

Grid Provenance

Guy K. Kloß (SC-VK)

Köln-Porz, 11.01.2006



Deutsches Zentrum
für Luft- und Raumfahrt e.V.
in der Helmholtz-Gemeinschaft



Überblick

- Projekt-Eckdaten
- Einführung und Ziele
- Provenance – Wie geht das?
- Status – Die Implementierung
- Arbeiten 2006 – Beispielanwendung



Projekt Eckdaten

- **Laufzeit:**
September 2004 – November 2006
- **Förderung:**
6. Europäisches Forschungsrahmenprogramm
- **Partner**
 - **IBM United Kingdom Ltd**
Koordination, Sicherheit, Skalierbarkeit, Implementierung
 - **Universität Southampton**
Architektur, Sicherheit, Skalierbarkeit
 - **Deutsches Zentrum für Luft- und Raumfahrt e. V.**
Applikation: Luft- und Raumfahrt
 - **Universität Wales, Cardiff**
Tools und Installation
 - **Universitat Politecnica de Catalunya, Barcelona**
Applikation: Organ-Transplantations-Management
 - **MTA SZTAKI, Budapest**
Applikation: Organ-Transplantations-Management





„Provenance“ – Was ist das?

➤ Duden Fremdwörterbuch

Provenienz [...*we*...; zu *lat.* provenire ‚hervorkommen, entstehen‘] *die*; -, -en: Herkunft, Ursprung

➤ Wiktionary (deutsch)

Provenienz

Bedeutungen:

[1] Herkunft, Ursprung

[a] die über die Vorbesitzer nachgewiesene Herkunft einer Ware.

[b] im Buchwesen den oder die Vorbesitzer eines Schriftstücks, meist private Sammlungen oder geistliche Institutionen, aus denen das Exemplar in den heutigen Besitz einer Bibliothek gelangt ist: Provenienz (Buch).

[c] im Archivwesen *die Herkunft betreffend*

[d] in der Medizin wird *Provenienz* auch einfach im Sinne von *Herkunft* benutzt.

➤ **Verfolgung und Analyse der Herkunft von Daten in Grids für komplexe Applikationen**



Hintergrund: Bestimmungen und Konformität

- Wir leben in einer regulierten Welt!
 - Audits, Konformität und Regulierung: Teil der Geschäftssprache
 - Geschäftliche „Spielregeln“:
Produkthaftung, ISO 9000, Basel II,
„Richtlinien zur Sicherung guter wissenschaftlicher Praxis“, ...
- Wie erzeugen und verwalten Organisationen Informationen?
 - Akademischer „peer review“ für Wissenschaft und Forschung
 - Audit-Regeln für finanzielle Transaktionen
 - Flugsicherungs-Regulierungen
 - Bestimmungen zur Sicherheit von Patienten-Informationen
 - Verfahren zur Zulassung pharmazeutischer Produkte
- Für Konformität müssen Prozesse, und die daraus gewonnenen Informationen ...
 - Offen, transparent und auditierbar sein
 - Eine geprüfte Integrität haben
- Eine Organisation will Konformität belegen.



Wie kann sie den Ursprung und die Authentizität der erhaltenen Informationen durch den Geschäftsprozess belegen?



Provenance: Was wir darunter verstehen ...

- Prozessinformationen müssen aufgezeichnet werden, während sie passieren
- Mit dieser Dokumentation kann folgendes ermittelt werden:
 - Der Ursprung der Daten
 - Die Konformität des (Daten-) erzeugenden Prozesses
- Das bezeichnen wir als „Provenienz“:
 - Die Provenienz einer Information ist die Geschichte ihrer Erzeugung.



Konformitätslösungen und deren Probleme

- Das nächste Problem der Konformität:
 - Wie stellen wir sicher, dass bei Einhaltung einer neuen Regelung nicht eine ältere gebrochen wird?
- Augenblickliche Konformitäts-Lösungen ...
 - ... sind proprietär, monolithisch
 - ... sind wie Silos, geschlossen
 - ... arbeiten nicht mit anderen Lösungen zusammen
 - ... sind nicht anpassbar an geänderte Bestimmungen





Projekt Provenance

Übergeordnete Ziele

Definition von Provenienz im Kontext der Endanwender-Applikationen:

Die Provenienz einer Information ist die Geschichte ihrer Erzeugung.

- Ziel: Eine computergestützte *Repräsentierung der Provenienz*, die uns erlaubt
 - Sinnvolle Analysen durchzuführen
 - Unsere Anwendungen zu belegen
- Die Anwendung von Provenienzen auf komplexe Probleme des Bereichs Grid-Computing und verteiltes Rechnen im Allgemeinen
- Standardisierung
 - Global Grid Forum
 - OASIS





Konkrete Projektziele und Umsetzung

1. Definition der Kernkonzepte für Provenienz
2. Spezifikation der geforderten Funktionalität:
 - Applikationen sollen *provenienz-bewußt* werden
3. Definition offener Datenmodelle und Protokolle
 - Provenienz-Operationen ermöglichen
4. Standardisierung dieser Datenmodelle und Protokolle
5. Bereitstellung einer Referenz-Implementierung



Wie funktioniert Provenienz?

Ein Beispiel

- Einfaches Prozess-Beispiel:
 - **Kuchen backen**
- Ziel: Aufzeichnung und Abfrage testen
- Basiert auf existierendem Code der Projektpartner

- **Victoria Sponge Cake** [<http://www.thefoody.com>]

- Zutaten:

- 110 g Butter
- 110 g Zucker
- 110 g Mehl
- 1 Tütchen Backpulver
- 2 Eier
- Vanille-Extrakt oder ein Teelöffel Zitronenschale

- Zubereitung:

- Ofen auf 190°C vorheizen (Gas Stufe 5)
- Die Butter und den Zucker schlagen, bis sich eine leichte, cremige Masse ergibt.
- Die geschlagenen Eier langsam mit etwas Mehl hinzugeben.
- Gesiebtes Mehl mit Backpulver und den anderen Zutaten unterheben.
- Teig gleichmäßig auf zwei flache Backformen (15 cm Durchmesser) verteilen.
- Für 20 – 25 Minuten backen.
- Zum Abkühlen auf einem Drahtgitter auslegen.

Wikipedia zum Sponge Cake:

Although simple to make, Victoria sponges recipes are **notoriously sensitive to cooking times and temperatures**. Because of this oven manufacturers often use a Victoria sponge recipe to test their ovens.



Backe, backe Kuchen ... „Mathematische“ Version

`mischung1 = Verquirlen(butter, zucker)`

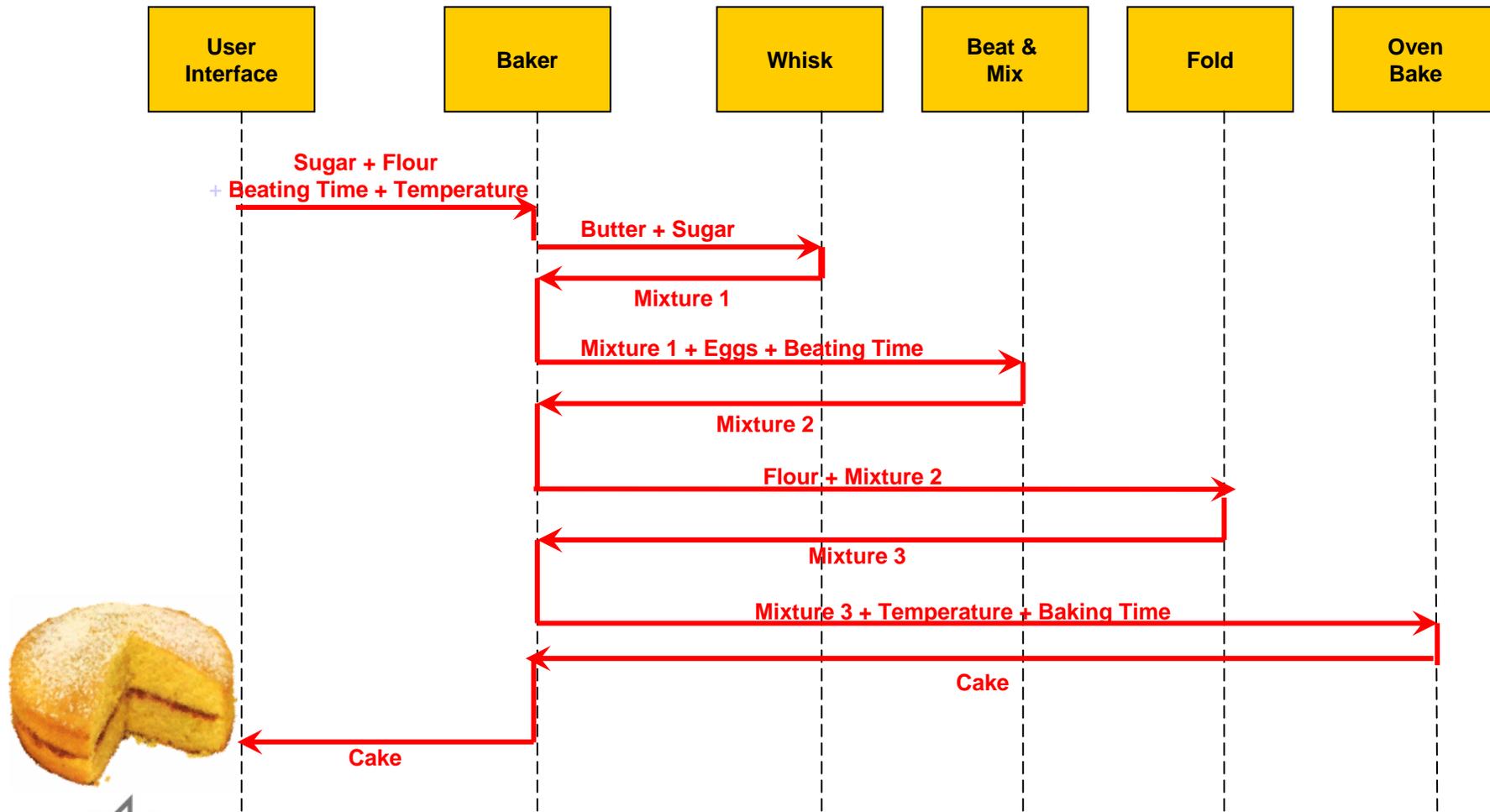
`mischung2 = Verrühren(mischung1, eier, rührzeit)`

`mischung3 = Unterheben(mischung2, mehl)`

`kuchen = OfenBacken(mischung3, temperatur, backzeit)`



Backe, backe Kuchen ... Informatiker-Version



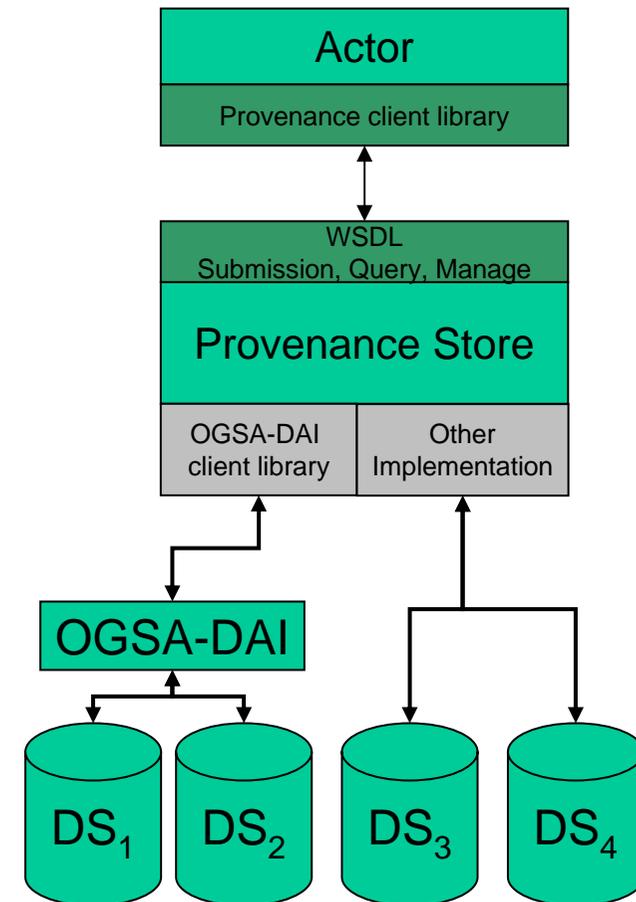


Kuchenanalyse mittels Provenance

- Nach dem Kuchenbacken können wir Fragen stellen!
 - Antworten können aus Aufzeichnungen während des Backprozesses gefunden werden – *der Provenienz*.
- Haben wir das Rezept akkurat verfolgt?
 - Haben wir die korrekten Zutaten verwendet?
 - Haben wir die korrekten Mengen verwendet?
 - Haben wir die Prozessschritte mit der richtigen Dauer ausgeführt?
- Prozess-Fragen
 - Haben wir für **kuchen1** und **kuchen2** die gleichen Mengen an Zutaten verwendet? Oder in gleichen Verhältnissen?
 - Was war der längste Schritt in der Ausführung des Rezeptes?
 - Warum haben wir den Prozess nicht beendet?
Wo haben wir aufgehört?

Provenance-Prototyp im Projekt

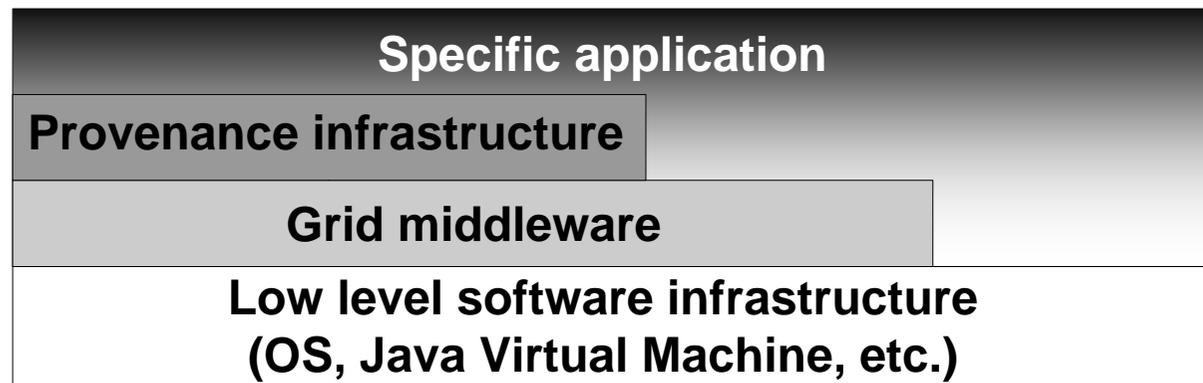
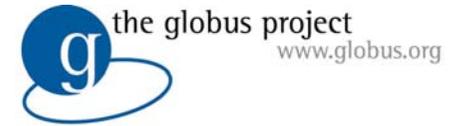
- Web-Services basierte Implementierung
- Das WSDL-Interface definiert einige Standard-Operationen für
 - Aufzeichnung
 - Abfrage
 - Management
- Die Implementierung basiert auf OGSA-DAI mit der XML-Datenbank eXist
- Andere Implementierungen sind möglich





Provenance-Implementierung

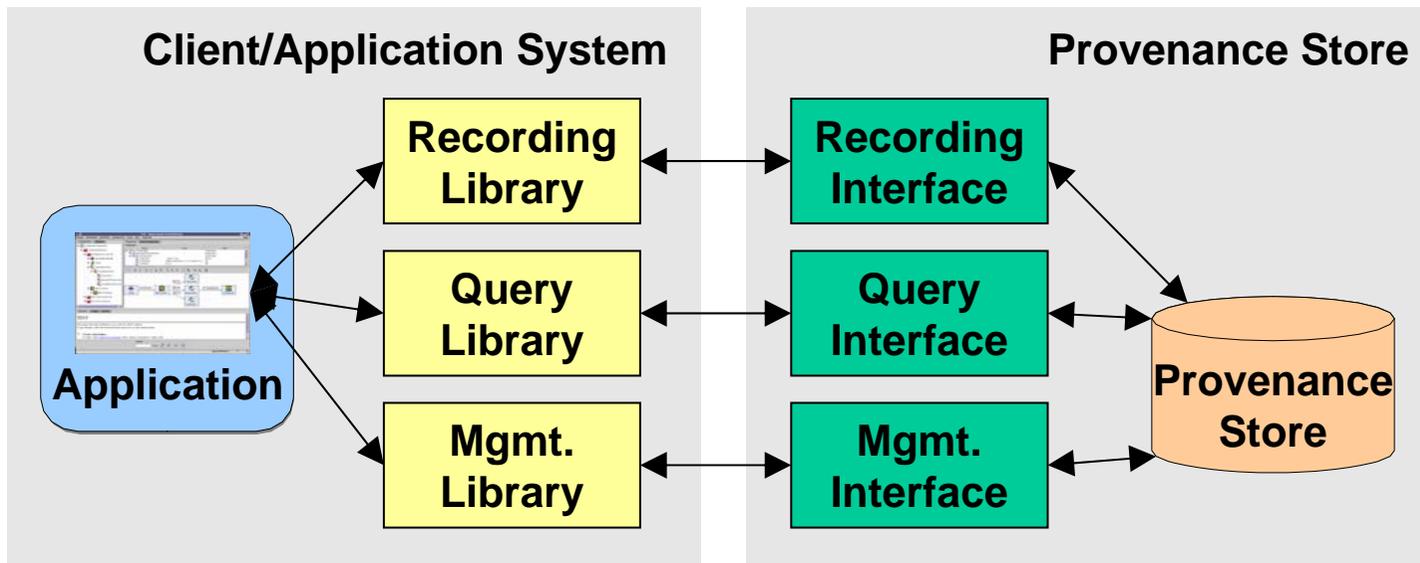
- Die Implementierung wird auf Globus GT4 umgestellt
 - Unterstützung für WS-Resource-Framework
 - Unterstützung für Sicherheits-Komponenten
 - Implementierung in 100% Java
- Verwendung offener und verabschiedeter Standards
- Software-Stack der Provenance-Architektur





Verwendung in Applikationen

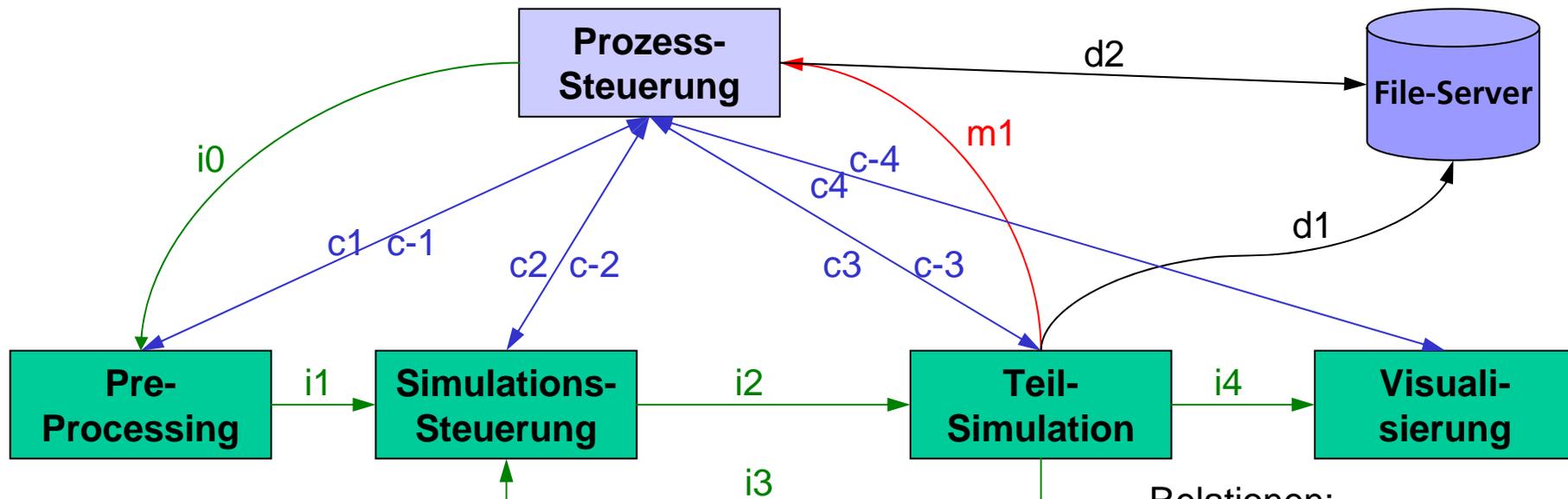
- Client-Interaktion mit dem Provenance-Store



- Verwendung existierender Bibliotheken
 - Geringer Anpassungsbedarf

Real-Beispiel für 2006

Simulations-Prozess



Interaktionen:

- Konfiguration
- Prozessablauf
- Monitoring
- Datenmanagement

Zustand:

der Akteure

Relationen:

- r0: i0 verursacht i1
- r1: i1 verursacht i2
- r2: i2 verursacht i3
- r3: i2 verursacht i4
- r4: i3 verursacht i2
- r5: i2 verursacht m1
- r6: i2 verursacht d1
- r7: i0 verursacht d2



Real-Beispiel für 2006

Prozess-Fragen

- Fragen zum Prozess:
 - In welcher Simulation wurde eine bestimmte Datei erzeugt?
 - In welchen Simulationen wurde ein bestimmtes Modell berechnet?
 - In welchen Simulationen wurde ein bestimmter Parameter verwendet?
 - Welche Monitoring-Informationen wurden in einer Simulation mit Parameter == x aufgezeichnet?
 - Welche Simulationen wurden mit einer bestimmten numerischen oder Modell-Konfiguration gerechnet?
 - Welche Ergebnisse und Unterschiede in der Provenienz ergaben sich bei zwei oder mehreren Simulationen mit demselben Input?



Beispielanwendungen im Projekt (1) Aerospace Engineering (DLR)

- Flugmanöver-Simulation
- Aus Projekt *SikMa* (Simulation komplexer Manöver)
- Gekoppelte Simulation
 - Aerodynamik
 - Strukturmechanik
 - Flugmechanik

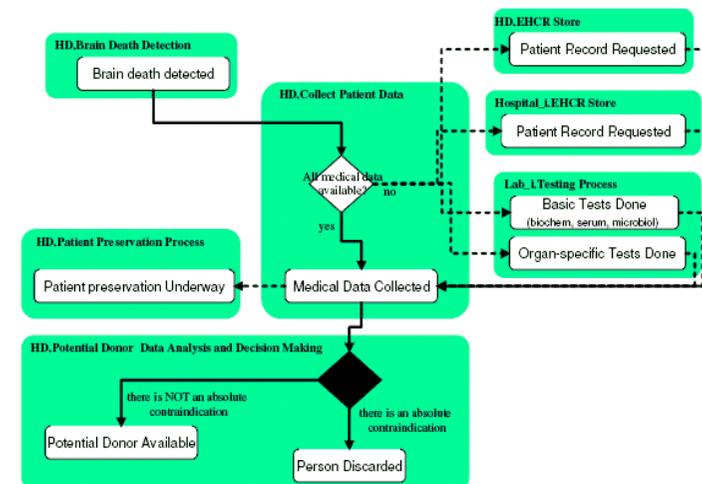
- Hohe Anforderung an Performance,
innerhalb gut geschütztem Netzwerk



Beispielanwendungen im Projekt (2)

Organ-Transplantations-Management

- Anwendung von
 - Universität Politecnica de Catalunya (Barcelona)
 - MTA SZTAKI (Budapest)
- Organ-Transplantations-Management
- Hohe Anforderung an Sicherheit, stark geographisch verteilte Anwendung



- Anwendungen mit komplementären Anforderungen



Vielen Dank für Ihre Aufmerksamkeit!

Weitere Informationen zum Projekt:

<http://www.GridProvenance.org>

oder

<http://www.dlr.de/sc/projekte/PROVENANCE>

Fragen und Anregungen?

