

Berliner Bibliothekswissenschaftlichen Kolloquium  
Institut für Bibliotheks- und Informationswissenschaft (IBI)

# **eResearch und die Rolle der wissenschaftlichen Bibliotheken**

Heike Neuroth  
SUB Göttingen / MPDL Berlin  
7. Juli 2009

# ToC

- eResearch, Virtuelle Forschungsumgebung (VRE)
- eScience, Wissenschaftsprozess
- Beispiele aus den Fachdisziplinen
- Forschungsinfrastrukturen, ESFRI
- Handlungsfelder VRE
- Schlußgedanken

# eResearch

- e-Research extends e-Science and cyberinfrastructure to other disciplines, including the humanities and social sciences, and denotes the use of ***information technology*** to support ***existing and new forms of research***. The main features are:
  - collaborative
  - uses GRID technologies
  - data intensive

# Oxford eResearch Center

- The Oxford e-Research Centre (OeRC) works with research units across the whole of Oxford University ***to enable the use and development of innovative computational and information technology in multidisciplinary collaborations.***
  - OeRC enables the exploration of new research areas in the sciences, and increasingly in the humanities and social sciences

# AU gov – Dep. of Education ...

- The term ‘e-Research’ encapsulates research activities that use a ***spectrum of advanced ICT*** capabilities and embraces ***new research methodologies*** emerging from increasing access to:
  - broadband communications networks, research instruments and facilities, sensor networks and data repositories;
  - software and infrastructure services that enable secure connectivity and interoperability;
  - application tools that encompass discipline-specific tools and interaction tools.

# Virtuelle Forschungsumgebung

## Virtual Research Environment (VRE)

### JISC Definition

- "The purpose of a VRE is to help researchers in **all disciplines** manage the increasingly complex range of tasks involved in carrying out research.
- A VRE will provide a **framework of resources** to support the underlying processes of research on both small and large scales, particularly for those disciplines which are not well catered for by the current **infrastructure**."

[http://www.jisc.ac.uk/programme\\_vre.html](http://www.jisc.ac.uk/programme_vre.html)

# ToC

- eResearch, Virtuelle Forschungsumgebung (VRE)
- eScience, Wissenschaftsprozess
- Beispiele aus den Fachdisziplinen
- Forschungsinfrastrukturen, ESFRI
- Handlungsfelder VRE
- Schlußgedanken

# eScience

John Taylor (UK): eScience is about **global collaboration** in key areas of science and the next generation of **infrastructure** that will enable it.”

- Virtuelle Integration verschiedener Dienste, Services und Tools
- Zusammenbringen unterschiedlichster Ressourcen wie (Forschungs)Daten, Digitalisate, Inhalte ...
- Resource Sharing: Rechenleistung, Speicherplatz, Dienste/Services/Tools, Wissen/Inhalte ...
- Interoperabilität: auf allen Ebenen (strukturell, technisch, semantisch ...)
- Kollaboration und Kooperation: neue Formen, virtuell und interdisziplinär
- basierend auf persistenten und nachhaltigen Infrastrukturen

# Paradigmenwechsel Wissenschaft

- Vor Tausend Jahren - **Experimental Science**
  - Beobachtung/Beschreibung natürlicher Phänomene
- In den letzten Jahrhunderten - **Theoretical Science**
  - Newton's Gesetze, Maxwell's Gleichung ...
- In den letzten Jahrzehnten - **Computational Science**
  - Simulationen komplexer Phänomene
- Heute - **eScience or Data-centric Science**
  - Zusammenführen von Theorie, Experiment und Simulation
  - “data exploration”, “data mining”, (“text mining”)

# Rolle der wiss. Bibliotheken

- Entwicklungen im wissenschaftlichen Prozess
  - „**Bench-top Science**“: Einzelne Wissenschaftler
  - „**Big Science**“: Mehrere Wissenschaftler an einer Fakultät
  - „**Team Science**“: Mehrere Wissenschaftler aus verschiedenen Fakultäten, Organisationen arbeiten an einer bestimmten Fragestellung zusammen (z.B. Human Genome Project), verankert in einem Forschungszentrum
  - „**Networked Science**“: Wissenschaftler aus unterschiedlichen Gebieten arbeiten an einer bestimmten Fragestellung zusammen, *verteilt, kooperativ, kollaborativ, vernetzt*

In: Library Support for „Networked Science“ (D-Lib Magazine, September/October 2007)

# ToC

- eResearch, Virtuelle Forschungsumgebung (VRE)
- eScience, Wissenschaftsprozess
- **Beispiele aus den Fachdisziplinen**
- Forschungsinfrastrukturen, ESFRI
- Handlungsfelder VRE
- Schlußgedanken

# Beispiele aus den Fachdisziplinen

- Teilchenphysik
- Astronomie
- Sprachwissenschaft

# Teilchenphysik

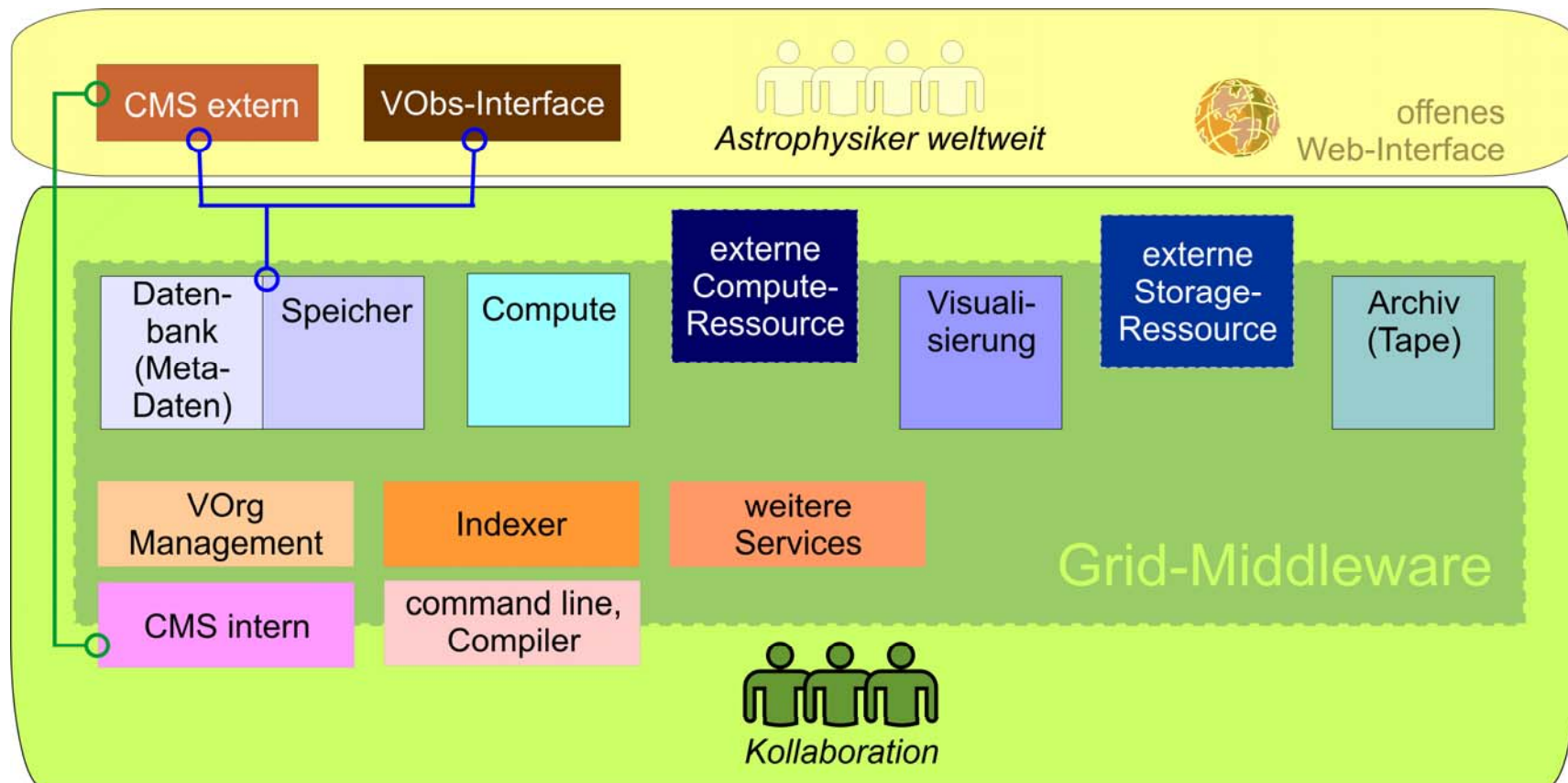
- Particle Physics – Large Hadron Collider
  - The largest machine in the world...
  - The fastest racetrack on the planet...
  - The biggest and most sophisticated detectors ever built...
  - The most powerful supercomputer system in the world...  
(<http://public.web.cern.ch/public/en/LHC/Facts-en.html>)
  - funded by and built in **collaboration with over 10,000 scientists** and engineers from over 100 countries as well as hundreds of universities and laboratories
  - The total cost of the project is expected to be €3.2 – 6.4 billion  
([http://en.wikipedia.org/wiki/Large\\_Hadron\\_Collider](http://en.wikipedia.org/wiki/Large_Hadron_Collider))

# Astronomie

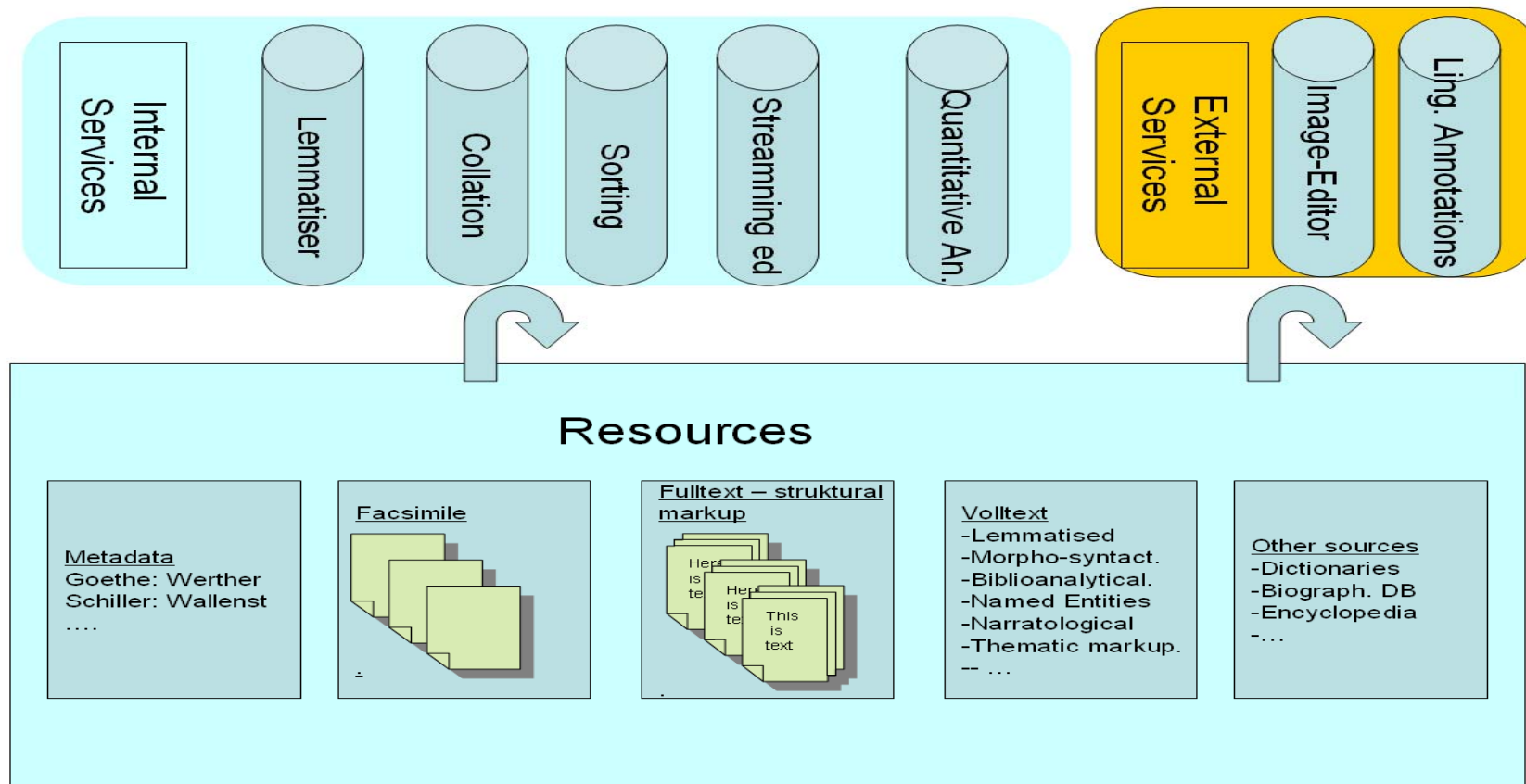
Traditionelle Arbeitsweise	Moderne Arbeitsweise
<ul style="list-style-type: none"><li>- Einzelbeobachtungen</li><li>- Individuelle Teleskopzeit</li><li>- ein Spektralbereich</li><li>- Theorie und Beobachtung</li><li>- Einzelauswertung</li></ul>	<ul style="list-style-type: none"><li>- Surveys, Satelliten, Teleskope</li><li>- Kollaborationen (Multinational)</li><li>- Daten aus vielen Spektralbereichen</li><li>- Theorie, Simulation, Beobachtung</li><li>- Daten-Pipelines</li></ul>
<p>Daten: ~ Gigabyte</p> <ul style="list-style-type: none"><li>- Photoplatten, Tabellen, Diagramme, Publikationen</li><li>- Einzelarbeitsplatz, Workstation</li><li>- Originaldaten nicht mit Publikationen verknüpft</li></ul>	<p>Daten: ~ Petabyte</p> <ul style="list-style-type: none"><li>- digitalisierte/digitale Daten</li><li>- Datenbanken</li><li>- Netzwerke von Datenservern, Rechnern</li><li>- Originaldaten und Publikationen verknüpft</li></ul>

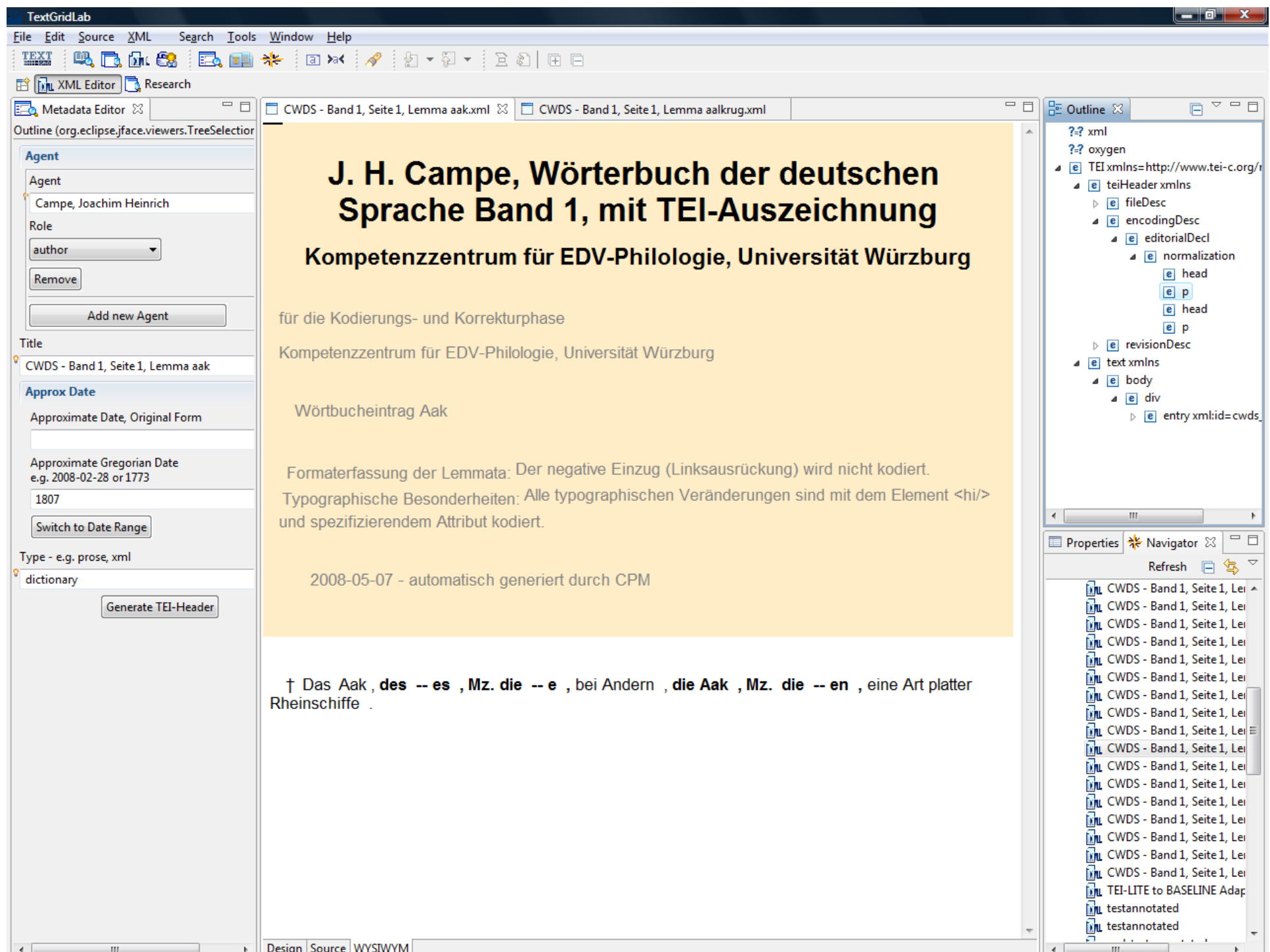
# Astronomie

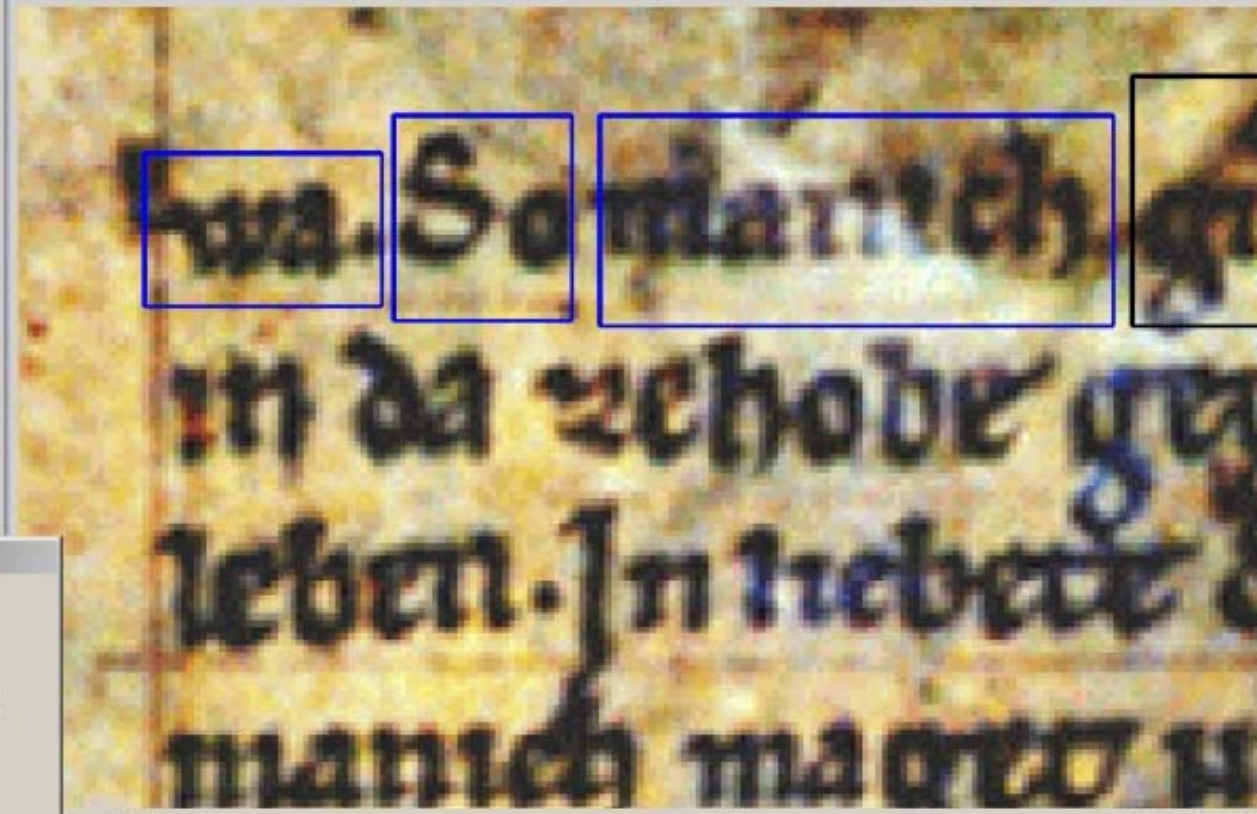
## Elemente der kollaborativen Forschungsumgebung



# Sprachwissenschaft







ar.select.description

a: 255, r: 181.

Texteingabe ✕

create Link

clear Text

Load Text

Save XML

Load XI

„So manich **juoc** ritter also da“ Oh wart  
la zehobe gegeben“ In allewis ein wns  
an“ In lebete den hob unde den lib, vil  
nich maget unde wib, Die scone ste van  
rich,“ Mich lameret warliche“ vnde  
sirz ith ich woldiz clagen“ daz nu bi  
en dagen, Solche uro wede niemer  
rien mah“ der man zwo den ziten nah“ (25

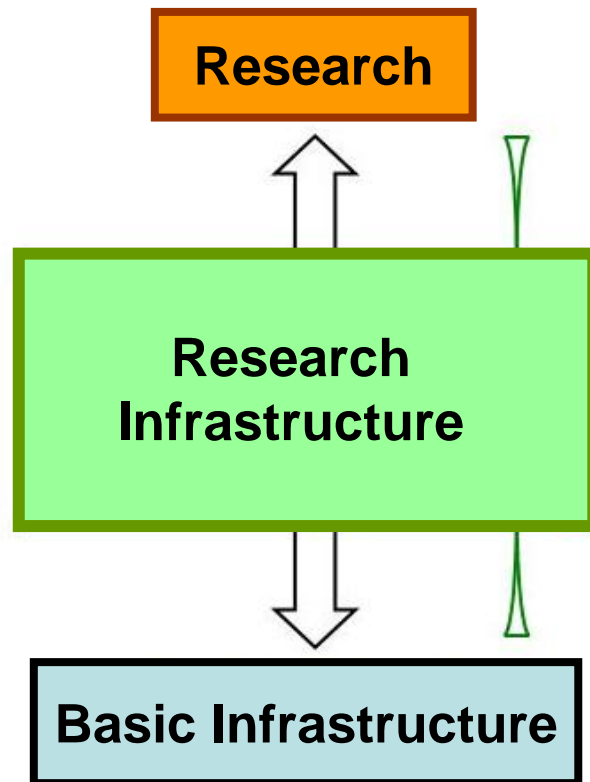
# ToC

- eResearch, Virtuelle Forschungsumgebung (VRE)
- eScience, Wissenschaftsprozess
- Beispiele aus den Fachdisziplinen
- **Forschungsinfrastrukturen, ESFRI**
- Handlungsfelder VRE
- Schlußgedanken

# Forschungsinfrastrukturen (RI)

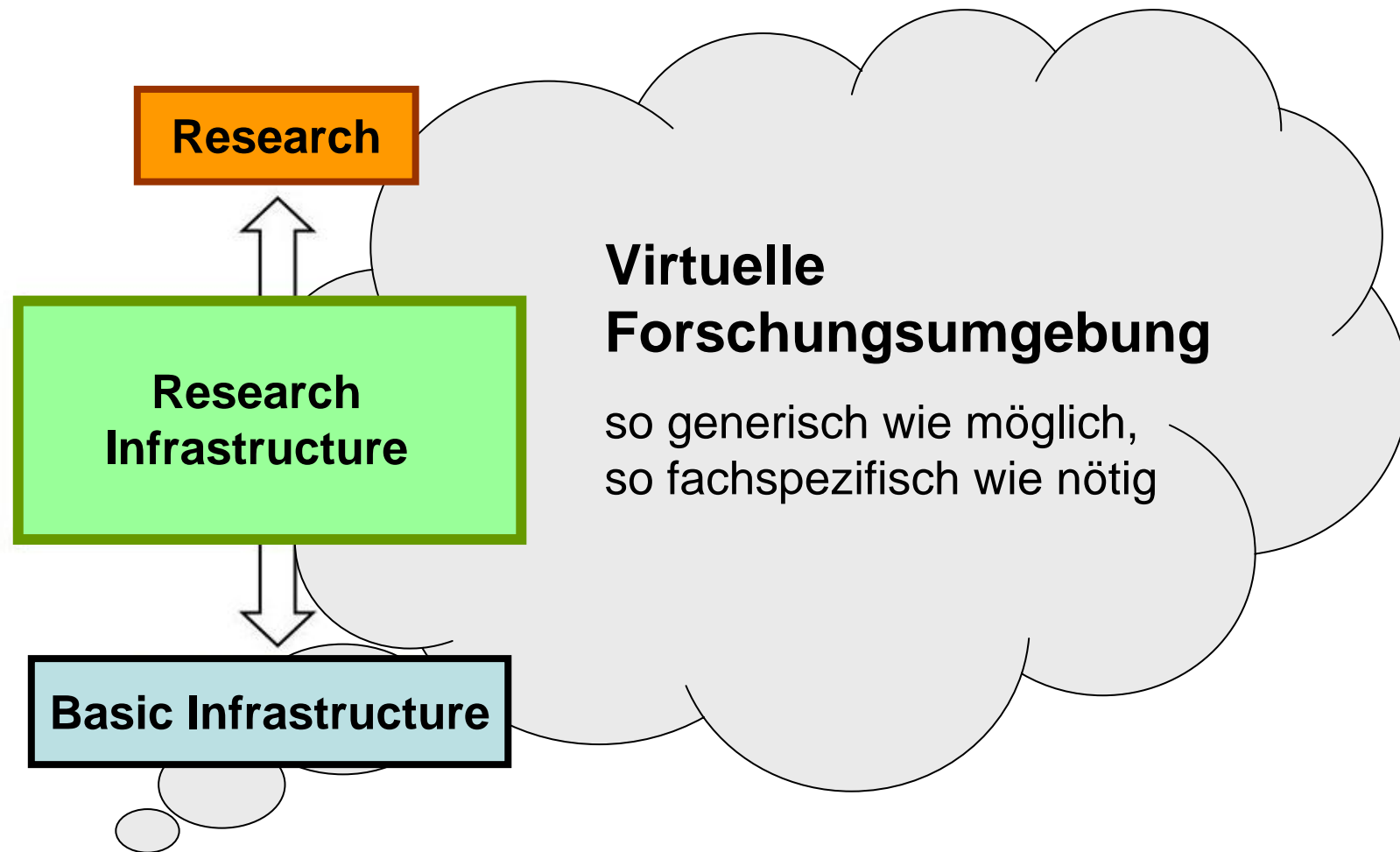
- Forschungsinfrastrukturen / Research Infrastructure (RI): Anlagen, Ressourcen und Dienstleistungen, die von Wissenschaftlern sämtlicher wissenschaftlich-technologischen Gebiete für die Forschung benötigt werden, z.B.
  - Arbeitsspeicher (z.B. Langzeitarchivierung), Rechenleistung (z.B. TextMining, Visualisierung)
  - Tools/Werkzeuge (z.B. Tokenizer, Lemmatizer ...)
  - Dienste/Services (z.B. Authentifizierung, Persistent Identifier ...)
  - Inhalte (z.B. Digitalisate), Forschungsdaten, Publikationen ...
  - Basis-Infrastruktur (z.B. Grid, Repositorien-Technologie)
  - ...

# Forschungsinfrastruktur (RI)



<b>R</b>	e.g. new research questions based on new methods like text mining, ...
<b>RI</b>	services, tools standards, training ...
<b>BI</b>	e.g. HPC/Grid infrastructure (compute, storage), repository infrastructure ...

# VRE



# ESFRI

- European Strategy Forum on Research Infrastructures
  - wurde im April 2002 initiiert und hat die Aufgabe, einen Beitrag zu einer europäischen Strategie der Forschungsinfrastrukturen zu leisten
  - ESFRI soll vor allen Dingen neue Forschungsinfrastrukturen von europäischem Interesse identifizieren, die für die nächsten 10-20 Jahre notwendig sind, um den Forschungsstandort Europa zu erhalten
  - ESFRI Roadmap
  - setzt sich aus Vertretern der Mitgliedsstaaten sowie einem Vertreter der Europäischen Kommission zusammen

# ESFRI Projekte

Clarín  
Dariah



<b>Social Sciences/Humanities (5):</b>	<b>ca. 200 Mio €</b>
<b>Environmental Sciences (11):</b>	<b>über 2 Mrd €</b>
<b>Energy (4):</b>	<b>über 2 Mrd €</b>
<b>Biological / Medical Sciences (10):</b>	<b>über 2 Mrd €</b>
<b>Materials / Analytical Facilities (6):</b>	<b>über 4 Mrd €</b>
<b>Physical Sciences / Engineering (8):</b>	<b>fast 6 Mrd €</b>

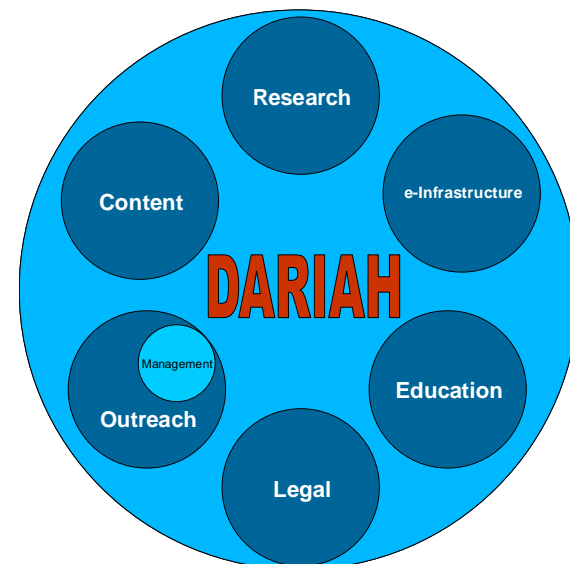
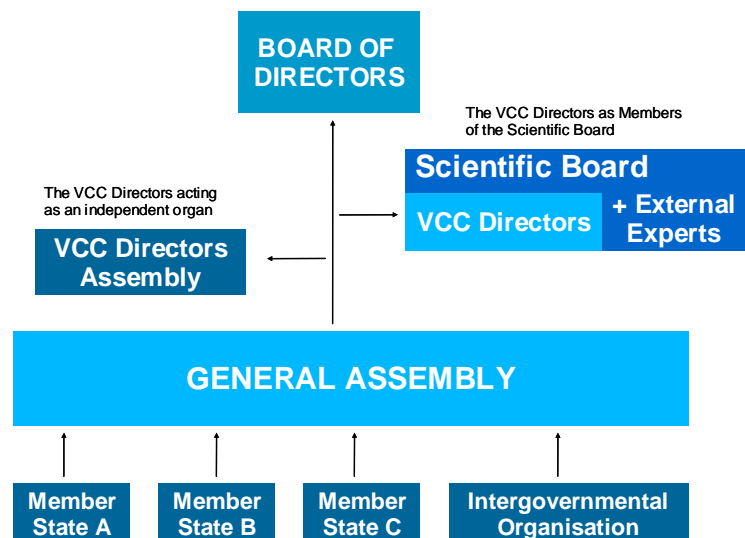
ESFRI Roadmap 2008

[ftp://ftp.cordis.europa.eu/pub/esfri/docs/esfri\\_roadmap\\_update\\_2008.pdf](ftp://ftp.cordis.europa.eu/pub/esfri/docs/esfri_roadmap_update_2008.pdf)

# Dariah

Digital Research Infrastructure for the Arts and Humanities

- Rechtsform
  - LEGAL FRAMEWORK FOR EUROPEAN RIs (ERIC)  
[http://ec.europa.eu/research/infrastructures/eric\\_en.html](http://ec.europa.eu/research/infrastructures/eric_en.html)
- Organisation über VCCs (Virtual Competence Center)
  - Research/e-Infrastructure/Education/Legal/Outreach/Content
- Kosten
  - Aufbauphase: ca. 4 Mio €/2 Jahre (davon 2.6 Mio € von EU), 2 Jahre Laufzeit
  - Umsetzung/Betrieb: ca. 6 Mio €/Jahr, 10 Jahre Laufzeit

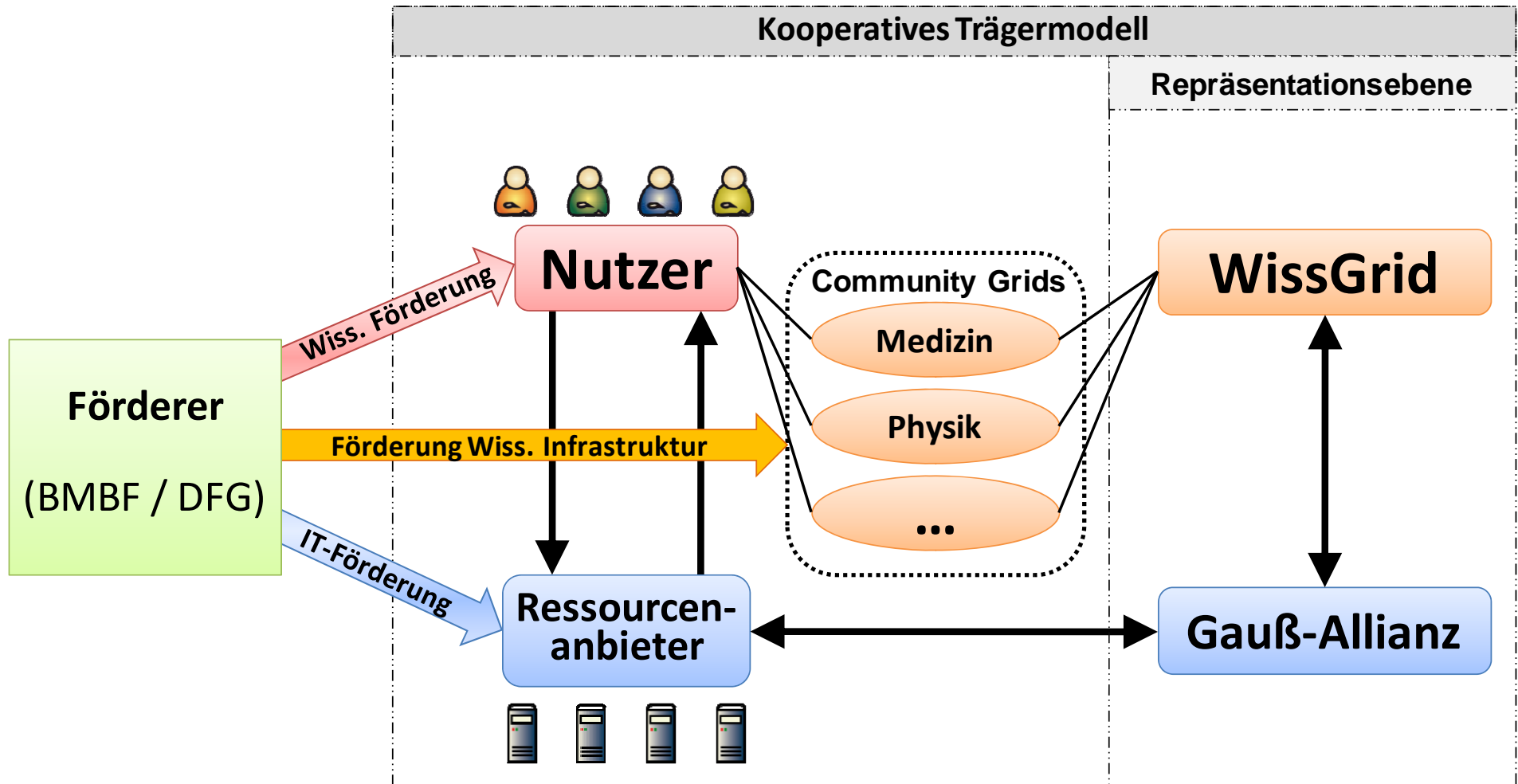


# TextGrid

Vernetzte Forschungsumgebung in den eHumanities

- Kosten:
  - 2. Förderphase BMBF (Juni 2009 bis Mai 2012)
  - über Grid/BMBF Sonderinvestitionen: 275 TB Speicher (inkl. Tape)
  - 1. Phase + 2. Phase inkl. Sonderinvestitionen: ca. 5 Mio / 6 Jahren
- Etablierung als Fach-Community mit Rechtsform
- Bereits im Verbund mit anderen Fach-Communities:
  - AstroGrid (Astrophysik, Astronomie)
  - C3Grid (Klimaforschung)
  - HepGrid (Teilchenphysik, Photonenphysik)
  - MediGrid (Medizin)
- **WissGrid – Grid für die Wissenschaft**
  - Verbund der akademischen Fach-Community Grids (CG)
  - Gemeinsame Interessensvertretung gegenüber den Ressourcen-Anbietern (Rechenzentren, HPC ..) = Gauß Allianz
  - Heranführen weiterer Fachdisziplinen (auch GW und SW) an das Grid

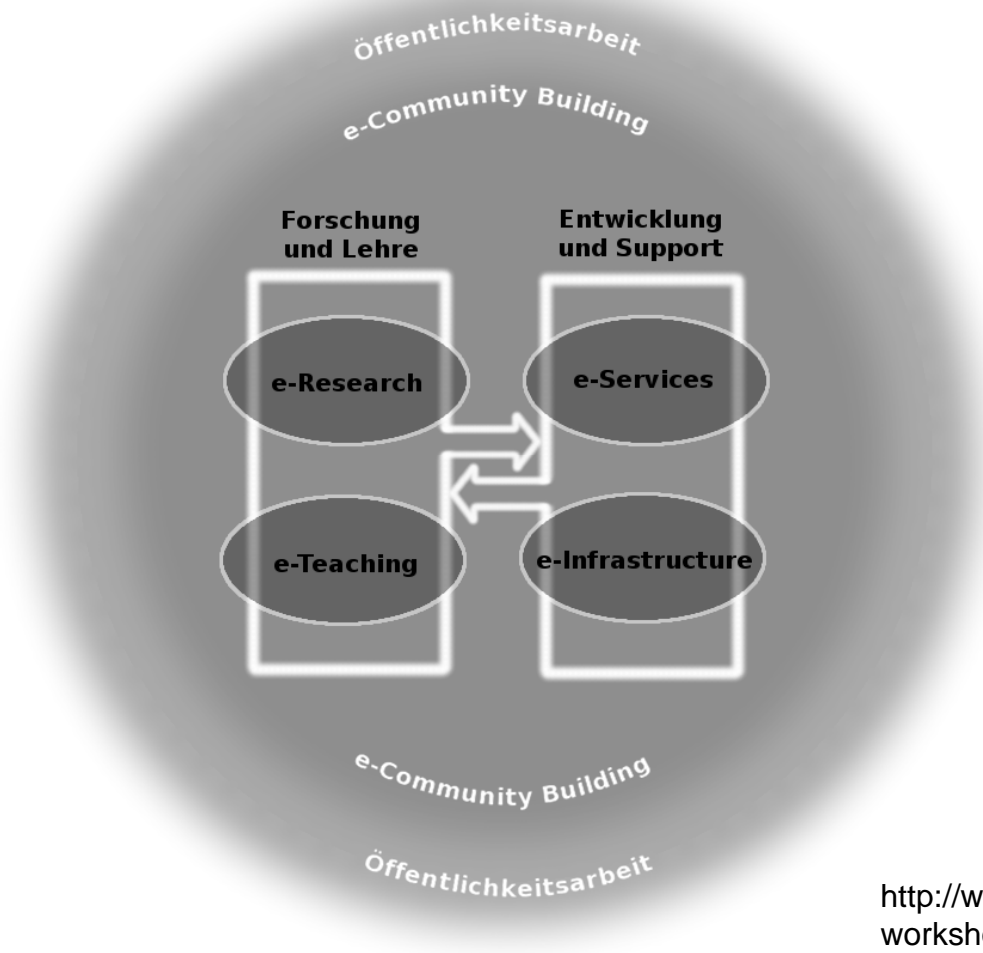
# Mögliches Betriebsmodell



# ToC

- eResearch, Virtuelle Forschungsumgebung (VRE)
- eScience, Wissenschaftsprozess
- Beispiele aus den Fachdisziplinen
- Forschungsinfrastrukturen, ESFRI
- Handlungsfelder VRE
- Schlußgedanken

# VRE Handlungsfelder



Ergebnisse des DFG  
Projektes „**Konzeptionelle  
Entwicklung einer  
Forschungsinfrastruktur  
für die e-Humanities in  
Deutschland**“ an der SUB  
Göttingen

Veröffentlichung in  
**Bibliothek Forschung und  
Praxis 2/2009**

<http://www.textgrid.de/konferenzen/e-humanities-abschluss-workshop-dfg/dfg-projekt-e-humanities.html>

# Handlungsfelder für wiss. Bibliotheken

- **e-Research:** Inhalte und Forschungsdaten
- **e-Services:** Dienste und Tools
- **e-Infrastructure:** Architekturen und Basis-Technologien
- **e-Teaching:** Lehre und Qualifizierung
- **e-Community Building:** Netzwerkbildung und Vernetzung

# e-Research

- Kollaboration und Kooperation durch Veröffentlichung, Verfügbarmachung sowie Nachnutzung und langfristige Speicherung von Publikationen UND Forschungsdaten
- Beispiel Forschungsdatenarchiv mit Publikation: **PANGAEA** (Publishing Network for Geoscientific & Environmental Data)

# e-Services

- Entwicklung und nachhaltige Bereitstellung von digitalen Diensten, Werkzeugen sowie die Implementierung und Weiterentwicklung einschlägiger Standards
- Beispiel: Dienstplattform wie ***Interedition*** (International initiative for digital scholarly editing infrastructure)

# e-Infrastructure

- (Weiter-)Entwicklung von Forschungsinfrastrukturen und Basis-Infrastrukturen sowie fachspezifische Anpassungen
- Beispiel RI: **ESFRI** (European Strategy Forum on Research Infrastructures) Projekte wie **PRACE** (A Petaflop Computing Research Infrastructure for European Scientists and Engineers) oder Dariah, Clarin ...

# e-Teaching

- Aus- und Weiterbildung des wissenschaftlichen Nachwuchts, z.B. kooperative Erstellung von nachnutzbaren e-Tutorials, Summer Schools etc.
- Beispiel: **GridKa** Schools, **nestor** Schools, EU Masterstudiengang in Digital Curation

# e-Community Building

- Vernetzung und Netzwerkbildung in (fach)wissenschaftlichen Communities, z.B. Aufbau eines ResearchVZ (interaktive Wer-Wo-Was-Datenbank) zur Vernetzung von Wissenschaftlern, Forschungsfragen, Projekten ...
- Beispiel: **CenterNet** (An international Network of Digital Humanities Centers) mit z.B.
  - exchanging information about tools development, best practices, organizational strategies, standards efforts, and new digital collections
  - developing collaborative teams

# ToC

- eResearch, Virtuelle Forschungsumgebung (VRE)
- eScience, Wissenschaftsprozess
- Beispiele aus den Fachdisziplinen
- Forschungsinfrastrukturen, ESFRI
- Handlungsfelder VRE
- **Schlußgedanken**

# DFG Ausschreibungen I

- Ausschreibung „**Repositorien** – Ausbau und Entwicklung vernetzter Informationsdienstleistungen für die Wissenschaft“ (30.04.2009)
  - Aufbereitung der Metadaten für die übergreifende Suche nach fachlichen und formalen Kriterien;
  - Optimierung der Interoperabilität von Repositorien und ihrer Kompatibilität mit webbasierten Diensten;
  - Entwicklung von Harvesting-Routinen;
  - Aggregation von Daten zur Nutzungsanalyse, Zugriffsstatistik und Zitanalyse;
  - Aufbereitung der Publikationen für eine übergreifende Volltextindexierung;
  - Verknüpfung von Ergebnissen aus einer maschinellen Aufbereitung von Volltexten mit Systemen zur Sacherschließung;
  - Aufbau von Repositorien für multimediale Inhalte;
  - wechselseitige Verzahnung institutioneller und fachbasierter Repositorien;
  - Vernetzung von Repositorien über Ländergrenzen hinweg;
  - aggregierende Zusammenführung und Übermittlung von Publikationen an Systeme zur Langzeitarchivierung;
  - Gestaltung von Funktionen, die sich in Lehr- und Forschungsumgebungen transparent integrieren lassen;
  - ...

<http://de.wikipedia.org/wiki/E-Science>

# DFG Ausschreibungen II

- Ausschreibung „**Virtuelle Forschungsumgebungen.** Infrastruktur und Demonstrationsprojekte“ (31. Mai 2009)

## **Ziel:**

Die Förderung zielt auf die Verbreitung virtueller Forschungsumgebungen zur **Unterstützung von Forschungsverbünden** auf nationaler als auch internationaler Ebene ab. Dabei sollen die Entwicklung neuer oder die Zusammensetzung verschiedener bereits bewährter interoperabler Komponenten datenbezogener Infrastruktur (z.B. Wikis, Blogs, Projekt Management, Virtuelle Organisationen, Primärdatenrepositorien, Authentifizierungsstrukturen, Data-Tagging usw.) gefördert werden. Weiteres wesentliches Ziel der Förderung ist die **verbesserte Zusammenarbeit zwischen den Forschungsverbünden und Informationseinrichtungen, vor allem Bibliotheken.**

# DFG Ausschreibungen

**DFG Ausschreibung zu Forschungsdaten  
Ende 2009, Beginn 2010 geplant,  
basierend auf den Ergebnissen der  
Allianz AG „Forschungsdaten“**

# Allianz Schwerpunktinitiative „Digitale Information“

## Allianz-Partnerorganisationen:

- Alexander von Humboldt-Stiftung
- Deutscher Akademischer Austauschdienst
- Deutsche Forschungsgemeinschaft
- Fraunhofer-Gesellschaft
- Helmholtz-Gemeinschaft deutscher Forschungszentren
- Hochschulrektorenkonferenz
- Leibniz-Gemeinschaft
- Max-Planck-Gesellschaft
- Wissenschaftsrat

## Handlungsfelder

- Nationale Lizenzierungen
- Open Access
- Nationale Hosting-Strategie
- Forschungsprimärdaten
- Virtuelle Forschungsumgebungen
- Rechtliche Rahmenbedingungen

<http://www.allianz-initiative.de/de/>

# Ausblick

Noch vieles ungelöst, z.B.:

- FrAGMENTIERUNG in den Wissenschaften ist extrem und Forschung, Ressourcen sind sehr verteilt
- Große Bandbreite an unterschiedlichen Ressourcen-Typen, Standards, Formaten und Strukturen
- Semantische Beziehungen zwischen Ressourcen ist weitestgehend unerforscht und komplex
- Forschung wird aber mehr und mehr “data driven” sein
- nur in vereinzelten Fach-Disziplinen ist wirklich IT Expertise vorhanden
- Wissenschaft nimmt neue (Forschungs)Methoden auch aus anderen Disziplinen nur sehr verzögert wahr/an
- “resource sharing” ist nicht immer gewünscht und möglich
- ...

# Ausblick

- gemeinsame **Strategien**/Policies benötigt, auf Standortebene, aber auch übergreifend, z.B. für Publikationen und Forschungsdaten
- technischer, organisatorischer Rahmen für Repositorien muss entwickelt werden
- Basis-Infrastrukturen und Forschungsinfrastrukturen werden dringend benötigt, für alle Fachwissenschaften, mit Hilfe von Service-Infrastruktureinrichtungen (Bibliotheken, Rechenzentren):
  - Digitale Langzeitarchivierung, Nachnutzung von Forschungsergebnissen (persistente Infrastrukturen, PID ...)
  - Vertrauensvolles AAI (Authentifikation-Autorisierungs-Infrastruktur, z.B. Shiboleth, GridShib/IVOM)
  - Entwicklung fachspezifischer und generischer Tools, Dienste, Services (z.B. WissGrid Projekt)

# Diskussion

- Was ist die Polarstern (Eisbrecher) oder der LHC (Teilchenbeschleuniger) in den Geisteswissenschaften, d.h. notwendige Infrastruktur, um eResearch zu machen?
- Mögliche Antwort: Kompletten Print Bestand der BRD im Volltext digitalisiert vorliegen zu haben (als ersten Schritt!), Kosten: ca. 1.8 Mrd Euro
  - Zahl der Bücher (die zwischen 1500 und 1965 gedruckt worden sind) = ungefähr mit sechs Millionen
  - Annahme: jedes Buch hat durchschnittlich ca. 300 Seiten hat, d.h. Kosten von ca. 360 Mio Euro für Bilddigitalisierung und Kosten von ca. 1,8 Mrd Euro für Volltexterfassung
  - Grundlage: ca. 0,20 € für Bilddigitalisierung/Seite und ca. 1 € Texterfassung/Seite

**Vielen Dank!**

Fragen, Anmerkungen?

Heike Neuroth  
7. Juli 2009, IBI/HU