THE HELMHOLTZ INVENIO REPOSITORY PROJECT :

BETWEEN ORGANIZATIONAL TASKS, NEEDS OF SCIENTISTS, AND CONSTRAINTS

Structure of the presentation

- Libraries of DESY, FZJ and GSI in the Helmholtz Association
- Needs & organisational tasks
- Helmholtz-INVENIO project as the selected solution
- Constraints
- Opportunities in regard to recent theses and discussions to information infrastructure

Libraries of DESY, FZJ and GSI in the Helmholtz Association

3 of 18 libraries in the Helmholtz association started a project with the Aachen university library to the field of publications database

Deutsches Elektronensynchrotron, Zentralbibliothek	≈ 2000 + 3000	
Forschungszentrum Jülich, Zentralbibliothek	≈ 5000 + 1000	
GSI Helmholtzzentrum für Schwerionenforschung, Biblitohek + Kern-IT	≈ 1050	
Maier-Leibniz-Zentrum, Garching	≈ 300	
RWTH Aachen, Hochschulbibliothek	≈ 9000	
Museum Zitadelle Jülich		
Institut für Experimentelle Kernphysik, Karlsruhe		

Needs & organisational tasks

- **publication database** as portal for the scientific output of the organisation
- portal of **open access** publications (Berlin declaration)
- portal of (small) **open data and supplementary material**
- portal for **multimedia content**
- integration of **own publishing house** (e.g.digital born, digitized old material)
- OAI-PMH interface
  - delivery to the Helmholtz central database
  - delivery to the national library
  - data exchange between partners
- [OPAC only GSI]
- **flexible and different kind of evaluations possibilties**
- ...

Helmholtz-INVENIO project as the selected solution

Principles

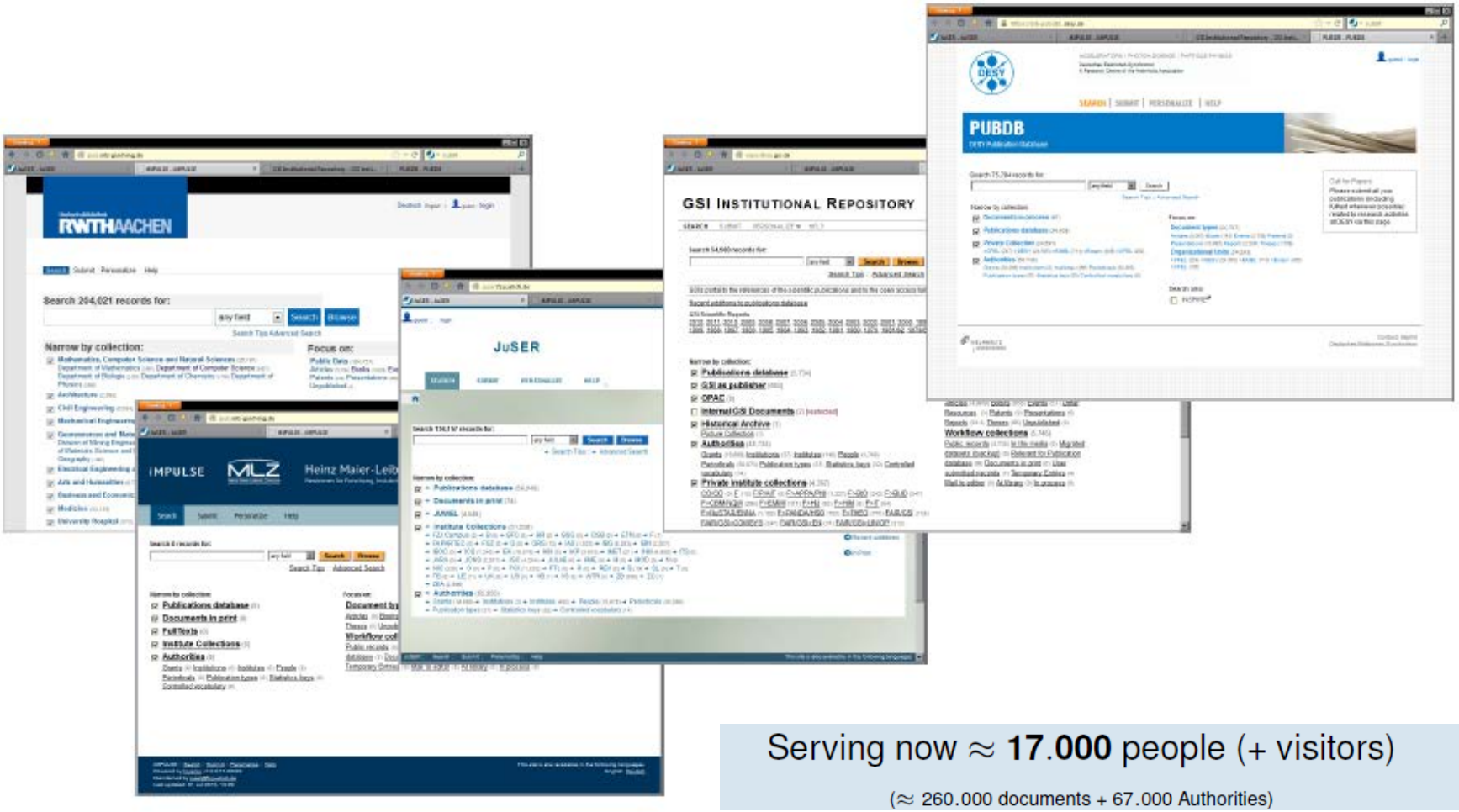
- co-operation between the libraries
- adopt an existing open system : INVENIO (CERN)
- same installation with different config files and styles

Helmholtz-INVENIO developments

- MARC21 authorities as core concept for different needs
- easy input/ingestion (arXiv, inspire, pubmed...)
- integration of workflows (scientists -> institute‘s editor -> library)
- consent to institutes, subject & grant assignments
- automatic statistics for the main evaluation tasks
- common infrastructure (git, authorities, ...)

  ...

Running systems



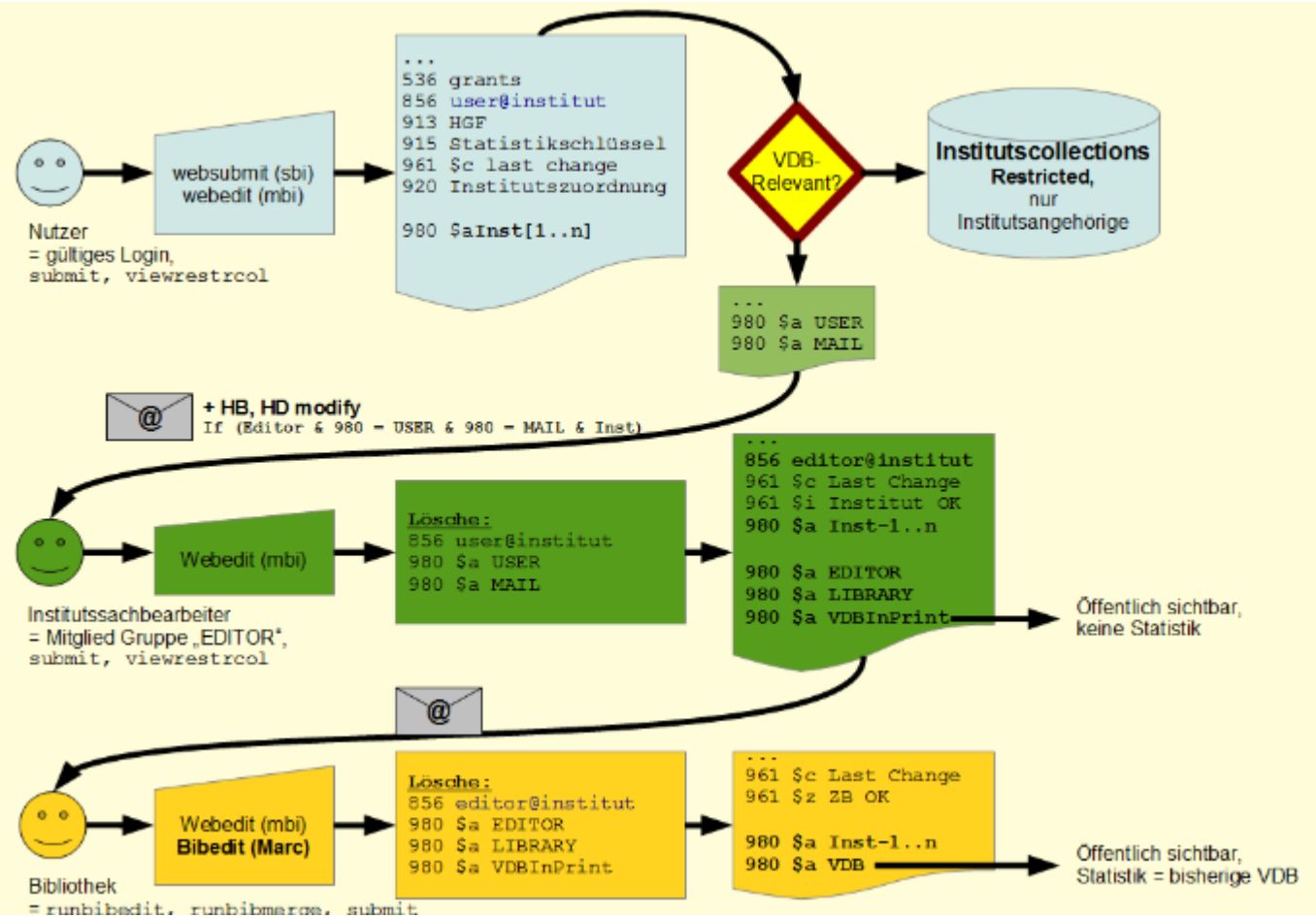
Serving now ≈ **17.000** people (+ visitors)

(≈ 260.000 documents + 67.000 Authorities)

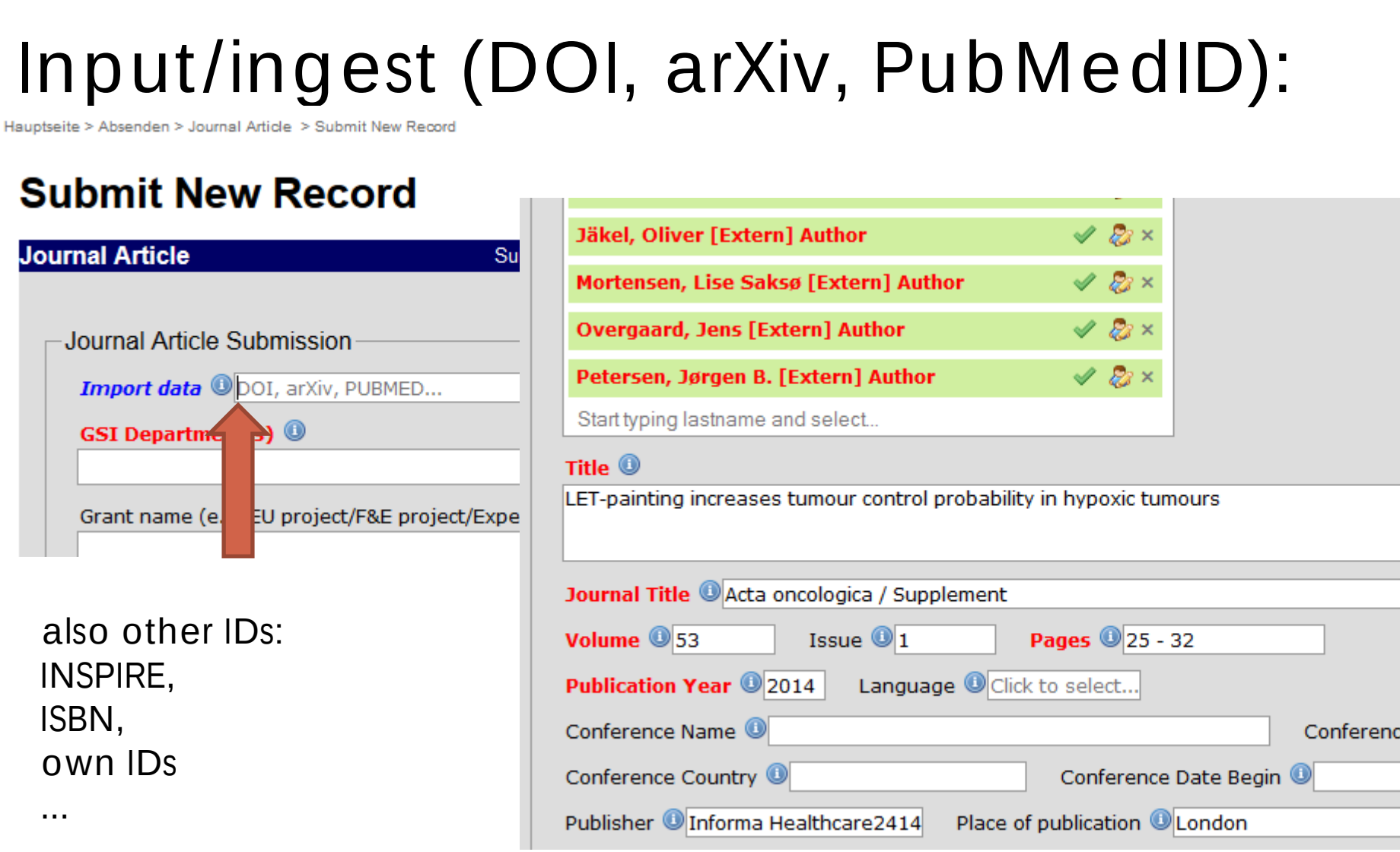
Tools used

- Workflow
  - Webbaskets (e. g. revision lists)
  - Alerts (e. g. revision lists)
  - Collections (e. g. private for institutes)
  - Webmessage (e. g. correction requests)
- Authority records (almost everywhere)
- OAI-PMH (authority exchange)
- High-level API (setup: e. g. collections, roles, groups, baskets. . . ; **no db-dump** sharing)
- jQuery/jQueryUI (websubmit)
- intbitsets (e. g. statistics)

Workflow:


...
536 grants
856 user@institut
913 HGF
915 Statistikschlüssel
961 $c last change
920 Institutszuordnung
980 $aInst[1..n]
VDB-Relevant?
Institutscollections
Restricted,
nur
Institutsangehörige
...
980 $a USER
980 $a MAIL
websubmit (sbi)
webedit (mbi)
Nutzer
= gültiges Login,
submit, viewrestrcol
+ HB, HD modify
If (Editor & 980 = USER & 980 = MAIL & Inst)
Webedit (mbi)
Institutssachbearbeiter
= Mitglied Gruppe „EDITOR“,
submit, viewrestrcol
Lösche:
856 user@institut
980 $a USER
980 $a MAIL
...
856 editor@institut
961 $c Last Change
961 $i Institut OK
980 $a Inst-1..n
980 $a EDITOR
980 $a LIBRARY
980 $a VDBInPrint
Öffentlich sichtbar,
keine Statistik
Webedit (mbi)
Bibedit (Marc)
Bibliothek
= runbibedit, runbibmerge, submit
Lösche:
856 editor@institut
980 $a EDITOR
980 $a LIBRARY
980 $a VDBInPrint
...
961 $c Last Change
961 $z ZB OK
980 $a Inst-1..n
980 $a VDB
Öffentlich sichtbar,
Statistik = bisherige VDB

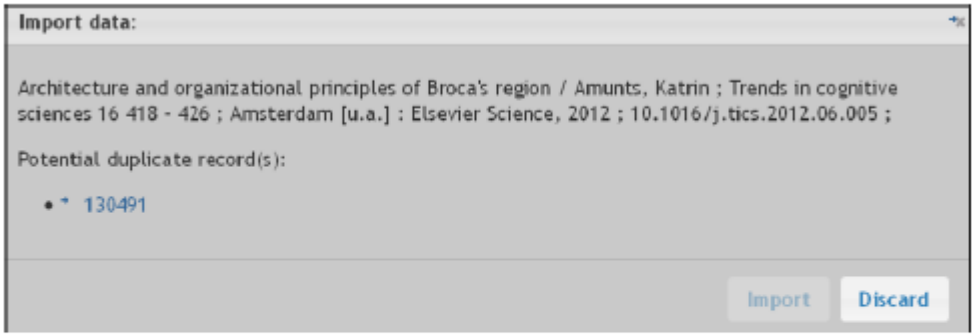
Input/ingest (DOI, arXiv, PubMedID):



also other IDs:
INSPIRE,
ISBN,
own IDs
...

Input/ingest (DOI, arXiv, PubMedID):

Duplicate entries



At import via doi, pmid, arXiv…

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

MARC21 authorities

- People
- Institutes
- Organisations
- Grants
- Experiments / Proposals
- Journals
- Statistic keys
- Types of OA

...

Why Authority Control?

- Scientific reporting
- Evaluations / Bibliometrics (link up with WoS/Scopus/Inspire)
- Publication lists (e. g. on the WWW: institutes, people, projects)

All this *needs precise answers* to complex questions:

Normalize as much as possible: in libraries = Authority Records

Example: History of an institutes name at Jülich

- till 07/31/2002: ICG-4 (Erdöl und Geochemie)
- till 12/31/2006: ICG-IV (Agrosphäre)
- till 10/31/2010: ICG-4 (Agrosphäre)
- since: IBG-3 (Agrosphäre)

MARC21 authorities

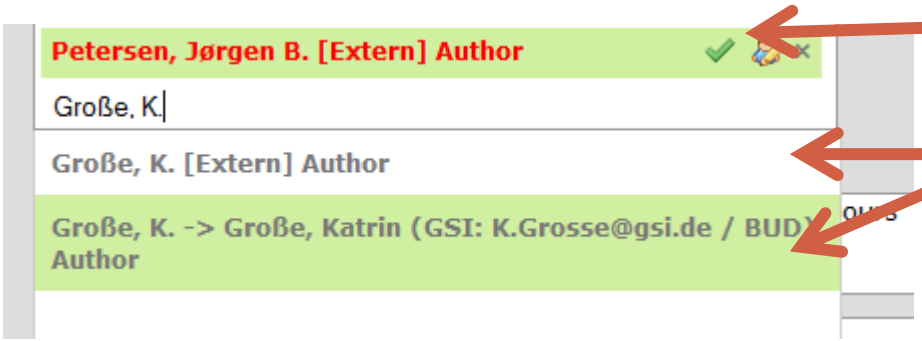
Authority Records: Requirements

1. n:m relations (e. g. splitting of institutes, grants with several top levels)
2. Tracking history (predecessors and successors)
3. Tracking hierarchy (dad/son relations; we could have several parents!)
4. Several identifiers (e. g. DOI + Handle + URN or ORCiD + other IDs)
5. Multiple types of authorities (e. g. people, grants, institutes, journals. . .)

Solved by MARC Authority

1. repeatable linking fields (MARC Authority 4xx, 5xx)
2. horizontal linking ($w control subfield: $wa, $wb)
3. vertical linking ($w control subfield: $wt)
4. 0247_ with $2 = source and/vs. 035__ (we do not repeat 035)
5. ignorant of the type but specify source (usually: $0 or $a and $2 subfields)

People authorities connections



confirmation

automatic suggestions


```
</datafield>
– <datafield tag="100" ind1="1" ind2=" ">
    <subfield code="a">Große, Katrin</subfield>
    <subfield code="0">P:(DE-Ds200)OR0387</
    <subfield code="b">0</subfield>
    <subfield code="u">gsi</subfield>
    <subfield code="e">Corresponding author</s
  </datafield>
```

```
– <datafield tag="700" ind1="1" ind2=" ">
    <subfield code="a">Wagner, A.</subfield>
    <subfield code="b">1</subfield>
    <subfield code="u">FZJ</subfield>
    <subfield code="0">P:(DE-Juel1)133832</subfield>
  </datafield>
– <datafield tag="700" ind1="1" ind2=" ">
    <subfield code="a">Barake, L.</subfield>
    <subfield code="b">2</subfield>
    <subfield code="0">P:(DE-HGF)0</subfield>
  </datafield>
– <datafield tag="700" ind1="1" ind2=" ">
    <subfield code="a">Diallo, A. S.</subfield>
    <subfield code="b">3</subfield>
    <subfield code="0">P:(DE-HGF)0</subfield>
  </datafield>
```

People authorities

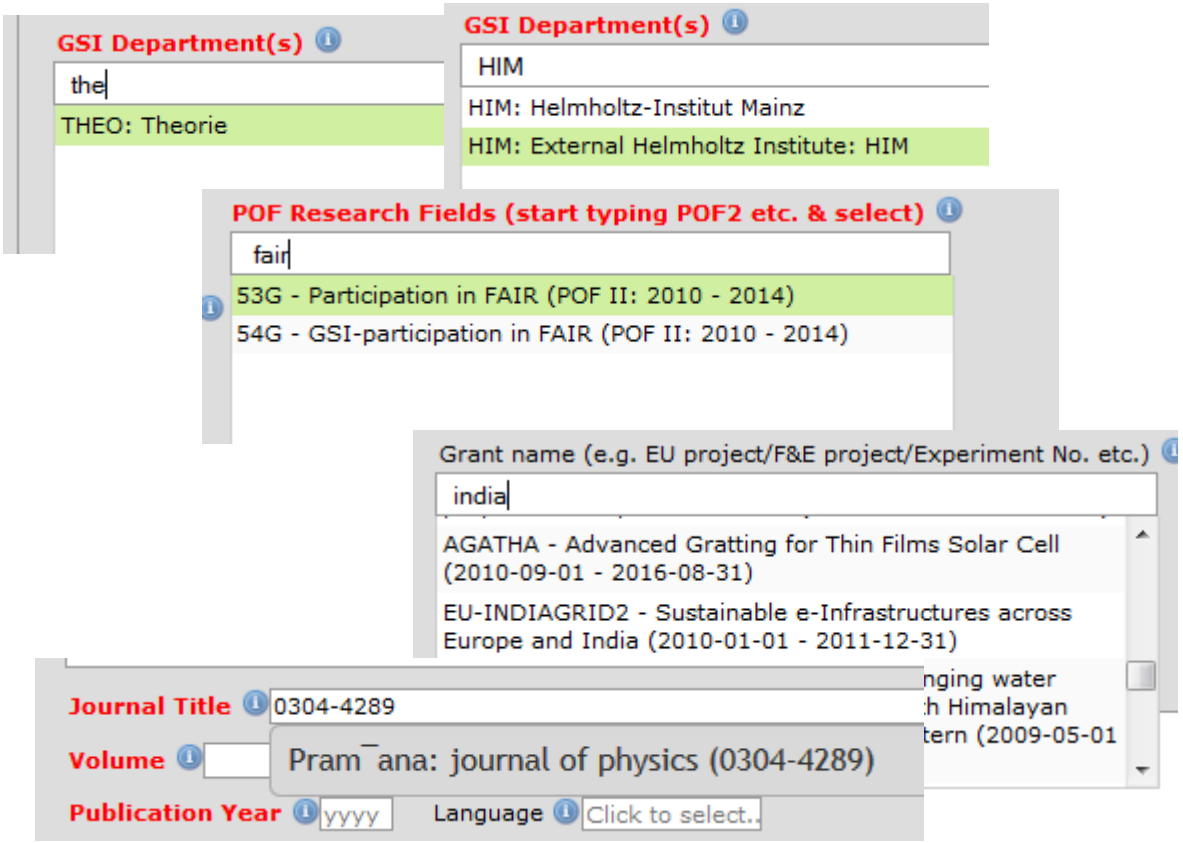
```
- <collection>
  - <record>
      <controlfield tag="001">45795</controlfield>
    - <datafield tag="024" ind1="7" ind2=" ">
        <subfield code="a">P:(DE-Ds200)OR0387</subfield>
        <subfield code="2">P:(DE-Ds200)</subfield>
      </datafield>
    - <datafield tag="035" ind1=" " ind2=" ">
        <subfield code="a">P:(DE-Ds200)OR0387</subfield>
      </datafield>
    - <datafield tag="100" ind1="1" ind2=" ">
        <subfield code="a">Große, Katrin</subfield>
      </datafield>
    - <datafield tag="371" ind1=" " ind2=" ">
        <subfield code="0">I:(DE-Ds200)20120319OR030</subfield>
        <subfield code="c">BUD</subfield>
        <subfield code="m">K.Grosse@gsi.de</subfield>
        <subfield code="v">GSI-OR-IdM</subfield>
      </datafield>
    - <datafield tag="373" ind1=" " ind2=" ">
        <subfield code="a">BUD</subfield>
        <subfield code="0">I:(DE-Ds200)20120319OR030</subfield>
        <subfield code="2">I:(DE-Ds200)</subfield>
      </datafield>
    - <datafield tag="400" ind1="1" ind2=" ">
        <subfield code="a">Grosse, Katrin</subfield>
      </datafield>
    - <datafield tag="980" ind1=" " ind2=" ">
        <subfield code="a">P</subfield>
      </datafield>
    - <datafield tag="980" ind1=" " ind2=" ">
        <subfield code="a">AUTHORITY</subfield>
      </datafield>
  </record>
</collection>
```

ORCID

- Handle multiple IDs

```
001__ 95749
0247_ $aP:(DE-Juel1)133832$2P:(DE-Juel1)
0247_ $a0000-0001-9846-5516$2ORCID
0247_ $aI-3159-2013$2ResearcherID
035__ $aP:(DE-Juel1)133832
1001_ $aWagner, Alexander$gmale
...
750_7 $0V:(DE-HGF)8$2HGFVOC$aAuthorized$d2013-11-09
980__ $aP
980__ $aAUTHORITY
```

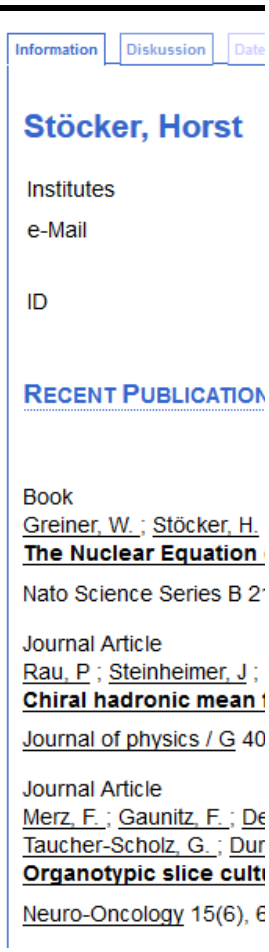
Institutes, Organisations, Grants, Journals

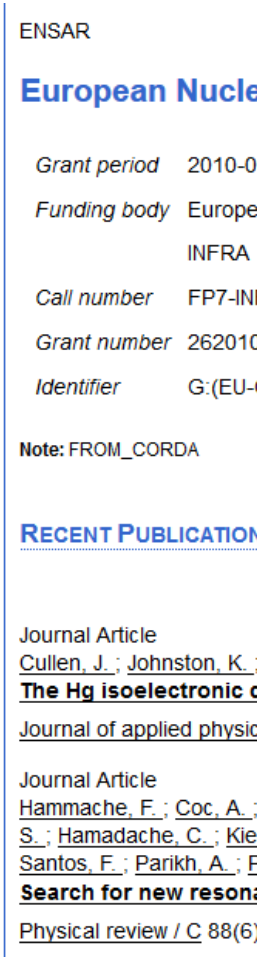



```
- <collection>
  - <record>
      <controlfield tag="001">44524</controlfield>
    - <datafield tag="024" ind1="7" ind2=" ">
        <subfield code="a">I:(DE-Ds200)20120319OR028</subfield>
        <subfield code="2">I:(DE-Ds200)</subfield>
      </datafield>
    - <datafield tag="035" ind1=" " ind2=" ">
        <subfield code="a">I:(DE-Ds200)20120319OR028</subfield>
      </datafield>
    - <datafield tag="110" ind1="1" ind2=" ">
        <subfield code="a">Theorie</subfield>
        <subfield code="g">Abteilung</subfield>
      </datafield>
    - <datafield tag="410" ind1="1" ind2=" ">
        <subfield code="a">THEO</subfield>
        <subfield code="w">d</subfield>
      </datafield>
    - <datafield tag="410" ind1="1" ind2=" ">
        <subfield code="a">Theory</subfield>
      </datafield>
    - <datafield tag="510" ind1="1" ind2=" ">
        <subfield code="a">Forschung</subfield>
        <subfield code="0">I:(DE-Ds200)20120123OR019</subfield>
        <subfield code="2">I:(DE-Ds200)</subfield>
        <subfield code="w">t</subfield>
      </datafield>
    - <datafield tag="961" ind1=" " ind2=" ">
        <subfield code="x">2005-12-14T20:57:44</subfield>
        <subfield code="c">2012-03-19T18:37:47</subfield>
      </datafield>
    - <datafield tag="980" ind1=" " ind2=" ">
        <subfield code="a">I</subfield>
      </datafield>
    - <datafield tag="980" ind1=" " ind2=" ">
        <subfield code="a">AUTHORITY</subfield>
      </datafield>
  </record>
</collection>
```


...

Because of ... web integration of lists





Because of ... further needs of scientists

- Web 2.0 functionalities e.g. commenting, recommendations, Alerts/RSS
- literature management (Endnote, BibTeX, ...)
- public and internal baskets e.g. for introductory works or working groups

- easy internal work spaces for collaborative works in the departments 

VDB-relevant no -> internal record in the department‘s collection with access only for department‘s staff
- multimedia possibilities e.g. conference photographs, videos ...
- delivery to EU grants (OpenAIRE)

Limits & Constraints

- The introduction of the system with complete new processes are more difficult than in research centers the scientists are used to do so
- data privacy: All persons‘ authorities are visible for staff only. (ORCID is planned)
- copyright laws: as professional librarian the introduction of group workspace has to be accompanied
- critics at the IT corner: stop normalizing, do more in search engines works (but there is the organizational demand for precise answers)
- resources at our libraries - need for co-operation
- limits for big and international collaborations

rather organizational limits and resources‘ constraints

Opportunities in regard to recent theses and discussions to information infrastructure in Germany

There are many political documents to structure, to relations of different kinds of information organizations and of course to copyright, and open access/open data.

- Tochterman/ZBW: Ten theses regarding the future of scientific information infrastructure institutions http://www.zbw.eu/e_news/e_2013-10-theses.pdf
- CRIS discussions (papers in progress)

Discussions to Current Research Informations Systems

Information to scientists, expertise, contacts, publications, projects, funding, patents, co-operations, projects, EU-projects, collaborations,, research data, ORCID, e-learning...
Metadata quality control, ingest for different organisational units,
connection of different information,
standardisation of evaluation,
...

Authorities as usable core

e.g. automatic evaluation statistics

```
Analyzing collection "VDB" in WEB year "2012":
6132 records in collection "VDB"
508 records in collection "VDB" for WEB year "2012"
wherein
385 records JCR refereed - 326 with external authors
1 records SCI/SCIe/AHCI/SSCI refereed (WOSnonJCR) - 1 with
386 records JCR or SCI/SCIe/AHCI/SSCI refereed (WOS) - 327
386 records JCR or SCI/SCIe/AHCI/SSCI refereed or with UT c
43 records otherwise refereed - 36 with external authors
0 habilitations
48 dissertations
0 master theses
0 bachelor theses
6 patents
438 journal articles - 368 with external authors
2 proceedings - 2 with external authors
2 contributions to proceedings - 2 with external authors
4 books - 3 with external authors
5 contributions to books - 4 with external authors
407 total documents with external authors

Statistics for main POF programs and refereed status:
Nr.          JCR  WOSnonJCR        WOS     WOS_UT      Othe
310            2           0          2          2
530          206           0        206        206
540          177           1        178        178         3

Statistics for all POF programs and refereed status:
(multiple counting of publications!)
Nr.          JCR  WOSnonJCR        WOS     WOS_UT      Othe
310            3           0          3          3
530          206           0        206        206
```

Tochterman/ZBW: Ten theses regarding the future of scientific information infrastructure institutions

‚8. Libraries offer supporting services for publishing:
Libraries will no longer exclusively act as information providers. Instead they will offer additional services (e.g. infrastructures for research data) that will support researchers in their publishing processes.'

To implement as first step complete publications processes into Helmholtz-INVENIO (working in group, release by the organization, open access postprints, supplementary materials, data...)

‚9. Conceptual connections add value to library catalogues: [...] *The modelling of conceptual connections, which consider scientific publications as a semantic composition of their components, will replace the cataloguing used in libraries today.*'

‚5. Future publications are complex, cross-medial and interconnected'

E.g. videos and e-learning material, photographs could be offered in Helmholtz-INVENIO (as already done at CERN-CDS).

Developing & building expertise within the Helmholtz-INVENIO project (limits for specialisted libraries)
'10. Libraries have high IT competence and/or high media competence'
‚2. Inhouse research raises the innovation level and the customer orientation of libraries'

'4. Content comes to the researcher'
‚3. Libraries provide crucial support for decentralised information provision'
Some things are done: website-integration of automatics lists, RSS
but more could/should be done to ORCID or to social media ...
(resource problem)

‘1. The traditional mission of libraries will remain; at the same time libraries will strongly internationalise’

INVENIO user community, INSPIRE ...
The Helmholtz-INVENIO project is interested in development of ideas and exchange of information with other specialized libraries.

Thank you!

- *Martin Köhler*[a]
- Zaven Akopov[a,b]
- *Tomasz Pazera*[a]
- Katrin Große[c]
- *Stefan Hesselbach*[d]
- Bernhard Mittermaier[e]
- Anna Fründ[e]
- Heike Lexis[e]
- *Cornelia Plott*[e]
- Christoph Holzke[e]
- *Alexander Wagner*[e]
- Jürgen Neuhaus[f]
- Connie Hesse[f]
- *Björn Pedersen*[f]
- Ulrike Eich[g]
- *Louai Barake*[g]
- *Abdoulaye Diallo*[g]
- Roland Rappmann[g]
- Dominik Schmitz[g]
- Edmund Wollgarten[g]

[a] DESY Library and Documentation; [b] Project Inspire; [c] GSI Library; [d] GSI Core IT;

[e] Forschungszentrum Jülich, Zentralbibliothek; [f] MLZ, Garching; [g] RWTH Aachen, Hochschulbibliothek