

Architektur der DARIAH-DE Forschungsinfrastruktur

1. Sitzung des Wissenschaftlichen Beirates von DARIAH-DE
14. September 2012, Göttingen

Rainer Stotzka,
Peter Gietz, Rene Smolarski,
u.v.a.

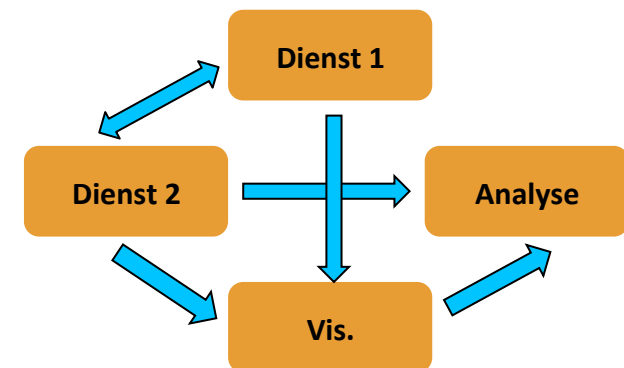
Dienste und Anforderungen

Dienst/Service:

- Dienstleistung
- Softwaredienst:
autarke Einheit zusammenhängender Funktionalitäten, verfügbar über eine klar definierte Schnittstelle
- ...

Anforderungen der Fachwissenschaften:

- Datenmanagement:
 - Heterogenität
 - Zitierbarkeit
 - Auffindbarkeit
 - einfacher Zugang
- Kommunikation
- Sicherheit und Kontrolle
- Erweiterbarkeit
- Zuverlässigkeit



Arbeitspakete

- AP 1 e-Infrastruktur
- AP 2 Forschung und Lehre
- AP 3 Forschungsdaten



Treffen DARIAH AP1 – CLARIN, 8.2.2012

- AP 1.1 AP 1 und VCC 1 Koordination (DARIAH-DE) und (DARIAH-EU)
- AP 1.2 Interoperabilität Meta-Daten (\leftrightarrow AP 3)
- AP 1.3 Interoperabilität Dienste (\leftrightarrow AP 2, AP 3)
- AP 1.4 Geisteswissenschaftliche Dienste (\leftrightarrow AP 2, AP 3)

Unterarbeitspakete: Bit Preservation, AAI und Hosting

DIGITALE FORSCHUNGSINFRASTRUKTUR FÜR DIE GEISTES- UND KULTURWISSENSCHAFTEN

DARIAH-DE-PORTAL

Beratung – Dokumentation – Standardisierung

LEHRE

- Summer Schools
- Expertenseminare
- Methodenworkshops
- Curricula

FORSCHUNG

Erkenntnisinteresse → Forschungsfrage → Recherche/Erschließung → Analyse → Interpretation → Publikation

Software Hosting Services

Demonstratoren und fachwissenschaftliche Dienste

- MEISE
- Archive in a Box
- Voyant Tools
- Monasterium
- Personendaten
- Relationen im Raum
- Interoperabilität Normdaten
- Differenzanalyse
- Virtuelles Skriptorium

Generische Dienste

- Entwicklung von Diensten auf Anforderung
- Geo-Browser
- Generische Suche
- Collection Registry
- Schema Registry

Forschungsdaten

Metadaten

Operative IT-Dienste

- AAI
- Sicherheit (operativ, strukturell)
- Qualitätssicherung

Software Hosting Services

Basisdienste

- Kollaborative Arbeitsumgebungen (Wiki, Developer Portal, Etherpad)
- PID Service
- Storage, Archive
- Datenbanken

Plattform Hosting Service

Bereitstellung VM mit Basissystem

Infrastructure Hosting Service

Bereitstellung virtueller Computer-Cluster

Operative IT-Dienste

- Ressourcen (Koordination, Bereitstellung)
- Helpdesk
- Monitoring

DARIAH AP1 Hosting Dienste

Software

Software
Hosting
Services

Software
Hosting
Services

Basissystem

Plattform Hosting Service

Bereitstellung VM mit Basissystem

Virt. Maschine

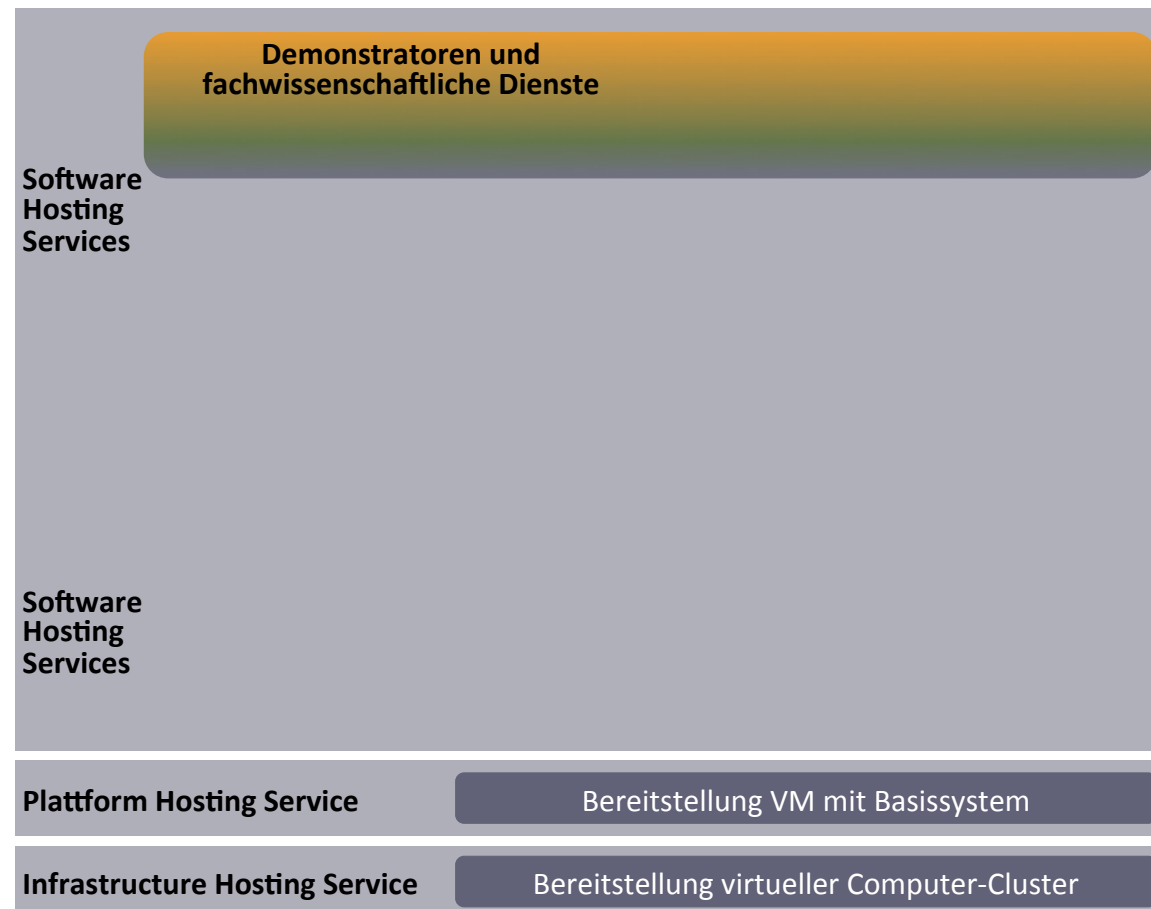
Infrastructure Hosting Service

Bereitstellung virtueller Computer-Cluster

Fachwissenschaftliche Dienste

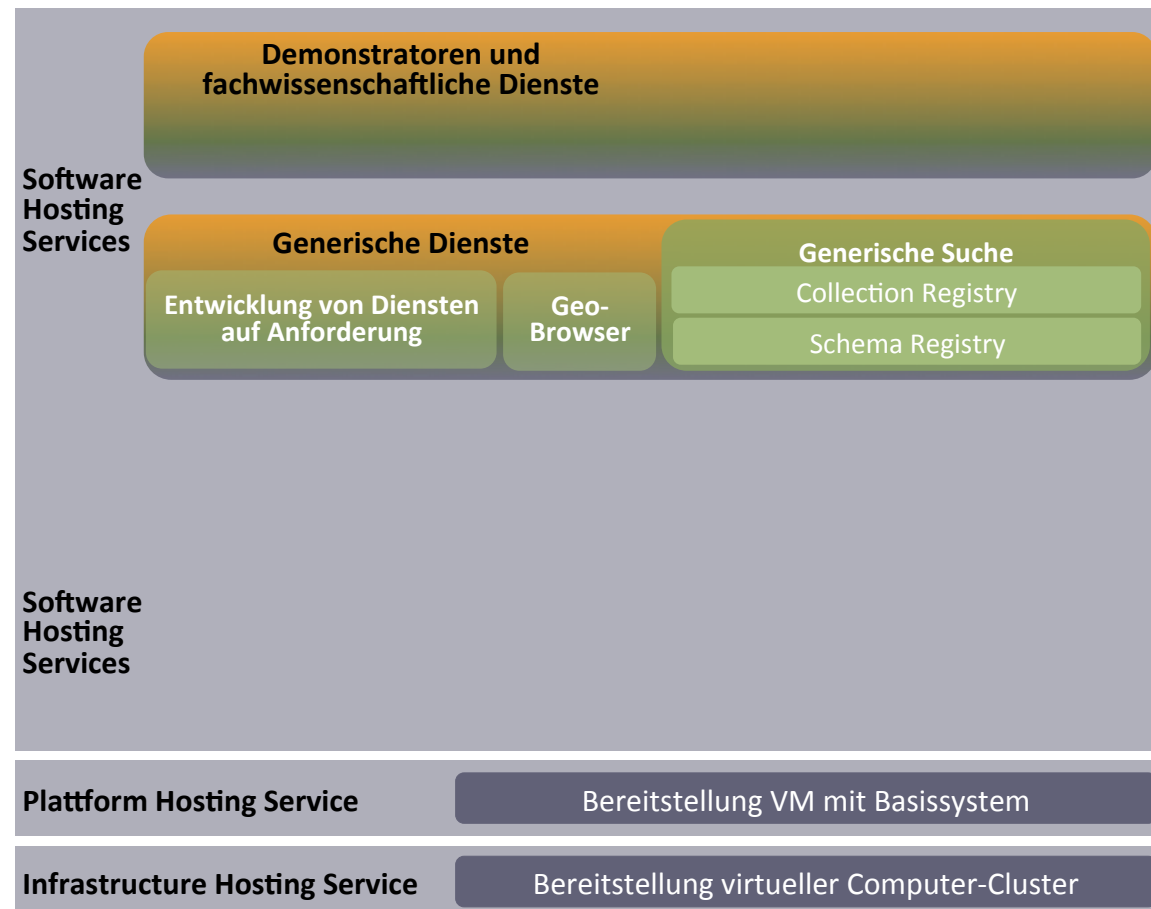
High-level
Dienste
von den
Fach-
wissenschaften
motiviert

AP 2 und
AG „Prozesse
und Dienste“



Generische Dienste

Von allen
Fach-
wissenschaften
benötigt

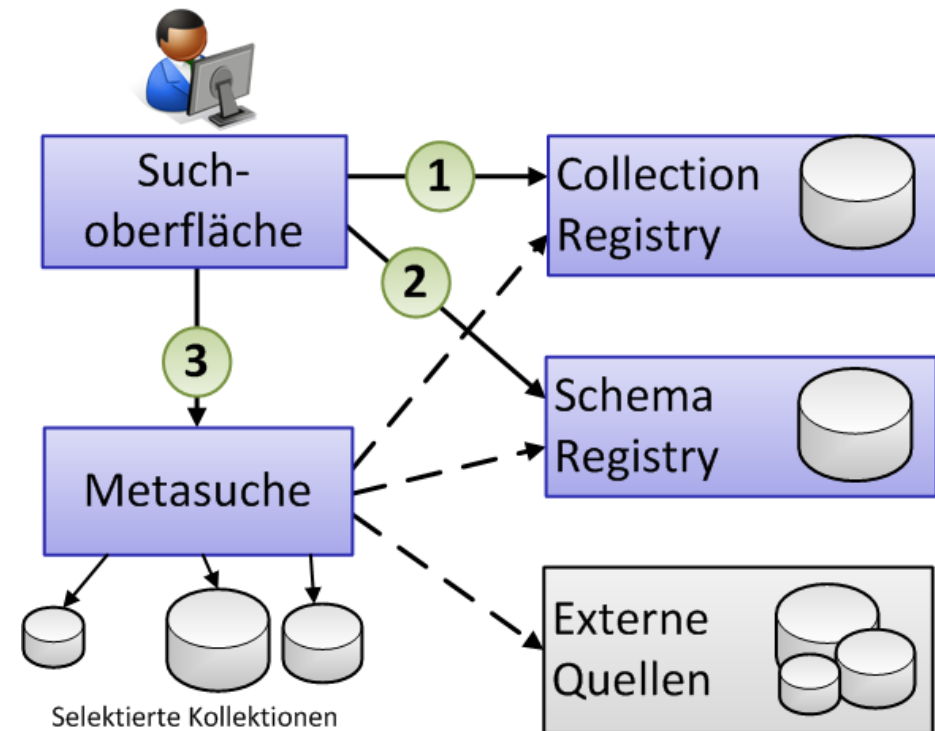


Generische Dienste: Suche

Collection Registry: Registrierung von Datensammlungen
Schema Registry: Beschreibung der internen Strukturen

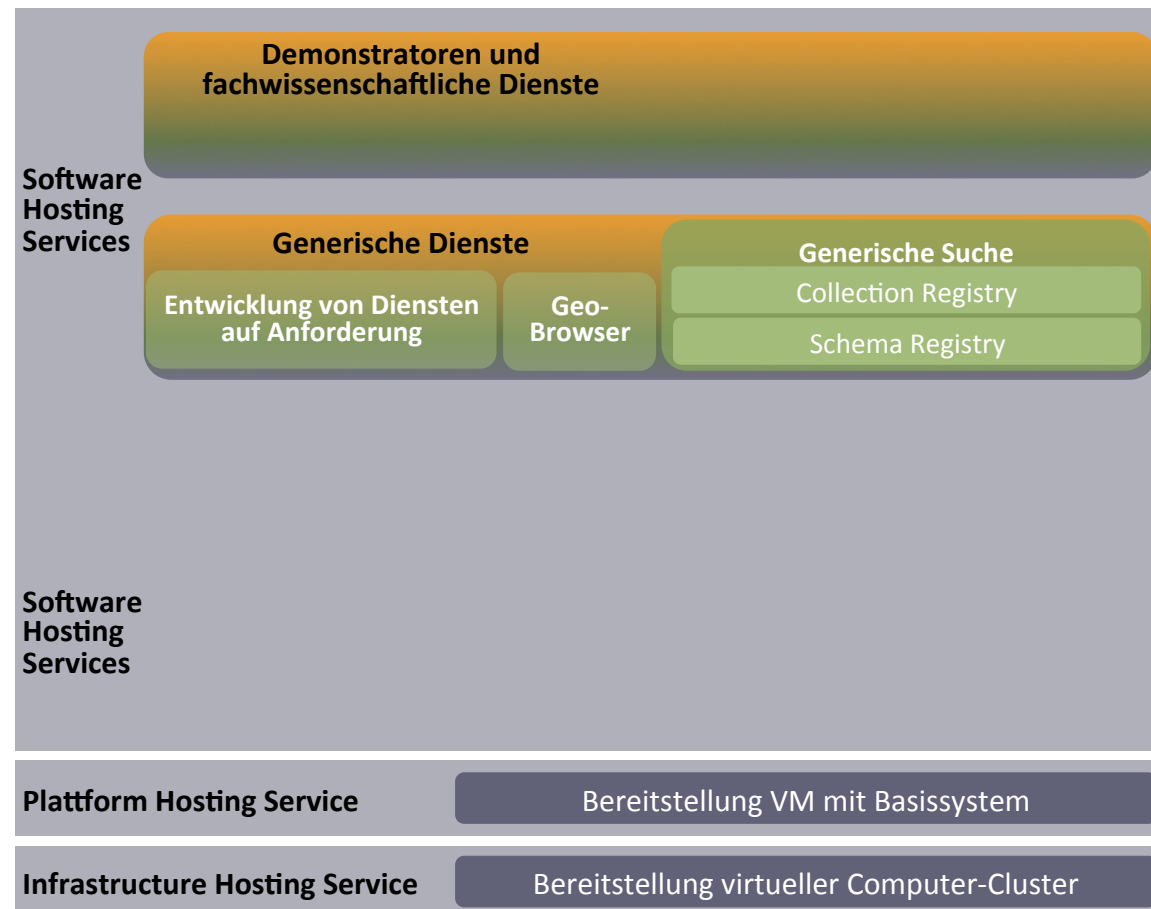
Zugriff auf die
Inhalte möglich!

1. **Auswahl der Sammlungen**, die für die Forschungsfrage des Benutzers von Bedeutung sind
2. **Crosswalks** zwischen selektierten Sammlungen und deren Schemata werden analysiert
3. **Suchfacetten und -filter**
Parametrisierung der Suche



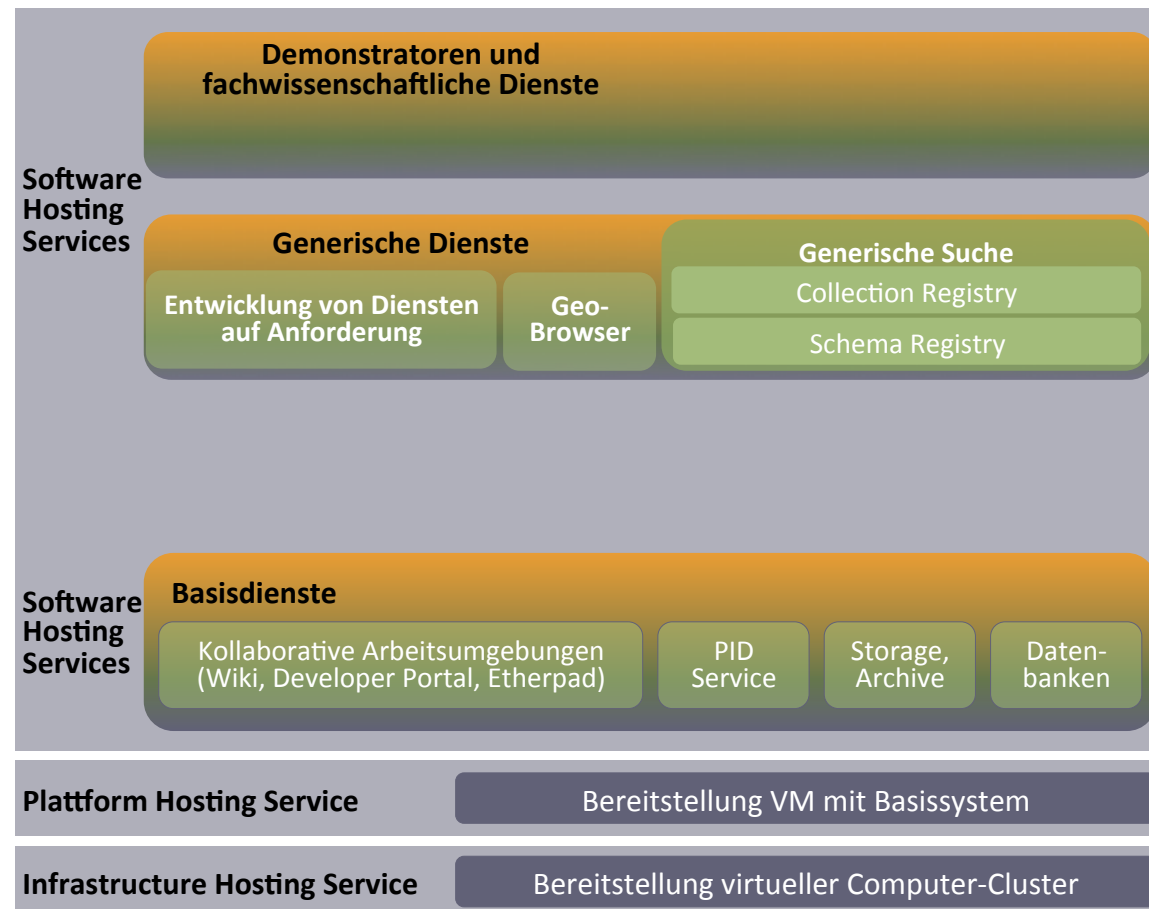
Generische Dienste

Von allen
Fach-
wissenschaften
benötigt



Basisdienste

Unabhängig
von den
Geistes-
wissenschaften



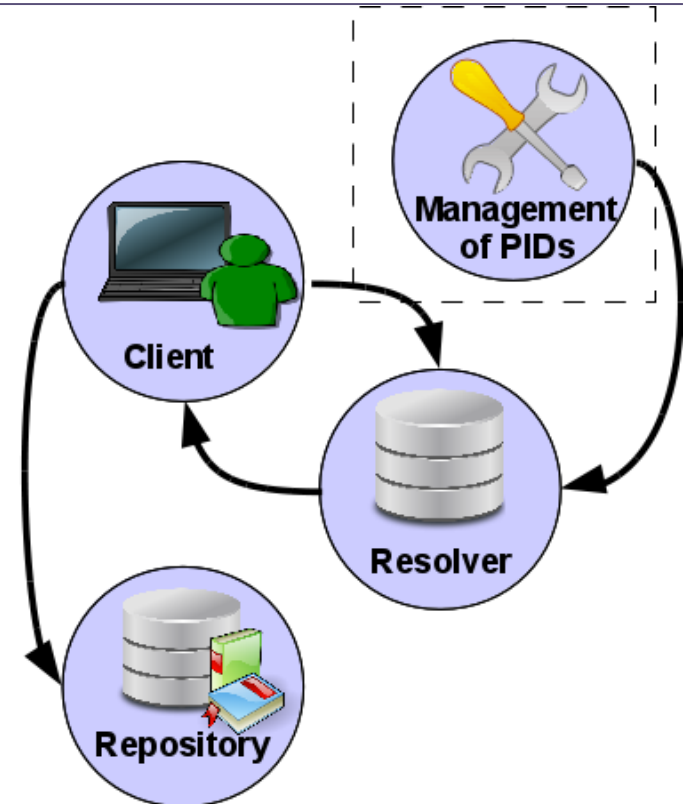
Persistent Identifier

Langfristige Referenzierung digitaler Objekte:

- Eindeutige und dauerhafte Erreichbarkeit
- Fester Bestandteil der Strategie zur Langzeitarchivierung und dem nachhaltigen Zugriff auf Forschungsdaten
- European Persistent Identifier Consortium (EPIC) PID-Service für die europäische wissenschaftliche Community

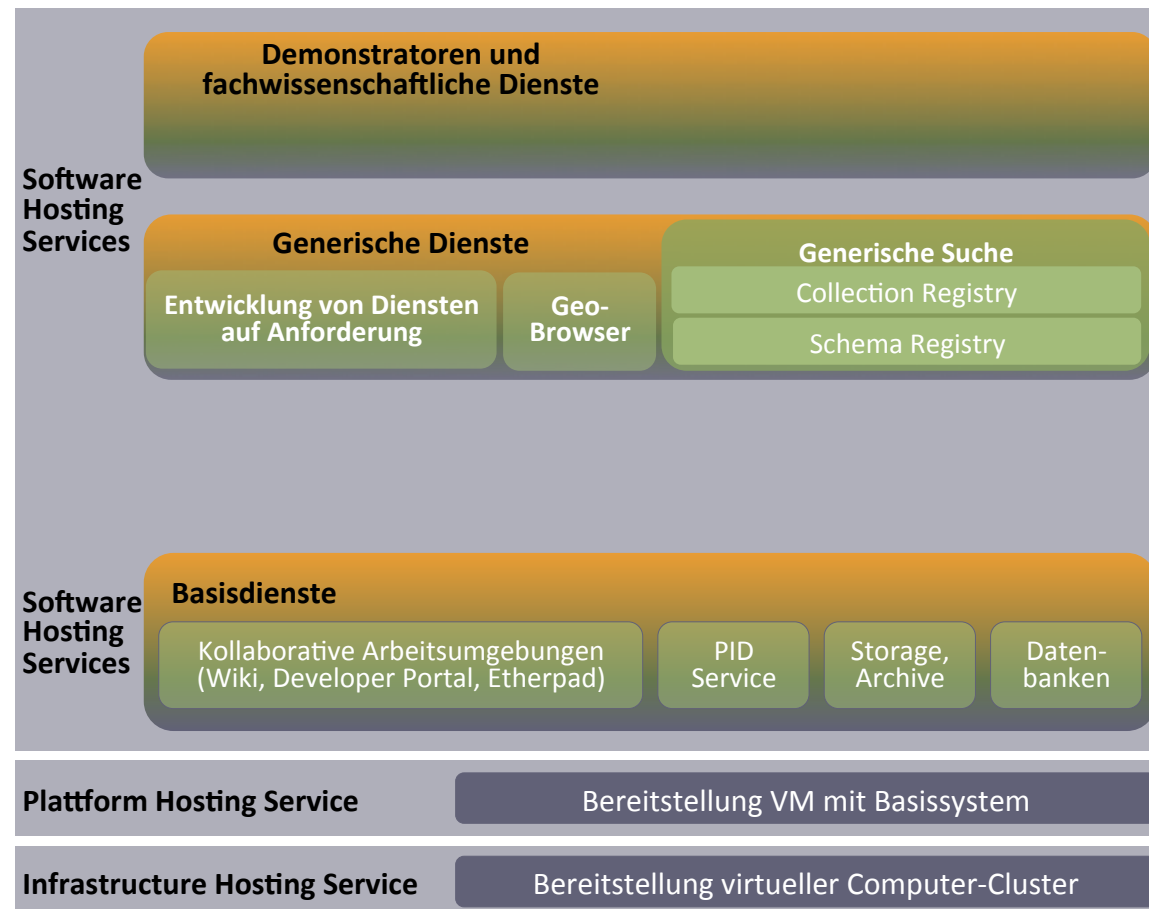
→ sehr starke Zusammenarbeit mit

- EPIC
- EUDAT

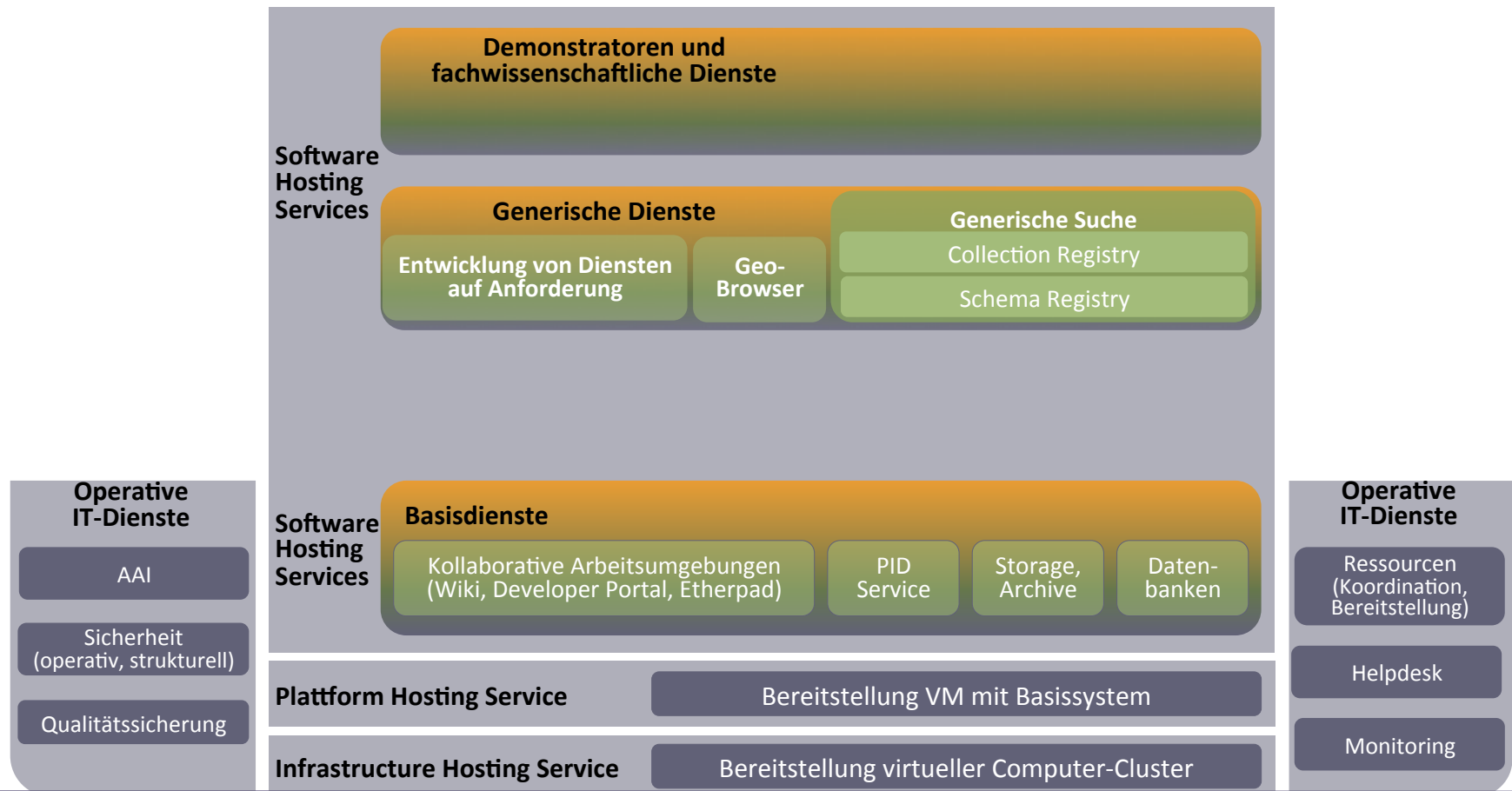


Basisdienste

Unabhängig
von den
Geistes-
wissenschaften



Operative IT-Dienste



Operative IT-Dienste: IDM und AAI

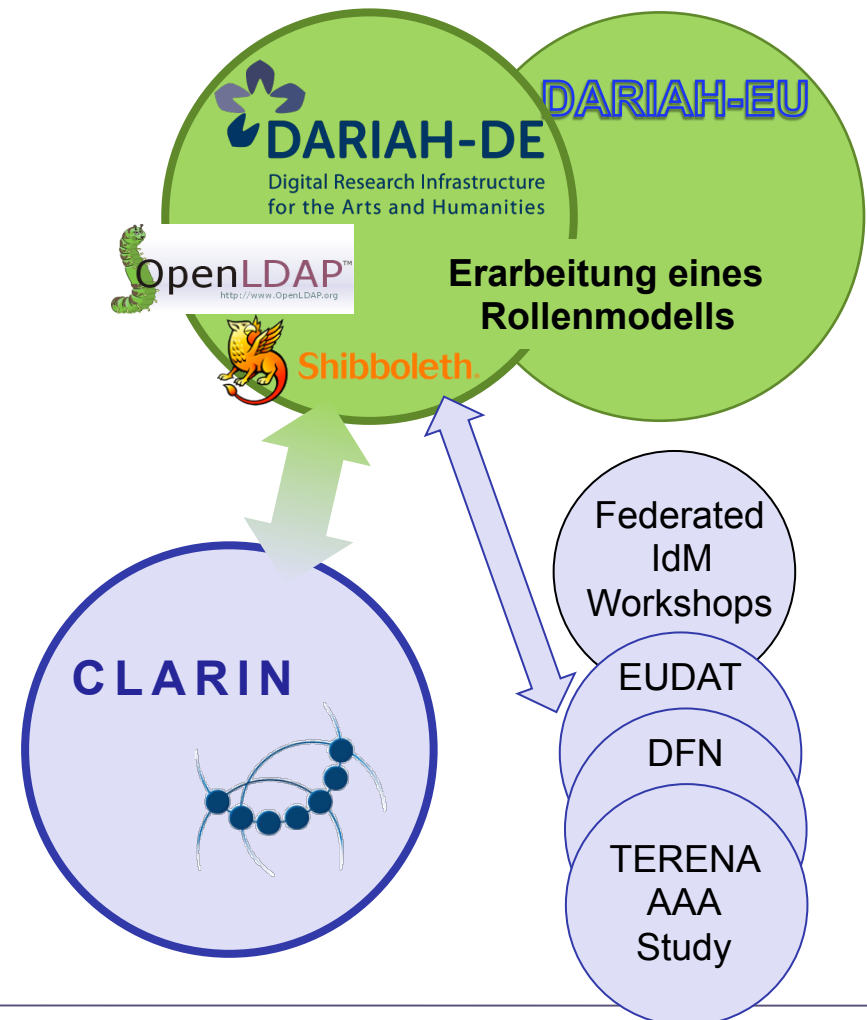
IDM: Identitätsmanagement
AAI: Authentifizierung- und
Autorisierungs-Infrastruktur

Föderierte DARIAH-AAI

- SAML, so dass alle Mitglieder einer Föderation Berechtigungen erhalten können
- LDAP-Server in Garching und Göttingen
- Integration von „Heimatlosen“
- Attribut-Aggregation aus der Heimatorganisation und DARIAH

Enge Koordinierung mit CLARIN

- Technische Interoperabilität
- Gemeinsame Föderation?
- Gemeinsamer „Call for Action“
(<http://www.clarin.eu/page/3500>)



DIGITALE FORSCHUNGSINFRASTRUKTUR FÜR DIE GEISTES- UND KULTURWISSENSCHAFTEN

DARIAH-DE-PORTAL

Beratung – Dokumentation – Standardisierung

LEHRE

- Summer Schools
- Expertenseminare
- Methodenworkshops
- Curricula

FORSCHUNG

Erkenntnisinteresse → Forschungsfrage → Recherche/Erschließung → Analyse → Interpretation → Publikation

Software Hosting Services

Demonstratoren und fachwissenschaftliche Dienste

- MEISE
- Archive in a Box
- Voyant Tools
- Monasterium
- Personendaten
- Relationen im Raum
- Interoperabilität Normdaten
- Differenzanalyse
- Virtuelles Skriptorium

Generische Dienste

- Entwicklung von Diensten auf Anforderung
- Geo-Browser
- Generische Suche
- Collection Registry
- Schema Registry

Forschungsdaten

Metadaten

Operative IT-Dienste

- AAI
- Sicherheit (operativ, strukturell)
- Qualitätssicherung

Software Hosting Services

Basisdienste

- Kollaborative Arbeitsumgebungen (Wiki, Developer Portal, Etherpad)
- PID Service
- Storage, Archive
- Datenbanken

Plattform Hosting Service

Bereitstellung VM mit Basissystem

Infrastructure Hosting Service

Bereitstellung virtueller Computer-Cluster

Operative IT-Dienste

- Ressourcen (Koordination, Bereitstellung)
- Helpdesk
- Monitoring

DARIAH Architektur: Fazit

Stand der Infrastruktur:

Kerndienste größtenteils vorhanden oder im Aufbau: AAI, Monitoring, Hosting, ...

Wesentliche **Basisdienste** vorhanden: Kollaboration, PID, Storage

Generische Dienste:

| | | |
|---------------------|---|---------------------------|
| Collection Registry | – | Version 1.0 abgeschlossen |
| Schema Registry | – | 1. Prototyp |
| Suche | – | im Aufbau |

Wesentliche Eigenschaften:

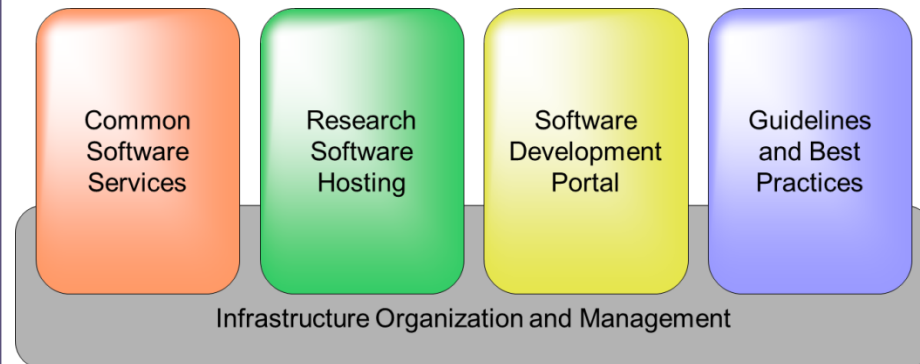
- Transparenz
- Langfristigkeit, Zitierfähigkeit
- Aufbau einer europäischen Infrastruktur:
 - ✓ Zuverlässigkeit der Komponenten
 - ✓ Design für Produktionsbetrieb
 - ✓ Erweiterbarkeit
 - ✓ Nachhaltigkeit



AP 1 e-Infrastruktur

- AP 1.1 AP 1 und VCC 1 Koordination (DARIAH-DE) und (DARIAH-EU)
- AP 1.2 Interoperabilität Meta-Daten (\leftrightarrow AP 3 Forschungsdaten)
- AP 1.3 Interoperabilität Dienste (\leftrightarrow AP 2 Forschung und Lehre, AP 3)
- AP 1.4 Geisteswissenschaftliche Dienste (\leftrightarrow AP 2, AP 3)

Unterarbeitspakete:
Bit Preservation, AAI und Hosting



Treffen DARIAH AP1 – CLARIN, 8.2.2012