Mitglied der Helmholtz-Gemeinschaft

Invenio@HGF – status and perspectives

[sic!] Jülich – 2nd Invenio User Group Workshop

Jülich, 18. November 2013 | Alexander Wagner, for the Collaboration

Overview

- Partner
- Initial TODO
- Accomplishments
- Lessons learned
- Project group

Project Partners

	Deutsches Elektronensynchrotron, Zentralbibliothek	$\approx 2000 + 3000$
	Forschungszentrum Jülich, Zentralbibliothek	$\approx 5000 + 1000$
	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	

Open for new Partners!

Serving now $\approx$ **17.000** people (+ visitors)

($\approx$ 260.000 documents + 67.000 Authorities)

Starting out

Goal

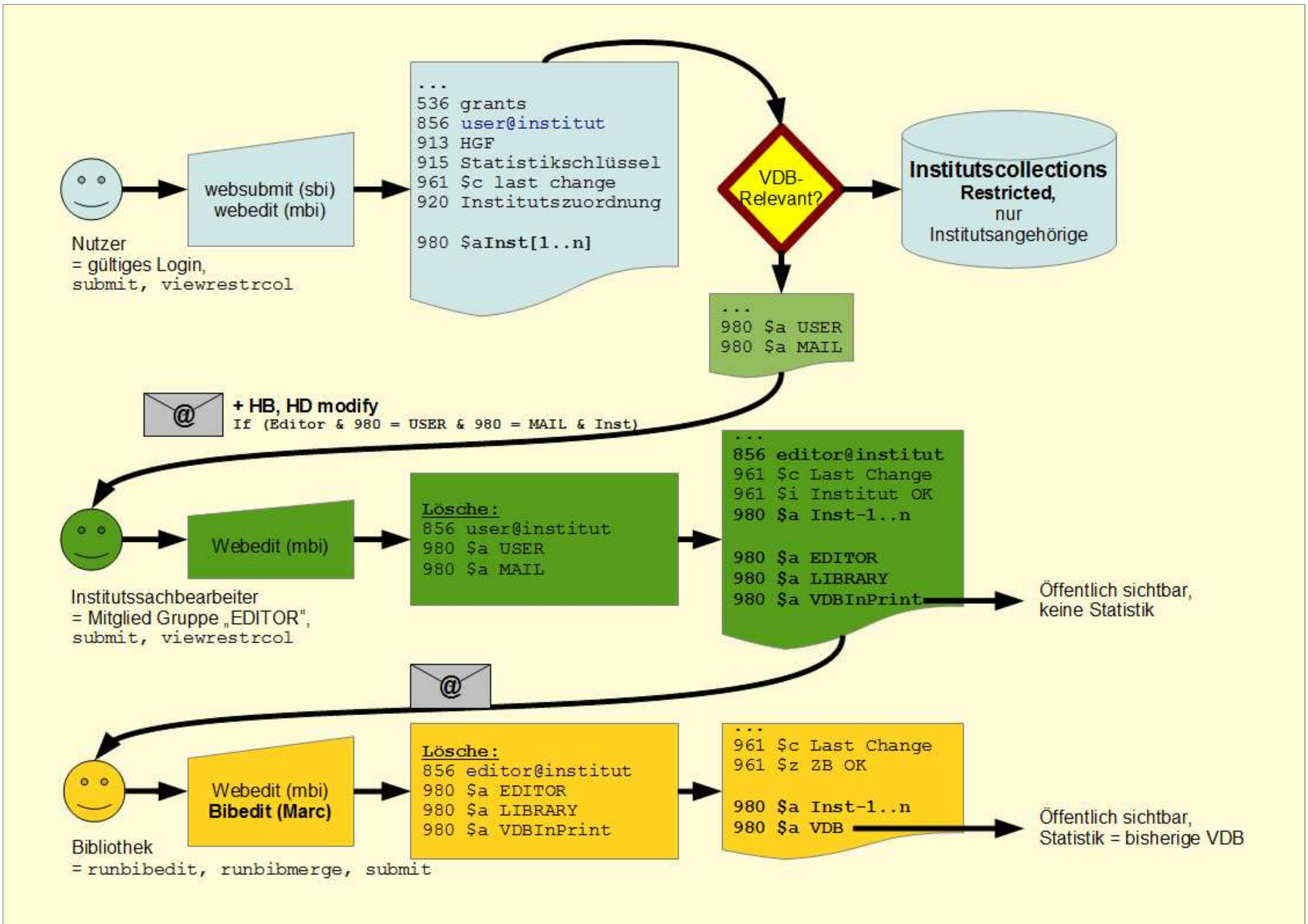
Replace existing systems, at GSI build up from scratch.
User-centric design (users aka scientists)

1. "Learn Invenio" (thanks to CERN ☺)
2. Define wording. . . (different insitutions!)
3. Build infrastructure: git and friends
4. Build more infrastructure: authorities and friends
5. Build a **deployment scheme**: InstallInvenio and friends

We need to roll out 10+ instances
with different data sets and keep them consistent on code level

Initial ToDo

- Design the system around web based literature management
- Design a document workflow



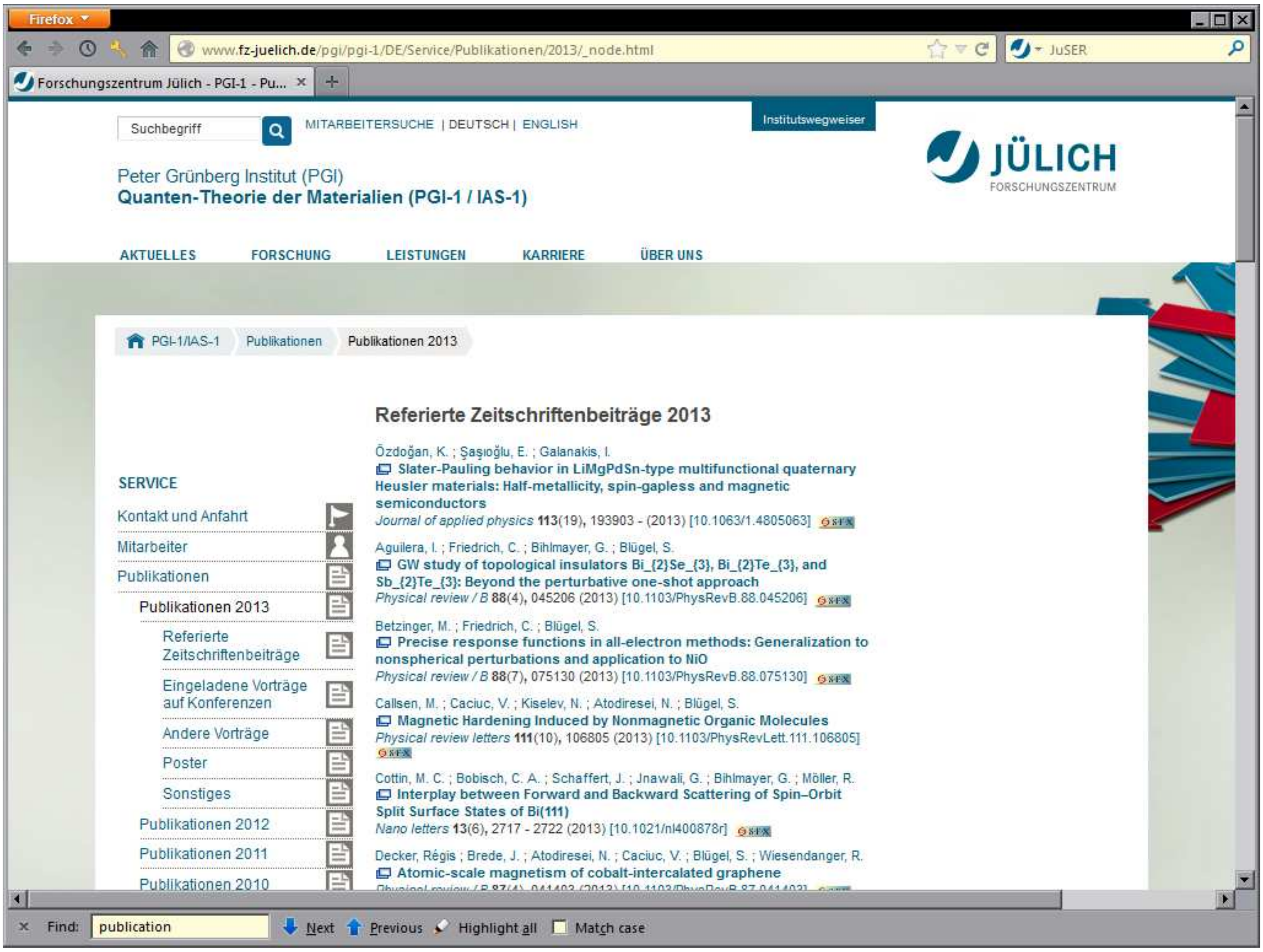
Initial ToDo

- Design the system around web based literature management
- Design document workflow (3 steps with privilege escalation)
- Design easy ingestion workflow (websubmit, imports, author disambiguation)

Finally we wrote some code. . .

Every unwritten line is a good line, still: ≈ 55.000 lines

- Migrate old data (various, proprietary sources)
- Train the inputters and users (secretaries, scientists, librarians)
- Hook up with content management system(s) (visibility!)



Initial ToDo

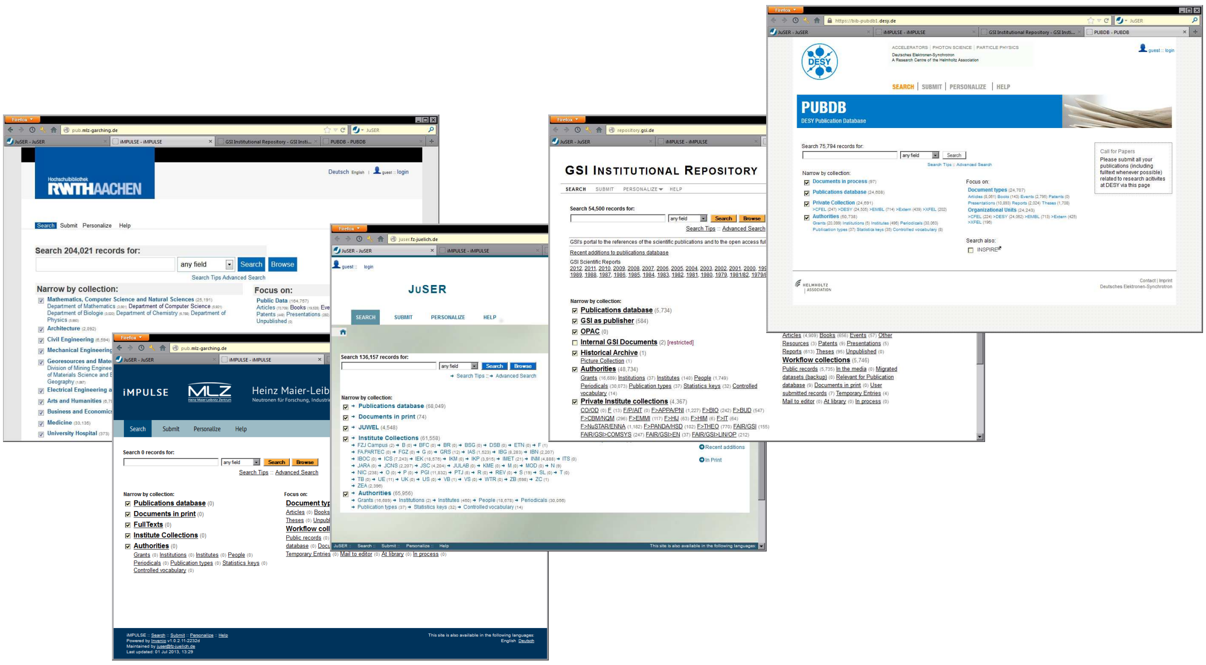
- Design a document workflow (3 steps with privilege escalation)
- Establish easy ingestion workflow (websubmit, imports, author disambiguation)

Finally we wrote some code. . .

Every unwritten line is a good line, still: ≈ 55.000 lines

- Migrate old data (various, proprietary sources)
- Train the inputters and users (secretaries, scientists, librarians)
- Hook up with content management system(s) (visibility!)
- Derive necessary reporting (details are way beyond this intro)
- Get it up and running (First Light: 11/19/2012)

Accomplishments and status



Accomplishments and status

- All partners have running systems (roll out works)
- Almost all partners are online
- Rich websubmit (including **repeatable field** handling)
- Importer routines (doi, pmid, arXiv, inspire, ISBN, own recs,... in **websubmit**)
- Authorities
  - Generate ($\approx$ 67.000 recs)
  - Use (e. g. JSON returns, statistics...)
  - Share (MarcXML OAI-PMH)
- Implement
  - Author identification (ORCiD ready!)
  - Output formats (JSON, BibTEX, EndNote...)
  - Reporting (publication statistics)
  - Delivery to content management systems

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)

Lessons learned / Next steps

- CERN is way to fast to keep up with
- Never use Dublin Core again (complex migration, to few data fields. . .)
- All libraries are the same ☺

- Upgrade to 1.2: get OAI-Server fixed!

However. . .

In our use case switching of the base system is non-trivial

(Remember: 10+ instances. . .)

- Open up for new partners
- Clean up our code and give it back

Contributors

- *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

Further “reading”

- Invenio @ HGF - Technical background

  Talk at Invenio Developer Forum

- Collaborative tools for an institutional repository

  Talk at Helmholtz OA Workshop

- JuSER – Publications Database

  Introductory course at Jülich

- JuSER - Autorenhandling

  Talk at HGF-ORCiD Meeting, Berlin (in german)

Thanks!

Alexander Wagner
Zentralbibliothek

Scientific Services /
Scientific Publishing

Tel.: +49-2461-61-1586
a.wagner@fz-juelich.de

This document is available as FZJ-2013-05410

Typeset by pdfLaTeX