

# JOIN<sup>2</sup>

A Publication Database and Repository based on Invenio

Alexander Wagner  
Budva Bečići, 02.10.2019

HELMHOLTZ RESEARCH FOR  
GRAND CHALLENGES



## 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  $\text{\LaTeX}$ )

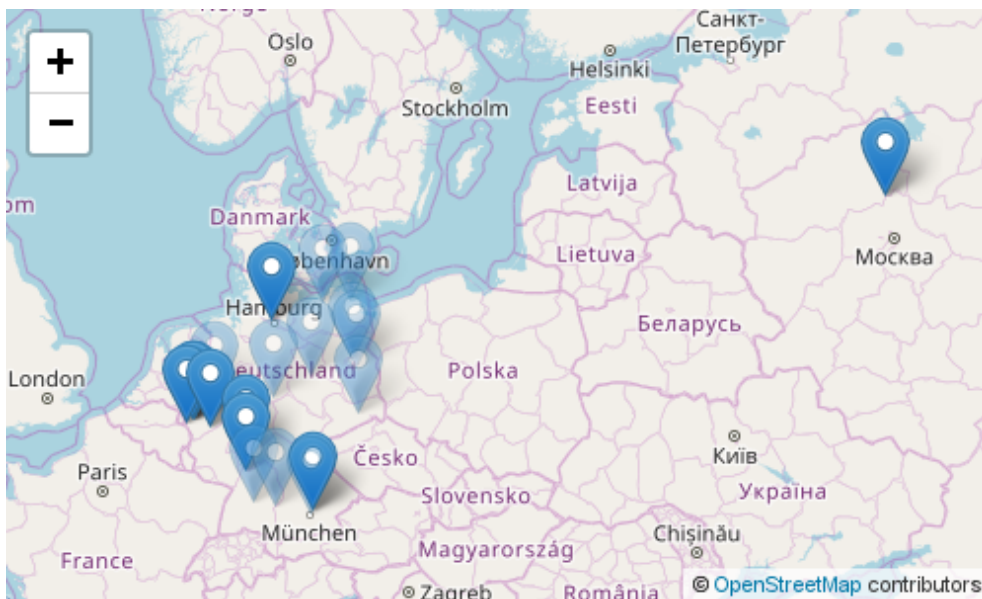


# JOIN² collaboration

Deutsches Elektronen-Synchrotron, DESY	≈ 2300 + 3000	
Deutsches Krebsforschungszentrum	≈ 3000	(internal access only)
<b>JOIN²: proud to serve</b>		
<b>9 institutions,</b> <b>&gt; 28.000 staff members</b> <b>&gt; 6000 visiting scientists</b> <small>(≈ 475.000 records, ≈ 68.000 OpenAccess, ≈ 135.000 shared authorities)</small>		
Heinz Maier-Leibnitz Zentrum, Garching	≈ 500 + 1000	
RWTH Aachen	≈ 9500	
Museum Zitadelle Jülich	≈ 3 + n	



# Put it on the map...



## Current challenge

### Cyrillic script

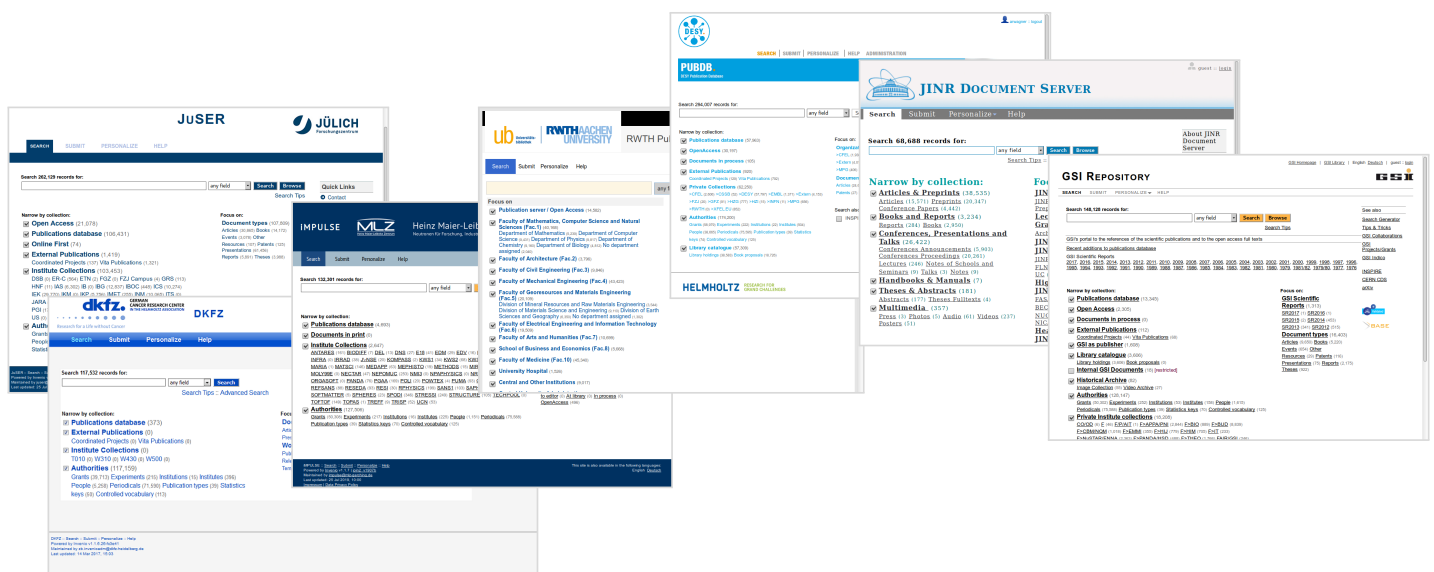
Wagner, Alexander  
or  
Александр Вагнер?

- 😊 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 )
- > 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)

**Import data** ⓘ e.g. DOI, arXiv, PUBMED... **Rele**

**Group(s) directly involved \*** ⓘ  
Select or type in name,shortcut (e.g. ATLAS, FS-PE, MI)

**POF III: Topic/Research Theme/Facility \*** ⓘ **Beamli**  
Select from list or type ID, Name of POF-Topic... e.g. RE

**Import data:**

Selectron production at an  $e^+e^-$  linear collider with transversely polarized beams / Bartl, A. ; Physics letters / B B 644 165 - 171 ; Amsterdam : North-Holland Publ., 2007 ; 10.1016/j.physletb.2006.11.038 ;

Carefully check and confirm (✓) or edit (✎) the authors **displayed in red** after the import and choose the roles where appropriate (e.g. Corresponding author)

**Import** **Discard**



## E. g. DOI Import

Submit New Record - PUBDB - Mozilla Firefox

Import data ⓘ e.g. DOI, arXiv, PUBMED... **Relevant for Reporting \*** ⓘ yes no

**Group(s) directly involved \*** ⓘ  
Select or type in name,shortcut (e.g. ATLAS, FS-PE, MI)

**POF III: Topic/Research Theme/Facility \*** ⓘ  
Select from list or type ID, Name of POF-Topic...

**Beamline/Experiment/Facility \*** ⓘ  
Antarctic Muon And Neutrino Detector Array (AMANDA)  
Any Light Particle Search (ALPS I/II)  
CNGS: OPERA (OPERA)  
Cherenkov Telescope Array (CTA)  
DESY NanoLab: Microscopy (NanoLab-04)  
DESY NanoLab: Sample Preparation (NanoLab-01)  
DESY NanoLab: Surface Spectroscopy (NanoLab-02)

**Grant name / Proposal No. ⓘ**  
e.g. EU project, FS proposal number (e.g. I-20120768)

**Author(s) / Editor(s) \*** ⓘ  
Bartl, A. [Extern] Author  
Fraas, H. [Extern] Author  
Hohenwarter-Sodek, K. [Extern] Author  
Kernreiter, T. [Extern] Author  
Moortgat-Pick, G. -> Moortgat-Pick, Gudrid (DESY: gudrid.moortgat-pick@desy.de / FLC) Author [Deutsches Elektronen-Synchrotron]  
Wagner, A. -> Wagner, Armin (Extern: armin.wagner@diamond.ac.uk / FS-PS|PhotonScience|HAS-User) Author [External Institute]

**Title \*** ⓘ  
Selectron production at an  $e^+e^-$  linear collider with transversely polarized beams

**Title preview:**  
Selectron production at an  $e^+e^-$  linear collider with transversely polarized beams

**Journal \*** ⓘ Physics letters / B **DOI** ⓘ 10.1016/j.physletb.2006.11.038

**Volume \*** ⓘ 644 **Issue** ⓘ 2-3 **Pages \*** ⓘ 165 - 171

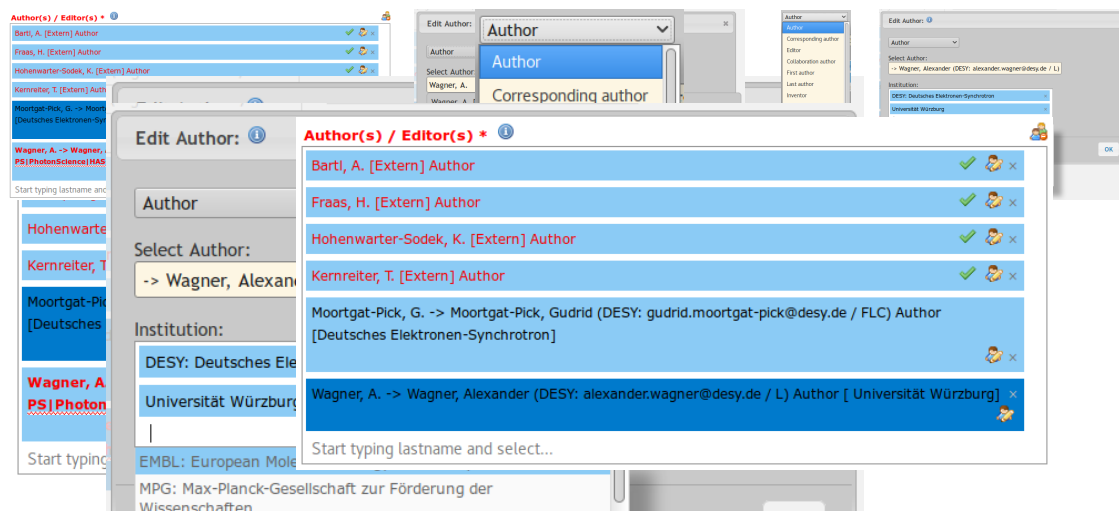
Most mandator (red) fields are already filled in!

But:

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




# Author disambiguation and roles



All this is automatic, if




the source of the import gives the

 ORCID

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

## What you do not see




- > Microformats (e. g. [schema.org](http://schema.org) to enhance visibility and indexing)
- > Normalization (Authorities and all that...)
- > Author identifier ( ORCID, Inspire...)
- > Data enrichment (e. g. uplinks to  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<sup>3</sup>, DEAL...)
- > Library system (DESY, GSI)
- > Setup and maintenance (semiautomatic roll-out)
- > Reporting tools

$\sum *.py > 150k$  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<sup>3</sup>, DEAL...)
- > Library system (DESY, GSI)
- > Setup and maintenance (semiautomatic roll-out)
- > Reporting tools

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



## 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)

The screenshot displays the DESY-PROC website interface. At the top, there are tabs for 'Information', 'Files', and 'Holdings'. The main content area is titled 'Book/Proceedings/Report' and 'PUBDB-2017-00171'. The title of the publication is 'Hamburg neutrinos from supernova explosions. Proceedings, Workshop, HANSE 2011'. Below the title, the editors are listed: Mirizzi, A. (Editor); Serpico, P. D. (Editor); Sigli, G. (Editor). The year 2011 is also indicated. The publisher is 'Verlag Deutsches Elektronen-Synchrotron Hamburg' with ISBN 978-3-935702-53-9. The publication date is 'Hamburg Neutrinos From Supernova Explosions, HANSE 2011, Hamburg, Germany: 19 Jul 2011 - 23 Jul 2011'. The publisher's website is 'Verlag Deutsches Elektronen-Synchrotron, DESY-PROC 1-166 (2011) [10.3204/DESY-PROC-2011-03]'. A note states 'Please use a persistent id in citations: doi:10.3204/DESY-PROC-2011-03'. The report number is 'Report No.: DESY-PROC-2011-03;'. Under 'Linked articles:', there are several entries, including 'Contribution to a conference proceedings/Contribution to a book' by Bruenn, S. W.; Lentz, E. J.; Lingerfelt, E. J.; Mezzacappa, A.; Hix, W. R.; Blondin, J. N.; Bronson Messer, O. E.; Marronetti, P. and 'Core-collapse supernovae: Explosion dynamics, neutrinos and gravitational waves' by Müller, Bernhard. Each entry includes a brief description and a link to the full text or related publications.



# 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)

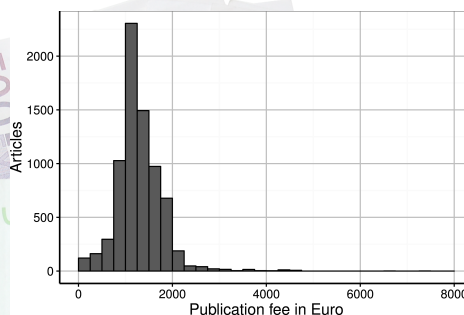
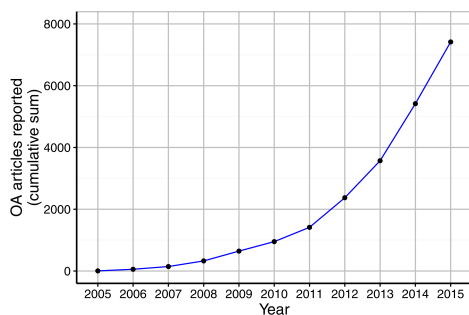
## Helpdesk for authors

(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



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	<b>€2.243</b>	2017: €2.169
Wellcome Trust	€11.966.250	2017: €10.750.792	<b>€2.408</b>	2017: €2.456
Forschungszentrum Jülich	€1.055.902	2017: €196.869	<b>€1.517</b>	2017: €1.246
<b>open@PC (full set)</b>	<b>€162.943.280</b>		<b>€1.941</b>	

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



# Users Display

Multi-messenger light curves from gamma-ray bursts in the internal shock model

DESY Publication Database

Home > Publications database > Multi-messenger light curves from gamma-ray bursts in the internal shock model

Information	Usage statistics	Files
Colour charges	148.50 USD	28.20%
Page charges	338.58 USD	64.29%
Other	39.60 USD	7.52%
TOTAL	526.68	

Report/Journal Article

PUBDB-2017-01283

Multi-messenger light curves from gamma-ray bursts in the internal shock model

Pulse sequences for efficient multi-cycle terahertz generation in periodically poled lithium niobate - PUBDB - Vimperator (Private Browsing)

DESY Publication Database

Home > Publications database > Pulse sequences for efficient multi-cycle terahertz generation in periodically poled lithium niobate

Information	Usage statistics	Files
APC	1771.04	54.42%
Page charges	1483.36	45.58%
TOTAL	3254.40	

Journal Article

PUBDB-2016-04378

Pulse sequences for efficient multi-cycle terahertz generation in periodically poled lithium niobate



# Outlook: Datapublications

Current infrastructure 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)





RIORD-2017-01394

2016  
Springer Open Be

This record in other databases: [Article in Web of Science](#), [Citing articles in](#)

Please use a persistent id in citations: doi:[10.1140/epjc/s40485-016-0033-2](#)

Report No.: DESY-16-055; arXiv:1603.09669;

**Abstract:** Test beam measurements at the test beam facilities of DESY have been conducted to check the EUEDET project. The beam telescopes are equipped with six sensor planes using MIMOSA26 modules. The stamp information on particle passage. Both data acquisition framework and offline reconstruction acquisition framework via predefined interfaces. The biased residual distribution is studied as a function of two centre pixel planes using all six planes for tracking in a 6 GeV electron/positron-beam is measured. Lines are performed to estimate the intrinsic resolution of the individual pixel planes. The mean intrinsic electron/positron beam, the track resolution halfway between the two inner pixel planes using an equal measured intrinsic resolution. Towards lower beam energies the track resolution deteriorates due to increased MIMOSA26 sensors at a sensor threshold of between five and six times their RMS noise. Measurement of the material traversed and allow for corrections to the predicted angular scattering for electron beams.

**Keyword(s):** detector: design ; electron: beam ; positron: beam ; semiconductor detector: pixel ; multiplexing ; track data analysis ; spatial resolution ; data analysis method

**Classification:**

- [ddc:540](#)

Contributing Institute(s):

1. LHC/CMS Experiment
2. LHC/ATLAS Experiment

Research Program(s):

1. 632 - Detector tech

1. LHC: CMS
2. LHC: ATLAS

**Dataset For The 'Performance Of The Eudet-Type Beam Telescopes' Project**

2016

zenodo, (2016) [[10.5281/zenodo.592551](https://doi.org/10.5281/zenodo.592551)]

Please use a persistent id in citations: doi:[10.5281/zenodo.59255](https://doi.org/10.5281/zenodo.59255)

**Abstract:** This repository holds the data on which the publication 'I

**Classification:**

- [ddc:540](#)

**Note:** For full data please check out <https://doi.org/10.5281/zenodo.10000000>

Contributing Institute(s):


1. LHC/CMS Experiment (CMS)
2. LHC/ATLAS Experiment (ATLAS)

Research Program(s):

1. 632 - Detector technology and systems (POF3-6)

Experiment(s):

1. LHC: CMS
2. LHC: ATLAS



August 2, 2016

Dataset
Open Access

# Dataset for the 'Performance of the EUDET-type beam telescopes' publication (EPJ TI)

Janßen, Hendrik; Dreyling, Eschweiler, Jan; Spangnagel, Simon

This repository holds the data on which the publication "Performance of the EUDET-type beam telescopes" is based on.

The substructure is as follows:

Files in the range /run000050.raw until /run000112.raw belong to a geometry with plane distance  $d = 150$  mm. Files in the range /run000114.raw until /run000171.raw belong to a geometry with plane distance  $d = 20$  mm. A run corresponds to a docket of beam energy and sensor threshold settings. The numbers are detailed in the README file.

This work is supported by the Commission of the European Communities under the 6th Framework Programme Structuring the European Research Area, contract number R3-G2-02-16. Furthermore, strong support from the Helmholtz Association and the BMBF is acknowledged.

Files (248.08)

Name	Size	
readme.txt	1.7 kB	
run5_38a0955d0a47b47654a0b4177678a0		
run000060.raw	869.2 MB	
run513491c688759f98a7236ac0e5d571		
run000091.raw	324.4 MB	
run513491c688759f98a7236ac0e5d571		
run000062.raw	140.8 MB	
run513491c688759f98a7236ac0e5d571		
run000093.raw	112.9 MB	
run513491c688759f98a7236ac0e5d571		

Share  
Site as

Export

[Bibtex](#)
[Cite](#)
[Data-Cite](#)
[Dublin Core](#)
[JSON](#)
[MARCXML](#)
[Mendeley](#)

Indiced in

Publication date:  
August 2, 2016

DOI:  
[10.5281/zenodo.52095](#)

Keywords:  
[information science](#), [beam telescope](#), [EUDET](#), [ATLAS](#)

Published in:  
EPJ: Techniques and Instrumentation 3 pp. 7

Related identifiers:  
Supplementary material:  
[10.5281/zenodo.49050](#)

License (for files):  
 Creative Commons Attribution 4.0



- > Jürgen Neuhaus<sup>j</sup>
- > Connie Hesse<sup>j</sup>
- > Björn Pedersen<sup>j</sup>
- > Jörg Pulz<sup>j</sup>
- > Ulrike Eich<sup>k</sup>
- > Louai Barake<sup>k</sup>
- > Corinna Brückener<sup>k</sup>
- > Abdoulaye Diallo<sup>k</sup>
- > Roland Rappmann<sup>k</sup>
- > Dominik Schmitz<sup>k</sup>
- > Edmund Wollgarten<sup>k</sup>

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

CERN Library, CERN IT, INSPIRE






# Thank you!

## Contact

**DESY.** Deutsches  
Elektronen-Synchrotron

[www.desy.de](http://www.desy.de)

Alexander Wagner  
 0000-0001-9846-5516  
Central Library  
[alexander.wagner@desy.de](mailto:alexander.wagner@desy.de)  
+49-40-8998-1758  
[10.3204/PUBDB-2019-02865](https://doi.org/10.3204/PUBDB-2019-02865)



*Typeset by lua<sup>A</sup>T<sub>E</sub>X*

