

Fighting Cancer with the (help of) Grid

Jarek Nabrzyski, PSNC
naber@man.poznan.pl

Outline

- Cancer informatics
- Cancer Grid projects
- European ACGT Project (eHealth)
- Summary

Cancer basics are not covered in this talk

The problem

- Individual cancer research groups and hospitals generate large amounts of data of diverse types
- Sharing of this data is not common today
- Value can be added by collecting data together and analysing it further
 - sharing data of one type
 - integrating data of different types
- **But** in most areas of research sharing and integrating data is invalid or error prone because of lack of standards and infrastructure

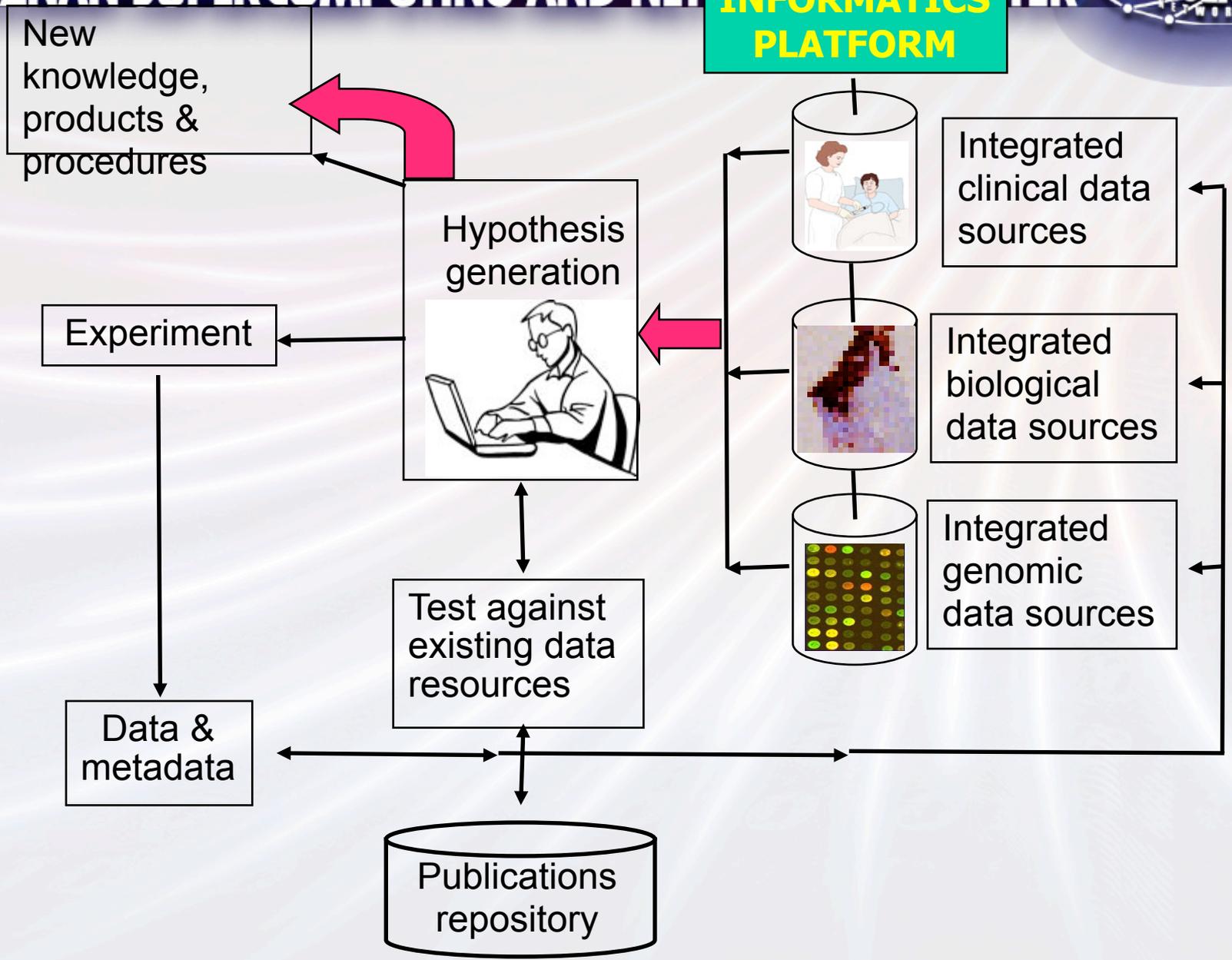
Does data sharing add value?

- Genbank grew from 606 sequences in 1982 to >30 million in 2003
- Data retrieved in first 6 months of 2004: equivalent to 5×10^{13} base pairs
- Cost of regenerating 0.01% of this: \$500 million

Essential standards & infrastructure

- Data elements
- Controlled vocabularies and ontologies
- Data exchange formats
- Protocol standardisation
- Implementation – architecture & databases
- Data mining tools
- Confidentiality & privacy enhancing technologies
- Knowledge management

INFORMATICS PLATFORM



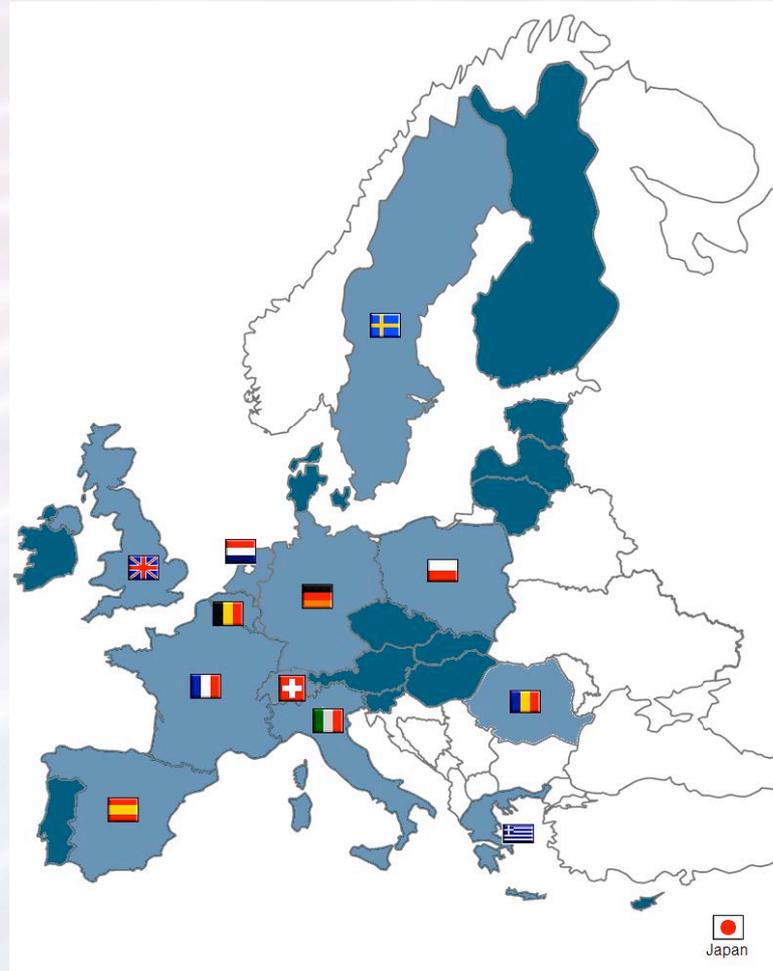
Cancer Grid Projects

- www.worldcommunitygrid.org
- CancerGrid (UK)
- Breast Cancer Grid in US (Pennsylvania)
- *Cancer Biomedical Informatics Grid (caBIG)*
- *Biomedical Informatics Research Network (BIRN)*

Advancing Clinico-Genomic Clinical Trials on Cancer (ACGT)

- New EU-funded IP project under eHealth unit
- Starting: February 2006, 4-years, 25 partners
- Budget: 16.7ME (11.8 requested)
- To manage and exploit the vast amounts of diverse information currently generated
- Improve prevention and treatment of cancer by effective use of informatics
- Increase the impact of European cancer research

ACGT Partner Countries



ACGT Goals

- To deliver to the cancer research community an integrated Clinico-Genomic ICT environment enabled by Grid infrastructure.
 - Grid: seamless mediation services for sharing data and data-processing methods and tools and advanced security
 - Integration: semantic, ontology based integration of clinical and genomic/proteomic data (standard clinical and genomic ontologies and metadata).
 - Knowledge discovery: Delivery of data-mining Grid services in order to support and improve complex knowledge discovery process.
- ACGT is trying to put in place, metaphorically, a World Wide Web of cancer research, a semantically interoperable World Wide Web of cancer research.

Clinical trials

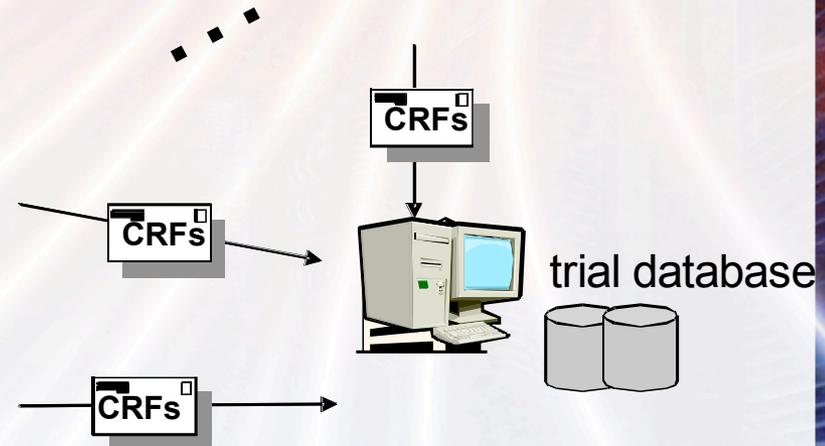
- Presence of clear research objectives, to validate the technological platform,
- Development of mechanisms for the smooth incorporation of the clinical-trials in an integrated, GRID-enabled and enriched with knowledge-discovery capabilities environment,
- Interpretation of results, as presented by the extracted knowledge, into standardized clinical guidelines and potential protocols

ACGT Pilot Trials

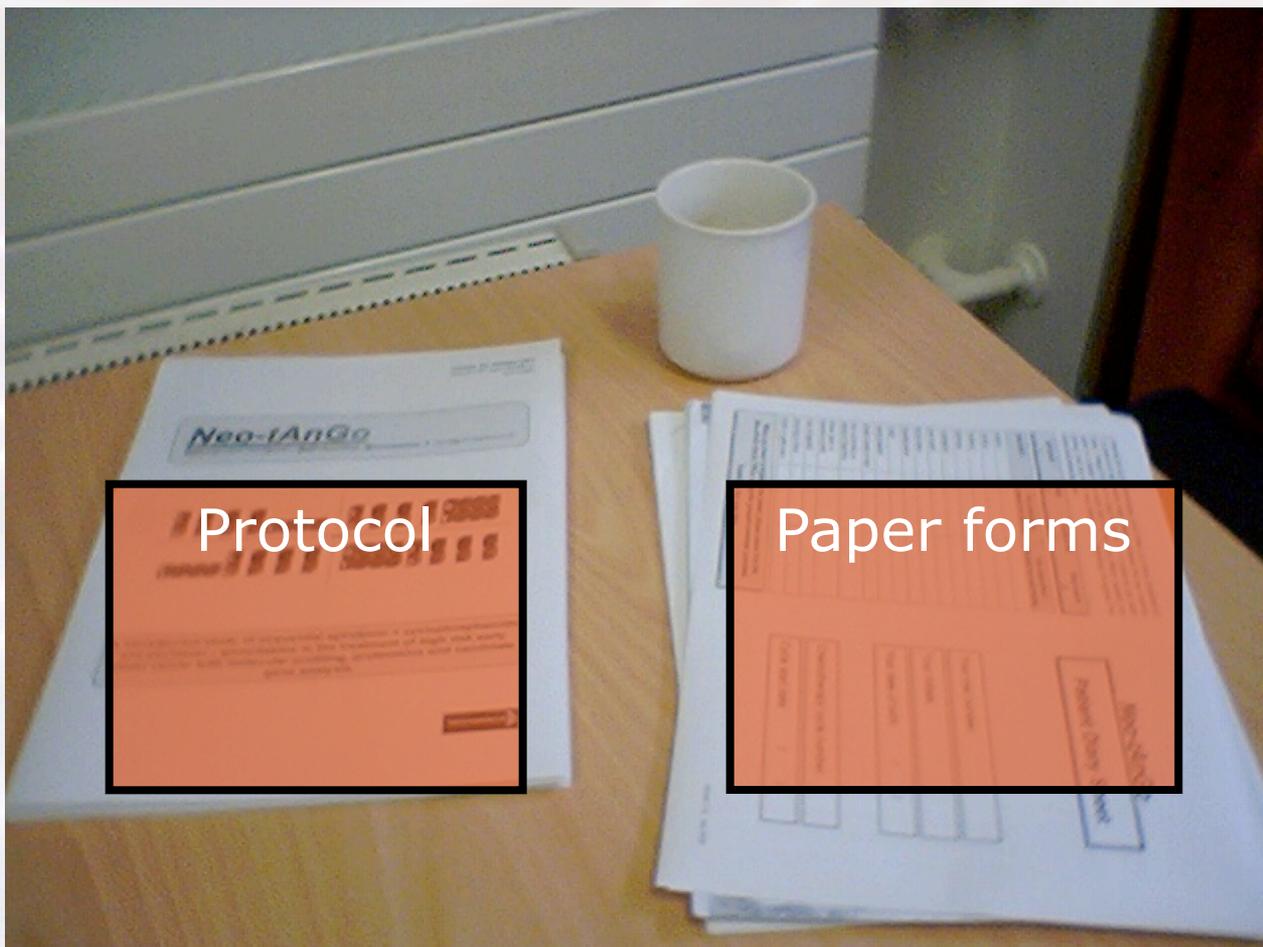
- Breast cancer
- Paediatric nephroblastoma or, Wilms tumour (PN)
- Development and evaluation of in silico tumour growth and tumour/normal tissue response simulation models

Current Approach

- Thousands case report forms (CRF) are collected to the central repository
- Data is only partially standardized



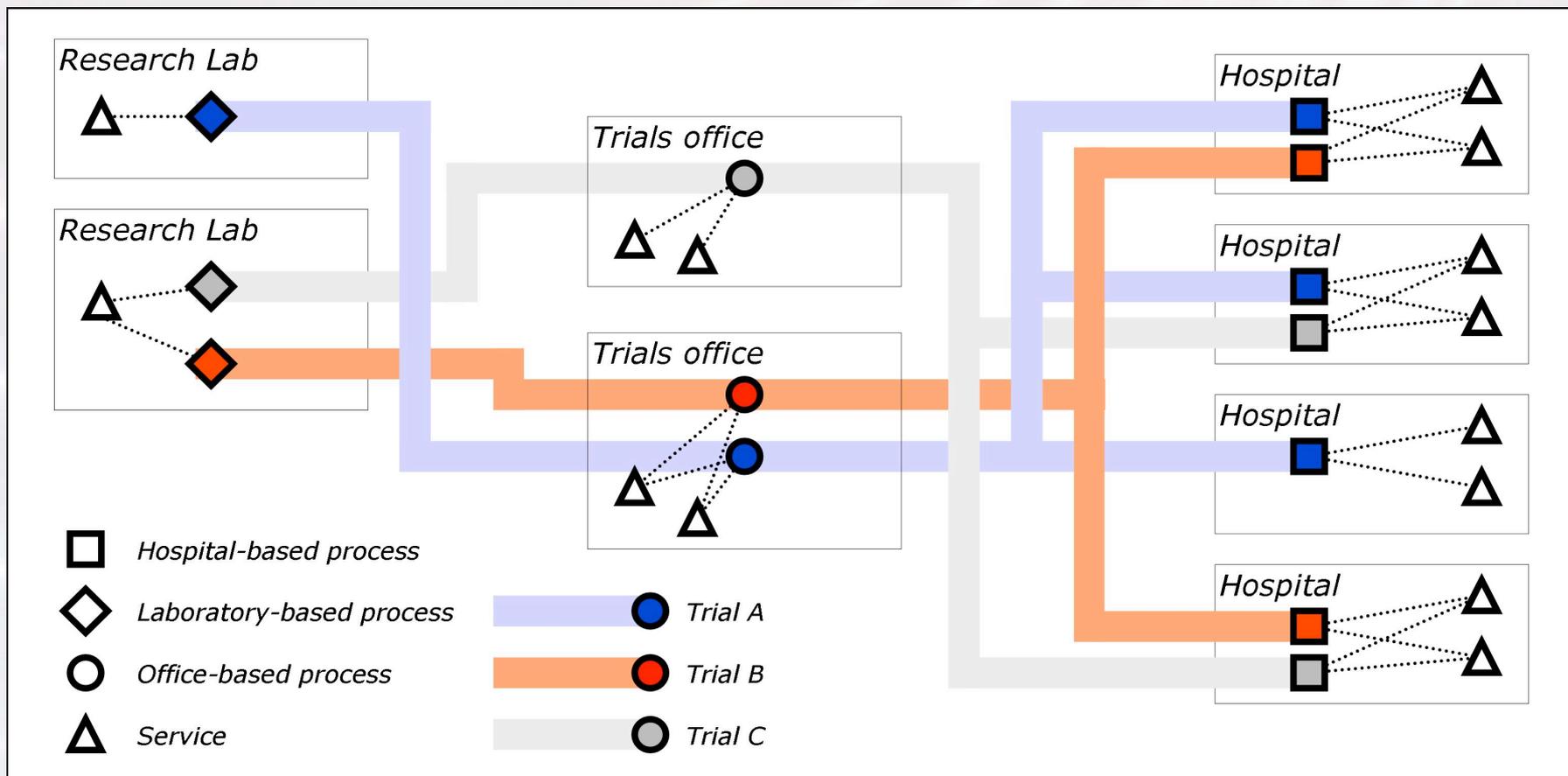
Clinical trials today



Protocol

Paper forms

Integrated trials management



Genomics Meets Medicine

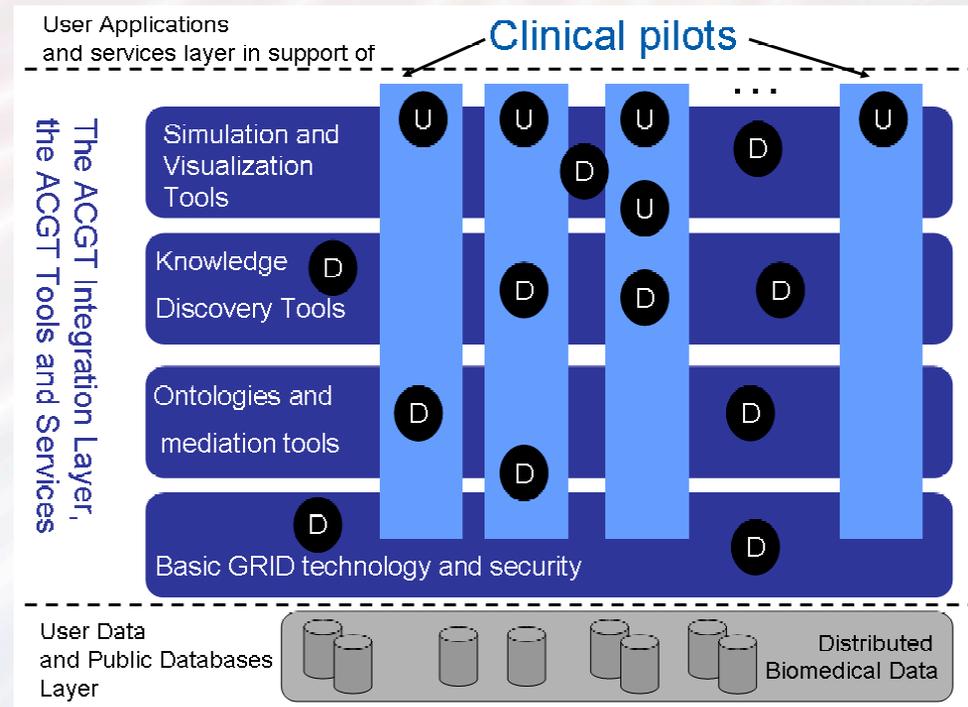
- Advancing and targeting microarray experiments in Clinical Trials
- Advancing and targeting polymorphisms identification in Clinical Trials
- Advancing and targeting systems biology approaches in Clinical Trials.

ACGT and Grid

- *ACGT* will deliver a biomedical GRID infrastructure able to *manage distributed databases* where *clinical applications*, *decision support* and *knowledge discovery* operations would access information, data and computational resources scattered over geographically dispersed databases and code-repositories,
- Special effort will be devoted for the **GRID-enabling** of the provisioned *data mining* components and *knowledge discovery* operations,
- Besides the computational empowerment, the provisioned *ACGT* GRID infrastructure will be also tuned in order to support the *seamless* aspects in *data management and processing*, thus meeting the needs of a *knowledge-GRID*.
- Grid deployment, adding extensions where needed

ACGT Architecture

- 6-layer architecture
- Conceptual model mapping applications and requirements to chosen technologies



ACGT Data Layer

- *Seamless* and *interoperable data access* services to the distributed data sources for efficient integration the clinico-genomic data.
- Include as many as possible institutions that want to share data and goals for cancer research
- Inclusive environment, VO for cancer research

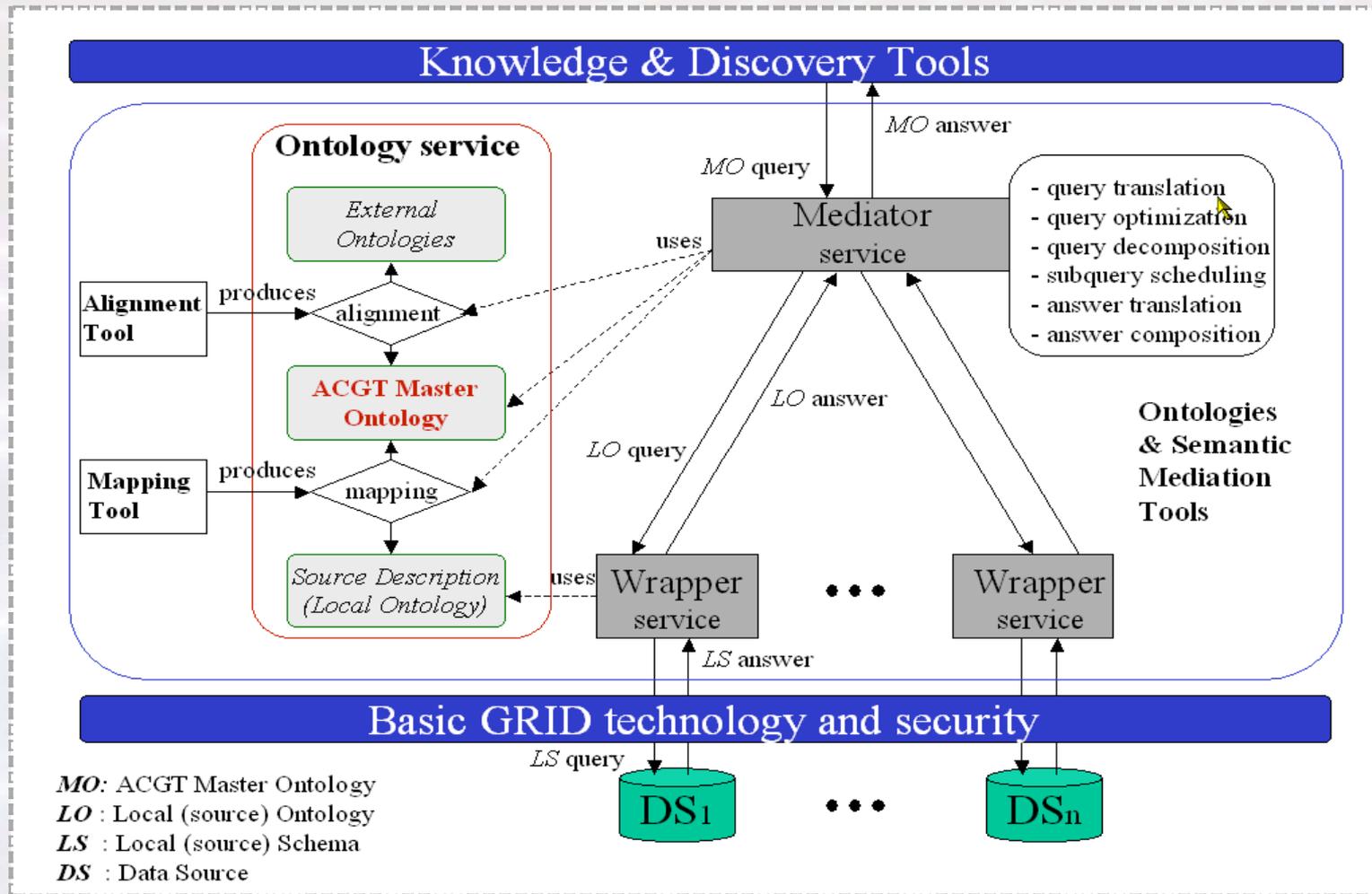
ACGT Grid Layer

- **Infrastructure services:** communication between disparate resources,
- **Resource management services:** monitoring, reservation, deployment, configuration, job mgmt, etc.
- **Data services:** database access and integration, data movement, replica management, data transformations etc.
- **Context services:** describe the resources, usage policies, etc.
- **Grid information services:** GPIR
- **Self-management,** autonomous services
- **Security services:** safe resource sharing, AAA, patient personal data security

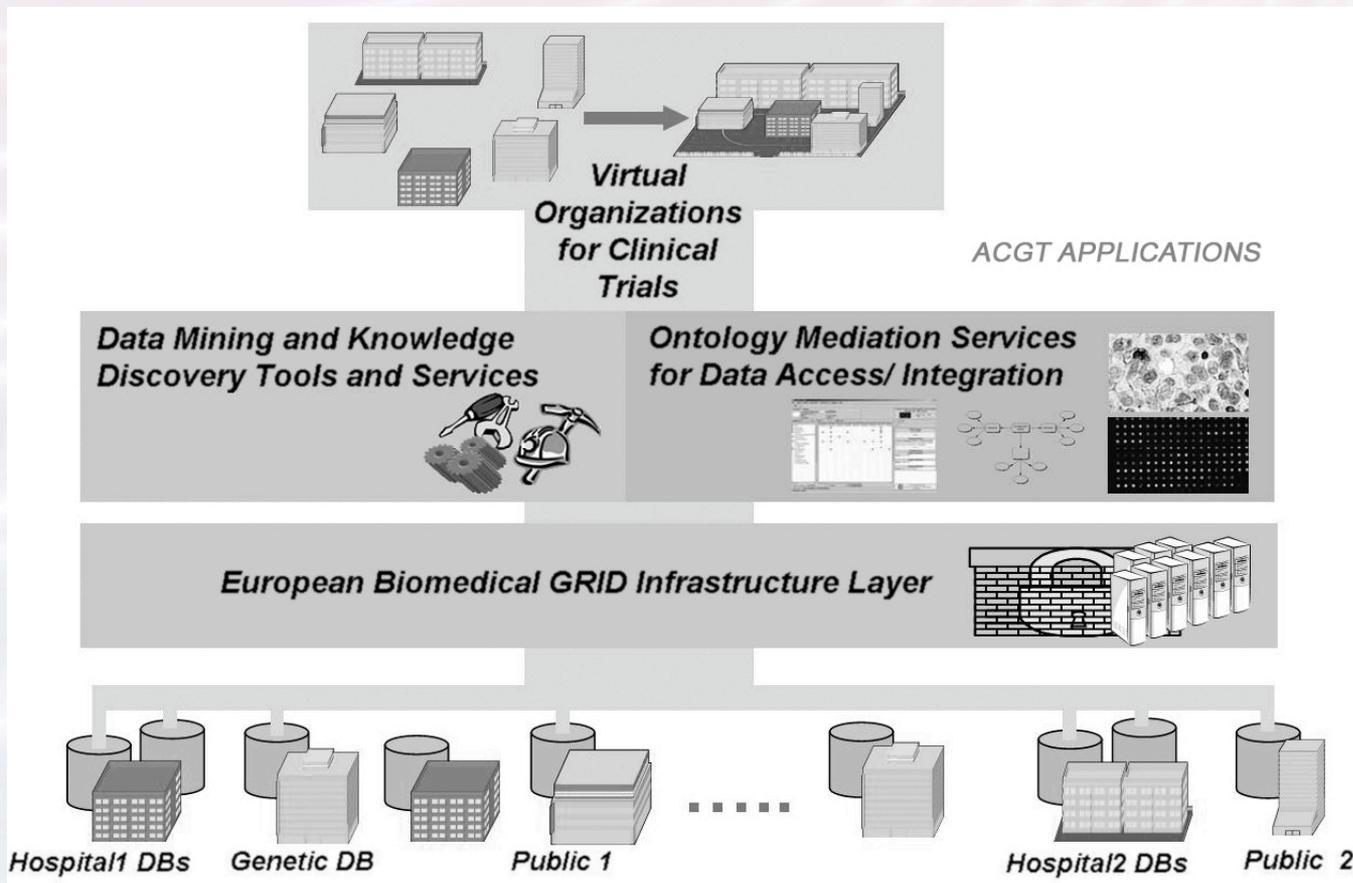
Research Challenges

- **Knowledge Mining and Clinico-Genomics:** to identify reliable *molecular-signatures* that correlate patients' *phenotypical* cancer features, i.e., medical history parameters, clinical observations, treatment outcome and disease recurrence, with the respective patients' *genomic / genotypical profiles*.
- **Knowledge Mining Tools: A common and unified suite: using R**

ACGT Integrated Environment



ACGT VOs



Benefits

- Patients
 - More rapid development of prevention & treatment
 - Greater safety of therapy
- Researchers
 - The informatics platform
- Clinicians
 - Greater knowledge base
 - Decision support systems
- Regulators & the NHS
 - Advances in care
 - Improved presentation of applicable research
- Funders
 - Increased cost-effectiveness and advance in cancer care

Summary

- Big hopes
- Will work with EGEE2 and other infrastructure projects (CrossGrid?, GridLab, DEISA)
-

Acknowledgements

ACGT partners are:

ERCIM	University of Crete
Foundation for Research and Technology Hellas	Unisersitaet Hannover
Institut National de Recherche en Informatique et en Automatique	Poznan Supercomputing and Networking Center
University van Amsterdam	Custodix
Philips Electronics Nederland B.V.	Healthgrid
Association Hospitaliere de Bruxelles – Centre Hospitalier Universitaire Bordet	Institute of Communications and Computer Systems
Institut Suisse de Bioinformatique	Universitaet des Saarland
Lunds Universitet	S.C. SIVECO ROMANIA SA
Universidad de Malaga	Facultes Universitaires Notre-Dame de la Paix
Universidad Politecnica de Madrid	Universitaet Hamburg
Fraunhofer-Gesellschaft zur Foerderung der angewandten Forschung	The Chancellor, Masters and Scholars of the University of Oxford
A. Persidis & SIA O.E.	Hokkaido University
	Istituto Europeo di Oncologia s.r.l