

Invenio @ HGF JuSER

June 17, 2013 | Alexander Wagner |

Overview



- History
- Basic Usage
- Installation
- Backend



Invenio @ HGF

Part I: History

June 17, 2013 | Alexander Wagner

VDB and JUWEL

Publications database **VDB**:

- Centralized database of the scientific output
- > 62.000 entries from ≈ 13 years ($\approx +4800/a$)
- **Obligatory input**¹ (editors at each institute, crosschecked by ZB)
- Basis for
 - Scientific Report
 - Evaluations
 - Publication lists (e. g. [WWW](#))

¹cf. Publication Guidelines

Forschungszentrum Jülich – Publikationen – Iceweasel

File Edit View History Bookmarks Tools Help Search News Bank Uni Bib Comp Cam

w.fz-juelich.de/inm/inm-1/DE/Service/Publikationslisten/publikationslisten_node.html JuSER

Suchbegriff MITARBEITERSUCHE | DEUTSCH | ENGLISH Institutsverzeichnis

Institut für Neurowissenschaften und Medizin
Strukturelle und funktionelle Organisation des Gehirns (INM-1)

AKTUELLES FORSCHUNG LEISTUNGEN KARRIERE ÜBER UNS

INM-1 Forschung Publikationen

FORSCHUNG

Architektur und Hirnfunktion
Multimodale Bildverarbeitung
Modellierung kortikaler Systeme
Genomic Imaging
Publikationen

SERVICE

Ansprechpartner
Mitarbeiter
Publikationen
Anfahrt
Downloads

Zeitschriftenbeiträge 2012

Amunts, K.; Zilles, K.
Architecture and organizational principles of Brocas region
Trends in cognitive sciences 16, 418 - 426 (2012) [10.1016/j.tics.2012.06.005]

Bis, J. C.; et, a.
Common variants at 12q14 and 12q24 are associated with hippocampal volume
Nature genetics 44, 545 - 551 (2012) [10.1038/ng.2237]

Boutros, N. N.; Gjini, K.; Eickhoff, S. B.; Urbach, H.; Pflieger, M. E.
Mapping repetition suppression of the P50 evoked response to the human cerebral cortex
Clinical neurophysiology 00, 00 (2012) [10.1016/j.clinph.2012.10.007]

Bzdok, D.; Laird, A.; Zilles, K.; Fox, P.; Eickhoff, S. B.
An investigation of the structural, connectonal, and functional sub-specialization in the human amygdala
Human brain mapping (2012) [10.1002/hbm.22138]

Bzdok, D.; Schilbach, L.; Vogeley, K.; Schneider, K.; Laird, A. R.; Langner, R.; Eickhoff, S. B.
Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy
Brain structure & function 217, 783 - 796 (2012) [10.1007/s00429-012-0380-y]

PUBLIKATIONEN INM-1

Publikationen 2011
Publikationen 2010
Publikationen 2009
Publikationen 2008

→ Mehr

In Print

VDB and JUWEL

Publications database **VDB**:

- Centralized database of the scientific output
- > 62.000 entries from ≈ 13 years ($\approx +4800/a$)
- Obligatory input**² (editors at each institute, crosschecked by ZB)
- Basis for
 - Scientific Report
 - Evaluations
 - Publication lists (e. g. WWW)

JUWEL:

- Berlin Declaration for **Open Access**
- Institutional repository (≈ 4500 full text files, $\approx +560/a$)

²cf. Publication Guidelines

Longitudinal deformation-based morphometry reveals spatio-temporal dynamics of brain volume changes in patients with corticobasal syndrome - JuSER - Mozilla Firefox

File Edit View History Bookmarks Tools Help Most Visited Search News Uni FR Bib Wiki SRV Chess Comp >> RSS

juser.fz-juelich.de/record/130981

admin :: Abmelden

JuSER

SUCHEN ABSENDEN PERSONALISIEREN HILFE VERWALTUNG

Hauptseite > Longitudinal deformation-based morphometry reveals spatio-temporal dynamics of brain volume changes in patients with corticobasal syndrome

Information Diskussion Dateien Plots

Journal Article FZJ-2012-00157

Longitudinal deformation-based morphometry reveals spatio-temporal dynamics of brain volume changes in patients with corticobasal syndrome

→ Pieperhoff, Peter^{*} ; → Ferrea, Stefano ; → Krause, Holger ; → Groiss, Stefan Jun ; → Elben, Saskia ; → Wojtecki, Lars ; → Zilles, Karl^{*} ;
→ Amunts, Katrin^{*} ; → Schnitzler, Alfons ; → Südmeier, M (Corresponding author)

2012
PLoS Lawrence, Kan.

Published in: → PLoS one 7 (7) e41873 -> [10.1371/journal.pone.0041873] →

→ Pubmed; → Pubmed Central Fulltext;
→ Web of Science (WOS); → Citing articles (WOS); → Related articles (WOS)

Report No.: FZJ-2012-00157

Abstract: Corticobasal syndrome (CBS) is a rare neurodegenerative disorder characterized by a progressive and asymmetric manifestation of cortical and basal-ganglia symptoms of different origin. The spatiotemporal dynamics of cerebral atrophy in CBS is barely known. This study aimed to longitudinally quantify the individual dynamics of brain volume changes in patients with CBS as compared

Contributing Institute(s):

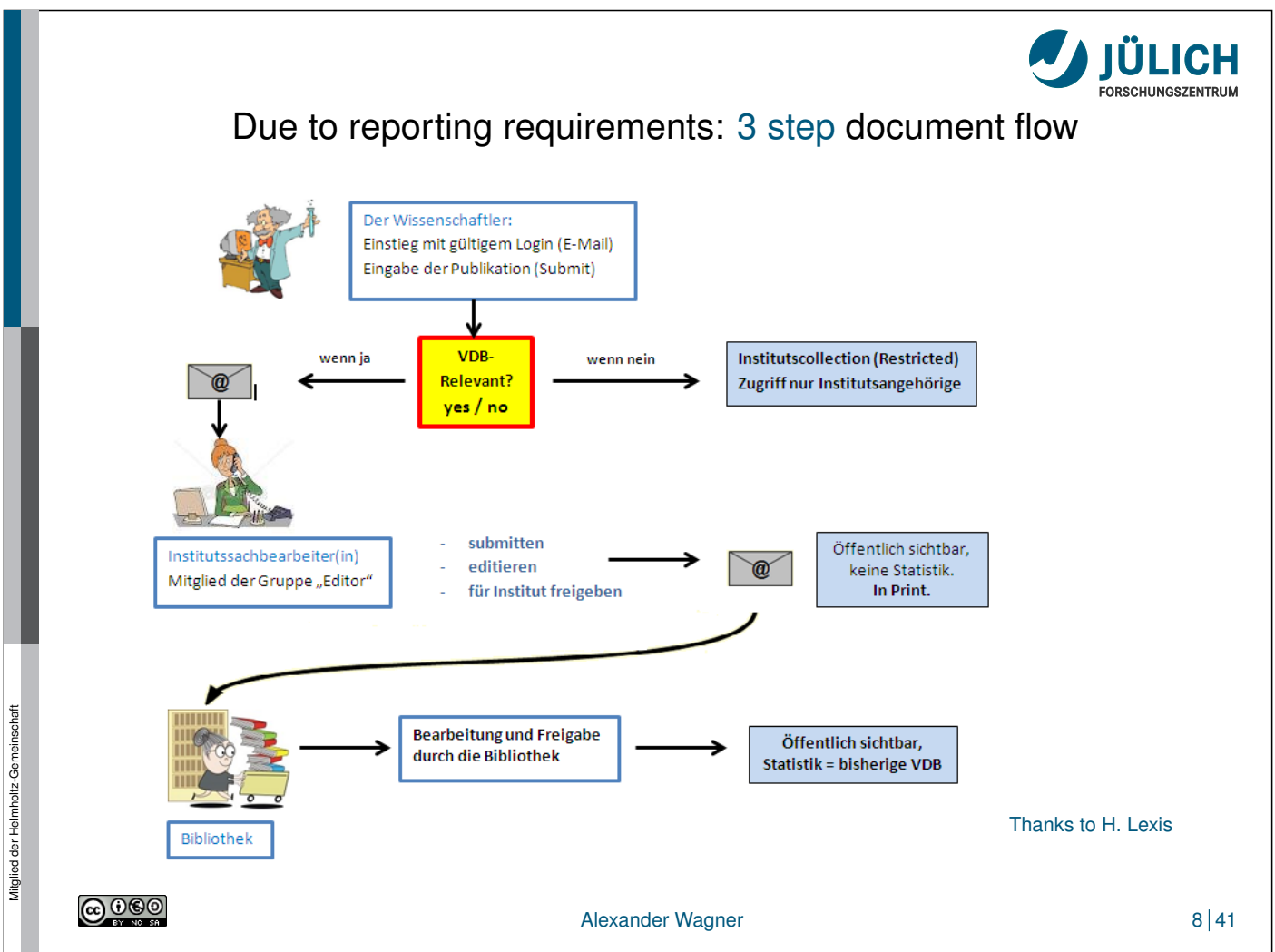
- Strukturelle und funktionelle Organisation des Gehirns (INM-1)
- Zentralbibliothek (ZB)

Research Program(s):

- 332 - Imaging the Living Brain (POF2-332) (Gesundheit)

Appears in the scientific report → 2012

Database coverage:



JuSER – Main features

- Import interfaces (improve data quality, ease up input)
- Exports to BibTeX/EndNote (Integrate with citations management)
- **Institute collections** (collect and share documents)
- Add full texts
- Normalize as much as possible (Key: Authorities)
 - **Authors** (tell apart Meier and Meier)
 - Institutes
 - Journals
 - Projects (POF, EU, ...)
 - ...

... Users don't need to care about technical details ...



Alexander Wagner

9 | 41

Collections

- **Publicationsdatabase**: Publications from Jülich
- **Documents in Print**:
 - approved by institutes editor
 - not approved by the library
 - e. g. papers just available Online (missing bibliographic data e. g. pages, volume)
 - already visible on institutes web pages
- OpenAccess repository **JUWEL**
- **Institute Collections**:
 - Institutes **private** workspace
 - e. g. drafts, collected literature, journal clubs ...
 - **Access for members of the institute only**
- **Authorities**



Alexander Wagner

10 | 41



Institute Collections

JuSER can hold documents **beyond own publications**

- **webbased** literature management
- document **exchange** at the institutes
- **centralized collection** of papers
- **easy export** to formatting tools (BibT_EX, EndNote)
- commenting (individual and in group)

- Institute collections require proper login
- Non-VDB relevant items do **not** show up on the webpage
- ZB does not care about usage

Invenio @ HGF

Part II: Basic Usage

June 17, 2013 | Alexander Wagner

Submit

- 1 Log in (LDAP based)
- 2 Select **Submit** from the main menu (► <http://juser.fz-juelich.de/submit>)
- 3 Select document type
- 4 Fill in the submit form
 - **import** if possible (doi, pmid, arXiv, inspire, own recs...)
 - take care of the authors
 - Own publication? ⇒ VDB-Relevant = yes
- 5 Add full text
- 6 Submit it:
 - “Finish & Release” (proceed to next level in workflow)
 - “Postpone” (generates TEMPENTRY)

Permissions & workflow: quite complex collection structure

Book Submit New Record page: 1 SUMM

Book

Import data ⓘ **VDB Relevant** ⓘ ☐ yes ☐ no Supported by FZJ-Employee ⓘ ☐ yes ☐ no

Institute(s) ⓘ **R&D Section** ⓘ

Grant Name ⓘ **Report Number** ⓘ

Author(s) ⓘ **Title** ⓘ

Publication Year ⓘ **Language** ⓘ

Conference name ⓘ **Acronym** ⓘ **City** ⓘ

Country ⓘ **Begin** ⓘ **End** ⓘ

ISBN ⓘ **Extent** ⓘ

Edition ⓘ **Publisher** ⓘ **Publisher's City** ⓘ **Series** ⓘ

Volume ⓘ

Abstract ⓘ

Additional Information ⓘ

Please upload your full text ⓘ

Finish & Release **Postpone**

Field designations

Red: mandantory for a full bibliographic description

Black: might not all apply

Blue: save manual work!

Import data

... is always available and allows to fetch data from external or internal sources.

PS: ⓘ is the manual (Wiki-based, german and english)

Import HowTo

- DOI: insert doi or dx.doi.org-url (e. g. 10.1016/j.physletb.2006.11.038)
-  pubmed : copy as displayed (e. g. PMID: 20923669)
-  arXiv.org : copy as displayed (e. g. arxiv:hep-ph/0610431)
- inspire: use URL
- own: recid: + record-Id or 037__a (e. g. recid:FZJ-2013-00499)
- ISBN: use the ISBN-field for this import

Journal Article

Import data  VDB Relevant 

Institute(s) 

Grant Name 

Import data:

Ultra-precision engineering: from physics to manufacturing / Jiang, X. J. ; Philosophical transactions of the Royal Society of London / A 370 3831 - 3834 ; London : Soc., 2012 ; 10.1098/rsta.2012.0178 ;



Alexander Wagner

17 | 41

Duplicate entries

Import data: 

Architecture and organizational principles of Broca's region / Amunts, Katrin ; Trends in cognitive sciences 16 418 - 426 ; Amsterdam [u.a.] : Elsevier Science, 2012 ; 10.1016/j.tics.2012.06.005 ;

Potential duplicate record(s):

-  130491

At import via doi, pmid, arXiv... JuSER

- can identify potential duplicates
- refuses the import
- shows links to the potential dupes



Alexander Wagner

18 | 41

E. g. DOI Import

Journal Article

Import data VDB Relevant ☒ yes ☐ no Supported by FZJ-Employee ☐ yes ☐ no

Institute(s) R&D Section

Grant Name

Author(s)

Jiang, X. J. (Extern) Corresponding author	✓	✕
Shore, P. (Extern) Author	✓	✕
McKeown, P. (Extern) Author	✓	✕
Whitehouse, D. J. (Extern) Author	✓	✕
Ruffles, P. C. (Extern) Author	✓	✕

Title Ultra-precision engineering: from physics to manufacturing

Journal Philosophical transactions of the Royal Society of London / A DOI 10.1098/rsta.2012.0

Volume 370 Issue 1973 Pages 3831 - 383

Publication Year 2012 Language

Conference name Acronym City

Country Begin End Publisher Soc. Publisher's City London

Abstract



Most red fields are filled in already!

Alexander Wagner

19 | 41

Authors

Upon import

authors are a mere **guesses** by the system. **Check them!**

- ✓: confirm the guess
- ✕: correct a wrong guess
- ✕: remove an entry
- 🧑: clear the whole list

- Associate only authors that show an **email address/institute**
- If unsure, leave them red



Alexander Wagner

20 | 41

Author input and association



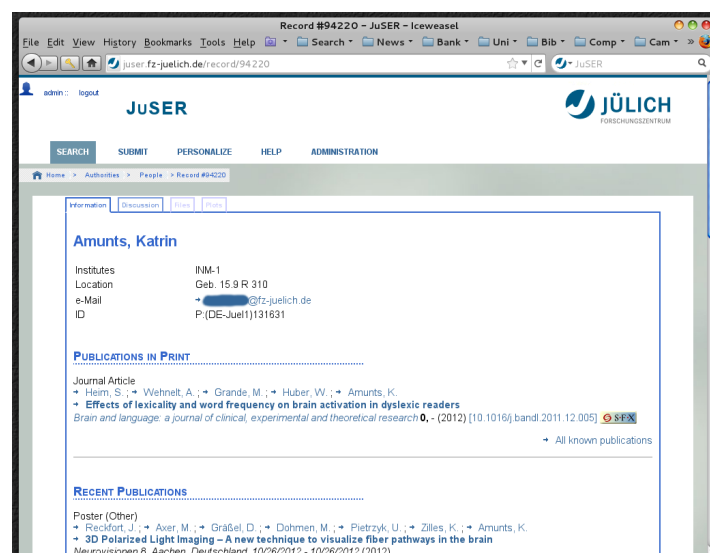
Note: ZB covers publication costs if the *corresponding author* is from Jülich.
Alexander Wagner

21 | 41

Own publications

Author association

allows easy extraction of **individual** and **exact** publication lists.



Alexander Wagner

22 | 41

Claim work (required only once!)

- 1 Log in
- 2 Navigate to: **Authorities / People** (<http://juser.fz-juelich.de/collection/People>)
- 3 Search for own name (e. g. 'Hofmann, D')
- 4 Note the IDs in question (e. g. P: (DE-Juel11)VDB63458 and P: (DE-Juel11)129471)
- 5 Open details and check the records found by the link
All known publications (below the list of recent publications)
- 6 notify juser@fz-juelich.de

(e. g. if the above two people are the same, we need both ids to join them)

To search own publications use aid: and own ID in quotes ("")

e. g.  aid: "P: (DE-Juel11)133794"

Want only first authorships? Use fai instead of aid.



Alexander Wagner

23 | 41

Invenio @ HGF

Part III: Installation

June 17, 2013 | Alexander Wagner

Repositories @ HGF

Roll out and disaster recovery

the **same code** on at least 5 different instances and **keep it consistent** by avoiding manual configuration.

- git: **cds-invenio** (centrally kept at DESY)
 - CERN git at “our” commit
 - used for base system
 - applied by `configure/make/make install`
- git: **hgf-invenio** (centrally kept at DESY)
 - overlay for our instances (our additions)
 - our patches (no replacements!)
 - directory structure like `/opt/invenio`
 - setup routines for roll-out (e. g. collections, roles, doctypes...)
 - instances configs



Alexander Wagner

25 | 41

InstallInvenio

InstallInvenio

sets up the **whole instance** at each partner in the proper layout including all **global** (hgf) and **local** configs

- assumes all deps are met and database exists
- sets up **from scratch** or to a **given point**
- compiles and installs **INVENIO**-src
- applies hgf patches and configs
- **applies global and local customizations**
- procedures based upon GUI functionality
- **no need** to mouse click in the GUI
- enables disaster recovery
- allows setup of an **identical** test environment



Alexander Wagner

26 | 41

Invenio @ HGF

Part IV: Backend

June 17, 2013 | Alexander Wagner

HGF “specifics”

Broad areas of research

some **INVENIO**-HEP-specifics just don't work

- Journal handling based on IDs, not names (hook up ZDB)
- allow for ingest of foreign data (e. g. ebook packages)
 - important in context of a **libraries catalogue**
 - stick to MARC definitions (Library of Congress)
 - avoid specialities (exceptions: 9xx)
 - implies adoption even of default defs (e. g. internal Marc, BibT_EX)
- Normalize all you can get
 - to be handled by users & librarians, not programmers
 - vitally important: Authorities

HGF-Modules

- **Importer:** (crossref, pubmed, GVK, arXiv.org, inspire...)
 - hook up with web services
 - reuse old code bascially in perl (to be rewritten to Python as time allows)
 - add new stuff in Python
- **HGFImport.py** (call externals but allow for permission checking)
- **HandleNames.py** (author guessing)
- **JSGetAllChildren.py** (recurse to the end of tree like MARC structures)
- **PubExporter.py** (handle specifics needed for web export)

Other functions

Always try to use **INVENIO** functions like alerts, baskets, ...
 Every unwritten line of code is a good line of code!



Alexander Wagner

29 | 41

Authorities: Types of Records

- **People** e. g. authors
 - **individualize(!)** local authors
 - entries are real people, not names (distinguish Wagner, A. and Wagner, A.)
 - connect to local institutes (incl. history)
 - allow interchange of records (e. g. Jülich with RWTH Aachen and MLZ)
- **Institutes** (people-like aggregates)
 - keep track of history (renaming, merging, splitting)
 - horizontal and vertical linkage
 - complex topography in names (e. g. loops)
- **Grants** (money-like aggregats)
 - horizoncal as well as vertical connections
 - different types (POF (HGF-specific), EU, DFG, ...)
 - flexibly extensible by **librarians**
 - visibility in websubmit depending on instance



Alexander Wagner

30 | 41

Authorities: Types of Records

- Periodicals, Journals
 - need ≈ 30.000 journals (links to german ZDB and EZB)
 - complex external tool to build up records
 - autogenerated/updated once a year (trigger: new JCR)
 - allow harvesting of records generated
 - contain statistics keys for evaluation
 - print and online edition are one journal (ISSN does not work as key)
- Statistics keys
 - database coverage (journal listed in Scopus or Web of Science, JCR or pubmed...)
 - OpenAccess designation
 - license identifier (Allianz-licensing)
- Vocabulary
 - allow for classifications (e. g. POF level)



Alexander Wagner

31 | 41

Frontend

- rich, browserbased GUI: (AJAX and CSS)
- JavaScript (jQuery + x)
- use few different elements (mainly tokeninput, autocomplete, datepicker)
- browser independent (needs to cover at least IE8+, Safari, FireFox on Win, Mac and Linux)
- Tokeninput (heavily extended and debugged(!) version of jquery.tokeninput)
 - institutes
 - authors
 - grants
- JSON-returns from the backend
- avoid program logic (backend has to deliver ready to use data)
- use encoding scheme for returns (allow processing in loops)
- define hidden subfields to keep structures (e. g. repeatable fields)
- invisibly enrich data from the backend (e. g. IDs, DDC, statistics keys...)



Alexander Wagner

32 | 41

New stuff for Websubmit

- repeatable fields (tokeninputs)
 - authors
 - grants
 - institutes
 - ...
- structured fields (mostly hidden)
 - Identifiers
 - statistics keys

Handling

Add hidden fields for all visible fields and do the fancy stuff there.
Pass on string encoded JSON structures.



Alexander Wagner

33 | 41

Backend

Expose records in JSON structures to JavaScript

- no repeatable subfields simplifies JSON
- define new output format: JS
- define format templates for each record type
 - use webbased frontend if possible
 - use pythonic BFE only for complex stuff
 - complex returns as ready to use text (i. e. escape substructures to strings)
- return hashes keyed as MARC:
 - 1 Field/Subfield: I245__a
 - 2 Structure: I536__
- innerrecordlinks: return from the backend (e. g. 536__ and 913__)



Alexander Wagner

34 | 41

Generate People Records

Use the local phone book (=LDAP, avoid privacy issues)

- LDAP harvesting (once a night, currently cron)
- unique IDs over time (email not suitable: limited TTL, recycling for common names)
- allow for arbitrary IDs
- allow for multiple IDs per entity (0247_ vs. 035__a)
- allow usage of IDs from several sources

Goal

Implement ORCID linkup with automagic registration/fetching



Alexander Wagner

35 | 41

Statistics

Use Marc 915__

Upon journal association: keys → bibliographic records

- Pythonic beancounters (use intbitsets)
- allow live Web query (if fast enough)
- allow complex queries, e. g.
 - output per institute and/or program
 - all JCR covered journal articles with external authors
 - all Scopus listed articles from John Smith (director of Inst X)
 - ...
- CLI is enough, but should be easily adoptable
- generate L^AT_EX-based reports
- generate structures outputs (for spreadsheet, literature management)



Alexander Wagner

36 | 41

Statistics/2: Bibliometrics

Mainly in Jülich: JuSER is source for bibliometric evaluations

- link up with  Web of Science (just another ID $\Rightarrow$ 0247_)
- semiautomatic  Web of Science-ID assignments (doi2ut)
- specific exports
- planned: direct link to workflow tools

In principle

... all queries can be done in the Webfrontend



Alexander Wagner

37 | 41

OpenAccess

- Acquire more fulltexts
- At least give a stable link (e. g.  pubmed-Central, OA-Journals)
 - Use statistics key for OA-Journals (for DOAJ-listed journals)
 - enrich upon ingest
- Implement “Allianz-Licensing” (ingest publishers versions)
- Make OA-Articles more prominent (e. g. logo, collection)
- OpenAIRE compliance
- OAI-Harvesting for BASE and friends



Alexander Wagner

38 | 41

Next Steps

- Upgrade from 1.0 to 1.1 or, more likely, 1.2 (fix OAI Server!)
- Feed Google Scholar (how?)
- Get pending instances online:
 - RWTH Aachen
 - MLZ
- At Jülich: implement crossharvesting (RWTH ↔ FZJ ↔ MLZ)
- Add workflow for “vita” (workflows for special collections)

Continually refactor/check code to get ready to give back



Alexander Wagner

39 | 41

Project partners



Deutsches Elektronensynchrotron, Zentralbibliothek



Forschungszentrum Jülich, Zentralbibliothek



GSI Helmholtzzentrum für Schwerionenforschung, Bibliothek + Kern-IT



RWTH Aachen, Hochschulbibliothek



Maier-Leibniz-Zentrum, Garching



Alexander Wagner

40 | 41

Questions?



Alexander Wagner
Zentralbibliothek

Fachinformation,
wiss. Publizieren

Tel.: 1586
a.wagner@fz-juelich.de

<http://www.fz-juelich.de/zb>

This document is available as ■ FZJ-2013-499



Alexander Wagner

Typeset by pdfL^AT_EX 41 | 41