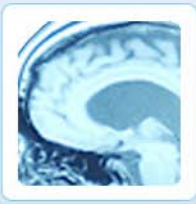
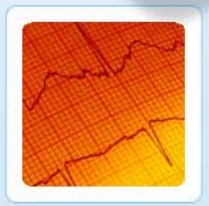
Health-e-Child

Health-e-Child:
A Platform for European Paediatrics

Tamás Hauer
University of the West of England, Bristol





HeC workshop, EGEE07

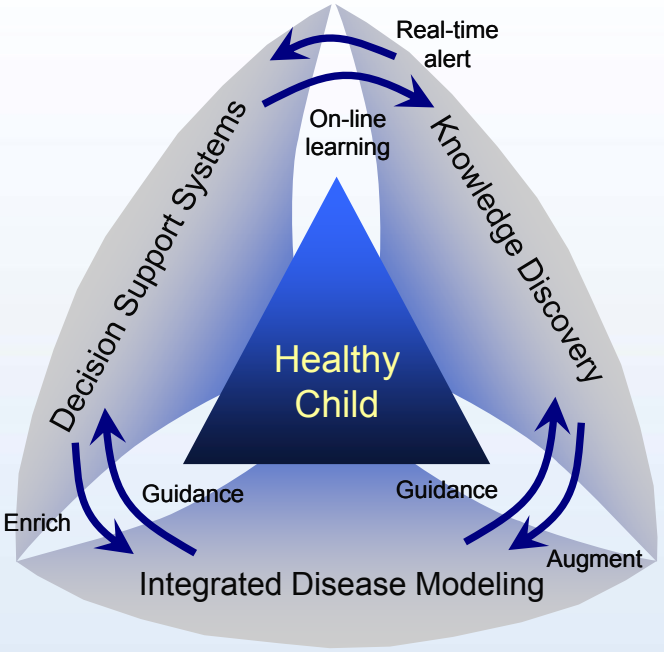
5 October 2007, Budapest

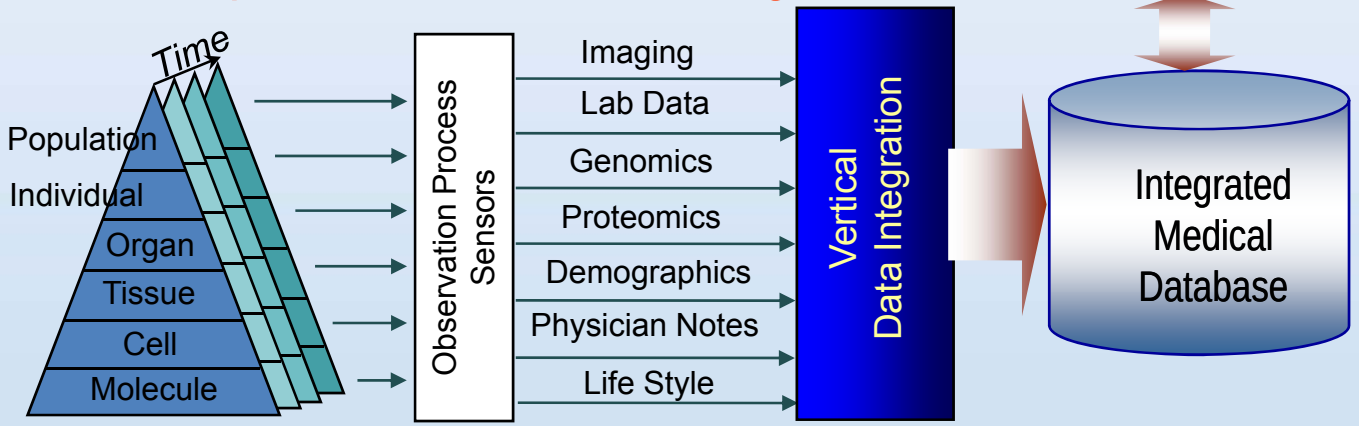
Motivation for the Project

- Clinical demand for integration and exploitation of heterogeneous biomedical information
  - vertical dimension – multiple data sources
  - horizontal dimension – multiple sites
- Need for generic and scalable platforms (Grid?)
  - integrate traditional and emerging sources
  - provide decision support
  - ubiquitous access to knowledge repositories in clinical routine
  - connect stakeholders in clinical research
- Need for complex integrated disease models
  - build holistic views of the human body
  - early disease detection exploiting in vitro information
  - personalized diagnosis, therapy and follow-up

Objectives of Health-e-Child

- Build enabling tools & services that improve the quality of care and reduce cost with
  - Integrated disease models
  - Database-guided decision support systems
  - Cross modality information fusion and data mining for knowledge discovery
- Establish multi-site, vertical and longitudinal integration of data, information and knowledge
- Develop a GRID-based platform, supported by robust search, optimisation and matching





What's unique about Health-e-Child?

- Paediatrics:
  - Temporal component
  - Some adult concepts do not (directly) apply, existing models might be misleading
  - Different examinations, treatments, some cannot be performed
  - Align with adult models (follow-up ?) ... Not in project scope
- Vertical Integration
  - Collect, represent and present the information, knowledge in an integrated way
  - Integration as a means of novel diagnosis/classification
- Extreme heterogeneity
  - Diseases, modalities, standards, interest...

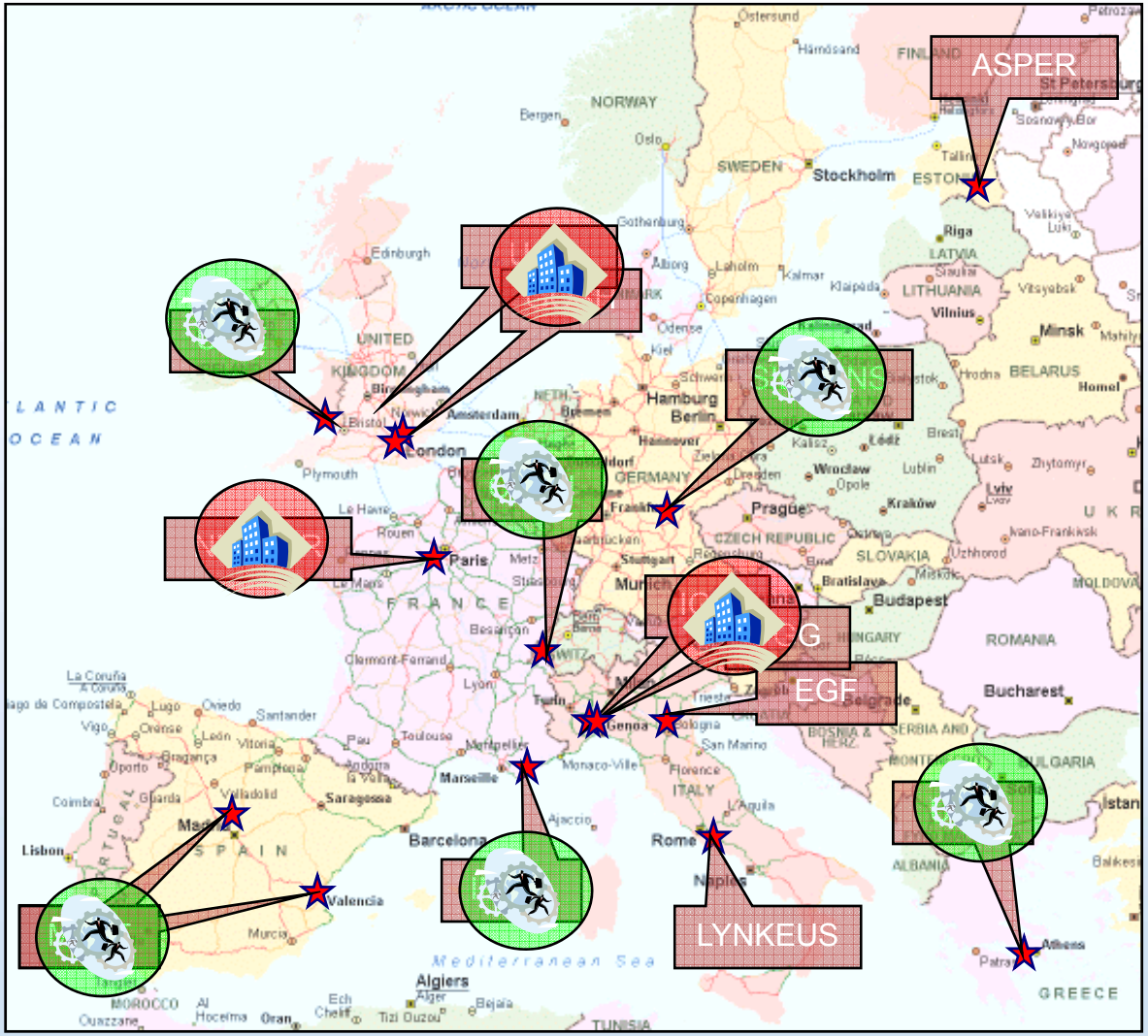
Focus on Paediatric Diseases

- Three Paediatric Diseases with at least partly unknown cause, classification and/or treatment outcomes
  - Heart diseases (*Right Ventricular Overload, Cardiomyopathy*)
  - Inflammatory diseases (*Juvenile Idiopathic Arthritis*)
  - Brain tumours (*Gliomas*)
- Many Clinical Departments
  - Cardiology
  - Rheumatology
  - (Neuro-)Oncology
  - Radiology
  - Lab (Genetics, Proteomics)
  - Administration, IT
- Main Modalities / Data Sources
  - Imaging (MR, US/echocardiography, CT, x-ray)
  - Clinical (Patient information, Lab results etc)
  - Genetics & Proteomics

A Geographically Distributed Environment

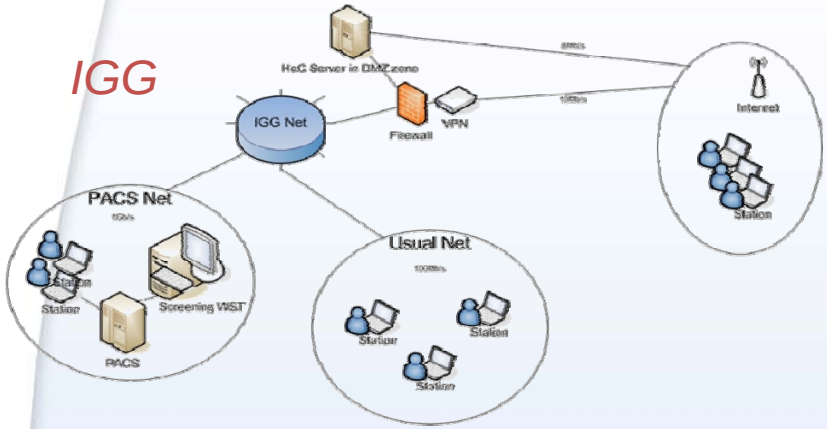
Clinical Site

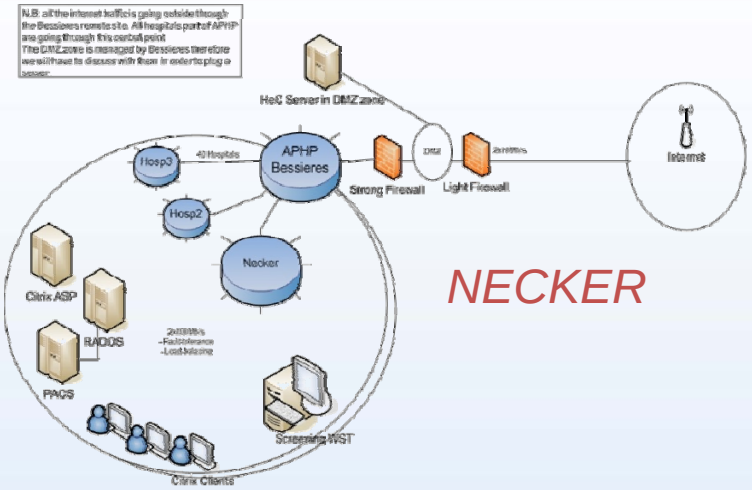
R&D Site



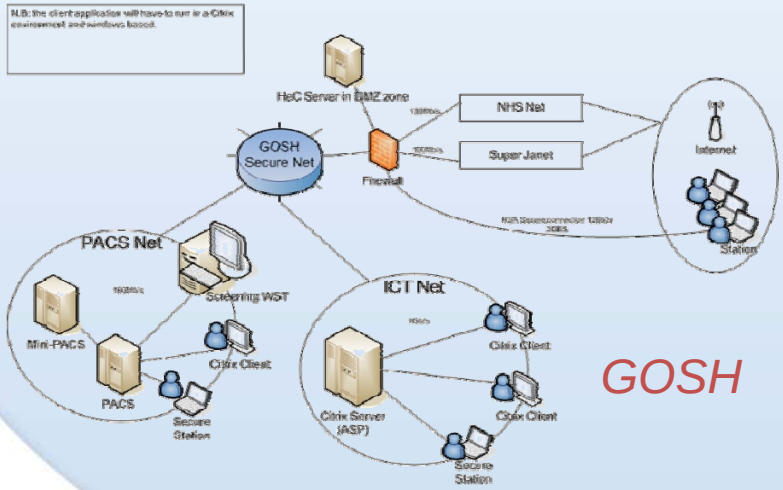
Integration Challenge: Applications

IGG





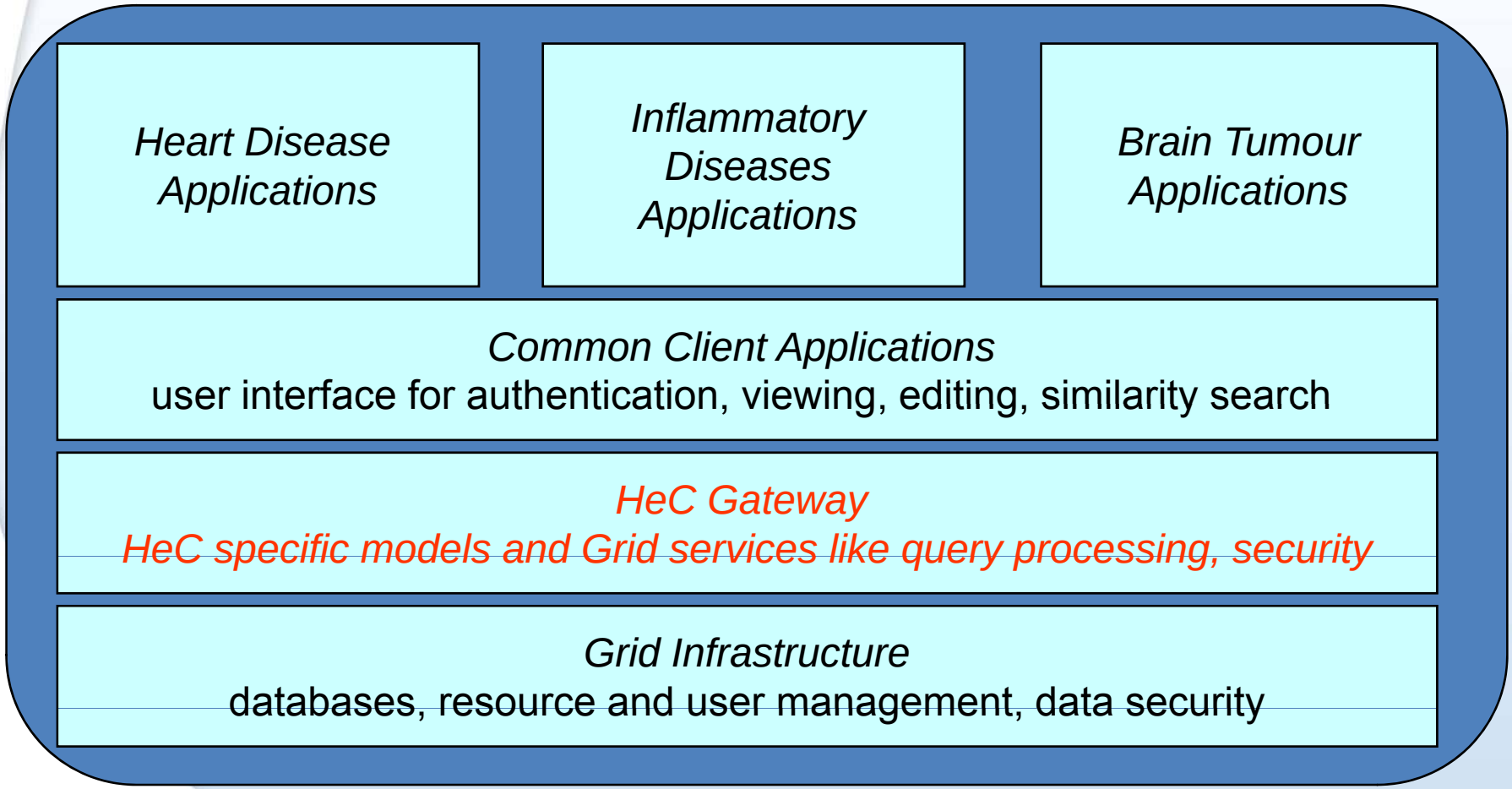
NECKER



GOSH

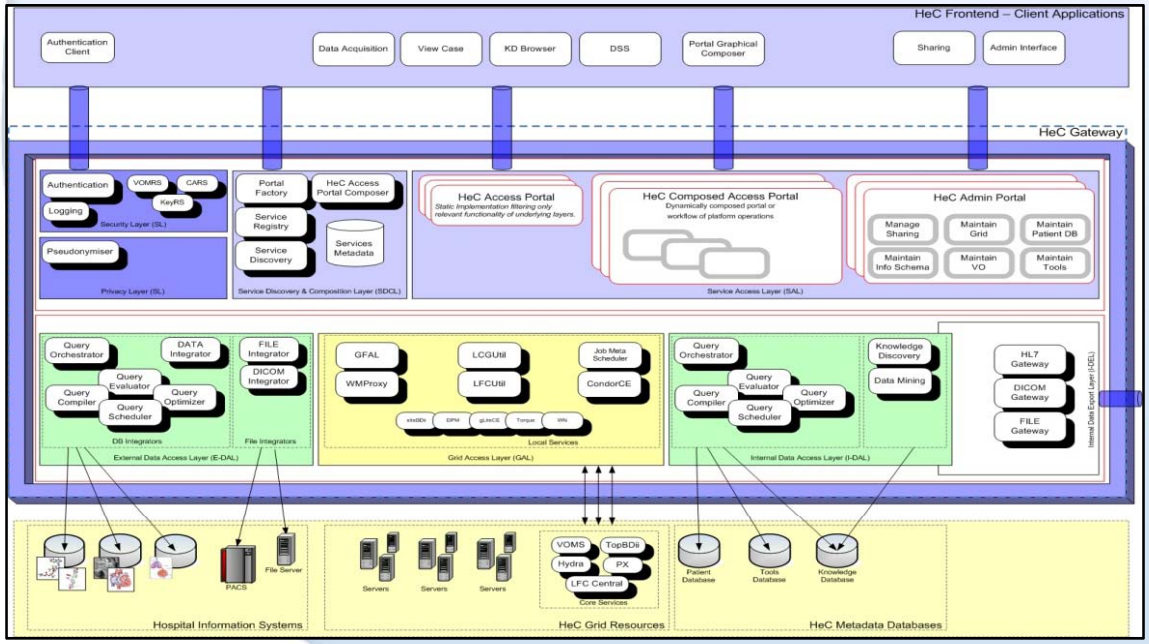
- Highlights
  - Different Networks: LANs, WANs, Internet
  - Security Constraints: Local & National Regulations
  - Bandwidth Limitations: LAN/WAN & Internet uplinks

HeC System Overview



Health-e-Child gateway

- The HeC Gateway
  - An intermediary access layer to decouple client applications from the complexity of the grid
  - Towards a platform independent implementation
  - To add domain specific functionality not provided by the middleware

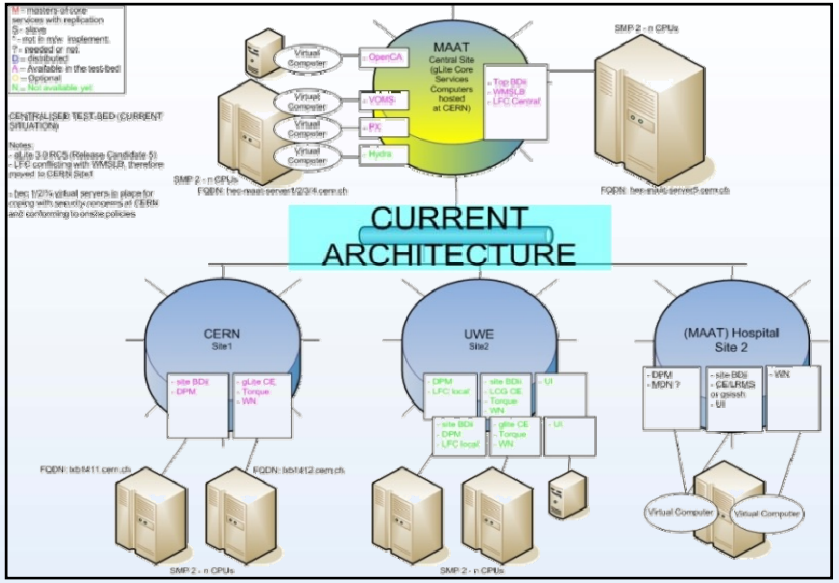


Status

- √ SOA architecture and design
- √ implementation of privacy and security modules

Architecture

- Grid technology (gLite 3.0) as the enabling infrastructure
  - A distributed platform for sharing storage and computing resources
- HeC Specific Requirements
  - Need support for medical (DICOM) images
  - Need high responsiveness for use in clinical routine
  - Need to guarantee patient data privacy:
    - ➢ access rights management
    - ➢ storage of anonymized patient data only



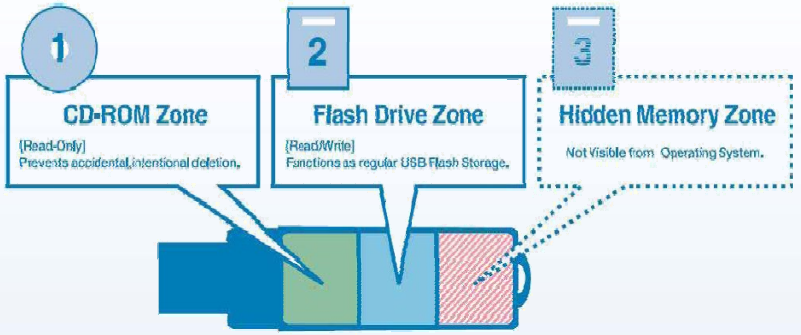
Status

- √ Testbed installation since May 2006
- √ HeC Certificate Authority
- √ HeC Virtual Organisation
- √ Security Prototype (clients & services)
- √ Logging Portal & Appender

Common Client Applications

USB key solution

- authentication
- basic functionality + applications that do not require specific resources
- available from any PC supporting USB in the hospital without SW installation
- autorun from Windows XP



Supports basic functionality

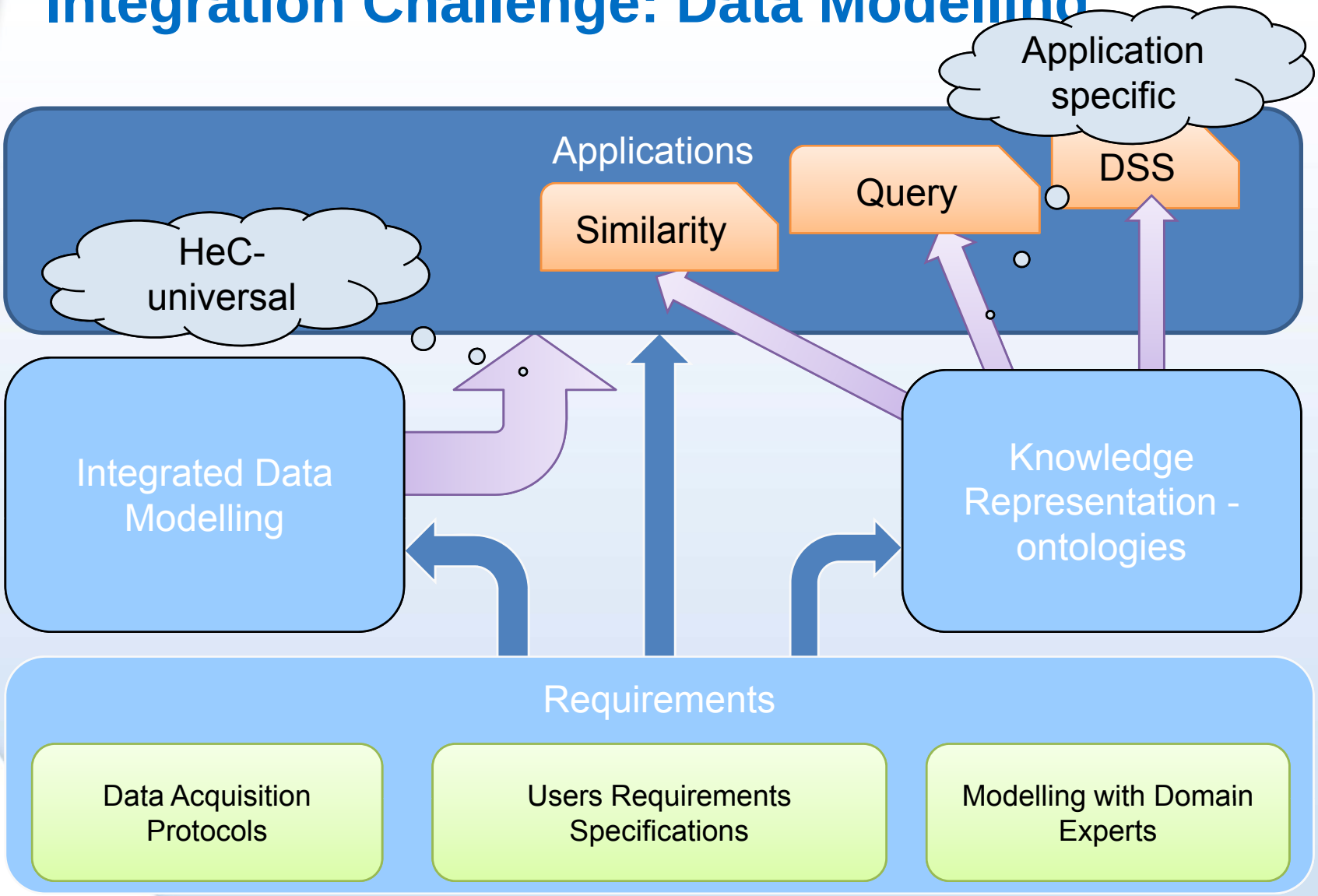
- browsing, viewing, editing patient data
- safety, security, privacy, anonymisation
- similarity search

Status

√ authentication (certificate-based single sign-on)

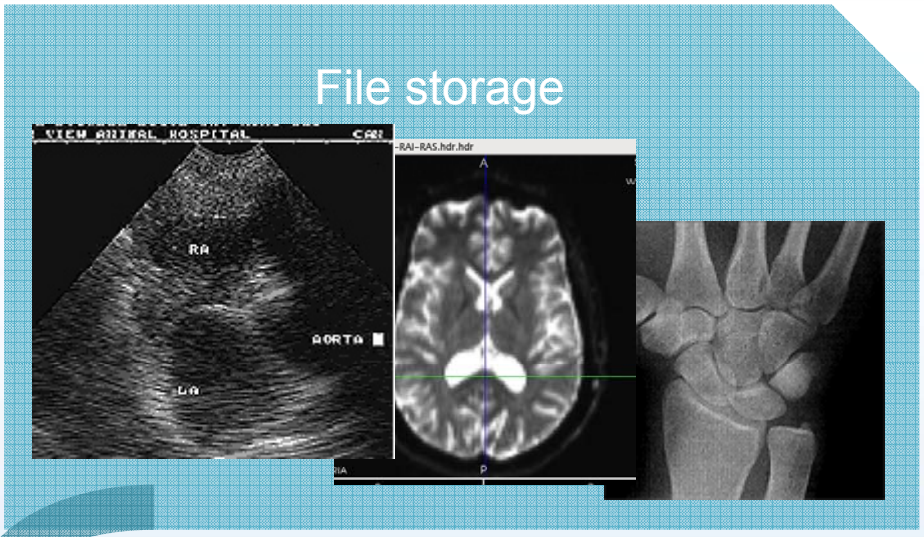
√ simple browsing and viewing

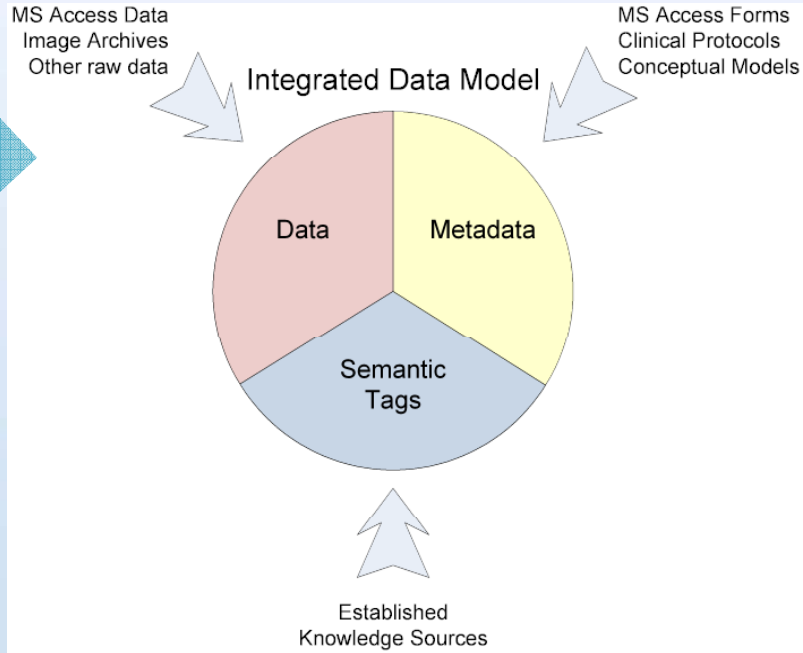
Integration Challenge: Data Modelling



Health-e-Child Data

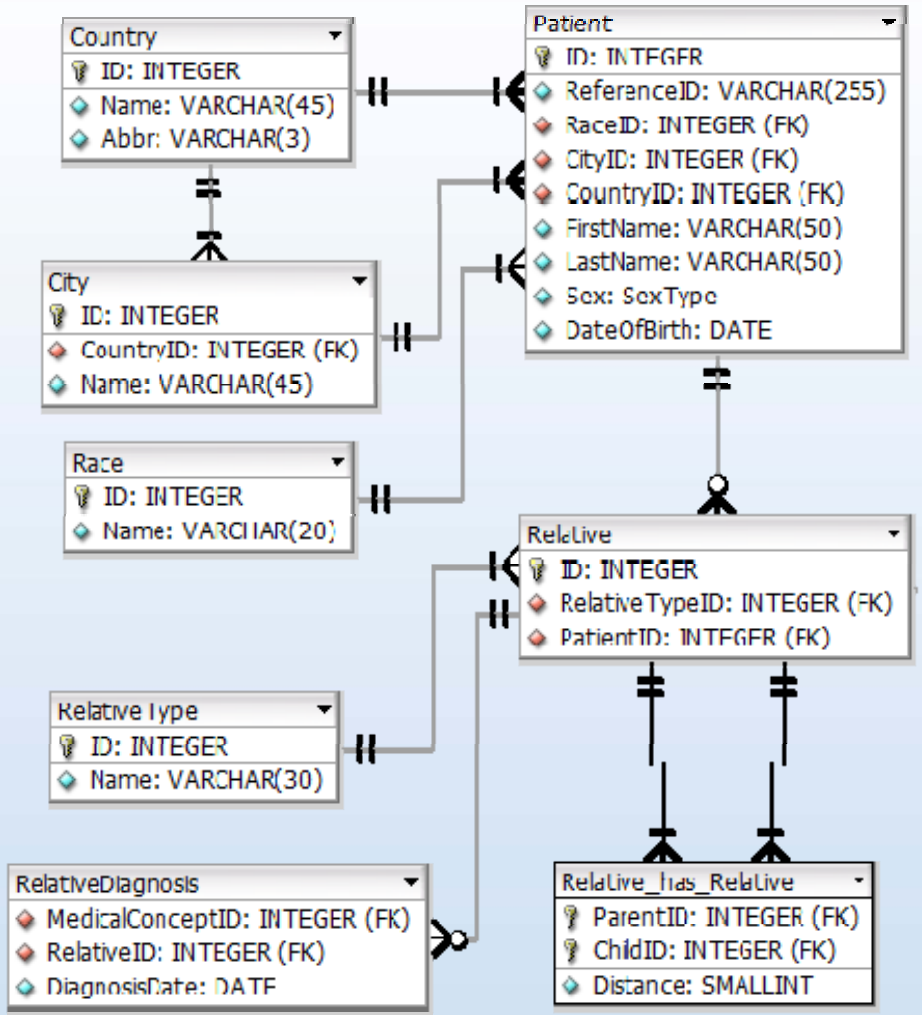
- Unstructured (file-based)
  - DICOM
    - Images (MRI, CT, x-ray)
    - Movies (US)
  - Molecular/Genetics data
- Semi-structured
  - Derived
  - Clinical data
  - Patient history
  - Diagnostics
  - Treatment
- Semantic annotations
  - Image annotations
  - Case annotations, Diagnosis
  - Links to external sources





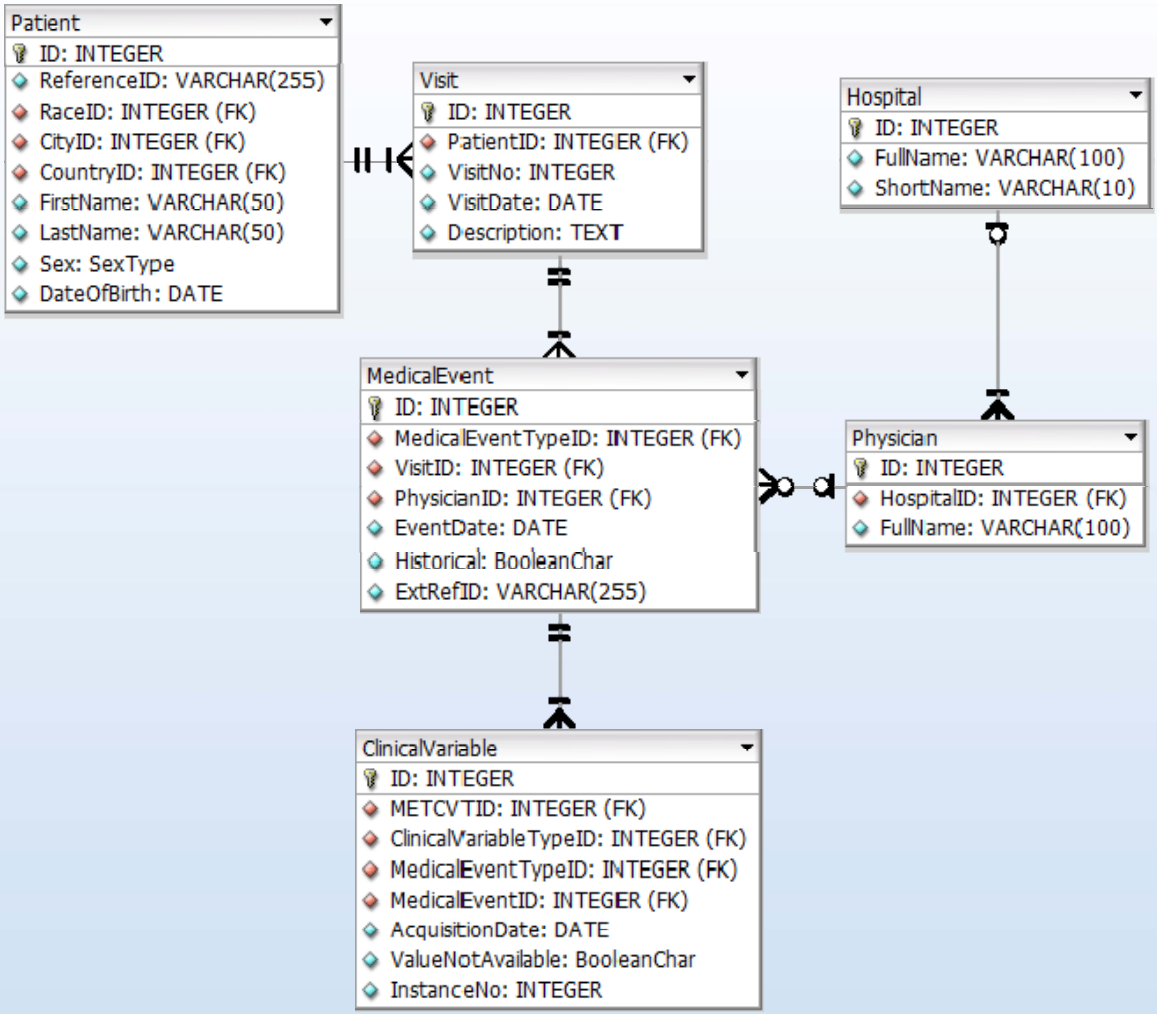
General Patient Information and Family History

- Patient
  - ReferenceID
- Family History
  - How to capture
    - Relative has/had a Disease
    - Disease in family
    - Pedigree up to 3 predecessors
  - Original vs Derived data
  - Incomplete, missing data



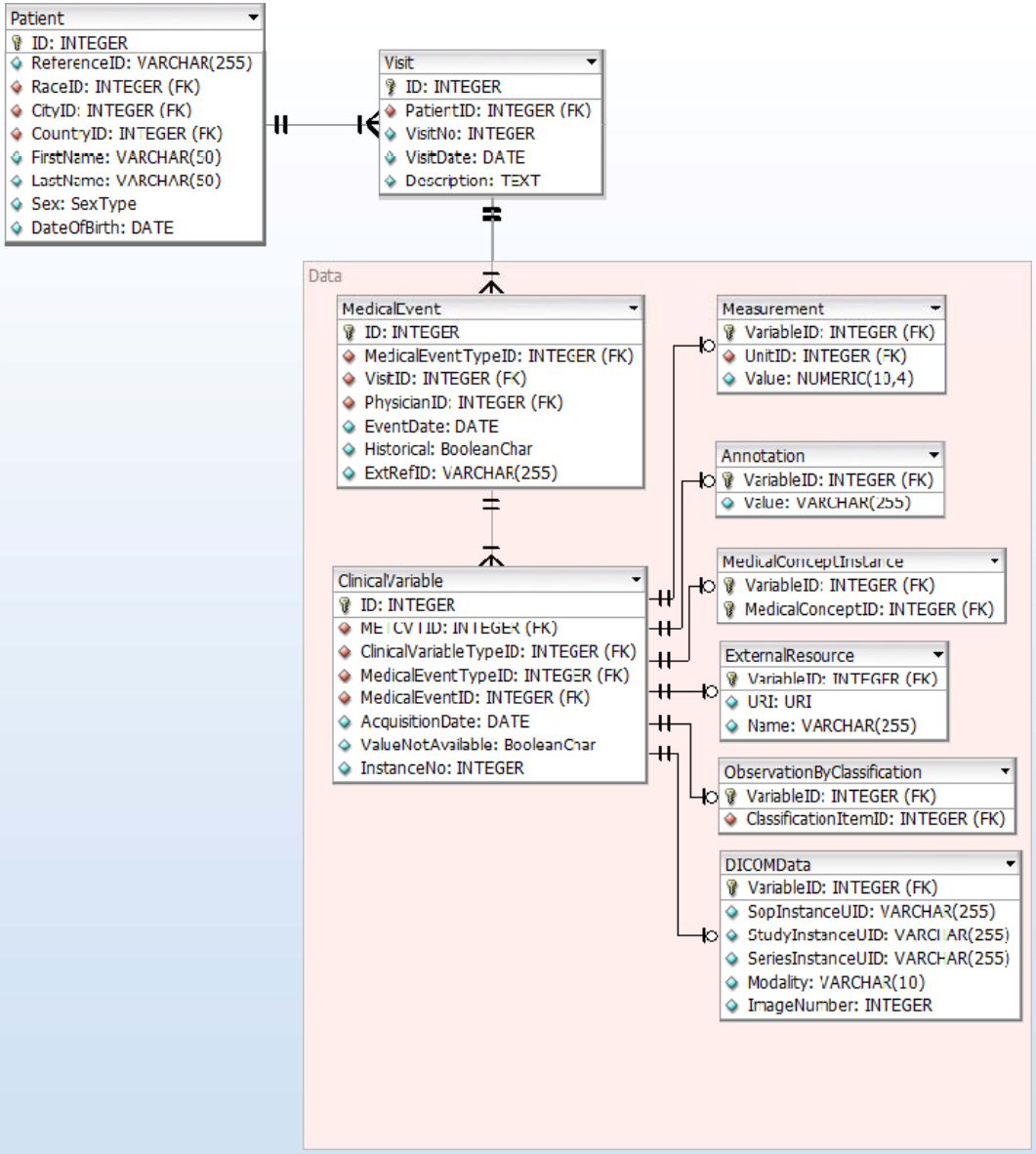
Patient Data Hierarchy

- Clinical Variable
  - Atomic piece of data
  - e.g. Joe's weight measurement - 50 kg
- Medical Event
  - Action on a patient
  - ExtRefID
    - e.g. DICOM StudyInstanceUID
  - E.g. Joe's physical examination
- Visit
  - Grouping/Context



Clinical Variables

- Actual atomic clinical data from clinical protocols – instance base
- Attached to Medical Events
- Described by Clinical Variable Types
- Can be related to each other
- Specialization/Categories of clinical variables
  - Measurement
  - Annotation
  - DICOM Data
  - Observation By Classification
  - External Resource
  - Medical Concept

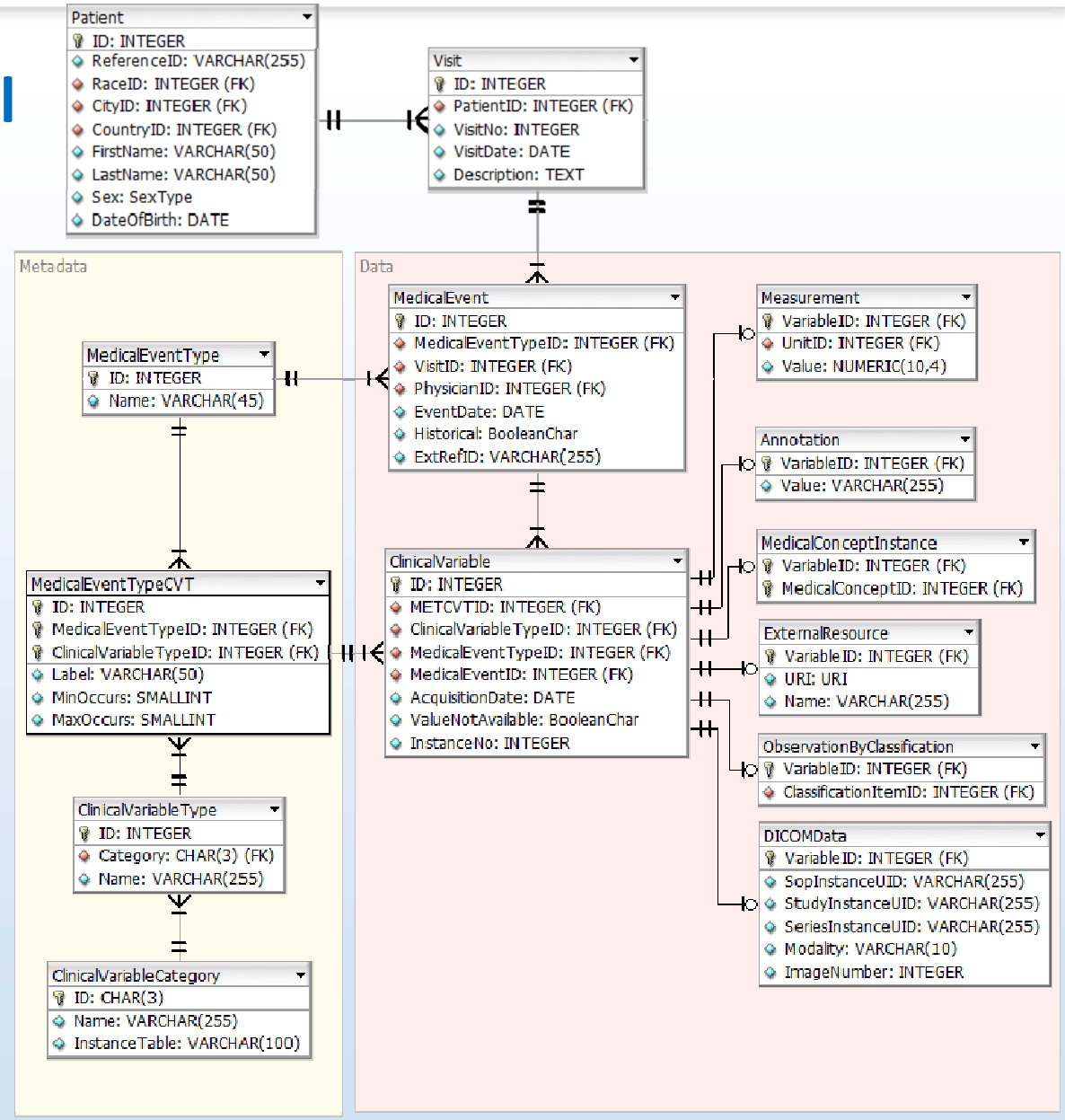


Clinical Variables Categories

- Measurement
  - any estimation of the physical quantity (e.g. height, weight, heart rate, RV volume etc.).
  - a numeric value associated with a unit of measurement (e.g. 170cm, 50kg, 72 bpm etc.)
- Annotation: any free text (e.g. comment, note, explanation etc.).
- Observation By Classification
  - classification-based assessment
  - Selection from a list of predefined values
  - Example: severity of RV dilation : ("no", "moderate", "severe")
- DICOM Data
  - Specialized container to store the relevant image associated data (image meta-data)
  - Currently - unique DICOM identifiers (e.g. SOPInstanceUID, StudyInstanceUID etc.) + a few DICOM tags (e.g. Modality)
- External Resource
  - any source of the binary data and identified by URI
  - no assumption on the structure of the data in the resource
  - Example: a file on the Grid identified by its Logical File Name (LFN)
- Medical Concept
  - “tagging” any medical event / other clinical variable with medical concept from the knowledge base
  - Example: Joe’s diagnosis “Oligoarthritis” is stored as a reference to the knowledge base (as opposed to recording as a string)

Metadata Model

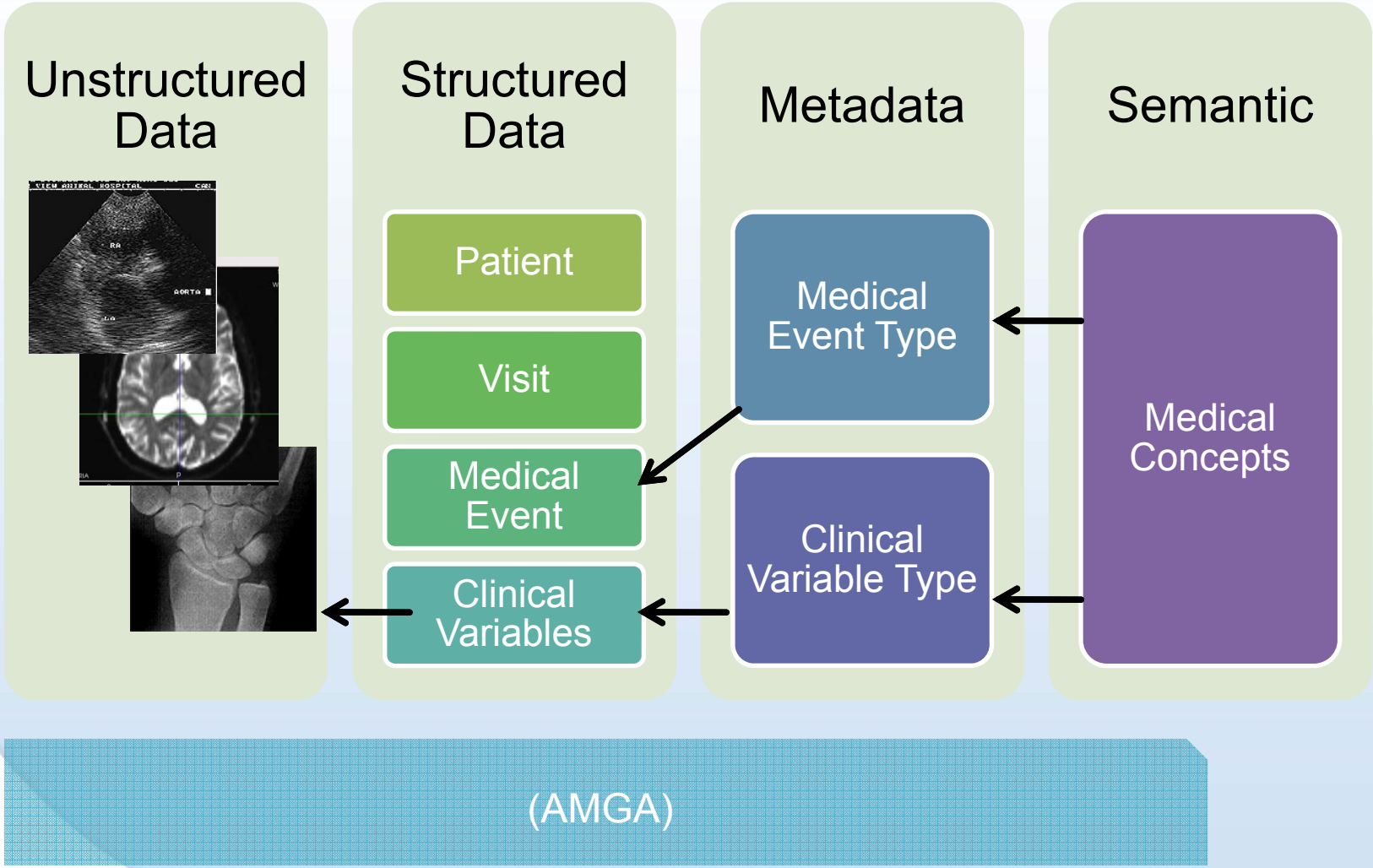
- Describes the data model
  - Kinds of data that can be stored (Clinical Variable Types)
  - How data is organized/grouped (Medical Event Types)



Example: Measurement

Layer	Statement	Meaning
MetaData	`ClinicalVariableType( ID=1, Name="Systolic LV volume", Caterogy="MSR" )`	"Systolic LV Volume" *is a ClinicalVariableType, the kind of which is "MEASUREMENT"*
MetaData	`Unit( ID=2, Name="Milliliters per square meter", Abbr="mL/m2" )`	"Milliliters per square meter", abbreviated as "mL/m2" *is a Unit*
MetaData	`MeasurementUnit( VarTypeID=1, UnitID=2, UseByDefault=Y )`	The unit "mL/m2" defined above *is used to measure the Clinical Variable Type* "Systolic LV Volume"
Data	`ClinicalVariable( ID=3, VarTypeID=1, AcquisitionDate=???, ValueNotAvailable=N )`	The Clinical Variable (identified with ID=3) *is of ClinicalVariableType "Systolic LV volume"*
Data	`Measurement( VariableID=3, UnitID=2, VarTypeID=1, Value=30.5 )`	The above clinical variable, whose type is "Systolic LV volume" *was measured in units of mL/m2 to be 30.5*

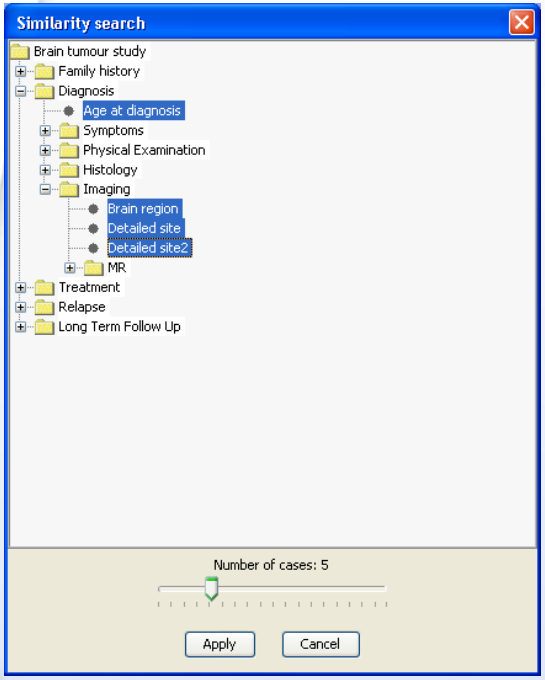
Content of data model layers

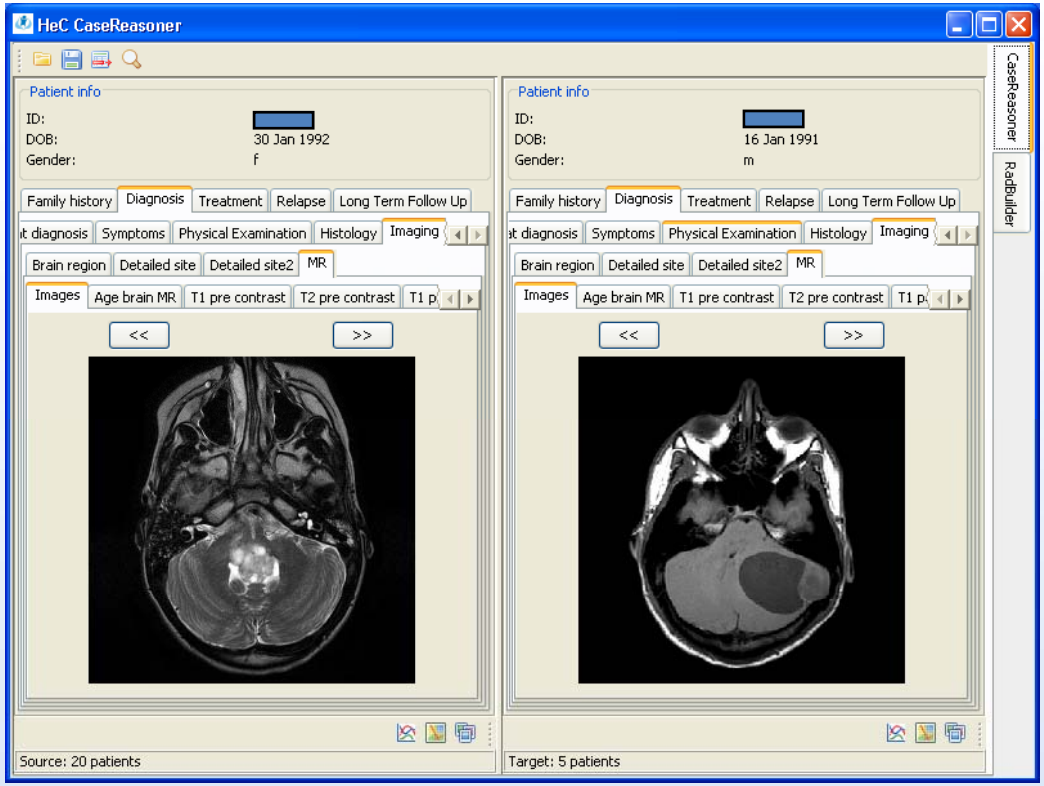


Data Access – Medical Query Processing

- Queries specified by
  - end-users
  - client applications
- Common entry point: HeC gateway
- Query workflow elements:
  - Semantic query rewriting
  - Query planning and distribution
  - Special operators -> workflow planning
  - Querying the dynamic schema
  - Database queries
  - Catalogue query, gLite SE request

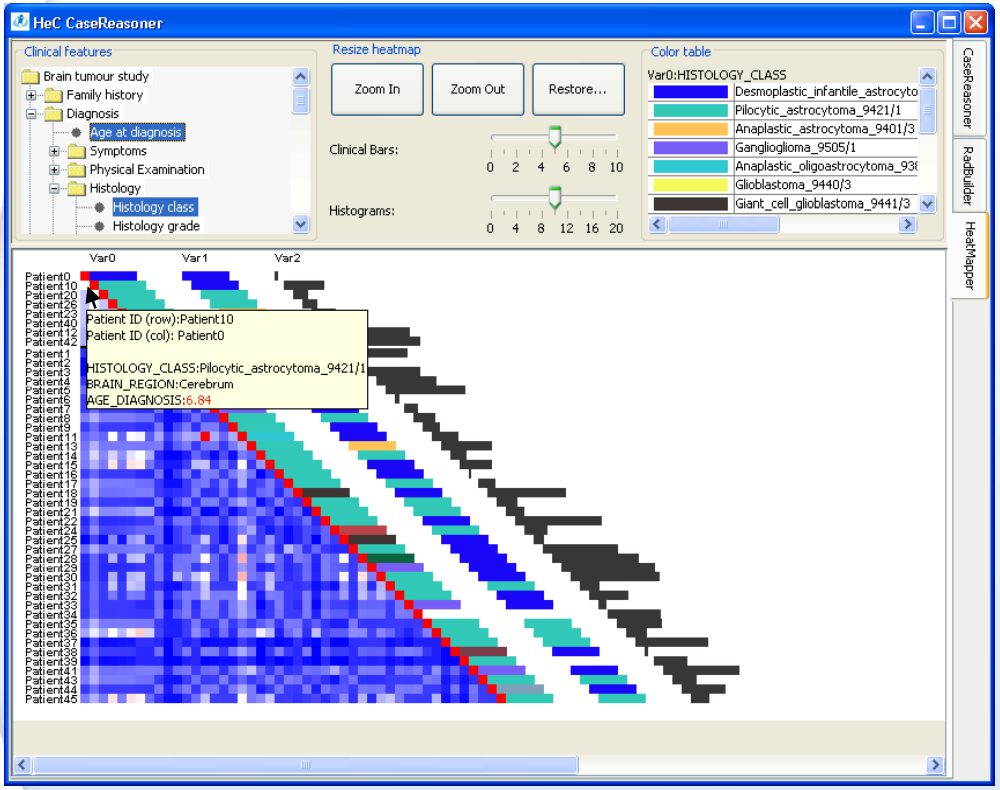
Demonstrator: Similarity Search

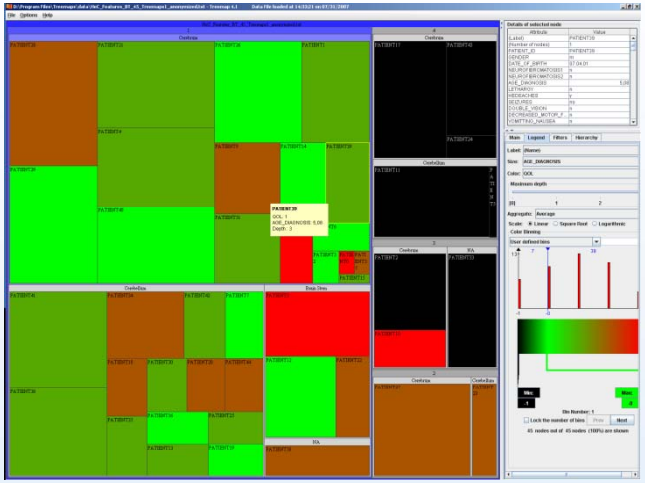




- search context is defined as a subset of (groups of) features of interest from the pre-defined feature hierarchy
- implementation in Java, Eclipse IDE, Window Builder Pro for GUI
- Weka open-source machine learning library for basic data management
- 2 initial domains: brain tumor and cardiology; extensible

Demonstrator: Visualization

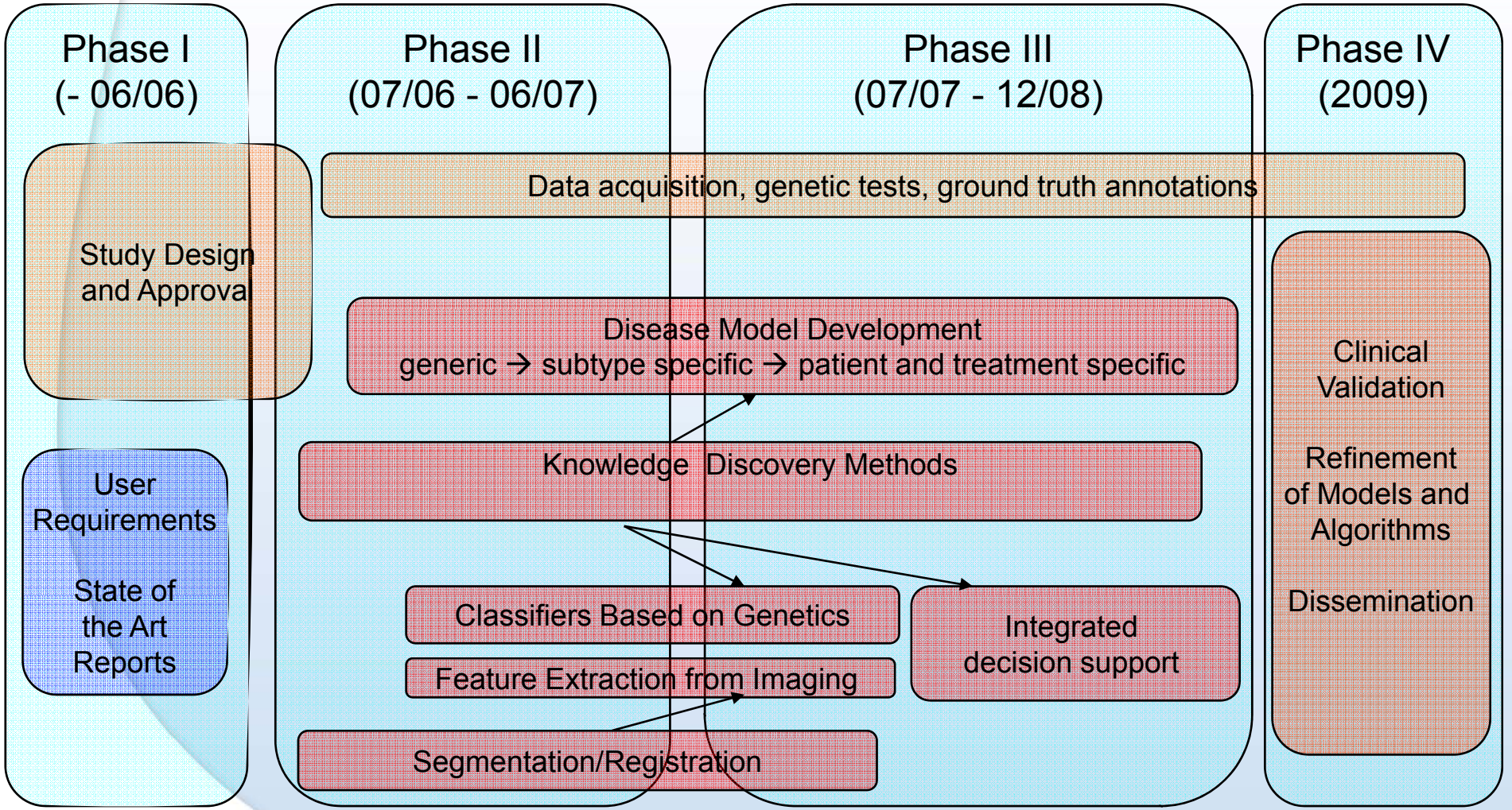




current prototype: distance maps and heatmaps are combined to visualise inter-patient distances, clinical, imaging and genetic features simultaneously

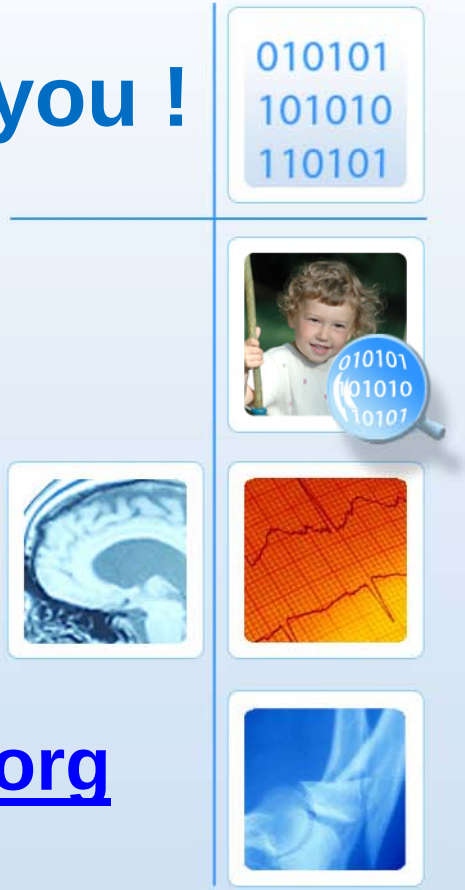
future work: treemaps and neighbour-hood graphs will be integrated for patient similarity visualization

Clinical and Application Roadmap



Health-e-Child

Thank you !



http://www.health-e-child.org