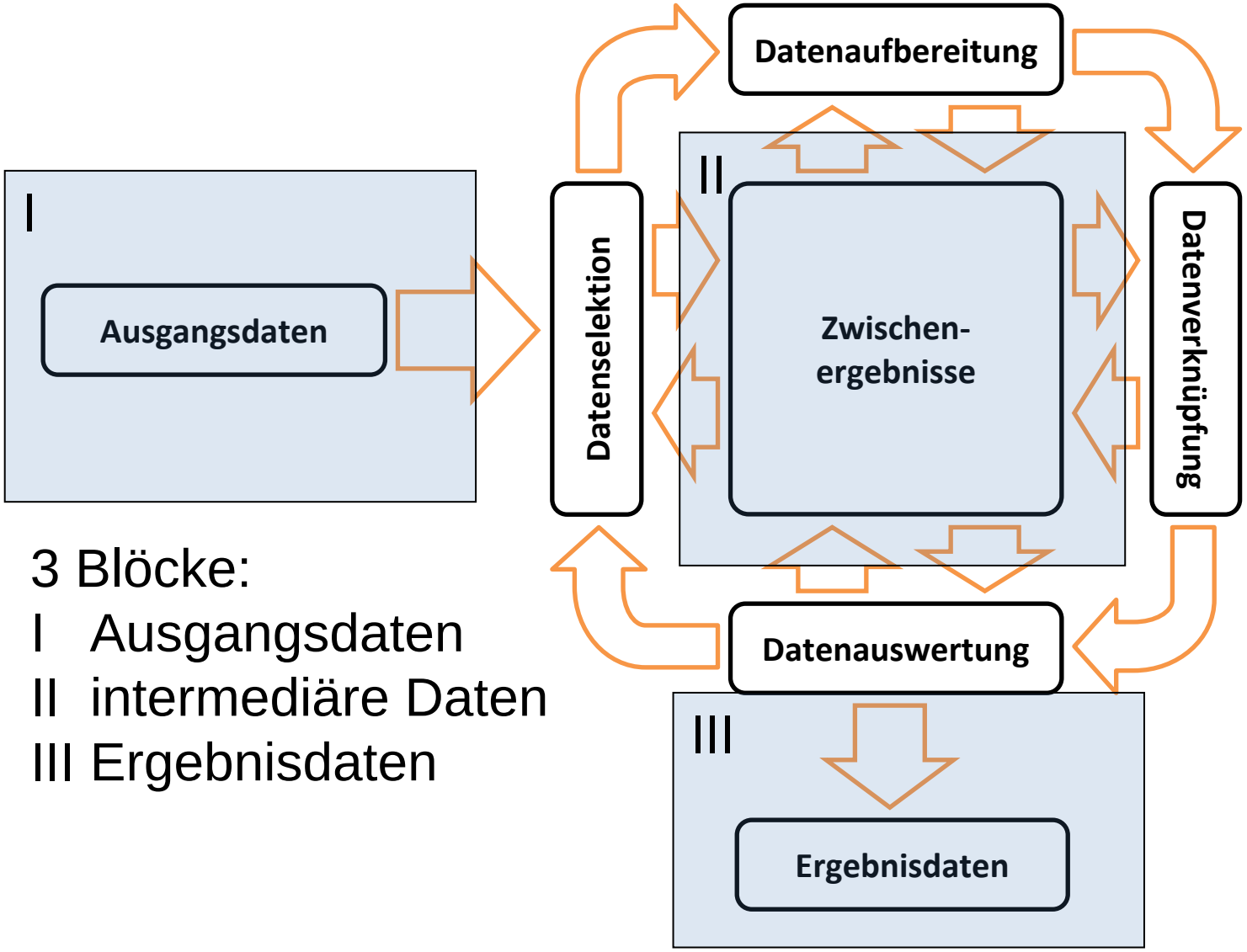
Workshop

»Virtuelle Arbeitsumgebung für Forschung und Berichterstattung mit Wirtschafts- und Sozialdaten«
Göttingen, 19. Juli 2010

Bundesministerium für Bildung und Forschung
WissGrid
D-GRID GmbH

Datenbereitstellung und –sicherheit



3 Blöcke:

I Ausgangsdaten

II intermediäre Daten

III Ergebnisdaten

Datenbereitstellung und –sicherheit: I

Mikrodaten aus

- Personen- und Haushaltssurveys
- Verwaltungsregistern
- Unternehmens-befragungen
- Erhebungen, etc.

Ausgangsdaten

Quellen:

- Forschungs-datenzentren
- Statistische Ämter
- Forschungs-einrichtungen
- Verwaltungen, etc.

Sicherheit der Daten in der VAU:

1. Nutzer-basierte Zugriffsrechte
   - Individuelle Rechte
   - Gruppenzugehörigkeiten
   - keine rollenbasierten Zugriffe
   - Zertifikats basierte Infrastruktur
2. Metadaten mit Informationen über Beschränkungen
   - zeitlich
   - personengebunden
3. Lokalität der Daten
   Service Level Agreements mit einzelnen Ressourcenprovidern für
   - Storage
   - Computing

Datenbereitstellung und –sicherheit: II

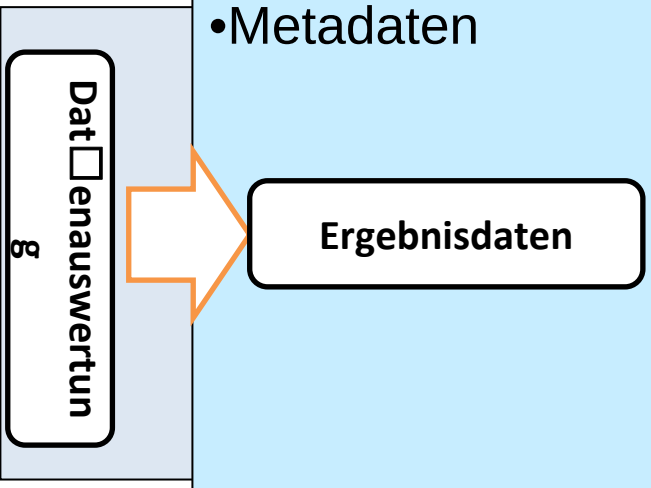
Typen:
- Arbeitsdatensätze
- Syntax-Dateien
- Grafiken,Dokumente
- Programm-spezi-fische Dateien

Zwischen-ergebnisse

Quellen:
- Workstation
- Remote Processing
- OnSite Processing
- Datenzentren
- VAU

Daten in der VAU:
- Vererbung der Beschränkungen
  Problem: wann sind Daten hinreichend unkenntlich, um Beschränkungen wegzunehmen?
- Syntax-Dateien:
  nur von individuellen Rechten bestimmt
- Grafiken, Dokumente, Programm-Output
  von individuellen Rechten bestimmt
  oder
  noch fortwirkende Beschränkungen?
- MetaDaten
- Lokalität der Daten
  Umgang mit Replikaten

Datenbereitstellung und –sicherheit: III



Typen:
- Ergebnisdatensätze
- Syntax-Dateien
- Grafiken,Dokumente
- Metadaten

Quellen:
- Workstations
- Remote-Processing
- Datenzentren
- VAU

Daten in der VAU:
- Vererbung der Beschränkungen nicht mehr vorhanden (Anonymisierung)
- Syntax-Dateien nur von individuellen Rechten bestimmt
- Grafiken, Dokumente, von individuellen Rechten bestimmt
- Metadaten erhalten Referenz und Information zu
  - Publikationen
  - Datensätzen, die in der VAU gelöscht sind (wg. Auflagen)
  - Struktur der gelöschten Datensätze
- Standardisierung von Formaten
  - Austausch-Dateiformate

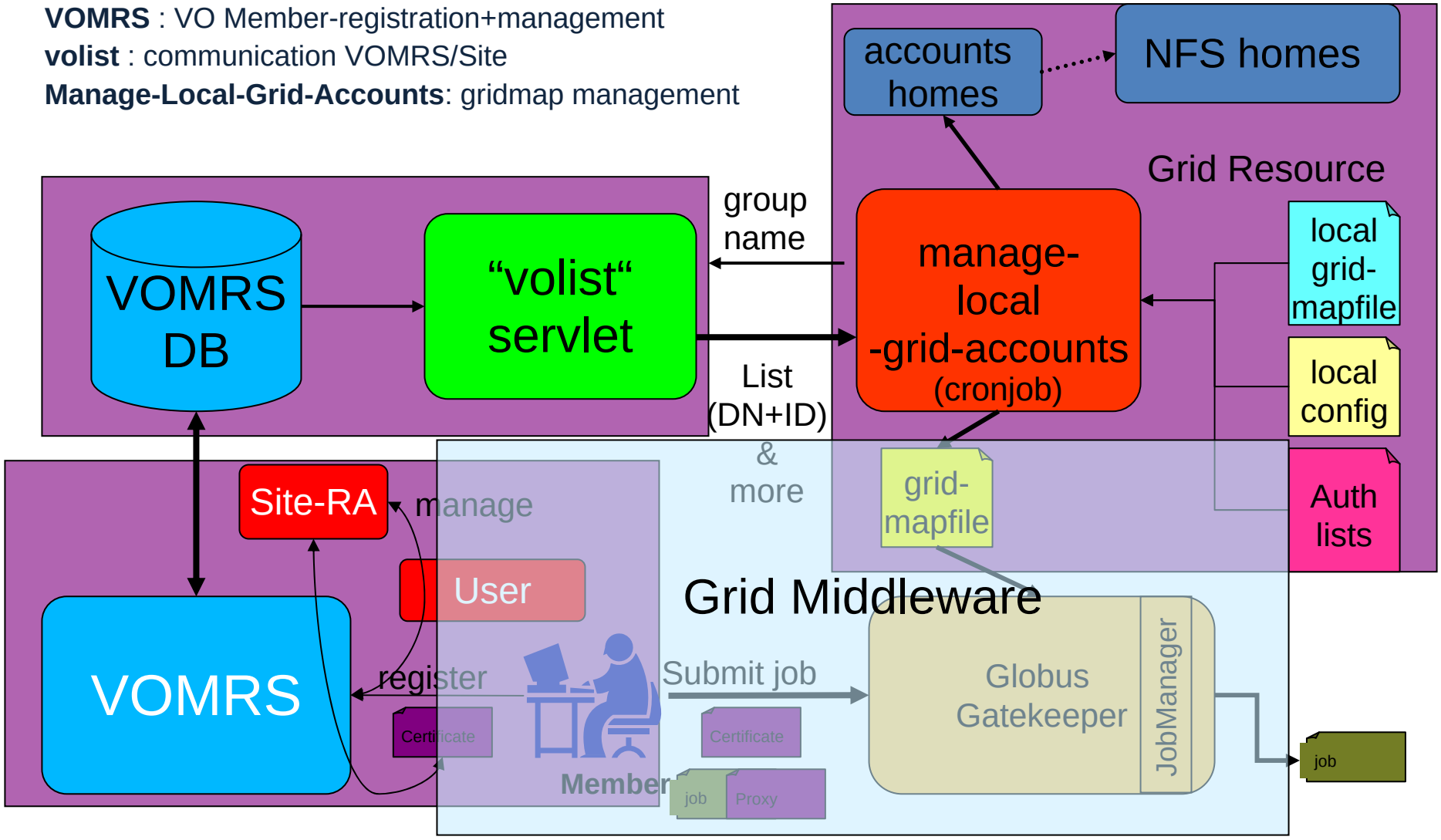
Datenbereitstellung und –sicherheit: DGrid

Grid Security Infrastruktur : (Middleware)

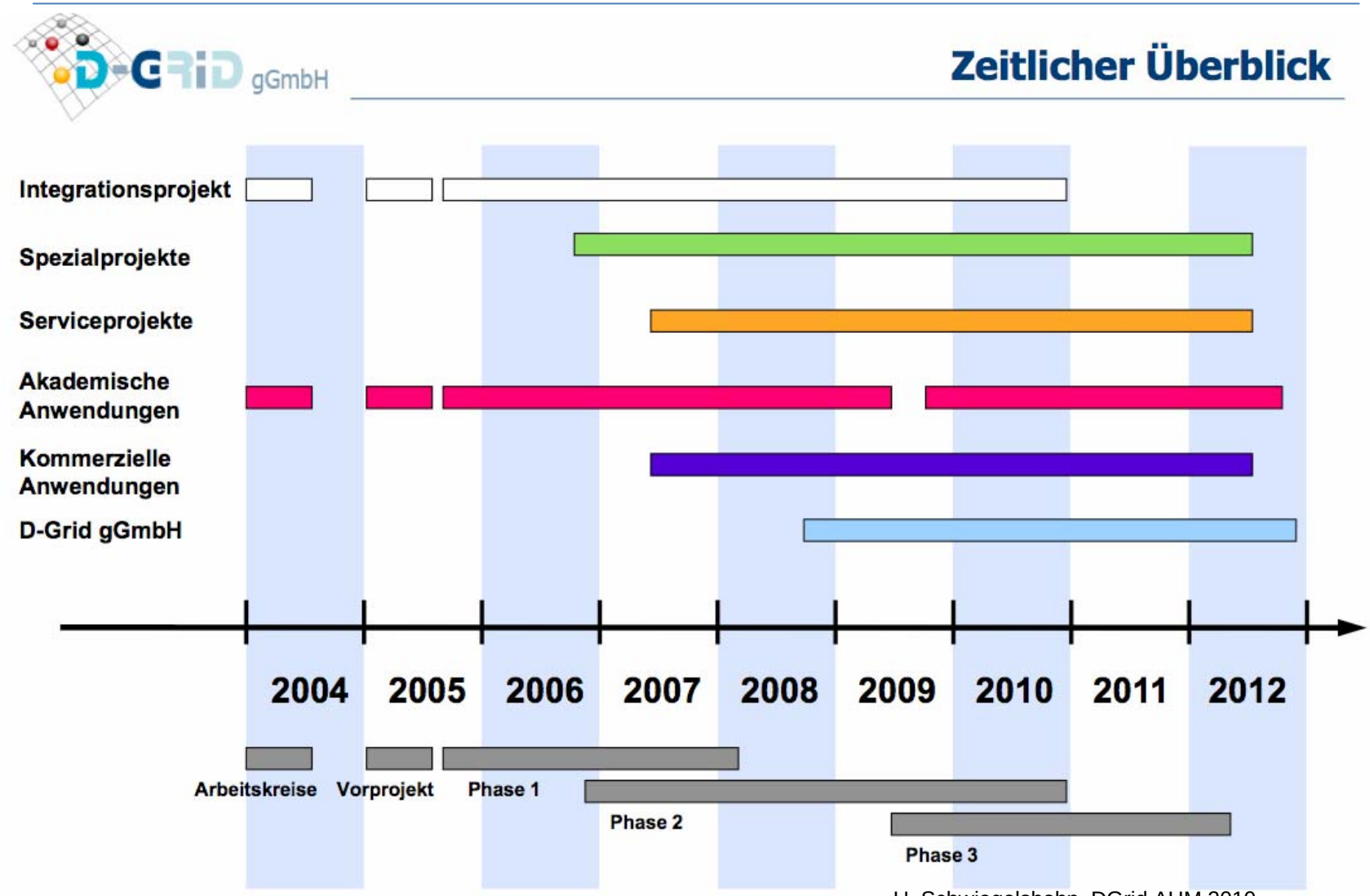
- Zertifikatsbasierte Authentifizierung
  - Durch zeitlich limitierte Proxyzertifikate Single-Sign-On
- Vorhandener „Chain of Trust“ durch GridKA und DFN mit (lokalen, universitären oder institutsweiten) Registration Authorities (RA) für Zertifikate
- Nutzer basierte Autorisierung auf (lokaler) Grid-Resource
- Gruppen- und Nutzerrechte bei lokaler Autorisierung
- Virtuelle Organisationen (VO) mit einer (hierarchischen) Struktur
- Einfaches Management der Mitgliedschaft in verschiedenen Substrukturen der VO über den Membership Registration Service (VOMRS)
- Verwaltung des VOMRS durch SOEB/Community

VO/Grid Account Management (Beispiel AstroGrid)

VOMRS : VO Member-registration+management
volist : communication VOMRS/Site
Manage-Local-Grid-Accounts: gridmap management



Einschub: Infrastruktur DGrid I



U. Schwiegelshohn, DGrid AHM 2010

D-Grid-Infrastruktur

- 3 Grid Middlewares:
  - Globus, gLite, Unicore
  - unterschiedliche Architekturen, einheitliche, PKI basierte Grid-Sicherheits-Infrastruktur (GSI)
- 2 Zertifizierungs-Organisationen (Certioficate Authorities,CA)
  - GridKA (KIT, ehem. FZK), DFN
  - Einheitliche Organisation des Virtuellen Organisations Managements
- Ressourcenanbieter (Rechen- und Speicherkapazitäten) aus
  - Instituten und Hochschulen
  - Hochschul- und Landesrechenzentren
  - HPC Einrichtungen

Datenbereitstellung und –sicherheit: Nutzung DGrid I

Nutzung der GSI und der D-Grid-Infrastruktur:

- Service Level Agreements mit DGrid-Providern für
  - die Bereitstellung von dedizierten Ressourcen
    - Datenspeicherplatz
    - Software für Datenzugriff / -bearbeitung
    - Rechenkapazität „on demand“, aber auf vereinbarten Ressourcen (Trusted Providers, SLA)
    - Professionelle Maintenance der Ressourcen
- Derzeit OpenSource R für Auswertungen auf Grid Ressourcen
- Langfristiger: Lizenzen für SPSS, Stata,.... auf Grid Ressourcen anwendbar

Datenbereitstellung und –sicherheit: Nutzung DGrid II

Nutzung der GSI und der D-Grid-Infrastruktur:

- Verwaltung der VO-Mitgliedschaft und der VO-Gruppen durch den VO-Management-Prozess
    - Community basierte Entscheidungsstrukturen
    - Schnelle und leichte Bildung von Arbeitsgruppen
    - Zertifikatsbasierte Authentifizierung
- Erfüllung von Sicherheitsanforderungen der FDZ, z.B. für die Online-Auslieferung von SUF, möglich
    - Langfristig ggf. weitere Vereinfachungen beim Fernrechnen?
    - Bessere Referenzierung von gelöschten SUF?
    - Storage von zu löschenden SUF bei FDZ
- Speicherung von und Zugriff auf Daten in einer Umgebung, die verschiedene Sicherheitskriterien zusammenführt und erfüllt
- Für den einzelnen Nutzer mit weniger Aufwand verbunden