



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)

The screenshot shows the dblp website interface. At the top, there is a navigation bar with links for 'home', 'browse', 'search', and 'about'. The main header features the dblp logo (computer science bibliography) and the Schloss Dagstuhl Leibniz Center for Informatics logo. A search bar contains the text 'open data'. Below the search bar, there is a dark grey banner with the text 'Search dblp for Publications' and 'powered by CompleteSearch, courtesy of Hannah Bast, University of Freiburg'. The main content area shows 'Publication search results' with a download icon and 'found 6,749 matches'. A 'Refine list' section on the right includes a bar chart showing the number of publications from 1962 to 2022, with a peak of 900 in 2022. The search results list two publications: one from 2022 by Yang Yang, Hongchen Wei, Zhen-Qiang Sun, Guang-Yu Li, Yuanchun Zhou, Hui Xiong, and Jian Yang, and one from 2021 by Matthias Frank, titled 'Knowledge-Driven Harmonization of Sensor Observations: Exploiting Linked Open Data for IoT Data Streams'. The authors of the 2021 publication are listed as Ansgar Scherp (19), Sören Auer (19), Sebastian Neumaier (18), and Silvio Peroni (18), with 18,978 more options.

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 **CCO license**

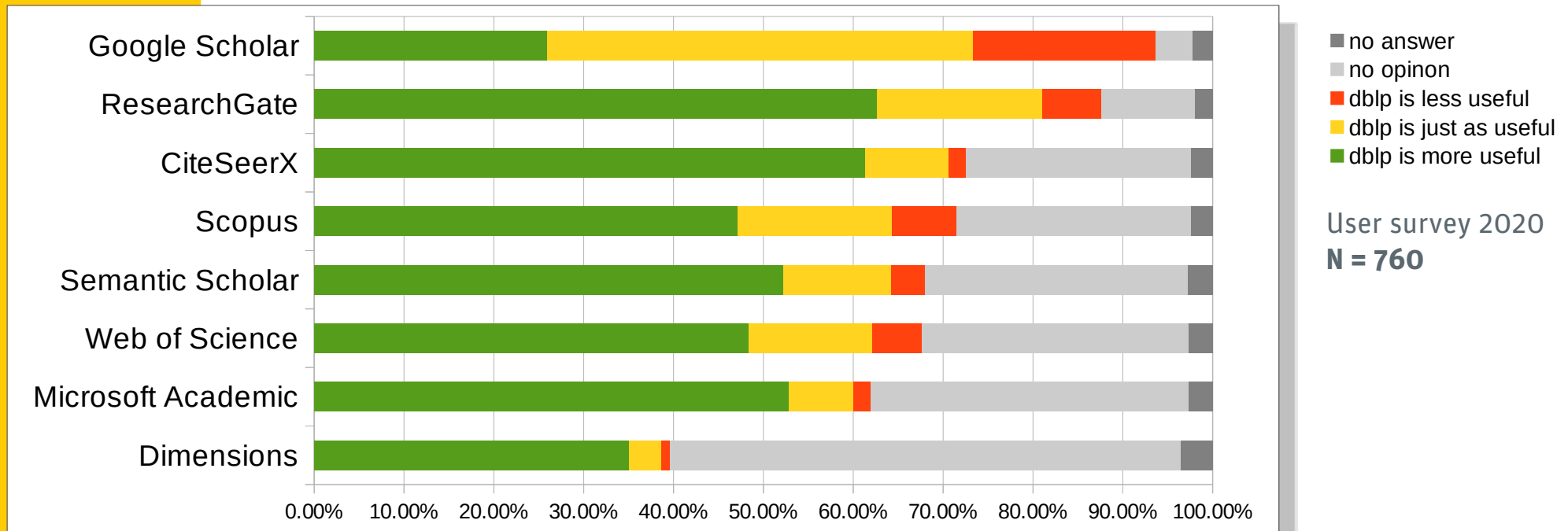
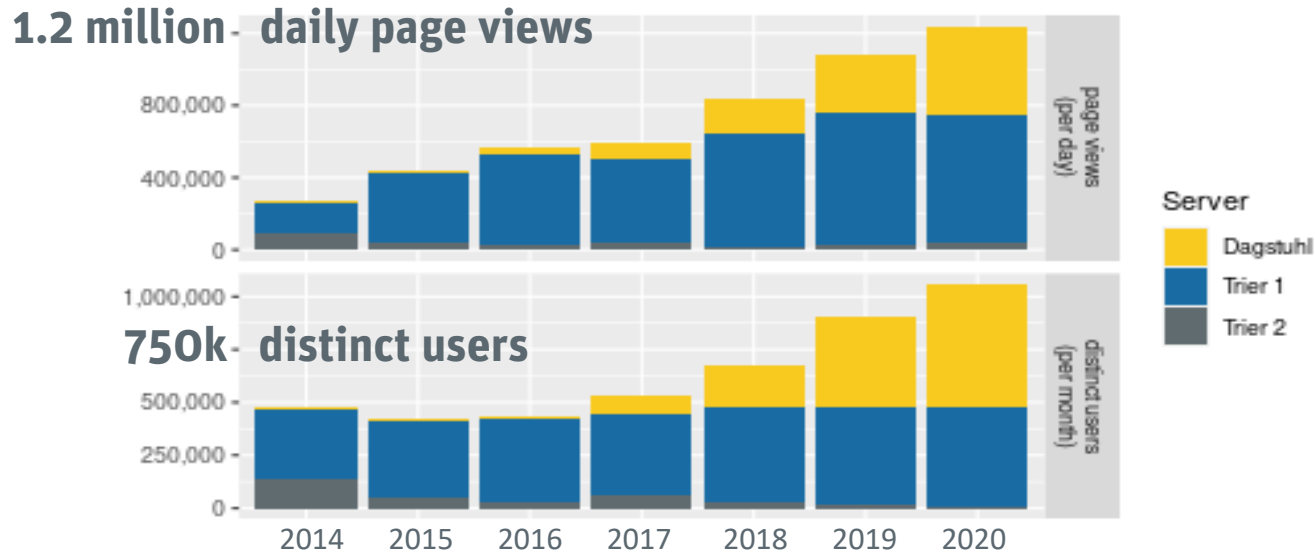


Our key principles

- **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



Author Disambiguation (1/2)

▪ Homonyms

Introduction of the Mitrack 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 Smart¹✉, Jia Du^{1,*}✉, Li Li¹✉, David Wang¹✉, Keith Leslie¹✉, Fan Ji²✉

**A hidden Markov model for detecting recombination
al detection microarrays**

Hilary Renshaw^{1,2}, Scott Weaver^{3,4,5}, Robert B. Tesh³ and David Wang¹

**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 [i] [d] [e] [s] [m]

> 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 [i] — 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 0012 — Beijing Jiaotong University, Beijing Key Laboratory of Security and Privacy in Intelligent Transportation, China (and 5 more)
- Wei Wang 0013 — Beijing Jiaotong University, Institute of Computational Linguistics, Beijing, China
- Wei Wang 0309 [i] — University, Nano Lab, Medford, MA, USA (and 1 more)
- Wei Wang 0310 — University, Institute of Intelligent Vehicles, School of Automotive Studies, Shanghai, China
- Wei Wang 0311 — University of Technology, School of Control Science and Engineering, China
- Wei Wang 0312 [i] — Jiao Tong University, School of Mathematical Sciences, China
- Wei Wang 0313 [i] — University of Technology, School of Computer and Information, China
- Wei Wang 0314 [i] — University of California, Palo Alto, CA, USA
- Wei Wang 0315 — Department of Aeronautics and Astronautics, Shanghai, China
- Wei Wang 0316 — Department of Mathematics and Physics, China

The dblp approach

1. Acquisition

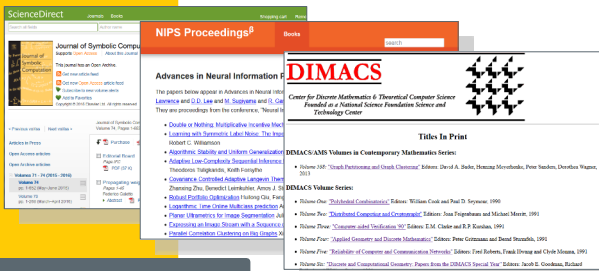
- web crawlers & data submission
- normalization

Publisher

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>
```

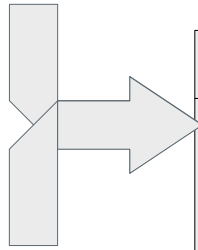




The dblp approach

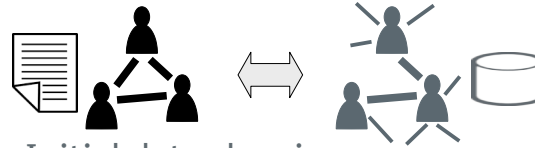
1. Acquisition

- web crawlers & data submission
- normalization



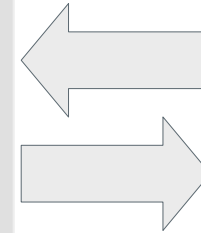
```
XML
<article key="1"
  <author>Gregor
  <title>Using ex
  <pages>161-181<
  <year>2016</yea
  <volume>72</vol
  <journal>J. Syn
  <ee>http://dx.d
  <url>db/journal
</article>
```

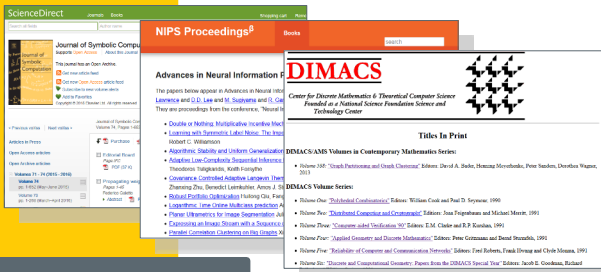
2. Initial metadata curation



- 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





1. Acquisition

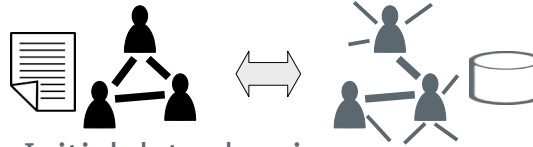
- web crawlers & data submission
- normalization

Publisher

XML

```
<article key="1"
  <author>Gregor
  <title>Using ex
  <pages>161-181<
  <year>2016</yea
  <volume>72</vol
  <journal>J. Syn
  <ee>http://dx.d
  <url>db/journal
  </article>
```

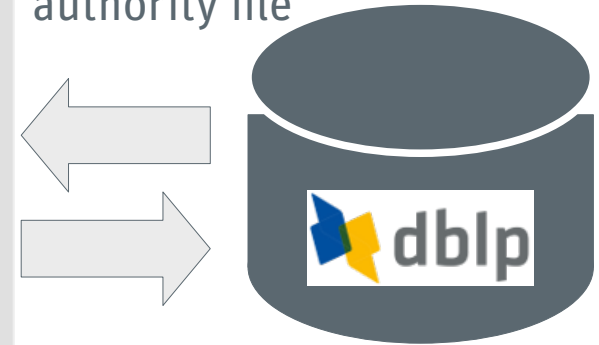
2. Initial metadata curation



- 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 ORCID, etc.)
- Uncover errors in existing data

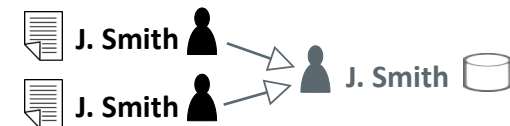
The dblp approach

“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

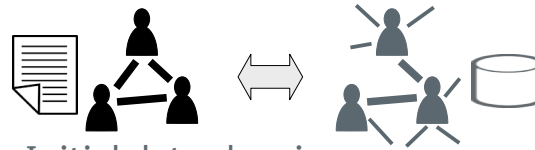
- web crawlers & data submission
- normalization

Publisher

XML

```
<article key="1"
  <author>Gregor
  <title>Using ex
  <pages>161-181<
  <year>2016</yea
  <volume>72</vol
  <journal>J. Syn
  <ee>http://dx.d
  <url>db/journal
  </article>
```

2. Initial metadata curation



- Initial data cleaning
- **Author disambiguation**
- **Homonym/synonym detection**
- Completing metadata (eg: abbreviations, missing parts, etc)
- Metadata enrichment (eg: homepage, ORCID PIDs like ORCID, etc)
- Uncover errors in existing data

the human curator

in the loop!

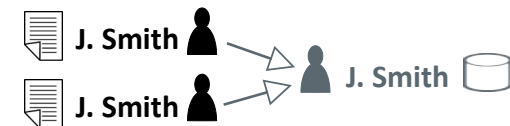
suggestions,
not solutions!

“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

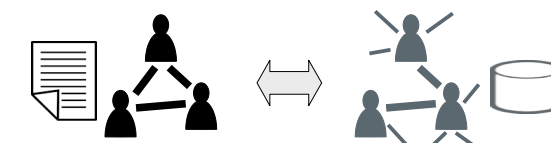
- web crawlers & data submission
- normalization

Publisher

XML

```
<article key="1"
  <author>Gregor
  <title>Using ex
  <pages>161-181<
  <year>2016</yea
  <volume>72</vol
  <journal>J. Syn
  <ee>http://dx.d
  <url>db/journal
  </article>
```

2. Initial metadata curation



- Initial data cleaning
- Author disambiguation**
- Homonym/synonym detection**
- Completing missing parts, etc
- Metadata enrichment (eg: homepage, ORCID, PIDs like ORCID, etc)
- Uncover errors in existing data

the human curator

in the loop!

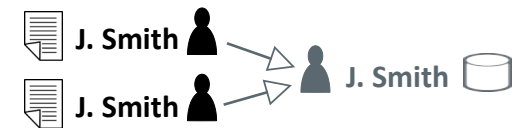
suggestions, not solutions!

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



semantic network analysis

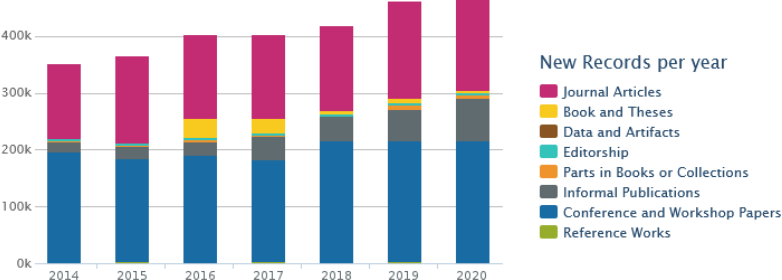
3. Continuous quality control



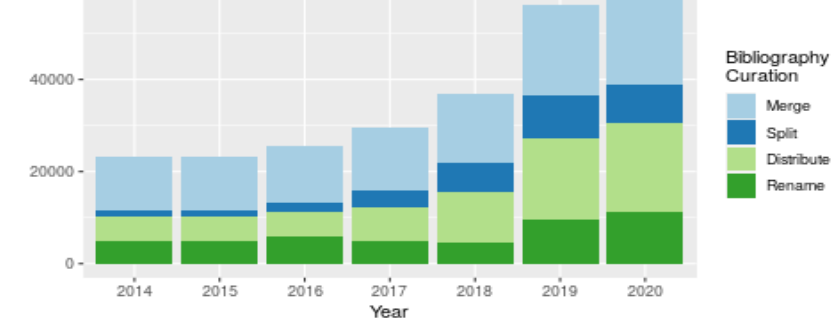
- Homonym/synonym detection**
- Fixing earlier mistakes in the data stock
- New hints due to newly added metadata

efficiency: high throughput, small team

500k



60k



The role of ORCID in dblp (1/3)



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

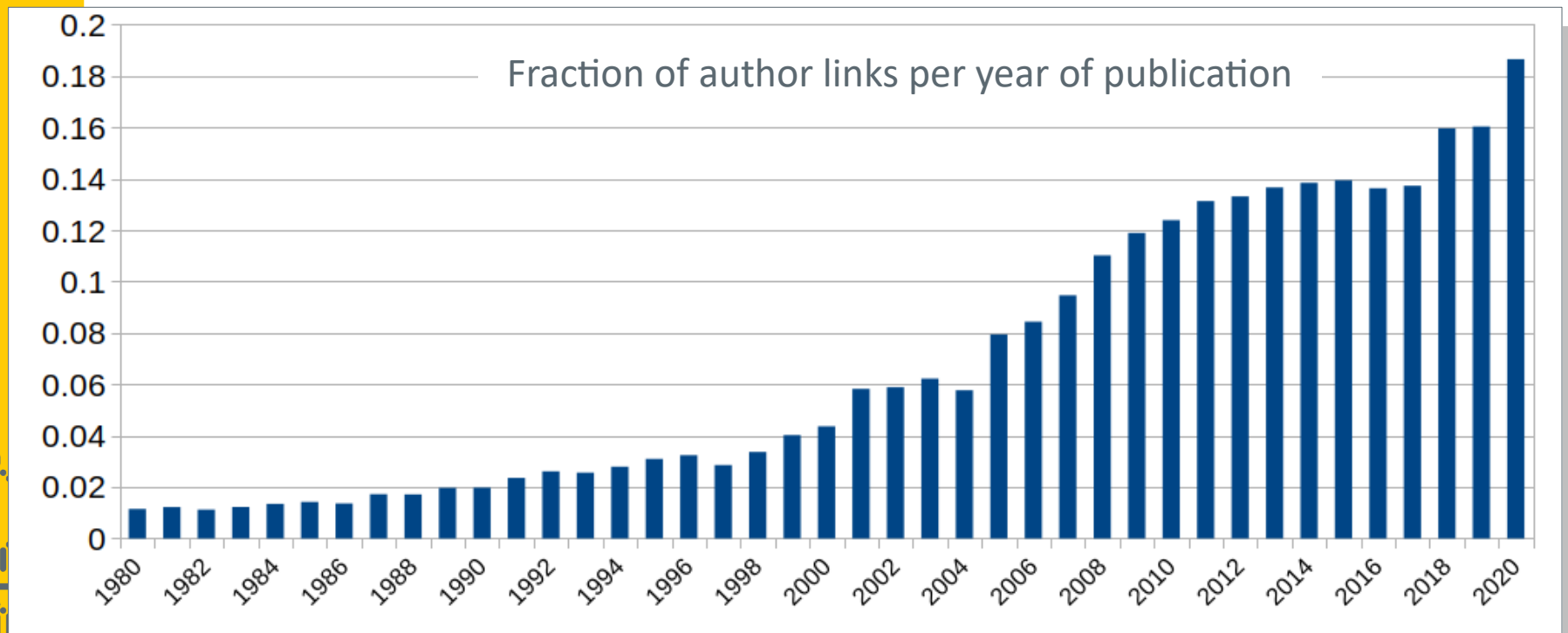


The role of ORCID in dblp (1/3)



- **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)

IEEE Access Special Section Editorial: Big Data Analytics for Smart and Connected Health

Publisher: IEEE

Cite This

PDF

Yuan Zhang  ; Lin Zhang  ; Eiji Oki ; Nitesh V. Chawla ; Anton Kos [All Authors](#)

2
Paper
Citations

1193
Full
Text

 0000-0002-0819-1644

 0000-0002-0819-1644



Problems created by publisher data



The role of ORCID in dblp (2/3)

IEEE Access Special Section Editorial: Big Data Analytics for Smart and Connected Health

Publisher: IEEE

Cite This

PDF

Yuan Zhang  ; Lin Zhang  ; Eiji Oki ; Nitesh V. Chawla

2
Paper
Citations

1193
Full
Text

 0000-0002-0819-1644

 0000-0002-0819-1644

Problems created
by ORCID data

ORCID 0000-0002-8719-1887

[-] Author lookup results

All 9 matches

- D. V. Satish Chandra
- Satish Chandra 0001 
Facebook, Menlo Park, CA, USA
- Satish Chandra 0002
Jaypee Institute of Information Technology, Uttar Pradesh, India
- Satish Chandra 0003
Jaypee Institute of Information Technology, Noida, India
- Satish Chandra 0004
Satyam Computer Services Limited, Hyderabad, India
- Satish Chandra 0005
National Aerospace Laboratories, Bangalore, India
- Satish Chandra 0006
CSIR - Central Road Research Institute, New Delhi, India
- Satish Chandra 0007
University of Lugano, Faculty of Informatics, Switzerland
- Satish Chandra 0008
Indian Institute of Technology Dhanbad, Department of Electronics Engineering, India

Problems created by publisher data



The role of ORCID in dblp (3/3)

The screenshot shows two overlapping ORCID profile cards. The top card is for the user 'Pay To Have Coursework Done' with ORCID ID '0000-0002-6788-7609'. The bottom card is for the user 'create website' with ORCID ID '0000-0002-2112-5756'. The bottom card's profile is more detailed, showing a 'Website and social links' section with links to 'Website Development - Express Services', 'Create Website Blog', 'Facebook', and 'Twitter'. It also has a 'Keywords' section with the text 'Create a website at lowest price, create your own website easily, responsive web design services'. A notice states 'No public information available' and 'Record last modified Oct 8, 2021, 9:53:26 AM UTC'. The ORCID logo and tagline 'Connecting research and researchers' are visible at the top of the cards.

The role of ORCID in dblp (3/3)

ORCID
Connecting research and researchers

iD
https://orcid.org/
0000-0002-6788-7609

Is this you? [Sign in to start editing](#)

Name
Pay To Have Coursework Done

iD
https://orcid.org/
0000-0002-2112-5756

Is this you? [Sign in to start editing](#)

Name
create website

Also known as
Website Development - Express Services

Website and social links >

- Website Development - Express Services
- Create Website Blog
- Facebook
- Twitter

Keywords >

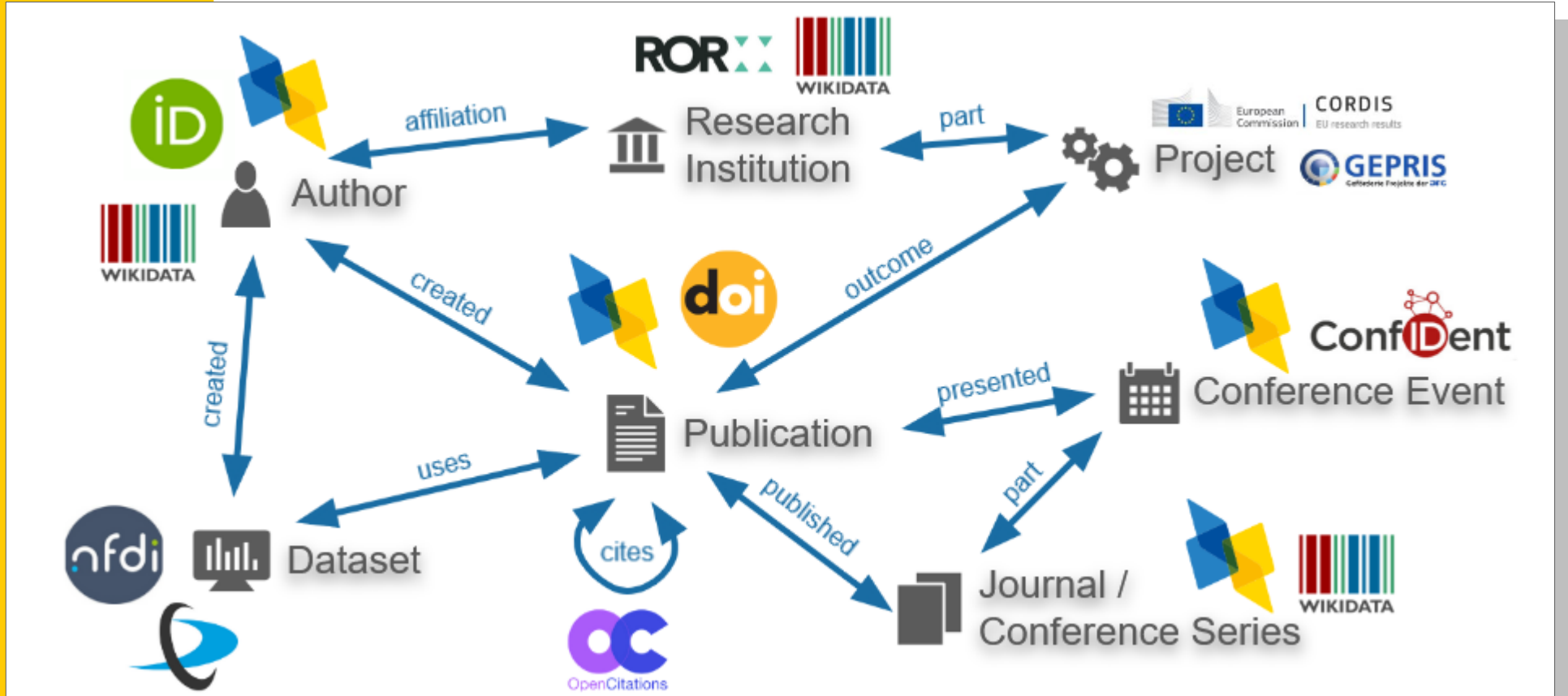
Create a website at lowest price, create your own website easily, responsive web design services

No public information available.
Record last modified Oct 8, 2021, 9:53:26 AM

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/>

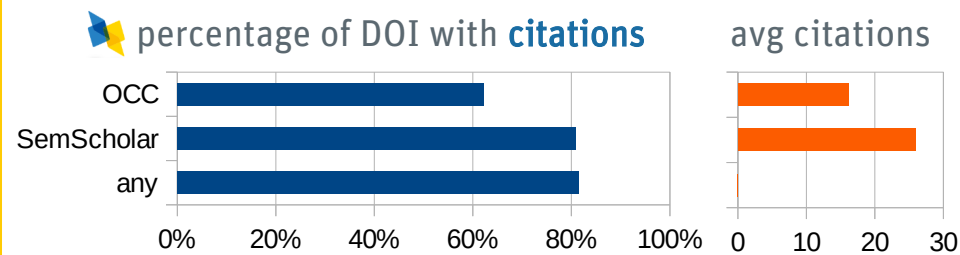
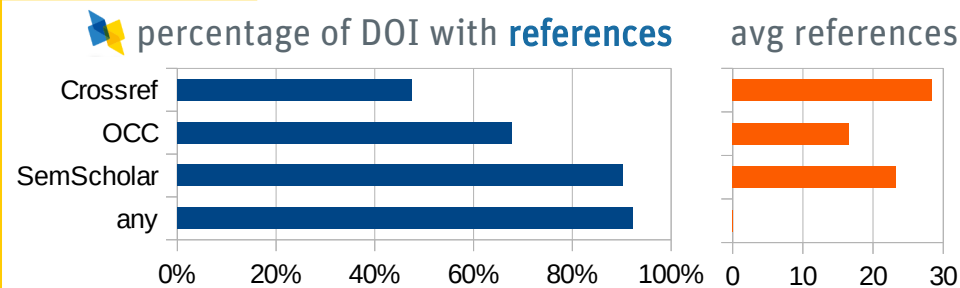
- **250+ million** RDF tripples
- 4.6+ million DOI links
- 525,000+ WikiData links
- 90,000+ ORCID links
- ...










Open Citations and References

- DOI look-up using open APIs






-  Crossref
-  OpenCitations
-  Semantic Scholar




















 Silvio Peroni , Alexander Dutton , Tanya Gray, David M. Shotton:
Setting our bibliographic references free: towards open citation

[-] References 


load references from crossref.org, opencitations.net, and semanticscholar.org





 Dag W. Aksnes :
A macro study of self-citation. *Scientometrics* 56(2): 235-246 (2002)









 Marcos Báez , Alejandro Mussi, Fabio Casati, Aliaksandr Birukou :
Liquid journals: scientific journals in the Web 2.0 era. *JCDL* 2010





 Thomas Baker, Sean Bechhofer , Antoine Isaac , Alistair Miles :
Key choices in the design of Simple Knowledge Organization Systems

Bergstrom, C. T., West, J. D., & Wiseman, M. A. (2008).
The Eigenfactor™ Metrics. *Journal of Neuroscience*, 28(45), 11433-11440

[-] Cited by 

load citations from opencitations.net and semanticscholar.org





 Bilal Hayat Butt, Muhammad Rafi, Muhammad Sabih:
A systematic metadata harvesting workflow for analysing scientific literature

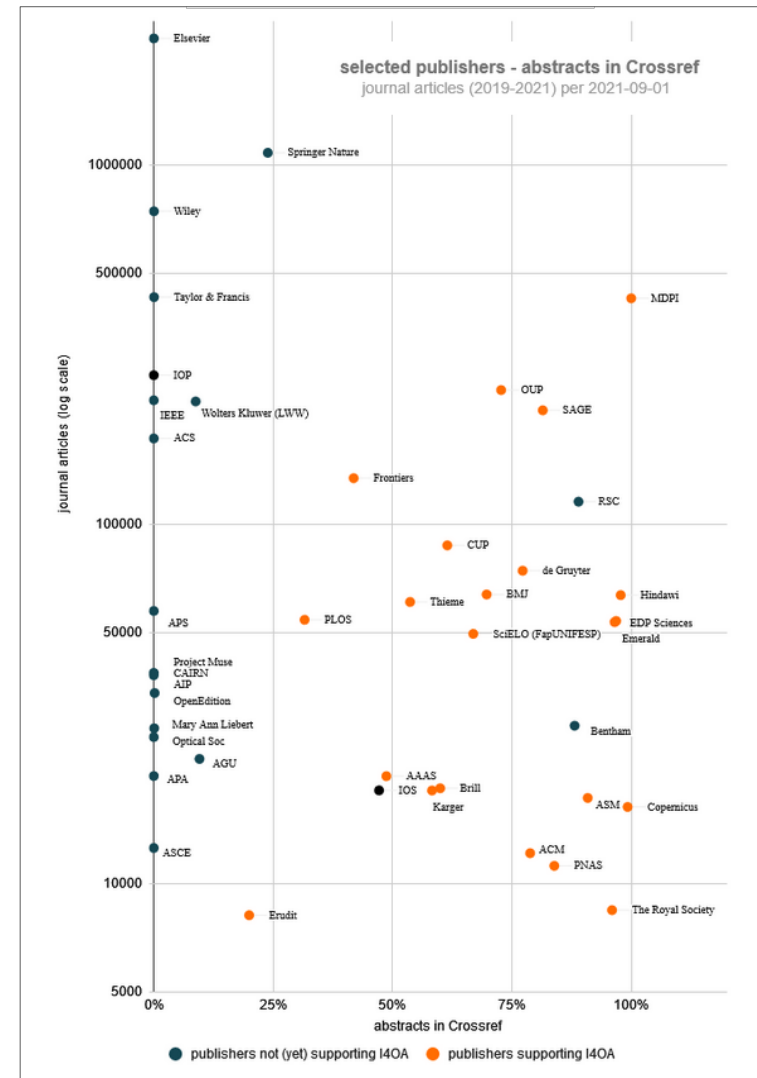
Saraite Sariene, L., Caba Pérez, C., & López Hernández, A. M. (2020).
Expanding the actions of Open Government in higher education
 doi:10.1371/journal.pone.0238801



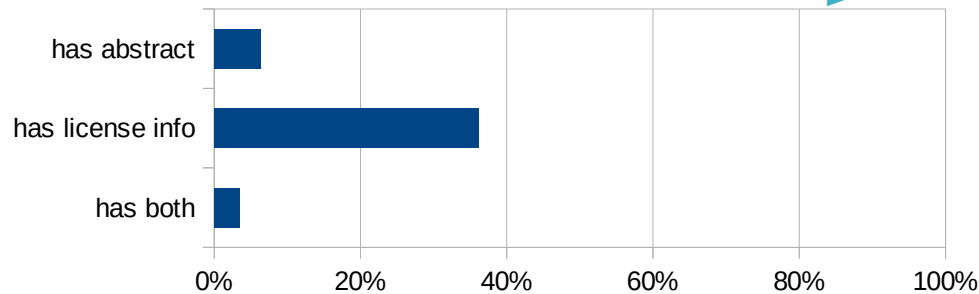


Open Abstracts

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



percentage of DOI with information deposited in Crossref



Thanks!

<https://dblp.org>

✉ dblp@dagstuhl.de

🐦 [@dblp_org](https://twitter.com/dblp_org)

