Opportunities and challenges to creating a federated global network of interoperable information systems

George A. Komatsoulis, Ph.D
Deputy Director
NCI Center for Biomedical Informatics and Information Technology
“A wonderful fact to reflect upon, that every human creature is constituted to be that profound secret and mystery to every other.”

Charles Dickens: *A Tale of Two Cities* 1859
Childhood Cancer: Dramatic Improvement in Survival

*5-year relative survival rates, based on follow up of patients through 2003.
High Participation in Clinical Research

Figure 1: Percentage of Childhood Cancer Patients on Protocol Treatment by Age and Type of Treatment Facility

- Pediatric Center
- Non-Pediatric Center

Winchester et al., CA Cancer J Clin 2001;51;119-130
Childhood ALL: Molecular Sub-types Impact Risk

Rubnitz and Pui, Oncologist (1997) 2:374-380
Childhood cancer is treated in a context that blends care delivery and clinical research

- Researchers and practitioners are able to correlate experimental laboratory data with clinical data (treatment, history, pathology, outcome, etc.)

- Clinical data are utilized to continuously evaluate outcomes

- Researchers develop and refine evidence-based strategies at an individualized level

- Care providers improve quality by adherence to care standards

Information flow is critical…
this model cannot be achieved without IT connectivity
The cancer Biomedical Informatics Grid (caBIG) was created by the US National Cancer Institute to enable the data sharing required by this molecular medicine paradigm.

caBIG Principles

- Open-development
- Open-access
- Open-source
- Transparency
- Federation
“And the Lord said, Behold, the people is one, and they have all one language; and this they begin to do; and now nothing will be restrained from them, which they have imagined to do. Go to, let us go down, and there confound their language, that they may not understand one another's speech.”

*Genesis, Chapter 11, 500-450 BCE*

“What we have here is… failure to communicate”

*Strother Martin in Cool Hand Luke 1967*
Interoperability: Defined

ability of a system to access and use the parts or equipment of another system

Syntactic interoperability

Semantic interoperability

Courtesy: Charlie Mead
## Four Pillars of Interoperability

<table>
<thead>
<tr>
<th>Maturity Model</th>
<th>Legacy</th>
<th>Bronze</th>
<th>Silver</th>
<th>Gold</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interface Integration</strong></td>
<td>- No Programming interfaces to the system are available. Only data transfer mechanisms such as XML can be read. - Some ad hoc data transfer mechanisms, such as FTP.</td>
<td>- Provide baseline* programmatic access to data. Data can be read from remote electronic sources or from common data servers. A standard format for data can be defined that allows the transfer of data from applications to other external data sources.</td>
<td>- Well-described API's that provide access to data objects. - System architecture separated into tiers and interoperable components.</td>
<td>- Interoperable with data grid architecture to be defined by caBIG. - Fully componentized. Provide access to individual resources in the form of grid services.*</td>
</tr>
<tr>
<td><strong>Vocabularies / Terminologies</strong></td>
<td>- Free text used throughout for all applications.</td>
<td>- Use of publicly acceptable standardized controlled vocabularies with public terminologies.</td>
<td>- Standard terminologies approved by public standards bodies or the caBIG. - caBIG CDE Workspace is used for all shared data-creation models.</td>
<td>- All features of Silver, plus: - Fully compliant with caBIG CDE and common standard models, services, and services and content exchanges.</td>
</tr>
<tr>
<td><strong>Data Elements</strong></td>
<td>- No Structured metadata is defined.</td>
<td>- Some type of metadata describing the information in the system is used for creation of metadata. Metadata is retrieved from an external repository shared by multiple applications. - Common Data Elements should be built using controlled terminology.</td>
<td>- Use common standard electronic representation for CDE's such as ISO 11179 or comparable standard. - Use the caBIG standard to represent information across the Domain Workspace. - Common Data Elements are built using standard controlled terminologies approved by public standards bodies or the caBIG Vocabulary/CDE Workspace.</td>
<td>- All features of Silver, plus: - Programmatic access to all metadata, including data class descriptions, site and source information, and any other caBIG-defined metadata requirements and use information models. - Use the caBIG standard or electronic representation of metadata and Common Data Elements.</td>
</tr>
<tr>
<td><strong>Information Models</strong></td>
<td>- No particular information model is used to represent data.</td>
<td>- Some type of diagrammatic model describing the data relationship is available in electronic format.</td>
<td>- Information models defined in a standard modeling language such as UML.</td>
<td>- All features of Silver, plus: - Information models are harmonized with other s across the caBIG Domain Workspace.</td>
</tr>
</tbody>
</table>

*Note: The asterisk (*) indicates that these features are optional and may not be present in all systems.
The caBIG Infrastructure

Data and Analytical Services
- caArray Service
- caGWAS Service
- Other Services

Metadata Services
- Index Service
- caDSR Service
- Vocabulary Service
- GME Service

Client Applications

Security Services
- Authentication Service
- Dorian
- GTS
- Grid Grouper

Higher Order Services
- Federated Query Processor
- Workflow Engine
- Delegation Service
30 Analytical Services
78 Data Services
108 Total Services
“Not less arduous must have been the task of marking the proper line of partition between the authority of the general and that of the State governments”

James Madison: *Federalist Papers #37*
January 11, 1788
Options for expansion

- Independent Grids
  - Weak center, powerful individuals
  - “Feudalism”

- One Grid
  - All Grids subsumed into one Grid
  - “The Borg”

- Federation
  - Local control of data, resources with central control over interoperability standards
  - “Grid of Grids”
Interoperation between Grids

A “Grid of Grids”
ἀνδρα μοι ἑννεπε, μοῦσα, πολύτροπον, ὤς μάλα πολλὰ πλάγχθη, ἔπει Τροίης ἱερὸν πτολίεθρον ἔπερσεν

‘Ομηρος: Οδύσσεια

“Tell me, O Muse, of the man of many devices, who wandered full many ways after he had sacked the sacred citadel of Troy”

Homer: The Odyssey (8th Century BCE)
Opportunities and Challenges

- **Technical**
  - Common Semantics across organizational and national boundaries
  - Authentication and Authorization Standards
  - Variations in interoperability standards
- **Legal**
  - Variations in rules surrounding privacy, reporting and patient protections
  - Global Trust Fabrics
- **Political**
  - Barriers to multinational agreements
Terminology and Metadata

- Differing standard terminologies
- Lack of community acceptance for standard terminologies
- Internationalization/Localization
- Graceful Evolution

The solution is both technical (federated terminology and metadata develop) and social (consensus based standards development)
Security Issues

• **Technical:**
  • Security infrastructure must be standards based and capable of accepting credentials from a wide variety of sources of identity

• **Legal:**
  • Security infrastructure must be capable of adapting to local laws and regulations surrounding identity management
  • Appropriate agreements are required to allow acceptance of credentials from multiple sources
Technical: GAARDS Security Infrastructure
Decision Tree for Privacy /Intellectual Capital Terms and Conditions

**Policy: Analysis Framework**

Data/Specimens

- **IP Value (Need for Protection)**
  - High
  - Medium
  - None/Low
  
  Examples: Is the data subject to a restrictive license? Is it related to an invention report you have or intend to file with your institution?

- **Data Sensitivity (Regulatory Status)**
  - Identifiable Data
  - Coded/Limited Data Set
  - De-Identified /Anonymized Data Set

- **IRB/ Institutional Restrictions (Policy/ Consent Limitations)**
  - Explicit Consent Limitations or Restrictions
  - Policy Limitations
  - Generic Registry or caGRID Permission

- **Sponsor Restrictions (Contract Terms & Conditions)**
  - Classified Research / Major Restrictions
  - Delays or Other Moderate Restrictions
  - No Restrictions

ANY of the following:
- high IP value
- high sensitivity data (e.g., PHI)
- significant IRB/consent restrictions
- major sponsor restrictions

ANY of the following:
- moderate IP value
- moderate sensitivity data (e.g., LDS)
- limited institutional or IRB policy restrictions
- moderate sponsor restrictions

ALL of the following:
- no IP value
- low sensitivity data
- no IRB restrictions
- no sponsor restrictions

Bi-Lateral or Multi-Lateral MTA

Standardized Click-Through Terms and Conditions

“EZ Pass” - General Website Terms of Use

Outside Institution or to caGRID?
Interoperability Standards

• Technical standards (even among interoperable systems) may vary.

• Assessing compatibility with technical standards is important for the user community, but does not scale if multiple organizations must certify interoperability levels.

• Cross-certification standards are essential to allow systems to flourish in a “Grid of Grids” environment.
Cross-certification

Alignment with caBIG™ certification scheme at top level
caBIG and NCRI (UK) Joint Activities

• **Technical collaboration**
  • NCRI Informatics Platform/caGrid development
  • Common portal technologies

• **Cultural collaboration**
  • Initial discussions of a US/UK trust fabric for cancer research
  • Cross-certification of compatibility
  • Joint endorsement of data standards

• **Scientific Collaboration**
  • Joint informatics research meetings
  • Imaging collaboration between the NCI and NCRI using the National Cancer Imaging Archive
Acknowledgements

• Ken Buetow, Ph.D.
• Peter Covitz, Ph.D.
• Leslie Derr, Ph.D.
• Mary Jo Deering, Ph.D.
• Wendy Patterson, J.D.
• Frank Hartel, Ph.D.
• Avinash Shanbhag
• Denise Warzel
• Charlie Mead, MD

• Robin Clark (NCRI-UK)
• Alan Hogg, Ph.D. (NCRI-UK)
• J. Max Wilkinson, Ph.D.
• Joel Saltz, MD, Ph.D.
• Scott Oster
• Stephen Langella
• Mark Adams, Ph.D.
• Marsha Young, J.D.

“the bravest are surely those who have the clearest vision of what is before them, glory and danger alike, and yet notwithstanding, go out to meet it.”

Thucydides, History of the Peloponnesian War, c. 400 BCE