Data Stewardship Maturity Report for NOAA Daily 0.25° Optimum Interpolation Sea Surface Temperature (OISST) CDR

Table 1 Legend				
Level 1	Level 2	Level 3	Level 4	Level 5
Ad Hoc	Minimal	Intermediate	Advanced	Optimal
Little or no management	Limited Management	Defined Management, partially implemented	Well-defined Management, fully implemented	Full Management, audited, measured, controlled

Table 1. Scores for the Nine DSMM Key Components at a Glance		
Preservability - 4	Accessibility - 2.5	Usability - 3.5
Production Sustainability - 4	Data Quality Assurance - 3	Data Quality Control/Monitoring - 3.5
Data Quality Assessment - 3	Transparency/Traceability - 3	Data Integrity - 3.5

NOAA National Centers for Environmental Information January 2020
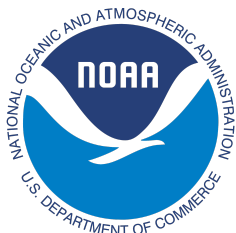
U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
National Environmental Satellite, Data, and Information Service

Cover Image: Data Stewardship Rating Diagram for NOAA Daily 0.25° Optimum Interpolation Sea Surface Temperature (OISST) CDR

Shades of green are used to represent level 1 through level 5 ratings; denoting Ad Hoc, Minimal, Intermediate, Advanced, and Optimal stages for each of the nine key components, respectively. The dark green level indicates all the practices are completely satisfied. The lighter green levels indicate only some of the practices are satisfied. The lightest green level indicates none of the practices are satisfied.

The stewardship maturity of NCEI data product, NOAA Daily 0.25° Optimum Interpolation Sea Surface Temperature (OISST) CDR, is assessed based on a reference stewardship maturity framework. The current maturity ratings of NOAA Daily 0.25° Optimum Interpolation Sea Surface Temperature (OISST) CDR are at Level 1 or higher for all nine key components with zero Level 1, one Level 2, six Level 3, two Level 4, and zero Level 5 key components.

NOAA Technical Memorandum Series
National Environmental Satellite, Data, and Information Service

The National Environmental Satellite, Data, and Information Service (NESDIS) manages the Nation's civil Earth-observing satellite systems, as well as global national data bases for meteorology, oceanography, geophysics, and solar-terrestrial sciences. From these sources, it develops and disseminates environmental data and information products critical to the protection of life and property, national defense, and the national economy, energy development and distribution, global food supplies, and the development of natural resources.

Publication in the NOAA Technical Memorandum series does not preclude later publication in scientific journals in expanded or modified form. The NESDIS series of NOAA Technical Reports is a continuation of the former NESS and EDIS series of NOAA Technical Reports and the NESC and EDS series of Environmental Science Services Administration (ESSA) Technical Reports.

Copies of earlier reports may be available by contacting NESDIS Chief of Staff, NOAA/ NESDIS, 1335 East-West Highway, SSMC1, Silver Spring, MD 20910, (301) 713-3578.

ASSESSMENT REVISION HISTORY

Revision	Description	Date
V01r00	Initial Release	11/02/2021

NOAA Technical Information Series NESDIS DSMR-00008

Version 1.0

Data Stewardship Maturity Report for NOAA Daily 0.25° Optimum Interpolation
Sea Surface Temperature (OISST) CDR

Ge Peng, Paul Lemieux III, Katy Luquire
NOAA's National Centers of Environmental Information (NCEI)
151 Patton Avenue, Asheville, NC 28801, (828) 271-4800

Recommended Citation

Table of Contents

List of Tables

Preface

In response to the President's Open Government Initiative and related policies, NOAA has committed to providing improved public access to all of its environmental information, to enable research and commercial innovation through ease of data discovery and use [*Casey*, 2016].

OneStop supports NOAA's efforts by leveraging existing access technologies and infusing specific innovations to provide improved discover, access, and visualization services for NOAA's data. Also, OneStop is viewed by a NESDIS as a pathfinder effort with an initial focus on selected high-priority datasets from NESDIS and other program data meeting OneStop standards, but eventually scalable across NOAA's data. Lastly, OneStop is implementing the USGEO Common Framework for Earth Observation Data and leveraging/supporting the NOAA Big Data Project (BDP) and Big Earth Data Initiative (BEDI) [*Casey*, 2016].

As with any process of improvement planning, agencies need to find out where they are in terms of their compliance to the federal regulations and what they need to do if any areas of non-compliance are identified. To this end, a unified framework would be beneficial for assessing the current stage of stewardship practices applied to individual datasets and for providing a road map that will guide future investments towards enhanced stewardship of environmental datasets. The value and quality of a dataset depends in part on the stewardship practices applied after its development and production. Therefore, a unified framework providing a holistic view of the quality of stewardship practices applied to individual datasets is beneficial to data stewards and users [*Casey*, 2016].

The data stewardship maturity matrix (DSMM), jointly developed by domain (data management, technology, and science) subject matter experts from NOAA's National Centers for Environmental Information (NCEI) and Cooperative Institute for Climate and Satellites – North Carolina (CICS-NC), provides such a consistent framework [*Peng et al.,* 2016]. The DSMM, leveraging institutional knowledge and community practices and standards, defines a graduated maturity scale for each of nine key components of scientific data stewardship to enable a consistent assessment of the measureable stewardship practices applied to a given data set or product.

The NOAA data stewardship maturity technical series captures stewardship maturity assessment results for individual datasets, provides consistent representation and citable documents of those assessments, ensures transparency, and allows better data quality information integration and content-based search and discovery of NOAA data.

NOAA Technical Report NESDIS DSMR-00008

Data Stewardship Maturity Report for NOAA Daily 0.25° Optimum Interpolation Sea Surface Temperature (OISST) CDR

1. Introduction

1.1 Purpose

The purpose of this document is to describe the results of stewardship maturity assessment for NOAA Climate Data Record for Mean Layer Temperature (Upper Troposphere & Lower Stratosphere from UCAR, Version 2, utilizing the Scientific Data Stewardship Maturity Matrix or DSMM [Peng, et al, 2016]. DSMM defines levels of stewardship maturity stages for Preservability, Accessibility, Usability, Production Sustainability, Data Quality Assurance, Data Quality Control/Monitoring, Data Quality Assessment, Transparency/Traceability, and Data Integrity key components. Each of these components is ranked from 'Ad hoc' to 'Optimal' (see Appendix I). This report is based on evaluation performed by NOAA OneStop metadata specialists working with Subject Matter Experts and utilizing the DSMM template [Peng, 2016].

1.2 Scope

Assessing stewardship maturity - the current state of how datasets are documented, preserved, stewarded, and made accessible publicly, is a critical step towards meeting U.S. federal regulations, organizational requirements, and user needs [Peng et al., 2016]. The goal of this document is to provide consistent and transparent stewardship maturity information to data users and decision-makers.

1.3 Dataset Abstract

This high-resolution sea surface temperature (SST) analysis product was developed using an optimum interpolation (OI) technique. The SST analysis has a spatial grid resolution of 0.25 (1/4) degree and temporal resolution of 1 day. The product uses Advanced Very High Resolution Radiometer (AVHRR) satellite data from the Pathfinder AVHRR SST dataset when available for September 1981 through December 2005, and the operational Navy AVHRR Multi-Channel SST data for 2006 to the present day. Pathfinder AVHRR SST was chosen because of good agreement with the in situ observation data. The product also uses sea ice datasets, in situ data from ships and buoys, and includes a large-scale adjustment of satellite biases with respect to the in situ data. In areas where sea ice is present, SST is estimated from sea ice concentration datasets from NASA GSFC before 2005 and then from NOAA NCEP from 2005 onwards. Preliminary products are produced daily in near real-time (1-day latency) and may be updated in the first few days if critical input data become available

after the initial production time. After two weeks, a complete or final product is generated with no additional changes expected. The preliminary products from near real-time data productions began in October 2008 for Version 2.0. The product is in netCDF-3 following the Climate and Forecast (CF) Conventions, and is accompanied by algorithm documentation, data flow diagram and source code under the NOAA CDR Program.

1.4 Document Maintenance

This document is generated and maintained by NOAA's National Centers for Environmental Information. More on policy is available at https://www.ncei.noaa.gov/.

2. Results

The data stewardship maturity assessment information is summarized in Table 1. Each component is displayed along with its corresponding score in a color-coded table.

Table 2. Dataset and Data Stewardship Maturity Assessment Metadata	
Dataset Title	NOAA Daily 0.25° Optimum Interpolation Sea Surface Temperature (OISST) CDR
Dataset Information URL	https://doi.org/10.7289/V5SQ8XB5
Data Provider POC (name; email; affiliation)	NOAA National Centers for Environmental Information, ncei.orders@noaa.gov
Dataset POC (name; email; affiliation)	NOAA Climate Data Record Program Office, oisst_contacts@noaa.gov
SMM Version (Document ID and Version Number)	NCDC-CICS-SMM_0001_Rev.1 12/09/2014
SMM POC (Name; E-mail; Affiliation)	Ge Peng, ge.peng@uah.edu, University of Alabama-Huntsville
SMM Template Version (Document ID and Version Numbers)	NCDC-CICS-SMM_0001_Rev.1 v4.0 06/23/2015
SMM Template POC	Ge Peng, ge.peng@uah.edu, University of Alabama-Huntsville
SMM Assessment Version (v<nn>r<mm>, e.g., v01r00)	v01r05
SMM Assessment Date (MM/DD/YYYY)	11/13/2015
SMM Assessment POC (Name; E-mail; Affiliation)	Viva Banzon, Viva.Banzon@noaa.gov, NOAA National Centers for Environmental Information; Ge Peng; ge.peng@uah.edu; University of Alabama-Huntsville
Stewardship Maturity Ratings (each key component) (kc1/kc2/kc3/kc4/kc5/kc6/kc7/kc8/kc9)	4 / 2.5 / 3.5 / 4 / 3 / 3.5 / 3 / 3 / 3.5
SMM Original Assessment Date (MM/DD/YYYY)	08/10/2015
SMM Original Assessment POC (Name; E-mail; Affiliation)	Ge Peng, Ge.Peng@noaa.gov, Cooperative Institute for Climate and Satellites, North Carolina (CICS-NC), North Carolina State University (NCSU)
SMM Last Modified Date (MM/DD/YYYY)	09/20/2021
SMM Last Modification POC (Name; E-mail; Affiliation)	Katy Luquire, catherine.luquire@noaa.gov , CASE Consultants International
SMM Modified Date (MM/DD/YYYY)	03/15/2018
SMM Modification POC (Name; E-mail; Affiliation)	Paul Lemieux III, paul.lemieux@noaa.gov, Earth Resources Technology, Inc.

DSMM Key Component	Stewardship Maturity Rating, Justification, and Comments
	Table 3. Stewardship Maturity Levels and Detailed Justifications for Each of Nine DSMM Key Components for the Dataset.

Preservability	Level 4
	  ▪ Archived at NOAA NCEI-NC
	  ▪ Following NOAA Climate Data Record (CDR) Research-2-Operation (R2O) transition process with the Initial Operation Capability (IOC)
	  ▪ Following OAIS RM
	  ▪ Conforms to ISO 19115-2 metadata standard
	  ▪ Conforming to NetCDF CF metadata conventions.
	  ▪ Conforming to CDR Program (CDRP) guidelines on coding and NCEI Archive Branch (AB) guidance on file and variable naming conventions per Submission Agreement (SA)
	  ▪ Plans to transition ISO metadata to newer 19115-1 standard
	
	Comments:
	No comments

Accessibility	Level 2.5
	  ▪ Collection level searchable online
	  ▪ Direct file download available via HTTP:
	  ▪ https://www.ncei.noaa.gov/data/sea-surface-temperature-optimum-interpolation/v2/access/avhrr-only/
	  ▪ THREDDS: https://www.ncei.noaa.gov/thredds/blended-global/oisst-catalog.html
	  ▪ Web Mapping Service available: https://gis.ncdc.noaa.gov/maps/ncei/cdr
	
	  ▪
	
	Comments:
	Not granular searchable
	Dissemination reports available internally, but not online
	New technology planned for OneStop and this dataset will be part of the first phase of the OneStop pilot project

Usability	Level 3.5
	  ▪ NetCDF format (CF compliant for the Obs4MIPs version)
	  ▪ Data Flow Diagram [Banzon and NOAA CDR Program, n.d.] is available online here: https://www.ncei.noaa.gov/products/climate-data-records/sea-surface-temperature-optimum-interpolation
	  ▪ C-ATBD [Banzon, 2013] is available online here: https://www.ncei.noaa.gov/products/climate-data-records/sea-surface-temperature-optimum-interpolation
	  ▪ Error estimates available in the C-ATBD [Banzon, 2013] available online here: https://www.ncei.noaa.gov/products/climate-data-records/sea-surface-temperature-optimum-interpolation
	  ▪ THREDDS allows aggregations of granules by virtually stacking files/timestamps as a single huge file.
	
	Comments:
	Data format for the new version will be NetCDF 4 and CF compliant with file attributes compliant to NOAA CDR and GHRSST guidelines

DSMM Key Component	Stewardship Maturity Rating, Justification, and Comments
	Table 3. Stewardship Maturity Levels and Detailed Justifications for Each of Nine DSMM Key Components for the Dataset.

Production Sustainability	Level 4
	  ▪ Under NOAA CDR Operation & Maintenance (O&M)
	  ▪ Updated annually
	  ▪ Funding is allocated yearly
	  ▪ Product improvement process in place
	  ▪ CDR program is managed by NCEI
	
	Comments:
	No comments

Data Quality Assurance	Level 3
	  ▪ - Coverage & outlier check for AVHRR (not run) and sea ice input (warning)
	  ▪ - In situ outliers check
	  ▪ - Land and lake point check
	  ▪ - Ice point check (>50%)
	  ▪ - Community-standard based data quality assurance metrics [Reynolds, Rayner, Smith, _et al_, 2002] available online here: https://doi.org/10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2 and [Reynolds, 2009] available online here: https://www.ncdc.noaa.gov/sites/default/files/attachments/Reynolds2009_oisst_daily_v02r00_version2-features.pdf
	
	Comments:
	No comments

Data Quality Control/ Monitoring	Level 3.5
	  ▪ - Product quality monitoring is done via the comparison tool by STAR: https://www.star.nesdis.noaa.gov/sod/sst/squam/
	  ▪ - UK Met Off SST Monitoring: https://ghrsst-pp.metoffice.gov.uk/ostia-website/fnd-monitoring.html
	  ▪ - Detection for outliers and ranges is implemented [Smith and Reynolds, 2003] available online here: https://doi.org/10.1175/1520-0442(2003)016<1495:EROGSS>2.0.CO;2
	
	Comments:
	Checksum is not provided at public FTP site but can request operations to do so

Table 3. Stewardship Maturity Levels and Detailed Justifications for Each of Nine DSMM Key Components for the Dataset.	
DSMM Key Component	Stewardship Maturity Rating, Justification, and Comments
Data Quality Assessment	Level 3 ▪ - Product algorithm is well validated [Banzon, Reynolds, Stokes, _et al_, 2014] see paper in literature available online here: https://doi.org/10.1175/JCLI-D-14-00293.1 ▪ - Research product assessed in literature [Reynolds and Chelton, 2010] available online here: https://doi.org/10.1175/2010JCLI3294.1 ▪ - Assessment carried out as the NCEI CDR R2O process ▪ - CDR IOC ▪ - Product Maturity Matrix assessment [Banzon and NOAA CDR Program, 2013] is available and online here: https://www.ncei.noaa.gov/products/climate-data-records/sea-surface-temperature-optimum-interpolation Comments: No data quality metadata Quality metadata not assessed
Transparency / Traceability	Level 3 ▪ - Product information available in literature [Banzon, Reynolds, Stokes, _et al_, 2014] which is online here: https://doi.org/10.1175/JCLI-D-14-00293.1 ▪ - C-ATBD [Banzon, 2013] available online here: https://www.ncei.noaa.gov/products/climate-data-records/sea-surface-temperature-optimum-interpolation ▪ - Source code package with Readme available online: https://www.ncei.noaa.gov/products/climate-data-records ▪ - DOI assigned: https://doi.org/10.7289/V5SQ8XB5 ▪ - NCEI OID assigned: NCEI DSI 3650_1 ▪ - Dataset Configuration Management is EIA-649-B standard compliant and diagrammed in this presentation document [Hutchins, 2015] available online here: http://www1.ncdc.noaa.gov/pub/data/sds/cdr/conferences/2015%20PI%20Annual%20Meeting%20-%20Presentations/Day_1/(A-2)%20Operations%20and%20Maintenance%20(O_M)%20of%20NOAA%20IOC%20CDRs%20-%20(Hutchins).pdf - Comments: In the process of transitioning to CDR FOC (Full Operational Capability)

DSMM Key Component	Stewardship Maturity Rating, Justification, and Comments
Table 3. Stewardship Maturity Levels and Detailed Justifications for Each of Nine DSMM Key Components for the Dataset.	
DSMM Key Component	Stewardship Maturity Rating, Justification, and Comments

Data Integrity	Level 3.5
	  ▪ - file-level checksum is generated and verified by archive at ingest and at archive
	  ▪ - using standard-based technology for generating checksum
	  ▪ - The file checksum is verified for staging sometimes but not all the time
	
	Comments:
	Recommending to verify checksum before staging the data for access and to provide checksum for each file at public ftp site

3. Acknowledgment

4. References

Casey, K. (2016), The NOAA OneStop data discover and access framework project, Version:June 3, 2016. https://cdn.ioos.noaa.gov/media/2017/12/OneStop-IOOS-DMAC-03-June-2016.pdf

Peng, G. (2015) The scientific data stewardship maturity assessment model template, Version: NCDC-CICS-SMM-0001-Rev.1 v4.0 6/23/2015. doi:10.6084/m9.figshare.1211954.

Peng, G., J.L. Privette, E.J. Kearns, N.A. Ritchey, and S. Ansari (2015), A unified framework for measuring stewardship practices applied to digital environmental datasets, *Data Science Journal*, 13, 231-253, doi: 10.2481/dsj.14-049.

Peng, G., J. Lawrimore, V. Toner, C. Lief , R. Baldwin, N. Ritchey, D. Brinegar, and S. A. Delgreco (2016) assessing stewardship naturity of the global historical climatology network-monthly (GHCN-M) dataset: use case study and lessons learned, D-Lib Magazine, 22, doi:10.1045/november2016-peng.

Banzon, V., and NOAA CDR Program, (n.d.), Daily Optimum Interpolation Sea Surface Temperature, NOAA National Centers for Environmental Information, Asheville, NC., retrieved online: https://www.ncei.noaa.gov/products/climate-dat a-records/sea-surface-temperature-optimum-interpolation (Accessed 08 December 2016).

Banzon, V., (2013), Climate Algorithm Theoretical Basis Document (C-ATBD) Daily ¼ Optimum Interpolation Sea Surface Temperature (OISST), _Rep. CDRP-ATBD-0303_, NOAA National Centers for Environmental Information, Asheville, NC., retrieved online: https://www.ncei.noaa.gov/products/climate-data-records/ sea-surface-temperature-optimum-interpolation (Accessed 08 December 2016

Reynolds, R., Rayner, N., Smith, T., Stokes, D., and Wang, W., (2002), An improved _in situ_ and satellite SST analysis for climate, _Journal of Climate_, 15(7), 1609—1625, doi: 10.1175/1520-0442(2002)015<1609:AIISAS>2.0CO;2.

Reynolds, R., (2009), OISST daily V02r00 version 2 features, NOAA National Centers for Environmental Information, Asheville, NC., retrieved online: http s://www.ncdc.noaa.gov/sites/default/files/attachments/Reynolds2009_oisst_d aily_v02r00_version2-features.pdf (Accessed 19 December 2016).

Smith, T., and Reynolds, R., (2003), Extended reconstruction of global sea surface temperature based on COADS data (1854-1997), _Journal of Climate_, 16(5), 1495—1510, doi: 10.1175/1520-0442-16.10.1495.

Hutchins, C. (2015), Operations and Maintenance (O&M) of NOAA IOC CDRs, http:/
/www1.ncdc.noaa.gov/pub/data/sds/cdr/conferences/2015%20PI%20Annual%20Meet
ing%20-%20Presentations/Day_1/(A-2)%20Operations%20and%20Maintenance%20(O_
M)%20of%20NOAA%20IOC%20CDRs%20-%20(Hutchins).pdf (Accessed 08 December
2016).

Appendix I: The Scientific Data Stewardship Maturity Matrix (DSMM)

Table A1: This matrix (Version: NCDC-CICS-SMM-0001-Rev.1. 12/09/2014) describes the criterion used to evaluate data stewardship maturity for each of the nine DSMM key components [*Peng et al.*, 2015].

DSMM Component	Level 1 *Ad hoc* Little or no management	Level 2 *Minimal* Limited management	Level 3 *Intermediate* Defined management, partially implemented	Level 4 *Advanced* Well-defined management, fully implemented	Level 5 *Optimal* Full management, audited, measured, controlled
Preservability *(The state of being preservable)*	Any storage location Data only	Non-designated repository Redundancy Limited archiving metadata	Designated archive Redundancy Community-standard archiving metadata Conforming to limited archiving standards	Level 3 + Conforming to community archiving standards	Level 4 + Archiving process performance controlled, measured, and audited Future archiving standard changes planned
Accessibility *(The state of being searchable and accessible publicly)*	Not publically available person-to-person	Publically available direct file download (e.g., via anonymous FTP server) Collection or dataset level searchable online	Level 2 + Non-standard data service Limited data server performance Granule/file level searchable Limited search metrics	Level 3 + Community-standard data service Enhanced data server performance Conforming to community search metrics Dissemination report metrics defined and implemented internally	Level 4 + Dissemination reports available online Future technology and standard changes planned

Usability *(The state of being easy to use)*	Extensive product-specific knowledge required No documentation online	Non-standard data format Limited documentation (e.g., user's guide online)	Community standard-based interoperable format & metadata Documentation (e.g. source code, product algorithm document, processing or/and data flow diagram) online	Level 3 + Basic capability (e.g., subsetting, aggregating) & data characterization overall/global, e.g., climatology, error estimates) available online	Level 4 + Enhanced online capability (e.g., visualization, multiple data formats) Community metrics of data characterization (regional/cell) online External ranking
Production Sustainability *(The state of data production being sustainable and extendable)*	Ad Hoc or Not applicable To obligation or deliverable requirement	Short-term Individual PI's commitment (grant obligations)	Medium-term Institutional commitment (contractual deliverables with specs and schedule defined)	Long-term Institutional commitment Product improvement process in place	Level 4 + National or international commitment Changes for echnology planned
Data Quality Assurance *(The state of data quality being assured)*	Data quality assurance (DQA) procedure unknown or none	Ad Hoc and random QA procedure not defined and documented	DQA procedure defined and documented and partially implemented	DQA procedure well documented, fully implemented and available online with master reference data Limited data quality assurance metadata	Level 4 + DQA procedure monitored and reported Conforming to community quality metadata & standards External review

Data Quality Control/ Monitoring *The state of data quality being controlled and monitored*	None or Sampling unknown or spotty Analysis unknown or random in time	Sampling and analysis are regular in time and space Limited product-specific metrics defined & implemented	Level 2 + Sampling and analysis are frequent and systematic but not automatic Community metrics defined and partially implemented Procedure documented and available online	Level 3 + Anomaly detection procedure well-documented and fully implemented using community metrics, automatic, tracked and reported Limited quality monitoring metadata	Level 4 + Cross-validation of temporal & spatial characteristics Physical consistency check Conforming to community quality metadata & standards
Data Quality Assessment *(The state of data quality being assessed)*	Algorithm/ method/model Theoretical basis assessed (methods and results online)	Level 1 + Research product assessed (methods and results online)	Level 2 + Operational product assessed (methods and results online)	Level 3 + Quality metadata assessed Limited quality assessment metadata	Level 4 + Assessment performed on a recurring basis Conforming to community quality metadata & standards External ranking
Transparency/ Traceability *(The state of being transparent, trackable, and traceable)*	Limited product information available Person-to-person	Product information available in literature	Algorithm Theoretical Basis Document (ATBD) & source code online Dataset configuration managed (CM) Unique Object Identifier (OID) assigned (dataset, documentation, source code) Data citation tracked (e.g., utilizing Digital Object Identifier	Level 3 + Operational Algorithm Description (OAD) online, OID assigned, and under CM	Level 4 + System information online Complete data provenance online

Data Integrity *(The state of data integrity being verifiable)*	Unknown or no data ingest integrity check	Data ingest integrity verifiable (e.g, checksum technology)	(DOI) system) Level 2 + Data archive integrity verifiable	Level 3 + Data access integrity verifiable Conforming to community data integrity technology standard	Level 4 + Data authenticity verifiable (e.g., data signature technology) Performance of data integrity check monitored and reported