

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**

Co-sponsored by Redwood MedNet & Mendocino County Public Health.

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

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

- *Beyond* document or message exchange (e.g. HL7¹)
- *More than* the prevailing distributed computing methodologies (middleware, REST², SOAP³, 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*

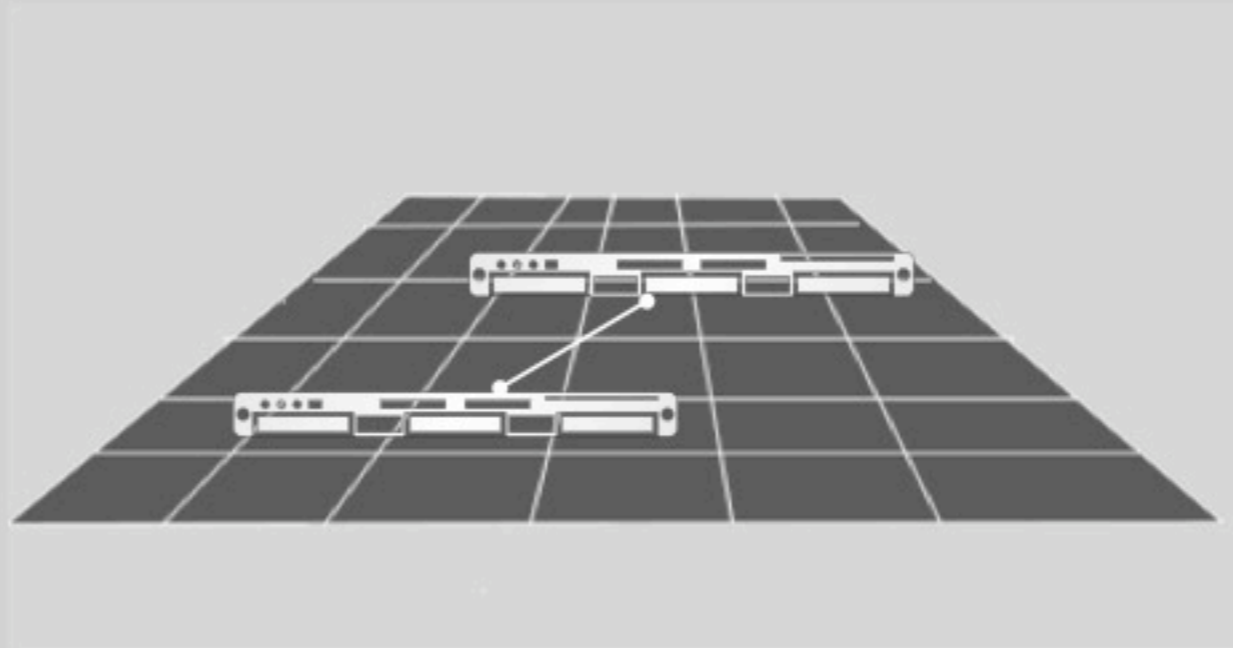
¹ Health Level Seven (HL7) ² Representational State Transfer (REST), ³ Simple Object Access Protocol (SOAP)

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

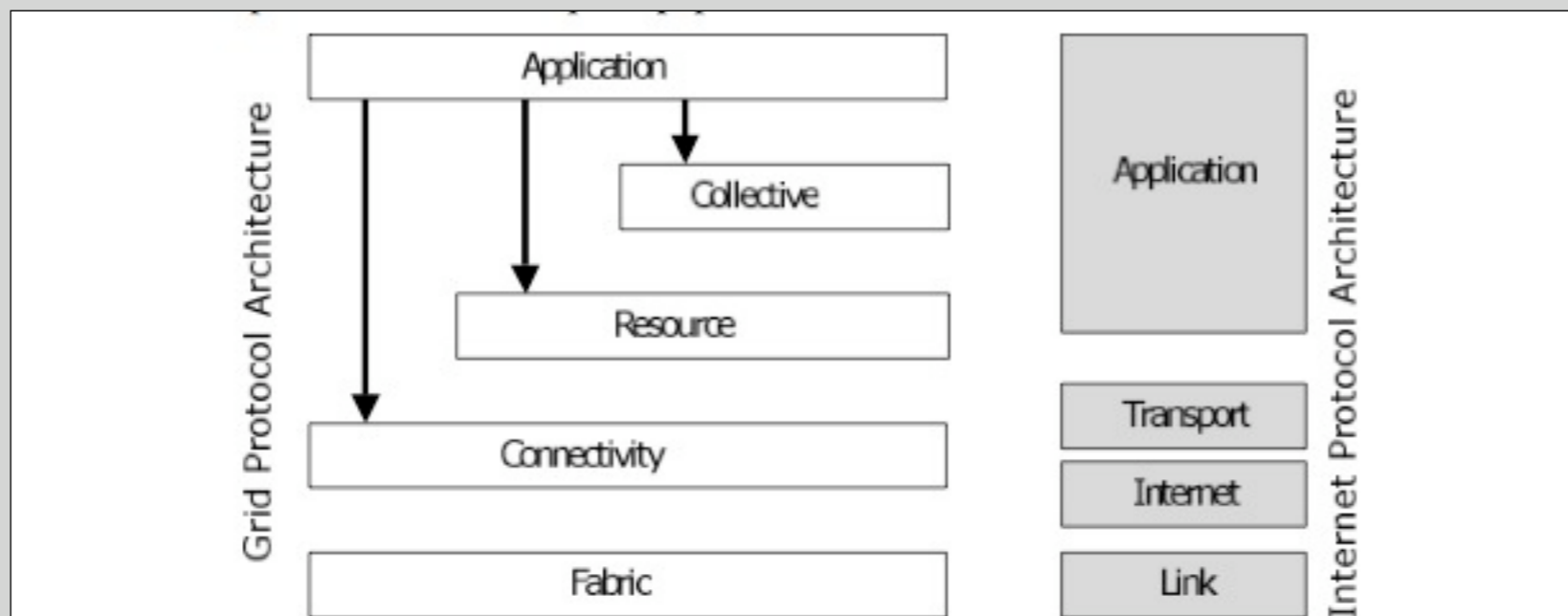


Figure 2: The layered Grid architecture and its relationship to the Internet protocol architecture. Because the Internet protocol architecture extends from network to application, there is a mapping from Grid layers into Internet layers.

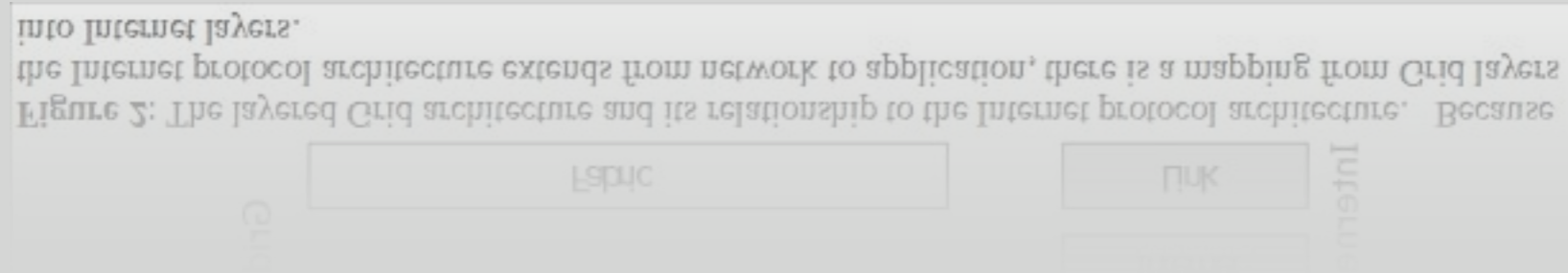


Figure 1: From Ian Foster, et al *Anatomy of the Grid - Enabling Scalable Virtual Organizations*

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.

See Foster I, *Globus Toolkit Version 4: Software for Service-Oriented Systems*, J. Comput. Sci. & Technol; July 2006, Vol. 21, No. 4, pp 513-520. More resources available at the end of this slideware document

Case Study

caBIG - Cancer Biomedical Informatics Grid

The screenshot shows the caBIG Community Website homepage. On the left is a navigation menu with categories like 'Getting Connected', 'Getting Support', 'About caBIG', and 'Workspaces & SIGs'. The main content area features a 'Welcome to the caBIG® Community Website' message, followed by a 'caBIG® Spotlight' section with three items: 'Solving Basic and Clinical Research Challenges in Cancer and Beyond', 'caBIG® Enterprise Support Network', and 'Upcoming Events'. A 'Featured Articles' section is also visible. On the right, there are promotional banners for the 'caBIG 2009 Annual Meeting' and 'Biomedical Informatics Without Borders' conference, along with a 'caBIG® News' section listing recent updates.

- Initiative under the National Cancer Institute Center for Biomedical Informatics and Information Technology (CBIIIT)
- 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 Language Services

Open-development

- Collaborative tools
- Object-oriented
- Unified Modeling

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

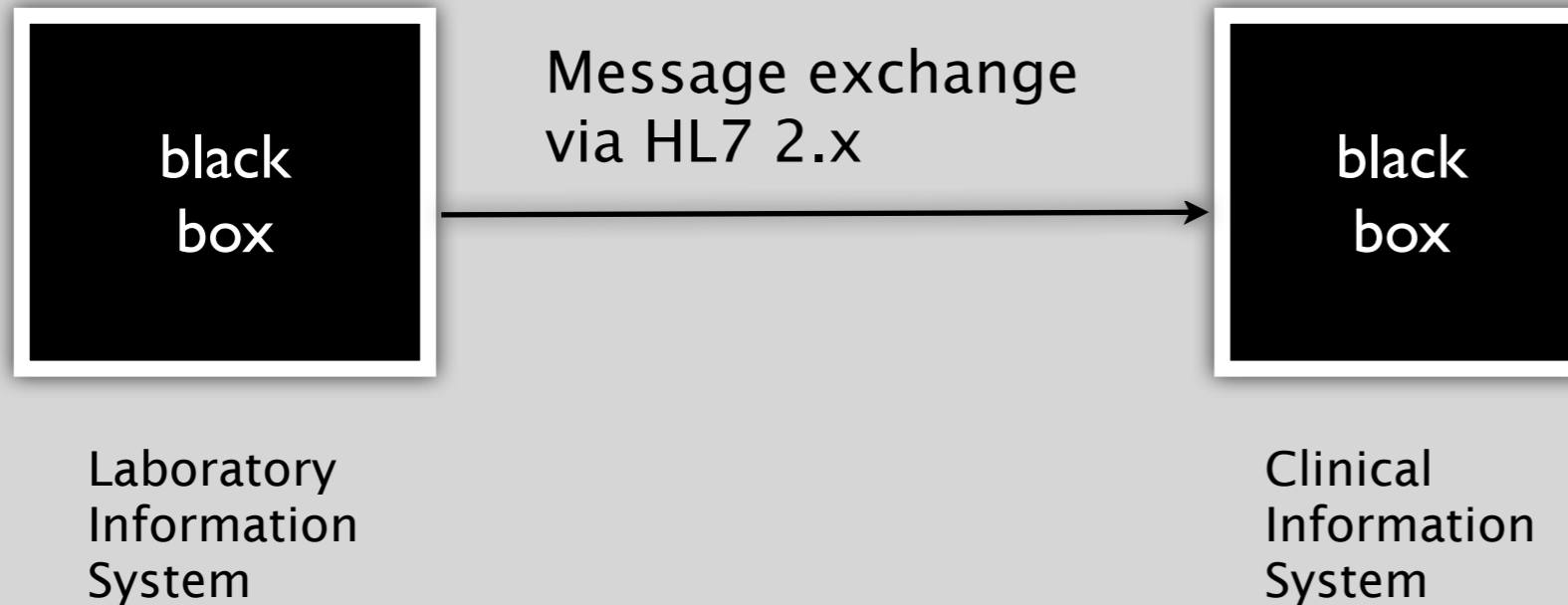
caCORE SDK
caGRID

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

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.).

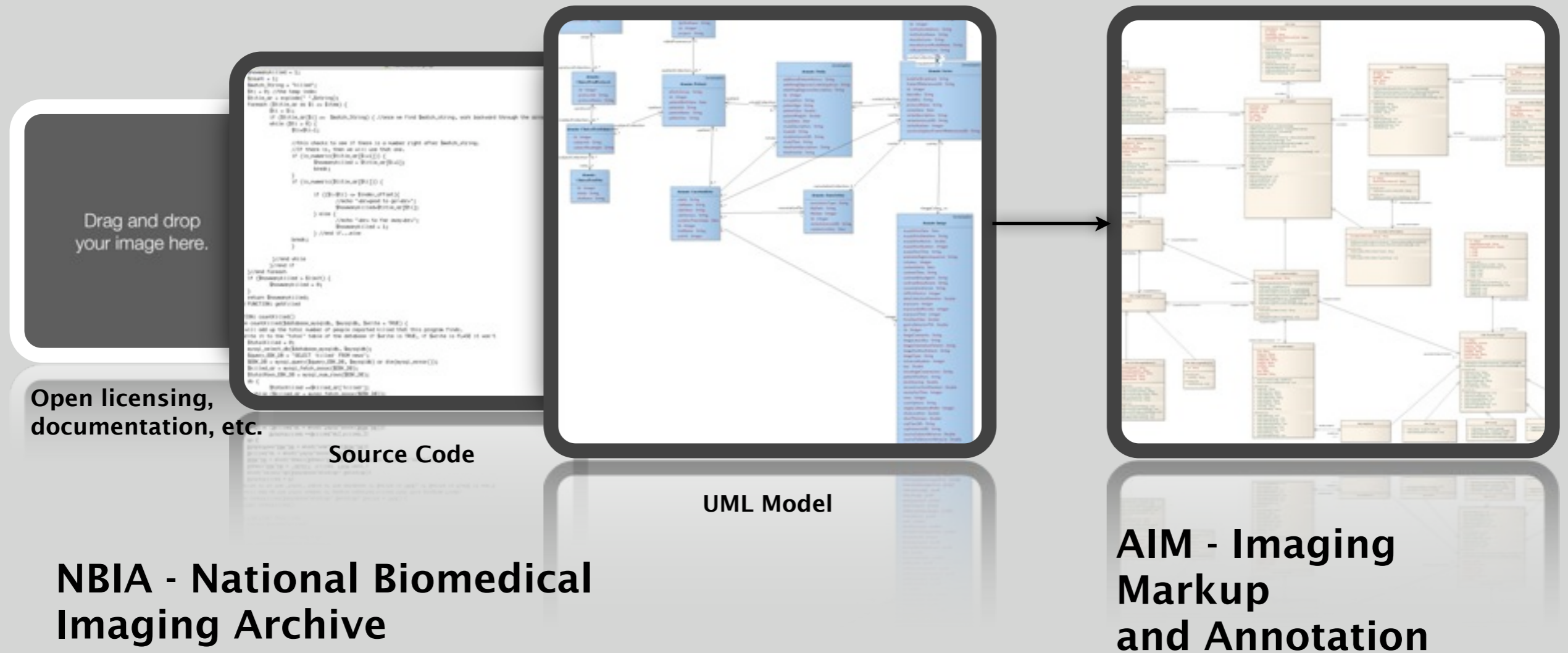
The LIS pulls relevant data elements to create an HL7 message and the CIS consumes this message, parses it, and instantiates it into its data model.

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



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 GILLMAN

Project Editor: Dan GILLMAN

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.

11179-1: Framework

Long Name	Preferred Question Text	Owned By	Used By Context
Agent Administration Potency Unit of Measure Unified Code for Units of Measure Code	Unit	caBIG	
Date	Date	caBIG	NIDCR
Day Date	Day Date	DCP	NIDCR,caBIG
Derived Person Age Value	The calculation of age based on the equation of Date of Event-Date of Birth.	caBIG	
Diet Performance Status Assessment Intake Scale	Normalcy of Diet Rating	CTEP	
Event DateTime	Date/Time of an event	caBIG	
Event Description Type	Description of a type of Event	caBIG	
Laboratory Test Result Unit of Measure Unified Code for Units of Measure Code	Lab Test Unit of Measure UCUM Codes	caBIG	
Month Date	Month Date	DCP	NIDCR,caBIG
Person Birth Date	Patient's Date of Birth	CTEP	caBIG
Person Gender Text Type	Gender of a Person	caBIG	CCR,CIP,DCP,NHLBI,SPORE
Person Height Unit of Measure Unified Code for Units of Measure Code	Height Unit of Measure UCUM Code	caBIG	NHLBI
Person Language Preference Rank Type	Rank of language used by a person	caBIG	
Person Weight Unit of Measure Unified Code for Units of Measure Code	Weight Unit of Measure UCUM Code	caBIG	NHLBI
Self Reported Person Sex Text Type	Sex of Person	caBIG	
Speech Performance Status Assessment Comprehension Scale	Understandability of Speech Rating	CTEP	
Year Date	Year Date	DCP	NIDCR,caBIG

Data Element Details

Public ID:	793
Version:	5.1
Long Name:	Person Birth Date
Short Name:	2193335v1.0:2018550v1.0
Preferred Question Text:	Patient's Date of Birth
Definition:	the month, day and year on which the person was born.
Value Domain:	Date
Data Element Concept:	Person Birth
Context:	CTEP
Workflow Status:	RELEASED
Origin:	
Registration Status:	Standard

Reference Documents

Document Name	Document Type	Document Text	Context	URL
CRF Text	Alternate Question Text	Patient Birth Date	CTEP	
Patient's Date of Birth	Preferred Question Text	Patient's Date of Birth	CTEP	
Source	DATA_ELEMENT_SOURCE	Data element and all values comply with Clinical Data Update System (CDUS) v3.0 Release 2; for CDUS reporting, only submit the year and month of the patient's birth (format: YYYYMM), do not submit the day of birth.	CTEP	http://cte
FBPP Question Text	Alternate Question Text	Date of Birth	NHLBI	
2007 DICOM Standard	REFERENCE	2007 DICOM Standard; Parts 3, 5 & 6; Detailed in Part 3 paragraph C.2.3 Patient Demographic Module	caBIG	http://me
Patient DOB	Alternate Question Text	Patient DOB	caBIG	

Common Data Elements (CDE's) Cancer Data Standards Repository (ISO11179 Metadata Compliant)

Cancer Data Standards Repository (caDSR) Common Data Element and UML Model Browsers

The screenshot shows the National Cancer Institute (NCI) website header with the logo and the text "National Cancer Institute" and "U.S. National Institutes of Health | www.cancer.gov". Below the header is the "CDE Browser" title and navigation links: "Admin Tool", "Curation Tool", "NCI Metathesaurus", "NCI Terminology Server", "Sentinel Tool", "UML Model Browser", and "What's new". There are also icons for "CDE Cart", "Home", "FormBuilder", and "Help".

On the left side, there is a "Refresh tree" button and a tree view of "caDSR Contexts" with expandable folders for various programs like BRIDG, caBIG, caBIG Data Standards, caCORE, CCR, CDISC, CIP, CTEP, DCP, EDRN, HITSP, and NCRI.

The main content area is titled "Data Element Search" and "Search for Data Elements". It includes a search input field, a "Name" dropdown menu, and radio buttons for search criteria: "Exact phrase" (selected), "All of the words", and "At least one of the words". A tip and note are provided below the search options.

Data Element Search

Search for Data Elements [Search preferences](#) [Advanced search](#)

caDSR Contexts

Exact phrase
 All of the words
 At least one of the words

Tip: This is an exact match search. To search for partial words or phrases use the * as a wildcard.
Note: Default settings exclude Test and Training Context views from the tree and certain 'non-released' Workflow and Registration statuses. Click the 'Search Preferences' link above to view or change the exclusion criteria. Search Preferences will be reset to default settings when the 'New Search' button is clicked on the search results page or 'caDSR Context' in the Tree.

This is a mirrored, inverted version of the screenshot above, showing the same interface elements from the caDSR CDE Browser.

Bioportal

The screenshot shows the BioPortal website interface. At the top is a navigation bar with the BioPortal logo and links for Browse, Search, Projects, Annotate, All Mappings, All Resources Alpha, Sign In, Register, Help/About, and Send Feedback. Below the navigation bar are three search boxes: 'Search all ontologies' with a 'Search' button and 'Advanced Search' link; 'Find an ontology' with an 'Explore' button and 'Browse Ontologies >' link; and 'Search resources' with a 'Search' button and 'Advanced Resource Search' link. The main content area is divided into three columns. The left column, 'Most Active Ontologies', contains a table with columns for Ontology, Version, Notes, and Mappings. The middle column, 'Latest Notes', contains two entries about EKG P Wave (Electrocardiography (ECG) Ontology) with dates and author names. The right column, 'Latest Mappings', contains two entries showing relationships between 'cellular component (Biological process)' and 'Cell component (Foundational Model of Anatomy)' with dates and author names.

Most Active Ontologies

Ontology	Version	Notes	Mappings
Human disease	1.39	0	17732
Mouse adult gross anatomy	1.194	0	3905
NCI Thesaurus	08.12d	9	3798

Latest Notes

[RE:Advice regarding the EKG P WAVE EKG P Wave \(Electrocardiography \(ECG\) Ontology\)](#) 06/16/09 whetzel

[Advice regarding the EKG P WAVE EKG P Wave \(Electrocardiography \(ECG\) Ontology\)](#) 06/15/09 sgranite

Latest Mappings

[cellular component \(Biological process\) => Cell component \(Foundational Model of Anatomy\)](#)
04/27/09 matthiassamwald

[Cell component \(Foundational Model of Anatomy\) => cellular component \(Biological process\)](#)
04/27/09 matthiassamwald

This block shows a faint, mirrored reflection of the BioPortal interface content shown above, appearing as a watermark or ghost image.

NCI Thesaurus (at Bioportal)

BioPortal | Browse | Search | Projects | Annotate | All Mappings | All Resources Alpha | Sign In | Register | Help/About | Send Feedback

NCI Thesaurus | Search

NCI Thesaurus Version 08.12d | Abnormal Cell | Link Here | Subscribe

View Ontology Details

Jump To: Go

- Abnormal Cell
- Activity
- Anatomic Structure, System, or Substance
- Biochemical Pathway
- Biological Process
- Chemotherapy Regimen or Agent Combination
- Conceptual Entity
- Diagnostic or Prognostic Factor
- Diagnostic, Therapeutic, and Research Equipme
- Disease, Disorder or Finding
- Drug, Food, Chemical or Biomedical Material
- Experimental Organism Anatomical Concept
- Experimental Organism Diagnosis

ID:	Abnormal_Cell
Full Id:	http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#Abnormal_Cell
Code:	C12913
Label:	Abnormal Cell
Full Syn:	Abnormal Cell PTNCI
Semantic Type:	Cell
Definition:	NCIA cell that occurs in human diseases or in experimental models of human cancer.
Preferred Name:	Abnormal Cell
Nci Meta Cui:	CL031062

NCI Meta Cui: CL031062

Preferred Name: Abnormal Cell

Definition: NCIA cell that occurs in human diseases or in experimental models of human cancer.

Semantic Type: Cell

Thanks
Questions?

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