

Federation Technology in DFC

Hao Xu
Mike Conway
Arcot Rajasekar
Reagan Moore
University of North Carolina at Chapel Hill

Mike Wan
Wayne Schroeder
University of California at San Diego



DFC October 2014 NSF Review Slide 1



Goals of DFC

- Enable scientific collaboration
- Support access to live research data
 - Data Sharing
- Support reproducible data-driven research
 - Workflow Sharing
- Establish national data cyber-infrastructure
 - Persistent and Extensible Architecture



DFC April 2013 NSF Review Slide 2



Need for Federation

Paradigm shift

- **Compute-Intensive** to **Data-Intensive**
- **Large Actions** on Small Amounts of Data to
Small Actions on Large Collections of Data
- **Move Data** to Processing Site (Warehouse Model)
Move Process to Data Site (Map-Reduce Model)
- **Function Chaining** (Programs) to
Service Chaining (Workflows and Web services)

*Data, Services and Workflows
need to be shared, discovered and seamlessly executed*

Federation is the key
And **Federation** is the middle name of DFC



Federation in DFC

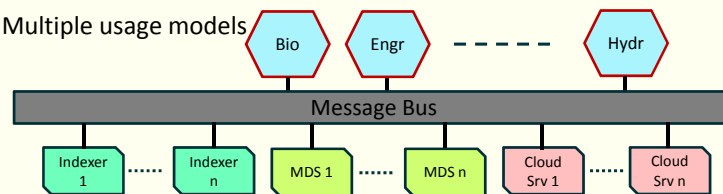
- DFC exposes **three** models of Federation
 - Strong Federation
 - Full and complete protocol-level federation across grids
 - Seamlessly Move from one grid to another
 - Mu Casa Su Casa
 - **Used in DFC to federate Science and Engineering domain grids**
 - Weak Federation
 - One-way DFC to External Resource/Grid Federation
 - Micro-services and **Workflows**
 - DFC needs to 'know' the external protocol - plug-ins & wrappers
 - Still seamless - external access problems hidden from user
 - **Used in DFC**
 - To access THREDDS (netCDF), Sensor system, federal data resources
 - To connect to SEAD and DataONE
 - To access Amazon Web Services (EC2, S3) and External workflows



Federation in DFC

- **Asynchronous Federation**

- Defined & Developed to meet needs of DFC
- Motivation Example: Provision access to indexing services
 - Multiple indexing services – each with its own protocols
 - DFC does not want to federate each separately – access is not seamless
 - Promote a common connectivity - based on message bus
 - Any indexer who can 'talk' this common connectivity can play
- Other examples: Metadata services, Ontologies, Formatting Services, Cloud services, etc.
- Multiple usage models



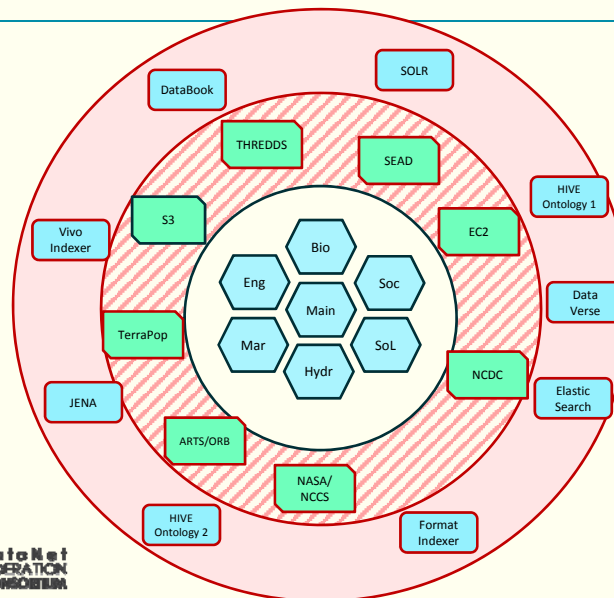
Standard-based Architecture in Asynchronous Federation

- Utilizes industry standard integration patterns and services
- Improves scalability and reliability by utilizing asynchronous, distributed architecture that includes load balancing and fault tolerance
- Lowers technical barriers to entry by easy integration with standard libraries and toolkits

Industry Standards/Components used in Asynchronous Federation

- Currently we support
 - OSGi (ServiceMix)
 - AMQP (ActiveMQ, Proton-J)
 - JSON (Jackson, Jansson, JSONSchema2POJO)
 - ElasticSearch
 - RDF (Standard Vocabularies)
 - Apache Aries Blueprint

Three Federations in DFC





Questions?



DFC April 2013 NSF Review Slide 9

National Science Foundation Cooperative Agreement: OCI-0940841