

Agenda

- OOI 1.0 Metrics
- CI 2.0 Structure
- CI Risks and Mitigations for OOI 2.0
- Strategic Roadmap
- Data Management
- QA/QC Outline
- Analysis of Alternatives (AoA)
- Questions
- MIOs

Agenda

- OOI 1.0 Metrics
- CI 2.0 Structure
- CI Risks and Mitigations for OOI 2.0
- Strategic Roadmap
- Data Management
- QA/QC Outline
- AoA

OOI 1.0 Metrics

- Visits to site
 - ooinet – 6,279
 - M2M – 18,345,567
 - opendap – 2,549
 - rawdata – 3,739
- Data transferred
 - Stored data – 1.9 Tb
 - Raw data – 1.78 Tb
- System Stability
 - 99.94%
 - Max uFrame CPU utilization 20%
 - Max Cassandra CPU utilization 10%
- Tickets
 - 701 total open tickets
 - 86 new tickets logged
 - 86 tickets closed – 30 of which were new
- 1190 Annotations - notes on quality tied to data in production
- 2,359 uncabled data sets available on ERDDAP

Metrics represent a 3 month snapshot

Source - OOI Q2 Quarterly Report (April 2018 – June 2018)

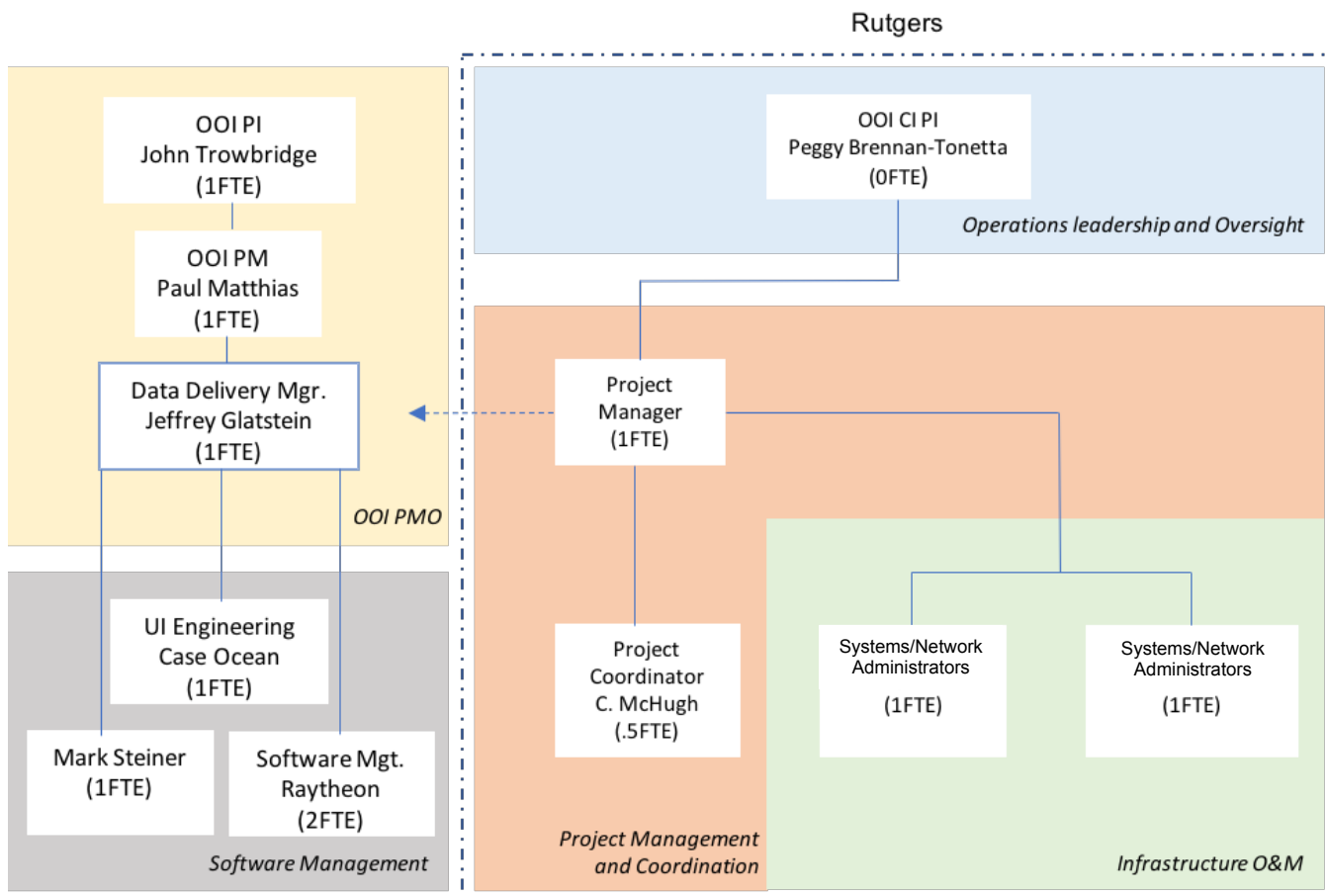
Agenda

- OOI 1.0 Metrics
- **CI 2.0 Structure**
- CI Risks and Mitigations for OOI 2.0
- Strategic Roadmap
- Data Management
- QA/QC Outline
- AoA

CI 2.0 : Data Delivery Manager

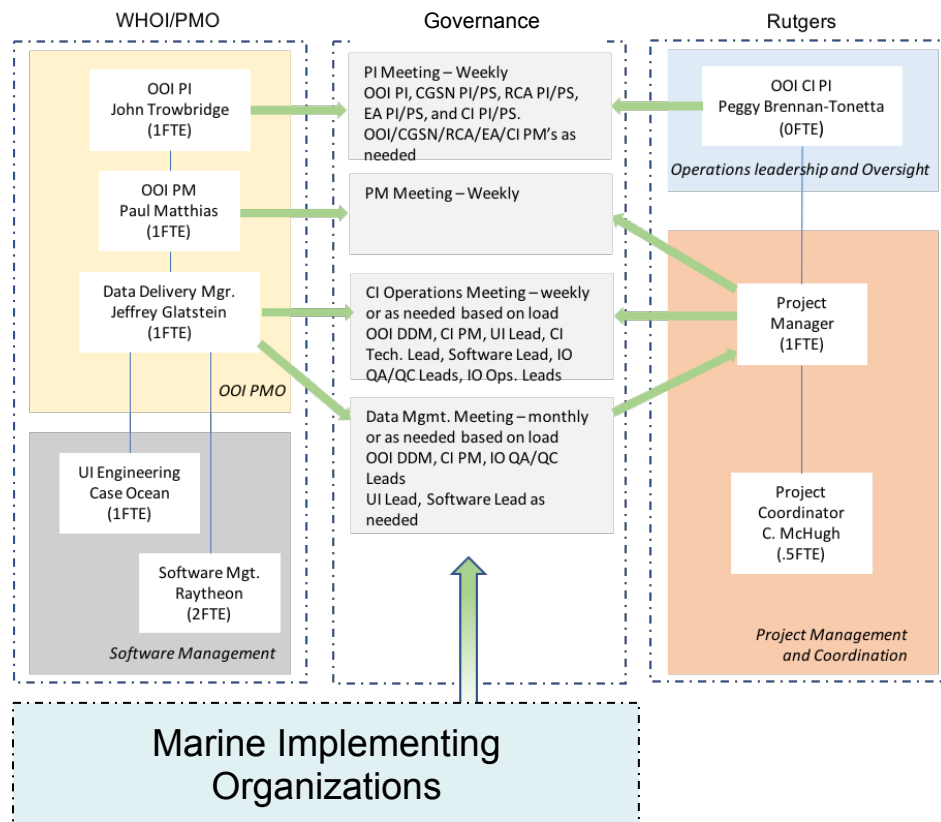
- Oversees development efforts for the CI platform
 - Data Ingestion process and logic
 - User Interfaces for data input and retrieval
 - Data Portal development and maintenance
 - General automation and simplification of current processes and systems
- Data management
 - Storage methods and structure
 - Establish uniform QA/QC process and procedures that are repeatable, automated to the extent possible and consistent across MIOs
 - Maintenance processes – data fixes, performance
- Acts as Operations Manager - point of contact for RU Project Manager incase of unplanned down time or emergency fixes
- Develops roadmap, in collaboration with stakeholders, for short term CI deliverables and long term CI strategic direction
- Cyber Security oversight
- Software licensing, versioning and upgrade analysis
- Redmine ticket review and prioritization lead
- CI self evaluation lead
- AoA lead

CI 2.0 Organization



CI 2.0 Operations

- Organized to foster collaboration, open and transparent
- MIOs are represented at all meetings
- Engage representation from all points along the data journey
- Consistent review of top priorities
- Consistent communications will allow the development and support teams to be nimble and responsive
- Bring together development, support and data creation resources on a regular basis



Agenda

- OOI 1.0 Metrics
- CI 2.0 Structure
- **CI Risks and Mitigations for OOI 2.0**
- Strategic Roadmap
- Data Management
- QA/QC Outline
- AoA

CI Risks for OOI 2.0

- Incomplete documentation
 - No documents at CCB level – goal is 70% - 80% complete for passing review
 - No single document outlines the system from end to end
 - Completing documentation will divert resources from enhancements
- Knowledge transfer of data team
 - Manual preparation of data
 - Ingestion problem resolution
- Knowledge transfer of software development team
 - No one individual knows all the interactions between the many subsystems – lack Interface Control Documents (ICD)
 - Port agents are not well known
- Approximately 600 outstanding Redmine tickets pertain directly to CI
 - Tickets in order of volume are: invalid and missing data, missing or needed functionality, uFrame bugs, UI bugs,
 - Review process needs to be more formalized and predictable – widening the levels of review
 - Approximately 198 tickets are 2 years or older (42 >= 3yrs)
 - 124 marked Immediate, Urgent or High
- 32% gap in requirements – 328 requirements researched across 18 categories
 - Data Access – 51% requirement met : Priority to registered users, identification of resource intensive requests
 - Asset Management – 68% requirement met : manual UID vs system generated UID creation, automated calibration ingest
 - Data Collection – 84% requirement met : QA/QC, data input monitoring of failures, interruptions and event alerts
- Gaps in user experience and system efficiencies
- QA/QC program not as mature as needed
 - Bad data does not find support, support needs to find data
 - 7 automated QA flag checks, 5 not working, results are not logged for reference and occur after user asked for data

CI Risks : Mitigation

- Temporary extension of development resources
 - Retain the expertise to cover majority of CI functions
 - Grow ‘in house’ development resources and knowledge
 - Bring documentation up to a level where it’s useful to on-boarding technical resources
- Build internal expertise in raw data preparation
 - MIOs attended data workshop
 - Short term contract of Data team personnel for problem resolution
 - Train development resources to understand the process
- PMO management of software development
 - Coordinated approach across functional areas (ingestion, UI, UX...)
 - Establish process and procedures that foster collaboration across technical resources and stake holders
 - Apply short term (tactical) decisions while managing and developing long term strategic plans
- Target high value tickets early
 - Defect resolution first (Immediate and Urgent)
 - System effectiveness with goal to reduce ability to create data issues
 - System efficiency with goal to reduce number of times data needs to be manually touched
 - Complete a review of Redmine tickets with goal of prioritization and cancelling those no longer relevant
- Develop repeatable method for system and data metric reporting
 - Need better insight to data and system usage to help inform where to apply resources. Current method is reactionary to ticket creation driven by issues.
 - While there are monitors in place, the distribution of the data and access to the logs needs to be expanded – issues should find us, not wait for being noticed on a screen
- Introduce formalized and uniform data quality program

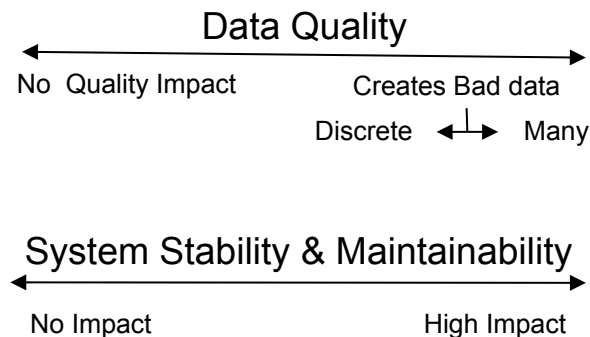
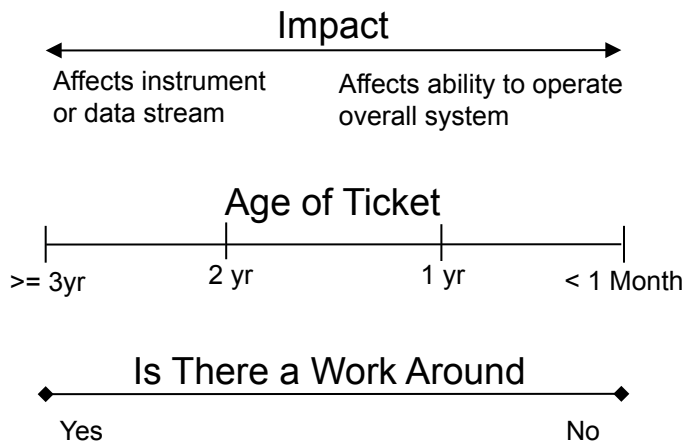
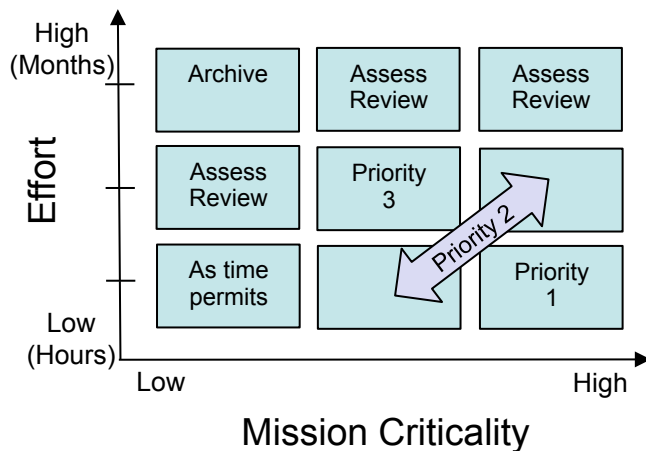
Agenda

- OOI 1.0 Metrics
- CI 2.0 Structure
- CI Risks and Mitigations for OOI 2.0
- **Strategic Roadmap**
- Data Management
- QA/QC Outline
- AoA

Strategic Roadmap

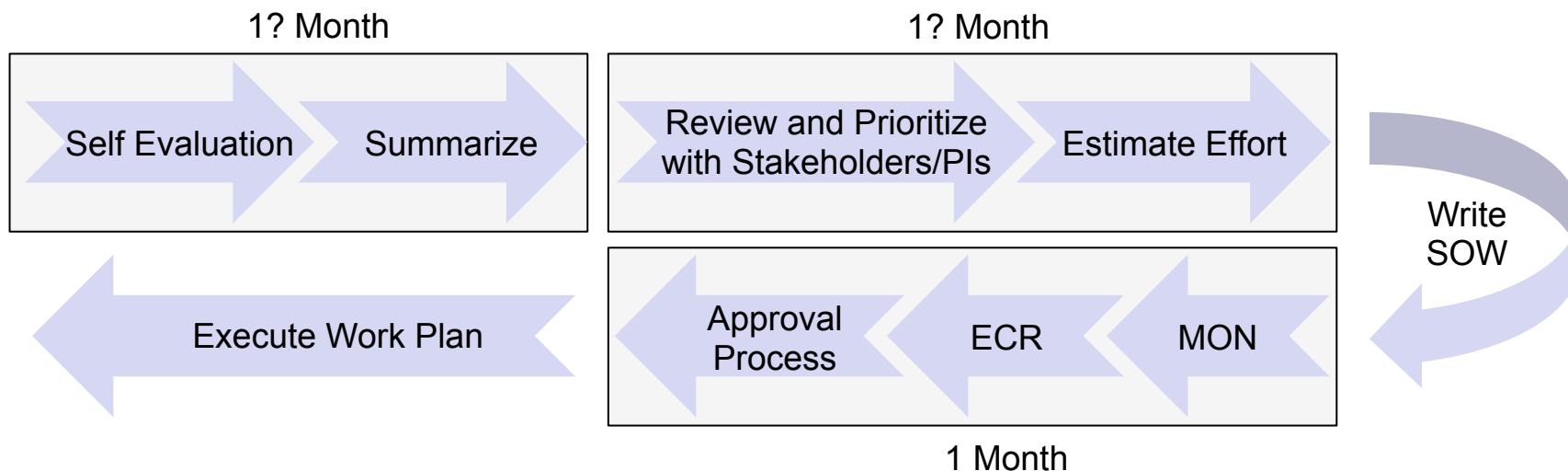
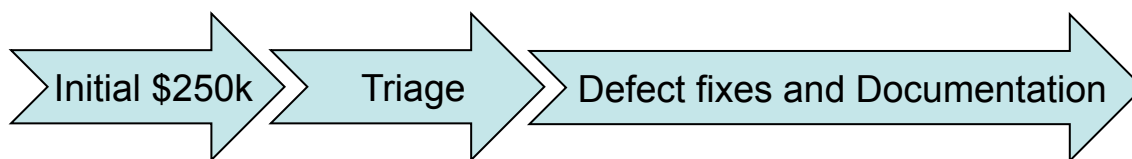
- Tactical – 2 - 3 months
 - Self evaluation – stakeholder interviews and user surveys
 - Functional Capabilities study of Requirements
 - TDP Documentation
 - Redmine tickets
 - User experience (UX) – how easy is the system to use – ex. ability to enter and find data
 - High priority issues and defects
 - Missing or bad data, system down, unable to retrieve data
- Tactical/Strategic – 4 - 5 months
 - Data input prior to ingestion
 - Mature user interface for entering and maintaining data (asset management, recovered and cruise)
 - Raw data repository maintenance
 - Reduction of .csv and manual scripts, clean-up and redesign of storage for processing assets
 - Data quality scripts, automation and reporting
 - User Interface and website
 - Documentation
- Strategic
 - Analysis of Alternatives (AoA)
 - Version upgrades
 - Data model redesign – analyze structure to simplify data maintenance
 - Design and development of original mission critical requirements not currently implemented

Redmine Ticket Assessment



External Resource Plan

- Raytheon IIS – uFrame framework of subsystems
- Raytheon IDS – Instrument agent drivers, data set parsers, integration and Cassandra database
- Case Ocean Services – OOINet integration and web interface



Agenda

- OOI 1.0 Metrics
- CI 2.0 Structure
- CI Risks and Mitigations for OOI 2.0
- Strategic Roadmap
- **Data Management**
- QA/QC Outline
- AoA

Data Management

- Data
 - Stored within the system but has a life outside of it as well
 - Affected by both tactical and strategic objectives
 - Management of it will drive system enhancements, policies and procedures
- Objectives
 - Develop tools to increase efficiency and effectiveness of loading data
 - Build library of algorithm and data quality checks
 - Build uniform repeatable automated QA/QC process
 - Report on metrics of data use
 - Report on metrics of data quality
 - Document methods of retrieving data from the system
 - Introduce ability to correct data in place
 - Data Governance
 - Cyber security

Agenda

- OOI 1.0 Q2 Metrics
- CI 2.0 Structure
- CI Risks and Mitigations for OOI 2.0
- Strategic Map
- Data Management
- **QA/QC Outline**
- AoA

QA (process)/QC (product) Program Outline

- Document data quality requirements
 - Define process of documentation - taking new data into account
 - Define rules for measuring
 - Define data point intervals
 - Normalize processes across organizations
- Repeatable process to assess data
 - Create quality baseline
 - Data profiling
 - Individual data element range
 - Load statistics
- Metadata management
 - Document schemas
 - Document data flow
 - Document process to update data elements
- Check data validity on an ongoing basis at determined intervals
 - Run QA/QC scripts
 - Implement automated data scans
 - Ad-hoc manual data checks
- Feedback
 - Document findings
 - Document fixes
 - Communicate to data owners
 - Communicate to users

Work output using one instrument – CTD

Common Framework

- Documented workflow following the life of the data
- Identified measures and values
- Identified tests
- Code base
- Stored results
- Statistics over time
- Communications method

Agenda

- OOI 1.0 Q2 Metrics
- CI 2.0 Structure
- CI Risks and Mitigations for OOI 2.0
- Strategic Map
- Data Management
- QA/QC Outline
- **AoA**

Analysis of Alternatives

- Goals and Output
 - Goal of the study is to determine whether the current system should continue as is, be modified, be replaced in part or in whole.
 - Work product will be recommendation of approach and the technologies and/or products to be used.
- Process
 - Initial research
 - Assemble analysis panel
 - Guidelines and criteria
 - Research alternatives
 - Gather facts using decision matrix
 - Deliver recommendations and gain consensus
 - Plan and execute product roadmap

AoA : Initial Research

- Understand the problem – assemble available documentation
 - Review self eval performed at start of OOI 2.0 in detail
 - Add any new requirements that have surfaced
 - Challenge any existing requirement, feature and/or function that no longer applies
- Current CI Architecture is very modular – think Lego bricks
- Parts are easier to deal with vs the ‘whole’ – CI can be broken into 4 basic functions
 - Data Staging
 - Data Preparation
 - Data Ingestion
 - Data Dissemination
- Research to identify applicable products and technologies
 - Speak with colleagues both in and out of Ocean Sciences, Sciences and Industry
 - Google, Safaribooks, various blogs and websites
 - Understand trending technologies and providers
 - Inquire into new products – both “end to end” or those that solve a specific need
 - Cloud hosting or virtualization options for operations

AoA: Preliminary Results

- **Speak with colleagues**
 - Doug MacFadden, Harvard Catalyst – all open source. One off solutions to solve specific data questions with some re-use. Uses RDF and ATSCALE to virtualize meta data. SPARQL to query the RDF, i2b2
 - Cimpres – microservices deliver data via dumb pipes. Virtual semantic data layer and query environment using combination of Redshift and Snowflake (full virtual data warehouse).
 - Juliane Schneider, eagle-i – Software Carpentry advocate, favors build over buy with use of POCs, eagle-i mistakes: over engineered, difficult to support, script driven. Feels pain of just connect data virtually across streams with some functions. Research Data Alliance
- **Blogs and websites**
 - Research Data Alliance
 - UX Magazine (website)
 - data.nasa.gov
- **Technologies - Providers - Partners**
 - Micro services with endpoint/dumb pipe communications
 - Amazon Redshift (based on PostgreSQL)
 - CyVerse – Limpid project – image processing, various scientific based platform/solutions, data platform. Provides some examples and concepts (some of what not to do)
 - Axiom Data Science
- **Products – both “end to end” or those that solve a specific need – mostly researched data management platforms**
 - JupyterHub – possible front end recovered and asset management data
 - Hortonworks – data flow and messaging. Looked at streaming analytics, not for the analytics but possibly for data set pre-calc distribution and cross stream data combinations
 - Pentaho/Hitachi Vantara - big data integration – zero code
 - SAP – University partnership, data platform
- **Thoughts**
 - Can keep logic and data storage in place but pre-calc into more traditional access methods, storage or tools
 - OOI DaaS is a very specific niche
 - Don't force functionality – ex. ZPLSC

AoA: Assemble Analysis Panel

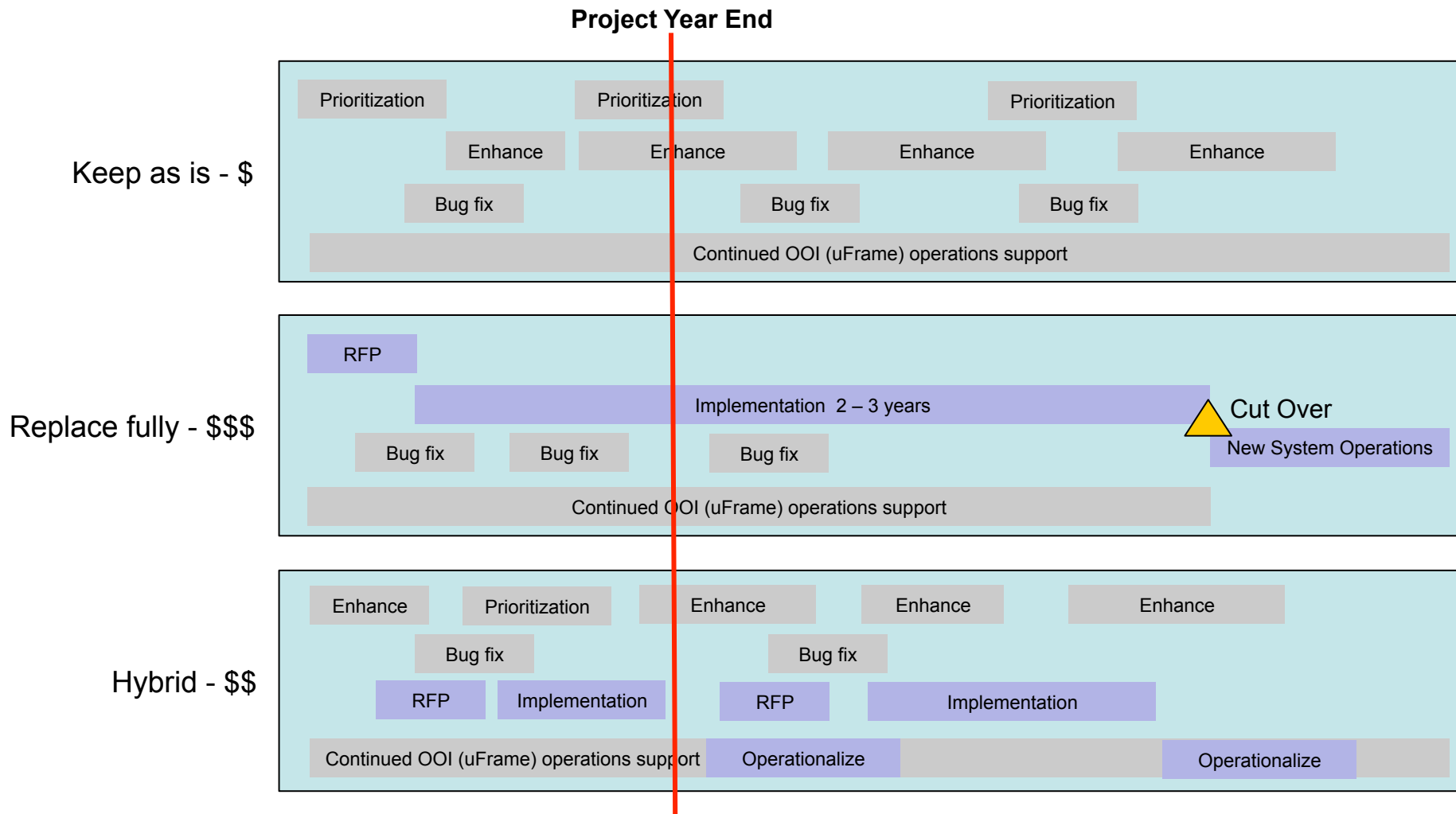
- Recruit system stake holders with varying focus
 - Individual responsible for entering data into a system
 - User of the transformed (L1, L2) and/or raw data through web user interface
 - Understanding of user experience (UX) needs
 - Data analyst with ability to understand QA, metadata and data interactions
 - Highly technical background in software, data architecture and product management
- Seek out differing backgrounds
 - Engineer – understands data at the instrument level
 - Scientist – correlates the data and understands its importance in research
 - QA lead – most familiar with wide set of data and implications of maintenance
 - Information Technologist – developer, architect and data architect
 - PMO - Management
- Keep committee size manageable and mindset collaborative
 - Large committees are not efficient – recommendation will be 7 or 9
 - Having a variety of background provides many points of view
- Be clear about time commitment
 - Everyone will need to contribute to the analysis and research as it pertains to their area of expertise
 - Attendance of meetings and presentations will not be the only time commitment

AoA: Process Timeline

Month	Task	Work Product
November	Assemble Panel	List of members, distribution list and initial meeting schedule
	Prepare and review background information	Document summarizing OOI features and functions in current state
December	Determine scope and set baseline objectives	Decision matrix with criteria, weights, priorities, alternatives
	Research vendors and technology	List of candidates with preliminary weights applied
January	Detailed research on viable candidates or new information	Edited list of candidates with more accurate matrix
	Vendor presentations	Schedule of demos and discussions
February	Vendor presentations	Schedule of demos and discussions
March	Compile findings and determine course	Summary document of process, research performed and summary matrix of overall values. Recommendation statement

Note: This is an aggressive schedule. Results need to be completed by June 2019

AoA Implementation Timeline - Illustrative



Summary

- OOI 1.0 Metrics
 - Redmine tickets opened vs closed is even
 - Understanding how the system is used will inform the management and enhancement of it going forward
- CI 2.0 Structure
 - Consolidation of oversight of resources and responsibilities
 - Designed to foster collaboration, transparency and consistency
- CI Risks and Mitigations for OOI 2.0
 - Knowledge transfer, QA/QC and gaps in requirements are biggest risks
 - Extension of resources, QA/QC approach, dev management and Redmine ticket approach will mitigate risk
- Strategic Roadmap
 - Target defect and system quality tickets immediately, perform a CI self evaluation
 - High value larger efforts such as automation, UX, data quality and continuing documentation
 - Plan to manage external resources and set work priorities for next year
- Data Management
 - Data has both physical and logical aspects that drive enhancements, policies and procedures
 - Need for QA/QC, metrics and reporting
- QA/QC Outline
 - QA/QC approach – do not need to complete all steps for every data element at once – gain value early
 - Need to find balance that is sustainable by resources
- Analysis of Alternatives (AoA)
 - Will deliver a recommendation of approach to CI future
 - Outlined research approach
 - Reviewed the 3 possible outcomes and impacts – Keep as is, Replace fully or Hybrid

Questions

MIO Activities

Current MIO Data and QA/QC Activities

- **How has the role of the MIOs changed?**
 - MIOs now have the responsibility for:
 - Asset Management
 - Data Ingestion
 - Data QA/QC
 - HelpDesk Support

Current MIO Data and QA/QC Activities

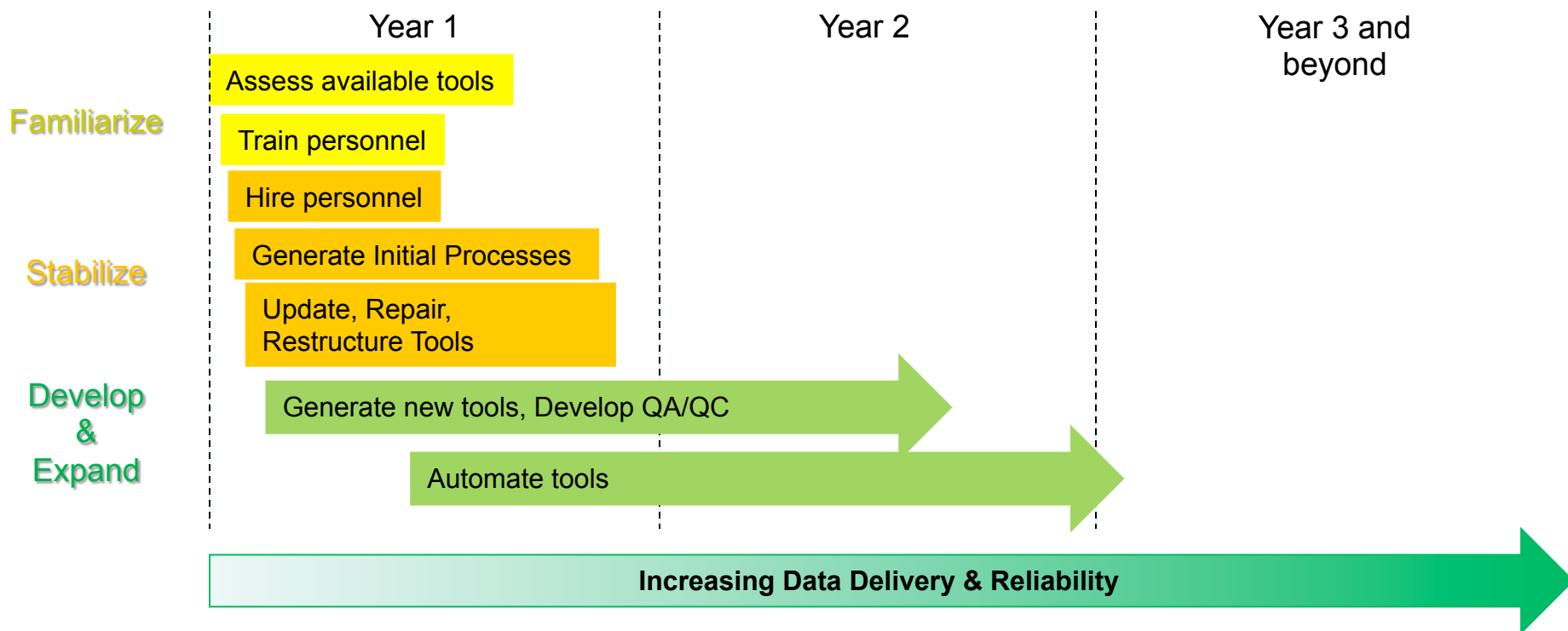
- **How has the role of the MIOs changed?**
 - MIOs are building teams including hiring of new personnel
 - Performing training within the teams to expand knowledge base
 - Reviewing, and utilizing, prior Data Team and CI tools, documentation, and workflows
 - Reviewing where gaps may be and how to address

Current MIO Data and QA/QC Activities

- **How has the role of the MIOs changed?**
 - RACI document in progress, defining responsible parties at PMO, CI, and MIOs
 - Communication is ongoing real-time between MIOs including use of Slack and Webex
 - PMO supports communication through weekly Data Management meeting to ensure uniformity and forward progress

Current MIO Data and QA/QC Activities

- How has the role of the MIOs changed?
 - Simplified Plan



Current MIO Data and QA/QC Activities

- **Additional Items being coordinated by PMO and supported by MIOs**
- **Formerly Data Team responsibilities under OOI 1.0, now coordinated by PMO in OOI 2.0**
 - Software, system, and UI testing.
 - ERDDAP development