

TMF-Workshop ID-Management 15.12.2008

Pseudonymisierungskonzepte in Grid-Anwendungen

Prof. Dr. Ulrich Sax

Frank Dickmann (S@M), Krister Helbing (AHF)

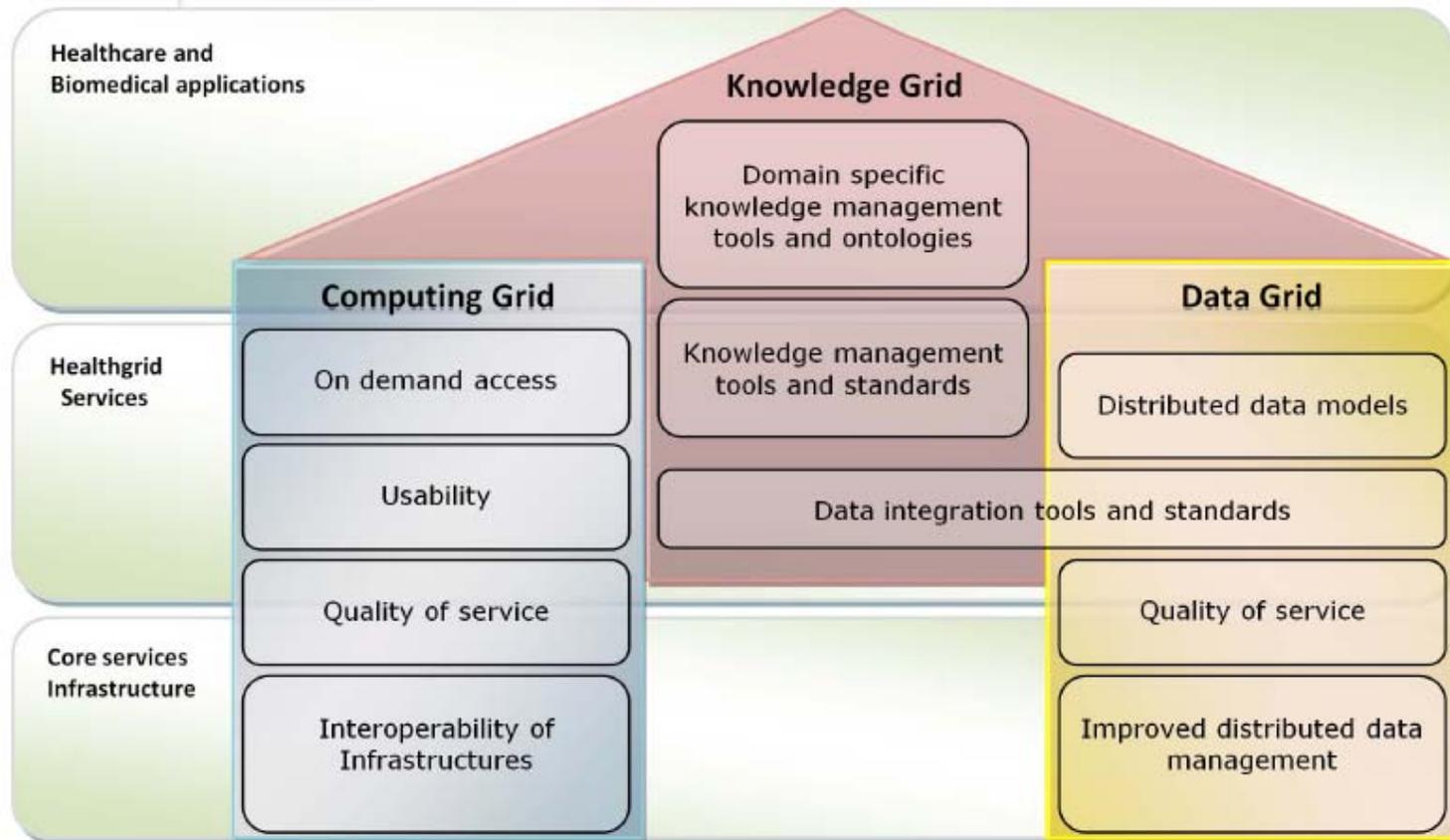
<http://www.mi.med.uni-goettingen.de>



Gliederung

- Anwendungsfälle
- Pseudonyme, Depseudonymisierung, Patientenbezogene IDs
- Datenhaltung
- Anforderungen der Datenschutzbehörden
- Stand in ausgewählten Projekten
- Zusammenfassung

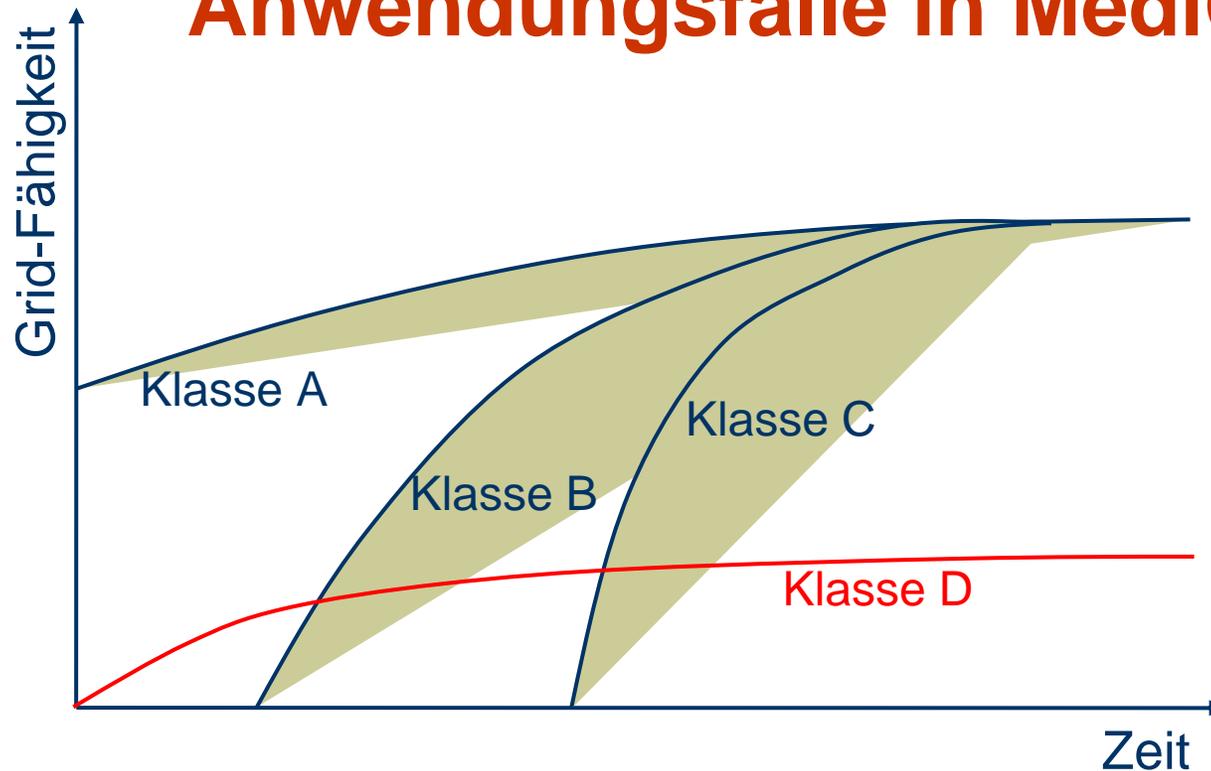
Einsatzgebiete des Grid Computing im Gesundheitswesen



Quelle: Solominides: UK eScience all Hands meeting 2008

Weiter führende Quelle: Iakovidis. (2005). "HealthGrid – 3 sided concept.", ISTAG WG, <http://www.who.int/classifications/terminology/iakovidis.pdf>.

Anwendungsfälle in MediGRID



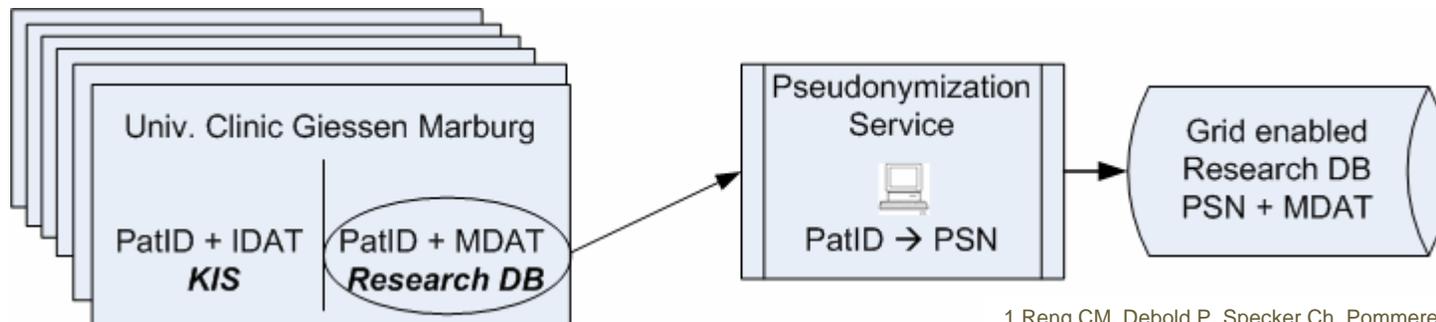
- Klasse A: z.B. Biomedizinische Informatik
- Klasse B: z.B. Bildverarbeitung
- Klasse C: z.B. Klinische Forschung (schneller durchführbar, aufgrund der Erfahrungen aus den Klassen A+B)
- Klasse D: keine Gridifizierung möglich



Anforderungen des Datenschutzes im GRID

allgemeine Anforderungen

- Trennung von medizinischen Daten (Signale) und identifizierenden Daten (Name, Geburtsdatum,...)
- Optimaler Schutz der ident. Daten: *Anonymisierung* verhindert eine Zuordnung der rückgemeldeten Analyse-ergebnisse
- Mehrstufige *Pseudonymisierung* als datenschutzrechtlich vertretbare Alternative¹



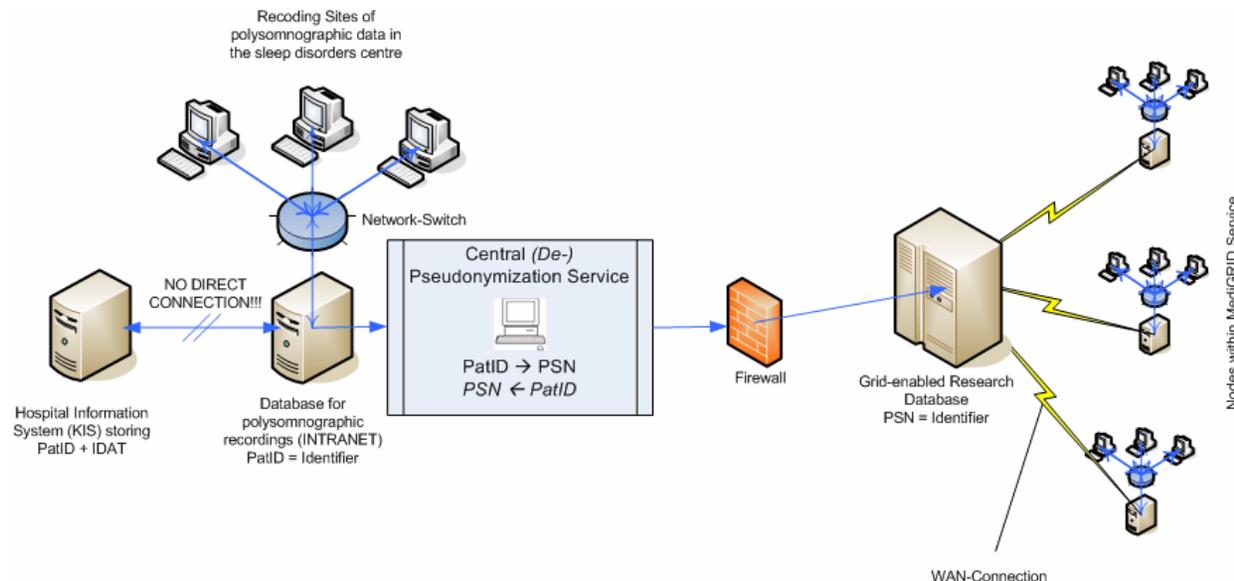
¹ Reng CM, Debold P, Specker Ch, Pommerening K.
 Pseudonymisierung von Forschungsdaten in wissenschaftlich
 fokussierten Forschungsnetzen. Generische Lösungen zum
 Datenschutz für die Forschungsnetze in der Medizin.
 Medizinisch Wissenschaftliche Verlagsgesellschaft; 2006. 43-65



Anforderungen des Datenschutzes im GRID

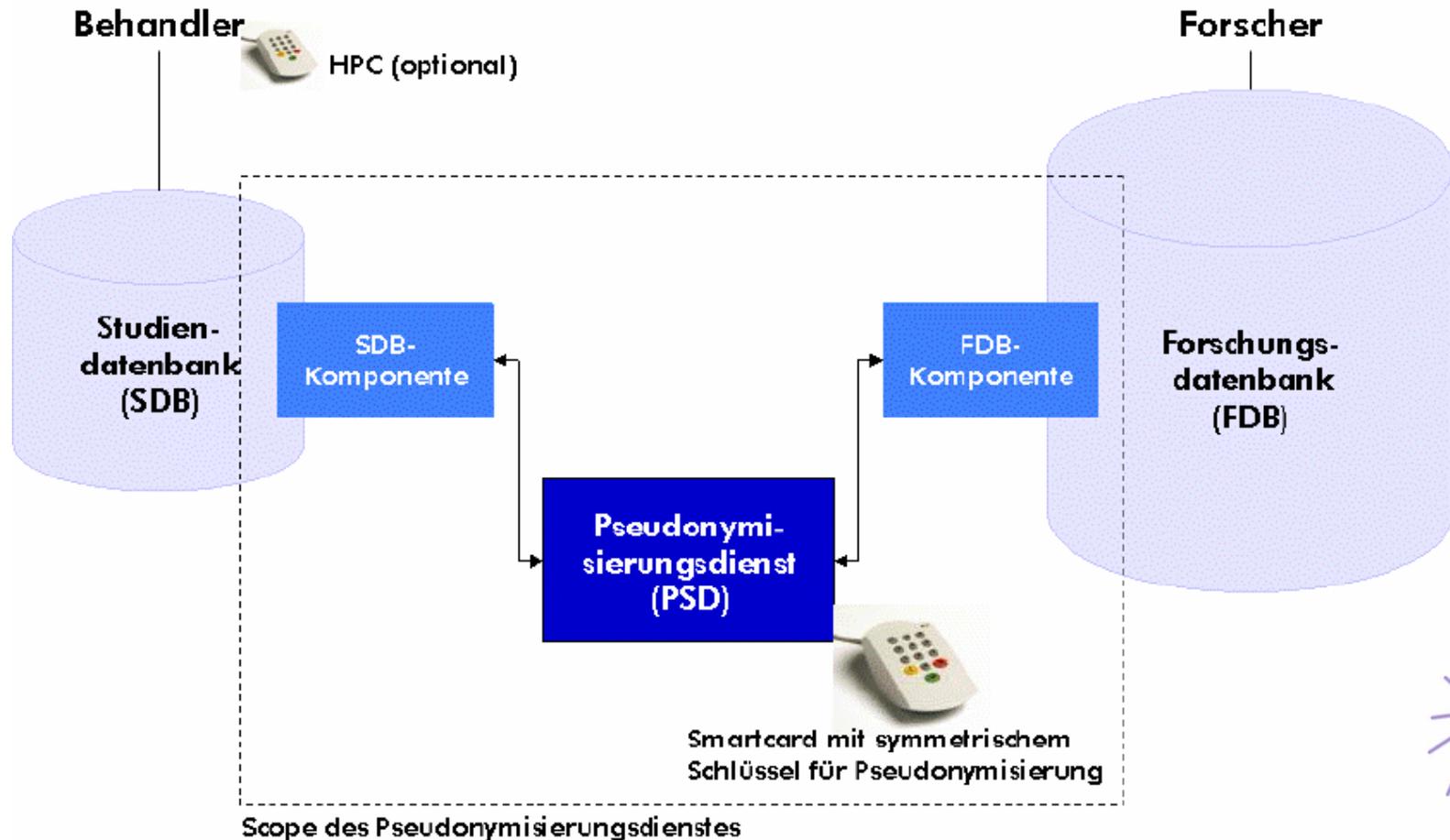
Anforderungen für die Nutzung in schlafmedizinischen Studien

- Automatisierte Verarbeitung von Biosignalaufzeichnungen in der GRID-Umgebung (z.B. automatisierte EKG Analyse)
- Nutzung des European Data Format (EDF) als Datenformat für die Signaldaten in der GRID-Umgebung



MediGRID

PSD der TMF (Tembit)

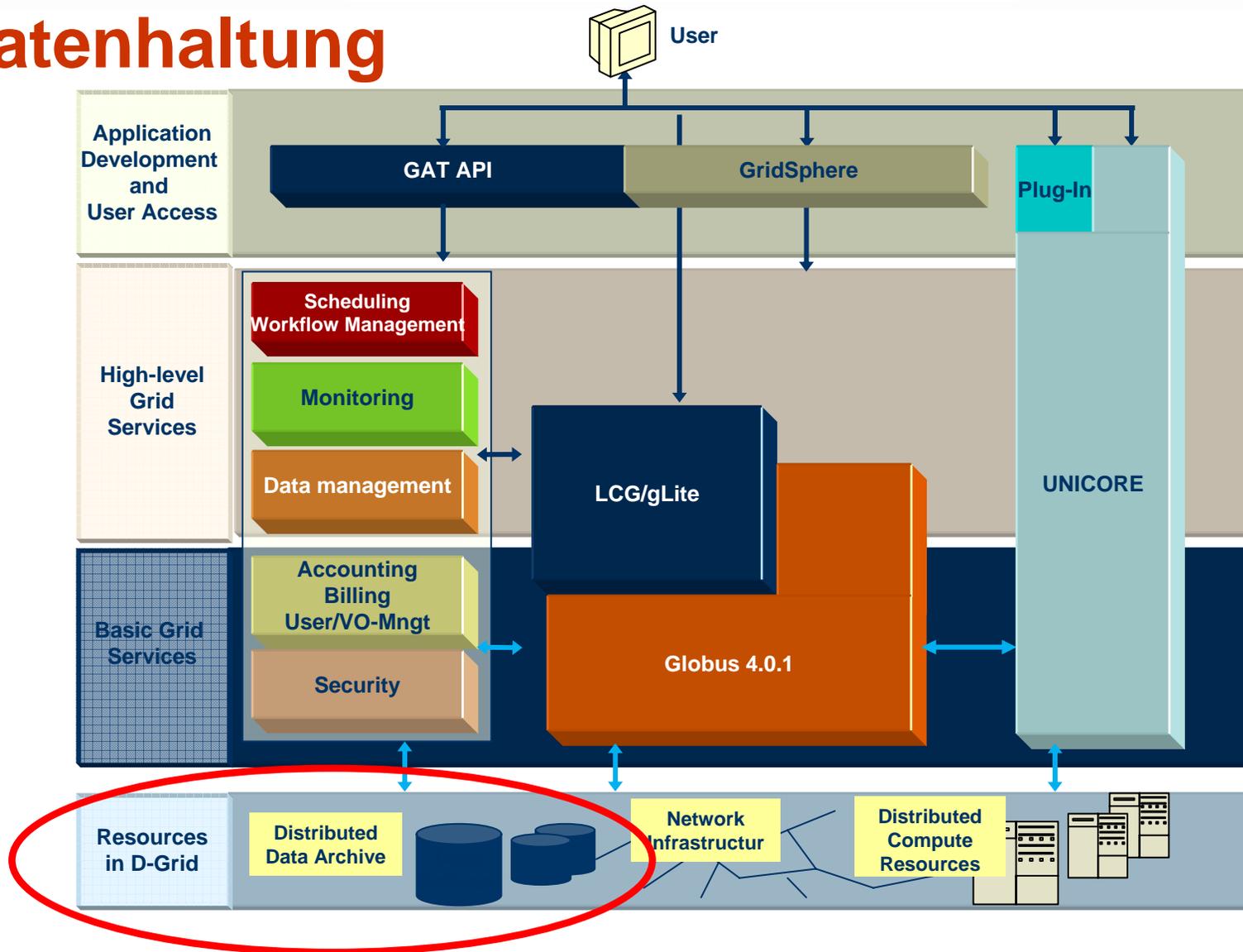


Pseudonyme, Depseudonymisierung, Patientenbezogene IDs (MediGRID)

- Anwendungsfall Computing Grid:
 - Alle Daten müssen vor dem Hochladen pseudonymisiert werden
 - Geeignete PID- bzw. PSD-Werkzeuge vorausgesetzt
 - Spezifisch für Datentyp (z.B. DICOM, CDISC ODM, andere)
- Anwendungsfall Storage Grid:
 - Derzeit nur mit Daten ohne Reidentifizierungsrisiko nutzbar
 - Skalierungsproblem bei Autorisierung in Grid-Datenhaltung
 - Spezialisierung der Ressourcenbetreiber notwendig?!



Datenhaltung



Anforderungen der Datenschutzbehörden an Grid-Anwendungen

- Derzeit nicht grid-spezifisch formuliert
- Standortabhängige Vorgehensweise
- Schlüsselkonzept: Einhaltung der generischen Datenschutzkonzepte der TMF bezüglich getrennter Datenhaltung
- Problem: Ressourcen sind in der Hand von Ressourcenprovidern
 - Beispiel: werden die (patientenbezogenen) Daten einer SNP-basierten Korrelation von einem Rechenknoten gelöscht, wenn die Berechnung abgeschlossen ist?
 - Derzeit nur möglich: bilateral abgestimmte Policies/SOPs mit Ressourcenbetreibern



MediGRID

Networked computing resources for biomedical research



Resource usage policy phase one - development

- Promotion number: 01AK803H -

- Deliverable -

„eScience and GRID-Middleware to support the scientific working“
of the federal ministry for education and research (BMBF)

Date	02.04.2007
Version	1.7
Type	For approval from MediGRID Steering Committee
Status	Final
Authors	Y. Mohammed, F. Vlezens, U. Sax, J. Falkner, Th. Steinke

MediGRID

Networked computing resources for biomedical research



MediGRID-developer/user/guest agreement phase one – development phase

- Promotion number: 01AK803H -

„e-Science and GRID-Middleware to support the scientific working“
of the federal ministry for education and research (BMBF)

Date	27.10.2006
Version	1.0
Type	For approval from MediGRID Steering Committee
Status	Final
Authors	Y. Mohammed, F. Vlezens, U. Sax, J. Falkner, Th. Steinke

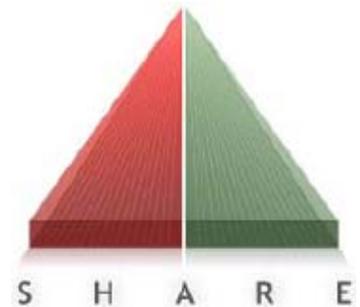
Stand in ausgewählten Projekten: caBIG

- DE-ID™ to Advance Progress of Cancer Research Data-Sharing Grid
- Esp. for Tissue Banks and Pathology Workspace
- „DE-ID does not simply eliminate any or all of the 18 HIPAA safe harbor patient identifiers, but maintains the research integrity and usability of the data through **name proxies** and **date offsets** within the de-identified file“

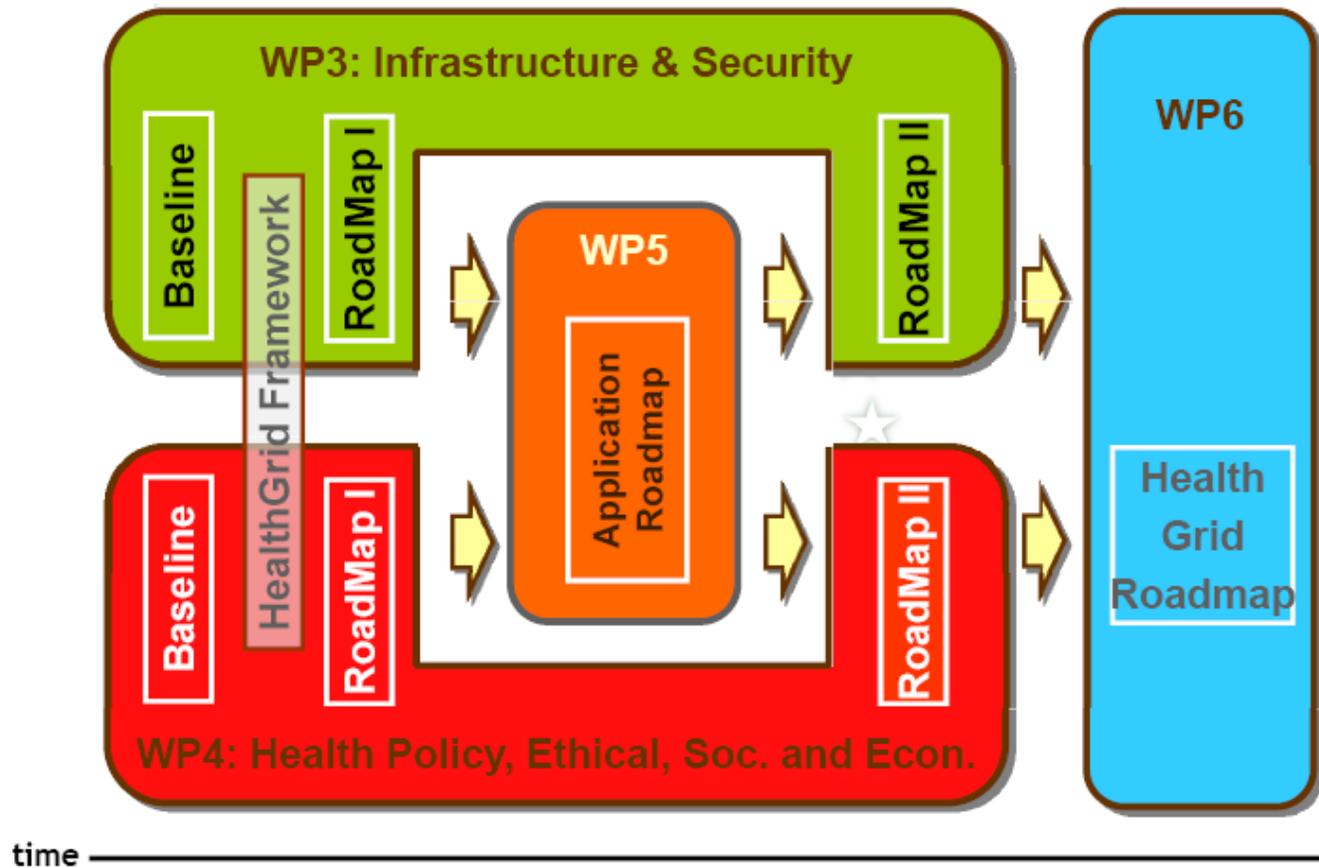


Stand in ausgewählten Projekten: SHARE

Challenge	Description of the requirement
RCDG1	<ul style="list-style-type: none"> • Easily installed grid middleware for health environments. • Low maintenance and administration.
RCDG2	<ul style="list-style-type: none"> • Data architectures/tools for private data dissociation, pseudo/anonymisation and encryption. • Automatic compliance with legal requirements.
RCDG3	<ul style="list-style-type: none"> • Exploitation models and guarantees QoS for services. • Advance resource reservation with pre-negotiated QoS.
RCDG4	<ul style="list-style-type: none"> • Scalable data cataloguing and data transfer.
RCDG5	<ul style="list-style-type: none"> • Storage services for easy upload/download of large binary objects.
RCDG6	<ul style="list-style-type: none"> • Distributed data models and repositories multiscale data.
RCDG7	<ul style="list-style-type: none"> • Enhanced standards for data protection in web services environments.



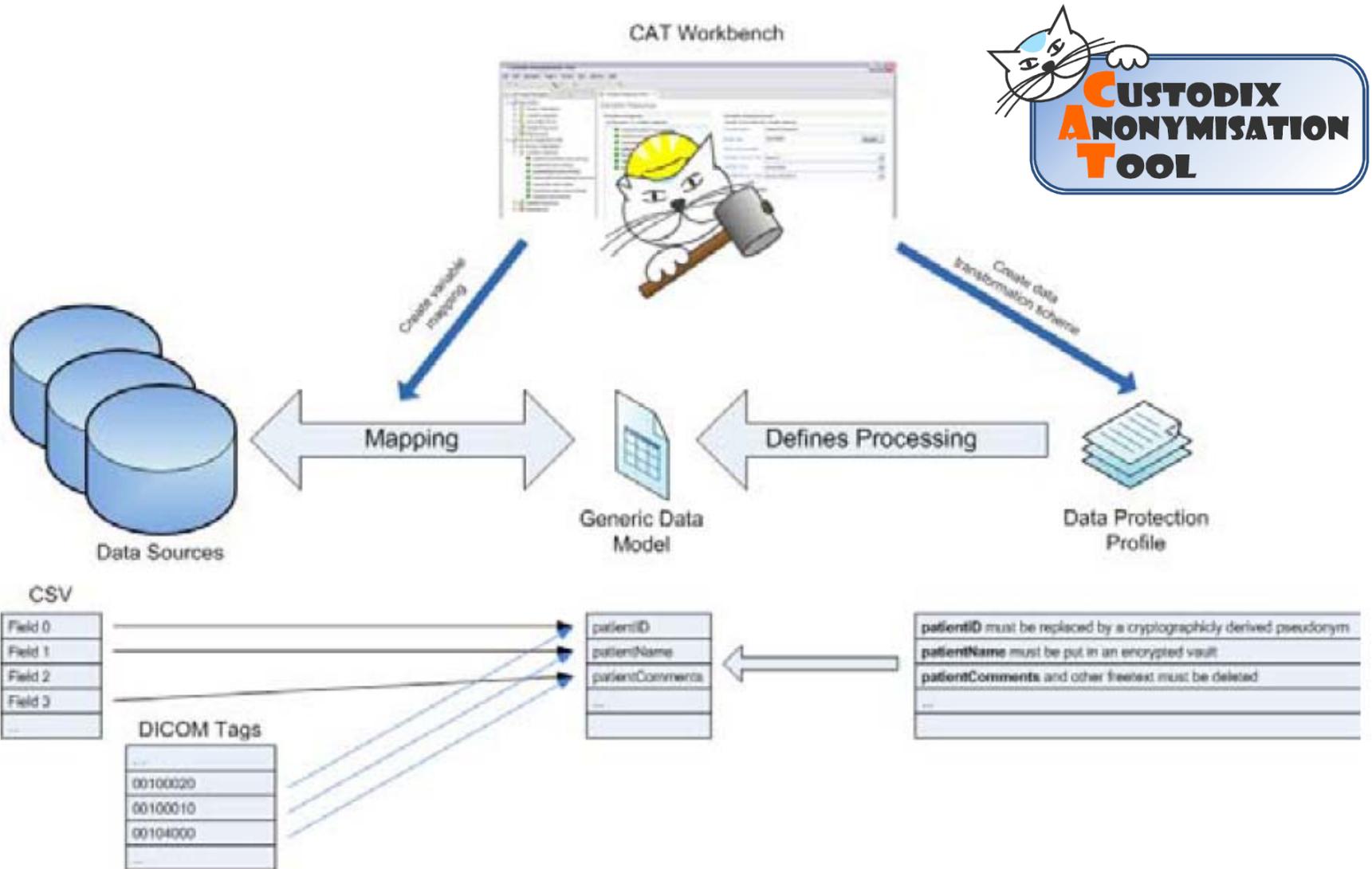
SHARE: Roadmap 2



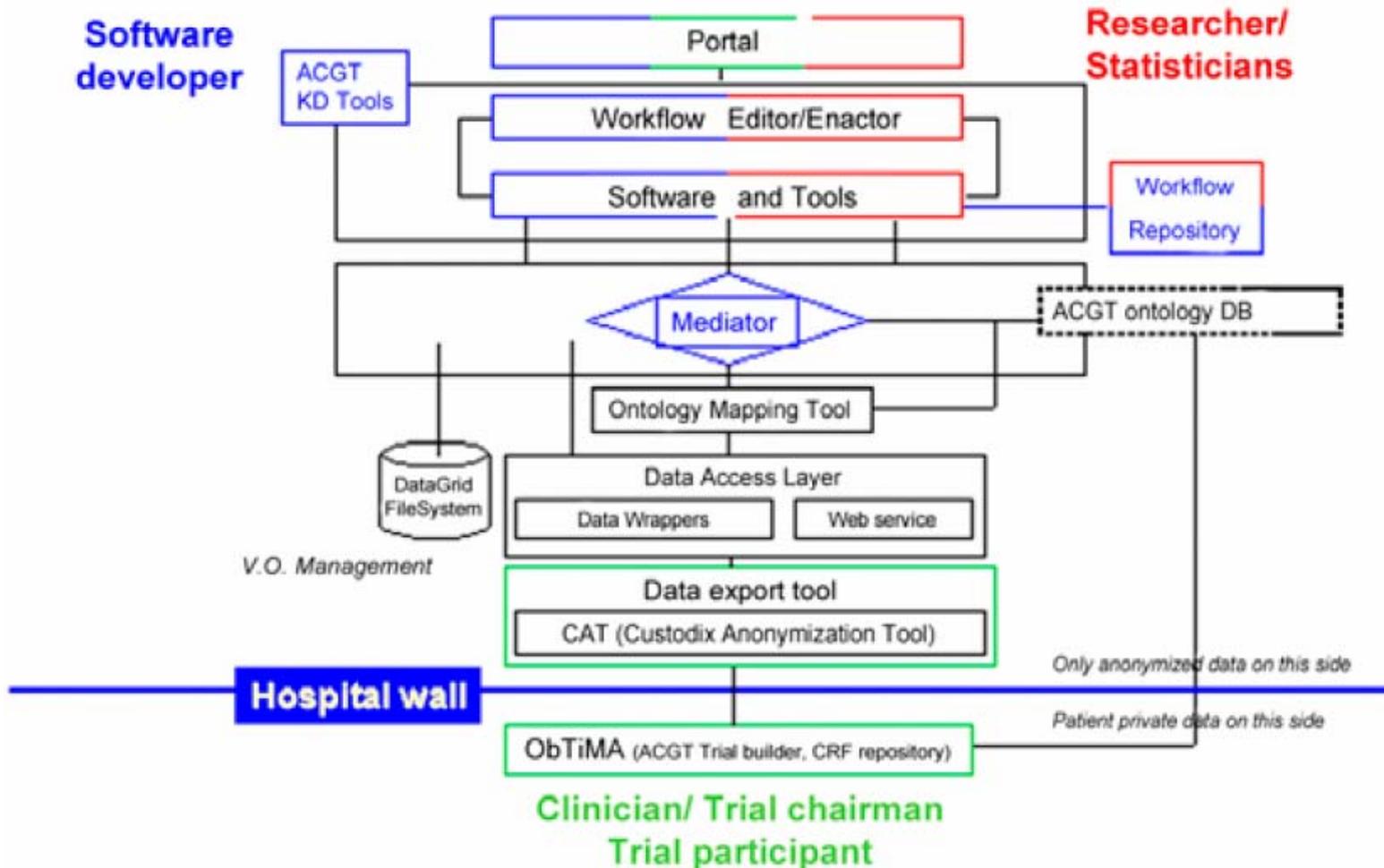
Stand in ausgewählten Projekten: ACGT

- ACGT Data Protection Framework
- “Custodix Anonymisation Tool” (CAT)
- „best possible equilibrium between the interests of patients giving their data and/or body material for ACGT and the researchers producing results out of these data/materials within the project.“





ACGT IT Platform



Zusammenfassung

- Derzeit keine Daten mit Reidentifizierungsrisiko in MediGRID
- Klassische Pseudonymisierungskonzepte greifen auch im Grid
- Spezifische PSD-Lösungen je Anwendungsklasse

- Computing Grid, Storage Grid im prä-Produktionsstadium (z.B. caBIG, MediGRID)
- Collaboration Grid im Projektstadium (z.B. TextGrid, SHARE)
- Knowledge Grid in statu nascendi (Ziel z.B. von ACGT, SHARE)

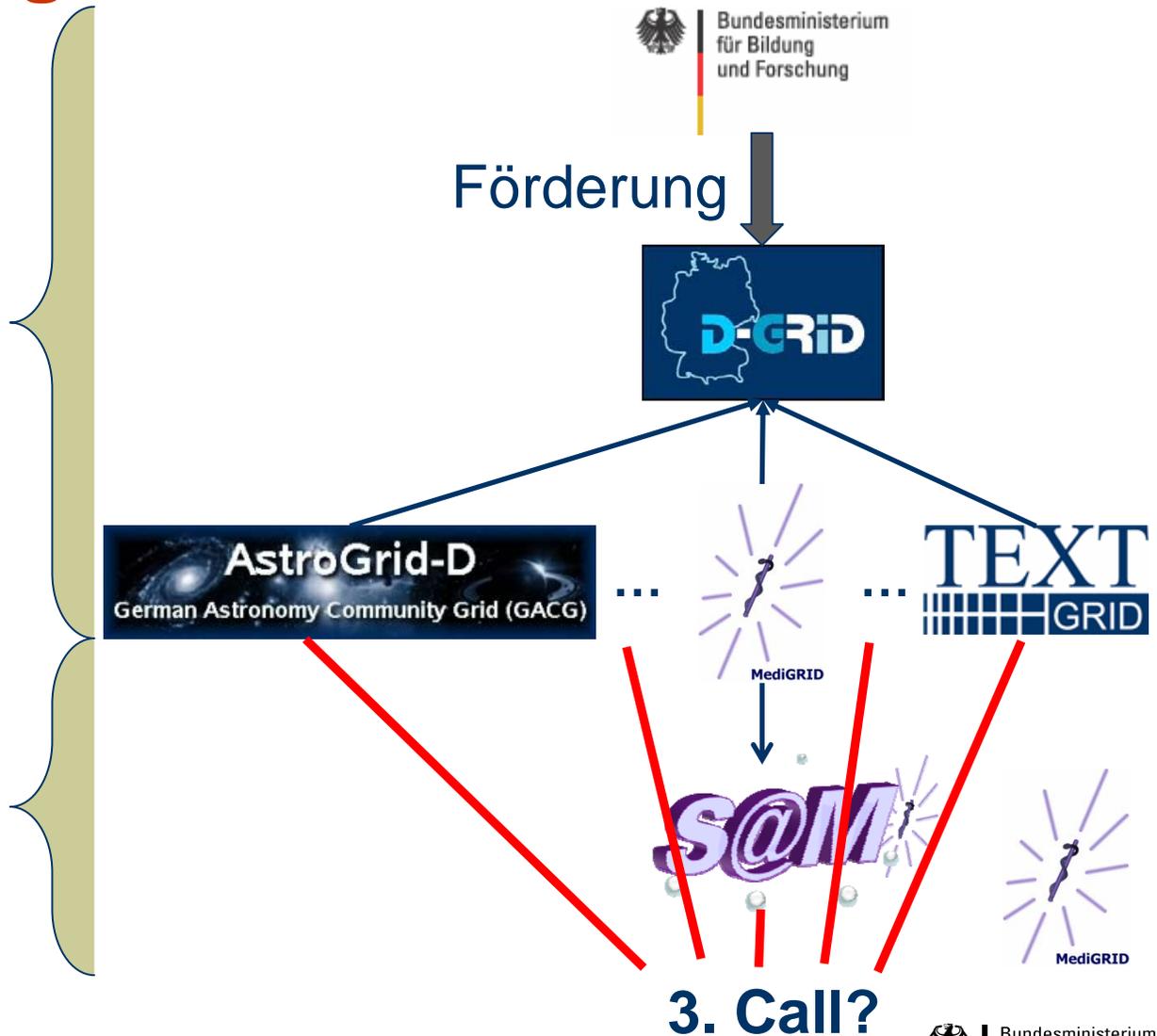
- **Für Grid-Anwendungen je nach Anwendungsklasse weitere Maßnahmen zur Senkung des Reidentifizierungsrisikos notwendig**



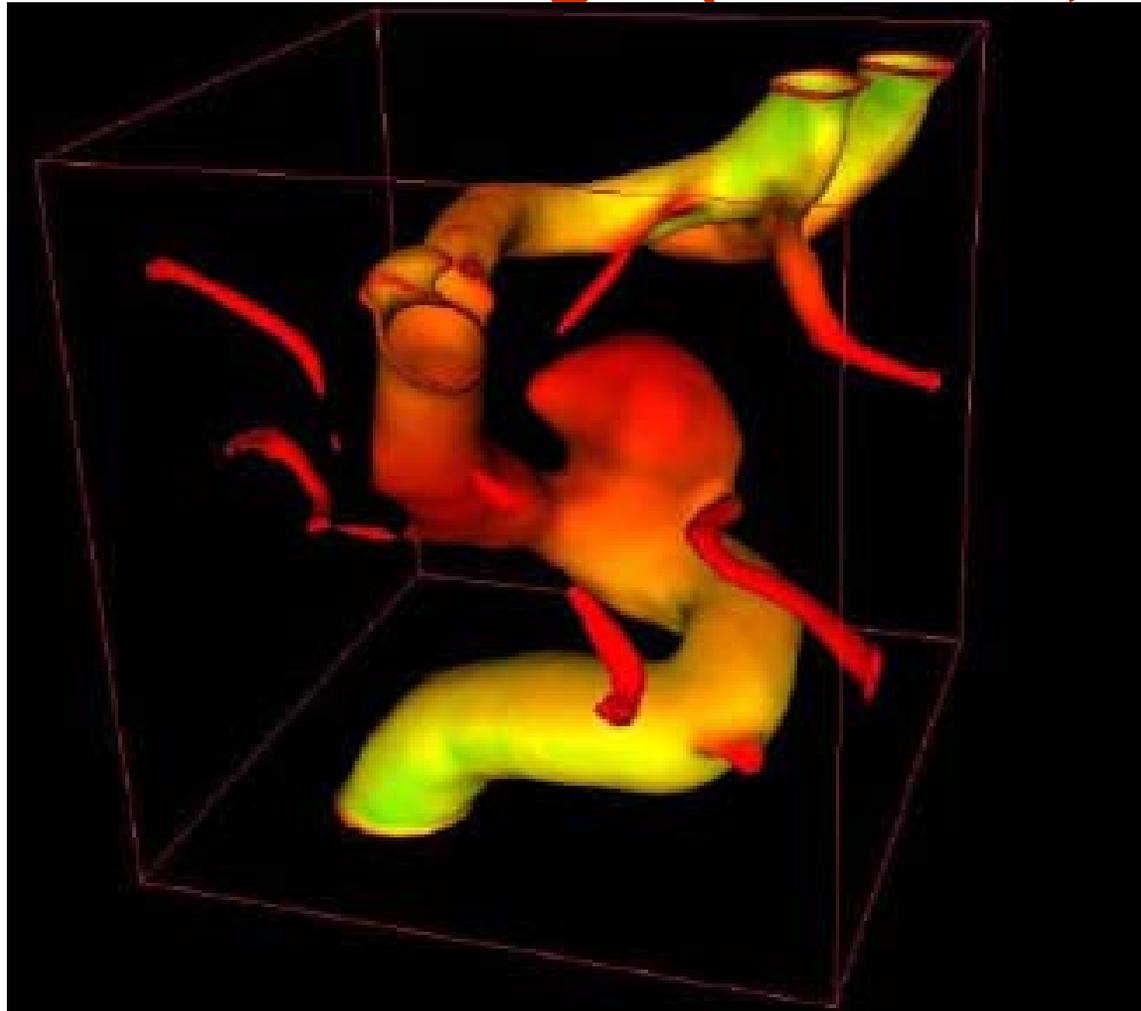
Grid Computing in Deutschland

„Experimentier“-
phase

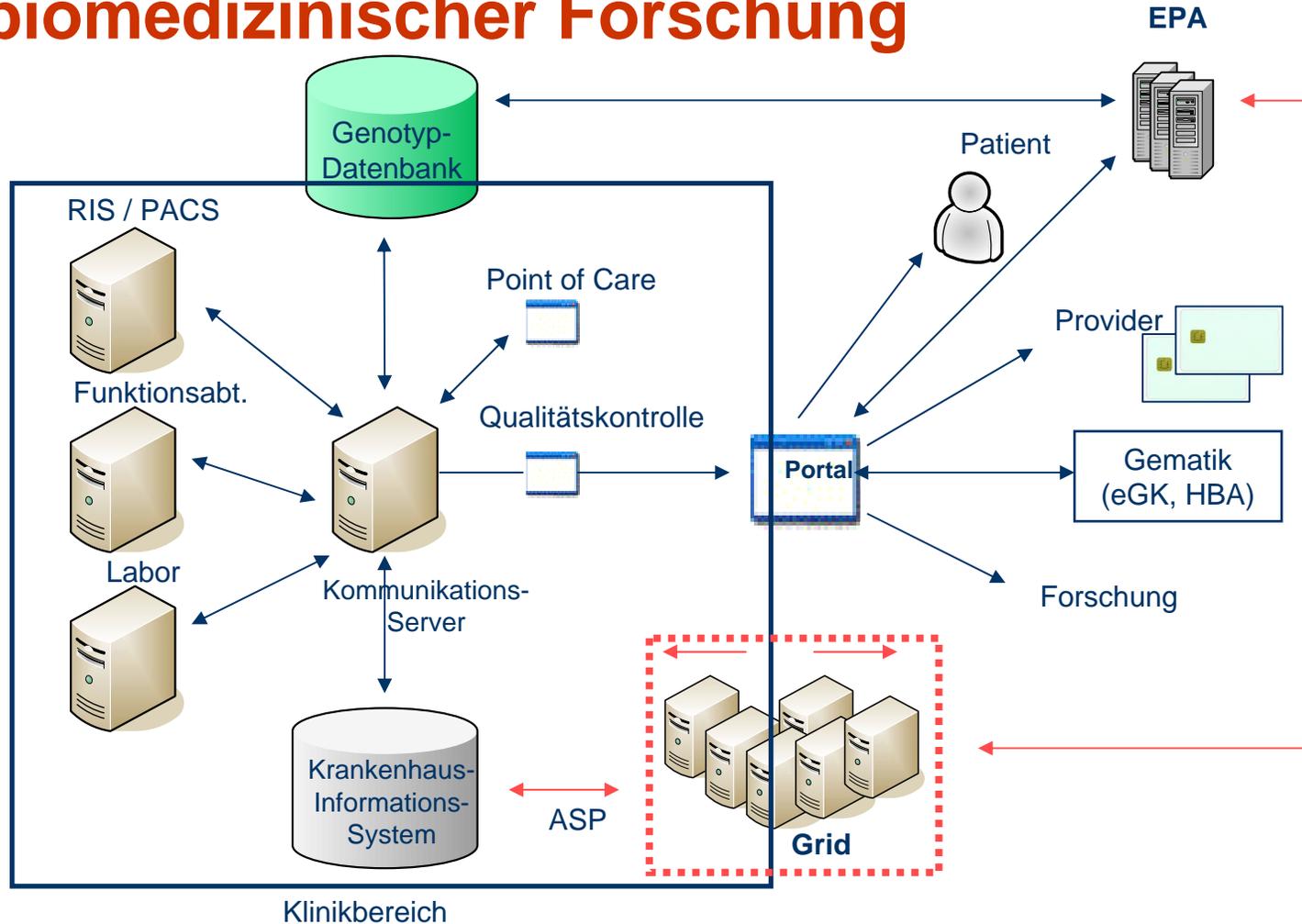
Reifephase



Virtuelle Gefäßchirurgie (Beronov, ER)



Vision: Grid Computing in klinischer Versorgung und biomedizinischer Forschung



Quelle: Sax, Weisbecker. (2007), „Auf dem Weg zur Individualisierten Medizin - Grid-basierte Services für die EPA der Zukunft“, Telemed 2007, <http://www.telemed-berlin.de/telemed2007/Beitraege/TELEMED-2007-10-Sax.pdf>.

