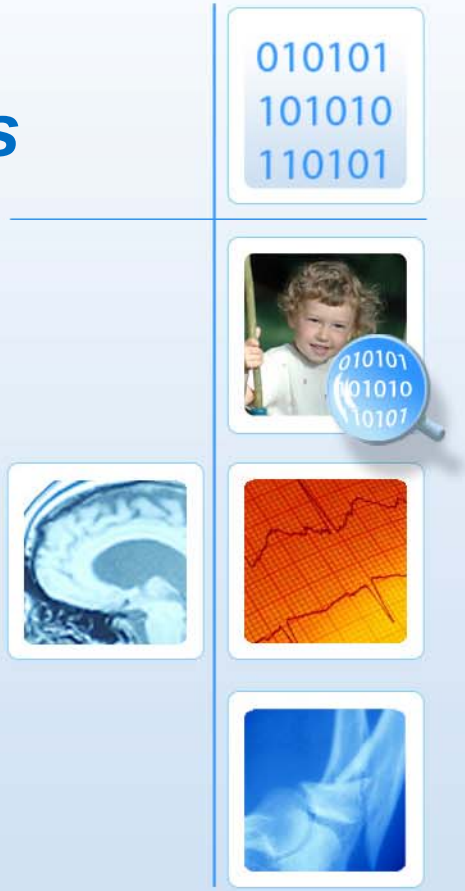
Health-e-Child

Health-e-Child

Project Requirements



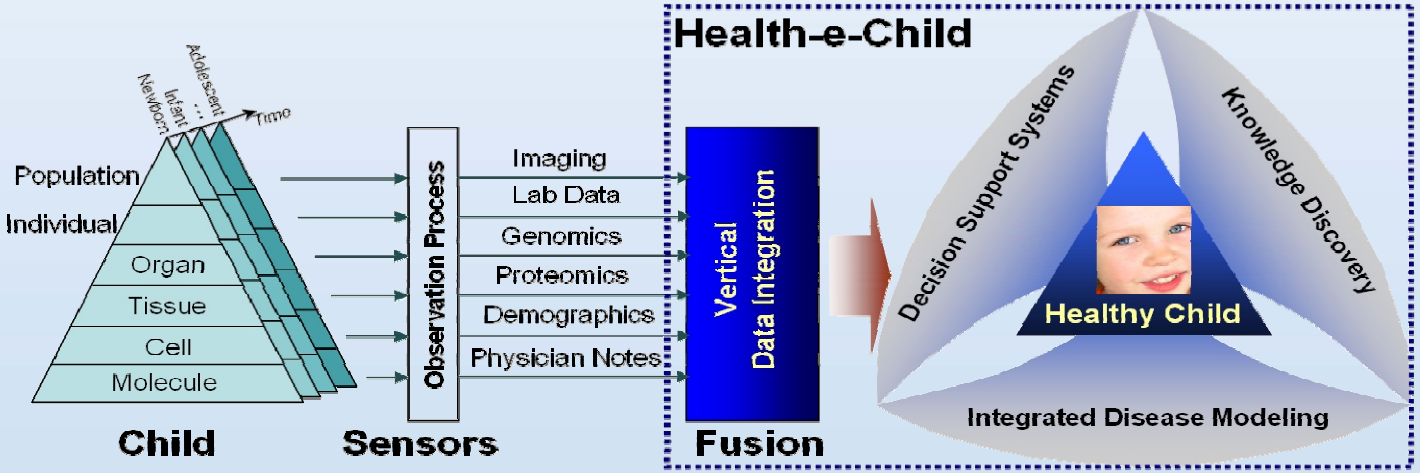
EGEE 2006 – HealthGrid

David Manset

MAAT GKnowledge

Project Objectives

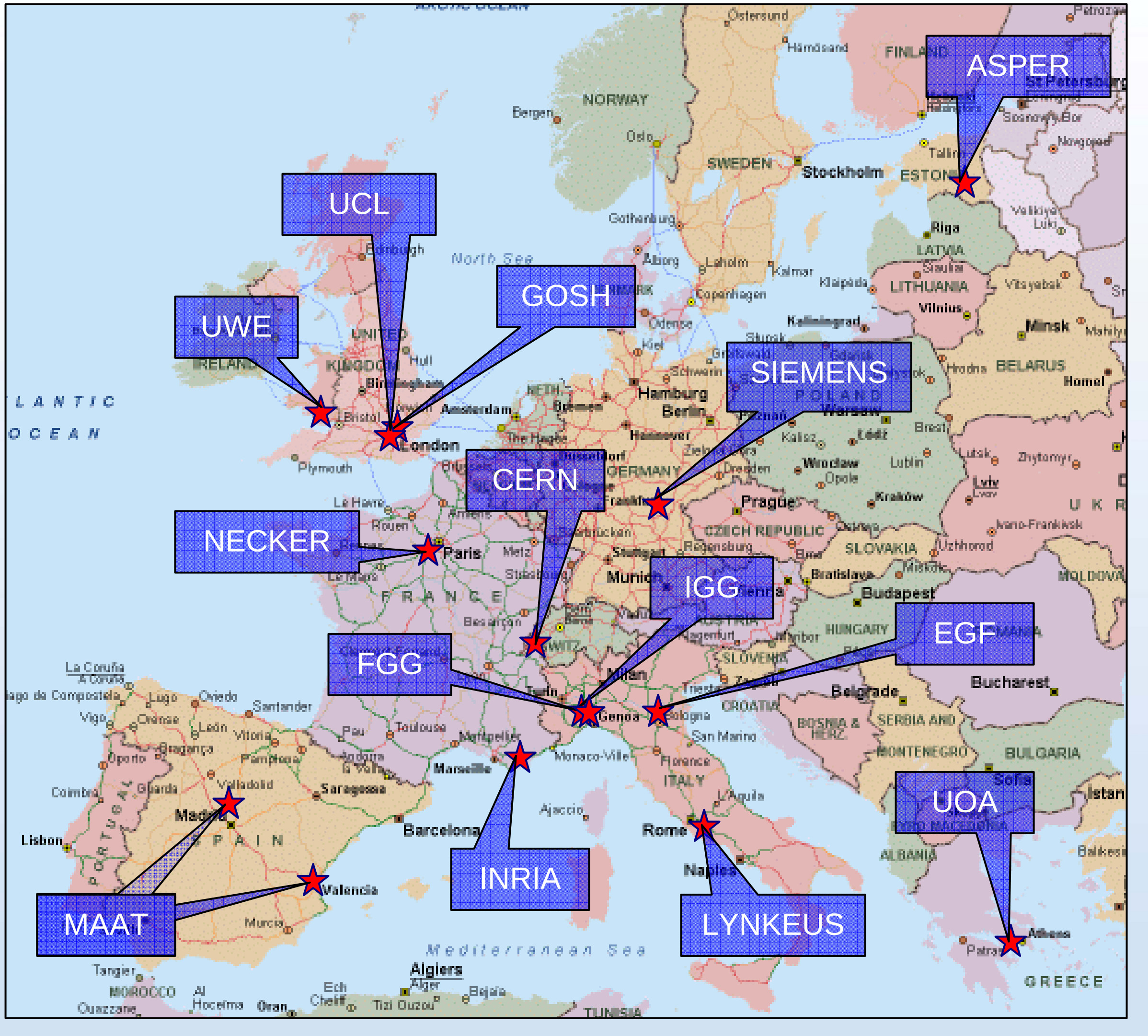
- Establish Horizontal and Vertical integration of data, information and knowledge
- Develop a grid-based biomedical information platform, supported by sophisticated and robust search, optimisation, and matching techniques for heterogeneous information,
- Build enabling tools and services that improve the quality of care and reduce its cost by increasing efficiency
  - Integrated disease models exploiting all available information levels
  - Database-guided decision support systems
  - Large-scale, cross-modality information fusion and data mining for knowledge discovery
- A Knowledge Repository?



Project General Info

- Instrument: Integrated Project (IP) of the Framework Program FP6
- Project Identifier: IST-2004-027749
- Coordinator: Siemens AG, Dr. Jörg Freund
- Partner: 14 European (companies, hospitals, institutions)
- Timetable: 01-Jan-06 to 31-Dec-09 (4 years)
- Total cost: 16.7 Mio. €
- EC funding: 12.2 Mio. €
- Web page: http://www.Health-e-Child.org

Project Map


ASPER
UCL
GOSH
UWE
SIEMENS
CERN
NECKER
IGG
EGF
FGG
UOA
INRIA
LYNKEUS
MAAT

Clinical Context

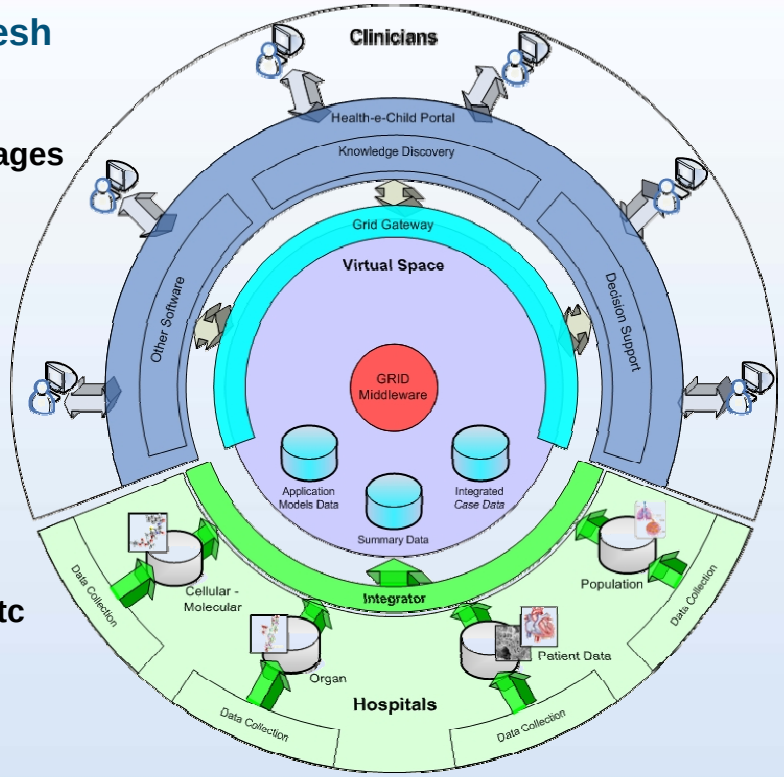
- **Diseases**
  - Heart diseases (*Right Ventricle Overload, Cardiomyopathy*),
  - Inflammatory diseases (*Juvenile Idiopathic Arthritis*), and
  - Brain tumours (*Gliomas*)

- **Clinical Institutions**
  - I.R.C.C.S. Giannina Gaslini (IGG), Genoa, Italy
  - University College London, Great Ormond Street Children's Hospital (GOSH), London, UK
  - Assistance Publique Hopitaux de Paris – NECKER, Paris, France

- **Clinical Departments**
  - Cardiology
  - Rheumatology
  - (Neuro-)Oncology
  - Radiology
  - Lab (Genetics, Proteomics, Lab)
  - Administration

Data Integration Challenge (1)

- **3 Hospital Nodes**
  - Integration of data stored in Hospital's IS + fresh new data to be acquired
  - Acquisition of large samples of Imaging data
    - 3 diseases X 300 cases X 2 modalities X 300 images
      - i.e. at most 540000 images ~ 270 GB
- **A Distributed Platform for sharing, manipulating and inferring data**
  - Decision Support System
  - Disease Modelling
  - Knowledge Discovery / Data Mining
  - Image Processing
    - Automatic segmentation of right ventricle
      - to determine volume, ejection fractions etc for cardiac MR and ultrasound images
    - Brain tumour segmentation/registration to determine volume, location etc
    - Volume of synovial fluid in wrist MR scans
- **Grid technology as the enabling infrastructure**



Data Integration Challenge (2)

		IGG	GOSH	NECKER
Cardiology	*DB*	MS ACCESS + Excel	TOMCAT	**NO** - Paper-based
	PACS	**YES** - But not operational	**YES**	**NO**
Rheumatology	*DB*	MS ACCESS + Excel	**NO** - Paper-based	**NO** - Paper-based
	PACS	**NO** - PACS in 2007	**YES**	**NO**
Radiology	*DB*	Not Available	RIS	RADOS
	PACS		**YES** - But not operational	**YES** - But being tested
Molecular Genetics	*DB*	MS ACCESS + Excel		
	PACS	**NO**		
NeuroOncology	*DB*	MS ACCESS + Excel		
	PACS	**NO**		
Proteomics	*DB*	MS ACCESS + Excel		
	PACS	**NO**		

Data Integration Challenge (3)

- Heterogeneous Data/Imaging Sources
  - DB Backends: from simple MS ACCESS to complex Patient Information Systems like TOMCAT, RIS …
  - No or few linkage bw department's IS
  - Various imaging modalities: MRI, CT, US, X-Ray…
  - Various imaging devices: Siemens Bi-Plan, GE Vivid7, Sequoia, HP128…
- Heterogeneous Connectivity
  - PACS not yet present in all Hospitals/Departments
  - Hospitals have different Hardware/Network/Security constraints
- A 3-Phase Data Integration Scheme
  - 1st: A temporary offline data acquisition application
  - 2nd: An online data acquisition application (interacting with the platform)
  - 3rd: A background data integration service (in the platform)

Early Faced Issues

Mainly Non-Functional since project has just started

- Selecting grid m/w services wrt project requirements
  - Lots of services/functionalities available
  - Different implementations with different levels of maturity
- Clustering grid m/w services
  - To reduce the h/w requirements & maintenance (1 server / Hospital)
  - To facilitate deployment (3 clinical sites + at least 5 institutional sites)
- Decentralisation of grid m/w services
  - Sites need to be as much as possible autonomous
- “Griddification” of Applications
  - Some of the HeC applications might be “griddified”
    - Griddification has to be balanced against runtime and development complexity criteria

Current Investigations

- Selecting grid m/w services wrt project requirements
  => Services selection based on URS + Grid Questionnaire
- Clustering grid m/w services
  => “Xenification” of OSs + clustering services wrt functionality
- Decentralisation of Grid Services
  => Dependent on gLite developments, but already some possibilities with Master/Slave configurations
- “Griddification” of Applications
  => Introduced a classification of applications. Grid Questionnaire will certainly help in making decisions
- Grid Access
  => Abstracting grid access through dedicated service

Remaining Challenges

- Data Integration in Hospitals (post phase 2)
  - What mechanisms to use? What will be the limitations (in particular with proprietary systems?
- Patient Data Distribution & Sharing
  - What technology/implementation?
- Patient Image Files Sharing
  - Enabling the sharing of large files over the internet
    - MRI @ GOSH = 500MB/patient
    - CT @ NECKER = 3.5GB/patient …*raises bandwidth problems*
- Griddification of Applications
  - Appears relevant for computation heavy algorithms or batch processing
    - However many clinical algorithms have short runtime (e.g. image processing, since clinicians need almost instantaneous results)

Conclusion - Middleware Requirements

Non-functional Requirements

- Hospital Sites should be autonomous
  - Sites should not depend on any central services
- Hardware requirements remain too high for Hospitals
  - Getting access to the grid through one box would be ideal
  - e.g. 1 Server per Hospital
- Fine-grained security mechanism for accessing data (at the record level?)

Functional Requirements

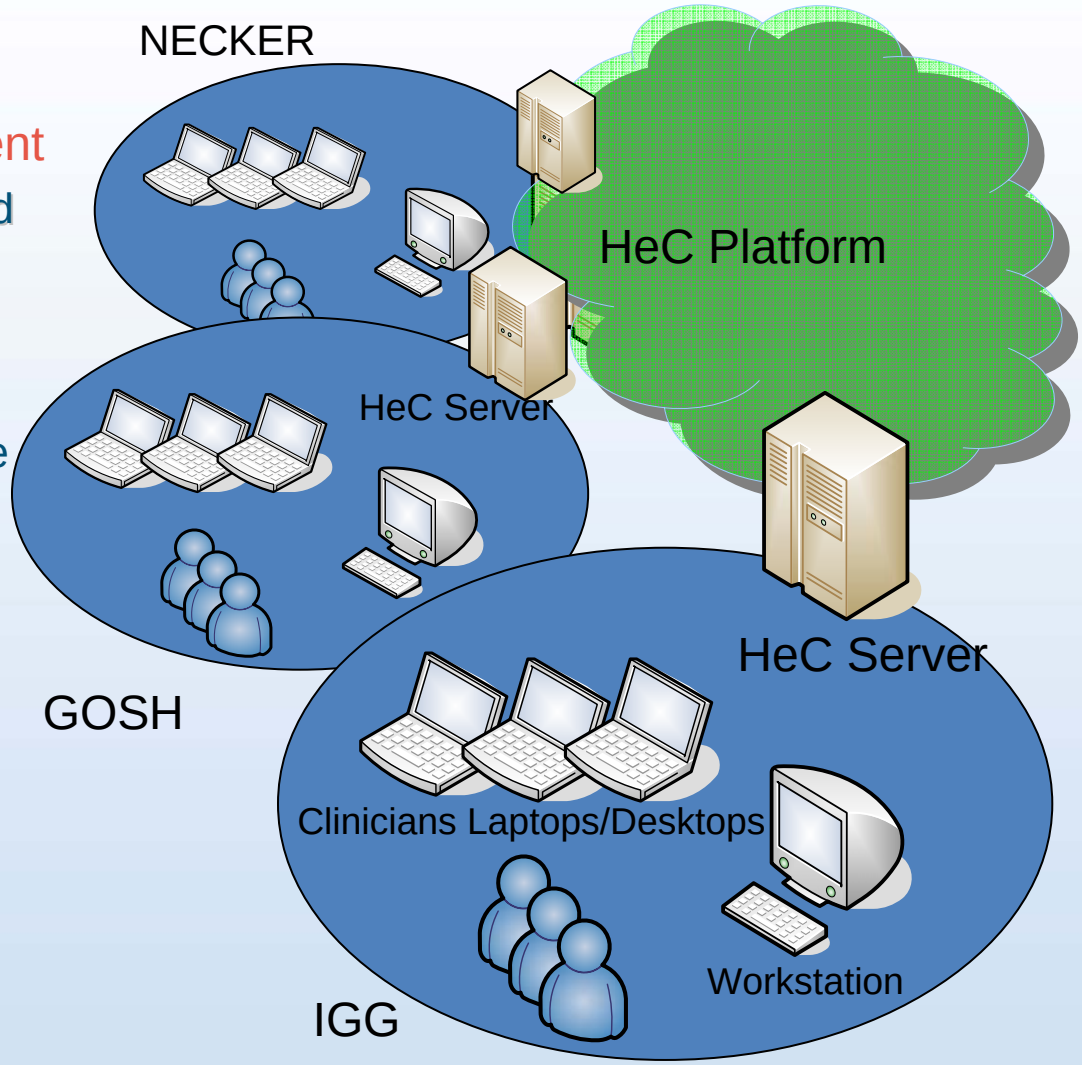
- Pseudonymisation as a native middleware service?
- Native Streaming facilities for sharing large DICOM files
- [Native patient-centric data model(s)
  - (flexibility) Optionally data model could be selected from existing standards (e.g. HL7…) or even created from scratch
  - (interoperability) Optionally a native commodity for exporting/exposing data through different data models would be nice (model-driven)
  - (interoperability) Optionally a data model (schema) discovery mechanism could help
- Native connectors to external backends for batch data integration]

1. Are HealthGrids likely to become the enabling infrastructure for Distributed PACS?

2. Is the Grid likely to become the enabling infrastructure for Knowledge Repositories?

Approach (1)

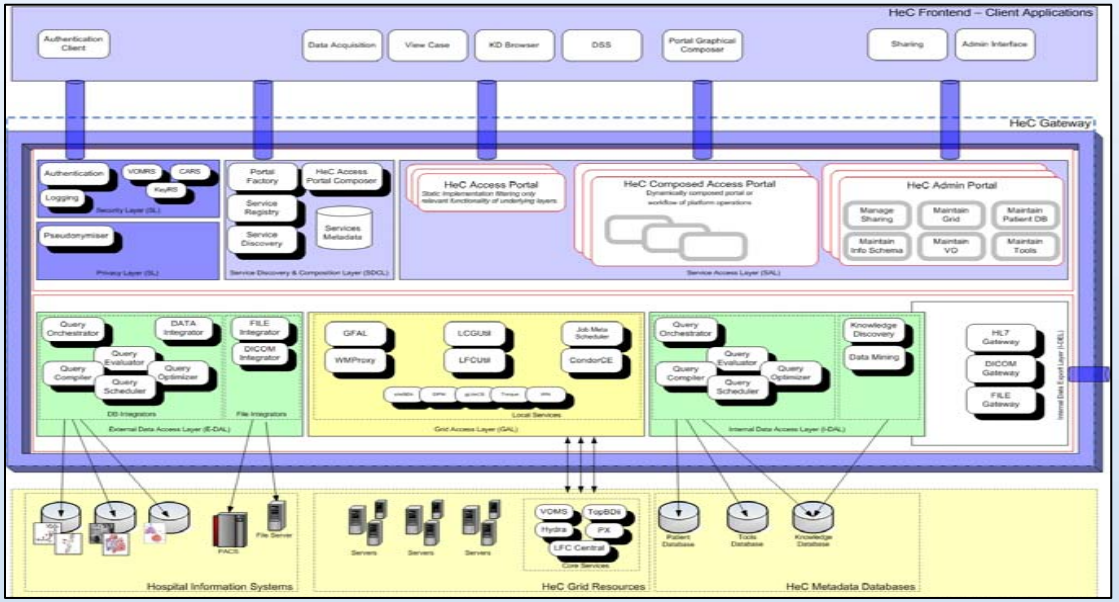
- One server per Hospital
  - Single entry point to HeC Platform
- One workstation per Department
  - For complex tasks a dedicated user interface is used
- Generic computers on Intranet
  - Most functionalities accessible from generic web browsers



Clinician's HeC Identity

Approach (2)

- An intermediary access layer: the **HeC Gateway**
  - To decouple client applications from the complexity of the grid and other computing resources
  - Towards a platform independent implementation
- Domain Specific Functionality exposed in the HeC Gateway
- Grid mainly used as a Distributed & Federated PACS
  - Different modalities of images to be anonymised and shared
  - Clinical Reports
  - Misc. Files



Platform Use Cases (1)

(high-level) Use Case	Comment	Scope
1. Collect Information		
Data Acquisition	--	Local
Data Annotation	--	Local & Global
2. Retrieve & Exploit Information		
View Case	Requires high responsiveness	Local & Global
Find Similarity	Requires high responsiveness	Local & Global
Query	Requires high responsiveness	Local & Global
Knowledge Mining	--	Global
Use Decision Support System	Requires high responsiveness	Local & Global
Use Disease Models	Requires high responsiveness	Local & Global
3. Maintain Platform		
Maintain Patient Database	--	Local
Maintain Information Schema	--	Local & Global
Maintain Tools	--	Global
Maintain VO	--	Global
Maintain Grid	--	Global
Manage Sharing	--	Global

1st Technical Accomplishments

- Establishment of a Common Development Environment
  - Indispensible to synchronise partners and leverage synergy
- Creation of the Health-e-Child Virtual Organisation (VO)
  - Establishment of a Certificate Authority (36 certs delivered so far)
  - HeC VO Structure in place, being tested
- 1st gLite Test-bed deployed in May 2006 on HeC dedicated servers
  - ~20 computers involved
  - Being refined according to project requirements
- 1st embryo HeC gateway
  - Authentication Client Application & Grid Service (VOMS enabled)
  - HeC Portal & Factory (exposing domain specific functionality)