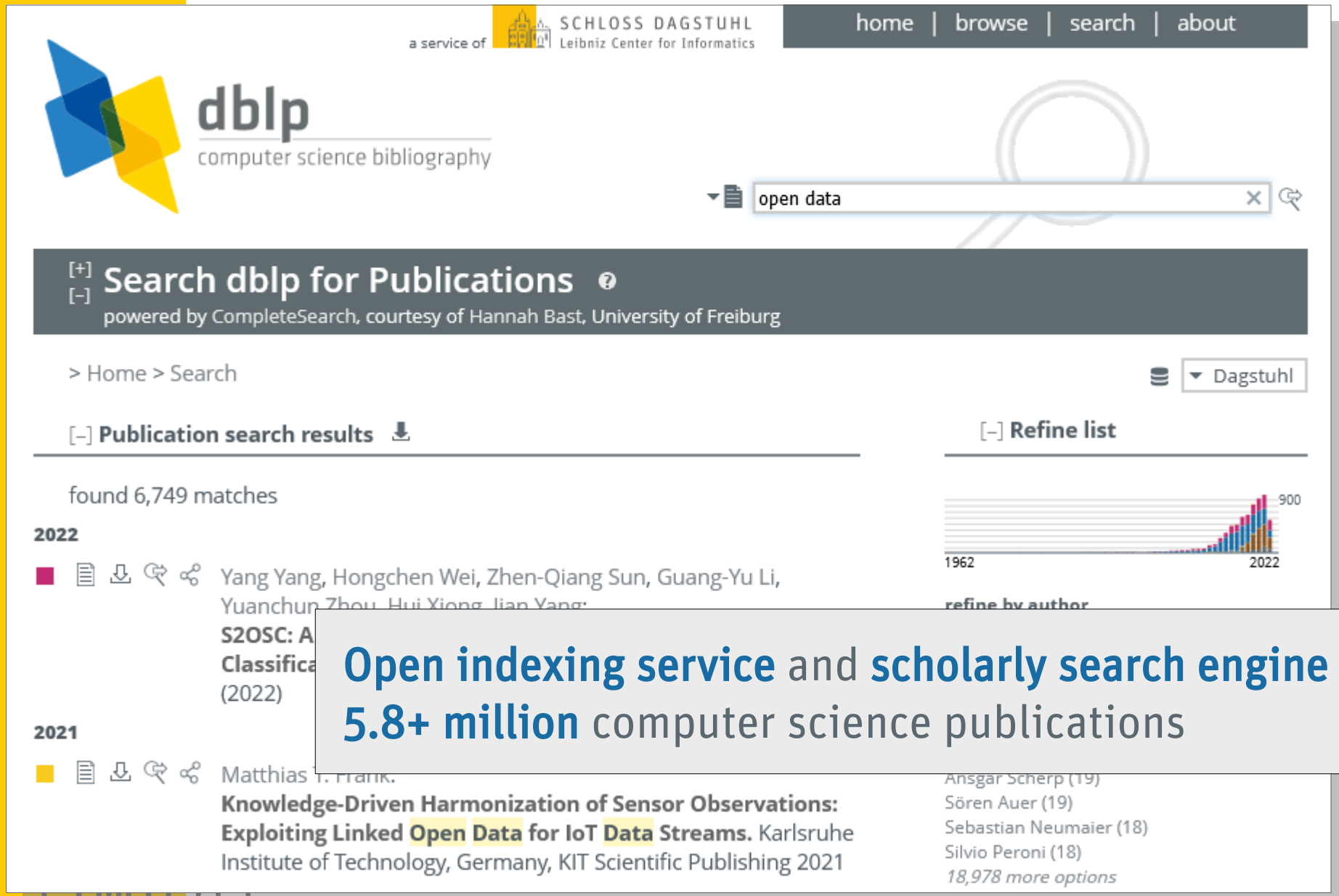
dblp
computer science bibliography

Open, Curated, and Semantically Meaningful Metadata for the Informatics Community

Marcel R. Ackermann

October 27, 2021

What is dblp? (1/4)



Open indexing service and **scholarly search engine** for **5.8+ million** computer science publications

What is dblp? (2/4)

[+] [–] Silvio Peroni

> Home > Persons

by year | Dagstuhl

[–] **Person information**

- *affiliation:* University of Bologna, Department of Classical Philology and Italian Studies,
- *affiliation:* OpenCitations.net

[–] **2020 – today**

2021

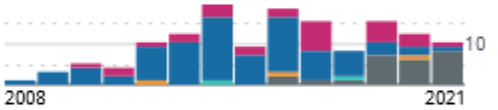
[j29] Ivan Heibi, Silvio Peroni:
A qualitative and quantitative analysis of open citations to retracted articles: the Wakefield 1998 et al.'s case.
Scientometrics 126(10): 8433-8470 (2021)

[c82] Zeyd Boukhers, Philipp Mayr, Silvio Peroni:
BiblioDAP'21: The 1st Workshop on Bibliographic Data Analysis and Processing. KDD 2021: 4110-4111

[i25] Federica Bologna, Angelo Di Iorio, Silvio Peroni, Francesco Poggi:
Do open citations inform the qualitative peer-review evaluation in research assessments? An analysis of the Italian National Scientific Qualification. CoRR abs/2103.07942 (2021)

[i24] Federica Bologna, Angelo Di Iorio, Silvio Peroni, Francesco Poggi:
Can we assess research using open scientific knowledge graphs? A case study within the Italian National Scientific Qualification. CoRR abs/2105.08599 (2021)

Curated bibliographies for **2.8+ million** authors and editors



refine by search term

refine by type
- Journal Articles (only)
- Conference and Workshop Papers (only)
- Parts in Books or Collections (only)
- Editorship (only)
- Informal Publications (only)

select all | deselect all

refine by coauthor
Fabio Vitali (63)
Angelo Di Iorio (35)
David M. Shotton (34)
Andrea Giovanni Nuzzolese (31)

What is dblp? (3/4)

[+] [–] ACM Conference on Human Factors in Computing Systems (CHI)

> Home > Conferences and Workshops

Trier 2

[–] Description

The **ACM Conference on Human Factors in Computing Systems** (**CHI**) series of academic conferences is generally considered the most prestigious in the field of human–computer interaction and is one of the top-ranked conferences in computer science. It is hosted by ACM SIGCHI, the Special Interest Group on computer–human interaction. CHI has been held annually since 1982 and attracts thousands of international attendees. CHI 2020, which was originally planned to take place on April, was cancelled due to COVID-19, and CHI 2021 will be held online as a virtual conference chaired by Yoshifumi Kitamura and Aaron Quigley. CHI 2021 "making waves, combining strengths" was originally scheduled to take place in Yokohama.

read full information at Wikipedia

[Text kindly provided by Wikipedia under CC-BY-SA-3.0 license.]

CHI 2021: Virtual Event / Yokohama, Japan

Yoshifumi Kitamura, Aaron Quigley, Katherine Isbister, Takeo Igarashi, Pernille Bjørn, Steven Mark Drucker:
CHI '21: CHI Conference on Human Factors in Computing Systems, Virtual Event / Yokohama, Japan, May 8-13, 2021. ACM 2021, ISBN 978-1-4503-8096-6 [contents] 7.6

Yoshifumi Kitamura, Aaron Quigley, Katherine Isbister, Takeo Igarashi:
CHI '21: CHI Conference on Human Factors in Computing Systems, Virtual Event / Yokohama Japan, May 8-13, 2021, Extended Abstracts. ACM 2021, ISBN 978-1-4503-8095-9 [contents] 5.9

[–] Venue statistics

records by year

frequent topics

design interact interfac social system user visual

(currently stemmed word rumps only)

frequent authors

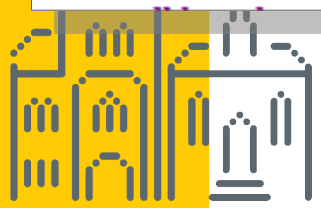
- Stephen A. Brewster (109)
- Hiroshi Ishii 0001 (101)
- Patrick Olivier (98)
- Florian 'Floyd' Mueller (95)
- Albrecht Schmidt 0001 (87)
- Patrick Baudisch (86)

Open directory of **1,700+** computer science journals and **5,500+** conference and workshop series

What is dblp? (4/4)

```
−<rdf:RDF>
  −<dblp:Person rdf:about="https://dblp.org/pid/48/1099">
      <rdf:type rdf:resource="https://dblp.org/rdf/schema#Creator"/>
      <owl:sameAs rdf:resource="https://orcid.org/0000-0003-0530-4305"/>
      <owl:sameAs rdf:resource="https://www.wikidata.org/entity/Q30536251"/>
      <owl:sameAs rdf:resource="https://id.loc.gov/authorities/names/no2014111866"/>
      <owl:sameAs rdf:resource="https://d-nb.info/gnd/105432011X"/>
      <rdfs:label>Silvio Peroni</rdfs:label>
      <dblp:orcid rdf:resource="https://orcid.org/0000-0003-0530-4305"/>
      <dblp:wikidata rdf:resource="https://www.wikidata.org/entity/Q30536251"/>
    +<datacite:hasIdentifier></datacite:hasIdentifier>
    +<datacite:hasIdentifier></datacite:hasIdentifier>
    +<datacite:hasIdentifier></datacite:hasIdentifier>
    +<datacite:hasIdentifier></datacite:hasIdentifier>
    +<datacite:hasIdentifier></datacite:hasIdentifier>
    +<datacite:hasIdentifier></datacite:hasIdentifier>
    +<datacite:hasIdentifier></datacite:hasIdentifier>
    −<datacite:hasIdentifier>
      −<datacite:Identifier>
          <datacite:usesIdentifierScheme rdf:resource="http://purl.org/spar/datacite/gnd"/>
          <litre:hasLiteralValue>105432011X</litre:hasLiteralValue>
        </datacite:Identifier>
      </datacite:hasIdentifier>
      <dblp:primaryFullCreatorName>Silvio Peroni</dblp:primaryFullCreatorName>
    −<dblp:primaryAffiliation>
        University of Bologna, Department of Classical Philology and Italian Studies, Italy
      </dblp:primaryAffiliation>
      <dblp:otherAffiliation>OpenCitations.net</dblp:otherAffiliation>
      <dblp:primaryHomepage rdf:resource="http://www.essepuntato.it/"/>
      <dblp:otherHomepage rdf:resource="https://www.unibo.it/sitoweb/silvio.peroni/en"/>
      <dblp:webpage rdf:resource="https://scholar.google.com/citations?user=uNyDqoUAAAAJ"/>
```

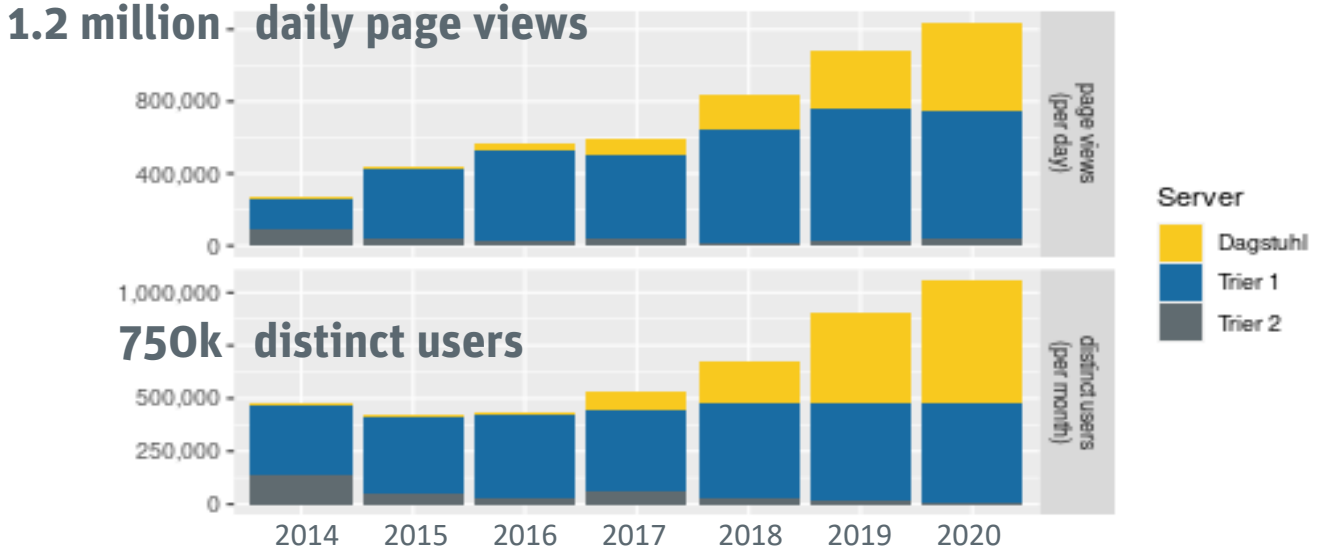
Open data APIs and full **dump download** of all of dblp's data for unrestricted reuse under **CC0 license**

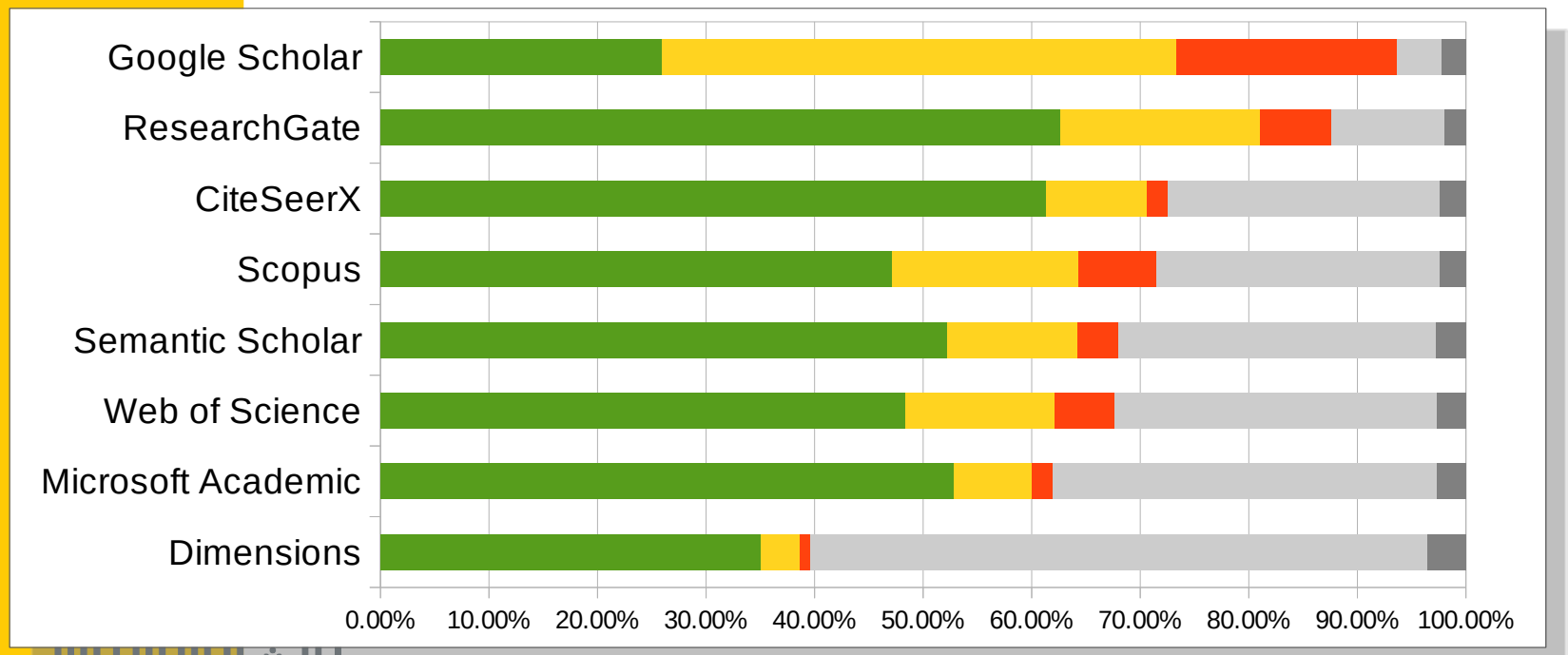


Our key principles

- **Openness** – all data is freely available for reuse
- **Neutrality of the data set** – additions selected according to significance to the community, no publisher preference, supervised by an advisory board
- **Data quality** – continuous curation & disambiguation
- **Semantic enrichment** – adding structure and linking to external resources
- **Enabling of research** – by providing means for literature search, and as a research data set
- **User orientation** – build to meet the needs of the international computer science research community

Usage statistics and user satisfaction





User survey 2020
N = 760

Author Disambiguation (1/2)

- Homonyms

Introduction of the Minitrack on Information Security and Privacy

Tung Bui
University of Hawaii at Manoa

Eric Clemons
University of Pennsylvania
clemons@wharton.upenn.edu

David Wang
University of Hawaii at Manoa
twwang@hawaii.edu

A Practical and Portable Solids-State Electronic Terahertz Imaging System

Ken Smart [1], Jia Du [1,*], Li Li [1], David Wang [1], Keith Leslie [1], Fan Ji [2]

a hidden Markov model for detecting recombination al detection microarrays

, Hilary Renshaw[1,2], Scott Weaver[3,4,5], Robert B. Tesh[3] and David Wang[1]

Real-Time Final Product Quality Prediction in Batch Process Operation

David Wang, *Member, IEEE*

Mechanisms, Overheads and Scaling

Brinda Ganesh†, Aamer Jaleel‡, David Wang†, and Bruce Jacob†

- Synonyms

Surface Prediction for Spatial Augmented Reality

Adam Gomes(✉), Keegan Fernandes, and David Wang

Degenerate Motions in Multicamera Cluster SLAM with Non-overlapping Fields

Michael J. Tribou[a,*], David W. L. Wang[b], Steven L. Waslander[a]

Collision Detection Using a Flexible Link Manipulator

S. J. Moorehead and D. Wang

Author Disambiguation (2/2)

[+] [−] **Wei Wang**

> Home > Persons

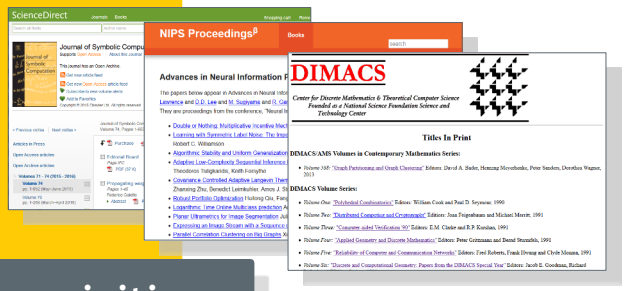
by year | Dagstuhl

This is just a *disambiguation page*, and is not intended to be the bibliography of an actual person. The links to all actual bibliographies of persons of the same or a similar name can be found below. Any publication listed on this page has not been assigned to an actual author yet. If you know the true author of one of the publications listed below, you are welcome to contact us.

[−] **Other persons with the same name**

- Wei Wang 0001 — University of Waterloo, David R. Cheriton School of Computer Science, ON, Canada
- Wei Wang 0002 — Nanjing University, State Key Laboratory for Novel Software Technology, Nanjing, China (and 1 more)
- Wei Wang 0003 — Chinese Academy of Sciences, Institute of Microelectronics, Laboratory of Nanofabrication and Novel Devices Integration Technology, Beijing, China (and 4 more)
- Wei Wang 0004 — Fudan University, School of Life Science, Shanghai, China
- Wei Wang 0005 — Zhejiang University, Center for Engineering and Scientific Computation, China
- Wei Wang 0006 — Language Weaver, Inc.
- Wei Wang 0007 — Chinese Academy of Sciences, ThinkIT Speech Lab, Institute of Acoustics
- Wei Wang 0008 — MIT, Nonlinear Systems Laboratory, Department of Mechanical Engineering, Cambridge, MA, USA
- Wei Wang 0009 — Fudan University, School of Computer Science, Shanghai Key Laboratory of Data Science, Shanghai, China
- Wei Wang 0010 — University of California Los Angeles, CA, USA (and 1 more)
- Wei Wang 0011 — The University of New South Wales, School of Computer Science and Engineering, Australia (and 1 more)
- Wei Wang 03[illegible] — Beijing Jiaotong University, Beijing Key Laboratory of Security and Privacy in Intelligent Transportation, China (and 5 more)

The dblp approach



1. Acquisition

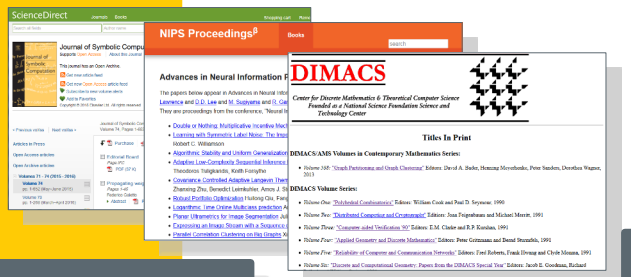
- web crawlers & data submission
- normalization

XML

```
<article key="journals/jsc/Kemper16">
<author>Gregor Kemper</author>
<title>Using extended Derksen ideals in computational
invariant theory.</title>
<pages>161-181</pages>
<year>2016</year>
<volume>72</volume>
<journal>J. Symb. Comput.</journal>
<ee>http://dx.doi.org/10.1016/j.jsc.2015.02.004</ee>
<url>db/journals/jsc/jsc72.html#Kemper16</url>
</article>
```

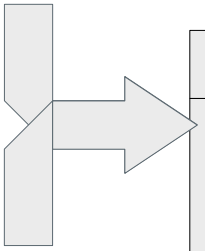
Publisher

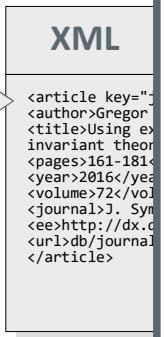
The dblp approach



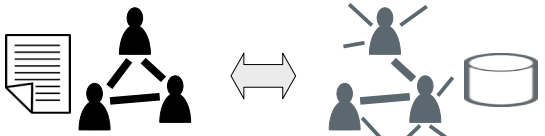
1. Acquisition

- web crawlers & data submission
- normalization



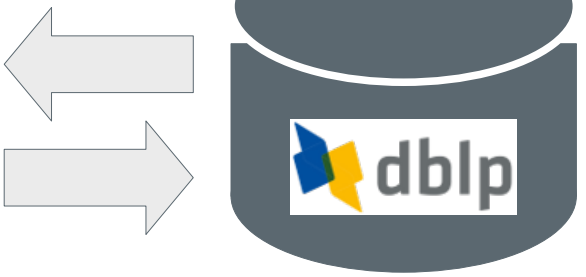


2. Initial metadata curation

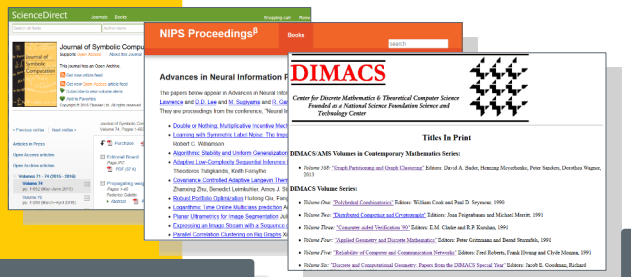


- Initial data cleaning
- **Author disambiguation**
- **Homonym/synonym detection**
- Completing missing data (eg: abbreviated names, missing name parts, etc.)
- Metadata enrichment (eg: homepages, affiliations, PIDs like ORCIDs, etc.)
- Uncover errors in existing data

“authority bootstrapping“
own, curated data stock as authority file

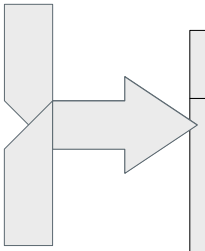


The dblp approach



1. Acquisition

- web crawlers & data submission
- normalization

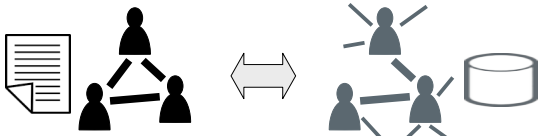


XML

```
<article key="
<author>Gregor
<title>Using e
invariant theor
<pages>161-181<
<year>2016</yea
<volume>72</vol
<journal>J. Sym
<ee>http://dx.d
<url>db/journal
</article>
```

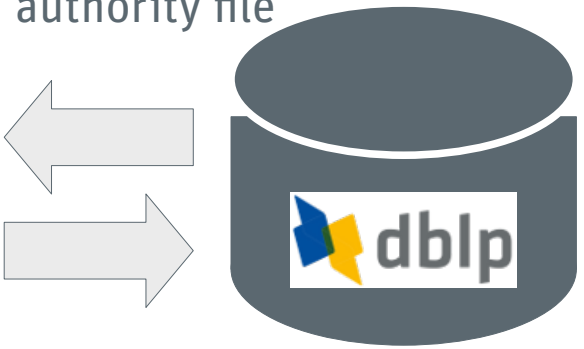
Publisher

2. Initial metadata curation



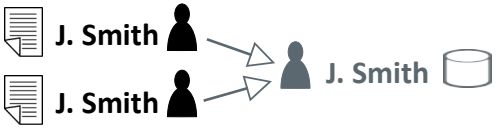
- Initial data cleaning
- **Author disambiguation**
- **Homonym/synonym detection**
- Completing missing data (eg: abbreviated names, missing name parts, etc.)
- Metadata enrichment (eg: homepages, affiliations, PIDs like ORCIDs, etc.)
- Uncover errors in existing data

“authority bootstrapping“
own, curated data stock as authority file



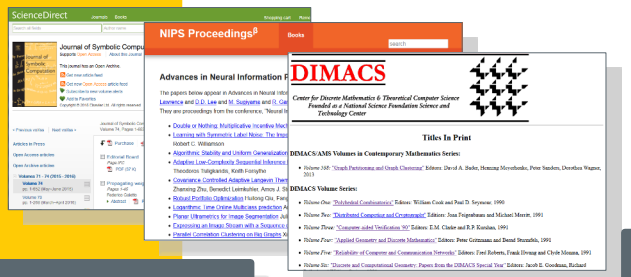
semantic network analysis

3. Continuous quality control



- **Homonym/synonym detection**
- Fixing earlier mistakes in the data stock
- New hints due to newly added metadata

The dblp approach



1. Acquisition

- web crawlers & data submission
- normalization

Publisher

XML

<article key="
<author>Gregor
<title>Using e
invariant theor
<pages>161-181<
<year>2016</yea
<volume>72</vol
<journal>J. Sym
<ee>http://dx.d
<url>db/journal
</article>

2. Initial metadata curation

- Initial data cleaning
- **Author disambiguation**
- **Homonym/synonym detection**
- Completing
  (eg: abb
  missin parts, etc
- Metad ric
  (eg: homepage at
  PIDs like ORCI .)
- Uncover error ting data

"authority bootstrapping"
own, curated data stock as authority file

dblp

semantic network analysis

the human curator in the loop!

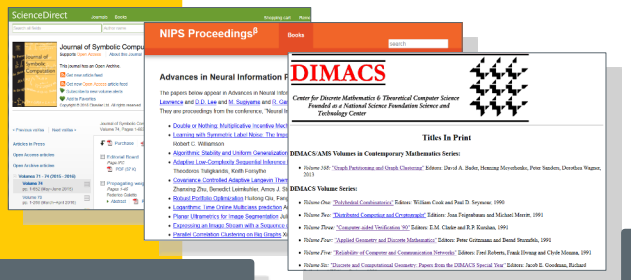
suggestions, not solutions!

3. Continuous quality control

J. Smith
J. Smith
J. Smith

- **Homonym/synonym detection**
  Fixing earlier mistakes in the data stock
  New hints due to newly added metadata

The dblp approach



1. Acquisition

- web crawlers & data submission
- normalization

XML

```
<article key="
<author>Gregor
<title>Using e
invariant theor
<pages>161-181<
<year>2016</yea
<volume>72</vol
<journal>J. Sym
<ee>http://dx.d
<url>db/journal
</article>
```

Publisher

2. Initial metadata curation

- Initial data cleaning
- **Author disambiguation**
- **Homonym/synonym detection**
- Completing
  (eg: abb
  missin parts, etc
- Metad ric
  (eg: homepage
  PIDs like ORCI .)
  Uncover error ting data

“authority bootstrapping“
own, curated data stock as authority file

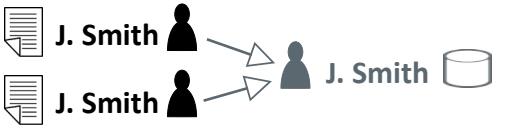
dblp

semantic network analysis

the human curator in the loop!

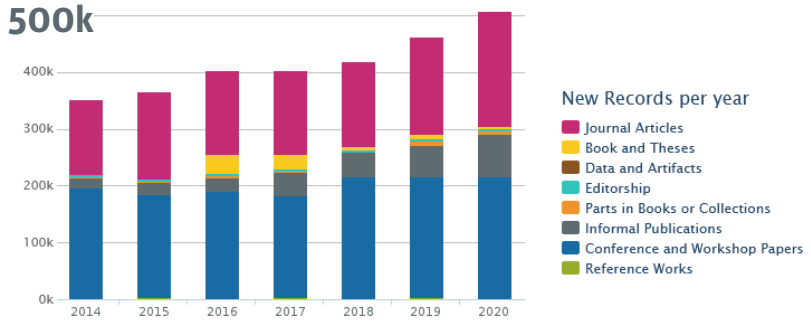
suggestions, not solutions!

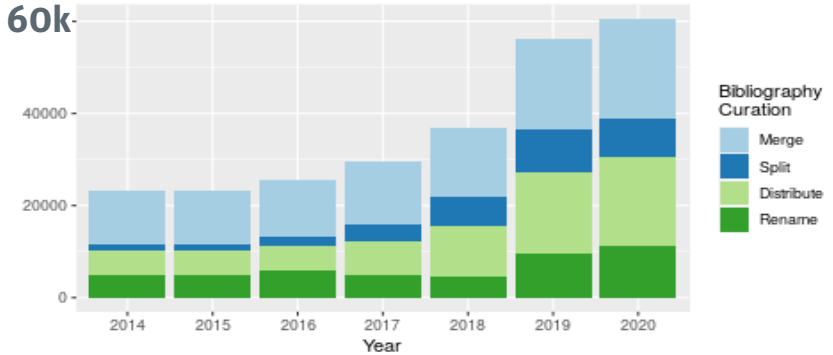
3. Continuous quality control



- **Homonym/synonym detection**
  Fixing earlier mistakes in the data stock
  New hints due to newly added metadata

efficiency: high throughput, small team





Marcel R. Ackermann

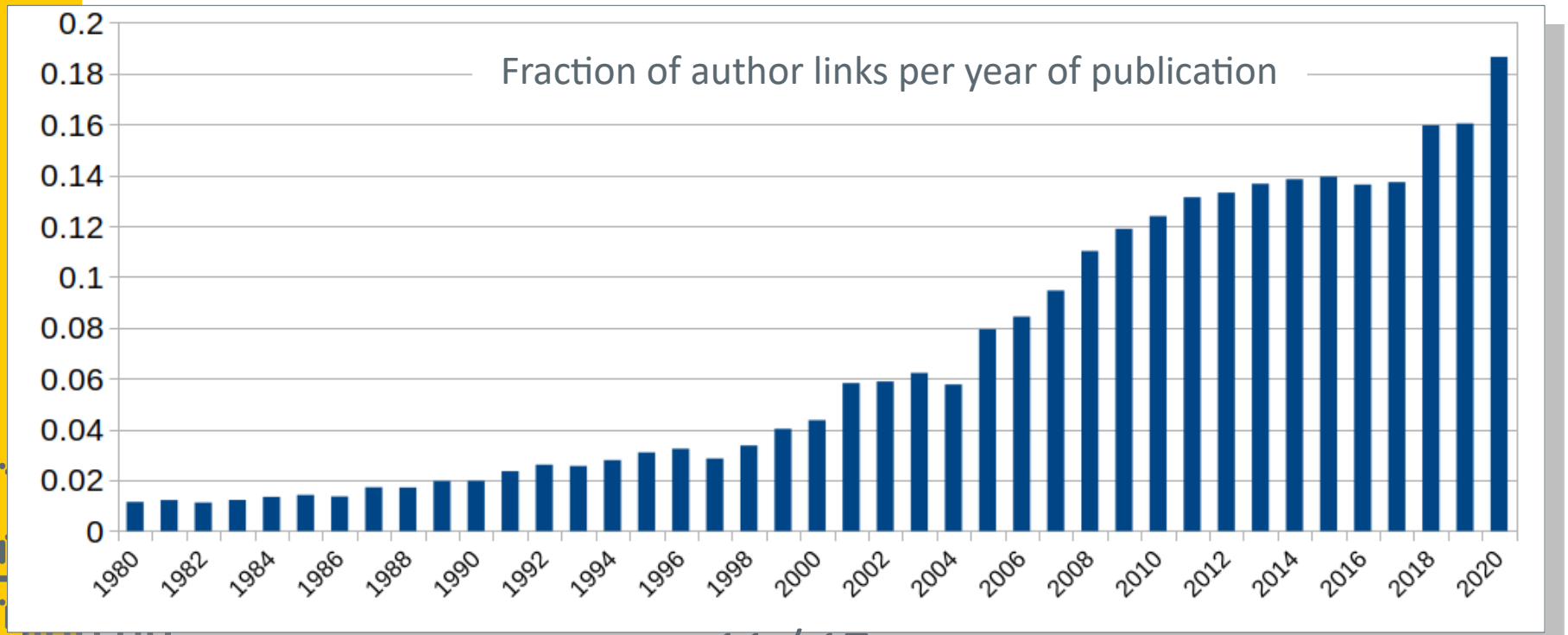
The role of ORCID in dblp (1/3)

- **3 million author links** with ORCID
- for **90k+ bibliographies** manually verified
- very helpful for author disambiguation!

The role of ORCID in dblp (1/3)

- **3 million author links** with ORCID
- for **90k+ bibliographies** manually verified
- very helpful for author disambiguation!

But ...



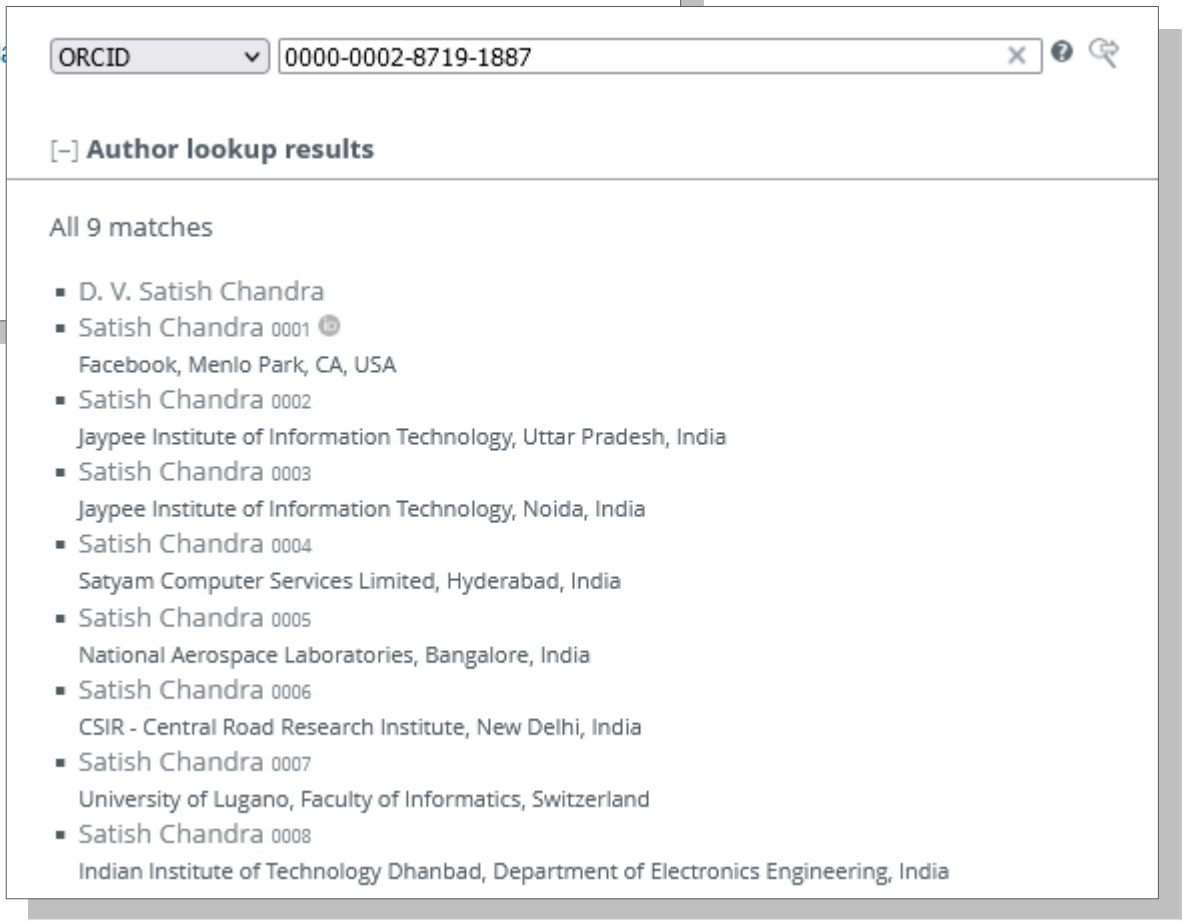
The role of ORCID in dblp (2/3)

Problems created by publisher data

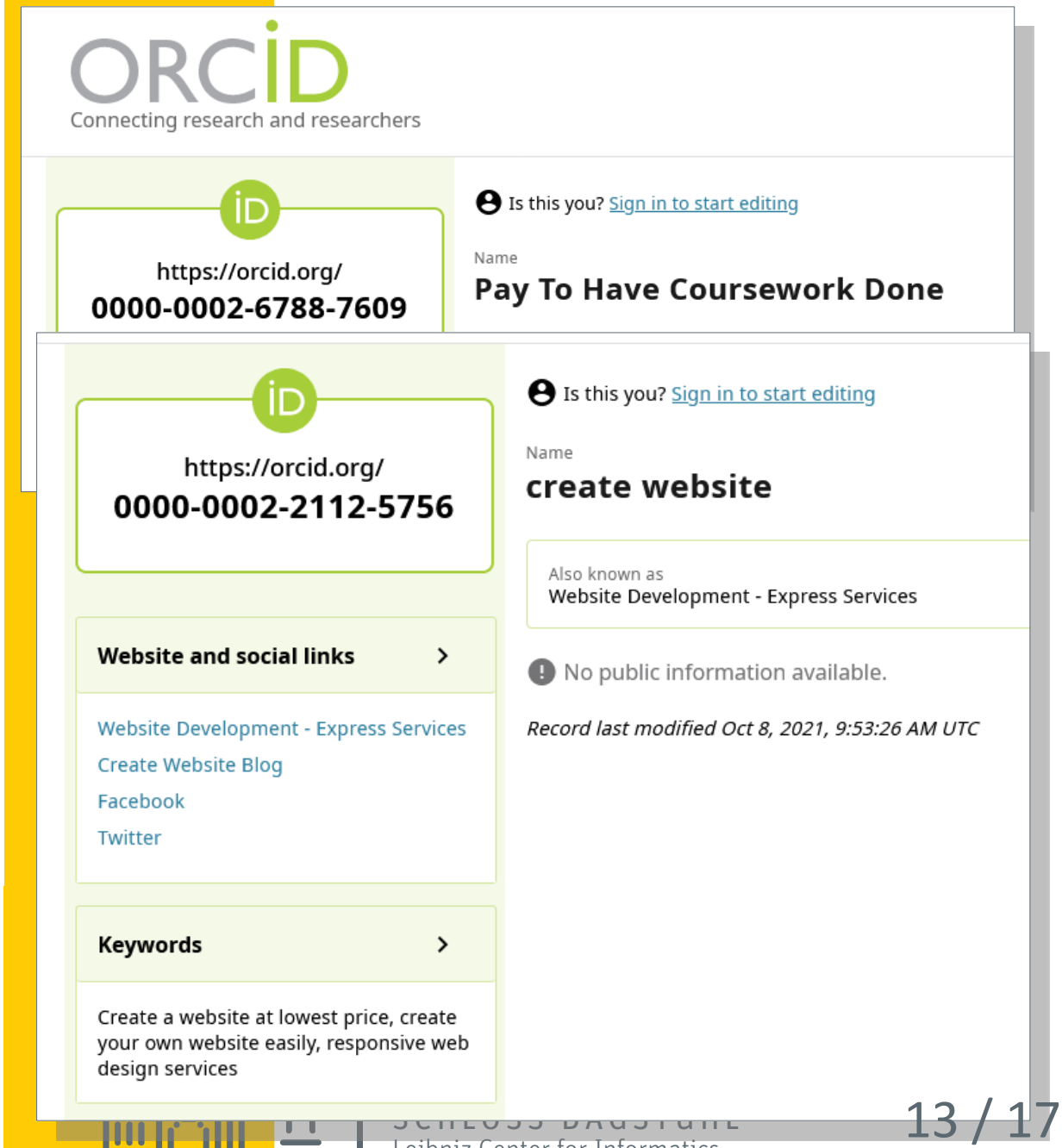
The role of ORCID in dblp (2/3)

Problems created by publisher data

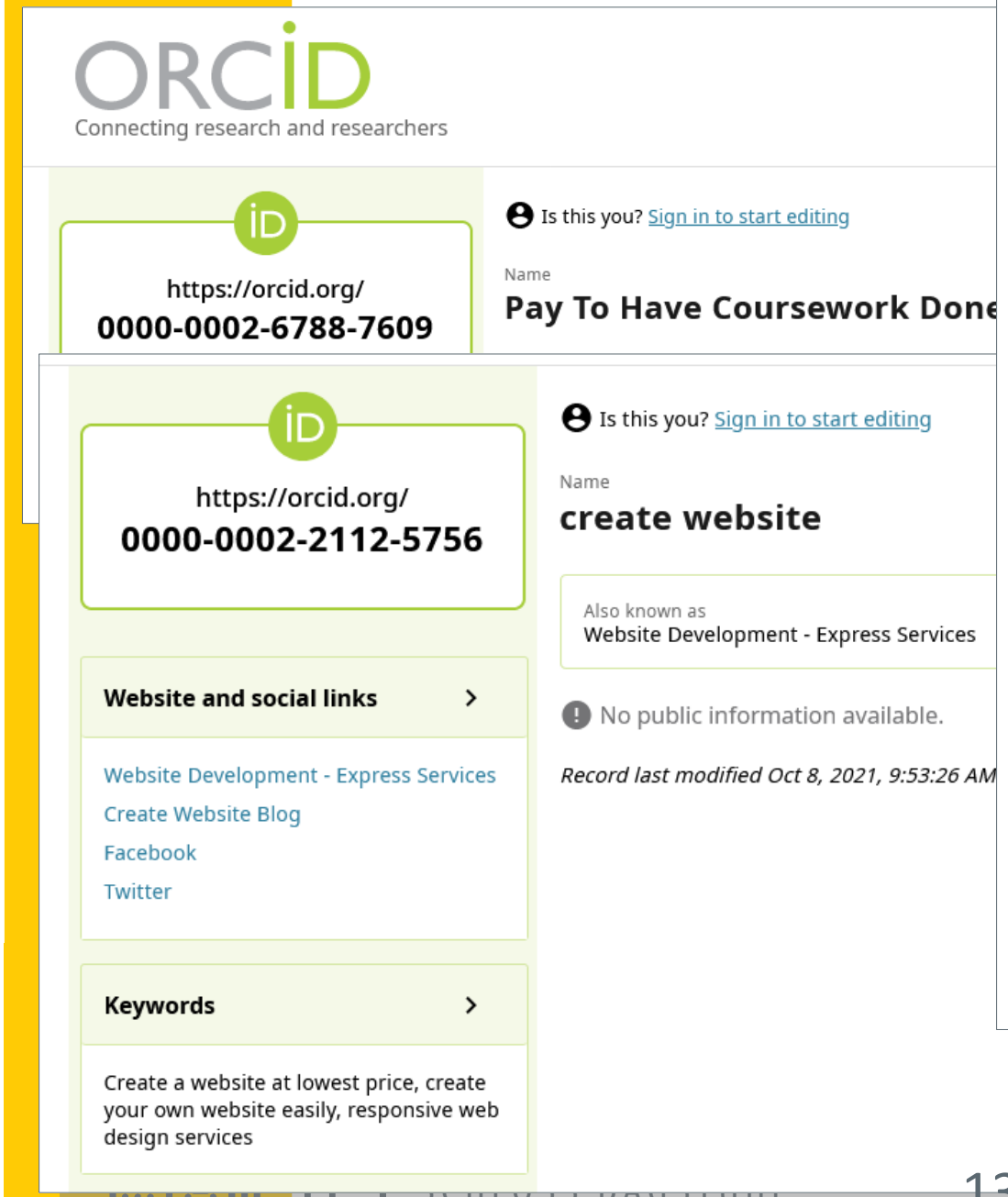
Problems created by ORCID data



The role of ORCID in dblp (3/3)



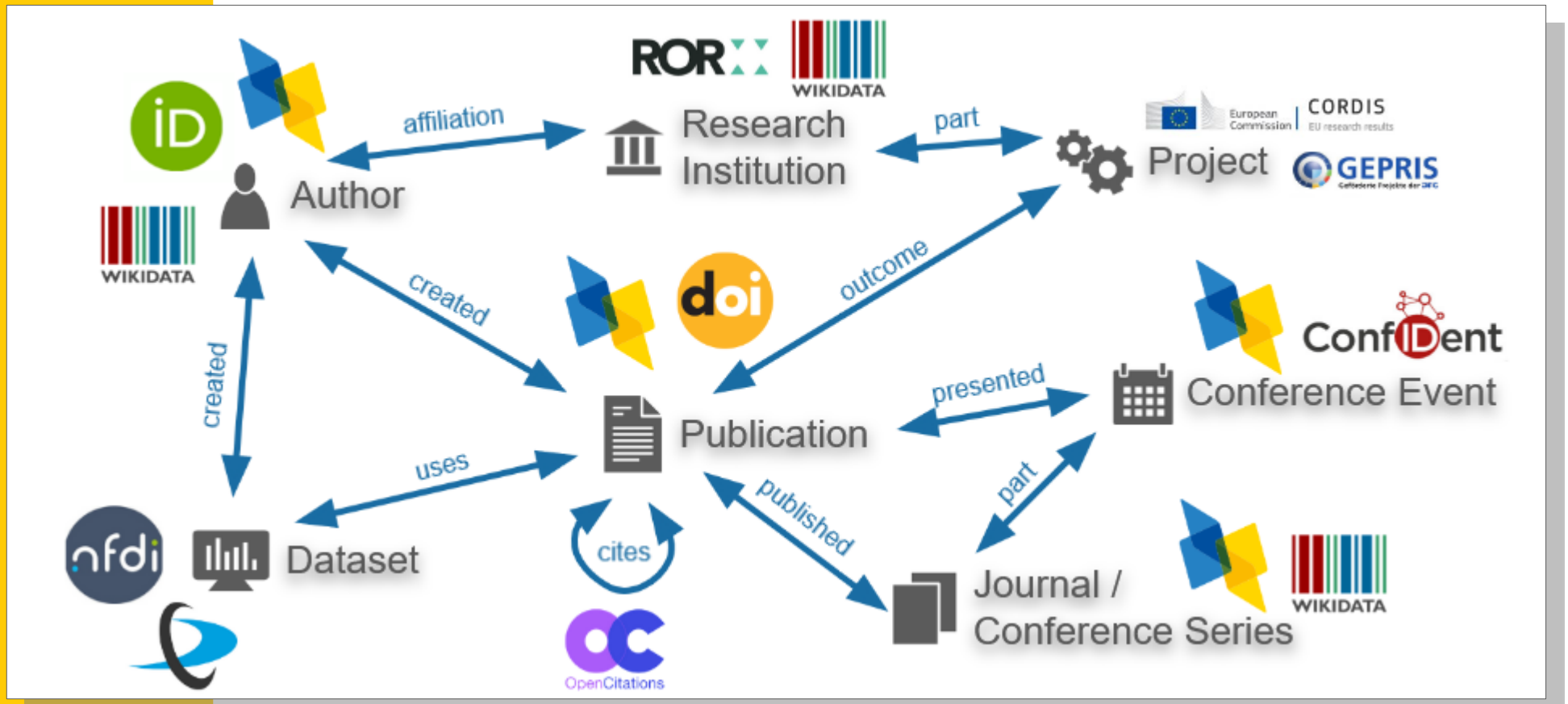
The role of ORCID in dblp (3/3)



ORCID ID	First Name	Last Name
0000-0002-7937-1095	test	test
0000-0003-2524-3853	test	test
0000-0003-2663-416X	test	test
0000-0002-5727-5895	test	test
0000-0002-2673-1763	test	test

ORCID ID	First Name	Last Name
0000-0002-9756-9046	Test	Account
0000-0002-9090-6708	Test	Account
0000-0002-6440-1538	Test564	Account
0000-0002-6646-5754	account	1001
0000-0001-7109-9253	Newsletter	Account
0000-0001-8385-7074	Deleted	Account
0000-0002-3739-2031	Guest	Account
0000-0002-4507-7429	dummy	account
0000-0002-8234-854X	Slick	Account

dblp as a Reserach Knowledge Graph



new: dblp as RDF dump

- **https://dblp.org/rdf/**

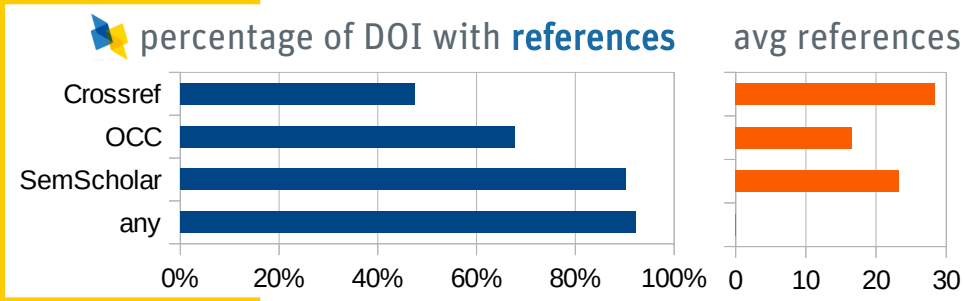
- **250+ million** RDF tripples
- 4.6+ million DOI links
- 525,000+ WikiData links
- 90,000+ ORCID links
- ...

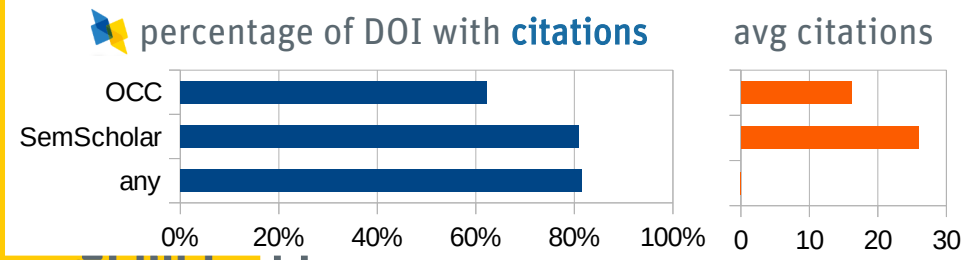
Open Citataions and References

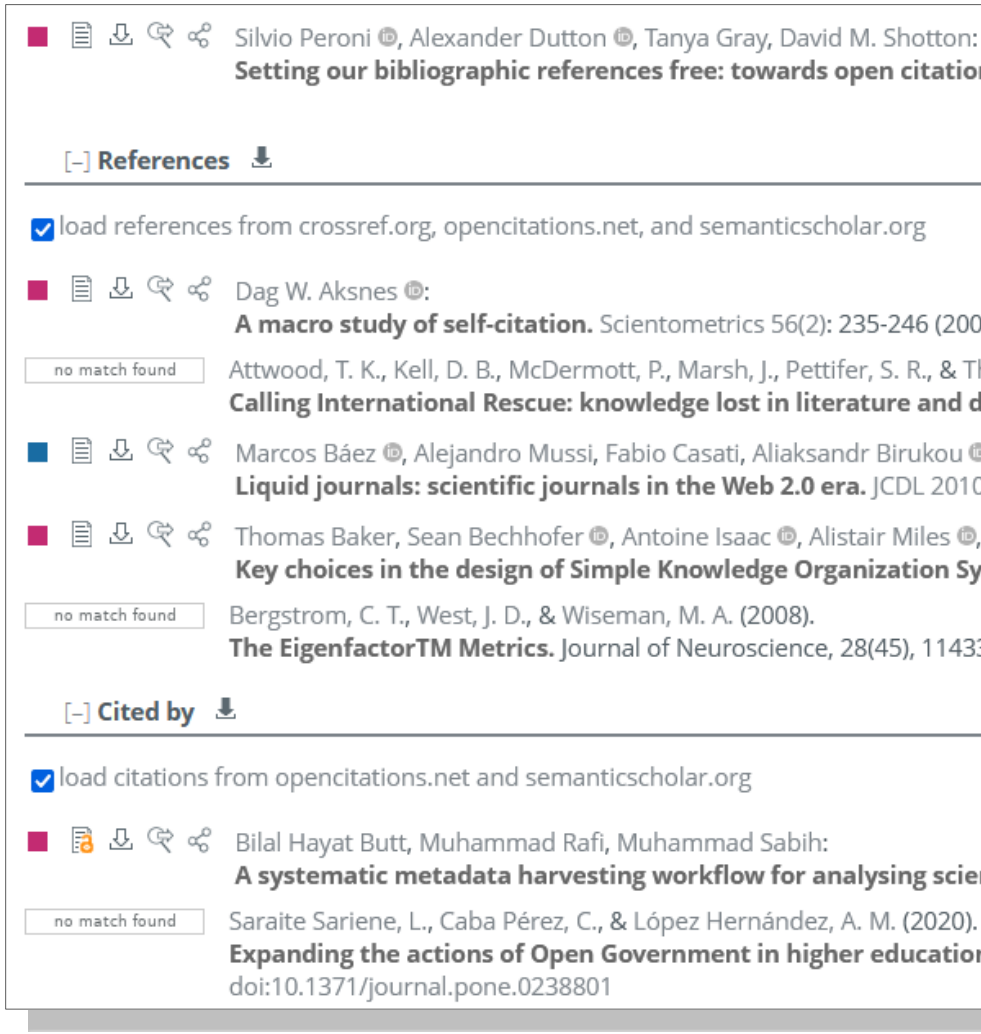
I4OC

https://i4oc.org

- DOI look-up using open APIs
  - Crossref
  - OpenCitations
  - Semantic Scholar





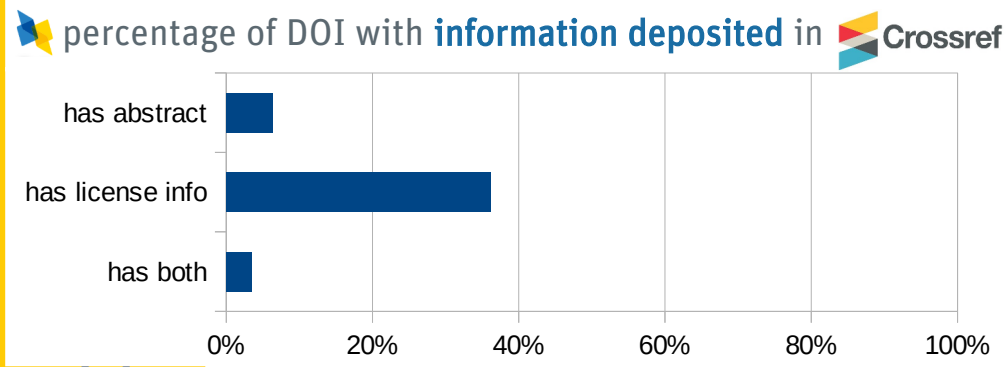


Open Abstracts

https://i4oa.org

ABS TRA CTS

- General limitation:
  - subject to copyright laws!
- Call on publishers to open the abstracts
  - machine-accessible
  - text and data mining



Thanks!

https://dblp.org

✉ dblp@dagstuhl.de
@dblp_org