OOI CYBERINFRASTRUCTURE
NSF OCEAN OBSERVATORIES INITIATIVE

OOI – CyberInfrastructure

Architecture & Design

Operational Node Connectivity Description

OV-2 PDR CANDIDATE

November 2007

Document owner: OOI CI System Architecture Team

Document History

Date	Version	By	Description of Changes
2006.03.09	0.1	John Graybeal	Template only
2006.05.28	1.1	Alan Chave	Initial draft
2006.05.30	1.2	Alan Chave	Initial release
2006.07.26	1.3	Alan Chave	Final release
2007-08-03	2.0	CI ADT	Conceptual Architecture Update Initial Draft. Reframed, replaced ORION with OOI
2007-10-04	2.1	CI ADT	Updated models; shows work in progress
2007-11-12	2.2	CI ADT	Updated models, adapted text, consolidated with OV-7

Preamble

The set of documents named AV*, OV*, SV*, TV* are all part of the OOI CyberInfrastructure Architecture & Design (CIAD), in the structure prescribed by the DoDAF (Department of Defense Architecture Framework). Each document has a designated title, an identifier (such as AV-1) and covers a specific topic in a self-contained way. Document AV-1 provides further explanations and a summary. A glossary of the terms used in these documents and their context can be found in AV-2.

The figure below suggests an intuitive reading flow through the provided documents. Other documents will be added to the figure as they emerge during the design of the CI (for the complete set of documents see AV-1). The thick arrow suggests a reading order through the core documents (AV-1, OV-1, OV-5 and SV-1). The red rectangle highlights the current document.

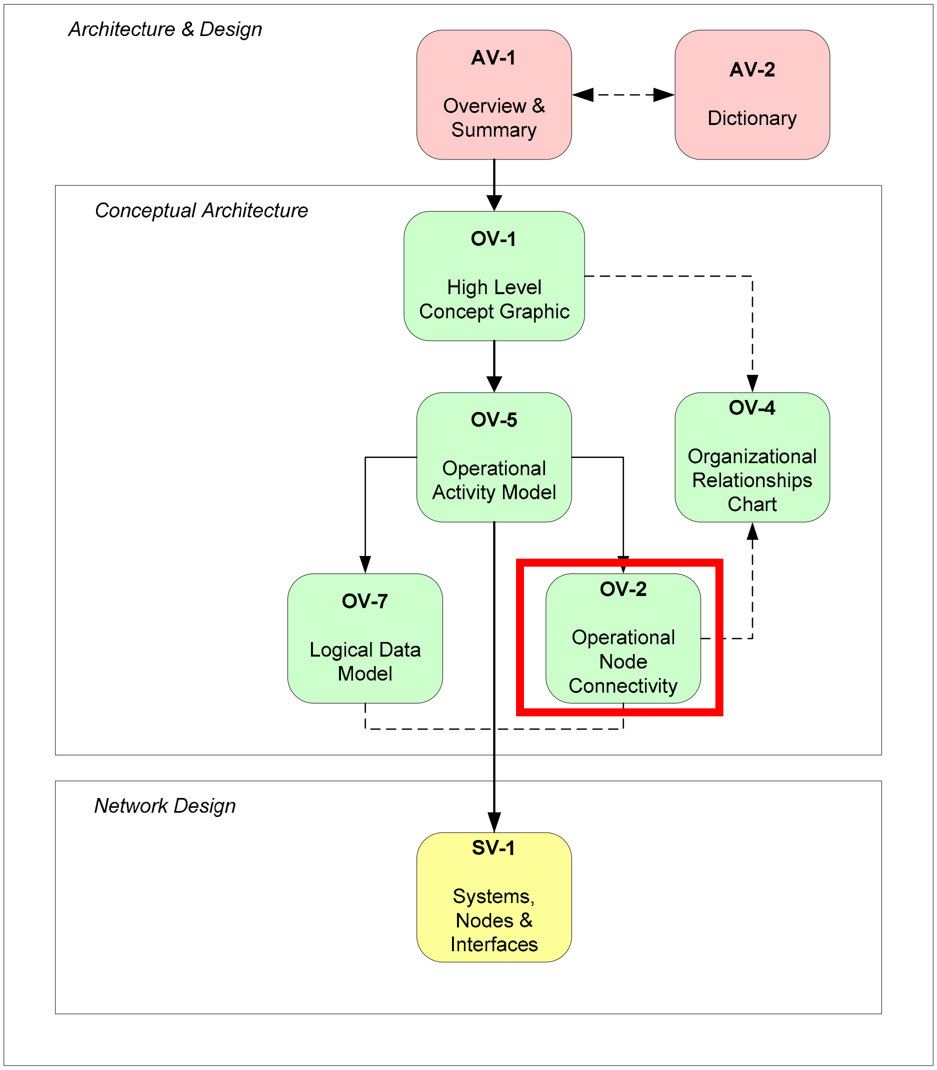

Table of Contents

OOI - CyberInfrastructure

Architecture & Design
Operational Node Connectivity Description (OV-2)

1 Introduction

1.1 Product Overview

The Operational Node Connectivity Description graphically depicts the operational nodes (or organizations) with needlines that indicate the requirement to exchange information between any pair of nodes. The graphic includes operational nodes internal to the architecture as well as those external to it. The OV-2 product may be modeled using collaboration diagrams, where actors (instances of those in OV-5's use case diagrams) represent the roles or organizations that communicate via UML links and UML messages (needlines and information exchanges). Operational nodes also correlate to classes on the corresponding OV-4 UML class diagram (adapted from [DoDAF-vII 2007]).

1.2 Product Purpose and Description

OV-2 is intended to track the need to exchange information between specific operational nodes that play key roles in the architecture and other nodes. OV-2 does not depict the connectivity between the nodes.

The main features of this product are the operational nodes and the needlines between them that indicate a requirement to exchange information. An operational node is an element of the operational architecture that produces, consumes, or processes information. What constitutes an operational node can vary among architectures, and includes, but is not necessarily limited to, representation of an operational/human role, an organization or organization type.

A needline documents the requirement to exchange information between nodes. The needline does not indicate how the information transfer is implemented. Needlines are represented by arrows (indicating the direction of information flow) and are annotated by a phrase that is descriptive of the principal types of information exchanged. It is important to note that the arrows on the diagram represent needlines only. There is a one-to-many relationship from needlines to information exchanges (adapted from [DoDAF-vII 2007]).

2 Operational Nodes

An overview of the OOI is depicted in Figure 1, showing the two observatory instances, potential other observatories, laboratories, classrooms, science portal and applications all connected to the CI core. The basic concepts of this overview exhibit are represented in Figure 2. The CI operational node represents all the functionalities within the cyber-infrastructure. The key operational nodes in the CyberInfrastructure are the five resources services networks: *Data, Modeling, Processing, Control* and *Instrument*, as well as *the Common Operating Infrastructure* and the *Common Execution Infrastructure.*

2.1 Resources and Resource Services Networks

A key purpose of the OOI CyberInfrastructure is provisioning of a distributed and federated capability to describe and utilize *resources* and *activities* or *processes* operating on and with resources. As defined in the *dictionary* (AV-2), the term *resource* encompasses representations of physical entities that can be utilized via the cyber-infrastructure; examples include instruments, networks, registries, repositories, or data. Any function associated with a resource is presented as a *service*, and is itself a resource. A *process definition* is a task or grouping of tasks that provides a specialized capability. A *process instance* is a running instance of a process. Multiple *process instances* can be interconnected by an *observation plan* to provide higher-complexity capabilities to the system. Since many resources will be shared and hence are subject to use conflicts, a mechanism that enforces *policy* to arbitrate rights and allocations is essential.

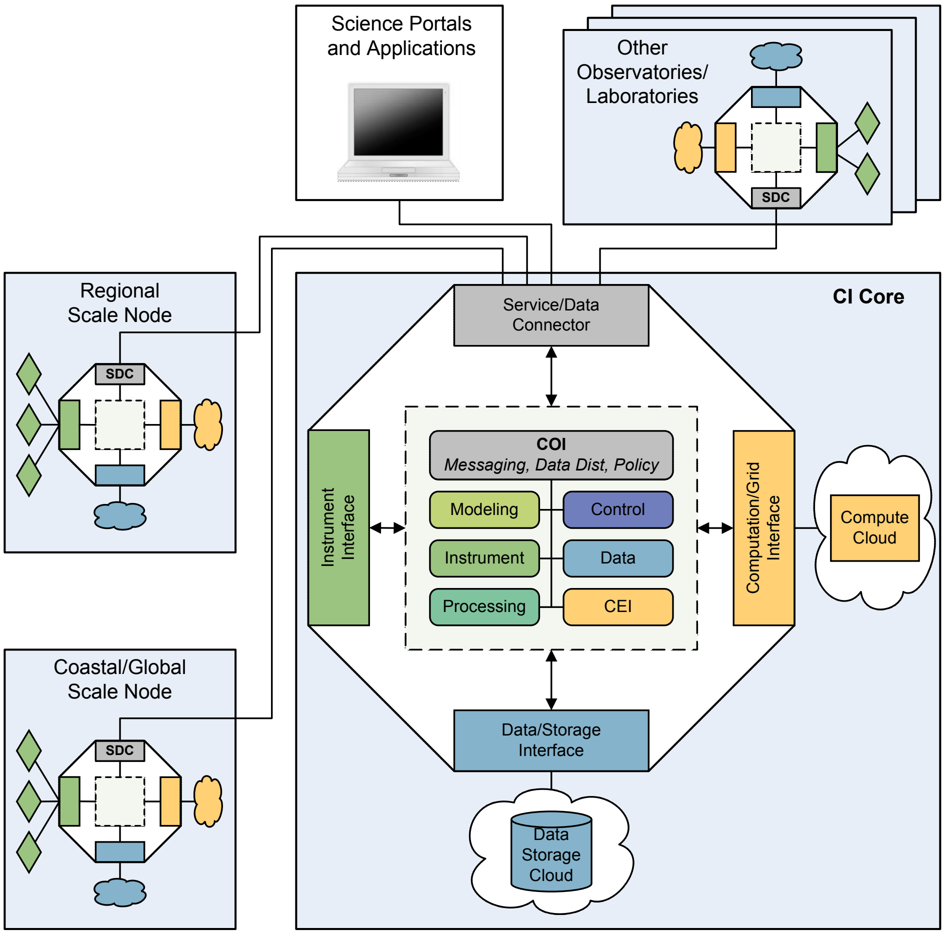

Figure 1. OOI High-Level Overview

These four elements – resources, processes, observation plans, and policy – operate at many levels with complex linkages throughout OOI, and beyond to non-OOI systems. Further, they are dynamic and multivariate; it is possible for a policy to change as resource utilization evolves. It is also possible for a single instance of a resource (e.g., a data stream) to be used by one or more processes (e.g., different numerical models) to produce multiple new resources (e.g., the data products produced by each model).

These concepts and their complex interactions are illustrated in OV-2 by a top-level CyberInfrastructure Model containing five resource services networks: *Instrument, Control, Processing, Data,* and *Modeling*, which will be presented, in detail, in the next sections. The *Instrument Services Network* describes the interplay of sensors and actuators with observatory physical nodes and the other networks. The *Control Services Network* contains the key elements required to couple observations with hypothesis testing and coordinates new observation missions, thus it describes the interactions needed to define and organize an observation program. The *Processing Services Network* refines an observation plan and submits individual process definitions for execution to the *CEI*, which manages and allocates distributed computational resources throughout an ocean observatory. The *Data Services Network* links the other services networks, archiving not only data and their resultant products, but also the experience and understanding gained through use of an observatory by means of specific science ontologies. The *Modeling Services*

Network provides the visualization of data and the validation of observations and the facilities to define new observations, in addition to Modeling Processes.

The CyberInfrastructure Model defines core CyberInfrastructure services required to implement its functionality.

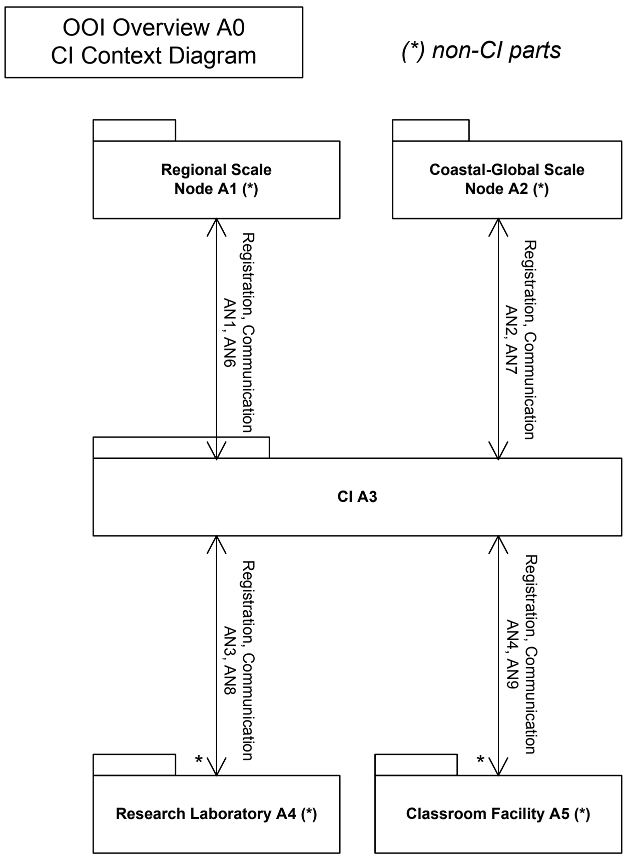

Figure 2. OOI Overview CI Context Diagram

2.2 Overview of OOI

Figure 2 presents an overview of OOI systems. *Regional, Coastal-Global Scale Nodes, Laboratories,* and *Classrooms* – instances of external *Facilities* – are connected to the CyberInfrastructure, which provides or integrates services for performing oceanographic observations.

A *Scale Node* operates a network of instruments and plays an oversight role through connection of resources with a wide range of management and control activities. A *Laboratory* establishes a virtual relationship between a group of Investigator actors and all of the resources and activities required to accomplish their set of scientific goals. The *Classroom* achieves educational goals by joining resources and teaching activities.

2.3 CyberInfrastructure

Figure 3 shows a high-level overview of the CI operation node. It displays it as a concept graphics with its interfaces to the environment and its internal structuring in services networks. As such, it can be represented not only in the CI core instances of the OOI but also inside observatories, laboratories, classrooms etc.

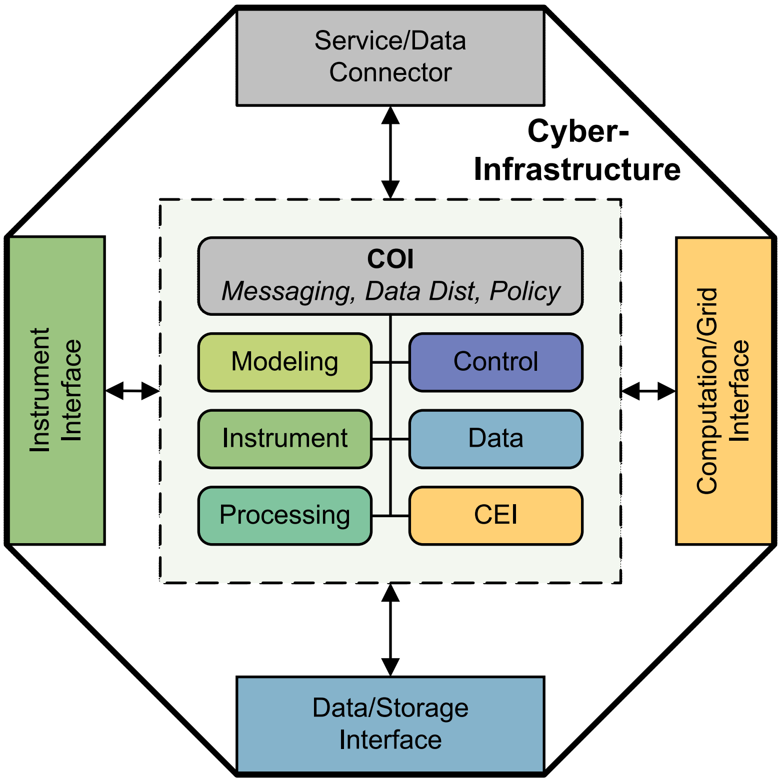

Figure 3. CyberInfrastructure concept graphics

Figure 4 shows the operational nodes and needlines for the CyberInfrastructure (CI). First, we present an overview of the services networks; then, we expand the operational nodes in this model in subsequent sections.

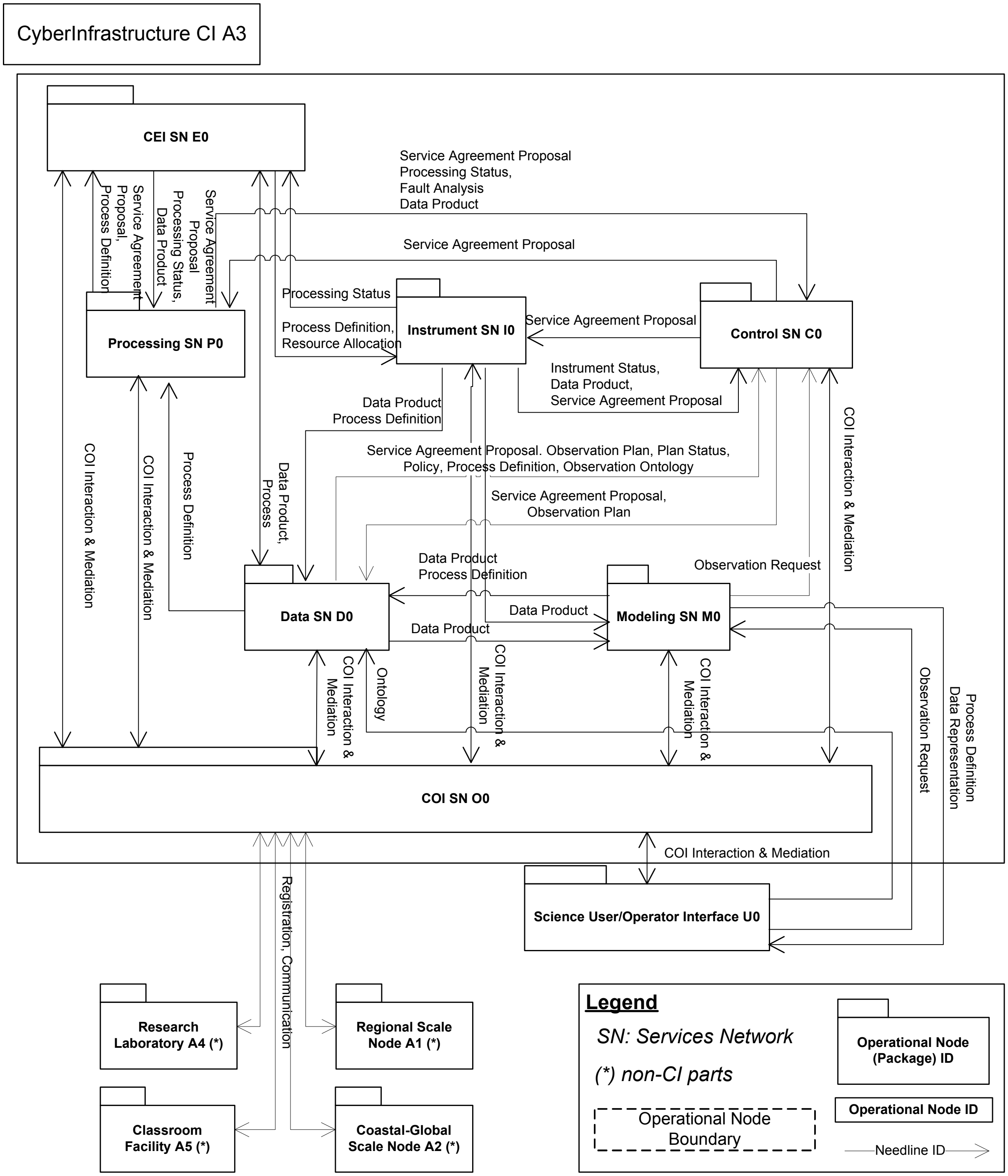

Figure 4. OOI CyberInfrastructure Model

2.3.1 Operational Node Overview

The CI is broken down into five resource services networks that contain different classes of resources and the capabilities needed to manage them. The *Instrument Services Network* is focused on the collection of observations. The *Data* and *Modeling Services Networks* are centered on the utilization of observations.

The *Control* and *Processing Services Networks* manage the observation plans and the allocation of resources. These resources services networks are tied together by crosscutting infrastructure elements that serve as the integration platform and communication conduit, and support the activity, resource, service, identity, communication, and presentation models that must be used across OOI.

The *Instrument Services Network* provides interactive and coordinated relations with real and/or synthetic environments through the use of transducers (sensors or actuators). It ensures the safe and secure operation of individual sensing platforms, and provides reliable delivery of acquired data with their associated metadata. These capabilities must be integrated with network-wide resource allocation and observation planning. The *Instrument Services Network* provides the command and control semantics for interacting with an Instrument resource. The *Instrument Services Network* node provides data from sensors, instrument status information, and requests for resources such as bandwidth or power. It receives commands to instruments and allocations of resources from the *Control Services Network* and the *CEI*.

The *Data Services Network* provides an automated data distribution and preservation network with pervasive and universal access subject to OOI Data Policy. It provisions a federated system of data streams, repositories, and catalogs that supports the distributed organization of resources.

The *Modeling Services Network* establishes baseline processes and tools comprising a coherent framework for the analysis and assimilation of data. It provides the command and control semantics for interacting with a Modeling Services resource. The capability to network and interact with multiple community-based numerical ocean models for parameter estimation/optimization and data assimilation is integrated into the framework. The network provides a wide range of event detection/response, quality control, and modeling capabilities linked together via the *Data Services Network*. It provides observation requests to the *Control Services Network* and receives observational data from the *Data Services Network*. It also interacts with human actors through analysis and visualization activities.

The *Control Services Network* establishes standard models for the management of computation resources. It provides the semantics to monitor and control the operating state of a computation. The *Control Services Network* configures resources to accomplish a task, usually within the context of an ongoing laboratory or classroom activity. It receives observation requests from the *Modeling Services Network*, and provides observation plans to the *Processing*, *Data* and *Instrument Services Networks*, a well as the *COI*. It receives Data and status information from these services networks. It also provides command and control instructions to the *Instrument Services Network*.

The *Processing Services Network* provides immediate-mode scheduling of processes at specified locations within the integrated network based on explicit time requirements and/or event triggers, during a service agreement protocol with the *CEI*. *CEI* provides process execution planning, and it couples processes to the streaming environment of the *Data Services Network*.

The *Scale Node*, *Laboratory*, and *Classroom* are instances (which are instantiated dynamically and may operate simultaneously) of the external *Facility Model* that utilize different classes of resources subject to distinct policy constraints to accomplish diverse goals.

2.3.2 Common Operating Infrastructure

A broad range of common services is required to bind the *OOI* into a coherent whole. *Common Operating Infrastructure (COI)* services integrate the resource services networks, enabling data to be published and consumed by all subsystems, allowing subsystem services to be composed in complex interactions, and implementing crosscutting aspects such as governance and security. Interactions define capabilities that modify and manage processes and link resources and users to accomplish a goal. Governance and security provide the capability to link resources within a policy management framework. Figure 5 depicts the *COI* architecture and services. A message-based communication infrastructure manages the service orchestration via two main layers. The *Messenger* layer is responsible for transmitting messages between services. The *Router/Interceptor* layer is responsible for intercepting messages placed on the Messenger and then routing them among all services involved in providing a particular capability. This allows the injection of policies governing the integration of a set of services. Thus, infrastructure services can modify interac-

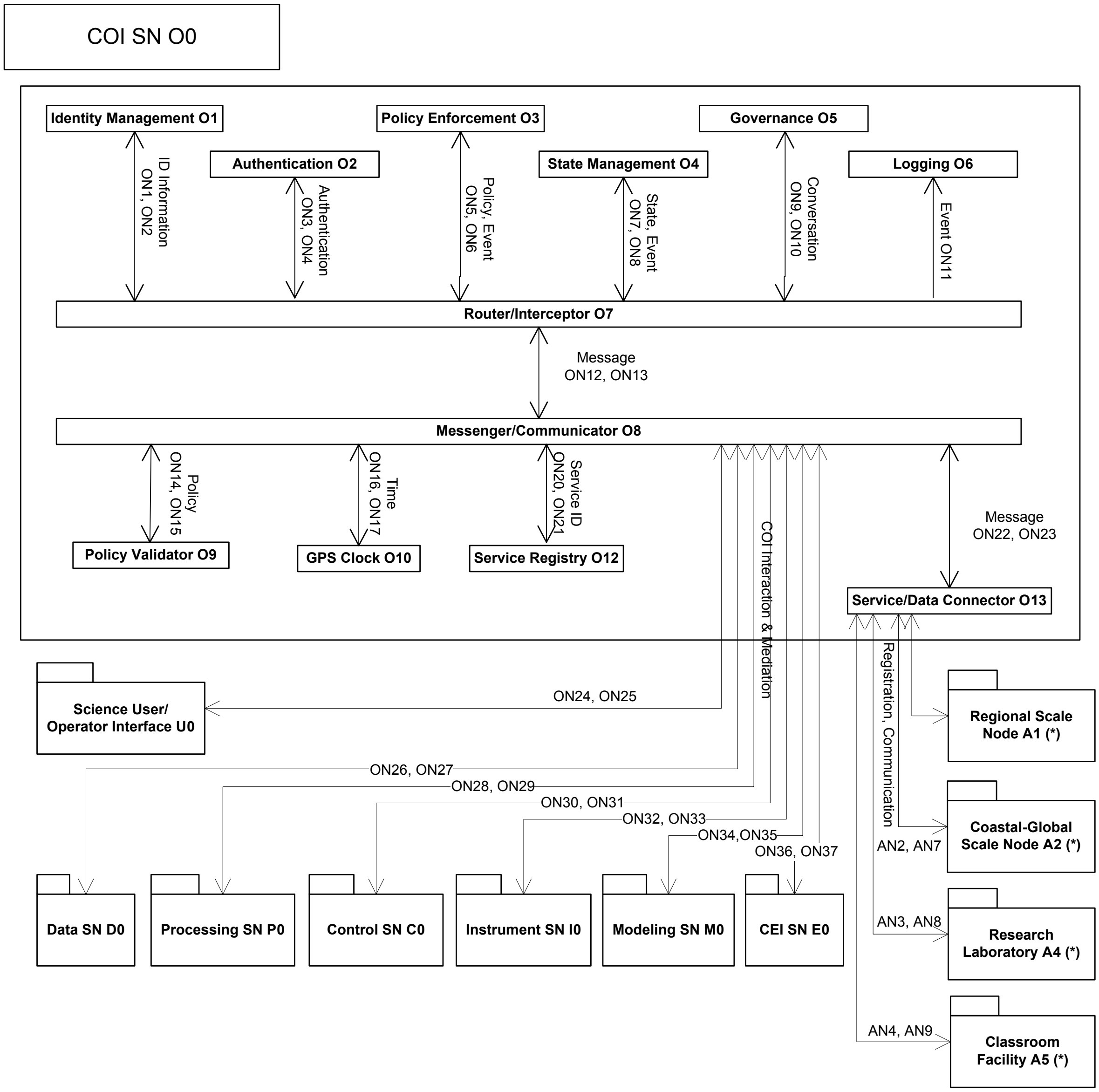

Figure 5. COI Services Network operational node and needlines

tions by rerouting, filtering, or modifying in other ways the messages exchanged. Infrastructure services include *Identity Management, Policy Enforcement, Authentication, Logging, Governance* and *State Management*. *Governance* defines the policy management framework that is implemented throughout the cyber-infrastructure. The *Service Registry* stores resources and associates them with their descriptions and relations with other resources. The *Policy* V*alidator* is responsible for processing new policies upon submission by human operators prior to storing and enforcement. The *GPS Clock* service provides a global timing capability. Resource Services Networks and other internal operational nodes of CyberInfrastructure communicate with COI directly via the *Messenger* layer. The operational node *Service/Data Connector* is the only way for operational nodes external to CyberInfrastructure to communicate with *COI.*

2.3.3 Instrument Services Network

The *Instrument Services Network* performs instrument management, mission execution, and other data acquisition tasks, and may include diverse sensor, actuator and mobile platform entities at multiple, dispersed physical locations. Figure 6 illustrates the operational nodes and needlines that are part of the *Instrument Services Network* operational node.

The *Instrument* device model consists of one or more physical sensors or actuators and a logical device that is typically implemented in firmware or software. The physical device provides sensor data and status information to the logical device (the *Instrument Proxy*), and receives configuration information and commands from the *Instrument Proxy*. The *Instrument Proxy* uses the *Data Acquisition* to provide *Engineering Data* (primarily power, bandwidth, and instrument status information) to the *Instrument Planner*, and *Science Data* to the filters that ultimately provide *Data Products* for the *Control* and *Data Services Networks*. The *Instrument Proxy* receives resource allocations and commands and possibly process definitions to define or modify instrument configuration and functionality from the *Instrument Planner*.

The *Instrument Planner* negotiates with the *Control Services Network* through the service agreement proposal protocol in order to agree on the resource allocations and partial plans. It also sends instrument status information to the *Control Services Network* for monitoring and higher level decision making and planning. The *Instrument Modeler* node provides information on instrument behaviors furnished by the Provider actors in facilities. It provides instrument behavior models to the *Instrument Process Repository* from the *Data Services Network*.

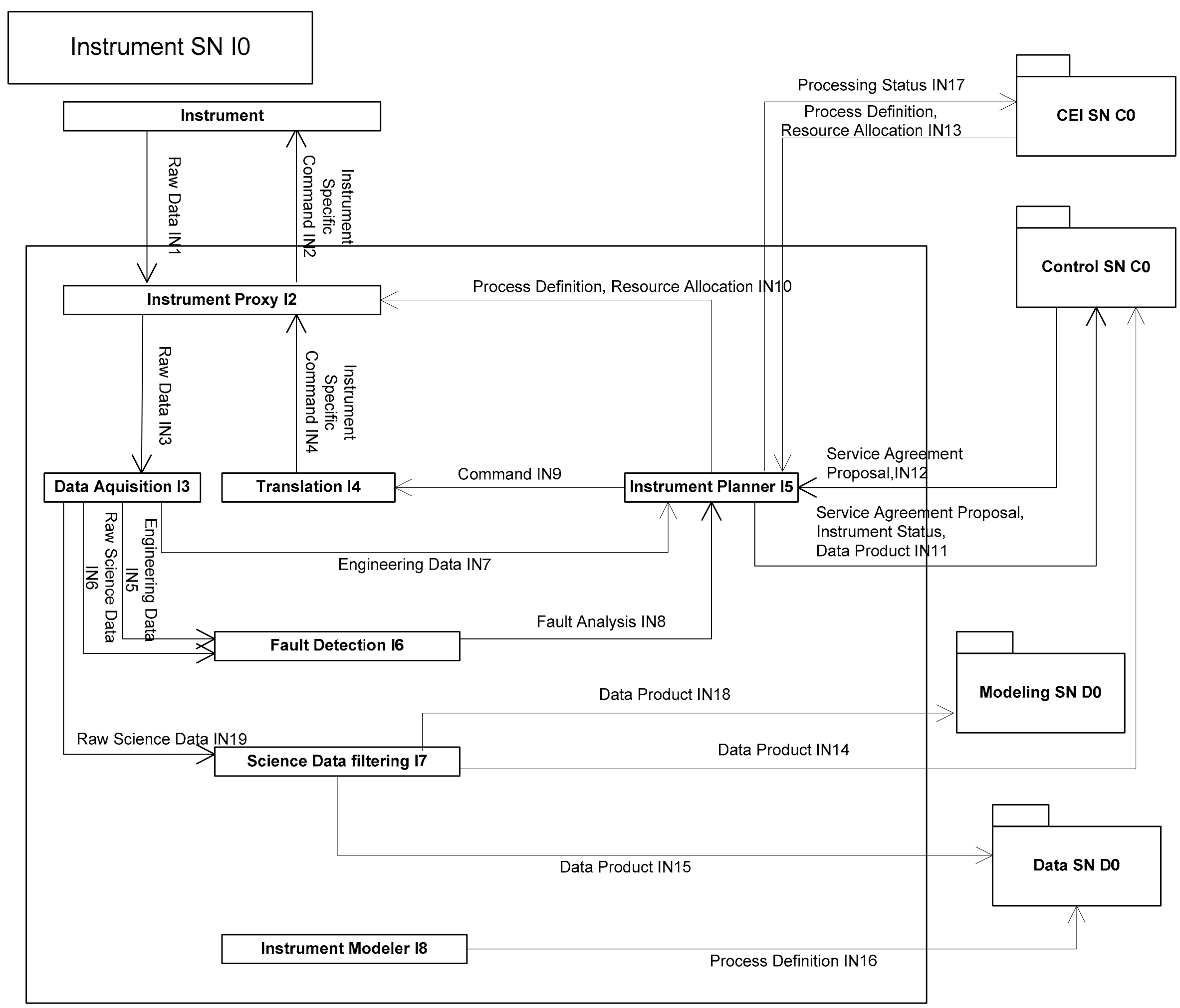

Figure 6. Instrument Services Network operational nodes and needlines

2.3.4 Control Services Network

Figure 7 illustrates the operational nodes and needlines that are part of the *Control Services Network* operational node. The *Resource Planner* provides diverse resource management and observation tasks. It receives observation requests from the *Modeling Services Network* and develops an observation plan; it

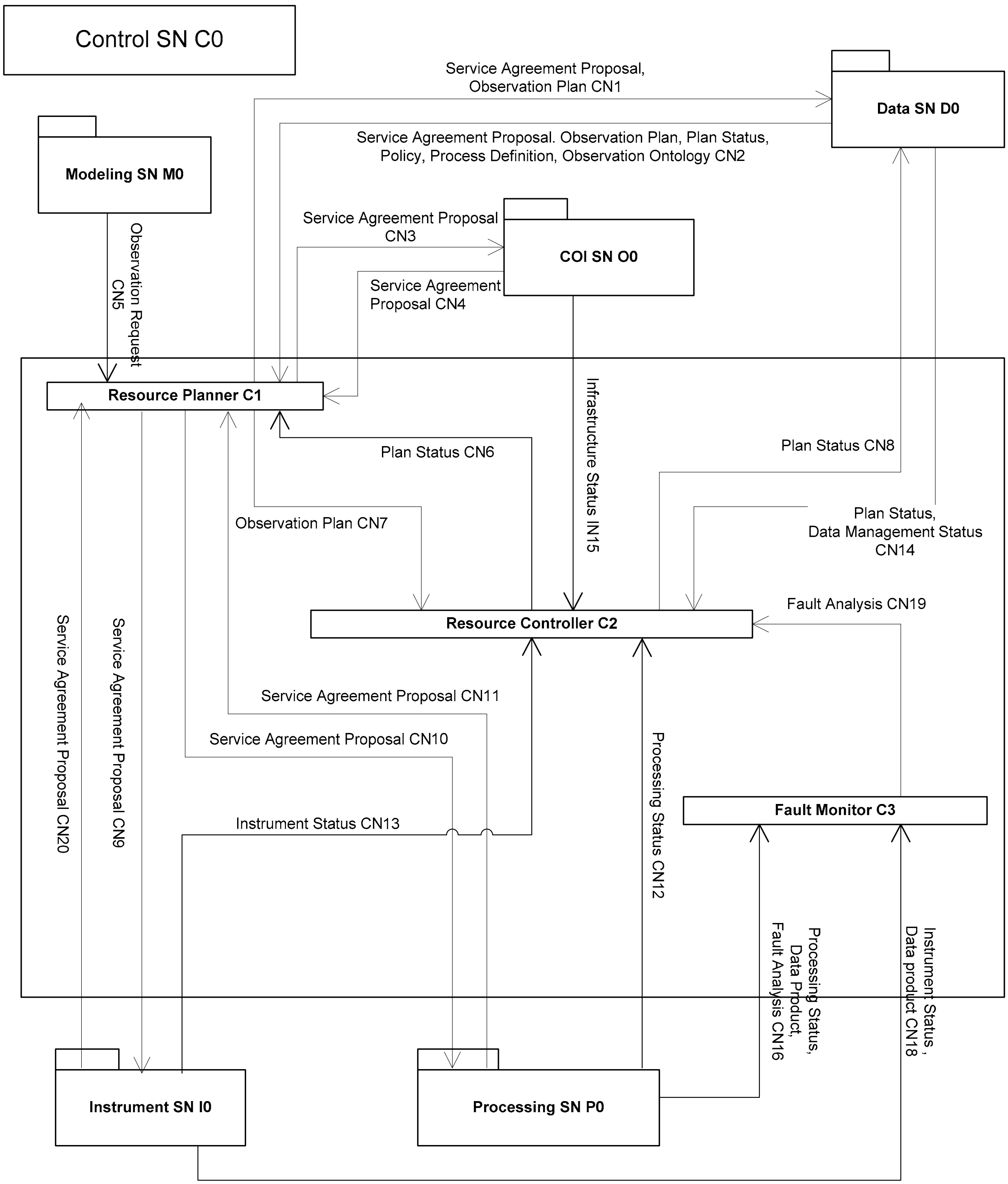

Figure 7. Control Service Network operational nodes and needlines

communicates with the *Data*, *Instrument*, *COI*, and *Process Services Networks* to reach an agreement for resource allocations and specifics of the observation plan. The allocation of resources is arranged by exchanging service agreement proposals with the *Resource Planner* until an agreement is reached. OOI policies dictate the protocol employed to reach such an agreement.

The *Resource Controller* is responsible for monitoring the execution of the final plan. The *Resource Controller* receives *Data Products* that are generated and status information from the respective Services Networks for monitoring and reconfiguration purposes. The observation plan and the intermediate status information are stored via the *Data Services Network* for possible later retrievals.

2.3.5 Processing Services Network

Figure 8 illustrates the operational nodes and needlines that are part of the *Process Services Network* operational node. The *Computation Scheduler* negotiates with the *Control Services Network* through the Service Agreement Proposal protocol in order to agree on resource allocations and partial plans. It refines and manipulates the agreed upon plan and provides a more detailed *Processing Plan* at the level of granularity that can be executed on the CEI. The Computation Scheduler interacts with the *Data Services Network* to store and retrieve instrument and model process instance information.

The *Process Controller* communicates with the *CEI Services Network* to reach an agreement for resource allocation, to execute the processing plan. The allocation of resources is arranged by exchanging service agreement proposals with *Process Controller* nodes until an agreement is reached. OOI policies dictate the protocol employed to reach such agreement. The *Process Controller* then submits *Process*

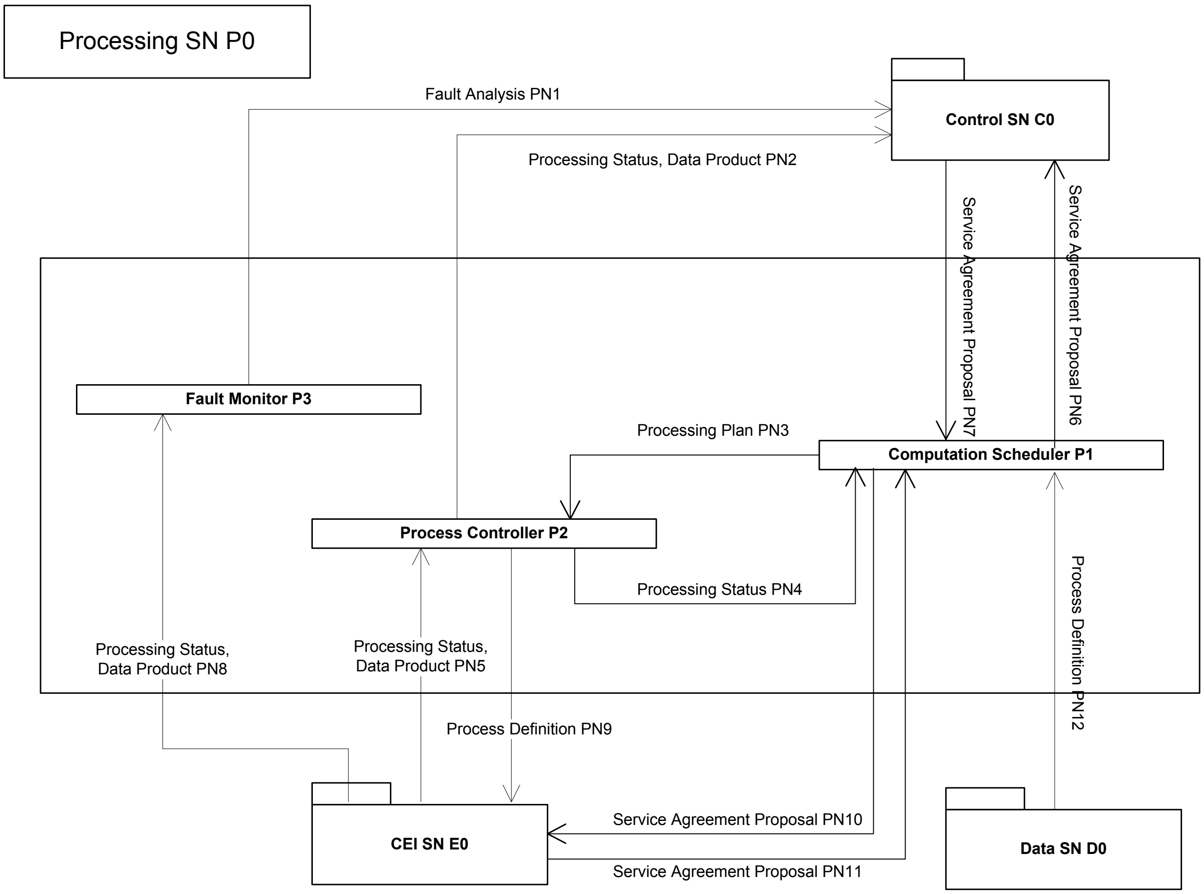

Figure 8. Processing Services Network operational nodes and needlines

Definitions to the *CEI* for execution. It receives processing status from the *CEI Services Network* and forwards a subset of it to the *Control Services Network*.

2.3.6 Common Execution Infrastructure Services Network

Figure 9 illustrates the operational nodes and needlines that are part of the *CEI Services Network* operational node. The *Dispatcher* receives *Process Definitions* from the *Processing Services Network*. It dispatches *Process Definitions* among various *Computation Nodes*. Each *Computation Node* aggregates a number of *Execution Engines*. *Process Definitions* along with resource allocation information are sent to each *execution engine* that is responsible for executing them. The *Execution Engine* performs process executions and interacts with the *Data Services Network* to store and retrieve data products. It provides data products and process status information to the *Processing Services Network* for monitoring.

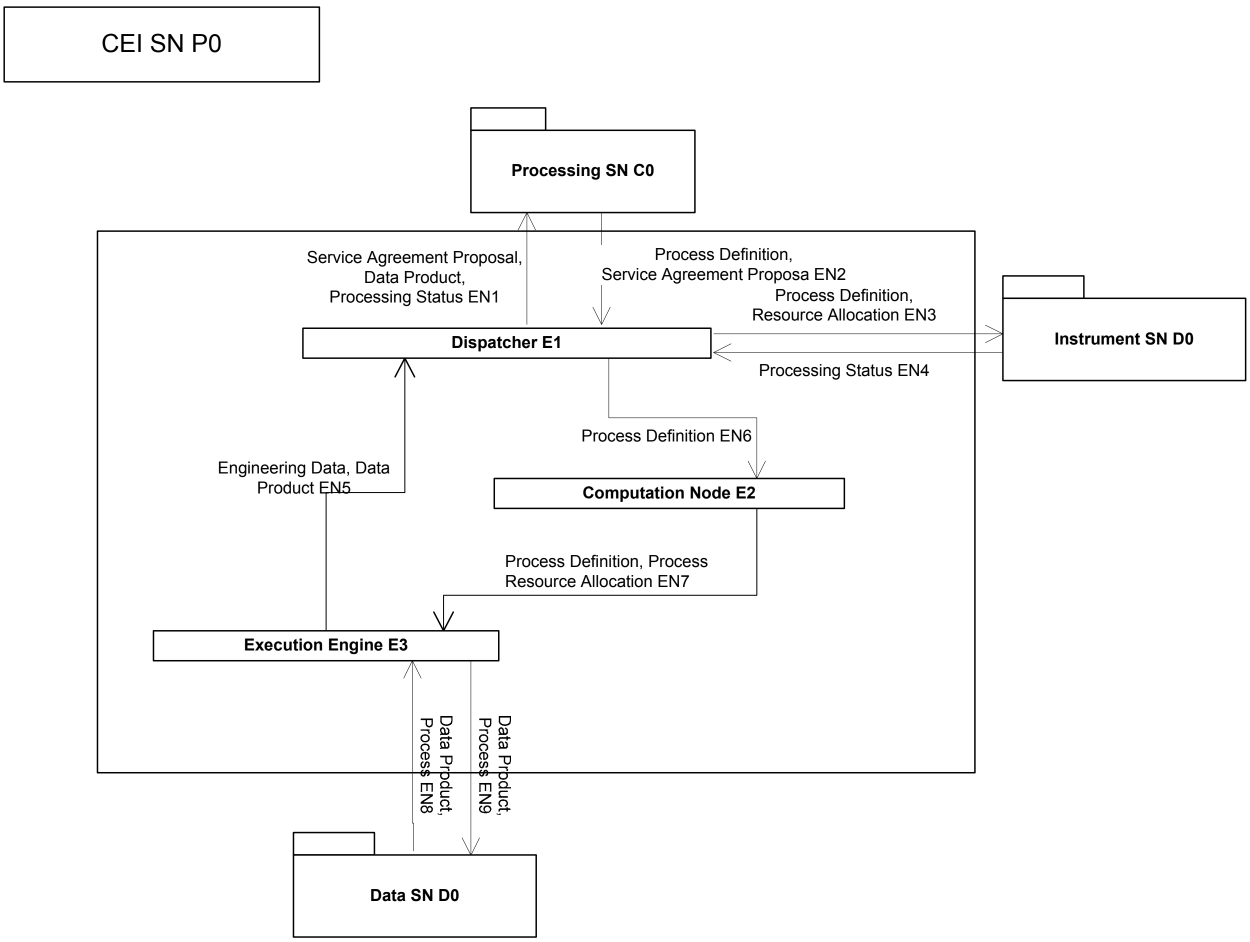

Figure 9. CEI Services Network operational nodes and needlines

2.3.7 Modeling Services Network

Figure 10 illustrates the operational nodes and needlines that are part of the *Modeling Services Network* operational node. The *Modeling Services Network* provides the core services needed for the analysis of observations and their synthesis into conclusions and testable hypotheses. It interfaces to science users and operators.

The science users and operators provide input, experience and oversight for analysis and synthesis activities using the *Modeling Services Network*. The users can submit new observation requests and new process definitions into the system and view various data representations.

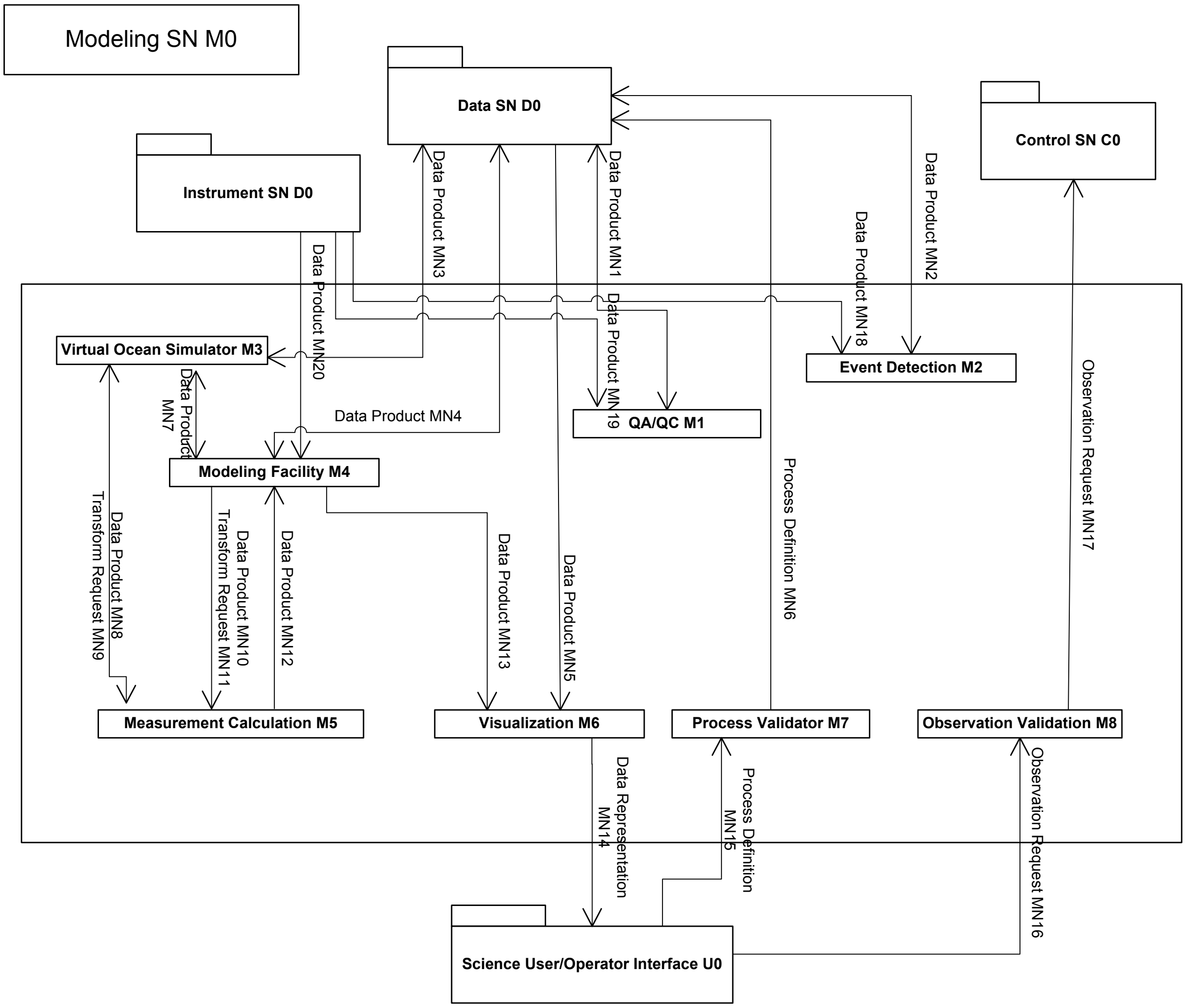

Figure 10. Modeling Services Network operational nodes and needlines

The *Quality Assurance/Quality Control (QA/QC), Modeling Facility* and *Event Detection* represent archetypical activities carried out in an analysis and synthesis effort. In this context, the nodes are activity centers that may be either ongoing or one-time in form. The *QA/QC* node performs automated data quality control. The *QA/QC* node functions as a filter that receives observations from the *Instrument Services Network* or retrospective data contained in repositories as improved *QA/QC* algorithms are devised.

The *Modeling Facility* node hosts prognostic and retrospective numerical models of observed processes and events, usually involving assimilation of real-time or retrospective data from the *Instrument* operational node or repositories, respectively. It receives specifications to define numerical experiments. It may also receive real-time data that has not been subjected to *QA/QC* directly from the *Instrument* node. The *Modeling Facility* publishes model products to the *Data Services Network* as well as to the *Visualization* node. In terms of process, the *Modeling Facility* only slightly differs from *QA/QC* or *Event Detection*. It differs in the magnitude and diversity of model products that are produced. These are both production and consumption resource-intensive, and as a consequence modeling tends to gravitate toward resource-rich environments.

The *Event Detection* is analogous to the *QA/QC* node, operating as a filter on real-time or retrospective data to provide detected and classified events as a product. It provides topic-based identified events and patterns to the *Data Services Network* node.

The *Visualization* node provides the services to define, generate and manage visual representations of data and data products. It receives algorithm information and data products from the *Modeling Facility* node or the *Data Services Network*, and presentation context information for the specific presentation platform node in use. The *Visualization* node renders data product representations based on one of these paradigms to the presentation platform node.

The *Measurement Calculation* node receives data transformation requests and data products from the *Modeling Facility* and the *Virtual Ocean Simulator* and returns processed data.

The *Process Validator* checks process definitions from science users and operators for compliance and consistency before providing them to the *Data Services Network* for storage.

The *Observation Validation* node validates observation requests coming from for science users and operators for compliance and consistency before providing them to the *Control Services Network* for processing.

2.3.8 Data Services Network

Figure 11 illustrates the operational nodes and needlines that are part of the *Data Services Network* operational node.

The *Data Product Repository* manages the flow of measurement data from the *Instrument Services Network* and the result of data processing from the *CEI* and *Modeling Services Network*. The *Data Services Network* provides persistence and data mediation services for the other networks. In particular, it provides a set of *Repositories* that associate some representation of the system *Resources* with *Metadata* expressed in a *System Ontology*. An *Ontology Validator* enables extensions to the ontologies used in OOI to capture metadata. We can identify three types of ontology: *Policy*, *Observation*, and *Science*. Policies are defined using an ontology that captures interactions and constraints at the OOI technical infrastructure level (e.g., the concept of *Authorization Role* in the *Authorization Policy* and its mapping to an *Interaction Role* in the *COI* Model, cf. OV-7). The *Observation Ontology* allows scientists to describe experiments in terms of data streams, filters, models, and triggers. Finally, the *Science Ontology* provides the dictionaries to describe various types of instruments, their actuators, capabilities, measurements types, and units. In general, the science ontology allows scientists to describe observations using their scientific language.

The *Data Planner* is responsible for negotiating with the *Control Services Network* through the service agreement proposal protocol in order to agree on the resource allocations and partial plans. It then provides resource allocation information to the *Data Product Repository*.

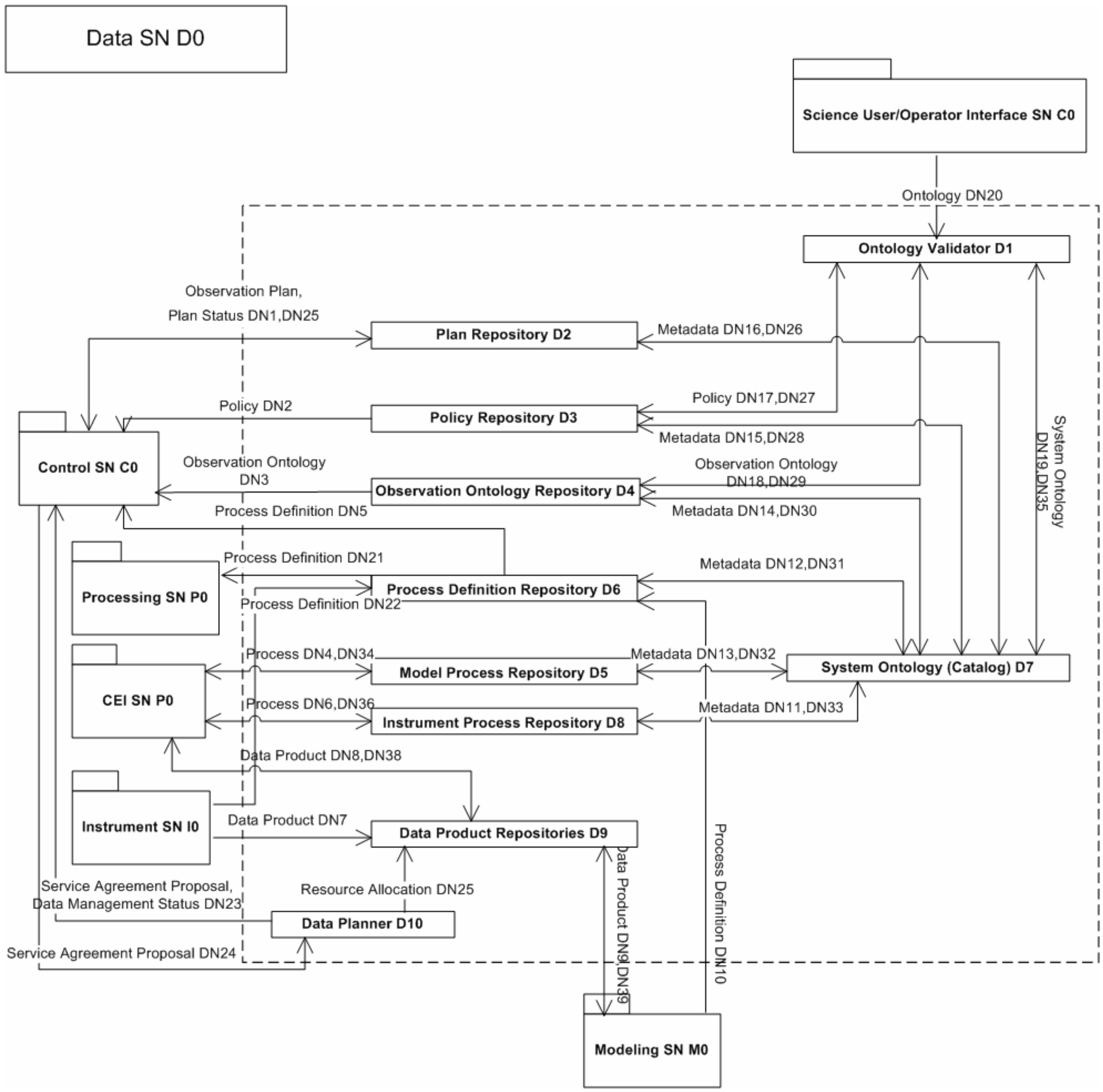

Figure 11. Data Services Network operational nodes and needlines