



Observatory Best Practices Overview

May 21, 2019





COI

Observatory Best Practices

- Best Practices Research and White Papers
- Best Practices and Self Assessment Tools
- Best Practice List Quick Overview







Best Practices Research and White Papers

- Under a work activity approved by the NSF, COL has recently examined observatory industry trends and best practices. These trends and best practices were iteratively documented using extensive literature research and website reviews of major observing systems.
- This information is presented in four white papers, describing ~46 best practices:
 - Data Product Quality (29)
 - Data Identification, Citation and Tracking (6)
 - Community Engagement (7)
 - Observatory Performance Metrics (7)
- These white papers provide best practice Self-Assessment Tools







Best Practices - Disclaimers

- What they represent:
 - Survey of existing observatory best practices
 - Idealized world of achievable best practices
 - Guide for self assessment and planning
 - Simplified, easy to understand and apply
- What they do not represent:
 - Technical assessments
 - All best types of best practices
 - Implementation guidelines







Best Practices Research and White Papers

Best Practices Research Projects

- Data Product Quality
- OOI Data Identification & Usage
- Community Engagement
- Performance Metrics

Research Best Practices

Synthesize Best Practices

Develop Self Assessment

Recommendations

Best Practice White Papers

- Data Product Quality
- Data Identification, Citation and Tracking
- Community Engagement
- Observatory Performance Metrics

Focus on Industry Drivers

Iterative Research to Confirm

Context for Findings

Reference Material







Best Practice White Papers - Scope and Methodology

- Scope
 - Examine best practice trends and drivers for current industry best practices, provide analysis, recommendations and reference material
 - Provide a best practice Self-Assessment Tool that enables an existing or new organization to assess their current best practice capabilities and maturity level
- Methodology
 - Best practices were iteratively researched, synthesized, refined and validated using extensive literature reviews and website reviews of major observing systems
 - Validate best practices and best practice self-assessment tools through interviews with 2-3 mature observatories







Best Practice Research Sources

Four Sources for Research Information:

- Major Observing Systems
 - Nomination Criteria: Oceanography Based; Direct Producer of Data; Large Facility; Mature Organization
- Aggregator/Data Centers
 - Nomination Criteria: Oceanography Based; Mature Organization
- Data Certification Organization List
 - Nomination Criteria: Oceanography Based; Widely Accepted
- Oceanography Conference Abstracts
 - Nomination Criteria: Oceanography Based; Large Conference







Best Practices and 7 Self Assessment Tools

- 46 Best Practices Identified, Grouped into 7 Categories
 - Data QA/QC Procedures (8)
 - Data Support/Services (7)
 - Metadata (4)
 - Interoperability (7)
 - Data Identification and Usage Tracking (6)
 - Community Engagement (7)
 - Observatory Performance Metrics (7)
- 7 Best Practices Self Assessment Tools
 - Each Best Practice Category has a Self Assessment Tool

Best Practices are described in 4 Best Practice White Papers





COL

Self Assessment Tool Overview









Best Practice Self-Assessment Tool Example

Data QA/QC Capability Maturity Levels

QA/QC performance standards defined, quality tracked, metrics reported; All data is reviewed by

Optimizing	HITL, with feedback for algorithm improvement; All relevant data uses in situ data comparison and is included in metadata				
Managed	Robust data mana procedures are ce improved frequent / post calibration c	Example To Show CMM Lev	Data rep thms are 'els II releva	oository and e automated and nt data uses pre	
Implemented	Data QA/QC proc implemented and reviewed by Huma	Next Slides Show Best Pract Lists and CMM Levels	tice QC proce mated; : calibrat	edures >50% of data is ion data	
Defined	Some elements of publicly accessible	e; Working towards implementing Data QA/QC co	QA/QC procedures defined and A/QC community best practices		
Initial	Aware that data product quality best practices are important; Initial stages of information gathering and planning				
	Benchmark/RankConcept, piloInadequate, aCapability Material	ing use "defined assessment levels" t, mature adequate, exemplar aturity Model * (shown on slide)		Self Assessment Tool Current Level Aspirational Level	

* Capability Maturity Model (CMM) developed by Software Engineering Institute (SEI) at Carnegie Mellon University in 1986 (Paulk et al., 1993)







Observatory Best Practices – Abbreviated Lists

- Four Areas Within Data Product Quality *
 - Data QA/QC Procedures (8)
 - Metadata (4)
 - Interoperability (7)
 - Data Support/Services (7)
- Observatory Performance Metrics (7)

High level lists follow, 4 best practice white papers provide:

- Industry drivers
- Current state of industry
- Detailed best practice descriptions
- Self-Assessment tools and instructions
- * Data Product Quality defined as "user trust in data quality"







Data Product Quality - Best Practices List

Data Quality Control

- 1. Data Management Plan is developed, current, and publicly accessible
- 2. Data QA/QC procedures are documented, maintained, and aligned with community recognized standards
- 3. Data repository procedures aligned with community recognized standards
- 4. Data QC algorithms are automated, frequently reviewed, and improved
- 5. Data QC use humans in the loop with relevant subject matter experience
- 6. Pre and post deployment calibrations are used to modify/annotate data, included in metadata
- 7. In situ data are collected, used to modify/annotate data, included in metadata
- 8. Versioning of modified datasets are available and accessible







Best Practice Self-Assessment Tool Example

Data QA/QC Capability Maturity Levels



QA/QC performance standards defined, quality tracked, metrics reported; All data is reviewed by HITL, with feedback for algorithm improvement; All relevant data uses in situ data comparison and is included in metadata

Managed Robust data management plan developed, current and publicly accessible; Data repository and procedures are certified by community recognized entity; All data QC algorithms are automated and improved frequently; >50% of relevant data uses in situ data comparison; All relevant data uses pre / post calibration data and is included in metadata

Implemented Data QA/QC procedures align with a community recognized standard, QA/QC procedures implemented and publicly accessible; >50% of data QC algorithms are automated; >50% of data is reviewed by Human In The Loop; >50%) of relevant data uses pre and post calibration data

Defined Some elements of data management plan developed; Data products have QA/QC procedures defined and publicly accessible; Working towards implementing Data QA/QC community best practices

Initial Aware that data product quality best practices are important; Initial stages of information gathering and planning









Data Product Quality - Best Practices List

Metadata

- 1. Metadata aligns with community recognized standards
- 2. Sufficient metadata always accompanies data products
- 3. Validation of the metadata has been performed
- 4. Metadata file creation processes are automated







Best Practice Self Assessment Tool

Metadata Capability Maturity Levels

Metadata performance standards defined, quality tracked, metrics reported; All data products come with a robust set of metadata including data DOI; Metadata provided is checked using automated file validations; All metadata files creation is automated, process defined and available

Managed Metadata SOPs align with community recognized standards; All data products come with a minimum set of metadata; Metadata provided is periodically manually checked for completeness; Significant (>50%) portion of metadata files creation is automated

Implemented Metadata procedures (SOPs) developed and publicly accessible; working towards implementing metadata using community standards; Metadata provided is periodically manually checked for completeness

Defined Some elements of Metadata procedures (SOPs) developed; working towards defining metadata using community standards

Initial Aware that metadata best practices are important; Initial stages of information gathering and planning









Data Product Quality - Best Practices List

Interoperability

- 1. Controlled vocabulary & defined ontology are used that adhere to community standards
- 2. Community recognized and supported data format standards are used
- 3. Dataset provenance transitions are clearly documented and available to users
- 4. Third-party systems are synchronized with evolving resources.
- 5. Provide clear data, metadata, and community software reuse statement (license)
- 6. Community aligned usability design concepts are employed for user interfaces
- 7. Actively participate in interoperability best practices communities

Note: White paper identifies10 Supporting Best Practices Previously Defined







Best Practice Self Assessment Tool

Interoperability Capability Maturity Levels

Optimizing

Community accepted controlled vocabulary and defined ontology defined and applied; All dataset provenance transitions are documented and publicly available; Provide support to third-party systems to remain synchronized; Provide community aligned usability design concepts for user interfaces

Managed All data products come with a robust set of metadata; Allow third-party systems to remain synchronized with evolving resources; Provide clear data, metadata, community software reuse statement (license); Actively participate in interoperability best practices communities

Implemented High level persistent data identifier provided with all data sets; All data products come with a minimum set of metadata; Basic controlled vocabulary and defined ontology implemented; Community recognized and supported data format standards implemented

Defined Basic controlled vocabulary and defined ontology defined; Community recognized and supported data format standards defined; Actively participate in interoperability best practices communities

Initial Aware that persistent interoperability best practices are important. Initial stages of information gathering and planning









Data Product Quality - Best Practices List

Data Support Services

- 1. Data product user training and reference materials are publicly accessible (user self service)
- 2. Enhanced user technical support for data products and system is available (Enhanced user support)
- 3. Enhanced data user services are enabled by unique user registration
- 4. Users are provided with the ability to request recurring data product downloads
- 5. Users are provided with the ability to access and execute saved community code and API functions
- 6. User performance metrics are defined, tracked, and reported
- 7. Data Centers and Aggregators provide data management services







Best Practice Self Assessment Tool

Data Support Services Capability Maturity Levels

Opti	mizing	Advanced data user services enabled by user registration, including curate data, read access to protected (embargoed) data, enter data user trouble tickets, recurring data product downloads, access to saved community code, run code on observatory hosted server; Advanced user help desk support, including live chat and phone support from trained staff with access to specific data product expertise	
Ма	naged	Online data products catalog with metadata information, online QA/QC manuals; Advanced user services, including live chat and phone support from trained staff with access to specific data product expertise; Help desk services, including user ability to view ticket status online; User support performance standards defined, quality tracked, metrics reported	
Imple	emented	Basic technical support self services tools such as online knowledge base, FAQs; User online interface to request support; Basic data product help desk services, including user ability to initiate trouble ticket and request ticket status	
Def	fined	Some elements of Data Support Service procedures (SOPs) developed; working towards defining implementation strategies and next steps; Some basic training material available, including tutorials, FAQs, videos, recorded webinars	
lı	nitial	Aware that data support services best practices are important; Initial stages of information gathering and planning Self Assessment Tool Aspirational Level	





Performance Metrics Best Practices

- 1. Develop and maintain 5 year Science Plan and ConOps strategic plan
- 2. Develop and maintain 1 year Annual Work Plan
- 3. Develop and maintain Key Performance Indicators (KPIs)
- 4. Develop and maintain Performance Metrics that are Specific, Measurable, Achievable, Relevant, Time-Bound
- 5. Develop, maintain, and distribute Performance Metrics Scorecard
- 6. Perform data-driven management reviews using KPIs and performance metrics
- 7. Perform performance gap analysis and continuous process improvement





COL

Questions?

Contact Information

Thomas Kearney OOI 1.0 Transition Manger <u>tkearney@oceanleadership.org</u> (503) 475-1406

Chris Rutherford OOI 1.0 Program Director and PI <u>crutherford@oceanleadership.org</u> (202) 448-1257



