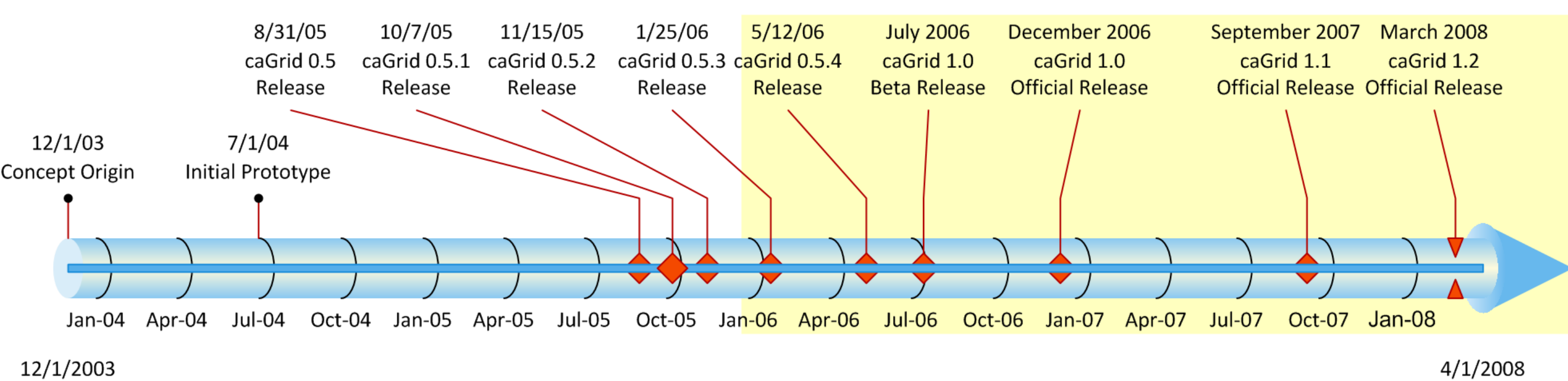# caGrid Overview

caBIG™ cancer Biomedical Informatics Grid ™

Technical Team: Ohio State University - Biomedical Informatics Department, University of Chicago/Argonne National Laboratory, SemanticBits, LLC., Ekagra Software Technologies Limited

caGrid History

- **Development project of Architecture Workspace, aimed at helping define and implement Gold Compliance**
- **Gold compliance creates the G in caBIG™**
  - **G**old => **G**rid => connecting Silver Systems
- **Initially developed as the Grid toolkit for caBIG, 2004**
- **caGrid 1.0 was a revolutionary release of the caGrid infrastructure, replacing the 0.5.x test bed stream**
- **The last stable release of caGrid was version 1.2, released mid March 2008**

Timeline: 12/1/03 Concept Origin; 7/1/04 Initial Prototype; 8/31/05 caGrid 0.5 Release; 10/7/05 caGrid 0.5.1 Release; 11/15/05 caGrid 0.5.2 Release; 1/25/06 caGrid 0.5.3 Release; 5/12/06 caGrid 0.5.4 Release; July 2006 caGrid 1.0 Beta Release; December 2006 caGrid 1.0 Official Release; September 2007 caGrid 1.1 Official Release; March 2008 caGrid 1.2 Official Release

12/1/2003   4/1/2008

Key Enhancements in caGrid 1.2

- **Usability Improvements**
  - a new component, caGrid Transfer Service, provides a simplified approach towards building services that need to handle large scale data.
  - the process of creating, deploying, and undeploying services in Introduce Toolkit has been greatly improved
- **Security**
  - a new framework has been developed, called the Web Single Sign On (WebSSO) that provides an out-of-box comprehensive Single Sign On solution for Web applications that use the caGrid GAARDS framework.
  - A newly developed Credential Delegation Service, to support the delegation of grid credential between users and services in a seamless manner.
- **Continuous Integration**
  - integration with the Apache Ivy build system to substantially improve dependency management of sub projects within the caGrid core infrastructure and externally.
  - supports the latest releases (version 4.0) of caCORE Software Development Kit
- **Workflow Improvements - Early Preview**
  - An *early preview* of the ability to orchestrate caGrid services in a workflow using the Taverna Workbench

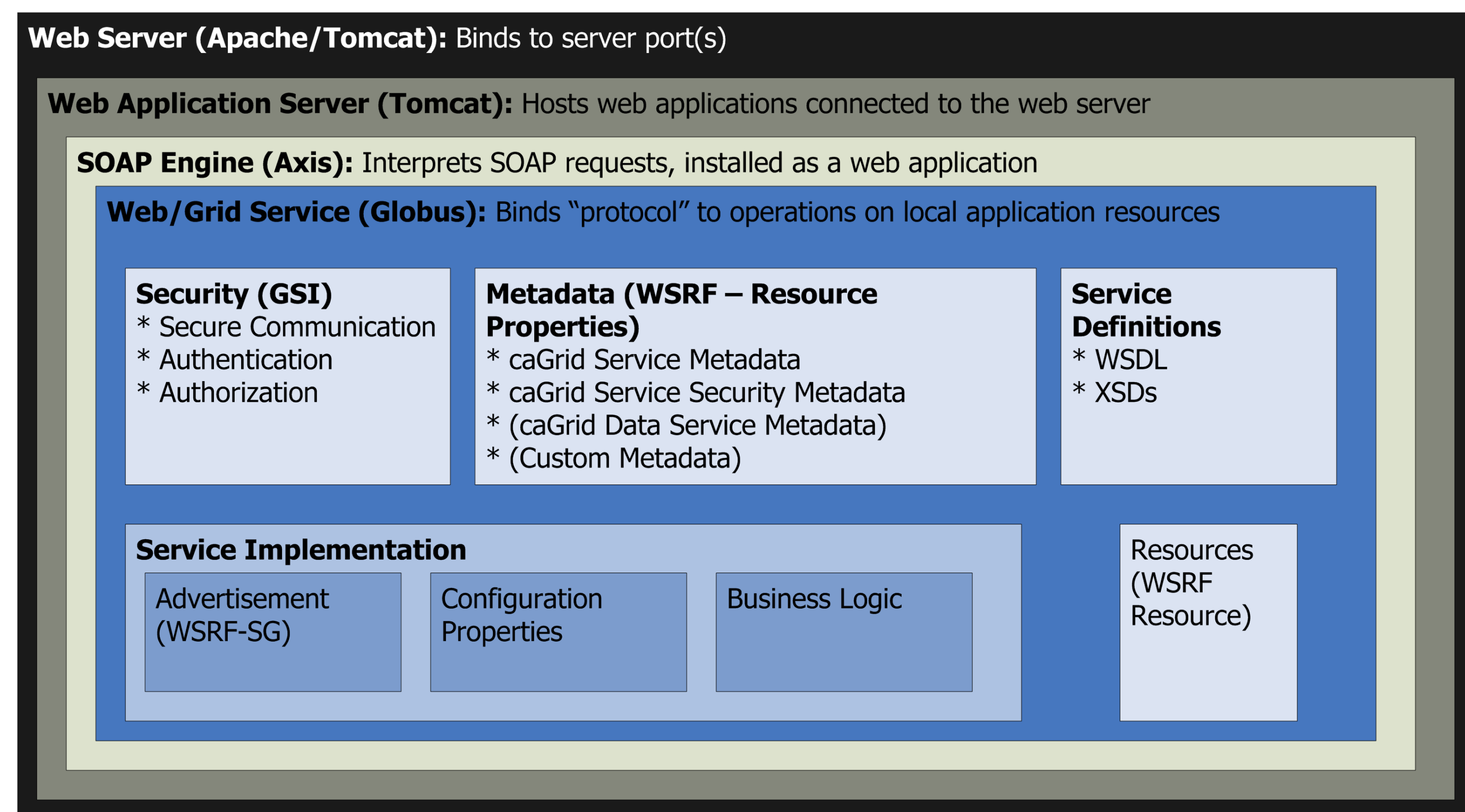### Production Grid Deployment
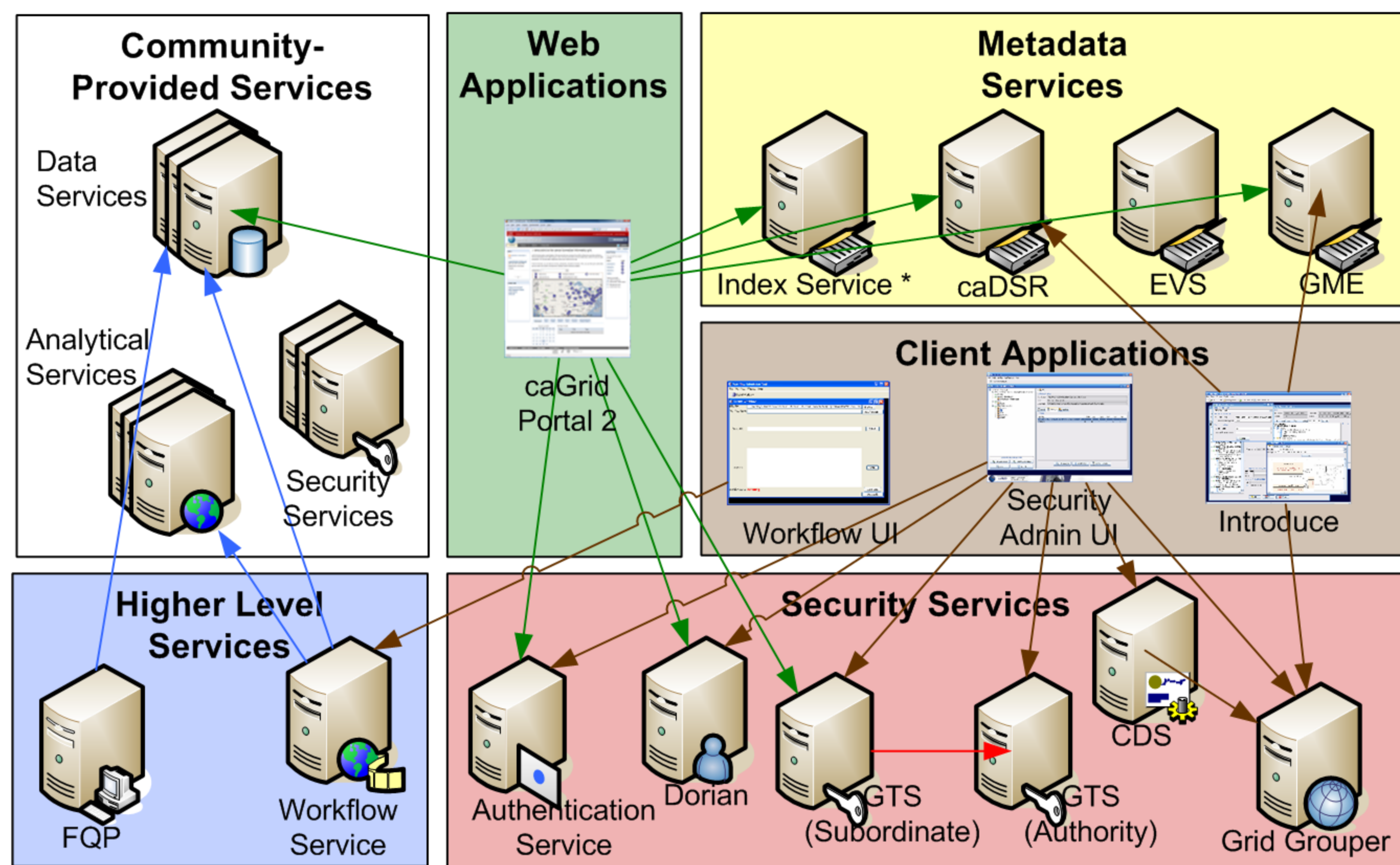
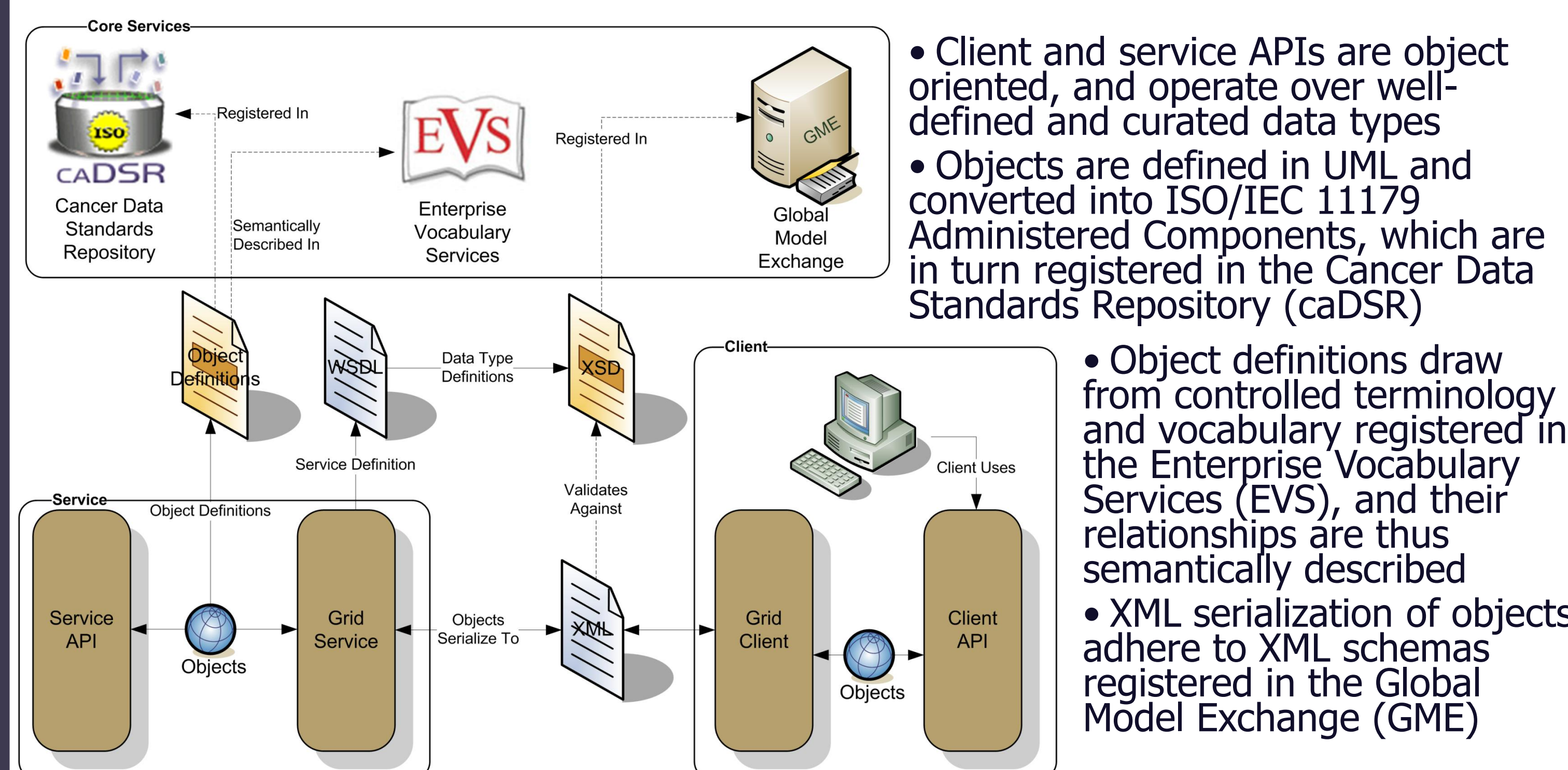

*All Services Register with the Index Service

Core Infrastructure Focus

- **Leveraging Grid technologies and standards as an interoperability platform**
  - Metadata Infrastructure
  - Surfacing wealth of existing caBIG data-oriented metadata on the grid
- **Providing new service-oriented metadata**
- **Security**
  - Integrating existing systems and applications with Grid security
  - Lowering burden of implementation of grid-wide and local policy
- **Service Developer Tooling**
  - Powerful platform for bringing applications and data to the grid
- **Facilitating Grid wide operations**
  - Federated query, workflow execution, resource discovery
- **Making the Grid more accessible**
  - Graphical installation and configuration, higher-level object-oriented APIs, web portals, graphical administrative applications
- **Quality**
  - Comprehensive testing infrastructure, automated builds and test execution on multiple platforms, dashboard with historical archive

caGrid Service Layers

Web Server (Apache/Tomcat): Binds to server port(s)

Web Application Server (Tomcat): Hosts web applications connected to the web server

SOAP Engine (Axis): Interprets SOAP requests, installed as a web application

Web/Grid Service (Globus): Binds "protocol" to operations on local application resources

Security (GSI)
* Secure Communication
* Authentication
* Authorization

Metadata (WSRF – Resource Properties)
* caGrid Service Metadata
* caGrid Service Security Metadata
* (caGrid Data Service Metadata)
* (Custom Metadata)

Service Definitions
* WSDL
* XSDs

Service Implementation
- Advertisement (WSRF-SG)
- Configuration Properties
- Business Logic

Resources (WSRF Resource)

caGrid Data Description Infrastructure

- Client and service APIs are object oriented, and operate over well-defined and curated data types
- Objects are defined in UML and converted into ISO/IEC 11179 Administered Components, which are in turn registered in the Cancer Data Standards Repository (caDSR)
- Object definitions draw from controlled terminology and vocabulary registered in the Enterprise Vocabulary Services (EVS), and their relationships are thus semantically described
- XML serialization of objects adhere to XML schemas registered in the Global Model Exchange (GME)
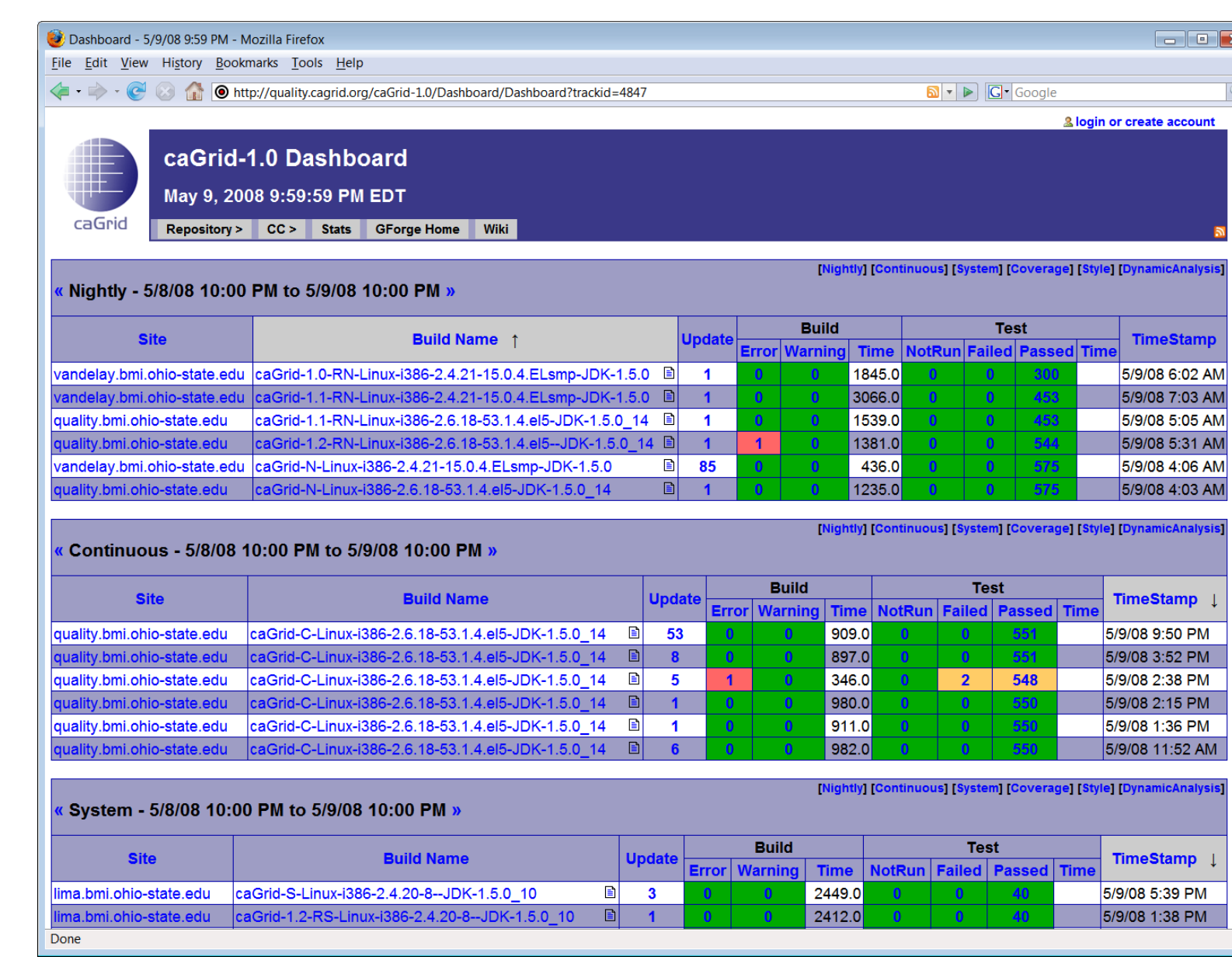
caGrid Suite

The "caGrid Suite" currently consists of 3 high-level products:
- **caGrid Core Infrastructure**
  - consisting of the core services, APIs, and tools
- **caGrid Workflow Client**
  - focusing on making workflows usable by the scientific community
- **caGrid Portal**
  - providing a web-based application enabling discovery and interaction with the services that are available on the caGrid infrastructure
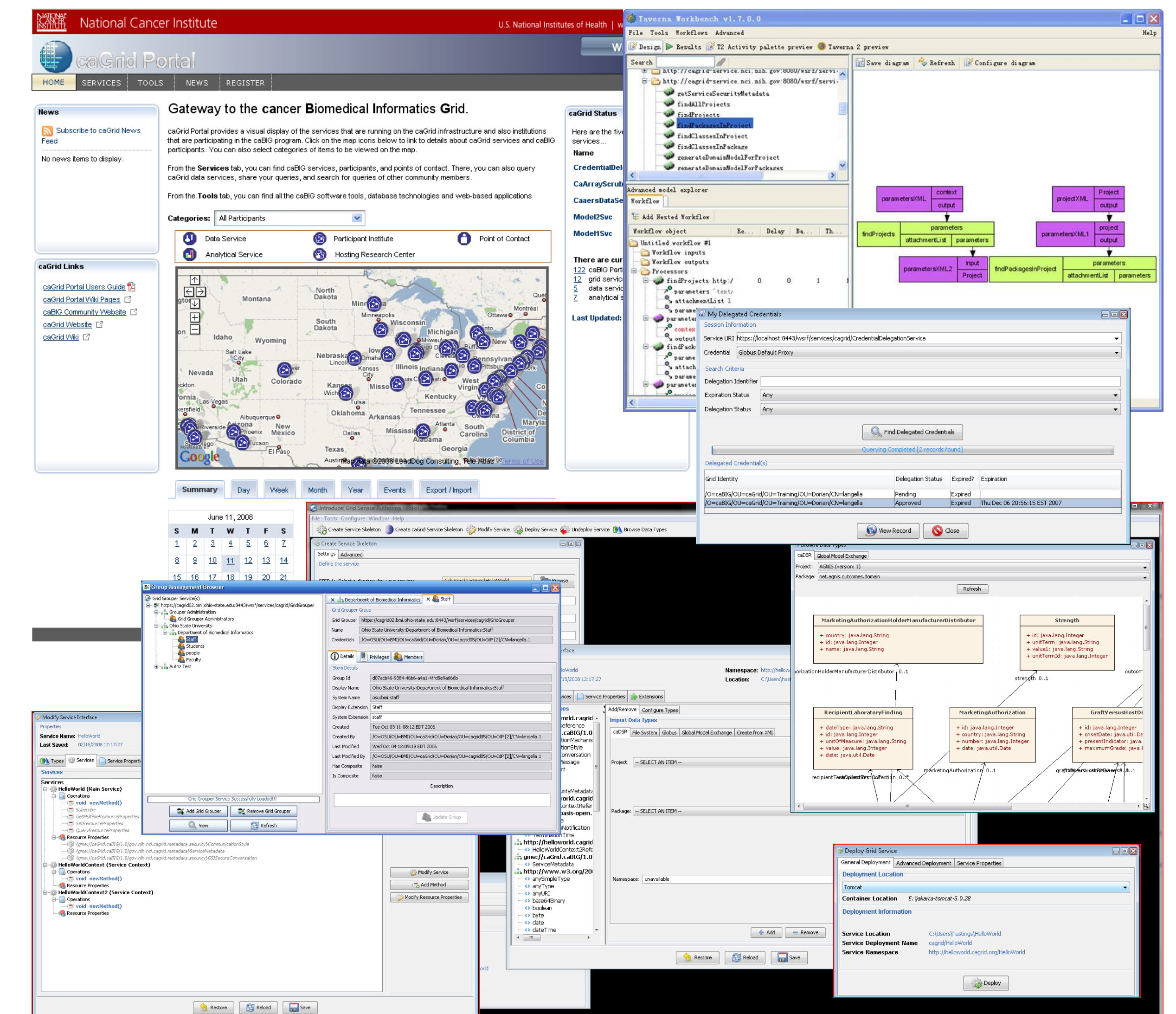
Emphasis on Quality

- **Testing**
  - Unit, System, Integration, and Validation
- **Automated Builds/Tests on multiple platforms**
  - Nightly (on a schedule)
  - Continuous (every CVS check in)
- **Quality Dashboards**
  - DART (multi-site, historical archive of quality)
  - CruiseControl (single site)

caGrid

- **caGrid Web Pages:**
  - http://cagrid.org
  - https://cabig.nci.nih.gov/workspaces/Architecture/caGrid
  - http://gforge.nci.nih.gov/projects/cagrid-1-0/
- **caGrid Users Mailing List**
  - https://list.nih.gov/archives/cagrid_users-l.html
  - cagrid_users-l@list.nih.gov