From Physics to Daily Life:
Data Management Challenges in Paediatric Information Systems

Professor Richard McClatchey
UWE, Bristol UK

Paediatric Information Characteristics I

- Huge amounts
  - Number of objects (images, datasets, test results etc.)
  - Size of objects (sets MR images can be 10→100Mb)
- Extreme heterogeneity
  - Biological granularity (genetics → epidemiological: vertical integration)
  - Distributed information sources (horizontal integration)
  - Semantic information models (data + meaning)
  - Information modalities (many forms of imaging)
  - Legacy information
- Uncertainty
  - Measurement errors
  - Text interpretation
  - Misclassification

Paediatric Information Characteristics II

- Continuous evolution
  - Child growth
  - Medical advances in disease understanding
  - Medical protocols
- Rapid generation and ingest
  - Human body and environment sensors
  - Large numbers of patient exams

Motivation for Health-e-Child

- Clinical demand for integration and exploitation of heterogeneous biomedical information
  - vertical dimension – multiple data sources
  - horizontal dimension – multiple sites
- Need for generic and scalable solutions
  - integrate traditional and emerging sources
  - offer decision support in diagnosis, therapy and follow-up
  - provide complex integrated disease models
  - ubiquitous access to knowledge repositories in clinical routine
  - connect stakeholders in clinical research
- Specific Needs in Paediatrics
  - Many medical disorders in children are little understood and some diseases are rare
  - Incentives to invest in research are low

Health-e-Child Project Structure & Partners

Framework FP6 Integrated Project (IP)

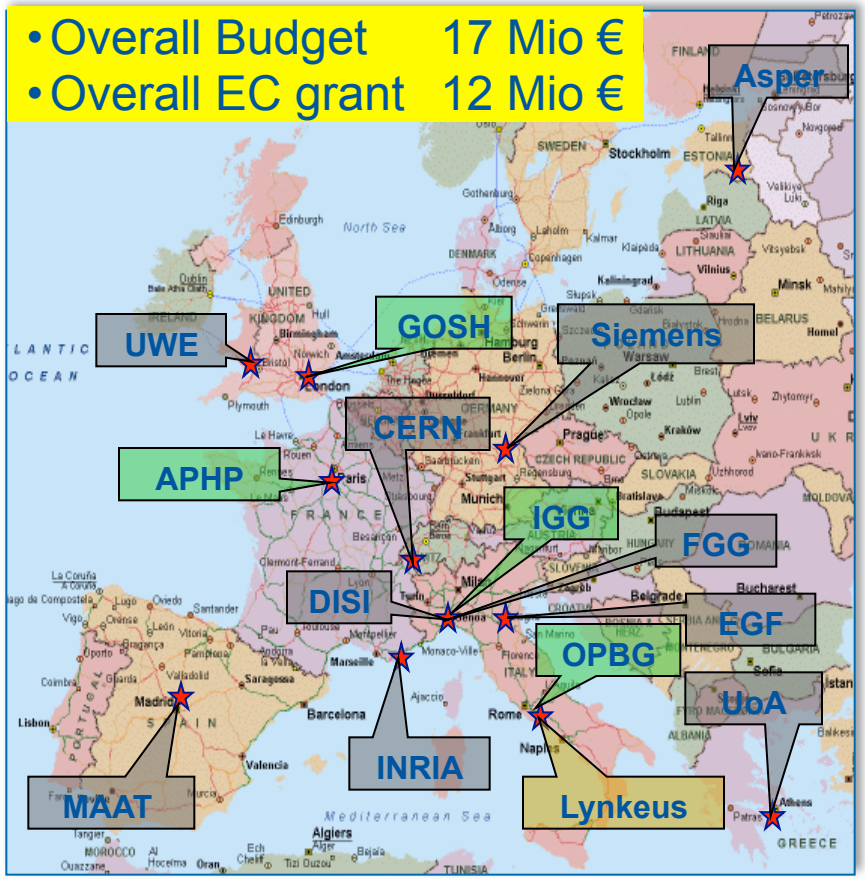
- Sponsor: European Commission
- Jan 2006 until Apr 2010

Coordinator – Siemens AG

- Siemens – H IM IKM ST
  - Project Coordinator: Jörg Freund
  - Governing Board Head: Dorin Comaniciu

Clinical Institutions

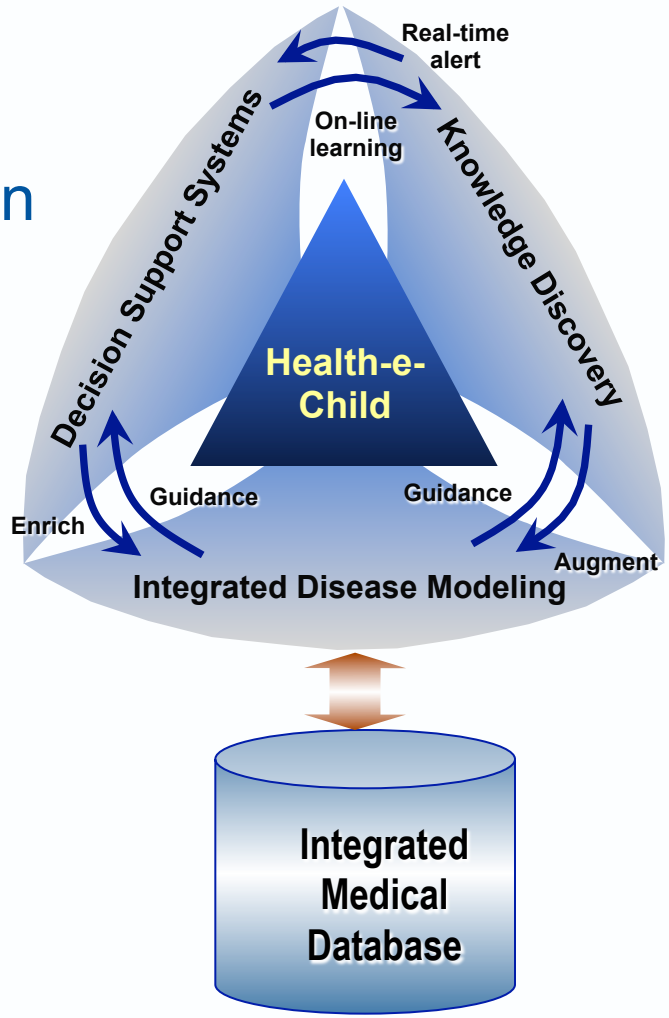
- Giannina Gaslini Hospital, Genoa, Italy
- Great Ormond Street Children's Hospital, London, UK
- Assistance Publique Hopitaux de Paris, Necker Hospital, Paris, France
- Ospedale Bambino Gesù, Rome, Italy



Other Partners: Lynkeus SRL (Rome, Italy), European Organisation for Nuclear Research **CERN** (Geneva, Switzerland), **Maat G Knowledge** (Toledo, Spain), **University of the West of England** (Bristol, UK), **University of Athens** (Athens, Greece), **Universita' degli Studi di Genova** (Genoa, Italy), **National Institute for Information and Automation Research** (Sophia Antipolis, France), **European Genetics Foundation** (Bologna, Italy), Aktsiaselts **ASPER BIOTECH** (Tartu, Estonia), **Gerolamo Gaslini Foundation** (Genoa, Italy)

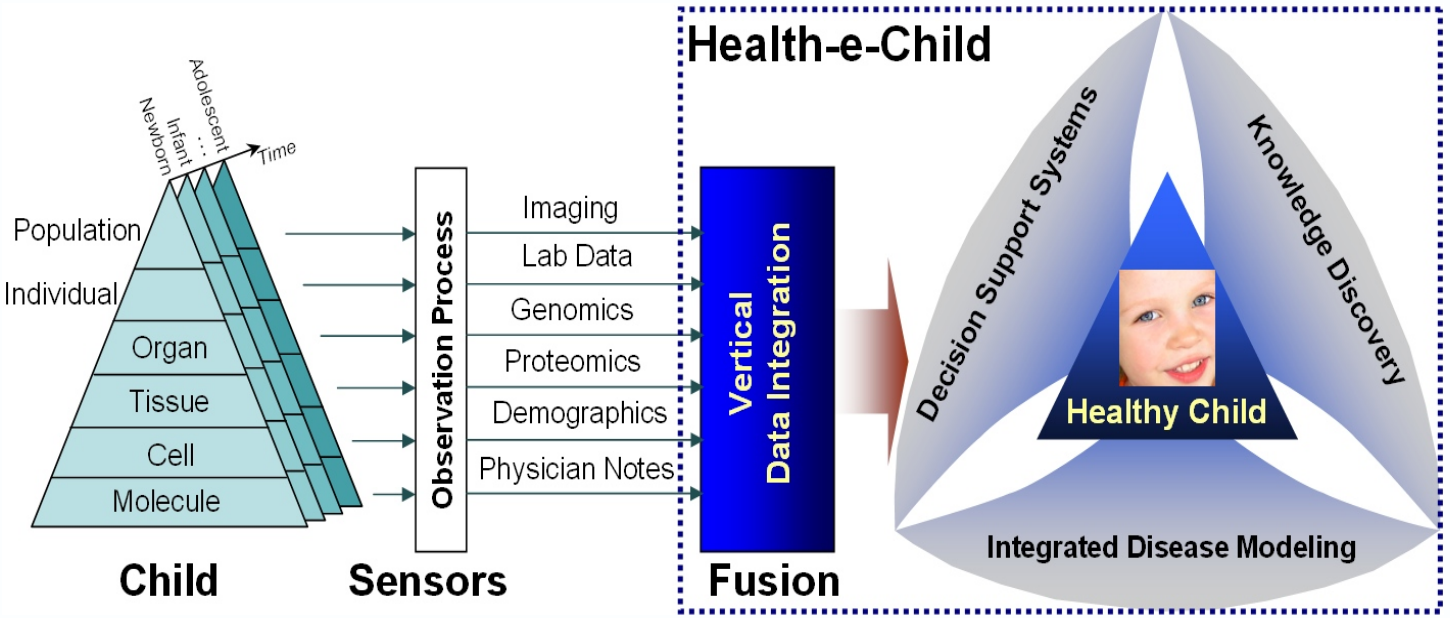
Health-e-Child at a Glance

- Establish multi-site, vertical, and longitudinal integration of data, information and knowledge
- Develop a GRID based platform, supported by robust search, optimisation and matching
- Build enabling tools and services that improve patient care
- Two main use case scenarios
  - "Aiding the Clinician in Decision Making"
  - "Clinical Studies"



Objectives of Health-e-Child

- Build enabling tools & services that improve the quality of care and reduce cost via:
  - Integrated disease models
  - Database-guided decision support systems
  - Cross modality information fusion and data mining for knowledge discovery

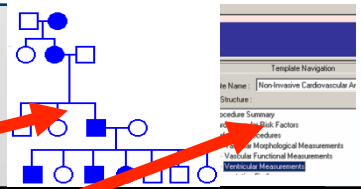


Focus on Paediatric Diseases

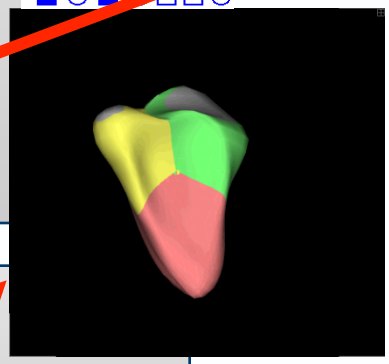
- Three Paediatric Diseases with at least partly unknown cause, classification and/or treatment outcomes
  - Heart diseases (*Right Ventricular Overload, Cardiomyopathy*)
  - Inflammatory diseases (*Juvenile Idiopathic Arthritis*)
  - Brain tumours (*Gliomas*)
- Many Clinical Departments
  - Cardiology
  - 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

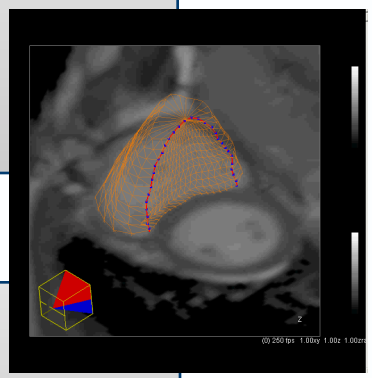
Example Disease: Right Ventricular Overload

Clinical Data
- Demographic, history & familial
- Lifestyle
- Clinical notes
- ECG

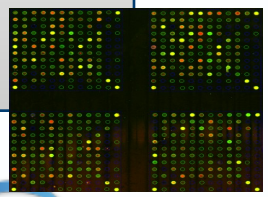


Imaging Data
- 2D/3D Echo
- Tissue Doppler
- MRI





Genetic Data
- Karyotyping
- Array-CGH



Clinical Features
- prolonged PR interval in electrocardiogram
- systolic ejection murmur on auscultation

Anatomical Features
- Hyperkinetic RV muscle
- Increased RV-LV ratio
- Ventricular septum defect
- Thickening (hypertrophy) of the RV muscle

Genetic Features
- candidates for gene mutations are e.g. 4p13-q12, 6p21.3, 1p31-p21, 3p25, 6q21-q23.2, 5q34

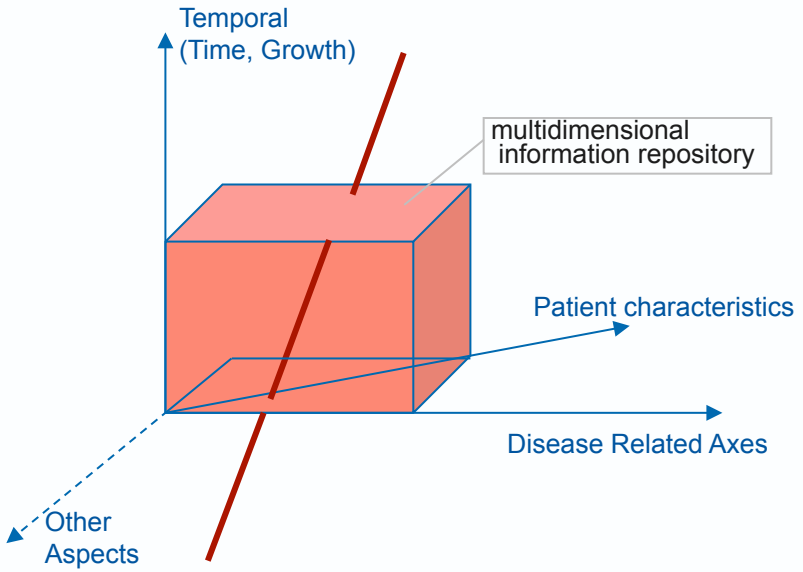
training and specific patient data

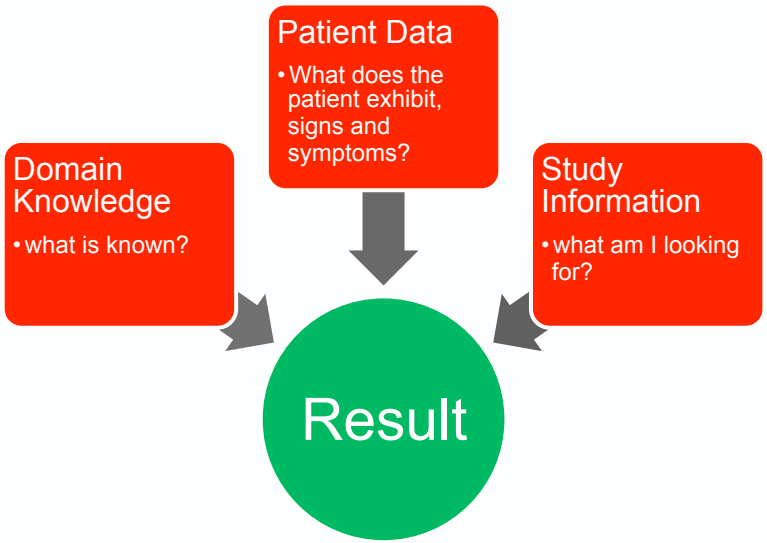
Decision Support
- prediction of type and timing of treatment
- classification of RV overload
- retrieval of similar cases

Knowledge Discovery
- classification of subtypes
- genotype/phenotype correlation

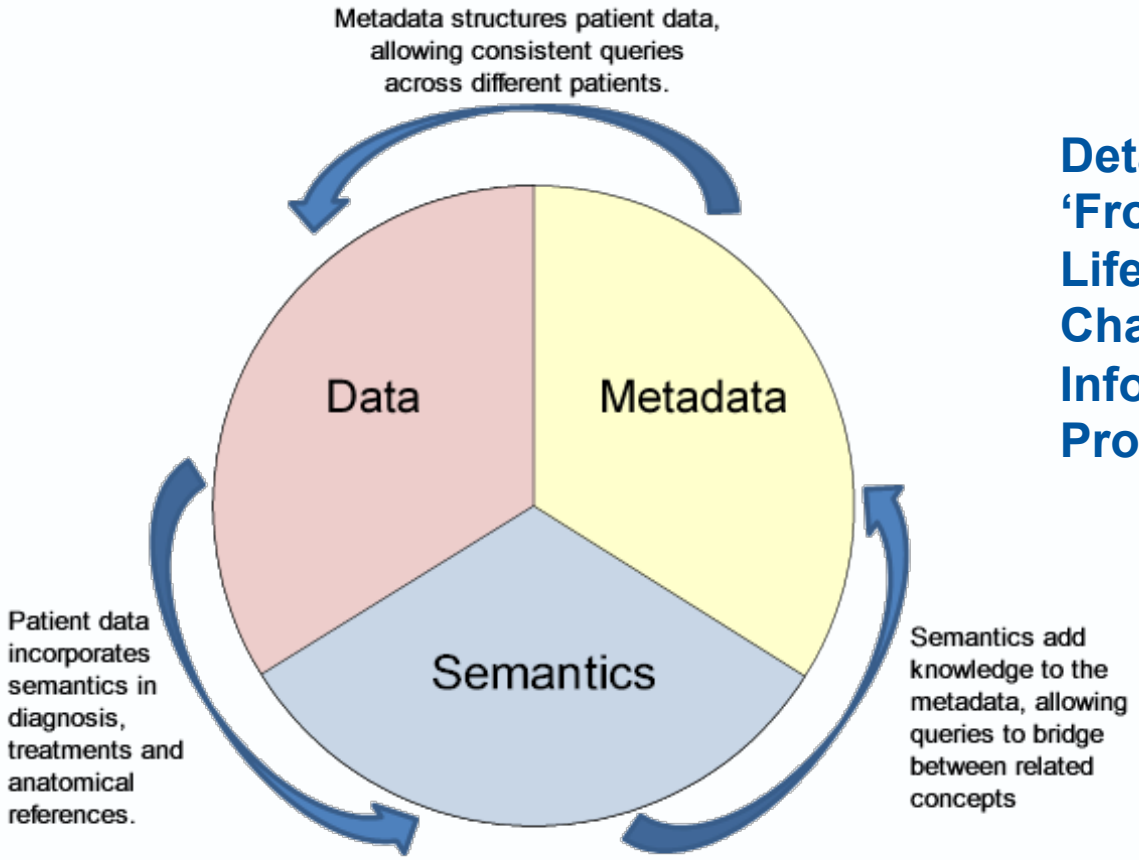
Integrated Data Modelling

- Modelling axes
  - disease, vertical levels, medical process, source/modality, temporal
- Requirements-driven
  - clinical protocols, user requirements
  - basis for data management in the platform
  - data access for applications
- Integration
  - Views / queries along multiple axes



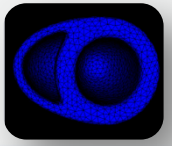


Health-e-Child : Integrated Data Model

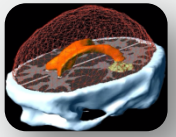


Details are outlined in : 'From Physics to Daily Life: Data Management Challenges in Paediatric Information Systems' by Prof Richard McClatchey

Health-e-Child Platform Overview



Cardiology
Right Ventricular Overload Cardiomyopathy



NeuroOncology
Brain Tumors / Gliomas

Rheumatology
Juvenile Idiopathic Arthritis

Knowledge Discovery

Decision Support

Disease Modelling

Gateway

Ontological Layer

Query Processing

Data Management

Grid – EGEE gLite

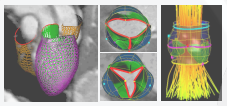
IGG GOSH NECKER OPBG JHU

Use of Grids for Biomedical Sciences

- Life Sciences
  - To address complexity of **databases interoperability** (e.g. Embrace)
  - To ease the design of **data analysis workflow** (e.g. MyGrid)
- Medical Research
  - To store and manipulate **large cohorts of medical images** (e.g. Mammogrid, neuGRID/N4U)
  - To bring together and to **correlate patient medical and biological** data (e.g. ACGT, Health-e-Child, @Neurist)
- Drug Discovery
  - First step of a full ***in silico* drug discovery process** successfully proven (e.g. Wisdom)
  - To **reduce time and save money** in the drug discovery process

Innovation Exchange

Grid-Enabled Platform for Simulations in Paediatric Cardiology
Toward the Personalized Virtual Child Heart



Sim-e-Child

Worldwide e-Infrastructure for Computational Neurosciences

neuGRID

Worldwide e-Infrastructure for Computational Neurosciences

outGRID

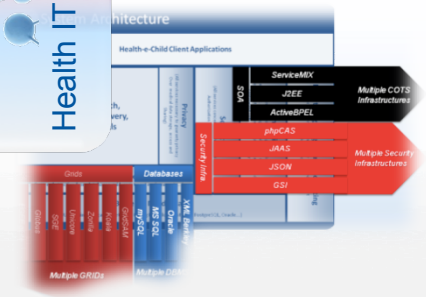
MammoGrid+

Digital Repository Infrastructure for Breast Cancer Research

Grid-enabled Sentinel Network for Cancer Surveillance

Sentinel

Health IT



SALTY

Self-Adaptive very Large Distributed Systems

Summary

- The importance of IT will increase the more personalized medicine becomes reality
  - to automatically process and analyze the data (genetics, proteomics but also imaging)
  - to provide access to large annotated patient data bases
- Health-e-Child has developed a multi-site system infrastructure supporting vertical data integration and offering both generic and specific tools
  - to discover new knowledge
  - to aid in decision making
- These are the first steps in a long journey towards support for effective, personalized healthcare in the 21st century.