

# Aufbau eines Netzwerkes von zertifizierten Open Access Repositories

## Ziele des Projektes und seine technische Umsetzung

Gefördert durch:

**DFG**

In Abstimmung mit:



Peter Schirmbacher,

Humboldt-Universität zu Berlin

Thomas Severiens,

Universität Osnabrück

# Aufbau eines Netzwerkes von zertifizierten Open Access Repositories

## Projektpartner:

- Humboldt-Universität zu Berlin,  
Computer und Medienservice, Institut für Bibliotheks-  
und Informationswissenschaft
- Universität Osnabrück  
Fachbereich Mathematik / Informatik
- Staats- und Universitätsbibliothek Göttingen

Gefördert durch:

**DFG**

In Abstimmung mit:



# Open Access - Repository-Netzwerk in Deutschland

Gliederung:

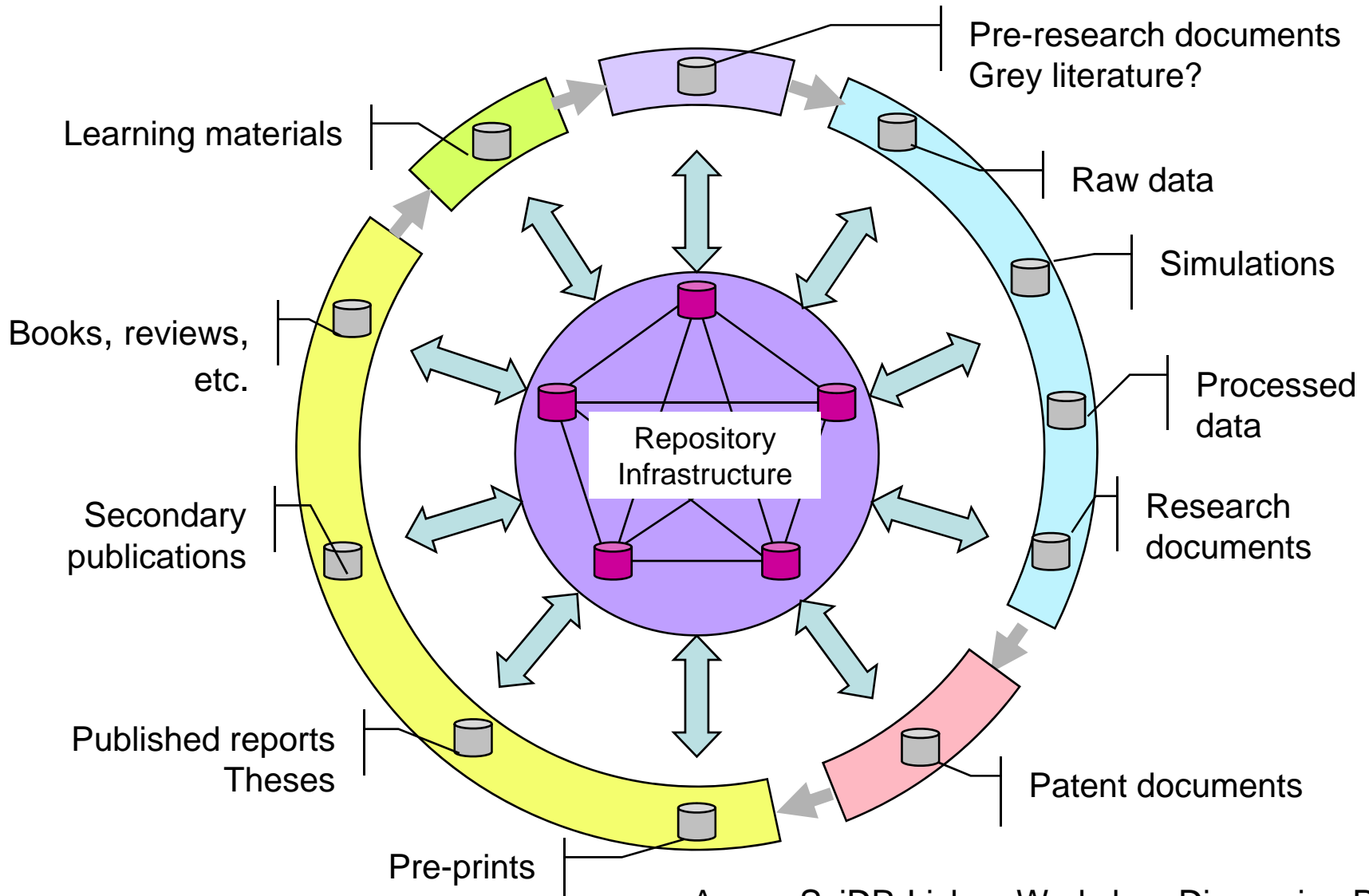
1. Stand der Repositories in Deutschland
2. Repositories im internationalen Kontext
3. Möglichkeiten zur Erhöhung der Sichtbarkeit von Repositories
4. Aufbau eines Netzwerks von zertifizierten Repositories

# Open Access Repositories: ihre Rolle im Publikationsprozess

- Einer von mehreren Publikationswegen
- Green Road to Open Access
  - individual Repository
  - institutional Repository
  - disciplinary Repository
- Bedeutung der Repositories
  - Präsentation des Wissenschaftsgebiets
  - Präsentation der wissenschaftlichen Einrichtung
  - Präsentation der Forschergruppe
- Vorteile von Repositories
  - Schneller Weg zu einer aktuellen Publikation
  - Weltweite uneingeschränkte Verfügbarkeit



# Research cycle & repositories



# Open Access Repositories: Stand in Deutschland

Ergebnisse der DINI-Umfrage bzw. Statistik:


- 124 Repositories bei DINI registriert
- 120 Meldungen bei Open DOAR
  - Directory of Open Access Repositories
  - Mehrfachmeldungen pro Institution
- 86 Meldungen bei ROARMAP
  - Registry of Open Access Repository Material Archiving Policies
- 61 Meldungen bei OAI als Dataprovider
- 21 zertifizierte Repositories
  - Davon drei Services mit dem Zertifikat 2007

# Open Access Repositories: im internationalen Kontext

Country	ROARMAP	Open DOAR
US	227	286
UK	109	116
Germany	86	120
Japan	46	66
Canada	44	38
France	41	37
Australia	34	53
Sweden	35	32
Niederlande	23	44

# Directory of World Repositories

January '08

[home](#) [world countries](#) [world rank](#) [rank by country](#) [european rank](#) [latin american rank](#) 
[> home > top 200 repositories](#)
[http://www.webometrics.info/top200\\_rep.asp](http://www.webometrics.info/top200_rep.asp)

## Data

[About Us](#)[About the Ranking](#)[Top 4000 Universities](#)[Premier League](#)[Top USA & Canada](#)[Top Latin America](#)[Top Europe](#)[Top Asia](#)[Top Middle East](#)[Top Oceania](#)[Top Africa](#)[Distribution by Country](#)[Top 1000 R&D Institutes](#)[Research Councils](#)[Best Practices](#)[Notes](#)

## Catalogue














[Universities by country](#)[R&D Centres by country](#)

## Academic Repositories

[Top 200](#)[Directory](#)[Methodology](#)

## Top 200 Repositories

[First](#) | [Previous](#) | [Next](#) | [Last](#) | [Repositories 1 to 50 of 200](#)

WORLD RANK	REPOSITORY	COUNTRY	POSITION			
			SIZE	VISIBILITY	RICH FILES	SCHOLAR
1	ARXIV.ORG E-PRINT ARCHIVE		2	2	37	3
2	RESEARCH PAPERS IN ECONOMICS		3	4	85	4
3	E LIS: RESEARCH IN COMPUTING AND LIBRARY AND INFORMATION SCIENCE		14	15	3	38
4	UNIVERSITY OF SOUTHAMPTON EPRINTS		40	14	7	26
5	ACADEMIC ARCHIVE ON LINE		23	16	83	8
6	CITSEER		1	1	175	6
7	ECOLE POLYTECHNIQUE FEDERALE DE LAUSANNE INFOSCIENCE		8	7	74	78
8	UNIVERSITY OF TSUKUBA REPOSITORY		13	28	27	61
9	INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE ARCHIVE OUVERTE		32	11	110	11
10	UPPSALA UNIVERSITY ACADEMIC ARCHIVE ON LINE		29	27	46	37
11	UNIVERSIDADE DO MINHO REPOSITORIO		16	48	22	47
12	UNIVERSITY OF MICHIGAN DEEP BLUE		26	45	52	15
13	UNIVERSITY OF SAINT GALLEN FORSCHUNGSPLATTFORM ALEXANDRIA		12	18	42	136



Information

Methodoloav

Links

Contact Us

Site Map

Statistics

Disclaimer

Search Webometrics for:

Submit

Category:

All

CybermetricsLab



14	MIT DSPACE		34	54	17	13
15	UNIVERSITY OF SOUTHAMPTON: DEPARTMENT OF ELECTRONICS AND COMPUTER SCIENCE		65	31	2	83
16	ORGANIC EPRINTS		95	29	8	44
17	HUMBOLDT UNIVERSITAT ZU BERLIN PUBLIKATIONSSERVER		24	57	11	45
18	THE ARTS AND HUMANITIES DATA SERVICE		10	6	9	241
19	UNIVERSITY OF CALIFORNIA ESCHOLARSHIP REPOSITORY		31	10	233	22
20	REVUES SCIENTIFIQUE EN SCIENCES HUMANES ET SOCIALES		9	20	247	7
21	SCIENTIFIC AND TECHNICAL INFORMATION NETWORK		4	12	301	2
22	GEORGIA TECH'S INSTITUTIONAL REPOSITORY		11	88	35	21
23	UNIVERSIDAD DE LOS ANDES REPOSITORIO INSTITUCIONAL		45	61	84	17
24	UNIVERSITAT HEIDELBERG DOKUMENTENSERVER		91	30	113	34
25	UNIVERSITY OF MARYLAND DIGITAL REPOSITORY		43	82	5	40
26	OHIO STATE UNIVERSITY KNOWLEDGE BANK		15	62	58	117
27	DIALNET		6	9	365	1
28	QUEENSLAND UNIVERSITY OF TECHNOLOGY EPRINTS		90	55	68	41
29	OREGON STATE UNIVERSITY SCHOLARSARCHIVE		151	43	19	50
30	UNIVERSITY OF OREGON SCHOLARS' BANK		33	75	14	108
31	GEORG AUGUST UNIVERSITAT GOTTINGEN DOKUMENTENSERVER		81	52	55	84

Fertig



World Universities' r...

Posteingang von sc...

Microsoft PowerPoi...

DE

Position	Repository	Ranking Visibility (50%)	Ranking Size (20%)	Ranking Rich Files (15%)	Ranking Scholar (15%)
17	HU Berlin edoc	57	(1) 24	(2) 11	(3) 45
24	Heidelberg HeiDok	(2) 20	91	113	(2) 34
31	Göttingen GOEDOC	52	(5) 81	55	84
36	Stuttgart HS-Schriftenserver	(1) 19	(2) 59	176	146
41	Münster miami	69	(4) 76	24	157
42	Chemnitz MONARCH	(3) 21	146	(5) 15	240
48	München LMU Dissertationen	53	121	196	(1) 32
56	Munich Personal RePEc Archive	47	129	219	(5) 68
58	Digitale Bibliothek Thüringen	66	92	(3) 12	281
61	Dortmund Eldorado	101	126	(1) 4	142
78	Duisburg-Essen Publications	141	(3) 74	(4) 13	218
90	München ePub	73	141	260	123
92	Marburg Archivserver	137	170	88	(4) 65
93	Giessen Elektronische Bib.	89	159	165	176
102	RWTH Aachen, HS-Schriften	108	254	93	100
115	Saarland SciDoc	78	202	210	231
123	Saarland PsyDoc	111	195	173	233
143	Bremen Elib	42	334	282	265
153	MPG Edoc	(4) 39	364	531	82
179	DLR Elib	74	206	360	482
185	NRW DIPP Digital Peer Publ.	44	311	340	474
192	Bielefeld Bieson	188	291	138	247
198	BBAW	(5) 41	367	373	431

# Probleme gegenwärtiger Open Access Repositories (Nutzersicht)

## Sichtbarkeit

- In der Vielfalt des Web-Angebotes gehen die IR unter
- Die Einordnung des Angebotes in den Gesamtzusammenhang bleibt unklar (Fehlende Statistik; Verhältnis zum Impact; ...)

## Angebotsqualität

- Kaum Aussagen zur Qualität des Angebotes
- Selten PeerReview

## Angebotsumfang

- Zahl der Dokumente ist meist zu gering
- Die Authentizität und Integrität wird bezweifelt

## Serviceumfang

- Die Suchfunktionalitäten sind meist unterentwickelt
- Es gibt nur wenige Zusatzfunktionen (Printing on Demand, ..)

# Probleme gegenwärtiger Open Access Repositories (Autorensicht)

## Sichtbarkeit

- Der Autor sucht die fachbezogene und weniger die Anerkennung innerhalb der Institution
- Der Autor will gelesen werden, hat jedoch meist zu wenig Kenntnisse über die Sichtbarkeit von Angeboten

## Open Access Gedanken steht nicht im Vordergrund

- Durch die gegenwärtig (noch) dominante Publikationskultur steht die Zeitschriftenpublikation im Vordergrund
- IR wird nur als Zweitpublikation gesehen und damit als zusätzliche Belastung

## Serviceumfang

- An die Vorgaben der Verlage hat man sich gewöhnt, die IR-Anforderungen werden als „Bevormundung“ gewertet
- Die Unterstützung für Autoren ist bei vielen IR und vor allem disziplinären Repositories zu gering

# Probleme gegenwärtiger Open Access Repositories (Betreibersicht)

## Organisatorische Verankerung

- Die Aufgabe wird noch zu selten als gleichwertige Tätigkeit anerkannt
- Vielfach fehlt die Qualifikation als Mischung aus bibliothekarischen und Kenntnissen der Informatik

## Inhaltliche Auskleidung

- Die Betreiber sind keine Fachwissenschaftler und haben so auf den Inhalt und die Autoren nur selten Einfluss
- Die Betreiber orientieren auf Quantität, weil sie auf die inhaltliche Qualität keinen Einfluss haben
- Es stehen Gesichtspunkte des Sammelns und nicht des Anbietens im Vordergrund

## Serviceumfang

- Bei der Gradwanderung zwischen perfektem System und zufriedenem Autor / Nutzer steht häufig das System im Mittelpunkt

# Netzwerk zertifizierter Open Access Repositories

## Ziele (allgemein)

- Erhöhung der Sichtbarkeit deutscher Repositories
- Deutscher Konzentrationspunkt für DRIVER Aktivitäten
- Verbesserung der Infrastruktur in Deutschland durch die Unterstützung bei der Zertifizierung weiterer Server
- Mitwirkung bei der Erarbeitung von Verfahren und Qualitätskriterien weltweit
- Beispiellösung für andere Länder

# Netzwerk zertifizierter Open Access Repositories

## Ziele (technologisch)

- Mehrwert-Dienste
  - Aufbau eines OAI-Serviceproviders
  - Übergreifenes Retrieval in Metadaten und Volltexten
  - Übergreifendes Browsing entlang der DDC und ggf. vertiefenden Fachklassifikationen
  - Exportschnittstellen
    - Printing on Demand
    - XML-Export
      - OAI-Dataprovider
      - Google-Scholar
      - Vascoda und internationale Dienste
  - Übergreifende Statistik
  - Zitationsanalyse und Browsing

# Netzwerk zertifizierter Open Access Repositories

## **Ziele** (technologisch, Fortsetzung)

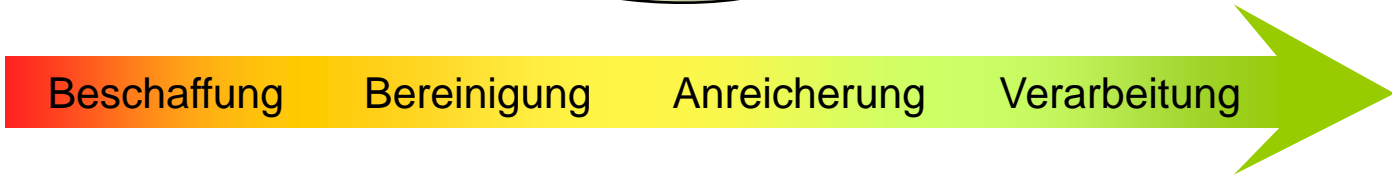
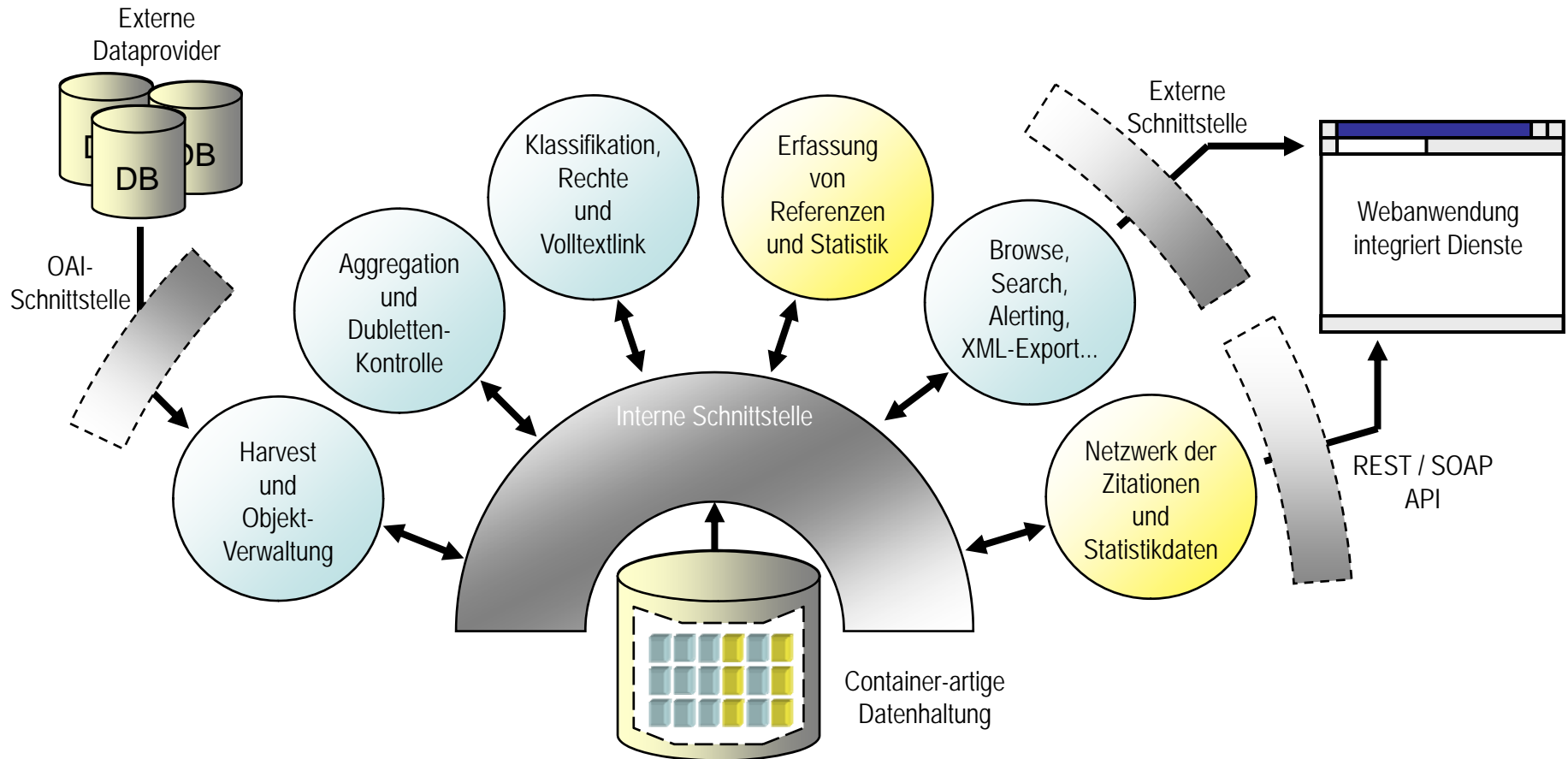
- Infrastruktur-Maßnahmen
  - Gemeinsames Wirken der Betreiber zertifizierter Server (national wie international),
  - Unterstützung beim Erreichen der Qualitätskriterien,
  - Rückkopplung zur Weiterentwicklung des Zertifikats.
- Verbreitung des KnowHow aus Deutschland



# Prinzipien der technischen Realisierung

- Aufbau eines Infrastrukturlayers
  - Modular,
  - Verteilte Dienste aber zentrale Datenhaltung,
  - Anreicherung und Clearing der Daten,
  - Strukturelle Aufbereitung auch durch Drittdaten
- Import via OAI-PMH
- Export über diverse Schnittstellen
- Einbindung von Modulen (Funktionalitäten über Webservices)
- Plattform-unabhängig, Sprach-übergreifend, Erweiterbar

# „Spektrum der Dienste“ in der modularen OA-Netzwerk-Architektur



## OAI-PMH Import

- Auswahl der Dataprovider
  - DINI-Zertifikat plus evtl. Einzelfälle
- Speicherung des Datenstreams
  - Tägliches inkrementelles Harvesting
  - Regelmäßiges komplettes Harvesting
  - Validierung des XML, sonst moderierte Rückmeldung an die Betreiber (Ansprechperson im DINI-Zertifikat benannt)
  - Validierung des OAI-PMH (funktionieren Resumption-Tokens? Datensätze vollständig?)
  - Validierung der Datensets (unlogische Datumsangaben? Sprachangaben sinnvoll? Volltext-Link vorhanden? DCMI-Fehler? Etc. sonst moderierte Rückmeldung)

## OAI-PMH Import, Fortsetzung

- Aggregation der Metadaten
  - Objekte werden anhand ihrer OAI-Identifizier reproduzierbar identifiziert
  - Ähnliche Objekte anhand einer Ähnlichkeits-Analyse zusammengeführt (Dubletten, Versionen, Formate), Repository-übergreifend
  - Ablage der aggregierten Metadaten in einer zentralen Objektverwaltung.

## API für externe Dienste

- Idee eines verteilten konkurrierenden Netzes der Objektorreicherung und Verwaltung
  - SOAP- und REST-Schnittstelle erlaubt
    - Auflistung der Objekte in der Objektverwaltung
    - Zugriff auf Datenbestand einzelner Objekte
    - Ändern einzelner Metadaten-Werte in einem Objekt
    - Ggf. Zugriff auf ergänzende Tables
  - Rechtevergabe durch Modulverwaltung

## Human friendly & XML-Export

- GUI
- RSS/ATOMS
- Alerting für neue Objekte nach Fachklassifikation abonnierbar
- OAI-Dataprovider
- Export an z.B. Google-Scholar, MSN books, vascoda etc.

## Koordination durch Kooperation

- Bei Fragen zu Schnittstellen, Konzepten, Kooperationen, Strategien, eben allem:  
Zögern Sie nicht, uns hier oder per E-Mail anzusprechen! **Es sind viele OA-Netzwerk-Ansprechpartner hier:**
  - Susanne Dobratz (HU-Berlin)
  - Manuel Klatt-Kafemann (HU-Berlin)
  - Robin Malitz (HU-Berlin)
  - Prof. Dr. Peter Schirmbacher (HU-Berlin)
  - Dr. Birgit Schmidt (SUB Göttingen)
  - Thomas Severiens (Univ. Osnabrück)