

# Publikationsdatenbanken: HGF, FZJ & DESY

Erfahrungen und Ausblicke

Martin Köhler

DINI-/Helmholtz-Workshop: Repositorien –  
Praxis und Vision  
Berlin, 30. November 2010

- Repositorien in der Helmholtz Gemeinschaft (HGF)
- Die gemeinsame HGF Publikationsdatenbank
- Kooperation Forschungszentrum Jülich (FZJ) und Forschungszentrum Deutsches Elektronen-Synchrotron (DESY)
- Weitere Entwicklungen
- Fazit





*„Die Helmholtz-Gemeinschaft leistet Beiträge zur Lösung großer und drängender Fragen von Gesellschaft, Wissenschaft und Wirtschaft durch wissenschaftliche Spitzenleistungen in sechs Forschungsbereichen. Sie ist mit 30.000 Mitarbeiterinnen und Mitarbeitern in 16 Forschungszentren und einem Jahresbudget von rund 3 Milliarden Euro die größte Wissenschaftsorganisation Deutschlands. Ihre Arbeit steht in der Tradition des großen Naturforschers Hermann von Helmholtz (1821-1894).“*

> Strategie **Programmorientierte Förderung**: Ressourcenbündelung in Programmen: Energie, Erde und Umwelt, Gesundheit, Schlüsseltechnologien, Struktur der Materie, Luftfahrt, Raumfahrt und Verkehr

- WissenschaftlerInnen der Zentren legen Inhalte ihrer Forschung in strategischen Programmen fest.

Kooperation über Grenzen von Institutionen und Disziplinen  
Wettbewerb um Forschungsgelder  
Forschungspolitische Vorgaben von den Zuwendungsgebern  
Begutachtung durch Spitzenwissenschaftler



> Forschungsmöglichkeiten für Gastwissenschaftler an einzigartigen Großgeräten

> Leistungen: **Publikationen**, Patente, Ausgründungen, Einnahmen aus Lizenzverträgen, Kooperationsprojekte mit Unternehmen

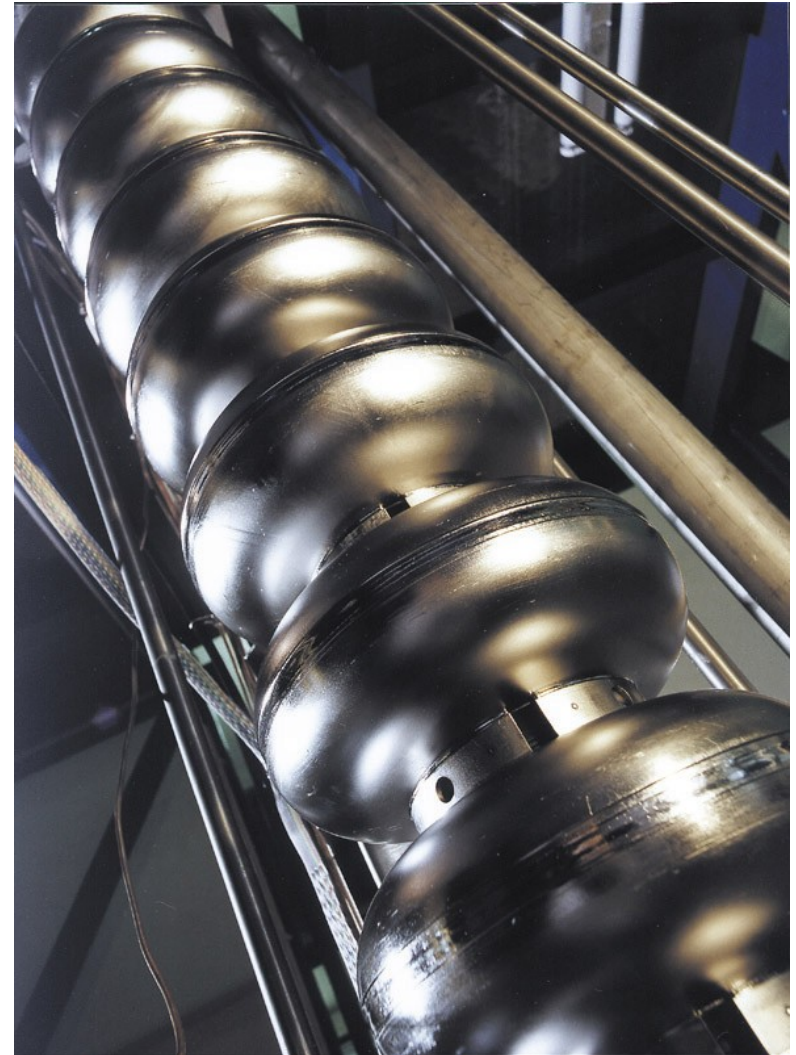


- > Die Helmholtz-Gemeinschaft gehörte 2003 zu den Erstunterzeichnern der "Berliner Erklärung"
- > **2004: Beschluss der Mitgliederversammlung der HGF zur Umsetzung der „Berliner Erklärung“**  
*"Publikationen aus der Helmholtz-Gemeinschaft sollen künftig ohne Ausnahme kostenlos zugänglich sein, soweit nicht ausdrückliche Vereinbarungen mit Verlagen und anderen dem entgegenstehen."*
- > **Grüner Weg:** Aufbau, Vernetzung und Standardisierung der HGF Repositorien



# Institutionelle Repositorien in der HGF

Name	OAI-PMH	Software
AWI	Ja	Fedora
DESY	Ja	Eigenentwicklung
DLR	Ja	EPrints
FZJ	Ja	DSpace
KIT	Ja	Eigenentwicklung
GFZ	Ja	Eigenentwicklung
GSI	Ja	Eigenentwicklung
HZG	Ja	Eigenentwicklung
HZB	Ja	Eigenentwicklung
HZI	Ja	BMC Open Repository
MDC	Ja	EPrints
IPP	Ja	Eigenentwicklung
UFZ	Ja	?



# Gemeinsame HGF Publikationsdatenbank

## > Vorhandene dezentrale Systeme sollen weiter genutzt werden

- Lizenzgründe („Grüner Weg“)
- Unterschiedliche Workflows in den Zentren
- Vorhandene Systeme komplex und divers
- Kein gemeinsames Vokabular,



## > Zentraler Nachweis aller HGF Publikationen (realisiert durch KIT)

- Harvester mittels OAI-PMH, da alle Systeme diese Schnittstelle haben
- Metasuchmaschine, aber keine reine Metasuche, da nicht performant

## > [www.helmholtz.de/aktuelles/publikationen/wisspublikationsdatenbank](http://www.helmholtz.de/aktuelles/publikationen/wisspublikationsdatenbank)

- Link direkt von [www.helmholtz.de](http://www.helmholtz.de)
- Suche von Titel, Autor, Jahr und Zentrum möglich
- Link zum Volltext via original Datenbank
- Inhalt: Metadaten von 11 HGF Zentren mit insgesamt ca. 250k Datensätzen

## Erfahrungen

### > Verschiedene freie Programmbibliotheken wurden untersucht

- Harvester: PKP, OJAX, Omeka

### > Entscheidung für: PKP & Omeka

- Open Source, aber Achtung: Umfeld beachten
- Zuverlässigkeit
- Skalierbarkeit

### > Probleme

- Zeichensätze
- Abbruch beim Harvesten (OJAX, PKP, Omeka)

Harvesten ist langwieriger und häufig instabiler, fehleranfälliger Prozess

- Unterschiedliche Ausgangssysteme (Technik und Inhalte)

Unterschiedliche Ausprägungen des Dublin Core Formates (→ Nur Kernfelder werden übernommen)



## 1. Veröffentlichungsdatenbank (VDB)

- Bibliographische Daten (ca. 45k Einträge)
- Nutzung seit über zehn Jahren

## 2. Jülicher Wissenschaftliche Elektronische Literatur (JUWEL)

- Volltexte (ca. 3,7k Einträge)
- Basis DSPACE

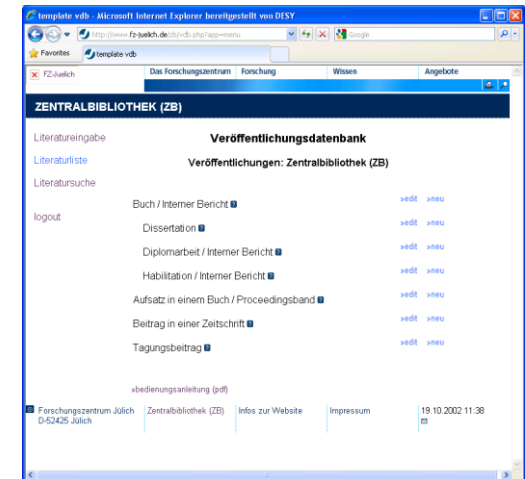
### > Workflow (komplex)

- Eingabe durch InstitutssachbearbeiterInnen
- Daten werden in der VDB und JUWEL erfasst
- Offline Synchronisation

### > Bibliothek sehr präsent im Zentrum

- OA: Newsletter, Webseite, Literaturlisten „Plugin“, ...

### > Systemtrennung u.a. schlecht für Nutzer



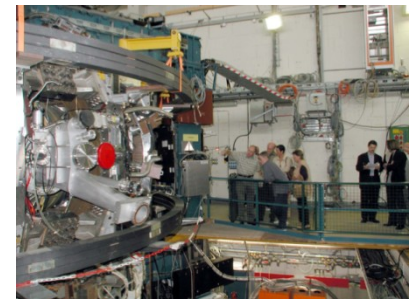


## 1. Zentrum: Wissenschaftlicher Ergebnisbericht

- Literaturlisten
- Kennzahlen

## 2. Nutzer: Publikationsunterstützung

- Publikationslisten („Plugin“ für individuelle Webseiten)
- Im- und Export von Literaturlisten



➤ Eingabedisziplin u.a. auf Grund des zentralisierten Workflows hoch

➤ „Fehlende“ Literatur wird bewusst zentral nicht nachrecherchiert

- Vermeidung von „Fehlzählungen“

# Beispiel Deutsches Elektronen-Synchrotron (DESY)

- > Bis August 2006 kein dediziertes System zur Erfassung von Publikationen
- > Volltexte nur als Linkliste(n) im Web bzw. in SPIRES
- > Ende 2004: Direktoriumsbeschluss zur

*„Einführung einer Publikationsdatenbank für Berichte zur programmorientierten Förderung (PoF) und zur Nutzung im Bereich „Forschung mit Photonen“*

- Einführung unter Nutzung von Synergien
  - Kooperation- und Lizenzvertrag mit FZJ
  - Institutionelles Repositorium mit „DOOR“-Anbindung
  - Anpassung, Entwicklung, Produktionsbetrieb des FZJ Systems mit Hilfe ext. Firma innerhalb von 1,5 Jahren ausschließlich mit vorhandenen Ressourcen
- > Stabiler Betrieb seit 1. August 2006
    - Verfügbarkeit ~ 99,95% in 2010 (Ausfallzeiten enthalten auch geplante Wartungen)



## > Eingaben, Vollständigkeit der Daten

- Klare Managementvorgaben wichtig
- Klare Information darüber, von wem die Daten wofür benötigt werden
- Klare Zuständigkeiten (bei DESY Mischung aus zentraler und dezentraler Eingabe)
- „Gesunde“ Mischung aus Pflichtfeldern und optionalen Eingaben
- Für die Bibliothek: Sinnvolle Prüfungen (z.B. Namen, TeX Syntax, Datum, ...) und Vorgaben (z.B. Ansetzungen bei Zeitschriften, Konferenzen)
- Importmöglichkeiten helfen bei der Akzeptanz

## > Volltexte

- Publikationsordnung
- Informationsveranstaltungen
- Individuelle Ansprache (z.B. per Mail) hilfreich



## > Eintrag der Datenbank in (Meta)Suchmaschinen (google, BASE, ...)

- > Eingaben erfolgen kurz vor der Deadline
  - Hohe Systemlast bei vielen (~50%) externen Nutzern (HASYLAB)
    - Hohe Zuverlässigkeit in dieser Zeit nötig
- > Ausgabe von Literaturlisten für Jahresbericht in LaTeX (Formeln, Autorennamen)
- > Direktes Anschreiben der Nutzer durch die Bibliothek zur Abgabe von Volltexten ist erfolgreich
  - Gesamt: 12600 Einträge davon 5800 Volltexte (~46%)
  - ISI Publikationen: 3400 Einträge, davon 1800 im Volltext (~50%)
- > NutzerInnen Wünsche
  - Suchfunktionen (inkl. Volltext)
  - **Korrekte** Publikationslisten (Müller, Wu) auch auf Gruppenebene/Beamline
  - Im und Export in verschiedenen Formaten



Viele Erweiterungswünsche bei FZJ und DESY gleich.

- Im- und Export, Suche, Statistik, Volltextworkflow, Tagging, ...

## > Stand der Dinge:

- Proprietäre Software (php Applikation)
- Trotz gemeinsamen Kerns mittlerweile stark divergente Codebasis
- Code veraltet und nicht standardisiert
- Aufwand sehr hoch, da vieles neu entwickelt werden müsste

## > Wunsch:

- Standardisierung (MARC21)
- Standardsoftware (Invenio)

DESY Bibliothek ist Partner bei SPIRES (Hochenergiephysikdatenbank): Migration zu INSPIRE (INVENIO)

- Gemeinsame Anpassungen (u.a. Dokumententypen, Feldzuordnungen, Statistik, ...)

## > Zusätzlicher Partner GSI



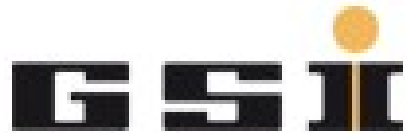
# Partner und Personen (Publikationsdatenbank)



> **C. Plott**, A. Wagner

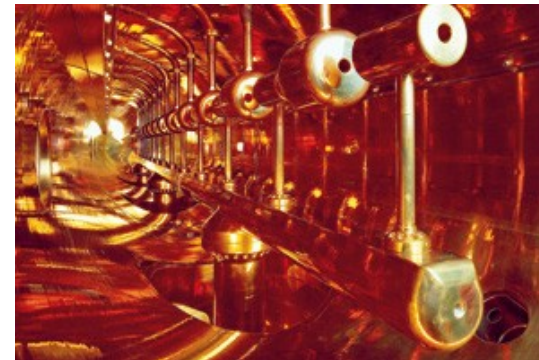


> Z. Akopov, T. Pazera



> S. Hesselbach, T. Badura

- > Kickoff (FZ Jülich) 5.7 & 6.7.2010
- > CERN Master Class 18.-20.10.2010
- > Enge Kooperation
  - INVENIO Entwickler
  - INSPIRE Collaboration



## > Kooperations- und Lizenzvertrag ✓

- Nachnutzung der beim FZJ entwickelten Software
- Trotz vieler Unterschiede im Detail relativ schnelle erfolgreiche Einführung des Systems bei DESY
- Aufgrund gemeinsamer Codebasis konnte Kosten für Erweiterungen (z.B. OAI-PMH Schnittstelle) geteilt werden



## > Zusammenarbeit bei Publikationsdatenbank erleichterte weitere gemeinsame Projekte ✓

- FATER (**F**ast **A**ccess **T**o **E**lectronic **R**essources)
- Gemeinsames „Vokabular“ für Publikationen und Kennzahlen
- Cern INVENIO als Nachfolgesystem (gemeinsam mit GSI)



## > Repositorien werden genutzt und fördern Zusammenarbeit ✓

