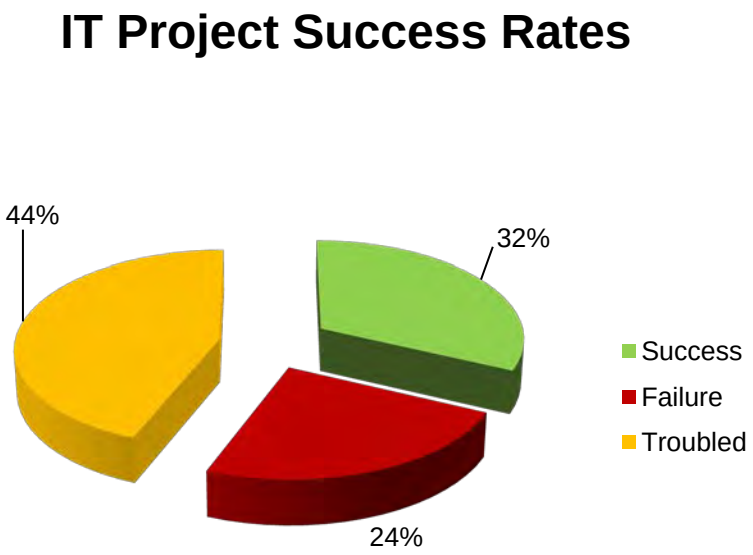
National Ecological Observatory Network

Cyber Infrastructure (CI)

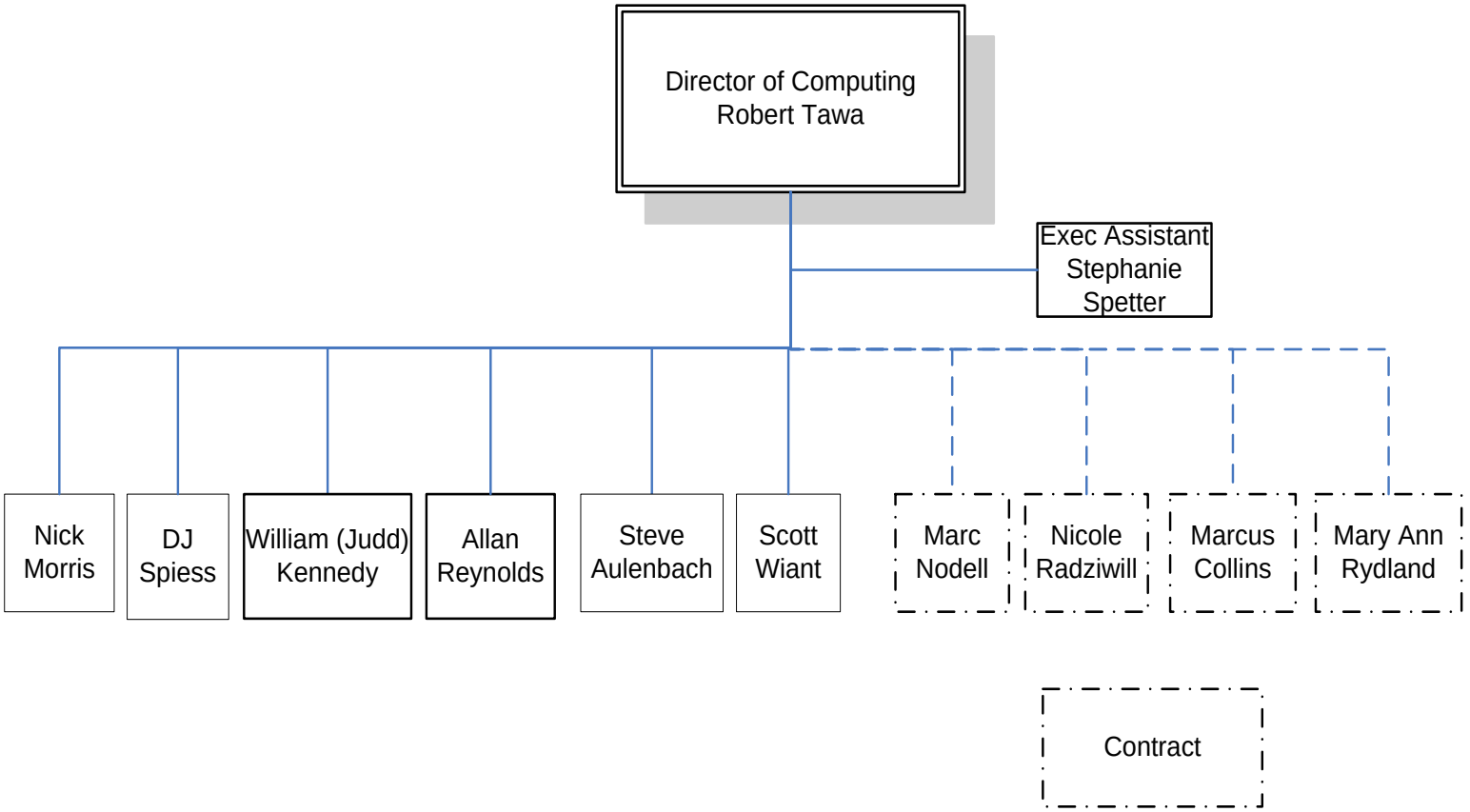
R. Tawa/NEON CI, PMCS, and Project Teams

NEON CI Mission

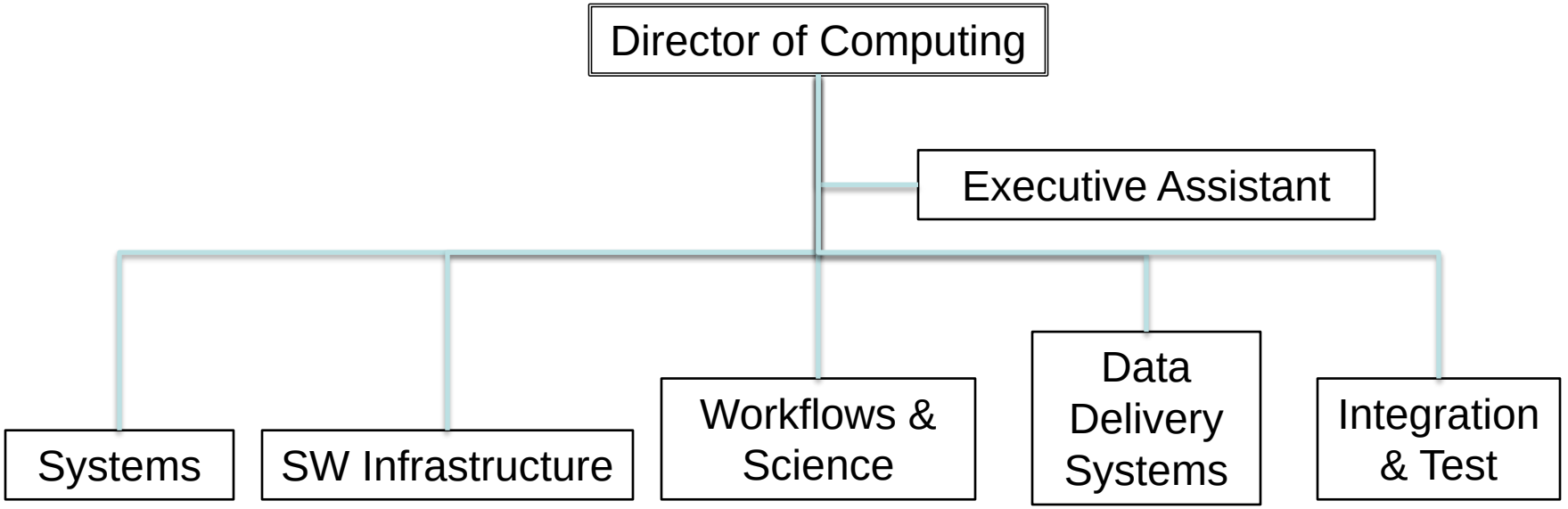
- Provide world-class cyber services in key areas
  - Ecological Data Production, Data Analysis, and Data Visualization
  - Directed and Efficient High Performance Computing
  - Technology Framework for Ecological Learning and Development
  - Virtual Meeting Space for Ecologically Minded Communities
- Promote a CI that broadens participation in all aspects of ecological science
- Provide an extensible, secure, and reliable cyber infrastructure that is sustainable over a 30 yr life



CI Organization & Staff



CI Construction Organization



- HW installation and support
- Data Center design & construction
- Network design & construction
- Computer and network security
- System Administration
- Database Administration

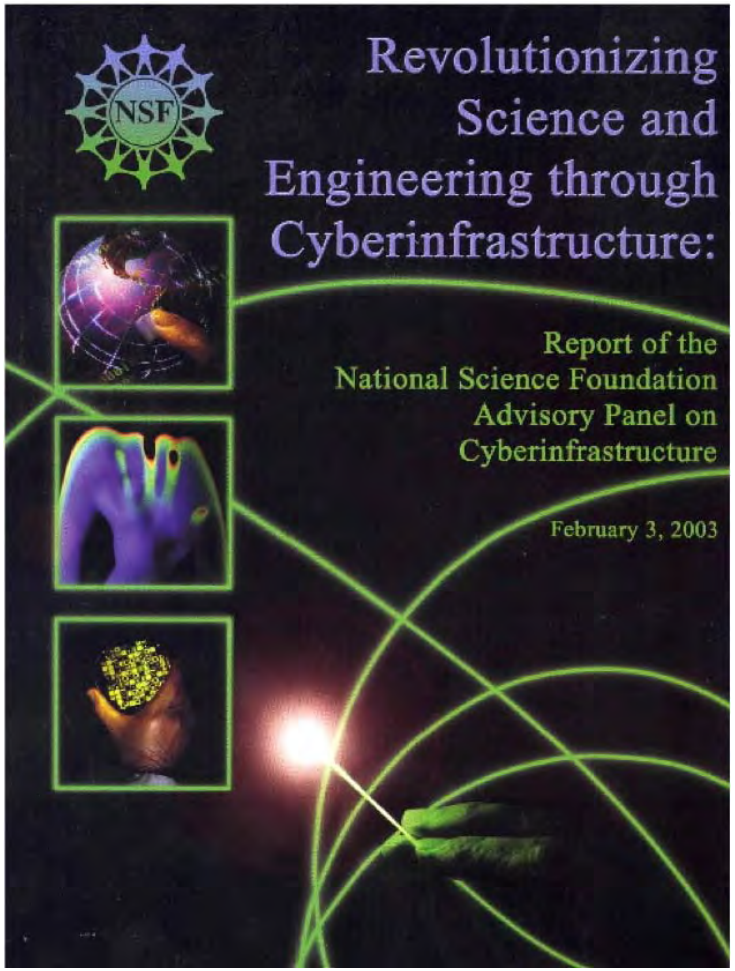
- Operations Support System Administration
- Messaging System Administration
- Ticketing System Administration
- Workflow Engine administration
- Software Configuration Management
- Database Development

- Develop data processing workflows
- Develop scientific software
- Modify Existing Scientific Software
- Develop Quality Assurance software for scientific data
- Develop Science Operations Software

- Science data Portal Software design
- Education and outreach web Portal Software design
- NEON Business web Portal Software design

- SW Unit Test development and conduct
- SW Build Integration
- Integration Test Development and Conduct
- Acceptance Test Development

Role of NEON Cyber infrastructure



NEON will address the foundations, applications, services, and social/collaboratory aspects of CI as recommended by the Atkins Report

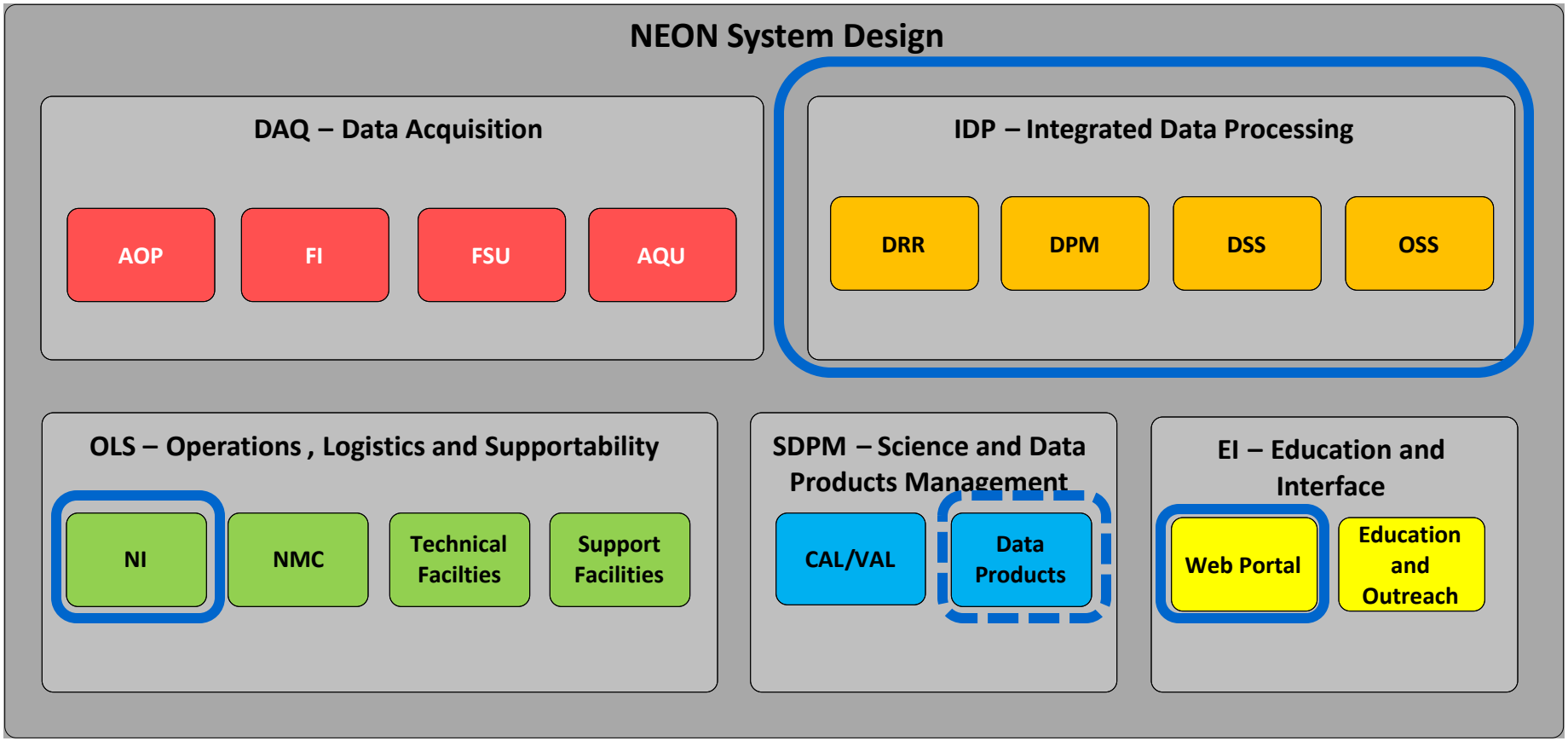
Data Manufacturing and Distribution

- Core infrastructure for data management
  - Messaging services, data access services
- Internal processes for data manufacturing
  - Tracking and managing NEON's data assets
  - Data processing & workflow development
  - Scientific data quality assurance
  - Managing FSU observations & samples
- Delivering NEON's Data Products
  - Portals for Data Discovery/Search

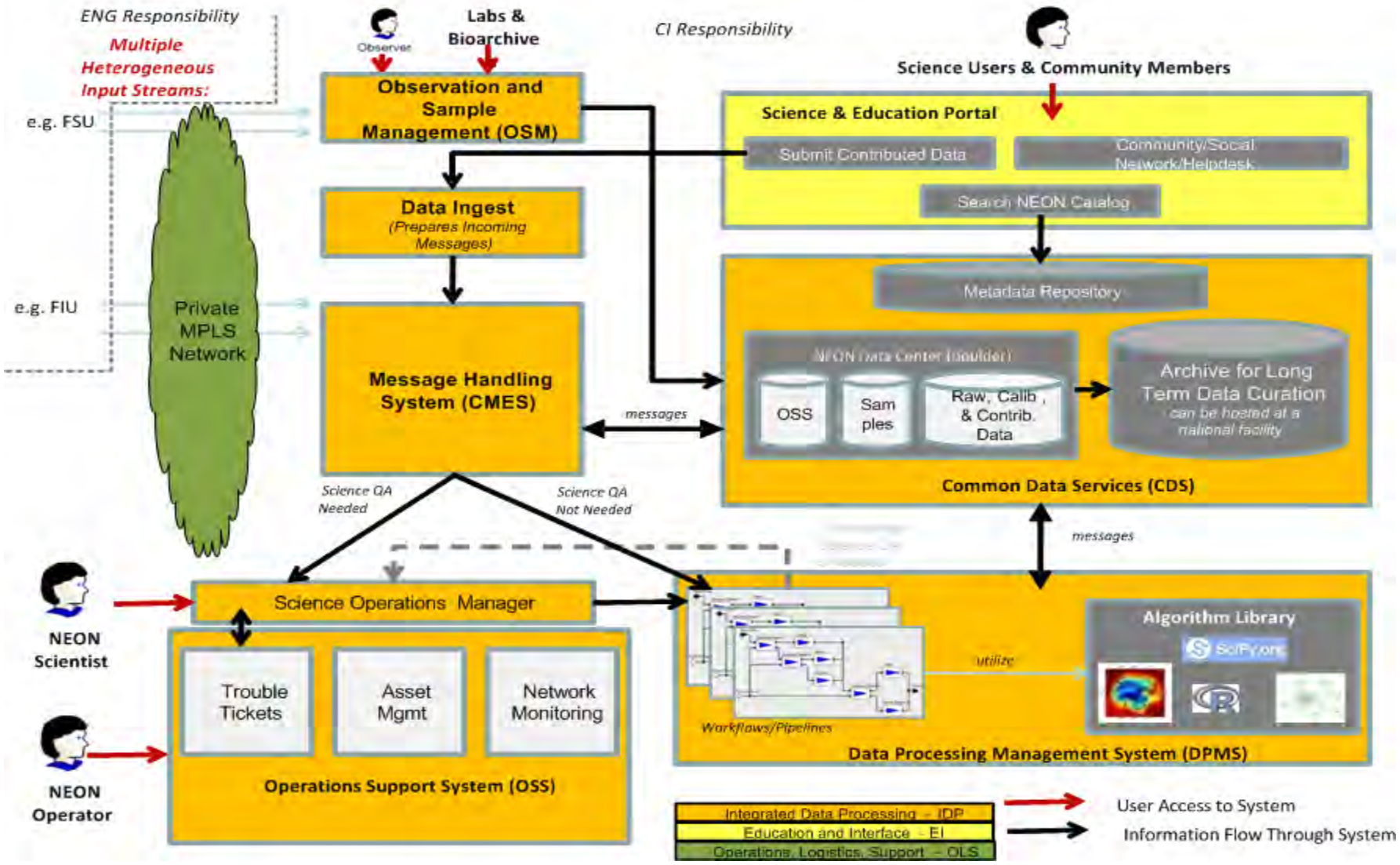
Operations and Support Systems

- Tracking and configuration management for NEON's physical and software assets
- Network Operations/Enterprise Management of Technology and Instrumentation
- Problem Tracking and Resolution

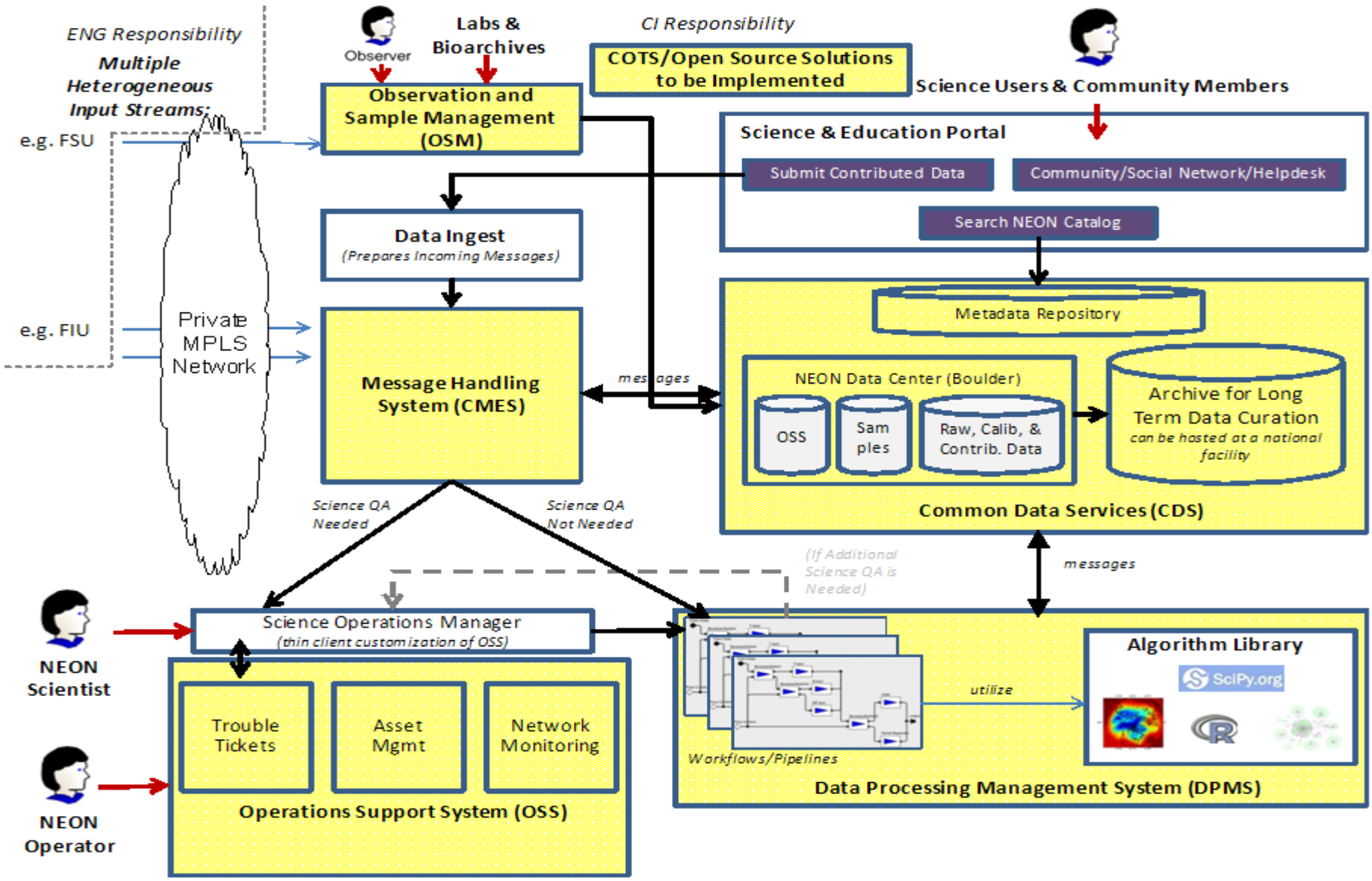
CI


NEON System Design
DAQ – Data Acquisition
AOP
FI
FSU
AQU
IDP – Integrated Data Processing
DRR
DPM
DSS
OSS
OLS – Operations , Logistics and Supportability
NI
NMC
Technical Facilties
Support Facilities
SDPM – Science and Data Products Management
CAL/VAL
Data Products
EI – Education and Interface
Web Portal
Education and Outreach

NEON Cyber Infrastructure



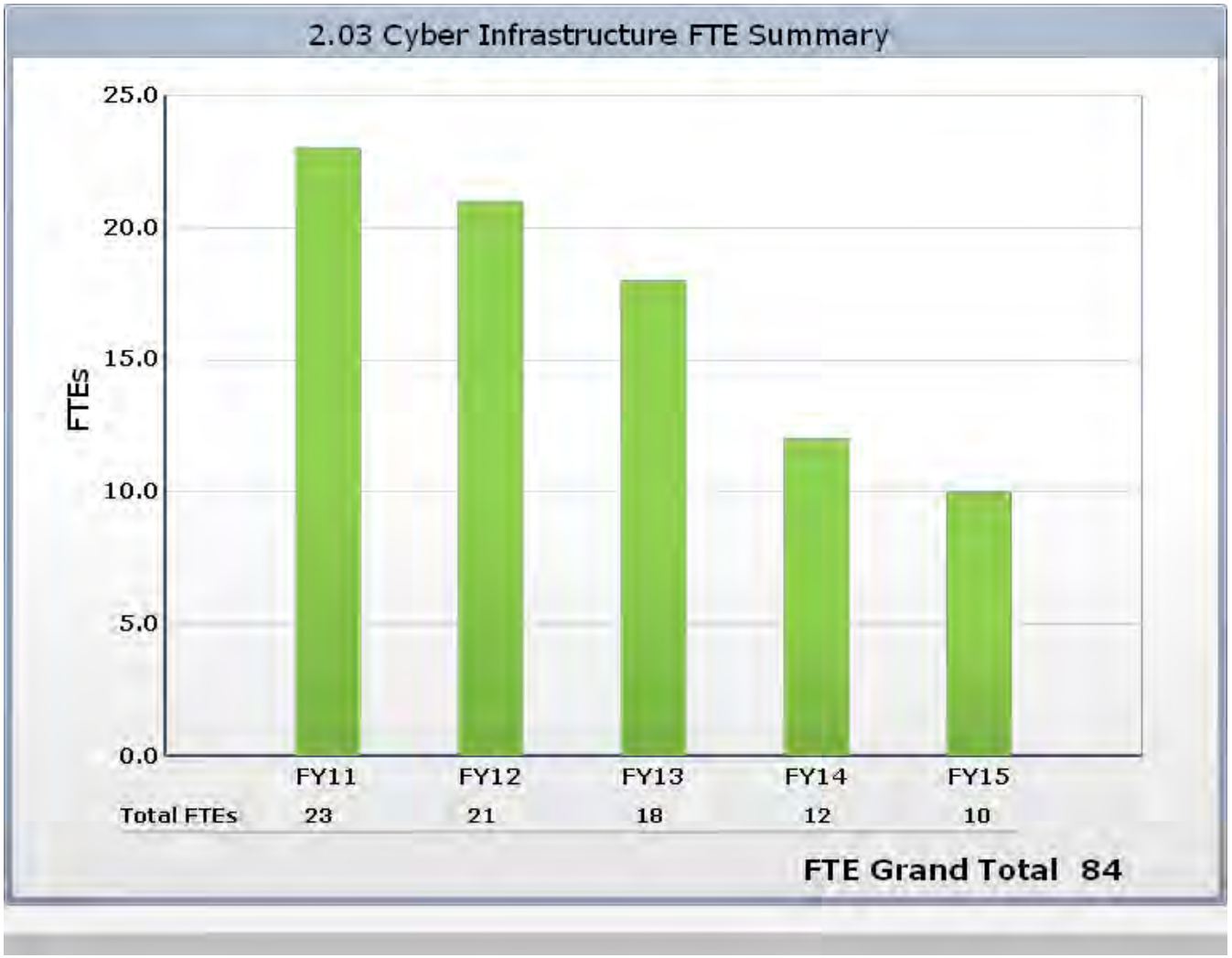
Use of COTS/Open Source Software (~70%)



2.03 Cyber Infrastructure - WBS

WBS	Title
2.03	Cyber Infrastructure
2.03.10	CI Management
2.03.20	CI Data Center Infrastructure
2.03.30	CI Software Data Products
2.03.30.10	CI Data Products 0, 1, 2 Application Implementation
2.03.30.20	CI Data Products 3, 4 Application Implementation
2.03.40	CI Domain Networks
2.03.50	CI Operational Support System
2.03.60	CI Data Processing
2.03.70	CI Data Handling
2.03.80	Domain Systems
2.03.90	Web Portal
2.03.95	CI Testing
2.03.98	CI LUAP

2.03 Cyber Infrastructure FTE Spread by FY



CI PT Risk Register

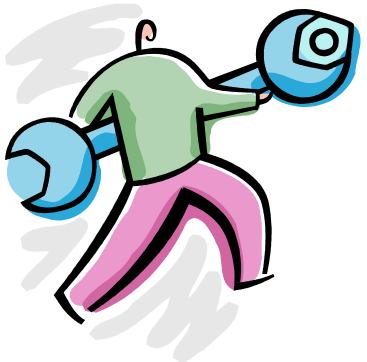
Risk ID	Risk Title	Description	RRS	Risk Exposure	Occurrence Cost	Program Area	Status
4	CI Construction Staffing	The staffing profile for the initial year of production is high with respect to the subsequent years. The risk is that NEON will not be able to hire the requisite personnel in a short time, leading to schedule slippage, excessively high workload for existing personnel, and possibly low morale and burnout as a result. Estimated cost is based on 8 months schedule slip, @ 16 FTEs @ 150K/yr + training	2.8	High	$ 2,000,000	CI	Mitigate
6	Scientific Software Integration	NEON expects to leverage existing scientific code bases for development of L3 L4 products. If the code base does not integrate efficiently within the NEON data production workflow engine in the CI architecture, then some amount of SW rework will become necessary. Estimated cost is based on 2 FTE @ 180K/yr	2.1	Medium	$ 360,000	CI	Mitigate
1	Level 4 data Product Complexity	The most complex L4 data product algorithms are not yet defined and may be much more complex than presently thought and budgeted for in development cost and schedule. This may cause the compute farm to be too small to process the product, the L4 DP may be late (into operations) or cut altogether from the catalog. Estimated cost is based on 3 FTE @ 180K/yr	1.5	Medium	$ 500,000	CI	Mitigate
7	Data Center Capability	The Computing cluster within the Data center has been parametrically estimated based on community processing benchmarks and estimated complexity of data product computations. However this is a 1st order approximation, and the actual compute load maybe bigger than estimated, thus driving up the cost and physical footprint of the CI. Cost is calculated as 1/4 of estimate data center cost or ~ 500K	1.5	Medium	$ 500,000	CI	Mitigate
10	Excessive volume of monitoring data	SNMP traps and the data from system health sensors result in an unmanageable amount of data to be reviewed and/or trouble tickets created. Estimated cost is based on 1 FTE @ 180K/yr	1.5	Medium	$ 180,000	CI	Mitigate

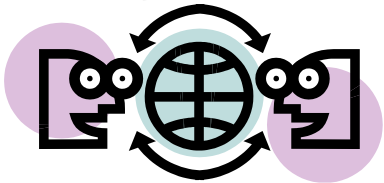
CI PT Risk Register

Risk ID	Risk Title	Description	RRS	Risk Exposure	Occurrence Cost	Program Area	Status
17	Open Source software doesn't work	Selected open source software is determined to be unfeasible, and commercial product must be selected. Estimated cost is based on cost of software licenses	1.5	Medium	$ 200,000	CI	Mitigate
18	COTS Application Integration	Integration of multiple different commercial and open source applications leads to lack of communication between systems and failure of certain CI processes. for instance, The JVM version required by one COTS or open source product differs from that required by another and is not backwards compatible. Estimated cost is based on 2 FTE @ 150K/yr	1.5	Medium	$ 300,000	CI	Mitigate
20	Reprocessing volume exceeds current compute capacity	Reprocessing due to refinement of the process or extended downtime of portions of processing infrastructure overwhelms data center compute facility. Estimated cost is based on cost of extra servers for compute farm	1.5	Medium	$ 300,000	CI	Mitigate
9	Data Center footprint may be too big for current facility	The environmental and space considerations in the current data center will be inadequate upon scaling up of compute hardware. Estimated cost of adding space/power and cooling	1.4	Medium	$ 300,000	CI	Mitigate
12	Hardware deliveries compromised	Critical hardware becomes difficult to obtain. Examples include large reduction in the supply of RAM due to a major supplier experiencing a catastrophic disaster. Estimate based on 0.25 FTE @100K + sourcing/buying of more expensive (available) HW	1.2	Medium	$ 200,000	CI	Mitigate

Progress Since PDR

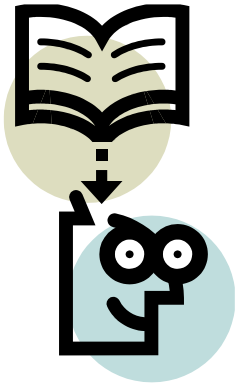
- New Hires
  - Senior Linux System Developer
  - Oracle Database Developer
- Communication
  - Met with NCEAS to discuss EML as distribution format.
  - Attended USGS Council for Data Integration meetings
  - NEON and CEN collaboration/Information Access meeting
- Community Software Research
  - Worked with Robert Clement of University of Edinburg on FIU Scientific Software

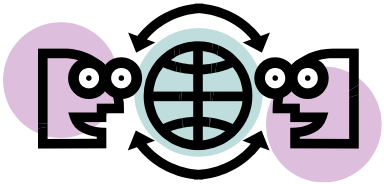




Progress Since PDR

- Documentation
  - Software Development Plan
  - Software Reuse plan
  - Metrics Plan
  - Software Configuration Management Plan
  - White paper providing overview of standards and formats for data product delivery
- Conferences
  - Cyber Security Summit 09
  - Ecological Society of America Annual Meeting
  - TeraGrid 09
  - Project Science Workshops





Progress Since PDR

- Prototyping
  - Hardware
    - C7000 Blade Chassis with 8 half height blades set up
    - Met with vendors to validate final hardware design
    - Received quotes and hardware specs design
  - Software Installed
    - Jazz
    - Liferay portal
    - LSF
    - Process Manager
    - IDL
    - Envi
    - Oracle
    - Understand Code Analysis Tool
    - R

Work Over Next 12-months

- Validate technology evaluations and mitigate risks by end to end prototyping
- Examine relationship between NEON Science & Education Portal and other Science Gateways
- Validate CI Computing Estimates based on Representative L0 to L4 Data Transformations
  - Use Table Mountain domain prototype as representative domain site
  - Use existing AOP-like data as AOP representative data
  - Generate simulated FSU data (as required)
  - Use actual LUAP datasets
- Support site characterization activities

CI Summary

- CI Team ready and excited to continue prototyping effort
  - Proof of architectural concepts
  - Proof of SW evaluations
  - Risk Reduction
    - Head start on Integration of COTS SW
    - Head start on Data Product generation
  - Contextual Scientific Operations Requirements Elicitation and Modeling
  - Portal Models and Mock-ups
- CI Expects to be well positioned to start construction

The National Ecological Observatory Network is a project sponsored by the National Science Foundation and managed under cooperative agreement by NEON Inc.