group, "FAIR Data Maturity Model: specification and guidelines — draft," 2020, doi: 10.15497/RDA00045.

[12]   EDAP, "DAP Best Practice Guidelines", [Online]. Available: https://earth.esa.int/eogateway/activities/edap/edap-best-practice-guidelines

[13]   Peng, Ge (2014). NCDC-CICSNC SDSMM Template. figshare. Dataset. https://doi.org/10.6084/m9.figshare.1211954.v23

[14]   2024. WGISS Data Management and Stewardship Maturity Matrix_Schema to be
        filled. https://ceos.org/document_management/Working_Groups/WGISS/Interest_
        Groups/Data_Stewardship/White_Papers/WGISS%20Data%20Management%20and
        %20Stewardship%20Maturity%20Matrix_Schema%20to%20be%20filled.xlsx

[15]   2024, CSW 2.0.2 Conformance Test Suite. https://cite.opengeospatial.org/te2/about/
        csw/2.0.2/site/

[16]   2021, CF Checker. https://github.com/cedadev/cf-checker

[17]   Geospatial Metadata Validation Service. https://www1.usgs.gov/mp/

[18]   diwg-data-compliance-test. https://github.com/eugenegesdisc/diwg-data-compliance-
        test

A

ANNEX A (NORMATIVE) ABBREVIATIONS/ACRONYMS

A ANNEX A (NORMATIVE) ABBREVIATIONS/ACRONYMS

ARD	Analysis Ready Data
CEOS	Committee on Earth Observation Satellites
CICS-NC	Cooperative Institute for Climate and Satellites-North Carolina
DMP-IG	Data Management Plan Interest Group
DMSMM	Data Management and Stewardship Maturity Matrix
DSMM	Data Stewardship Maturity Matrix
EDAP	Earthnet Data Assessment Pilot
EO	Earth Observation
FAIR	Findable, Accessible, Interoperable, and Reusable
NASA	National Aeronautics and Space Administration
NCEI	National Centers for Environmental Information
NOAA	National Oceanic and Atmospheric Administration
PFS	Product Family Specifications
RDA	Research Data Alliance
SMM-CD	Stewardship Maturity Matrix for Climate Data
WGCV	Working Group on Calibration & Validation
WGISS	Working Group on Information Systems and Services
WMO	World Meteorological Organization

B

ANNEX B (NORMATIVE) DATA MATURITY MATRIX

———

ANNEX B
(NORMATIVE)
DATA MATURITY MATRIX

NOTE: This annex reviews existing data maturity matrices used in assuring the FAIRness of data, especially climate datasets and build up an updated data maturity matrix.

B.1. Data Stewardship Maturity Matrix (DSMM)

The Data Stewardship Maturity Matrix (DSMM) is a comprehensive framework designed to assess and enhance the maturity of data stewardship practices for digital datasets[2]. Developed by the National Centers for Environmental Information (NCEI) and the Cooperative Institute for Climate and Satellites-North Carolina (CICS-NC), the DSMM ensures that data is accessible, usable, current, preserved, documented, and of assured scientific quality[3].

B.1.1. Key Components

The DSMM evaluates nine critical components (KCs) of data stewardship[4] [7]:

1. Preservability: Measures the ability to preserve data over the long term.

2. Accessibility: Assesses how easily data can be accessed by users.

3. Usability: Evaluates the ease with which data can be used and understood.

4. Production Sustainability: Looks at the sustainability of data production processes.

5. Data Quality Assurance: Ensures that data meets quality standards before being released.

6. Data Quality Control/Monitoring: Involves ongoing monitoring and control of data quality.

7. Data Quality Assessment: Provides a comprehensive evaluation of data quality.

8. Transparency/Traceability: Ensures that data provenance and processing steps are well-documented.

9. Data Integrity: Protects data from unauthorized changes and ensures its accuracy.

Each component is assessed on a five-level maturity scale, providing a comprehensive view of the data stewardship practices in place. The maturity levels range from "Not Assessed/ Not Available" (Level 0) to "Optimal" (Level 5), representing a progression from basic to highly developed and optimized practices. This scale allows data stewards to identify areas for improvement and implement best practices to enhance their data management processes.

B.1.2. Benefits and Applications

A standout feature of the DSMM is its dual utility. It not only helps data stewards improve their practices but also assists data users in evaluating the suitability of datasets for their needs. This dual functionality ensures that datasets meet high standards of quality and reliability. The DSMM has been applied to over 800 datasets at NCEI, demonstrating its robustness and versatility. Its adoption by several Earth Science organizations further underscores its value. Notably, the World Meteorological Organization (WMO) uses the Stewardship Maturity Matrix for Climate Data (SMM-CD), based on the DSMM, as a cornerstone of its global climate data management framework.

The DSMM is a comprehensive and effective tool for enhancing data stewardship practices, particularly within NOAA data management. Its structured approach to assessing and improving data quality, integrity, and usability makes it an essential resource for organizations managing large datasets. The widespread adoption and adaptation of the DSMM by leading scientific organizations highlight its importance in the field of data stewardship.

B.2. CEOS Analysis Ready Data

The CEOS Analysis Ready Data (ARD) specification is designed to streamline the use of satellite data by ensuring it is pre-processed and ready for immediate analysis. Here are the major components[8] [5] [9]:

1. General Metadata: Provides overall information about the dataset, allowing users to assess its suitability for their needs

2. Per-Pixel Metadata: Includes detailed information for each pixel, enabling users to make informed decisions about which data points to use or discard

3. Radiometric and Atmospheric Corrections: Ensures that data is corrected for sensor and atmospheric effects, providing accurate measurements

4. Geometric Corrections: Aligns data to a common grid, ensuring spatial accuracy and consistency

5. Product Family Specifications (PFS): Defines specific requirements for different types of data products, including threshold and target levels for quality and accuracy

6.   Product Assessments: Evaluates how well a product meets the specified criteria, including peer reviews by the CEOS Working Group on Calibration & Validation (WGCV)

B.2.1. Key Differences from DSMM

The following summarizes the major differences between CEOS ARD and DSMM.

- Focus: CEOS ARD is primarily concerned with making satellite data ready for analysis by standardizing preprocessing steps, while DSMM focuses on the overall maturity of data stewardship practices, including preservation, accessibility, and quality control.

- Components: CEOS ARD emphasizes specific technical corrections and metadata requirements for satellite data, whereas DSMM covers a broader range of stewardship aspects, including sustainability and transparency.

- Assessment: CEOS ARD includes detailed product assessments and peer reviews for compliance with specifications, while DSMM provides a framework for rating the maturity of data stewardship practices across multiple dimensions.

B.3. WGISS DMSMM

The WGISS Data Management and Stewardship Maturity Matrix (DMSMM) is a comprehensive framework designed to assess and enhance the management and stewardship of Earth Observation (EO) datasets. Developed by the CEOS Working Group on Information Systems and Services (WGISS), this matrix aims to ensure the long-term preservation, curation, accessibility, discoverability, and usability of EO data[10].

The DMSMM represents the culmination of a combined analysis of the "Long-Term Scientific Data Stewardship" Maturity Matrix and the Data Management Plan Interest Group (DMP-IG). This analysis was performed in consultation with a Data Access and Preservation Working Group at the European level and the WGISS Data Stewardship Interest Group. Additionally, the matrix incorporates the Research Data Alliance (RDA) FAIR principles, which emphasize making data Findable, Accessible, Interoperable, and Reusable[11]. It also includes quality input from the Earthnet Data Assessment Pilot (EDAP), produced as part of WGISS cooperation with the CEOS Working Group on Data Calibration and Validation (WGCV)[12].

B.3.1. Key Features

B.3.1.1. Five Pillars of Data Stewardship

- Discoverability: Ensuring data can be easily found.

- Accessibility: Making data available and retrievable.

- Usability: Enhancing the ease of use and understanding of data.

- Preservation: Maintaining data integrity over time.

- Curation: Managing data to ensure its quality and relevance.

B.3.1.2. Four Levels of Maturity

- L0 (Not Managed): No formal management practices.

- L1 (Partially Managed): Some management practices in place.

- L2 (Managed): Comprehensive management practices implemented.

- L3 (Fully Managed): Best practices fully integrated and optimized.

B.3.1.3. Twelve Components

- Metadata for Discovery

- Online Access

- Data Encoding

- Data Documentation

- Data Traceability

- Data Validation

- Data Metrology (e.g., Uncertainty)

- Data Quality Control

- Data Preservation

- Data and Metadata Verification

- Data Processing/Reprocessing

- Persistent and Resolvable Identifies

B.3.2. Benefits

- Quality Assurance: Provides a structured approach to evaluate and improve data quality.

- Resource Allocation: Helps in planning and allocating resources for data stewardship.

- Stakeholder Confidence: Enhances trust among data users, stakeholders, and decision-makers by ensuring high standards of data management.

Overall, the WGISS DMSMM is a valuable tool for data managers and stewards, offering a clear roadmap to achieve and maintain high standards in data stewardship.

B.4. Updated Data Maturity Matrix

This updated Data Maturity Matrix integrates the strengths of both the CEOS Analysis Ready Data (ARD) specification, the NOAA Data Stewardship Maturity Matrix (DSMM), and the CEOS WGISS Data Management and Stewardship Maturity Matrix (DMSMM). It provides a comprehensive framework for assessing the maturity of data products and stewardship practices.

The maturity levels are adopted from the WGISS Data Management and Stewardship Maturity Matrix (DMSMM), with four levels: 0 (Not Managed), 1 (Partially Managed), 2 (Managed), and 3 (Fully Managed). The revised definitions are:

- L0 (Not Managed):

    - Definition: No formal management practices.

    - Considerations for Analysis-Ready Data: Data is not organized or documented, making it difficult to use for analysis. There is no standardization or quality control.

- L1 (Partially Managed):

    - Definition: Some management practices in place.

    - Considerations for Analysis-Ready Data: Basic management practices are implemented, such as minimal documentation and some level of data organization. However, data may still lack consistency and thorough quality checks.

- L2 (Managed):

    - Definition: Comprehensive management practices implemented. Considerations for Analysis-Ready Data: Data is well-documented, organized, and standardized. Quality control measures are in place, ensuring data is reliable and ready for analysis. Metadata is comprehensive and accessible.

- L3 (Fully Managed):

    - Definition: Best practices fully integrated and optimized.

- Considerations for Analysis-Ready Data: Data management practices are fully optimized and integrated into the organization's workflows. Data is consistently high-quality, well-documented, and easily accessible for analysis. Continuous improvement processes are in place to maintain and enhance data quality.

The following shows how the levels of DSMM map to the levels of updated data maturity matrix.

Table B.1 — Mapping of Maturity Levels between the Updated Data Maturity Matrix and DSMM

MATURITY LEVEL	DSMM	UPDATED DATA MATURITY MATRIX
Level 0 (not managed)	Level 1	L0
Level 1 (partially managed)	Level 2 and 3	L1
Level 2 (managed)	Level 3	L2
Level 3 (fully managed)	Level 4	L3

The key components for the updated data maturity matrix are:

1. General Metadata and Accessibility: Ensures that datasets have comprehensive metadata and are easily accessible to users. Components are: General Metadata (CEOS-ARD and DMSMM), Accessibility (DSMM and DMSMM).

2. Per-Pixel Metadata and Usability: Provides detailed information for each pixel and ensures data is user-friendly. Components are: Per-Pixel Metadata (CEOS-ARD), Usability (DSMM), Data Encoding (DMSMM), and Data Documentation (DMSMM).

3. Radiometric, Atmospheric, and Geometric Corrections: Ensures data is corrected for sensor and atmospheric effects, providing accurate measurements. Components are: Radiometric Corrections (CEOS-ARD), Atmospheric Corrections (CEOS-ARD), and Metrology (DMSMM).

4. Data Quality Assurance: Aligns data to a common grid and ensures it meets quality standards. Components are: Geometric Corrections (CEOS-ARD), Data Quality Assurance (DSMM), and Data Validation (DMSMM).

5. Product Family Specifications (PFS) and Data Quality Control/Monitoring: Defines specific requirements for different data products and involves ongoing quality monitoring. Components: Product Family Specifications (CEOS-ARD), Data Quality Control/Monitoring (DSMM and DMSMM).

6. Product Assessments and Data Quality Assessment: Evaluates how well a product meets specified criteria and provides a comprehensive quality evaluation.

Components are: Product Assessments (CEOS-ARD), Data Quality Assessment (DSMM), and Data Processing/Reprocessing (DMSMM).

7.   Preservability and Production Sustainability: Measures the ability to preserve data over the long term and the sustainability of production processes. Components are: Preservability (DSMM and DMSMM) and Production Sustainability (DSMM), and Persistent/Resolvable Identifier (DMSMM).

8.   Transparency/Traceability: Ensures that data provenance and processing steps are well-documented. Components are: Transparency/Traceability (DSMM and DMSMM).

9.   Data Integrity: Protects data from unauthorized changes and ensures its accuracy. Components are: Data Integrity (DSMM) and Data Verification (DMSMM).

Table B.2 — Updated Data Maturity Matrix

CATEGORY	LEVEL 0 (NOT MANAGED)	LEVEL 1 (PARTIALLY MANAGED)	LEVEL 2 (MANAGED)	LEVEL 3 (FULLY MANAGED)
GENERAL METADATA AND ACCESSIBILITY	• No sufficient metadata • Not registered and released in a recognized catalogue • Data and metadata are not accessible online	• No standard adopted for metadata • Registered and released in a specific catalogue • Catalogue search at product-specific level • Basic online service for data and metadata	• Detailed and comprehensive metadata • Product metadata considered international standard • International standard adopted for collection metadata • Registered and released in an online catalogue for searching against data and metadata • Search interface partially adopted international standard	• International or community agreed standard for detailed and comprehensive metadata • Both product and collection metadata adopted international standard • Registered and released in a well-known online catalogue for searching against data and metadata • Information for periodic updates of metadata available online • Search interface adopted international standard and accessible online • Advanced services available online, such as visualization, summary, and customizable retrieval

CATEGORY	LEVEL 0 (NOT MANAGED)	LEVEL 1 (PARTIALLY MANAGED)	LEVEL 2 (MANAGED)	LEVEL 3 (FULLY MANAGED)
				• Dissemnination reports and data policy available online
PER-PIXEL METADATA AND USABILITY	• No per-pixel metadata • Data not structured • Non-standard or propriertary data format • Incomplete documentation • No document online	• Partial per-pixel metadata • Schema available for automated data use • Data in open or well-known formats • Data documentation avalable online • No link between mission documentation and data records	• Per-pixel metadata but not completely following international standard • International standards for data encoding • Basic interoperable capabilities (such as subsetting, aggregating, retrieval prtocol) • Data in internationally recognized standard formats • All documents are available online • Link between mission documentation and data records created and managed	• Per-pixel metadata in internationally recognized standards • Standard vocabularies are adopted • International standards for data semanatic encoding • Advanced interoperable capabilities (such as semantic discovery and retrieval, interoperable visualization) • Standarad-based metadata for doucmentation and accessible online • Link between mission documentation and data records published online
RADIOMETRIC, GEOMETRIC, AND ATMOSPHERIC CORRECTIONS	• No radiometric and atmospheric corrections • No geometric correction • No algorithm/processing documentation • No uncertainty characterization • No information about ancillary data	• Low level radiometric and atmospheric corrections • No geometric correction, insufficient information for geometric calibration • Partial algorithm/processing documentation • Limited uncertainty characterization, not in cross-comparison • Limited information about ancillary data	• Radiometric and atmospheric corrections • Basic geometric correction • Algorithm/processing documentation is available online, major steps are documented. • Uncertainty characterization are available for processes and calibrations applied • Information about ancillary data is available, but ancillary data may not be available in open standards	• Radiometric and atmospheric corrections are complete. Calibrations adopted international or community standard procedures. • Geometric correction is complete. • All algorithm/processing documentation is available online and following international or community protocols. • Complete ncertainty characterization.

CATEGORY	LEVEL 0 (NOT MANAGED)	LEVEL 1 (PARTIALLY MANAGED)	LEVEL 2 (MANAGED)	LEVEL 3 (FULLY MANAGED)
				• Per-pixel uncertainty is available. • Ancillary data are available online in international standards. Meet the standard requirements of high level processed products.
DATA QUALITY ASSURANCE	• Data quality assurance procedure is unknown • No validation activity performed • No validation results	• Ad hoc data quality assurance procedure is applied. No ducumentation is available. • Simple comparison. Limited validation activity performed • Simple validation results without solid support	• Data quality assurance procedure is performed and documented. Fully available online with master reference data. Limited data quality assurance metadata. • Complete methodology is documented with characterized uncertainties. Reference measurements are well representative. • Independent validation results are available. Mission prefmance shows excellent agreement with validation results.	• Cmplete data quality assurance procedure is performed and documented following community or international protocols. External review is applied. • Comprehensive validation activities are performed. Reference measurements independently assessed to be fully represenative. • All validation results are available online. Mission prefmance shows excellent agreement with validation results with validated uncertainty characterization. Analysis performed independently.
PRODUCT FAMILY SPECIFICATIONS (PFS) AND DATA QUALITY CONTROL/ MONITORING	• No product family specificaion • Sampling unknown • Analysis unknown • No control and monitoring check	• Product family specificaion in draft, not fully developed and approved • Sampling is frequent and available	• Product family specificaion is developed and approved within the community • Sampling is extensive • Analysis is automatic following	• Product family specificaion is fully developed, tested, and approved as a standard through an internationally recognized standard body

CATEGORY	LEVEL 0 (NOT MANAGED)	LEVEL 1 (PARTIALLY MANAGED)	LEVEL 2 (MANAGED)	LEVEL 3 (FULLY MANAGED)
	• No quality indicator in metadata • No procedure documentation	• Analysis is systematic but not automatic • Basic control and monitoring check • Community quality metrics are partially implemented • Procedure documentation is limited to implemented basic procedures	community metrics • Complete control and monitoring check • Quality indicator is available in metadata • Quality control rocedure is documented and available online	• Sampling is cross-validated • Analysis is automatic, comprehensive, and independent • Full control and monitoring check compliant with an international standard • Quality indicator, both pre- and post-processing, are available in metadata • Full procedure documentation is available online. Quality metadata is accessible online.
PRODUCT ASSESSMENTS AND DATA QUALITY ASSESSMENT	• Algorithm / method / model theoretical basiss assessed (available online) • No calibration algorithm document • No Geometric processing algirhtm document • No Retrieval algorithm document • No additional processing step document	• Operational product assessed (available online) • Calibration algorithm is partially documented • Geometric processing algirhtm is documented, but missing all or part of the calibration parameters. Confidence in the calibration qulaity is minimal. • Retrieval algorithm is documented, but the retrieval algorithm is with limited performance. • Additional processing steps are documented, but they are limited.	• Quality metadata assessed (available online), but limited assessments. • Calibration algorithm is documented to cover all expected use cases of the mission. • Geometric processing algirhtm is documented with all input calibration parameters and meet the performance of the mission for all expected use cases. Quality flags indicate good gemetric accuracy with less than 5% excceptional. • Retrieval algorithm is documented and meet the	• Assessment performed on a recurring basis conforming to community quality metadata and standards (available online) • Calibration algorithm is well-documented with state-of-the-art calibration algorithm applied and considered to support the mission's stated performance • Geometric processing is well-documented with state-of-the-art methodology used to support the mission's stated performance. Quality flags indicate excellent

CATEGORY	LEVEL 0 (NOT MANAGED)	LEVEL 1 (PARTIALLY MANAGED)	LEVEL 2 (MANAGED)	LEVEL 3 (FULLY MANAGED)
				geometric accuracy.
			mission's stated prefmance in all epxected use cases and validated perfmance against similar algoriithms or with empirical evidence • Additional processing steps are documented and fit for the stated purpose of the mission.	• Retrieval algorithm document is well-documente with state-of-the-art retrieval to support the mission's stated performance with full uncertainty derived and validated. • Additional processing steps are well-documented for the mission's stated purpose.
PRESERVABILITY AND PRODUCTION SUSTAINABILITY	• Uncontrolled storage location • Only data are stored • No sustainability information is available • No persistent and resolvable identifiers available	• Basic archiving repository • Metadata preserved • Midium term, institutional sustainaining plan (contractual deliverables with specs and schedule defined) • Persistent identifier asssignment only available for particular data records collection • Basic landing page management	• Preservation repository certified internally • Community-standard-conformed metadata preservation • Long term, institutional commitment to sustainain data preservation • Persistent identifier asssignment available for all desiminated data records collections and metadata • Automatic landing page generation, updating, and maintenance	• Preservation repository certified officially • International standard for metadata preservation • Future archiving standard chnages planned • Long term, national or international commitment to sustain the data preservation • Persistent identifier for accessible data and metadata • Data include the identifier in its metadata • Metadata follows international standards that can be harvested and indexed automatically
TRANSPARENCY, TRACEABILITY	• Limited product information available online • Detailed product information is scattered	• Theoretical basis documents on algorithms / method / model and code are available online	• Operational algorithms and documents are available online • Data are trackable	• All data product information is available online • Automatic metadata generation and

CATEGORY	LEVEL 0 (NOT MANAGED)	LEVEL 1 (PARTIALLY MANAGED)	LEVEL 2 (MANAGED)	LEVEL 3 (FULLY MANAGED)
	and difficult to obtain	• Overall data product citation trackable with unique identifier • Detailed product information is not accessible online • Detailed product information is available unpon inquiry	through unique identifiers • Detailed product information is accessible online • Data tested for presence of correct provenance metadata	updating for provenance documentation • Complete and updated data provenance available online
DATA INTEGRITY	• No data ingest integrity check • No data/ associated information integrity, authenticity and readibility check	• Data archive integrity verifiable • Data records/ associated information integrity basic check	• Data access integrity verifiable • Conforming to community data integrity technology standard • Data records/ associated information content integrity check and verification • Media readability and accessbility testing	• Data authenticity verifiable • Performance of data integrity check monitored and reported • Automatic data records/ associated information content integrity check and verification • Data authenticity verifiable • Automatic verification process, including monitoring and reporting

The following table shows how the key components of the updated data maturity matrix mapped from those in DSMM, DMSMM, and CEOS-ARD.

Table B.3 — Mapping of Key Components between Updated Data Maturity Matrix and DSMM, DMSMM, and CEOS-ARD

COMPONENT	DSMM	DMSMM	CEOS-ARD
1. General Metadata and Accessibility	Accessbility	MMP1 Metadata for Discovery, MMP2 Online Access	General Metadata
2. Per-Pixel Metadata and Usability	Usability	MMP3 Data Encoding; MMP4 Data Dcoumentation	Per-Pixel Metadata

COMPONENT	DSMM	DMSMM	CEOS-ARD
3. Radiometric, Geometric, and Atmopsheric Corrections		MMP7 Data Metrology	Radiometric Corrections; Atmopsheric Corrections; Geometric Corrections
4. Data Quality Assurance	Data Quality Assurance	MMP6 Data Validation	
5. Product Family Specification (PFS) and Data Quality Control / Monitoring	Data Quality Control / Monitoring	MMP8 Data Qulaity Control	Product Family Specifications
6. Product Assessment and Data Quality Assessment	Data Quality Assessment	MMP11 Data Processing / Reprocessing	Product Assessment
7. Preservability and Production Sustainability	Preservability; Production Sustainability	MMP9 Data Preservation; MMP12 Persistent and Resolvable Identifier	
8. Transparency / Traceability	Transparency / Traceability	MMP5 Data Traceability	
9. Data Integrity	Data Integrity	MMP10 Data Verfiication	

B.4.1. Key Advancements

The following summarizes the major advancements of the updated data maturity matrix:

- Comprehensive Coverage: By combining the technical focus of CEOS-ARD with the broader stewardship aspects of DSMM, this matrix provides a holistic approach to data maturity. Enhanced Usability: Ensures that data is not only ready for analysis but also easy to access, understand, and use.

- Quality Assurance: Integrates rigorous quality control and assessment practices to maintain high data standards.

- Sustainability and Preservation: Emphasizes the long-term sustainability and preservability of data, ensuring its value over time.

C

ANNEX C (NORMATIVE) DATASETS

C ANNEX C (NORMATIVE) DATASETS

NOTE: This annex reviews datasets for disaster resilience and climate assessments and gives evaluations of data maturity and data analysis readiness on selected example datasets.

C.1. Reviewing and Evaluating Datasets for Analysis Data Readiness and Data Maturity

Evaluating the readiness and maturity of datasets is crucial for effective disaster analysis and climate assessment. This process involves assessing the datasets' readiness level and maturity level, particularly focusing on Essential Climate Variables (ECVs). The updated data maturity matrix, developed by CEOS and the Coordination Group for Meteorological Satellites (CGMS), is used for this evaluation.

The ECV Inventory is a structured repository of climate data records, providing information on existing and planned Climate Data Records (CDRs). It is an open resource that offers insights into climate data records from various space agencies, making it highly relevant for disaster risk response and climate assessment.

- Version 5.0 of the ECV Inventory includes:

  - 918 existing CDRs

  - 371 planned CDRs

- The NCEI Geoportal catalogs 768 ECVs.

Many CDRs in these repositories have DSMM quality information, which can be mapped to the updated maturity matrix with minimal effort.

C.2. Example Evaluations Using the Updated Data Maturity Matrix

C.2.1. Example 1: NOAA Climate Data Record (CDR) of Passive Microwave Sea Ice Concentration

This dataset from NOAA is evaluated to determine its maturity level. Each component is assessed and justified, revealing that most components are managed or fully managed, with only one component being partially managed.

Dataset: NOAA Climate Data Record (CDR) of Passive Microwave Sea Ice Concentration, Version 2

URL: NOAA CDR

Data Maturity Levels:

1.   General Metadata and Accessibility (L2)

2.   Per-pixel Metadata and Usability (L2)

3.   Radiometric, Geometric, and Atmospheric Corrections (L3)

4.   Data Quality Assurance (L2)

5.   Product Family Specification (PFS) and Data Quality Control / Monitoring (L1)

6.   Product Assessments and Data Quality Assessment (L2)

7.   Preservability and Production Sustainability (L3)

8.   Transparency / Traceability (L2)

9.   Data Integrity (L3)

Table C.1 — Data Maturity Scoreboard for NOAA Climate Data Record (CDR) of Passive Microwave Sea Ice Concentration

COMPONENT	JUSTIFICATION	MATURITY
General Metadata and Accessibility	1. ISO 19115-2 Metadata 2. Released into NCEI portal 3. Support OGC catalogue / STAC / THREDDS 4. FTP server / AWS S3	L2
Per-pixel Metadata and Usability	1. NetCDF 4 2. Documents online	L2

COMPONENT	JUSTIFICATION	MATURITY
	3. Link to documents	
Radiometric, Geometric, and Atmospheric Corrections	1. Model outputs 2. Corrections	L3
Data Quality Assurance	1. Quality information is available 2. Quality assurance intermediate	L2
Product Family Specification (PFS) and Data Quality Control / Monitoring	1. Minimal data quality control and monitoring	L1
Product Assessments and Data Quality Assessment	1. Intermediate data quality assessment	L2
Preservability and Production Sustainability	1. NOAA official geoportal 2. ISO standard and OGC standard adopted 3. Regularly updated 4. Long-term geoportal & AWS storage 5. Data identifier	L3
Transparency / Traceability	1. Documents online 2. Data product citation 3. Lineage is available online	L2
Data Integrity	1. Data access integrity verifiable 2. Advanced data integrity	L3

C.2.2. Example 2: Landsat Collection 2 Surface Reflectance

This EO dataset is evaluated as a CEOS ARD dataset with a fully developed Product Family Specification. Its data maturity is managed or fully managed for all components.

Dataset: Landsat Collection 2 Surface Reflectance

URL: Landsat Collection 2 Surface Reflectance

Data Maturity Levels:

1. General Metadata and Accessibility (L2)

2. Per-pixel Metadata and Usability (L2)

3. Radiometric, Geometric, and Atmospheric Corrections (L3)

4. Data Quality Assurance (L2)

5. Product Family Specification (PFS) and Data Quality Control / Monitoring (L2)

6. Product Assessments and Data Quality Assessment (L2)

7.   Preservability and Production Sustainability (L3)

8.   Transparency / Traceability (L2)

9.   Data Integrity (L3)

Table C.2 — Data Maturity Scoreboard for Landsat Collection 2 Surface Reflectance

COMPONENT	JUSTIFICATION	MATURITY
General Metadata and Accessibility	1. STAC 2. Released into CMR and Earth Explorer 3. Support STAC / OpenSearch 4. NASA CMR / LP DAAC	L2
Per-pixel Metadata and Usability	1. GeoTIFF 2. Documents online 3. Link to documents	L2
Radiometric, Geometric, and Atmospheric Corrections	1. Processed from low level data 2. Radiometric, Atmospheric, and Geometric corrections	L3
Data Quality Assurance	1. Quality information is available 2. Quality assurance intermediate	L2
Product Family Specification (PFS) and Data Quality Control / Monitoring	1. PFS CARD4L-SR version 5.0 2. Data quality control	L2
Product Assessments and Data Quality Assessment	1. Intermediate data quality assessment	L2
Preservability and Production Sustainability	1. LP DAAC 2. STAC / OpenSearch in CMR 3. Regularly updated 4. Long-term DAAC and cloud storage 5. Data identifier	L3
Transparency / Traceability	1. Documents online 2. Data product citation 3. Lineage is available online	L2
Data Integrity	1. Data access integrity verifiable 2. Advanced data integrity	L3

C.3. Recap

Several existing datasets for disaster and climate studies were analyzed, including NCEI CDRs, WGClimate ECVs, and EO data. ARD datasets are still in development, focusing on

specifications, interoperability, and readiness for AI/ML/cloud, along with the evolution of technologies.

D

ANNEX D (NORMATIVE) TOOLS FOR DATA MATURITY

D ANNEX D (NORMATIVE) TOOLS FOR DATA MATURITY

NOTE: This annex reviews existing tools and libraries for evaluating data maturity and data analysis readiness for disaster resilience and climate assessments, especially in the context of artificial intelligence and machine learning.

Data stewardship and analysis-ready data maturity evaluation are critical components in ensuring data quality, accessibility, and usability. Various tools and frameworks have been developed to support these processes, each offering unique features and benefits. Below is an extensive review of some of the key tools and methodologies used in this domain.

D.1. Data Maturity Assessment Templates

Data maturity assessment is often supported by templates and structured frameworks. These tools help organizations evaluate and tag datasets with maturity quality information, ensuring that data is managed effectively throughout its lifecycle.

- Scientific Data Stewardship Maturity Assessment (DSMM) Model Template[13]: The template provides a structured approach to assess the maturity of data stewardship practices. It includes worksheets to evaluate various aspects of data management, such as documentation, quality control, and accessibility.

- WGISS Data Management and Stewardship Maturity Matrix (DMSMM) Schema[14]:The schema offers a comprehensive framework for assessing data management and stewardship maturity. It utilizes a matrix to evaluate datasets across multiple dimensions, including data quality, metadata, and preservation.

- CEOS Analysis Ready Data (ARD) Self-Assessment User Guide[6]: The document guides users through the process of evaluating the readiness of data for analysis.It provides detailed instructions for manually assessing various components of data, ensuring it meets the standards for analysis-ready data.

D.2. Compliance Test Tools

Compliance test tools are essential for verifying that datasets adhere to specific standards and conventions. These tools help ensure that data is consistent, reliable, and meets the required quality benchmarks.

- OGC Compliance Test Suites[15]:This compliance test tool checks compliance with Open Geospatial Consortium (OGC) standards. It validates catalogue interfaces and metadata, ensuring interoperability and standardization.

- CF-Checker[16]: The checker verifies compliance with the Climate and Forecast (CF) Metadata Convention. It ensures that datasets conform to CF standards, which are crucial for climate and weather data interoperability.

- Geospatial Metadata Validation Service (ISO 19115)[17]: The Web service validates geospatial metadata against ISO 19115 standards. It ensures that metadata is accurate, complete, and adheres to international standards.

- ESDSWG DIWG Compliance Test[18]: This compliance test provides compliance testing for various data standards. It supports a range of standards and conventions developed by the Dataset Interoperability Working Group, ensuring comprehensive validation of datasets.

D.3. Summary

Effective data stewardship and maturity evaluation are crucial for maintaining high-quality, reliable, and accessible data. The tools and frameworks reviewed above provide comprehensive solutions for assessing and improving data management practices. By leveraging these tools, organizations can ensure that their data is well-managed, compliant with standards, and ready for analysis.