JOIN²

A Publication Database and Repository based on Invenio

Alexander Wagner
Budva Bečići, 02.10.2019

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

DESY.

Overview

- Motivation and project layout
- Main features
- What you do not see
- Publishing your work
- Data publications

Motivation

1. Visibility (on the Web, in search engines, OpenAIRE, BASE…)
2. Web integration (integrate publications with institutes pages)
3. OpenAccess (increases visibility and indexing)
4. Document collections (including private areas)
5. Publication lists (on the Web, in reports, for PR)
6. Search & Browse (via database homepage)
7. Reporting (Institutional, EU, other funders…)

Authority control (librarians slang)

- unify and normalize spelling (still: names are but sound and smoke…)
- allow for “exact matches” (based on unique IDs)
- handle different languages and scripts (UTF-8 and LaTeX)

JOIN^2 collaboration

Deutsches Elektronen-Synchrotron, DESY	$\approx 2300 + 3000$	
Deutsches Krebsforschungszentrum	≈ 3000	(internal access only)
Maier-Leibnitz-Zentrum, Garching	$\approx 300 + 1000$	
RWTH Aachen	≈ 9500	
Museum Zitadelle Jülich	$\approx 3 + n$	

JOIN^2: proud to serve

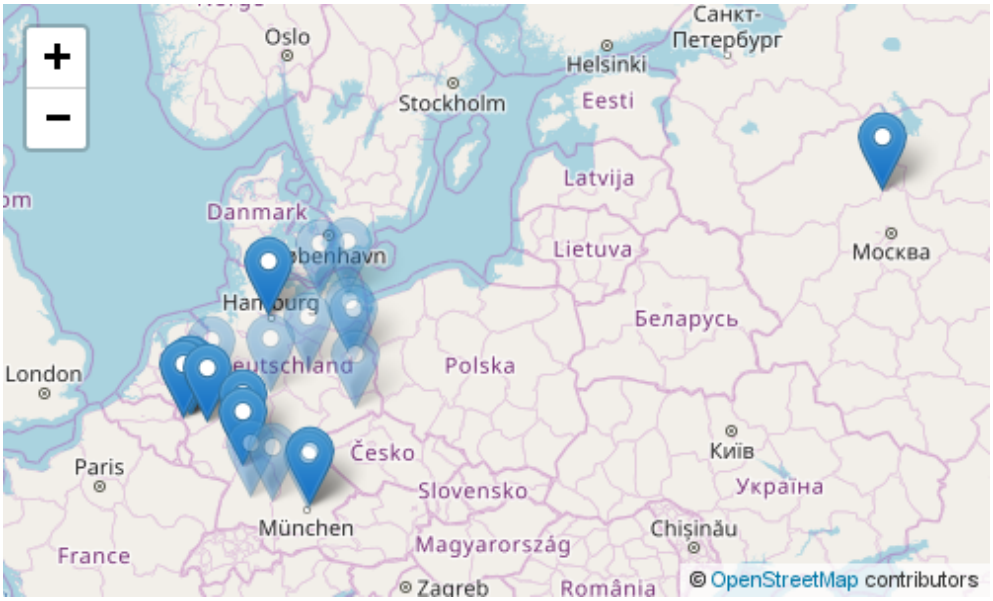
9 institutions,

$>$ **28.000** staff members

$>$ 6000 visiting scientists

(≈ 475.000 records, ≈ 68.000 OpenAccess, ≈ 135.000 shared authorities)

Put it on the map…



Current challenge

Cyrillic script

Wagner, Alexander

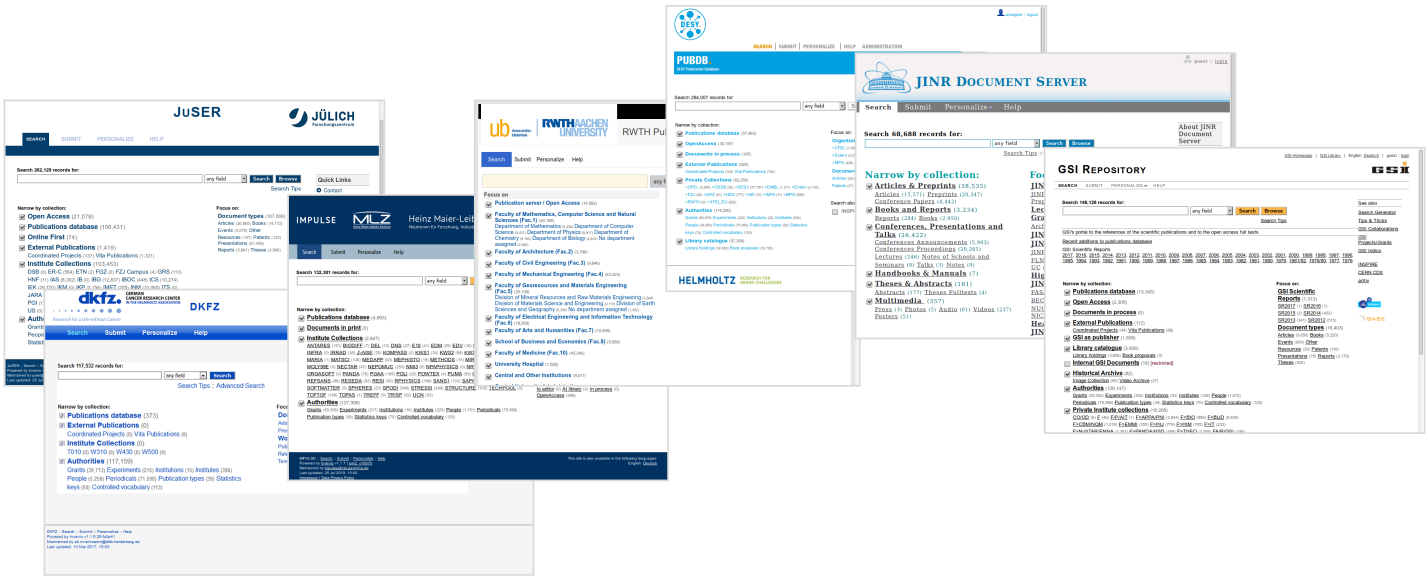
or

Александер Вагнер?

- ☺ both forms searchable
- ☺ authority display
- ☹ bibliographic display

(Affects authors, journals, grants, experiments…)

Installations online



Individual instances, powered by the partners on site

Easy, live Weboutput

Integrates with your current web pages

Publication lists need data as detailed as for evaluations…

(aid, cid, fai pof, pub, sid, typ …)

JOIN²-- Main features

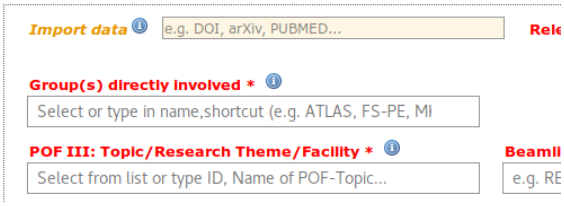
- Import interfaces (ease up input, improve data quality)
- Exports to BibTEX/EndNote/RIS (Integrate with citations management)
- Full text handling (private and OpenAccess, OpenAIRE Validated)
- Private collections
  - collect and share documents (in your institutes/groups)
  - integrate collection and reporting
- Normalize as much as possible (Key: Authorities)
  - Authors (tell apart Smith and Smith, handle Wagner and Вагнер, ORCiD enabled)
  - Institutes (group level)
  - Journals
  - Projects (EU, (inter-)national, local funding…)
  - …

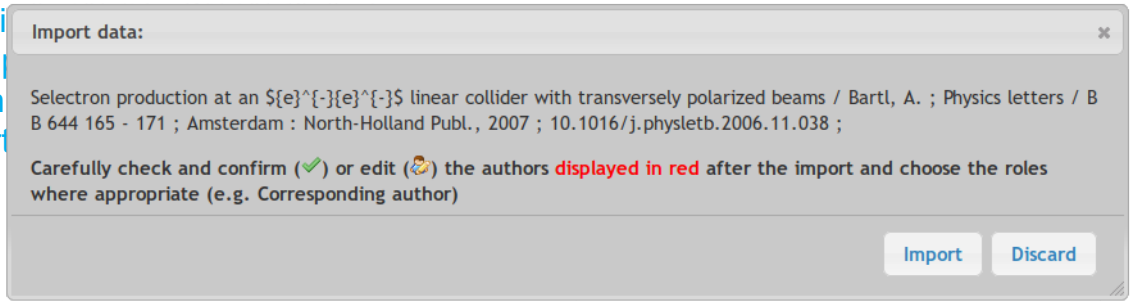
Scientists

…**don't need to care** about technical details …

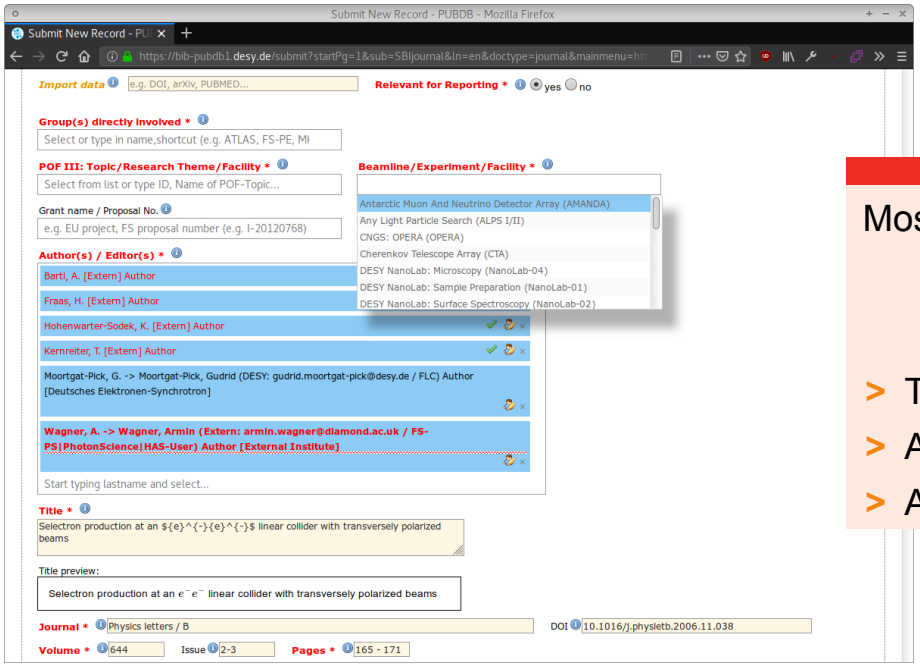
Import

- DOI: insert doi or doi.org-url (e. g. 10.1016/j.physletb.2006.11.038)
- INSPIRE: use URL (e. g. https://inspirehep.net/record/730481)
- `arXiv.org`: copy as displayed (e. g. arxiv:hep-ph/0610431)
- `pubmed`: copy as displayed (e. g. PMID: 20923669)
- ISBN: use the ISBN-field for this import
- …
- reuse existing data (e. g. conferences)
  - reci
  - pre
  - join
  - part





E. g. DOI Import

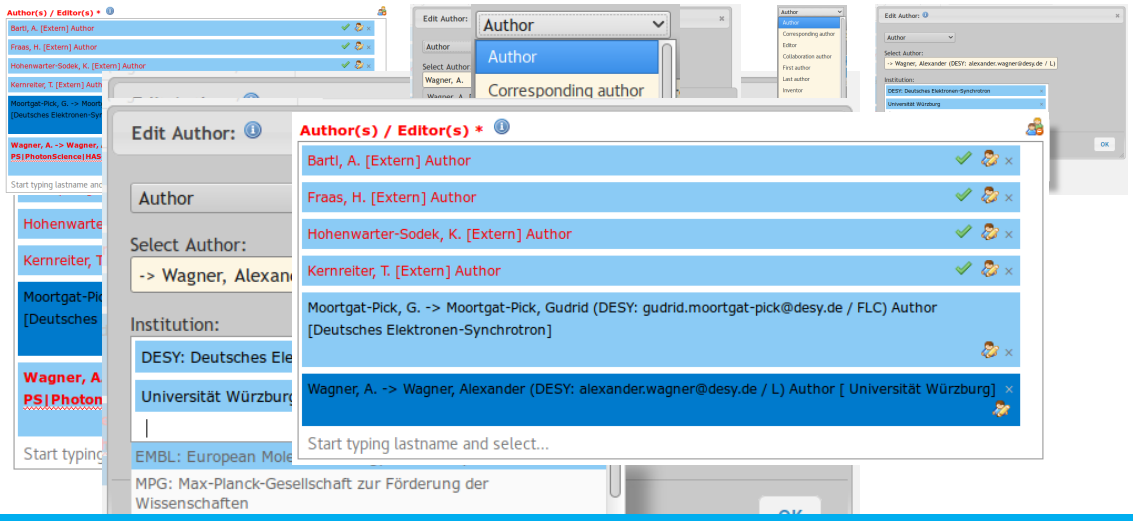


Most mandator (red) fields are already filled in!

But:

- Take care of authors
- Add your funding
- Add facilities used

Authordisambiguation and roles



All this is automatic, if

the source of the import gives the
ORCiD

(or INSPIRE-ID, any other ID the system knows of)

What you do not see

- Microformats (e. g. schema.org to enhance visibility and indexing)
- Normalization (Authorities and all that…)
- Author identifier (ORCiD, Inspire…)
- Data enrichment (e. g. uplinks to INSPIRE or WoS, feed)
- OAI-PMH (Feed search engines etc, OpenAIRE delivery)
- publishing workflow (more than *putting it on some page*)
- DOI minting (commitment to availability for $t \gg 0$)
- APC handling (Gold OpenAccess, SCOAP^3, DEAL…)
- Library system (DESY, GSI)
- Setup and maintanance (semiautomatic roll-out)
- Reporting tools

$\sum *.\text{py} > 150\text{k lines}$

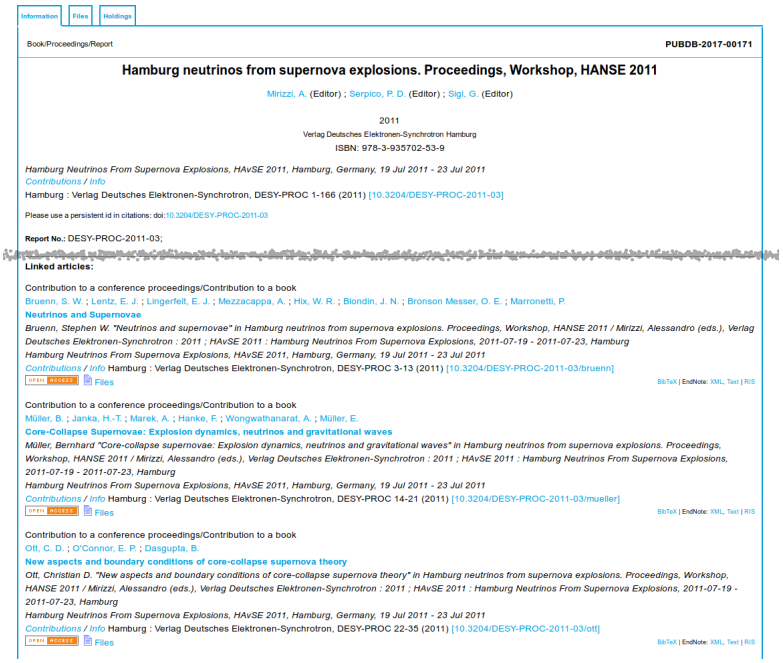
What you do not see

- Microformats (e. g. schema.org to enhance visibility and indexing)
- Normalization (Authorities and all that…)
- Author identifier (ORCiD, Inspire…)
- Data enrichment (e. g. uplinks to INSPIRE or WoS, feed)
- OAI-PMH (Feed search engines etc, OpenAIRE delivery)
- publishing workflow (more than *putting it on some page*)
- DOI minting (commitment to availability for $t \gg 0$)
- APC handling (Gold OpenAccess, SCOAP^3, DEAL…)
- Library system (DESY, GSI)
- Setup and maintanance (semiautomatic roll-out)
- Reporting tools

$\sum *.\text{py} > 150\text{k lines}$

Publishing workflow: DESY-PROC

- create and publish the book
  (including printing and ISBN assignment)
- deposit archival copies (dnb, TIB)
- DOI minting
  (make it citeable, commit to availability and update)
- clean up metadata
- publish individual articles online
- DOI minting
  (make them citeable, commit to availability and update)
- notify DESY groups
  (add funding to own articles)
- push data to INSPIRE
  (via DESY ingest workflow)



Information | Files | Holdings

Book/Proceedings/Report PUBDB-2017-00171

Hamburg neutrinos from supernova explosions. Proceedings, Workshop, HANSE 2011

Mirizzi, A. (Editor) ; Serpico, P. D. (Editor) ; Sigl, G. (Editor)

2011

Verlag Deutsches Elektronen-Synchrotron Hamburg

ISBN: 978-3-935702-53-9

Hamburg Neutrinos From Supernova Explosions, HAvSE 2011, Hamburg, Germany, 19 Jul 2011 - 23 Jul 2011

Contributions / Info

Hamburg : Verlag Deutsches Elektronen-Synchrotron, DESY-PROC 1-166 (2011) [10.3204/DESY-PROC-2011-03]

Please use a persistent id in citations: doi:10.3204/DESY-PROC-2011-03

Report No.: DESY-PROC-2011-03;

Linked articles:

Contribution to a conference proceedings/Contribution to a book

Bruenn, S. W. ; Lentz, E. J. ; Lingerfelt, E. J. ; Mezzacappa, A. ; Hix, W. R. ; Blondin, J. N. ; Bronson Messer, O. E. ; Marronetti, P.

Neutrinos and Supernovae

Bruenn, Stephen W. "Neutrinos and supernovae" in Hamburg neutrinos from supernova explosions. Proceedings, Workshop, HANSE 2011 / Mirizzi, Alessandro (eds.), Verlag Deutsches Elektronen-Synchrotron : 2011 ; HAvSE 2011 : Hamburg Neutrinos From Supernova Explosions, 2011-07-19 - 2011-07-23, Hamburg

Hamburg Neutrinos From Supernova Explosions, HAvSE 2011, Hamburg, Germany, 19 Jul 2011 - 23 Jul 2011

Contributions / Info Hamburg : Verlag Deutsches Elektronen-Synchrotron, DESY-PROC 3-13 (2011) [10.3204/DESY-PROC-2011-03/bruenn]

OPEN ACCESS Files BibTeX | EndNote: XML, Text | RIS

Contribution to a conference proceedings/Contribution to a book

Müller, B. ; Janka, H.-T. ; Marek, A. ; Hanke, F. ; Wongwathanarat, A. ; Müller, E.

Core-Collapse Supernovae: Explosion dynamics, neutrinos and gravitational waves

Müller, Bernhard "Core-collapse supernovae: Explosion dynamics, neutrinos and gravitational waves" in Hamburg neutrinos from supernova explosions. Proceedings, Workshop, HANSE 2011 / Mirizzi, Alessandro (eds.), Verlag Deutsches Elektronen-Synchrotron : 2011 ; HAvSE 2011 : Hamburg Neutrinos From Supernova Explosions, 2011-07-19 - 2011-07-23, Hamburg

Hamburg Neutrinos From Supernova Explosions, HAvSE 2011, Hamburg, Germany, 19 Jul 2011 - 23 Jul 2011

Contributions / Info Hamburg : Verlag Deutsches Elektronen-Synchrotron, DESY-PROC 14-21 (2011) [10.3204/DESY-PROC-2011-03/mueller]

OPEN ACCESS Files BibTeX | EndNote: XML, Text | RIS

Contribution to a conference proceedings/Contribution to a book

Ott, C. D. ; O'Connor, E. P. ; Dasgupta, B.

New aspects and boundary conditions of core-collapse supernova theory

Ott, Christian D. "New aspects and boundary conditions of core-collapse supernova theory" in Hamburg neutrinos from supernova explosions. Proceedings, Workshop, HANSE 2011 / Mirizzi, Alessandro (eds.), Verlag Deutsches Elektronen-Synchrotron : 2011 ; HAvSE 2011 : Hamburg Neutrinos From Supernova Explosions, 2011-07-19 - 2011-07-23, Hamburg

Hamburg Neutrinos From Supernova Explosions, HAvSE 2011, Hamburg, Germany, 19 Jul 2011 - 23 Jul 2011

Contributions / Info Hamburg : Verlag Deutsches Elektronen-Synchrotron, DESY-PROC 22-35 (2011) [10.3204/DESY-PROC-2011-03/ott]

OPEN ACCESS Files BibTeX | EndNote: XML, Text | RIS

Article Processing Charges / Publication Fees

Traditional publication model

- Closed Access publication
- Pay to read
- Fees: journal subscriptions
  (subscriptions ↑↑↑ inflation, "serials crisis")
- Additional fees:
  - article buy out (hybrid Open Access)
  - page charges
  - colour charges
  - cover charges
  - submission fees

APC based model

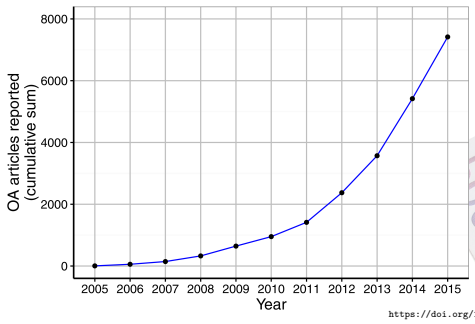
- Open Access publication
- Pay to publish
- Fees: several components
  (e. g. base fees, licences, reductions, global/national contracts)
- Changes:
  - new, different workflows
  - publisher and journal dependent
  - one bill per publication
  - several budgets (e. g. project funding)
  - complex contracts (e. g. DEAL)

Helpdesk for authors

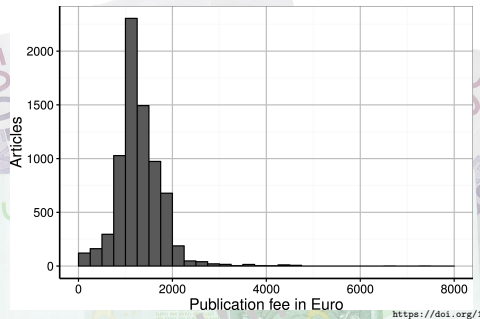
(copyright, authors rights, necessary fees, handling, funding possibilities, funder requirements etc.)

Manage "the paperwork" centrally at the libraries.
Keep track of the total costs of publishing for future contracts.

APC Handling



https://doi.org/10.7717/peerj.2323/fig-1, CCBY4



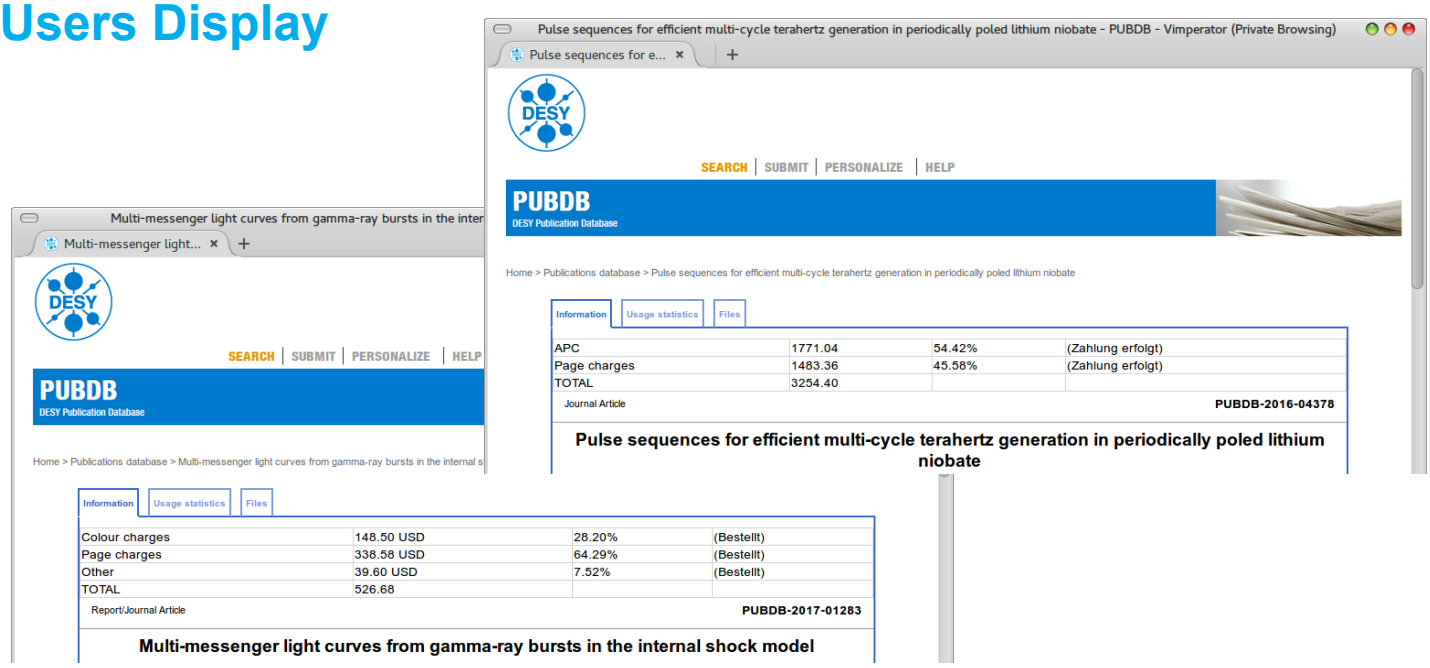
https://doi.org/10.7717/peerj.2323/fig-2, CCBY4

Jahn N., Tullney M. (2016) **A study of institutional spending on open access publication fees in Germany.** *PeerJ* 4:e2323 `https://doi.org/10.7717/peerj.2323`

Austrian Science Fund (FWF)	€13.445.549	*2017: €7.187.295*	**€2.243**	*2017: €2.169*
Wellcome Trust	€11.966.250	*2017: €10.750.792*	**€2.408**	*2017: €2.456*
Forschungszentrum Jülich	€1.055.902	*2017: €196.869*	**€1.517**	*2017: €1.246*
open APC (full set)	**€162.943.280**		**€1.941**	

Source: OpenAPC-DE/intact project, 2019-09-03

Users Display



Outlook: Datapublications

Current infrastrucure is not suitable for intermediary to large data sets (GB and up).

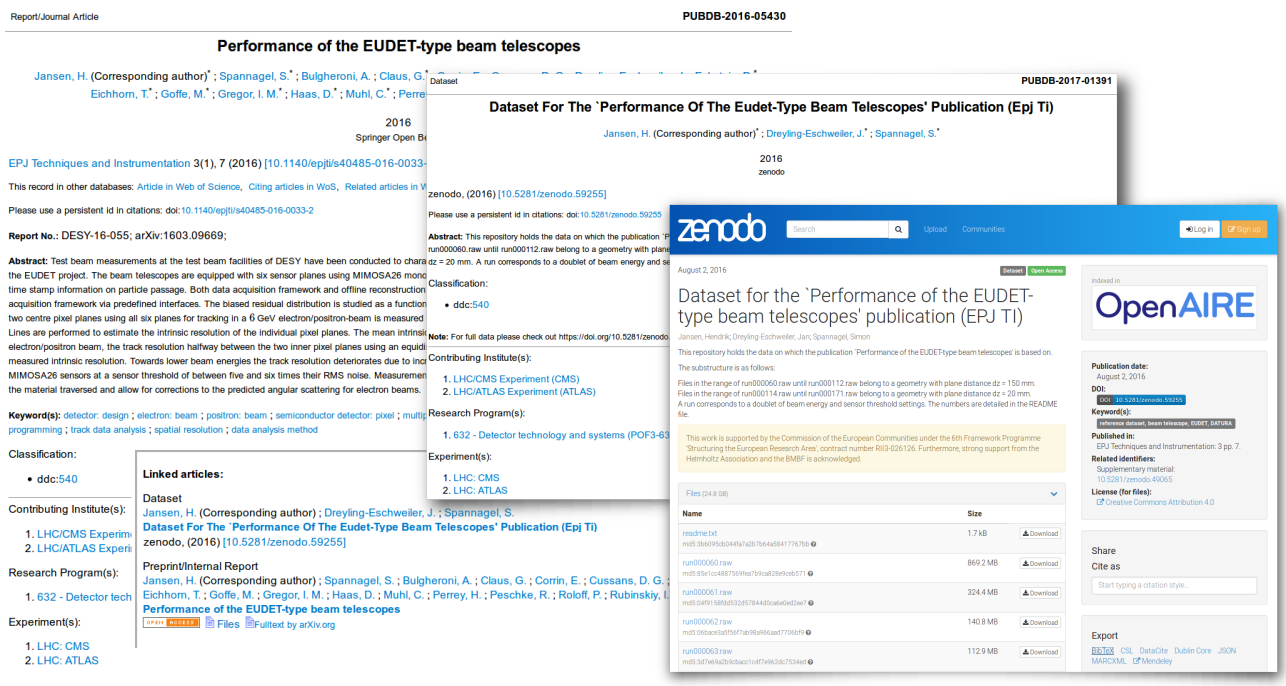
(Slow up-/download, limited storage backends)

- small data sets: like normal publications
- others: link to data repositories (e.g. Zenodo)
- future
  - Invenio 3.x
  - improve storage backend
  - improve general infra structure

Datasets may include complex permission handling as well as referencing unpublished/embargoed data sets, even on external storage solutions.

(see e. g. implementation at RWTH Aachen)

Data @ PubDB



People behind JOIN^2

- Dagmar Sitek^a
- Gudrun Friedburg^a
- Martin Köhler^b
- Zaven Akopov^{b,d,1}
- Tomasz Pazera^{b,1}
- Kirsten Sachs^{b,h}
- Florian Schwennsen^{b,h}
- Robert Thiele^b
- Alexander Wagner^{e,b}
- Andrea Ciocchetti^c
- Laura Baracchi^c
- Cüneyt Uysal^c
- Bernhard Mittermaier^e
- Torsten Bronger^e
- Claudia Frick^e
- Anna Fründ^e
- Christoph Holzke^e
- Heike Lexis^e
- Cornelia Plott^e
- Katrin Große^f
- Stefan Hesselbach^g
- Irina Filozova^i
- Tatiana Zaikina^i
- Galina Shestakova^i
- Jürgen Neuhaus^j
- Connie Hesse^j
- Björn Pedersen^j
- Jörg Pulz^j
- Ulrike Eich^k
- Louai Barake^k
- Corinna Brückener ^{k}
- Abdoulaye Diallo^k
- Roland Rappmann^k
- Dominik Schmitz^k
- Edmund Wollgarten^k

^{a} DKFZ Central Library, ^{b} DESY Central Library, ^{c} DZNE, ^{d} European XFEL, ^{e} Forschungszentrum Jülich, Central Library, ^{f} GSI Library, ^{g} GSI Core IT, ^{h} INSPIRE, ^{i} JINR LIT, ^{j} MLZ Garching, ^{k} RWTH Aachen, University Library, ^{1} (retired)

plus a little help from our friends

CERN Library, CERN IT, INSPIRE

Thank you!

Contact

DESY. Deutsches
Elektronen-Synchrotron

www.desy.de

Alexander Wagner
0000–0001–9846–5516
Central Library
alexander.wagner@desy.de
+49–40–8998–1758
10.3204/PUBDB-2019-02865

Typeset by lualaTeX