On the Usability of Solutions from the caBIG™ Project in MediGRID

Thomas Steinke and Ulrich Sax

eHealth Conference 2007

Berlin, April 18th 2007

Outline

- The German MediGRID project
- The U.S. caBIG™ project
- Re-using caBIG™ technology in MediGRID?
- Outlook

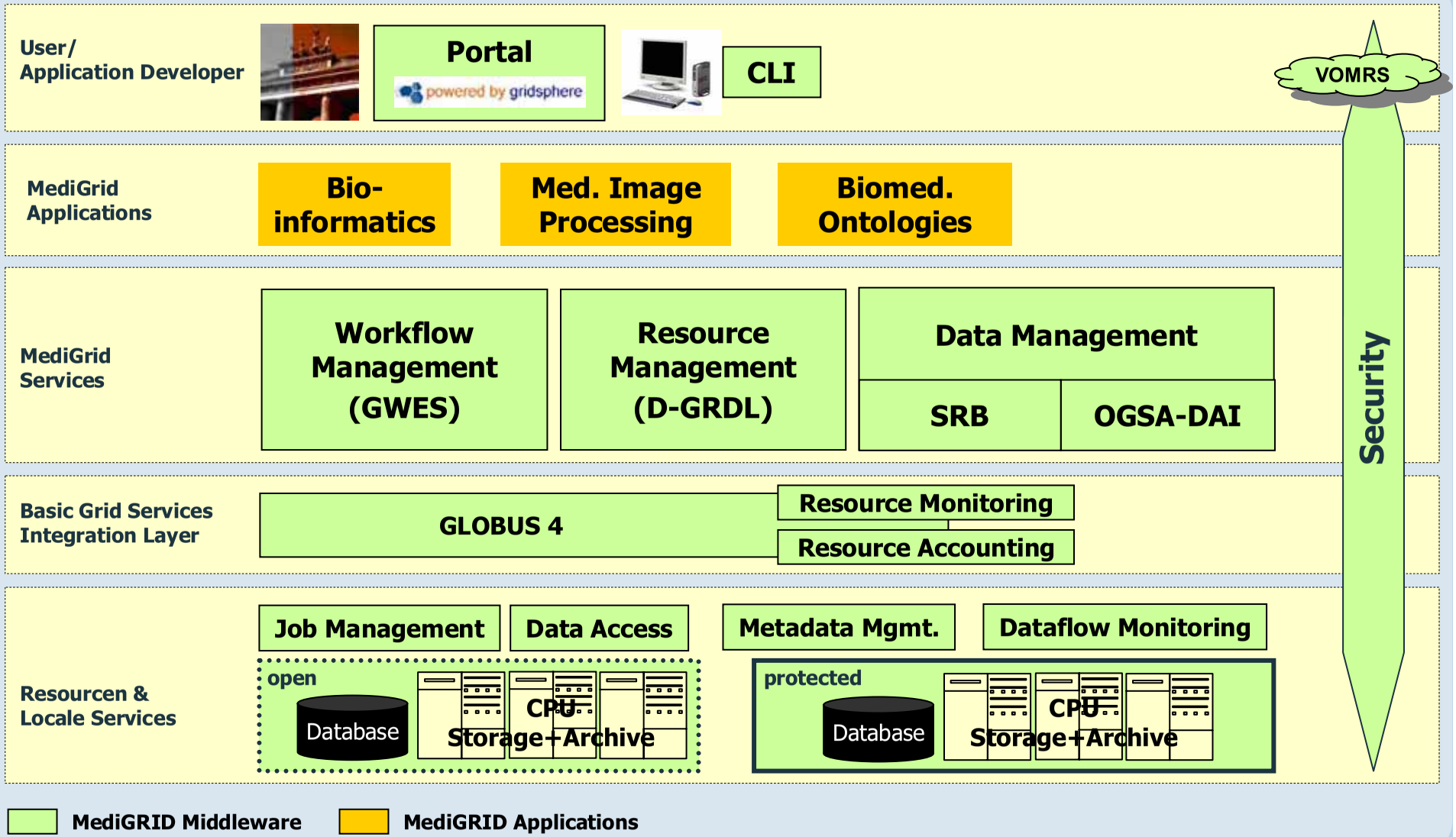
Why Using Grids in Medicine - Example: Medical Image Analysis

- **today:**
  - decentralized data repositories (Federalism)
    - e.g. global accessible research databases planned
    - e.g. each state in Germany operates its own mammogram database
  - research and medical care are disconnected
- **tomorrow:**
  - vast amount of data
    - **today: 7 TByte/year** (in a 1000 beds hospital)
    - **tomorrow: 6000 TByte/year**
  - linking academic and clinical research

→ **Objectives of MediGRID**:

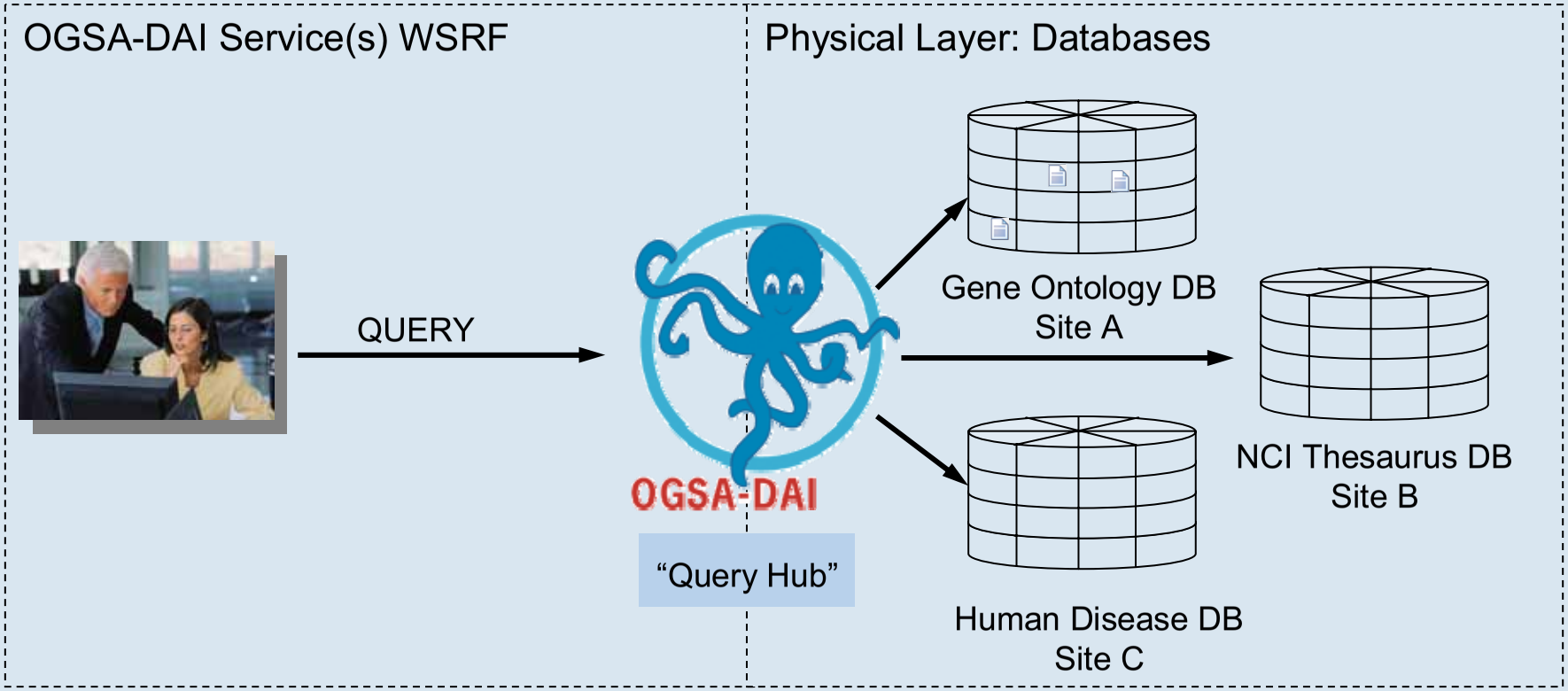
- infrastructure for collaborative platforms in medical research
- migration of pilot applications into the Grid
- establishing sustainable operational structures for applications

MediGRID Software Architecture



Database Access in MediGRID via OGSA-DAI

Open Grid Service Architecture - Data Access and Integration



End user deals with query site (OGSA-DAU service) only, virtualization from physical storage location(s)

caBIG™ is an innovative bioinformatics program at the NIH's National Cancer Institute

50 Cancer Centers are working towards a common goal of integrated data, tools and methodologies to accelerate cancer research goals at the National Cancer Institute for Bioinformatics (NCICB), the cancer Biomedical Informatics Grid (caBIG™)

The goal of caBIG™ is to create a virtual web of interconnected data, individuals, and organizations which will:

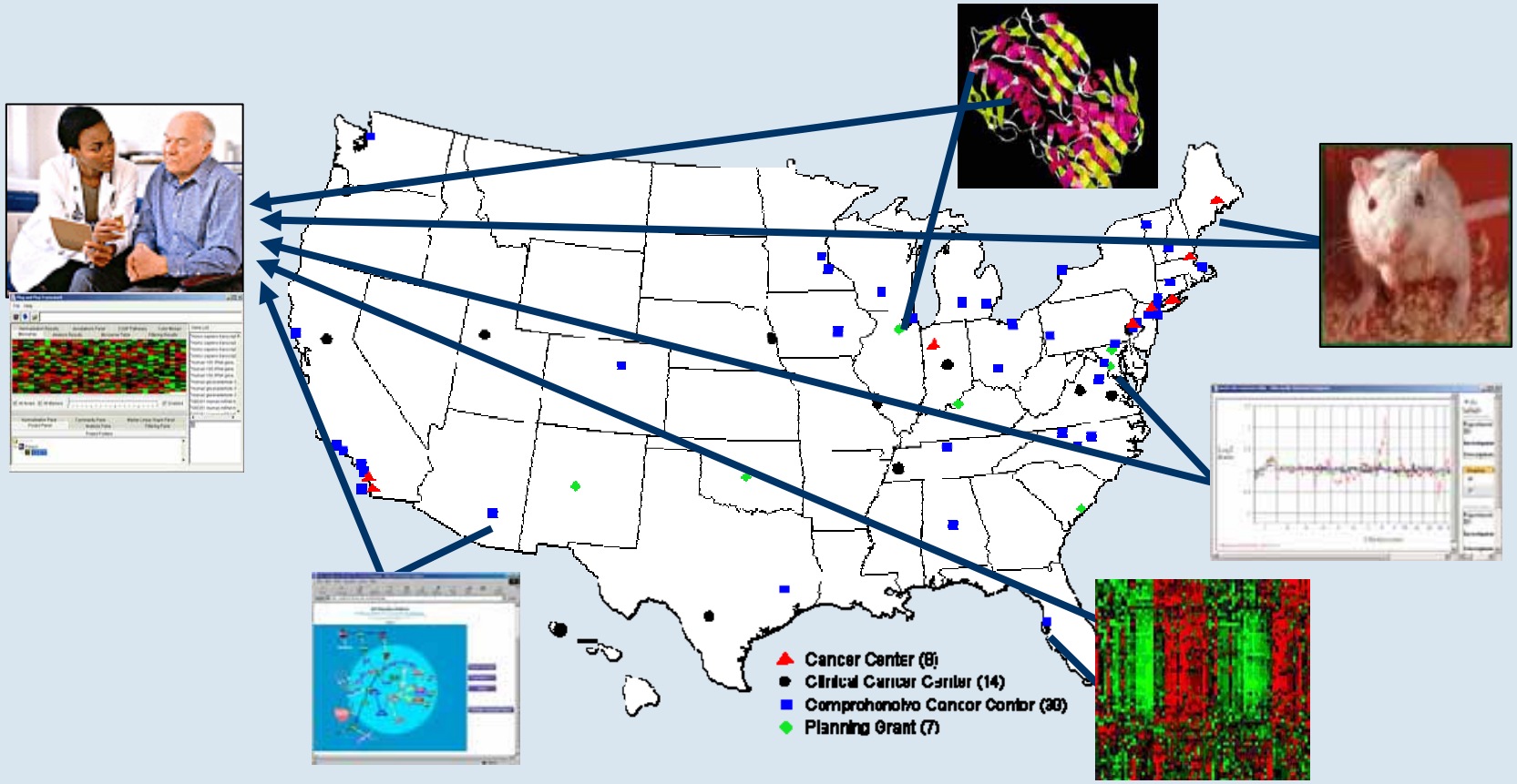
- redefine how research is conducted
- care is provided
- patients / participants interact with the biomedical research enterprise

The principles driving caBIG™ are:

- Open Source
- Open Access
- Open Development
- Federated Model

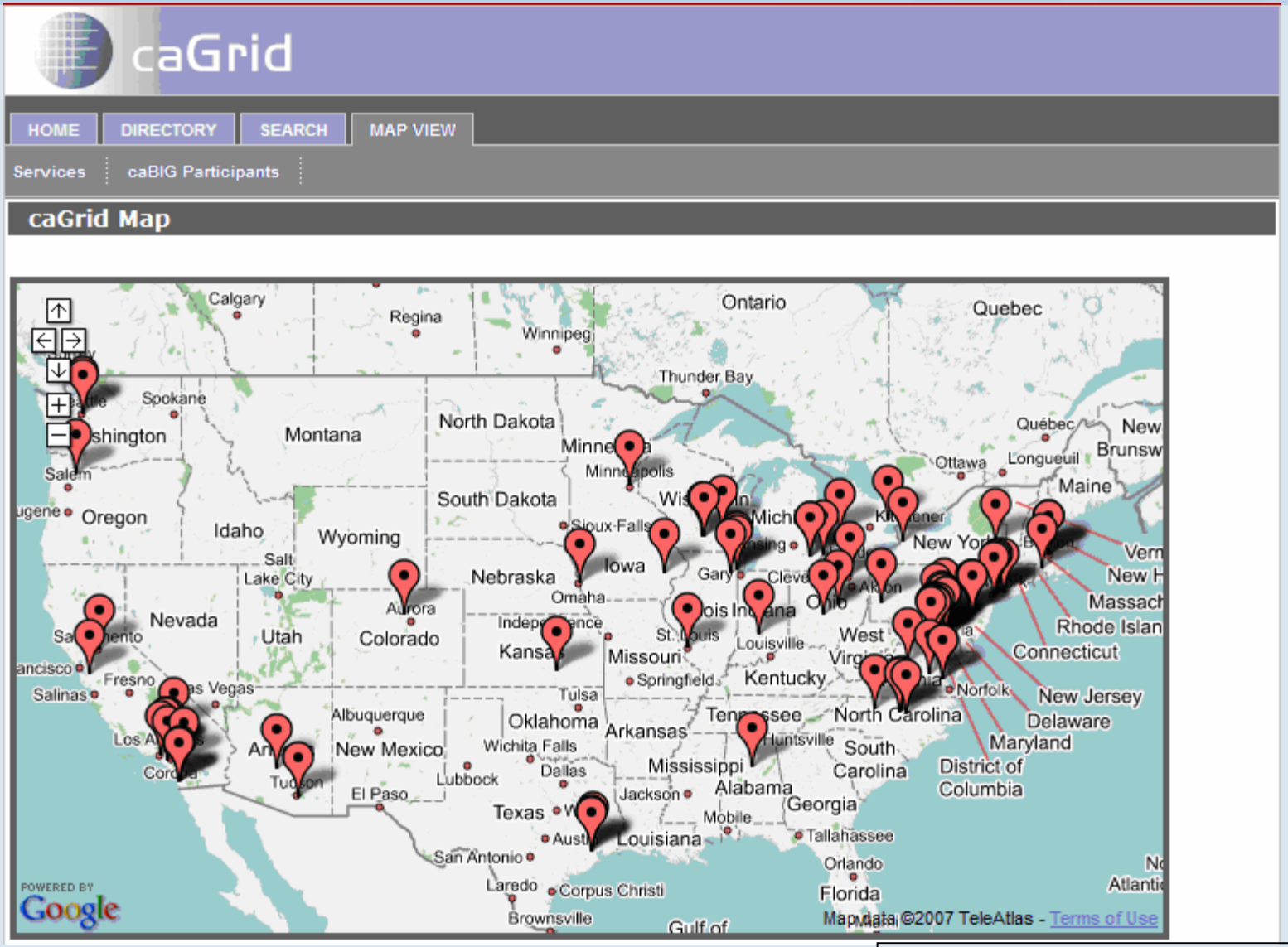
source: R. Mark Adams, caBIG™

caBIG will facilitate sharing of infrastructure, applications, and data across multiple cancer research programs



source: R. Mark Adams, caBIG™

...and the realization of that goal



source: R. Mark Adams, caBIG™

Four Domain Workspaces and two Cross Cutting Workspaces have been launched

DOMAIN WORKSPACE 1
Clinical Trial Management Systems

addresses the need for consistent, open and comprehensive tools for clinical trials management.

DOMAIN WORKSPACE 2
Integrative Cancer Research

provides tools and systems to enable integration and sharing of information.

DOMAIN WORKSPACE 3
Tissue Banks & Pathology Tools

provides for the integration, development, and implementation of tissue and pathology tools.

DOMAIN WORKSPACE 4
Imaging

provides for the sharing and analysis of in vivo imaging data.

CROSS CUTTING WORKSPACE 1
Vocabularies & Common Data Elements

responsible for evaluating, developing, and integrating systems for vocabulary and ontology content, standards, and software systems for content delivery

CROSS CUTTING WORKSPACE 2
Architecture

developing architectural standards and architecture necessary for other workspaces.

source: R. Mark Adams, caBIG™

caGrid Metadata Infrastructure Goals

Support *strongly typed* grid

- Syntactic and Semantic interoperability
  - **Programmatic!**
- Smooth transition from Application to Grid and back

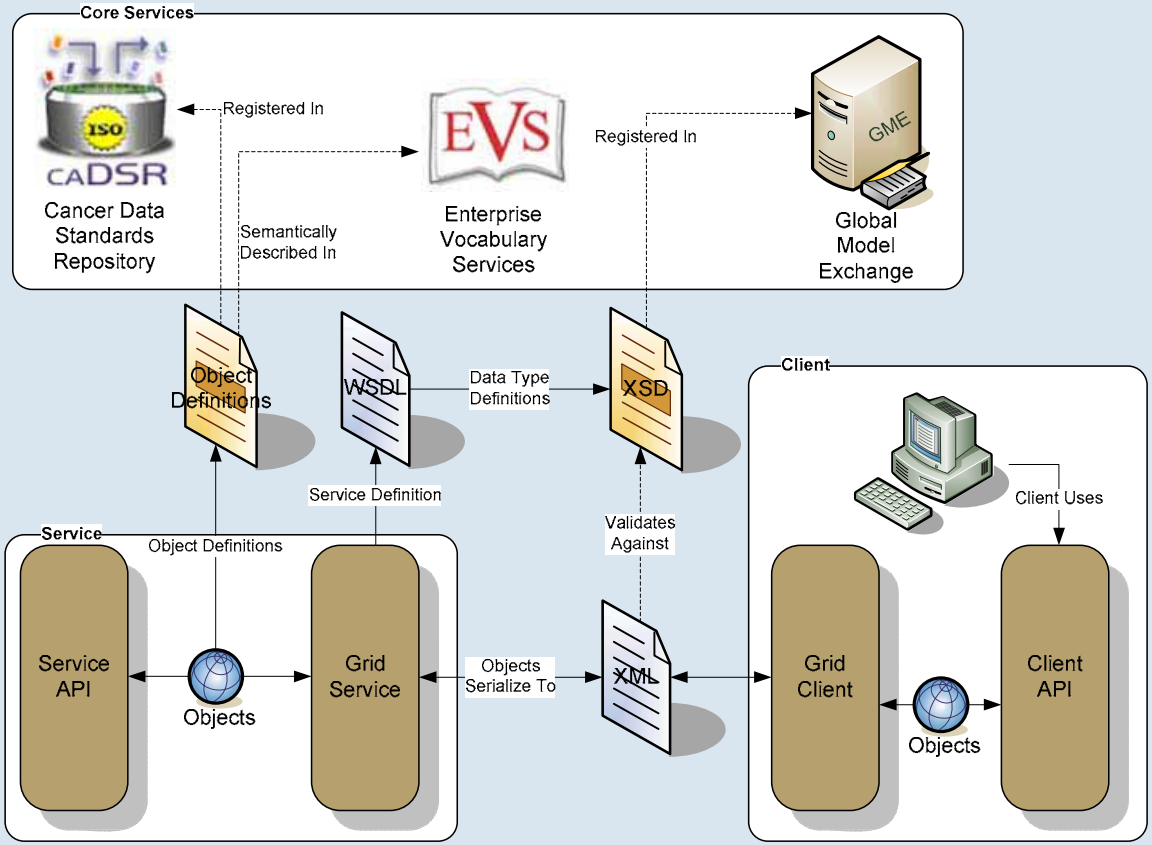
Leverage wealth of existing metadata

Enable service Advertisement and Discovery

source: Scott Oster, caBIG™

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)

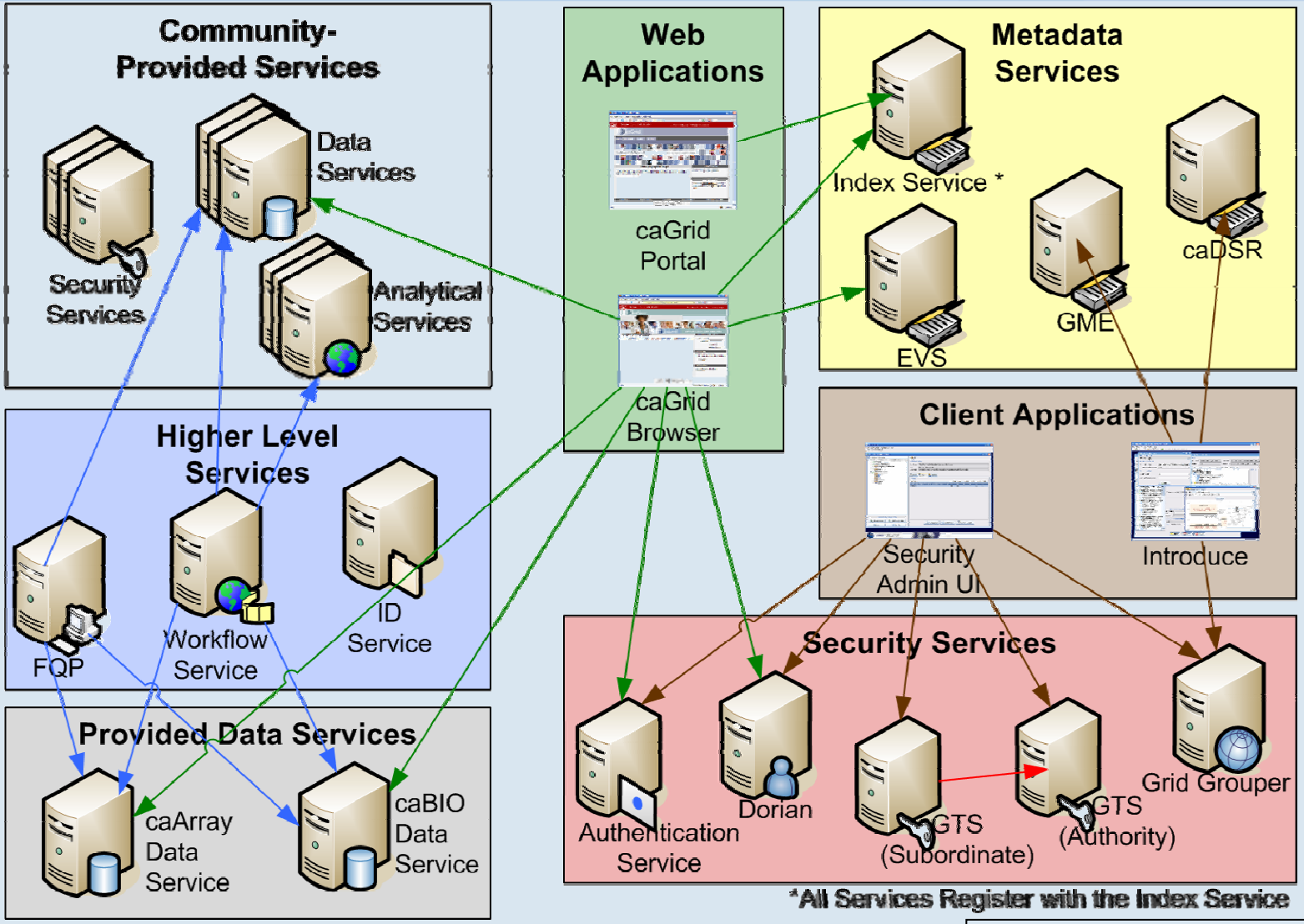


source: Scott Oster, caBIG™

caGRID Technology Standards

- WSRF Service Layer
- Common Service Metadata
  - Provided by all services
  - Details service's capabilities, operations, contact information, hosting research center
  - Service operation's inputs and outputs defined in terms of structure and semantics
- Data Service Metadata
  - Provided by all data services
  - Describes the Domain Model being exposed, in terms of a UML model linked to semantics
  - Provides information needed to formulate the Object-Oriented Query
- Service Security Metadata

caGrid Production Environment



source: Scott Oster, caBIG™

The Need for Interoperability

- solving complex problems requires the interaction of many experts
- sharing of resources, data, knowledge
- at a global scale world-wide

data can "travel" much easier around the globe than human beings

→ **standards, standards, ...**

Middleware Standards in MediGRID

WSRF & W3C standards:

- Grid Middleware Services: Globus TK 4 (Globus Alliance)
- Data Virtualization: OGSA-DAI (UK eScience)
- X.509 Certificates
- User Portal / Portlets: JSR168 (GridSphere project)

Community & de-factor standards:

- Data Virtualization: SRB (SDSC)
- Workflows: Grid Workflow Description Language (FhG)

The Best of caBIG™ - Our Perspective

- „bottom-up" approach
- strict certification → compatible guide

Compatibility Match MediGRID – caBIG™

Common technologies:

- Grid MW: Globus
- Database virtualization: OGSA-DAI

Challenges for MediGRID:

- Ontology: infrastructure to support of strongly typed data objects at all levels (services, user data, …)
- Security infrastructure:
  - centralized (D-Grid) vs. federated Grid Trust Service (GTS)
- Workflows: BPEL vs. GWDL

Questions?

MediGRID Related Talks @ eHealth

- Interoperability in Grid Computing
- Integrating Portals for Collaboratories and Grid Services
- Ontology Applications at MediGRID
- Importance of HL7 – Terminology TC for HealthGrids
- Cost Analysis of Grid Infrastructure
- Research Projects in Biomedicine
- Research Projects in Image Analysis
- Research Projects in Clinical Medicine

More MediGRID Information at:

http://www.medigrid.de/