Medical Image Processing in MediGRID

Thomas Steinke

Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB)

German e-Science Conference 2007

Baden-Baden, May 3rd 2007

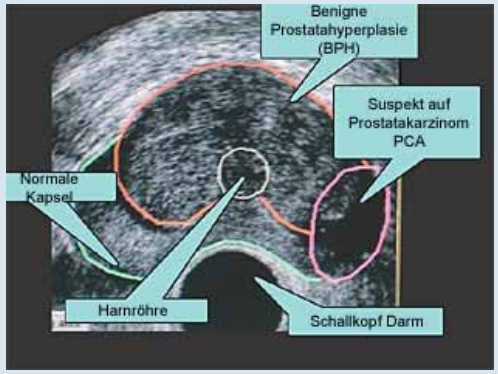
Why Using Grids in Medicine - Example: Medical Image Analysis

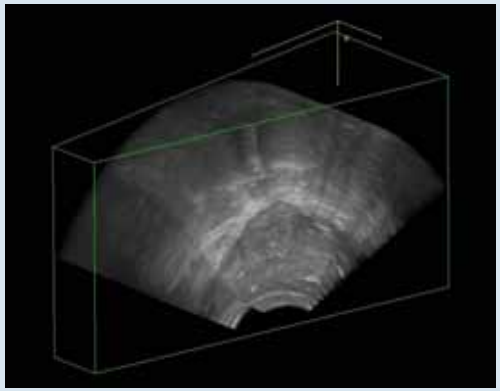
- **today:**
  - decentralized data repositories (Federalism)
    - e.g. global accessible research databases planned
    - e.g. each state in Germany operates its own mammogram database
  - research and medical care are disconnected
- **tomorrow:**
  - vast amount of data
    - **today: 7 TByte/year** (in a 1000 beds hospital)
    - **tomorrow: 6000 TByte/year**
  - linking academic and clinical research

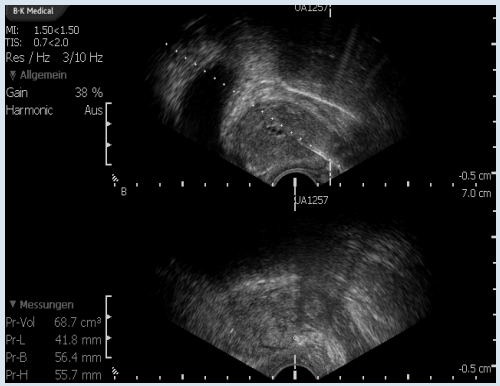
→ **Objectives of MediGRID**:

- infrastructure for collaborative platforms in medical research
- migration of pilot applications into the Grid
- establishing sustainable operational structures for applications

3D Transrectal Ultrasound (TRUS) guided Prostate Biopsy







Prostate cancer is most common cancer in men.

Ultrasound guided prostate biopsy is gold standard for prostate cancer diagnosis.

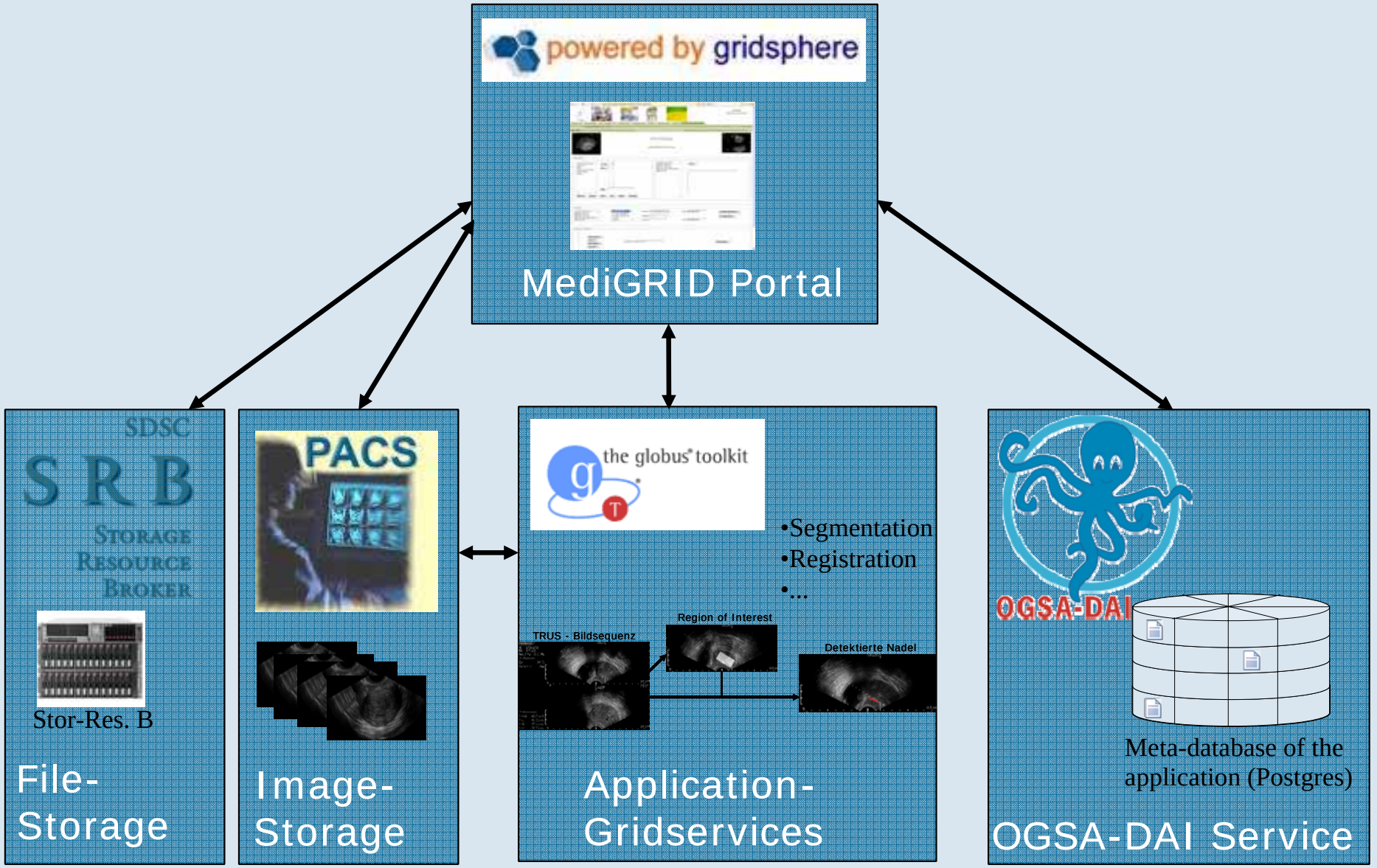
Tissue probes are taken from different parts of the prostate.

Present application determines and visualizes the position of the tissue probes within the prostate 3D volume.

Processing workflow:

- automated segmentation of the biopsy needle in the guiding 2D ultrasound images
- registration of 2D image in 3D volume
- visualization
- optional: classification, consultation of image retrieval systems

Overview


powered by gridsphere
MediGRID Portal
SDSC
SRB
Storage Resource Broker
Stor-Res. B
File-Storage
PACS
Image-Storage
the globus toolkit
•Segmentation
•Registration
•...
TRUS - Bildsequenz
Region of Interest
Detektierte Nadel
Application-Gridservices
OGSA-DAI
Meta-database of the application (Postgres)
OGSA-DAI Service

MediGRID Software Architecture

User / Application Developer

Portal

powered by gridsphere

CLI

VOMRS

MediGrid Applications

Bio-informatics

Med. Image Processing

Biomed. Ontologies

MediGrid Services

Workflow Management (GWES)

Resource Management (D-GRDL)

Data Management

SRB

OGSA-DAI

Security

Basic Grid Services Integration Layer

GLOBUS 4

Resource Monitoring

Resource Accounting

Resourcen & Locale Services

Job Management

Data Access

Metadata Mgmt.

Dataflow Monitoring

open

Database

CPU

Storage + Archive

protected

Database

CPU

Storage + Archive

MediGRID Middleware

MediGRID Applications

Challenges in Data Management

Types of Data:

- Metadata in relational databases
- 2-D, 3-D movies, images respectively in DICOM format

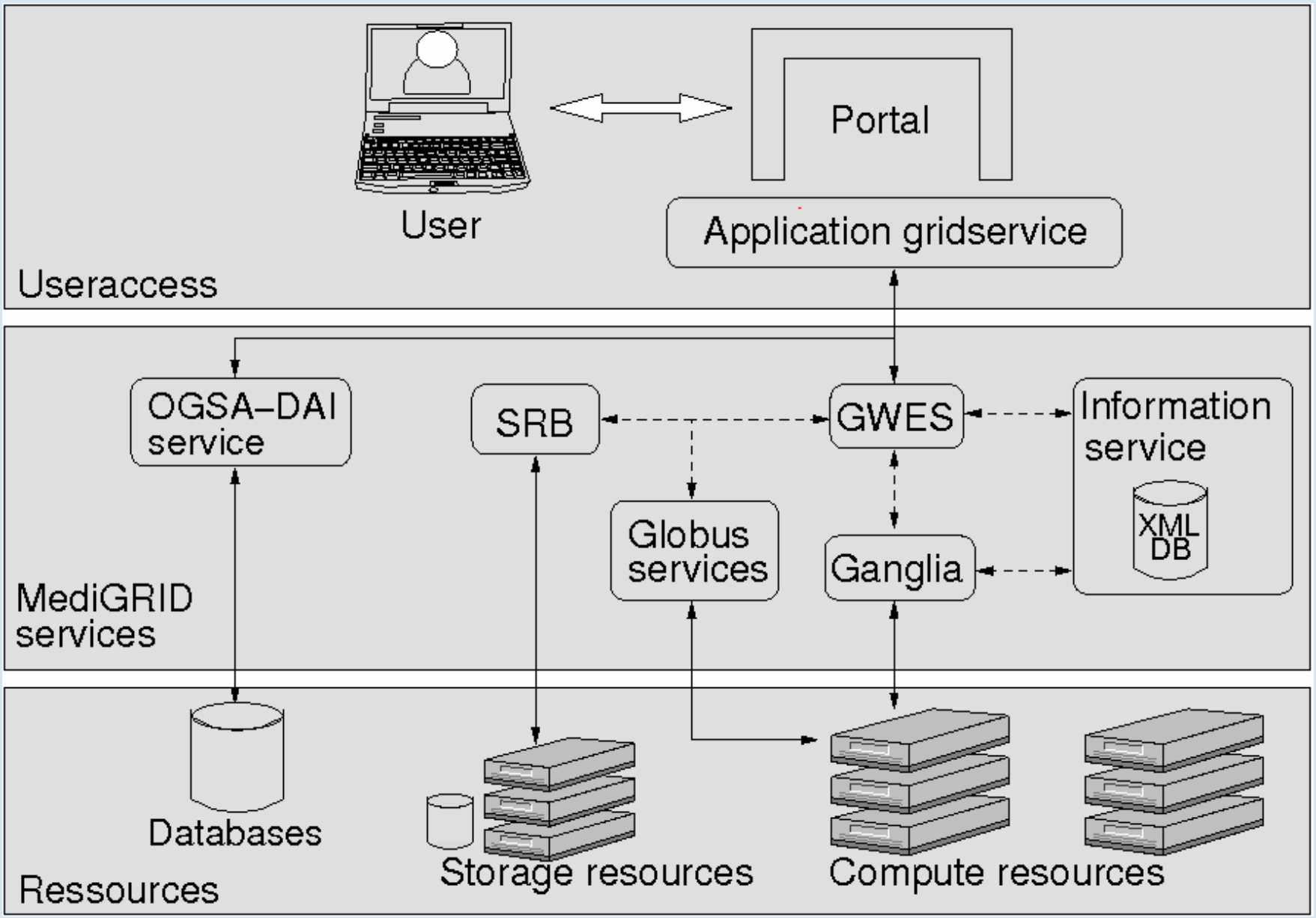
Requirements:

- Strict patient privacy laws: Pseudonymous data
- The patient data should never leave the hospitals and only authorized persons should access data.
- Data integrity, security, and authorizations are most important (end-to-end).
- Data has to be encrypted during transfer to Grid.

Data Management with OGSA-DAI and SRB

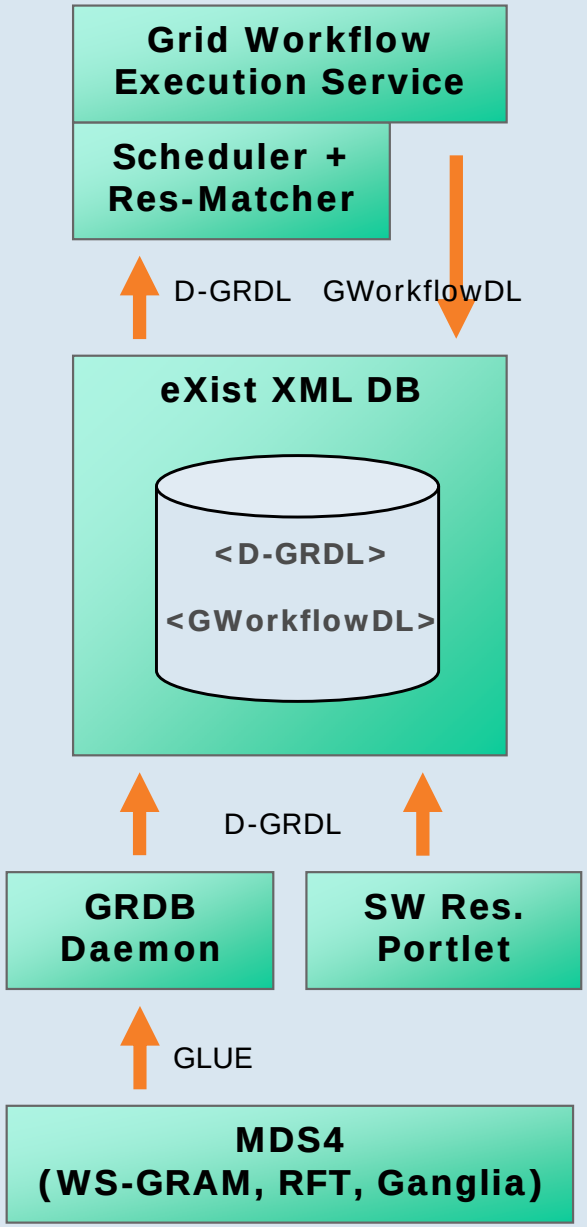
- **OGSA-DAI** (for relational data):
  - Transparent access to databases from any grid resource
  - Fine granular authorizations
  - Database servers are safe from outside world
  - Guarantee of data integrity and security end-to-end
- **SRB** (for images and other files):
  - Global namespace, global identifiers
  - Replica management
  - GSI user authentication
  - User/Group access rights
  - User-defined metadata (attribute-value-unit triples)
  - User interface: SCommand client, JAVA, and C APIs

Data Management for Medical Image Processing



Information Service and Metadata Management

- Task: gather and store necessary resource information
- Central metadata repository – eXist XML database
- Types of metadata:
  - Software resources:
    - programs & web services
    - properties & dependencies on other resources
  - Hardware resources:
    - static information: system, batch system
    - monitoring data (utilization)
    - GRDB daemon queries MDS4; WS-GRAM, RFT and Ganglia are information provider
    - GLUE schema output of MDS4 is converted into D-GRDL
  - Workflows:
    - active/ completed workflows (GWorkflowDL)
- HW and SW metadata are used by GWES for resource matching and program execution
- Metadata for data is handled by the SRB.



Challenges in Security

The legal framework implies special requirements regarding data security and data protection:

(1) Confidentiality

(2) Integrity and authenticity

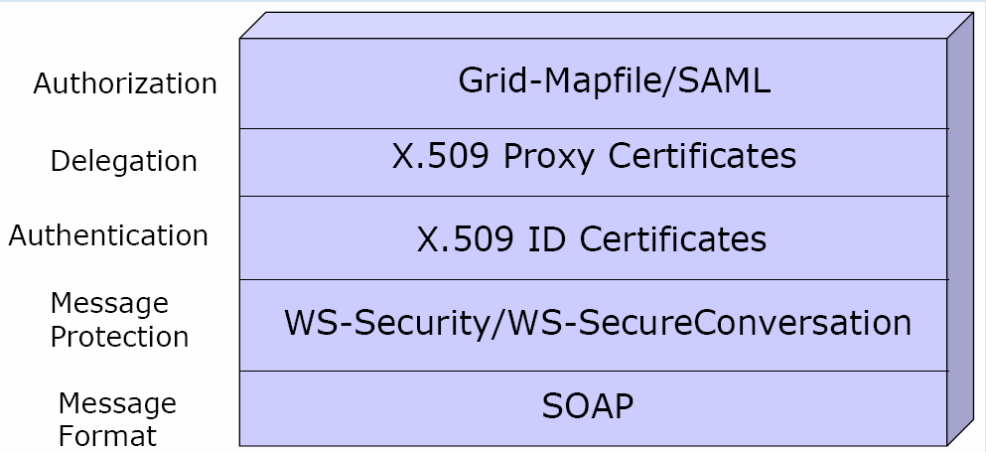
(3) Accessibility

(4) Accountability

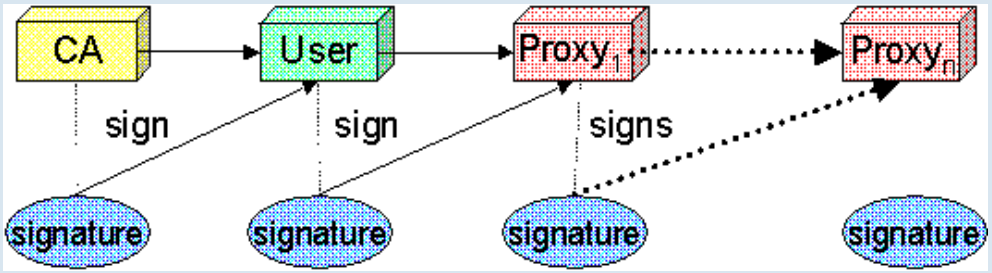
Globus Security Provides ...

- Authentication,
- communication security,
- Authorization, and
- supporting functions for managing user credentials and maintaining group membership information.

Security Layers in GT4*



Proxy Certificates (MyProxy) **



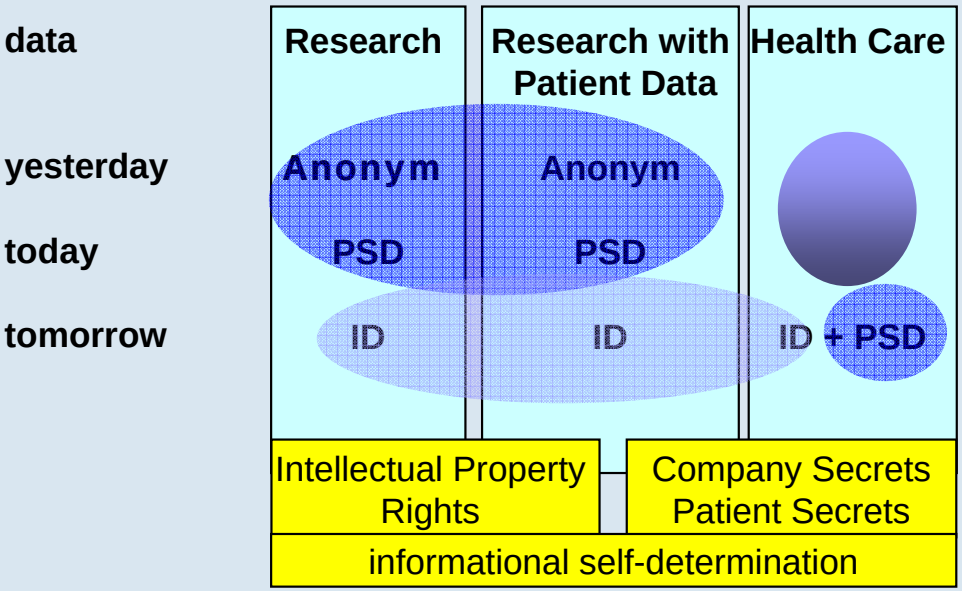
* F. Siebenlist, Von Welch. *Grid Security: The Globus Perspective.* In *GlobusWORLD 2005, Feb 7-11, Boston, MA.*

** The GT4 Securtiy Team, GT 4.0 Security: Key Concept. http://www.globus.org/toolkit/docs/4.0/security/key-index.html [10.12.2006]

„Enhanced Security“ AP InGrid / MediGRID / DGI

- **Audit:** *a posteriori* logs; data provenance & annotation during processing steps
- **Trackability:** *a priori* information or policies; about when/where transfers, transactions, processing and storage of patient data will occur
- **Fine-granular access control:** not only at file level but within subtrees of a document tree
- **Confidentiality:** according to fine-granular access control
- **Trust & trust delegation:** at level of software instances but additional for people, organizations etc. as well
- **Safety:** physical control of data in unsecured Grids, policy-based storage, → Data & Information Management

Outlook: Enhanced Security



ID: identifiable data

PSD: pseudononymous data

Very Enhanced Security

- automatic splitting of data sets into non-identifiable parts (→ genome)
- implementation in common environments via standardization (e.g. Open Grid Forum)
- on-going development of generic security concepts for hardly identifiable data sets

Questions?

More MediGRID Information at:

http://www.medigrid.de/

BACK UP SLIDES

Other Data Storage and Management Tools

dCache/SRM:

- Huge amounts of data
- Distributed among a large number of heterogeneous server nodes
- Single virtual file system tree
- Variety of standard access methods
- Hot spot determination, replication, and recovery etc.

DataFinder:

- Management software for scientific and engineering data
- Existing or newly created data can be structured and provided with metadata
- Self-defined data models
- Based on open and flexible standards like XML and WebDAV
- Interface to openAFS and GridFTP