

Accessing MBARI Data Systems

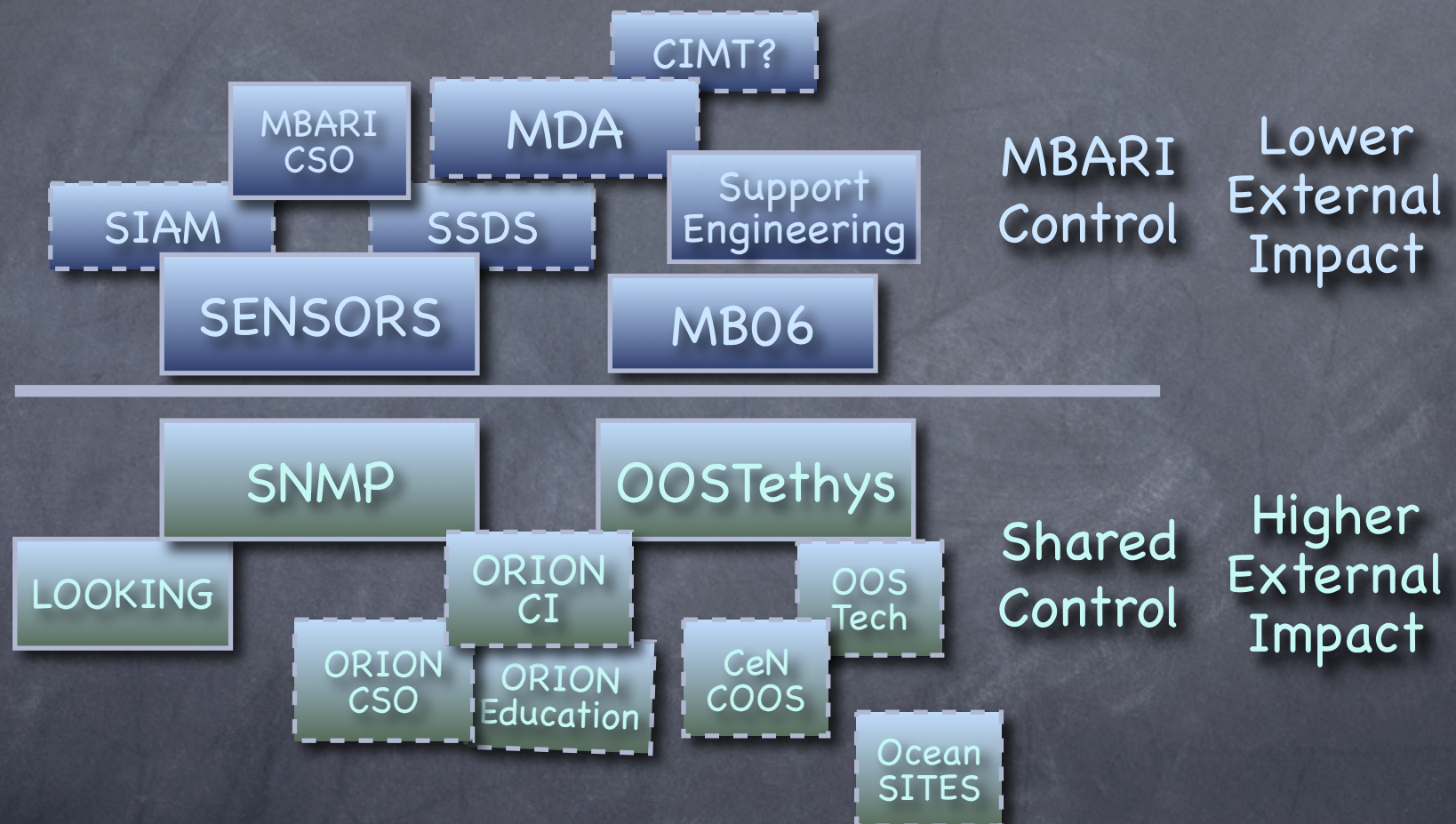
16 November 2006

Version 2

Existing Activities

Resources
Allocated

No Resources
Allocated



These are projects or activities that MBARI is involved in now, or likely to be involved in soon.

The size of the box reflects its estimated potential contribution, over the next two years, to an integrated data access system. The potential contribution is a function of the work that is already planned for that project, the funding that is available, and the relevance to a comprehensive integrated solution.

Projects closer to the center line are more likely to overlap or interoperate with each other.

Note that CSO has two components: one ("ORION CSO") is the externally funded IO, which MBARI hopes to lead; the other ("MBARI CSO") is the MBARI project for 2007.

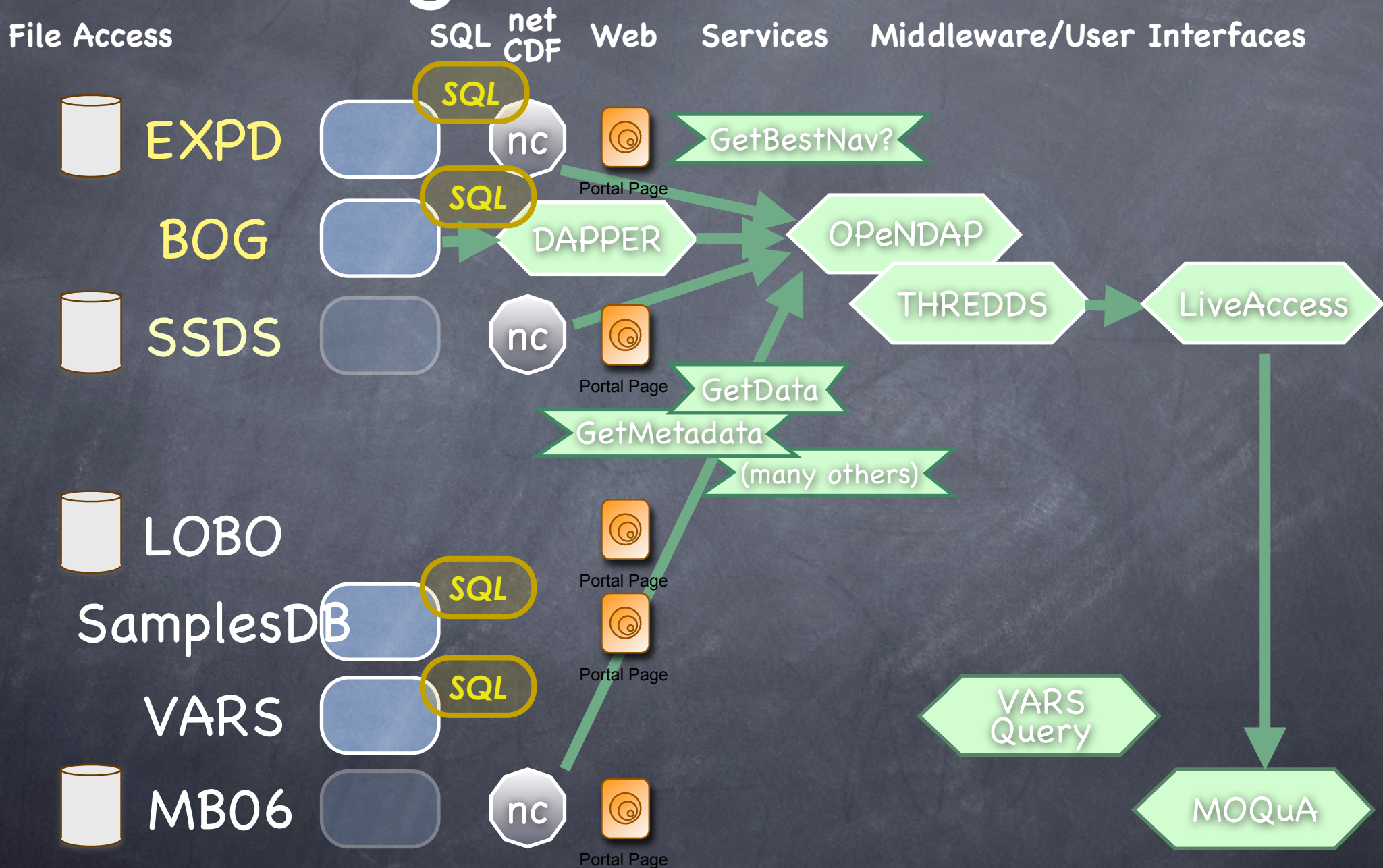
Existing Data Systems



This does not show all of MBARI's data systems, but it shows the most significant systems with readily available data.

The icons represent the data stores that contain key data from each system.

Existing Services & Files



Here we emphasize the data access services provided by the different data stores.

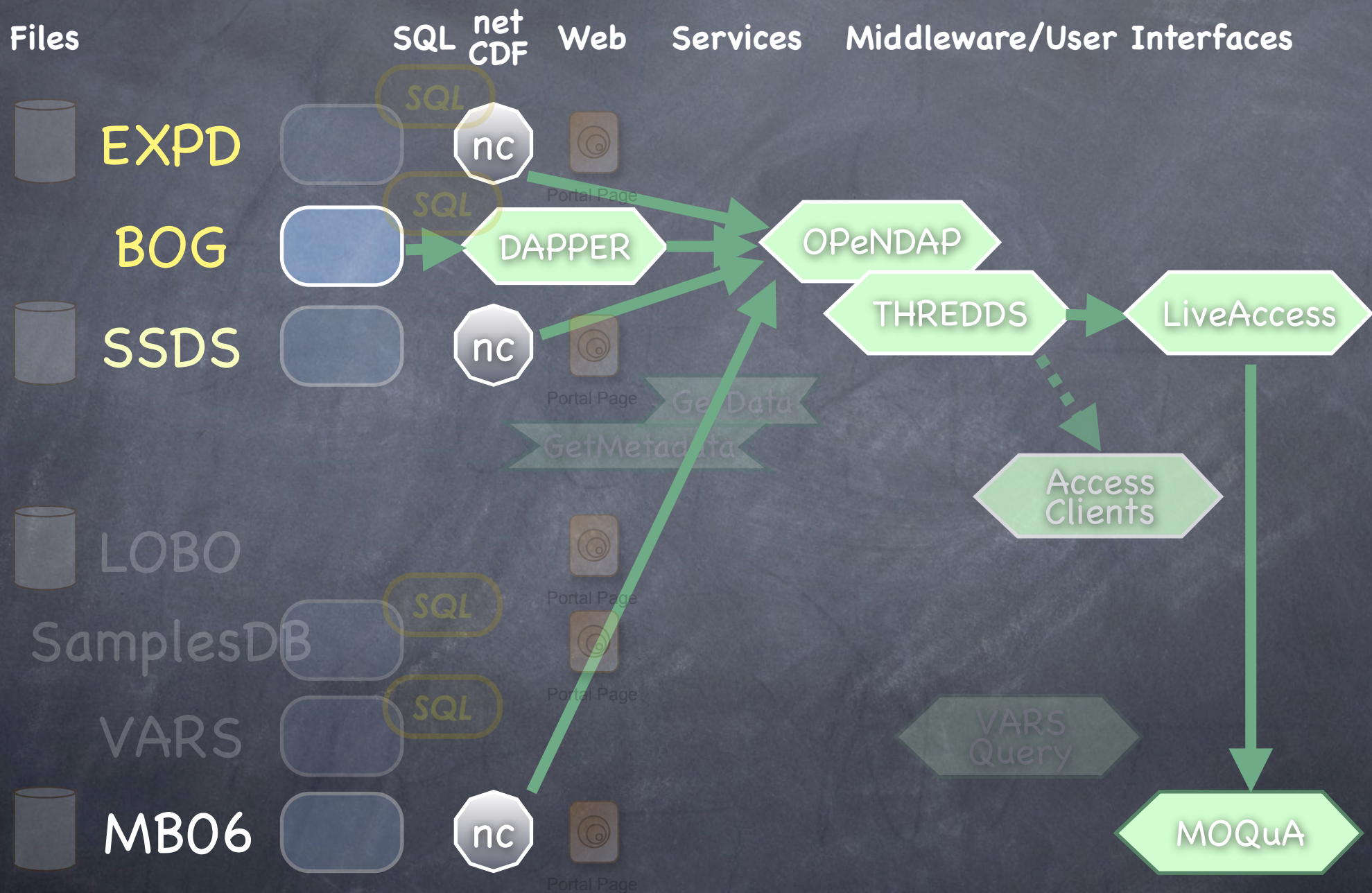
The types of services available are listed across the top, with the icons arranged below (loosely) in columns.

In some cases a service was not shown if it was not heavily emphasized or utilized. For example, SQL queries may be made to SSDS and MB06, but are not a supported method for system users to obtain data.

Web portals provide varying levels of access, but all provide significant searching capabilities. Note, however, that the services provided by each portal's search function vary in significant ways. For example, SSDS (and others?) do not provide a 'search by bounding region' capability.

netCDF Phase — Data

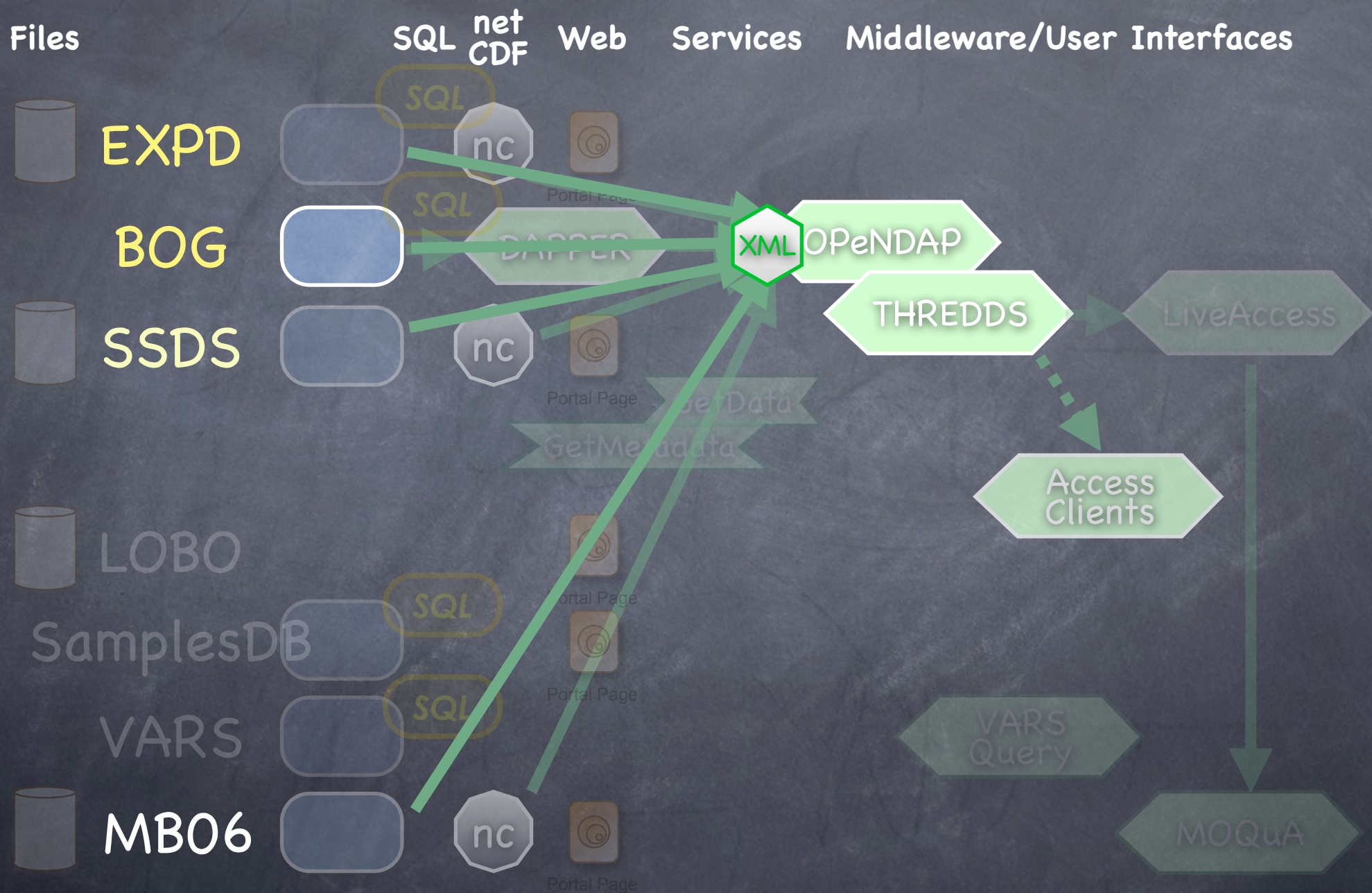
→ = Datapath
....→ = Future



In the netCDF capabilities being proposed, the data aggregation services would be similar to what is available now. (The BOG DAPPER connection is being worked on as this presentation is being developed. “Access Clients” for THREDDS is a potential future addition.)

Existing plans to change the THREDDS-to-MOQuA interface, eliminating the LiveAccess component, are addressed after the next slide.

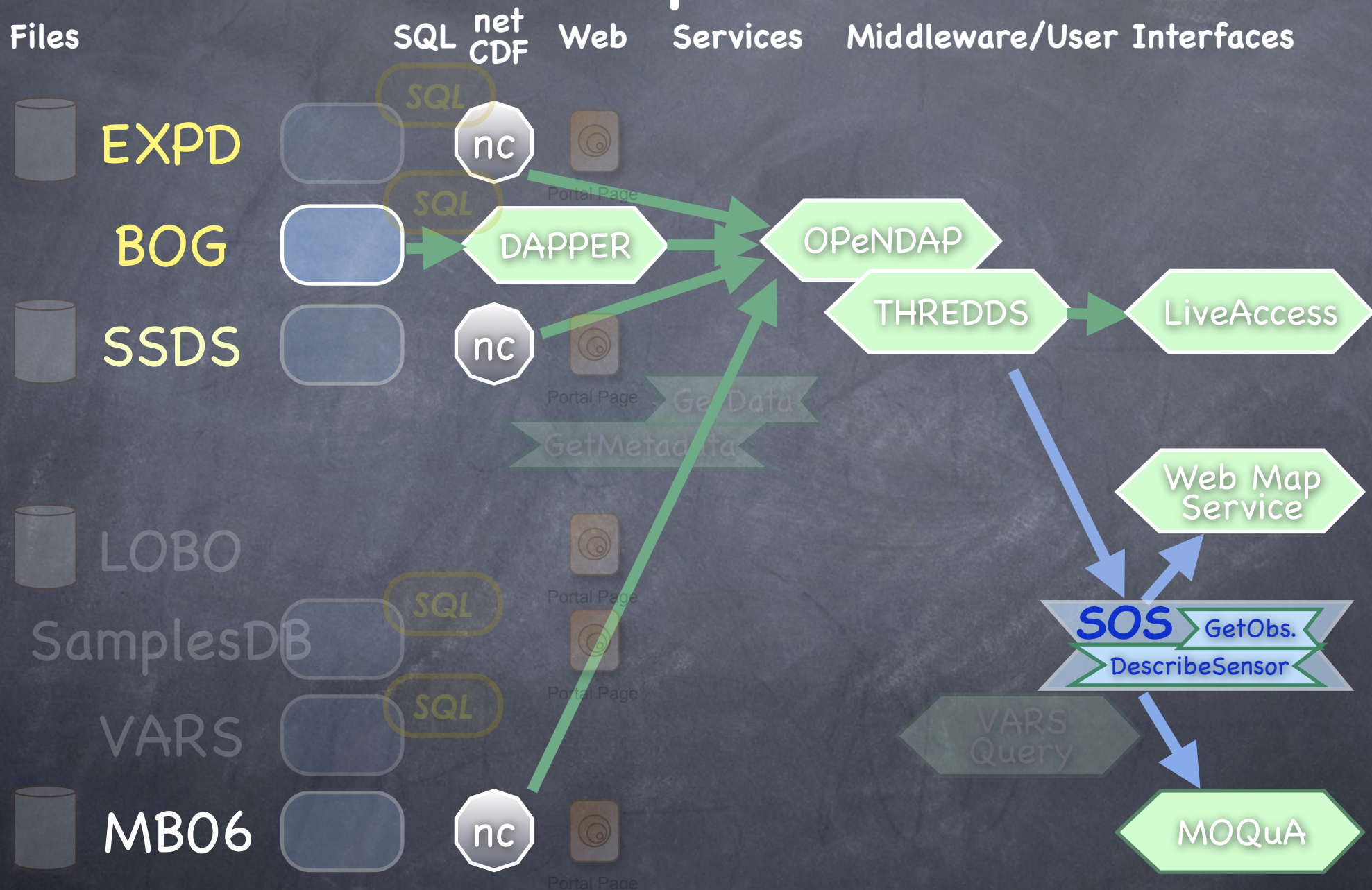
netCDF Phase — Metadata



To improve the ability of clients to find and select data sets, the metadata contained in netCDF files can be augmented using auxiliary XML files. These files describe characteristics of the observing system that are not captured by the netCDF files, but can be processed and made available by OPeNDAP servers and THREDDS catalogs. This also allows more flexibility to specify the files that are available to OPeNDAP (the default is all the files in a specific set of folders).

The definition of suitable XML standards ('schema') for this purpose has been long sought by the OPeNDAP/THREDDS community, and is on the list of important MMI goals.

OGC-SOS Proposal Phase 1



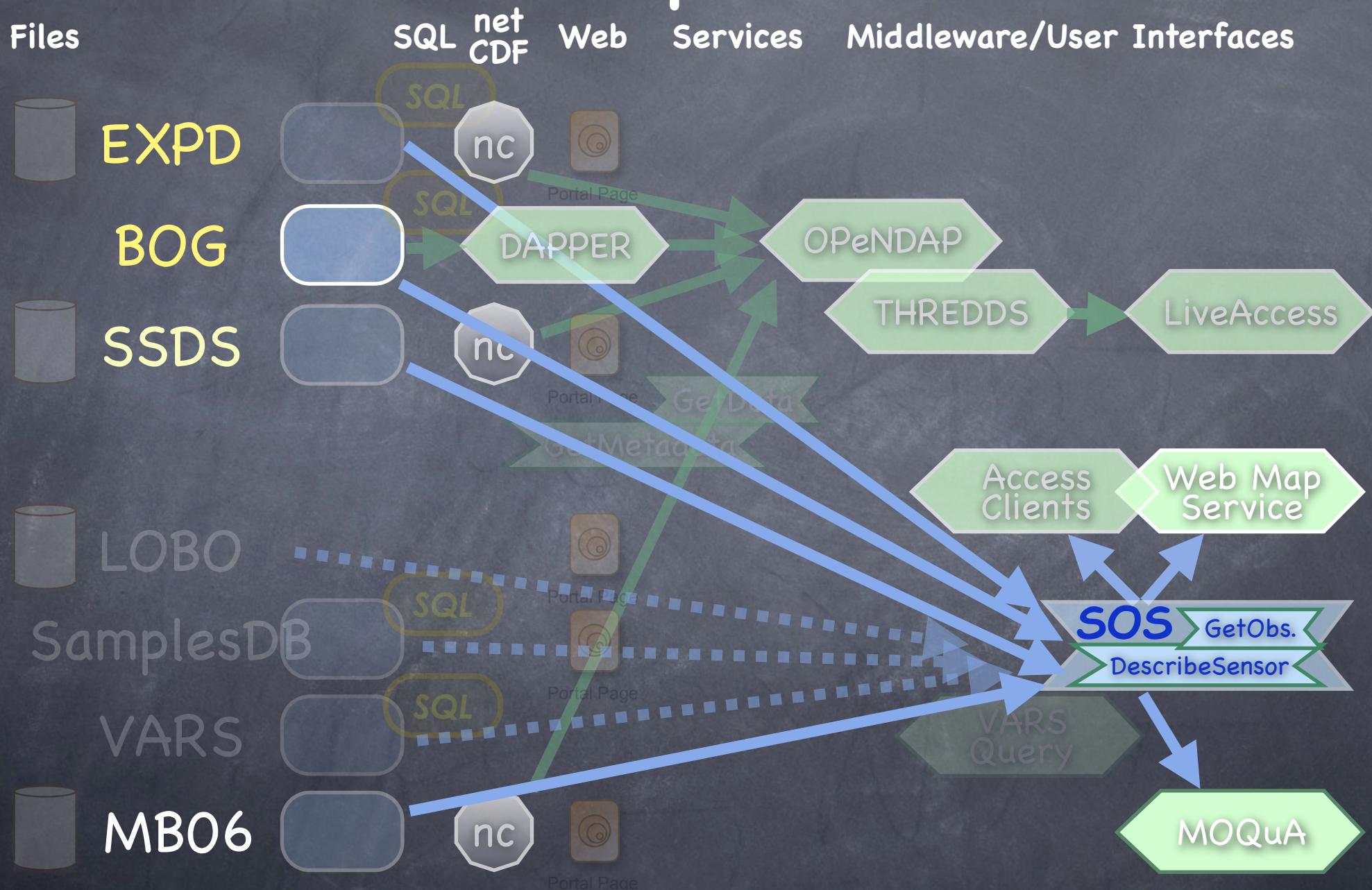
This modification proposes the introduction of OGC's Sensor Observation Standards as a mediator between THREDDS and other services. The existing netCDF-OPeNDAP-THREDDS capabilities would not be altered, but a new interface would be developed to leverage the information in the THREDDS catalog. (Some of that information would be newly available thanks to the netCDF developments described previously.)

The SOS interfaces, which include the capabilities to get the observations, and to describe the source of those observations, can provide sufficient functionality to serve a sophisticated interface like MOQuA. In this case, work would be required to modify MOQuA to use the interface.

Once such an interface is in place, other tools like Web Map Services, can also leverage the same interface. Examples of this tool using SOS have already been developed elsewhere, and MBARI has some experience with the Web Map Services.

To make this possible, either an SOS 'wrapper' would need to be created for THREDDS, or THREDDS itself would need to be modified to provide the SOS capabilities. It is possible external funds or support could be obtained to do this.

OGC-SOS Proposal Phase 2



In a second phase, existing data systems will be modified to support the SOS interface (similarly, either by modifying them or creating wrappers). This interface has been designed from the ground up to enable access to sensor data from both archival data sets and streaming data providers. (See SOS description later in this presentation.) Luis Bermudez is already leading an MMI project to produce a working, comprehensive SOS-based data streaming system.

Some systems are likely to be easily adapted or wrapped to support SOS, but more sophisticated ones may require more subtlety. As a rough estimate, the simpler the system, the easier providing SOS support will be. (On the other hand, the more complex the system, the more valuable SOS support will be.)

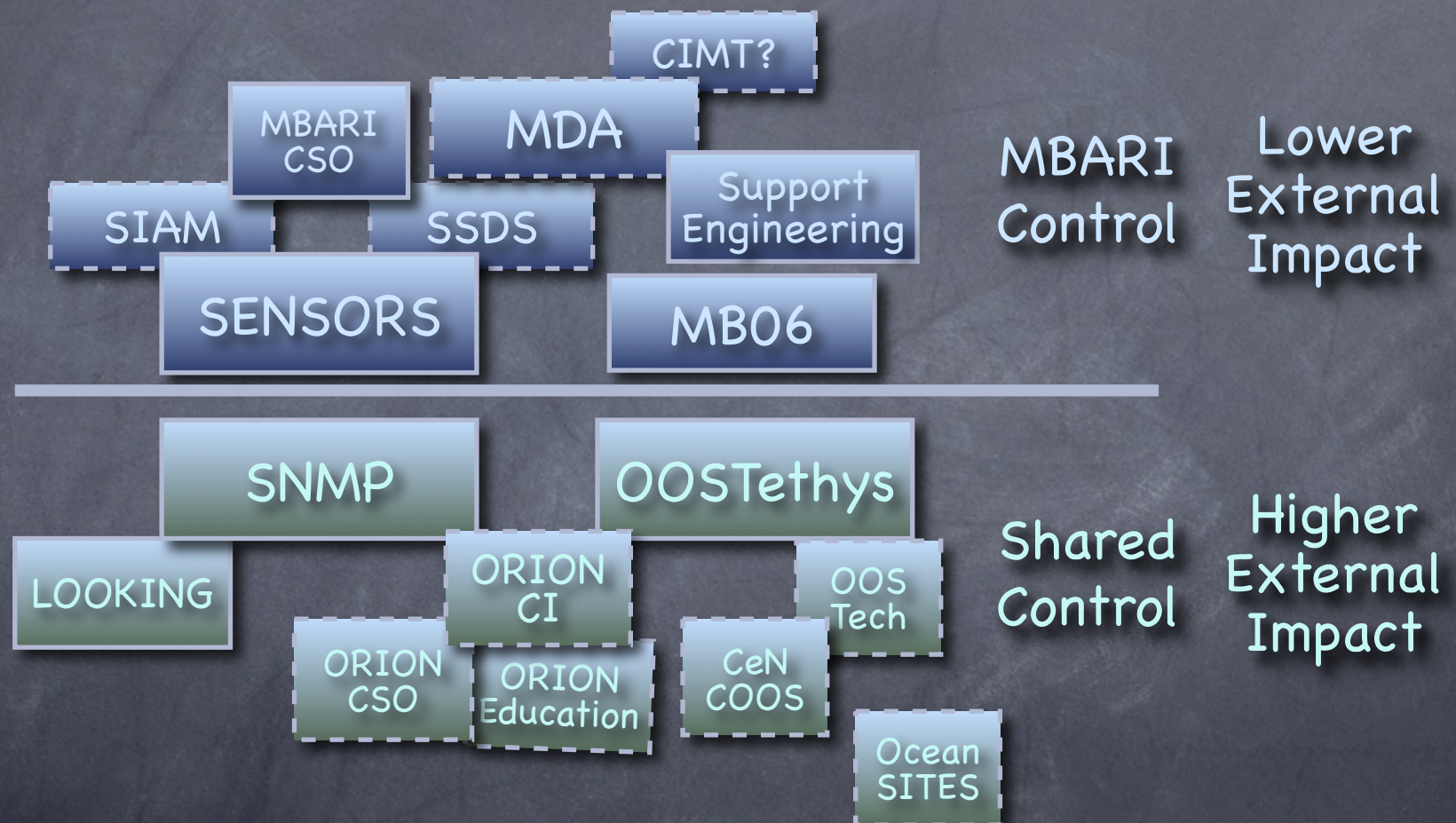
With SOS as a central framework, multiple data access applications can access the corpus of data available at MBARI, and elsewhere.

A question to consider is how to aggregate access so that a single request contacts multiple data systems. This will be evaluated as part of the MMI SOS project (OOSTethys), as well as other activities, and may be straightforward to address before too long.

Existing Activities

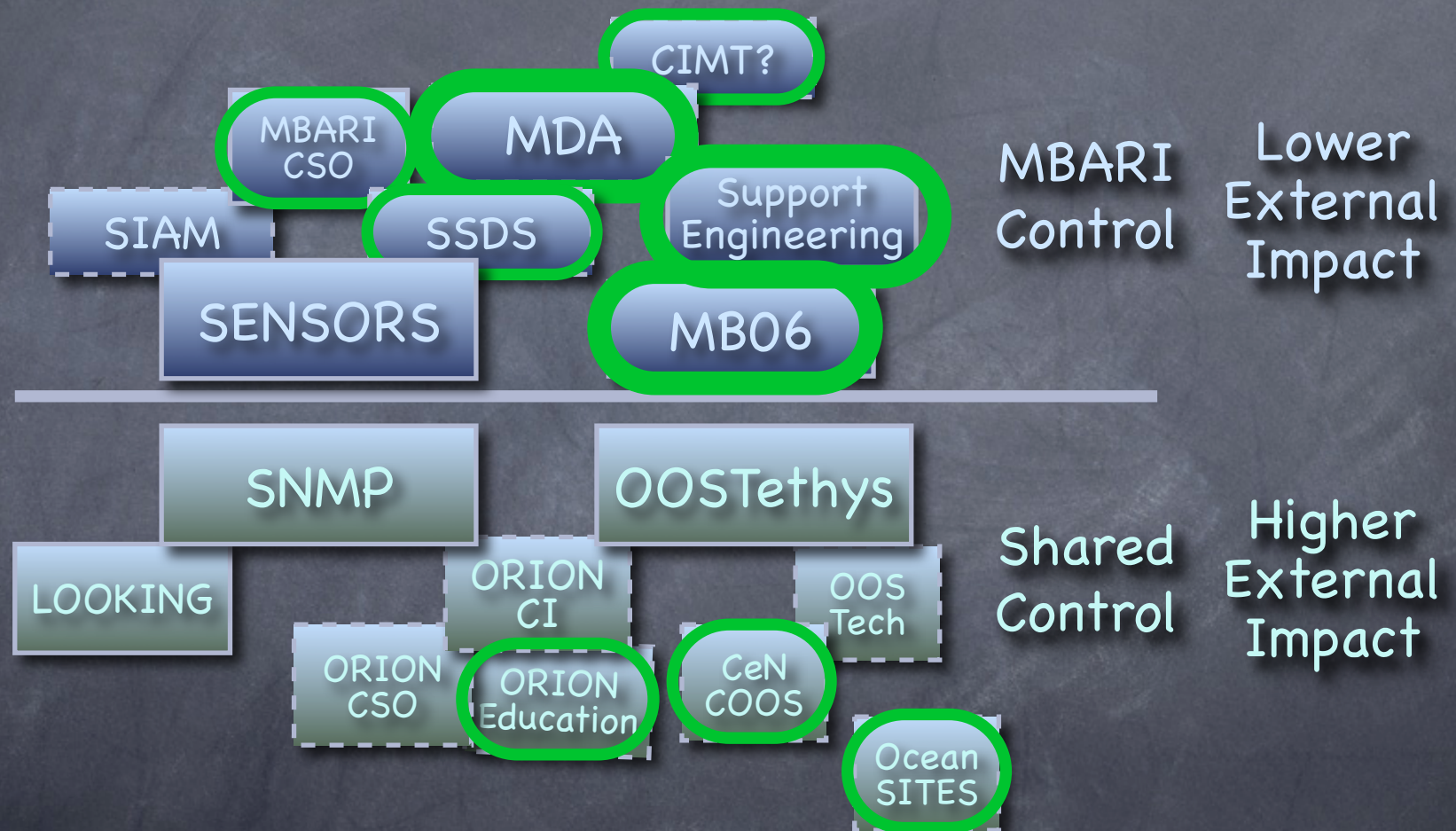
Resources
Allocated

No Resources
Allocated



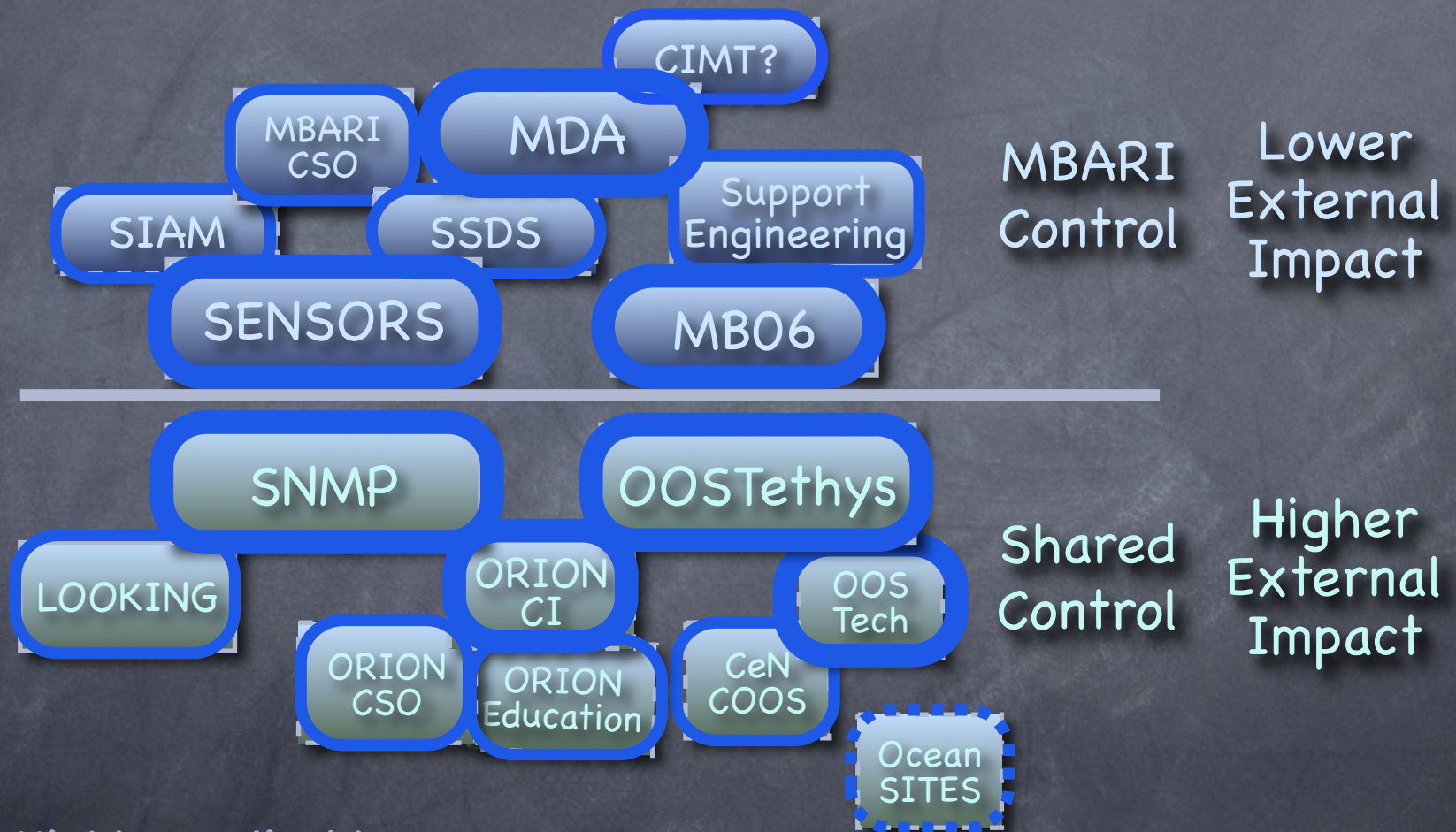
We return to the list of activities introduced earlier.

netCDF work applies to...



The netCDF work applies to a number of the existing activities and projects. It also has direct value in partially addressing some short-term requests for functionality in a common science use case.

OGC-SOS work applies to...



The OGC work will be applicable to a wide variety of projects going forward, based on conversations with likely leaders of those projects.

Partly this increased applicability is due to the natural evolution of 'favored' software solutions, in this case, SOS supplanting OPeNDAP as a favorite architecture.

However, it is also due to the flexibility and sophistication of the SWE approaches, which are much more comprehensive and agile than the netCDF and OPeNDAP frameworks.

Note that with those gains, something is lost. OPeNDAP is highly tuned to structures and arrays, and to selecting efficiently from large data sets. Only some of these features will be available through SOS mechanisms. Thus, the blended OPeNDAP/SOS architecture may provide doubly added value to the community, providing SOS-style access to OPeNDAP-served data.

What is SOS?

Sensor Observation Services

- an Application Programmer Interface (API)—specifically, a web service interface—to obtain information from sensors (a.k.a. instruments)
 - enables requesting, filtering, and retrieving observations and sensor system information
 - the intermediary between a client and an observation repository or near-real-time sensor channel
- part of Open Geospatial Consortium's (OGC) Sensor Web Enablement (SWE) initiative

Sensor Web Enablement

A set of specifications for describing key sensor-related systems (hardware, data, and processes).

- Observations and Measurements (O&M: describe data)
- Sensor Model Language (SensorML: describe sensors)
- Transducer Model Language (TransducerML: like SensorML)
- Sensor Observations Service (SOS: access observations)
- Sensor Planning Service (SPS: control sensors)
- Sensor Alert Service (SAS: subscribing to information/events)
- Web Notification Service (WNS: delivery protocols)

While these are in varying degrees of maturity, they are all quite sophisticated, and SensorML is an OGC-adopted standard. This means it has already been demonstrated in some end application, as OGC requires for all of its standards.

The MMI Sensor Interoperability Workshop developed considerable material documenting and evaluating SensorML and TransducerML.

Note that MMI intends to lead a project to reconcile SensorML, TransducerML, and 1451 data models, and maybe someday the standards themselves. This was also a recommendation of the ACT workshop.

SOS Interfaces

• GetCapabilities

- identify such things as offerings (files, areas observed), observed properties (variables), and information on data sources (sensors or processes)

• GetObservations

- retrieve observation data in the form defined in the Observation and Measurement specification (O&M)
- select based on time, property, location

• DescribeSensor

- return detailed description based on SensorID

The GetCapabilities interface allows clients that do not know what services are offered by a data system to find out about that system, and whether it is likely to be useful.

GetObservations is the service that provides data, according to the criteria specified. (The client can decide what criteria to specify by analyzing the response to its GetCapabilities request.)

DescribeSensor actually is capable of describing whatever the component is that produced the data. This can be a sensor, a sensor system, or a process. Here the SOS interface falls back on the SensorML standard, which offers a comprehensive way to describe sensors or processes.

SOS Status and Usage

(this page in progress)

- Status of standard
 - Submitted as candidate specification (version?)
 - In work by SWE Working Group
- In early use or recommended by many
 - MMI's OOSTethys (Luis Bermudez)
 - CI IO (proposer Matthew Arrott)
 - SEACOOS (developer Jeremy Cothran)

I'm not sure the exact status of the standards, but all are actively being pursued by members of the SWE Working Group. Most are very far along, and starting to attract wide attention.