The Mobius Project

Shannon Hastings, Stephen Langella, Scott Oster

Tahsin Kurc, Joel Saltz

Ohio State University

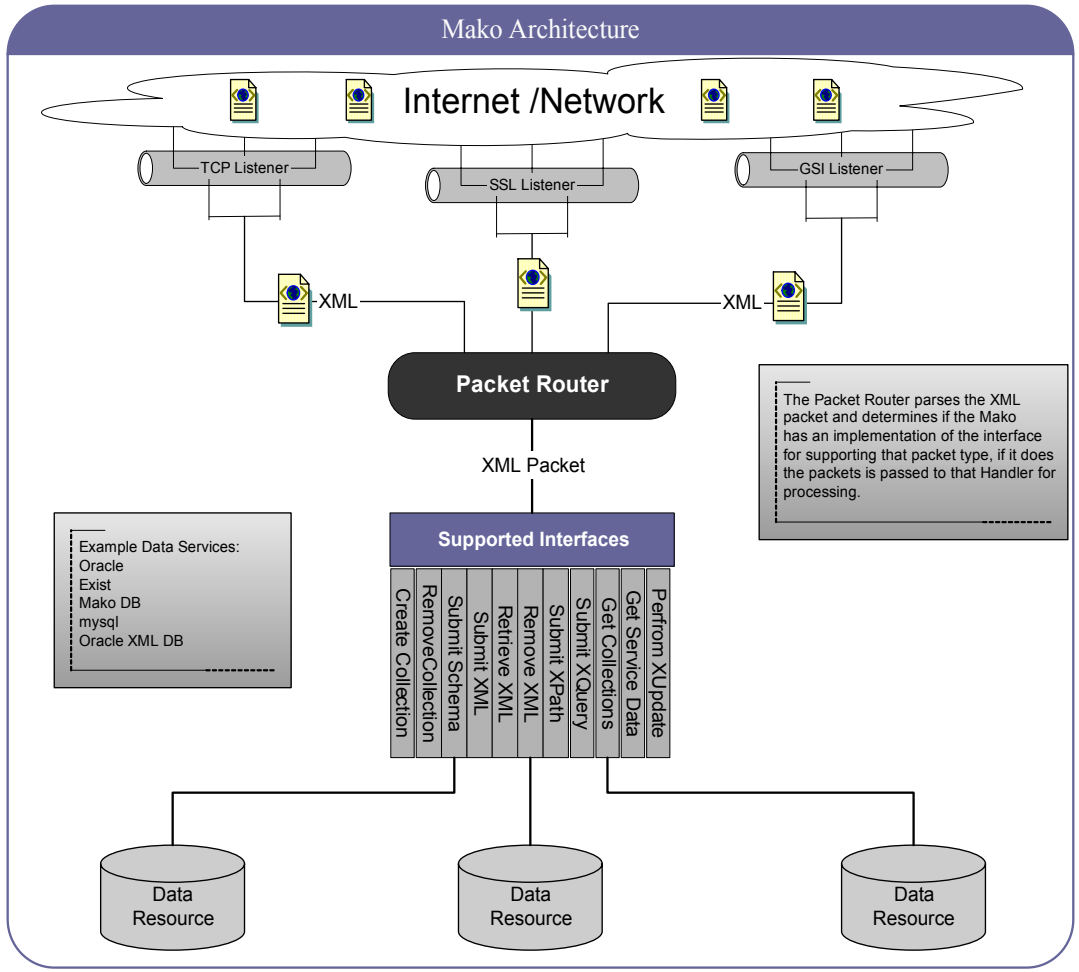
Department of Biomedical Informatics

Multiscale Computing Laboratory

- **Mobius Overview**
  - The Mobius project identifies, defines, and implements a set of services and protocols enabling the management and integration of both data and data definitions.
  - Everything is broken into two components:
    - The protocol definitions and service interfaces.
    - The service implementation of those definitions.
  - Mobius intends to conform to all standards set forth by the GGF and will attempt to leverage relevant existing grid technologies.
    - Currently support the GGF/DAIS XML Realization Proposed Specification
- **Why are we interested in grid data management?**
  - Seamless data services integration.
  - Large scale data management and processing.

- **Mobius Core Services**
  - Mako (Federated Adhoc Storage Services)
  - Data Translation Service
  - Global Model Exchange (GME)
- **Mobius Extension Services**
  - VMako (Single virtual service view of a federation of Makos or VMakos)
  - Mentat (Semantic data management services)
- **Mobius Future Services**
  - Other Higher level query services (semantic query, inference services etc.)
  - Data Transportation Services.
- **Other Needed Grid Services**
  - Namespace Management
  - Service Naming
  - Data Replication
  - Security

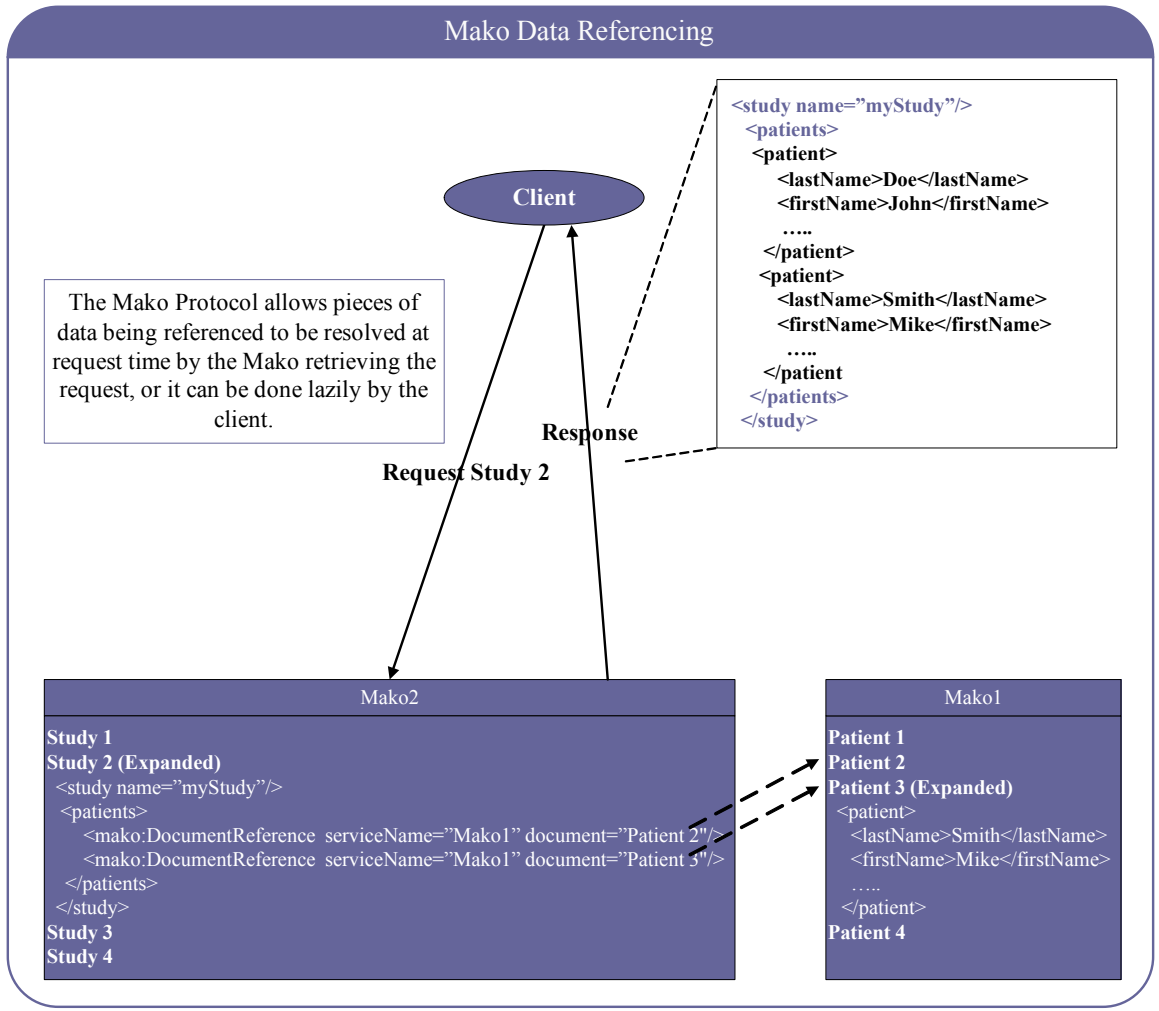
- **Mako Service**
  - Exposes existing/new data services as XML data services through a set of well defined interfaces based on the Mako protocol. (GGF/DAIS XML Realization Specification).
  - Configurable Protocol Handling
    - Admin specifies handlers to process incoming packets based on type.
    - Specific implementations of handler can easily be written and integrated.



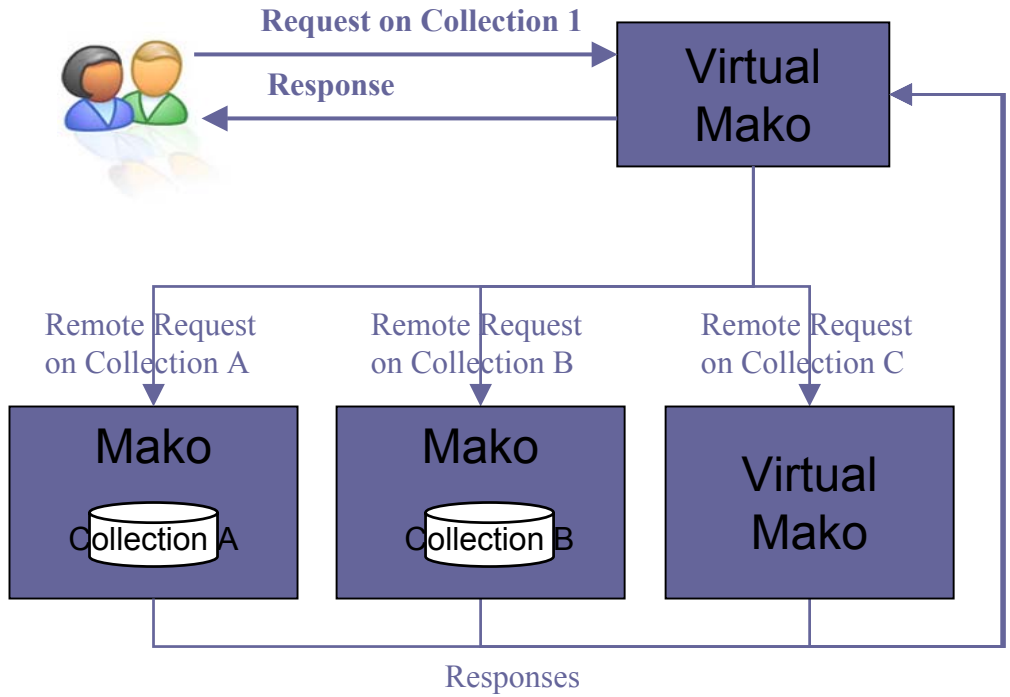
- **Data Resource Support**
  - Mako DB
    - In house XML database.
    - Optimized for federated adhoc usage of XML.
    - Plugs into Mako framework.
  - XML DB Support
    - Built in support for XML databases that support the XML DB API.
  - Other Data Resources
    - Easily integrated, by implementing a small set of handlers for them
    - In the future these handlers will be publicly available.

- **Mako Specialized Features**
  - Data Validation
  - Element Referencing
    - Static
    - Dynamic through xpath queries
  - Lazy Retrieval
  - Distributed Document Object Model (DOM)

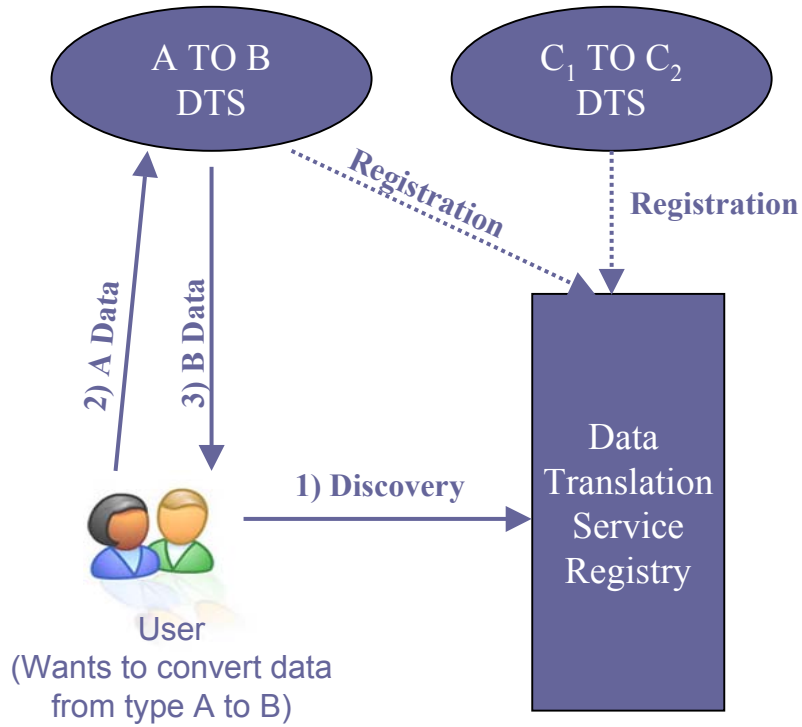
- **Mako Data Referencing Use Case**
  - Shows retrieving data that is federated across multiple Mako services.



- **Virtual Mako**
  - Simplifies client-side complexity of interfacing with multiple Makos by presenting a single virtualized interface to a collection of federated Makos.
    - "Is a" Mako service implementation
    - Acts as a data integration point for distributed queries
    - Pluggable algorithms for XML instance ingestion/distribution
    - Protocol request broadcast and response aggregation
    - Maps a Virtual Collection to a number of remote standard Collections



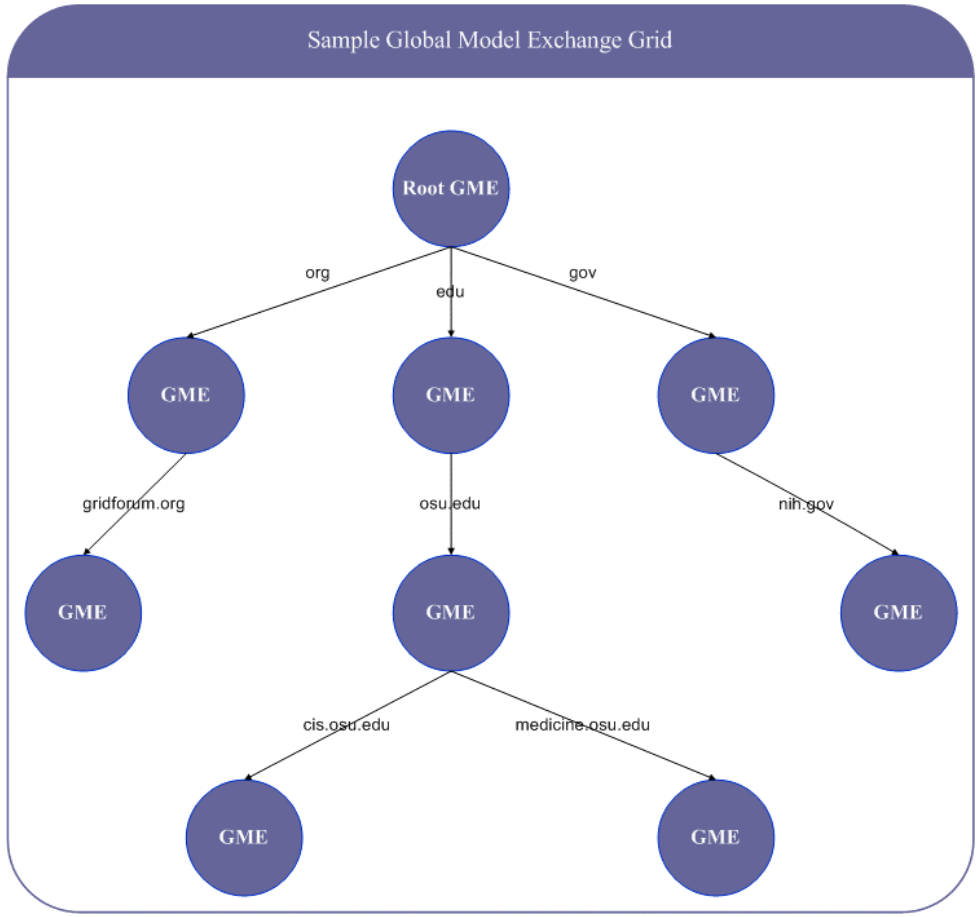
- **Data Translation Service**
  - Issues
    - How do I translate one data type to another?
    - How do I convert an old version of a data type to a newer one?
  - Protocol and service framework for handling the mapping of one data instance or data definition to another should exist.
  - Allows two protocol disjoint services to communicate
  - Enables translating between changing data types.



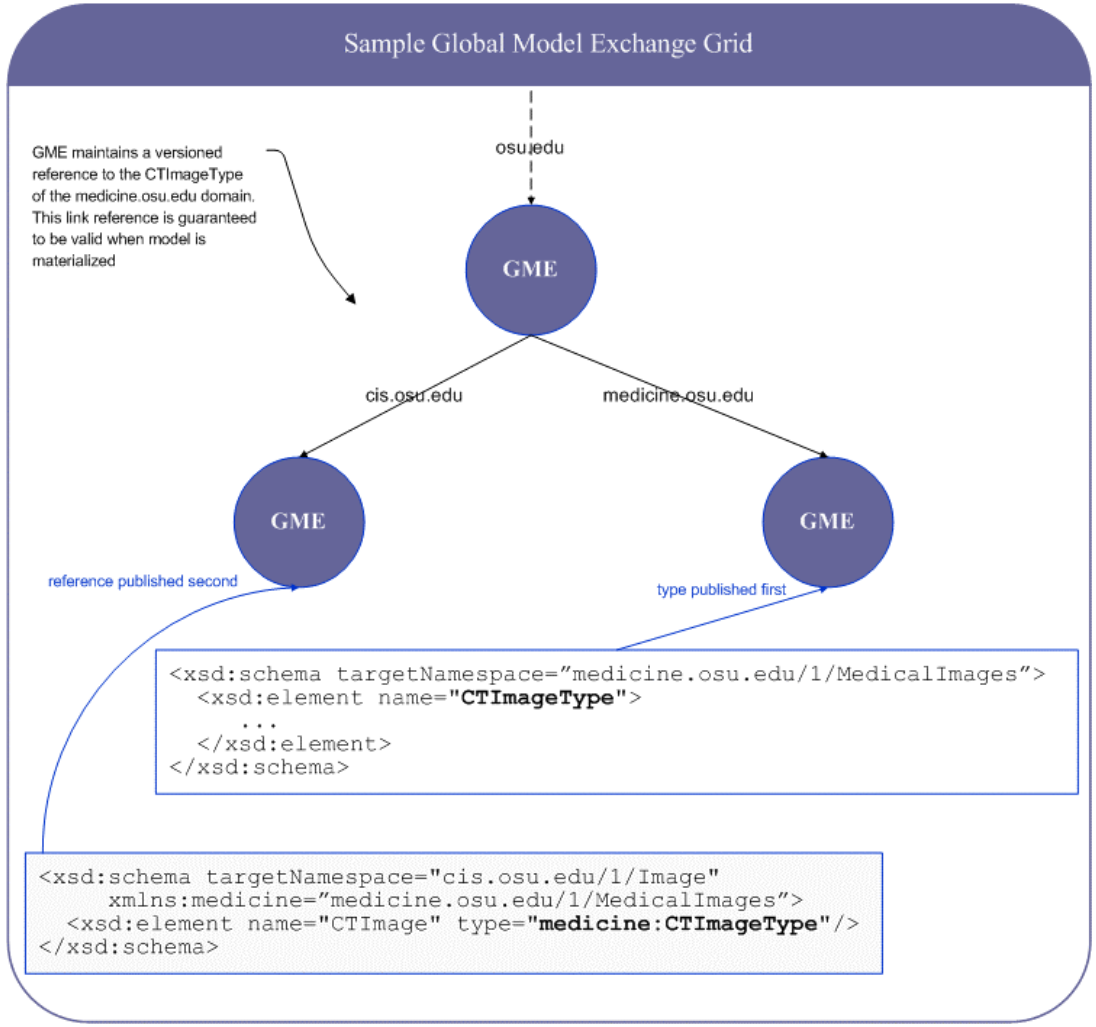
- **Need for a global data definition management!**
  - What is “global data definition” (Global Models)?
  - Promote creation and evolution of standard definitions of data types.
  - For communication between multiple institutions they must agree on a common structure or a mapping between structures.
  - Allow for sharing and discovery of data definitions in a grid environment.

- **Global Model Issues**
  - User/Organization defined entities
    - ex: *my "person"* != *your "person"*
  - Changing models
  - Models disappear
  - Prevent conflicting models
  - Discovering models
  - Multiple definitions of similar models for different communities (syntactic / semantic mapping)

- **Global Model Exchange**
  - Manages the Global Model
    - handles presented issues
  - Provides submission and discovery protocol
  - DNS like architecture
    - Hierarchical authority subordinate domain division
    - Model resolution query messaging
    - Local cache with TTL.

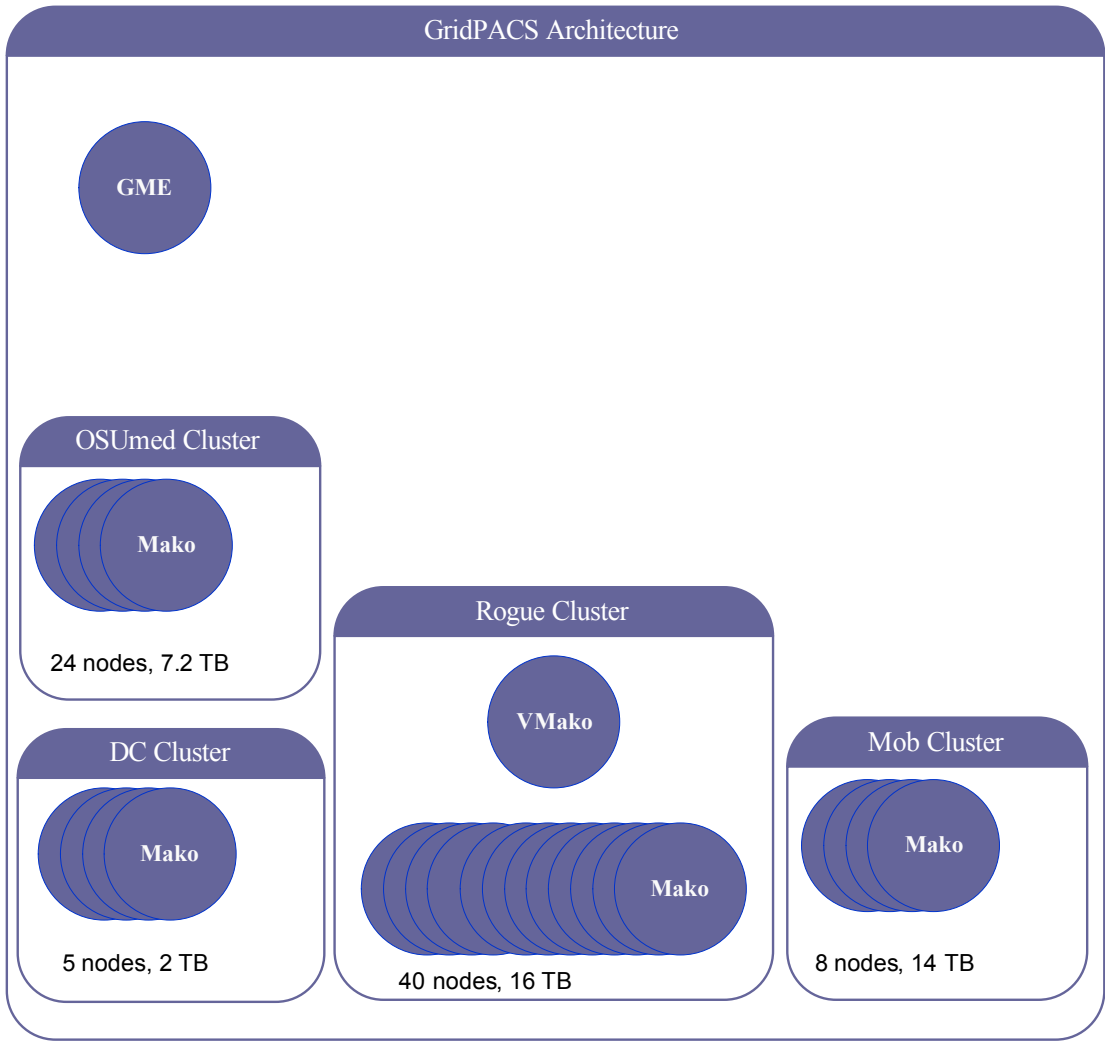


Users/Services publish schemas to the authoritative GME of that namespace. Then any other services/users from similar or different organizations who have the authority will be able to reference, use, alter, etc the data definitions of that schema.

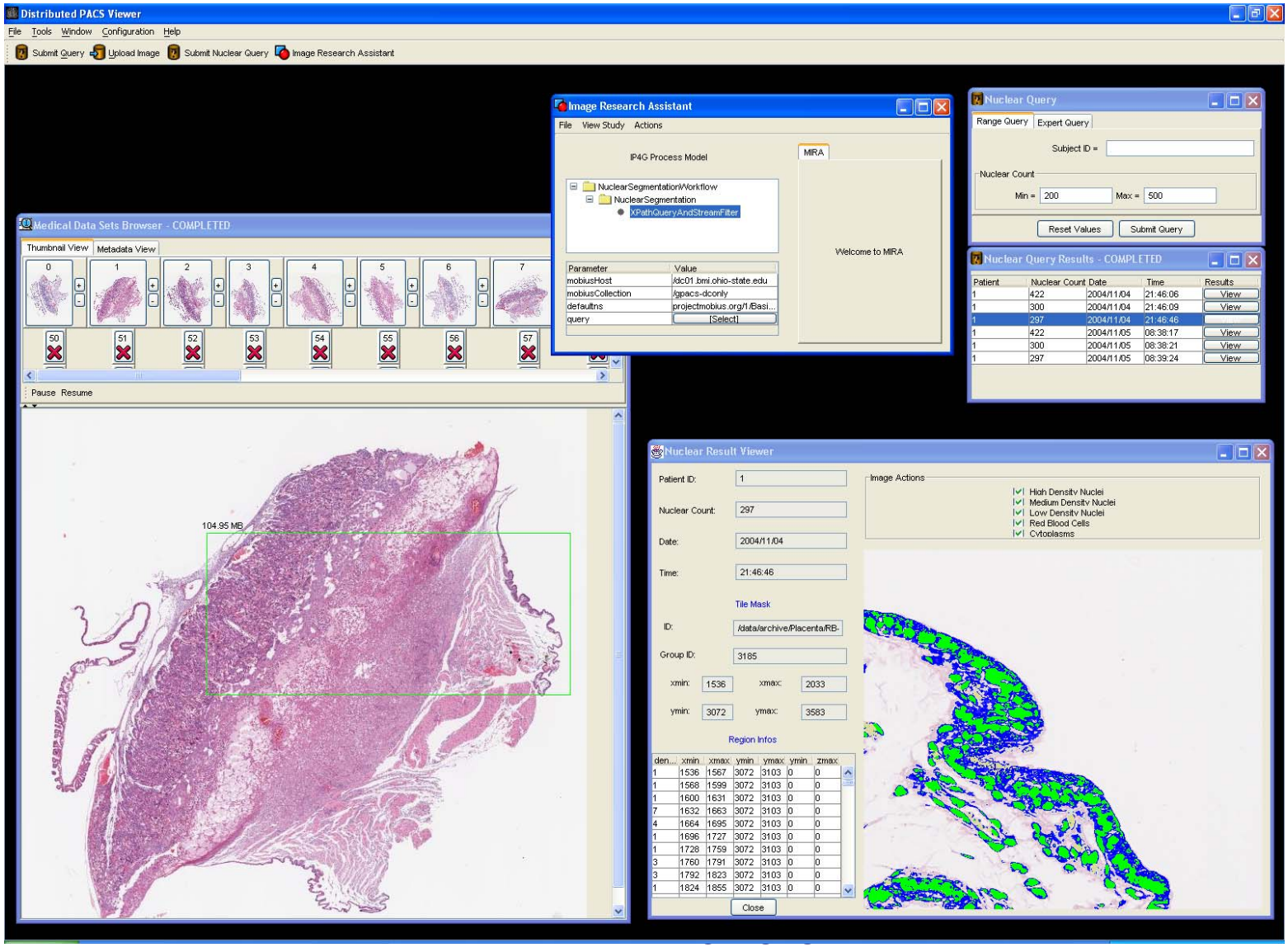


- **Globus Integration**
  - Mobius Services have been integrated with Globus and can be run as Globus Grid Services
    - Mobius protocol layered on top of the OGSI protocol
    - Mobius services can be leveraged and discovered using Globus toolkit.
  - Integrated Mobius Services take advantage of Globus Security Infrastructure (GSI) .
    - Secure Communication
    - Authentication.

- **Mobius Application Examples (www.projectmobius.org)**
  - GridPACS
    - Distributed image management and analysis application.
  - BlockMan
    - Large-scale data de-clustering and querying application.



The MOBIUS PROJECT



Mobius in the Community

- **NIH-CaBIG**
  - Mobius has been adopted as a core middleware component of the NIH Cancer Grid effort.
- **OGSA-DAI**
  - Mobius has been selected as a principle sister project to the E-Sciences project OGSA-DAI effort.
- **GGF**
  - Active members of Data Access and Integration Services Working Group (DAIS-WG)
- **Biological Research Technology Transfer (BRTT)**
  - Mobius is currently funded under a State of Ohio technology transfer grant.