GRID COMPUTING IN HEALTHCARE

*a primer and case study*

**caBIG** (TM) The National (USA) Cancer Institute’s Biomedical Informatics Grid project

Wednesday 17 June 2009
The Redwood Health Information Collaborative

DR STUART TURNER
Health Informaticist | stuart@leafpath.org | +1.916.596.0255

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License
What is Grid Computing?
Grid computing
Grid computing

“coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations”

- Beyond document or message exchange (e.g. HL7)
- More than the prevailing distributed computing methodologies (middleware, REST, SOAP, WSDL etc.)
- Coordinated and integrated direct access to systems - their unique or specialized data and analytical services and other resources
  - coordinates resources that are not subject to centralized control
  - using standard, open, general-purpose protocols and interfaces
  - to deliver non-trivial qualities of services
- Allows for novel discovery and interaction with new resources
- Beyond defining syntactic exchange methods, grid computing better enables rich semantic integration of resources.
- User-centric

* Ian Foster, et al Anatomy of the Grid - Enabling Scalable Virtual Organizations

1 Health Level Seven (HL7) 2 Representational State Transfer (REST), 3 Simple Object Access Protocol (SOAP)
4 Resources include computational, storage, network, source code, vocabulary or ontological, catalogs, etc.
Grid computing

Computing cluster

Distributed grid resources
Grid computing

Think about the following...

Heterogeneity
Collaboration
Cooperation
Novel discovery
User centric
Semantically aware
Distributed, federated
Resource sharing
Virtual organization
Harnessing distributed expertise
Reuse
Data aggregation
The Parts: Architecture

The **Grid Fabric**

The grid fabric is a layer that mediates shared access (publish/consume) to a diverse array of resources that for the most part are agnostic to the underlying (local/internal) applications.

*Figure 1: From Ian Foster, et al. Anatomy of the Grid - Enabling Scalable Virtual Organizations*

*Figure 2: The layered Grid architecture and its relationship to the Internet protocol architecture. Because the Internet protocol architecture extends from network to application, there is a mapping from Grid layers into Internet layers.*
The Parts: Architecture

The **Globus Toolkit**

- An open source toolkit to support grid computing
- Multiple functions for authentication, authorization; managing credentials; data replication and management, resource allocation management, information services, global indexing, resource discovery, reusable libraries, common runtime services, etc.

Case Study

caBIG - Cancer Biomedical Informatics Grid

- Initiative under the National Cancer Institute Center for Biomedical Informatics and Information Technology (CBIIT)
- Large-scale, multi-institutional, multi-year effort
- Researchers, clinicians, and patients
- Mitigate challenges in resolving rapid growth and diversity of clinical and research data collection, management and analysis
- 46 NCI-designated Cancer Centers and 16 Community Cancer Centers (Aug 2008)
- > 1,000 software developers from ~ 200 organizations
- Evolving suite of caBIG Tools
caBIG

Federated
- Based on Globus
- caBIG Tools
- caCORE SDK
- Enterprise Vocabulary Services

Open-development
- Collaborative tools
- Object-oriented
- Unified Modeling Language

Open-access
- Vocabulary Knowledge Center
- GForge Portal (subversion)

Open-source
caBIG Model License (not OSI-approved)

Open-standards
ISO and others
caBIG Workspaces and Special Interest Groups

**Domain Workspaces**

- **Clinical Trials Management Systems**
  - C3PR (Study participant registration)
  - Patient Scheduling (PSC)
  - caAERS (Adverse Event Management and Reporting)
  - caXchange (Laboratory Exchange)
  - NCI Enterprise Services (Protocol, person and organization management)

- **Integrative Cancer Research**
  - Bioconductor
  - caArray
  - caIntegrator
  - caBIO
  - caB2B
  - caMOD
  - caNanoLab
  - caTRIP
  - DWD
  - caFE
  - GeneConnect
  - GenePattern
  - geWorkbench
  - GOMiner
  - gridPIR
  - ProteomicLIMS

- **In Vivo Imaging**
  - NBIA
  - caIMAGE
  - XIP
  - AVT
  - AIM
  - Middleware
  - Query Formulation
  - DICOM Ontology
  - caMicroscope

- **Tissue Banks and Pathology Tools**
  - caTissue Core
  - caTissue Suite
  - caTIES

**Cross-Cutting Workspaces**

- **Architecture**
  - caCORE SDK
  - caGRID

- **Vocabulary and Common Data Elements**
  - CDE Browser
  - UML Model Browser
  - Vocabulary Knowledge Center
  - Vocabulary (terminology/ontology) Standard
  - Governance and Review Process
  - Compatibility Review
  - Common Data Element Standards Review
  - Enterprise Vocabulary Services
  - CTCAE
  - Ontology Development Projects
  - BiomedGT (Semantic Mediawiki)
  - LexEVS
  - LexBIG

---

Wednesday, June 17, 2009
Information sharing/exchange comparison in healthcare

Scenario 1

What is known:
A reference laboratory has made an explicit data sharing agreement with a partner hospital information system. The LIS and CIS are closed (i.e. no API’s) proprietary systems and their information models or knowledge representation methods are unknown and therefore not relevant.

The message is syntactically correct (explicit) with a minimal level of semantics (presumed) embedded in the message (e.g. encoding to a terminology, adherence to datatypes, etc.).

The LIS pulls relevant data elements to create an HL7 message and the CIS consumes this message, parses it, and instantiates it into its data model.

What is unknown if not uncertain
The data elements now stored in each system faithfully represents the source concepts, relationships, its context, character encoding, etc.
Information sharing/exchange comparison in healthcare

Scenario 2

NBIA - National Biomedical Imaging Archive

Open licensing, documentation, etc.

Source Code

UML Model

AIM - Imaging Markup and Annotation
Interoperability

- Controlled vocabularies and ontologies
- Strict adherence to metadata standards (ISO 1117)
- Compatibility Reviews (“certification”) for all applications/models that will be exposed on the grid
- Certification of vocabularies (ontologies)
- Development of domain analysis models (DAMs)
  - References models (think HL7 RIM)
  - “Backbone models”
- Common Data Element Standards Process
  - Fierce loyalty to reuse of CDE’s and UML models. Reuse of existing standard CDE’s for “Gold Level” Compatibility
- Automated processes to help developers build, harmonize and certify their applications
  - Semantic Integration Workbench (SIW)
  - Mentoring
  - Introduce Toolkit
ISO 11179

ISO/IEC JTC1 SC32 WG2 Development/Maintenance

Last Update: 2009-03-25

ISO/IEC 11179, Information Technology -- Metadata registries (MDR)

The 11179 standard is a multipart standard that includes the following parts:

- Part 1: Framework
- Part 2: Classification
- Part 3: Registry metamodel and basic attributes
- Part 4: Formulation of data definition
- Part 5: Naming and identification principles
- Part 6: Registration

11179-1: Framework

This part of ISO/IEC 11179 introduces and discusses fundamental ideas of data elements, value domains, data element concepts, conceptual domains, and classification schemes essential to the understanding of this set of standards and provides the context for associating the individual parts of ISO/IEC 11179.

Project Editor: Dan Gilligan
Common Data Elements (CDE’s)
Cancer Data Standards Repository (ISO11179 Metadata Compliant)
Cancer Data Standards Repository (caDSR)
Common Data Element and UML Model Browsers
Bioportal
NCI Thesaurus (at Bioportal)
Thanks
Questions?

Note: This slideware is a draft. An updated version with more annotations and references will be made available.