LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY

TIB

Findable

Angelina Kraft, Katrin Leinweber

TIB, 9. July 2018 Recording: doi.org/10.5446/37823

FAIR Data & Software (Carpentries-based workshop) #TIBFDS

Leibniz
Association

to be Findable

F1. (meta)data are assigned a globally unique and eternally persistent identifier

F2. data are described with rich metadata

F3. (meta)data are registered or indexed in a searchable resource

F4. metadata specify the data identifier

Your institution’s / repository’s role

Findable

- assign a globally unique PID upon publication (or draft upload)
- provide metadata schema in human- & machine-readable format
  - PID, author names, subject areas, etc.
  - support structured input of metadata (submission forms or XML schema)
  - index (meta)data to enable effective searching
  - allow metadata upload & assign corresponding PID

Your role as a scientist

Findable

- check datasets that you use for a PID & cite it
- ensure that your datasets get published with a PID
  - choose repositories that automate this
  - report this requirement to repos that don’t
- add rich metadata (describe dataset’s context, quality, condition & characteristics)
  - should be understandable by researchers from different discipline (ask a friend to proofread)

Example of paper - data citation using PIDs

Paper:

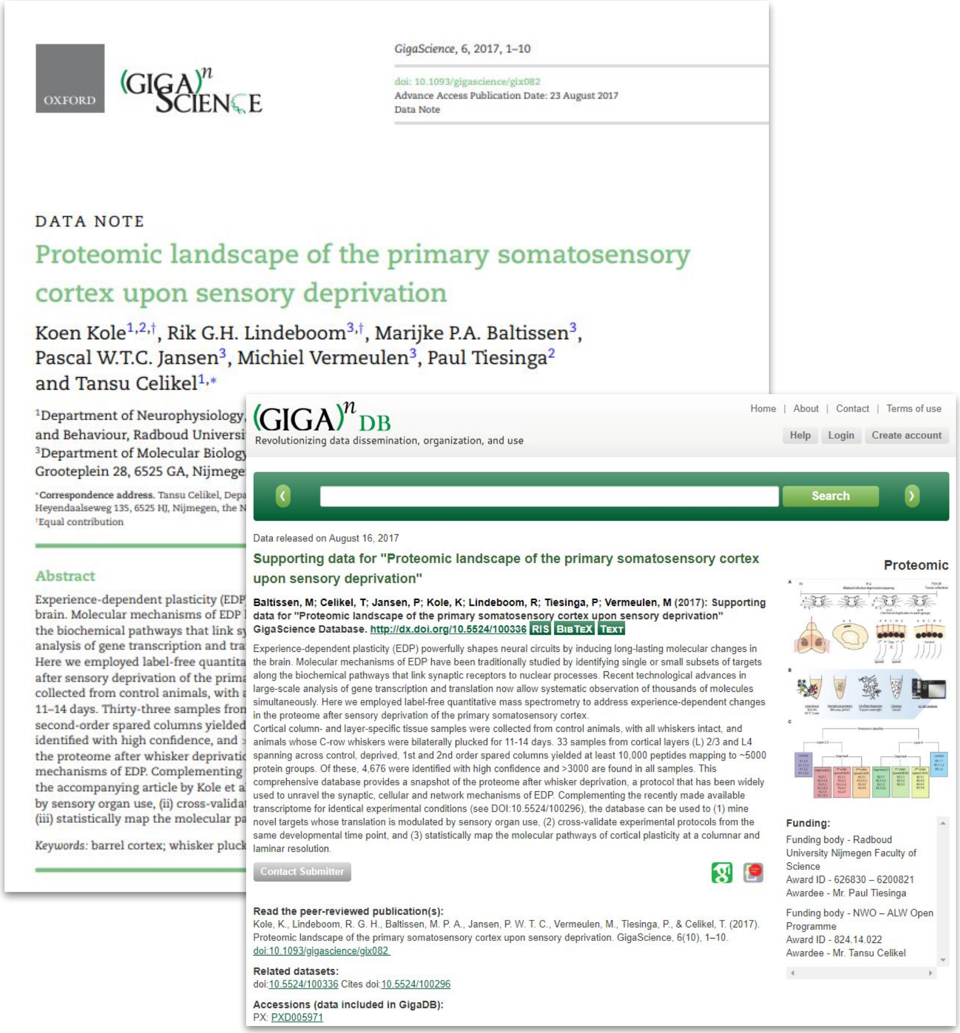
Koen Kole, Rik G.H. Lindeboom, Marijke P.A. Baltissen, Pascal W.T.C. Jansen, Michiel Vermeulen, Paul Tiesinga, Tansu Celikel (2017):
Proteomic landscape of the primary somatosensory cortex upon sensory deprivation, *GigaScience*, Volume 6, Issue 10, 1 October 2017, Pages 1–10. DOI https://doi.org/10.1093/gigascience/gix082

Note in the paper:

"**Availability of the supporting data**
Data supporting this work are available in the *GigaScience* repository, *Giga*DB [14]. The raw mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium via the PRIDE partner repository [15] with the dataset identifier PXD005971"

Reference:

[14] Kole K, Baltissen M, Lindeboom R et al. Supporting data for "Proteomic landscape of the primary somatosensory cortex upon sensory deprivation." GigaScience Database 2017. http://doi.org/10.5524/100336



GigaScience, 6, 2017, 1–10

doi: 10.1093/gigascience/gix082
Advance Access Publication Date: 23 August 2017
Data Note

DATA NOTE

Proteomic landscape of the primary somatosensory cortex upon sensory deprivation

Koen Kole[1,2,†], Rik G.H. Lindeboom[3,†], Marijke P.A. Baltissen[3], Pascal W.T.C. Jansen[3], Michiel Vermeulen[3], Paul Tiesinga[2] and Tansu Celikel[1,*]

[1]Department of Neurophysiology,
and Behaviour, Radboud Universi
[3]Department of Molecular Biology
Grooteplein 28, 6525 GA, Nijmege

*Correspondence address. Tansu Celikel, Depa
Heyendaalseweg 135, 6525 HJ, Nijmegen, the N
†Equal contribution

Abstract

Experience-dependent plasticity (EDP
brain. Molecular mechanisms of EDP
the biochemical pathways that link sy
analysis of gene transcription and tra
Here we employed label-free quantita
after sensory deprivation of the prima
collected from control animals, with a
11–14 days. Thirty-three samples from
second-order spared columns yielded
identified with high confidence, and >
the proteome after whisker deprivatio
mechanisms of EDP. Complementing
the accompanying article by Kole et a
by sensory organ use, (ii) cross-validat
(iii) statistically map the molecular pa

Keywords: barrel cortex; whisker pluck

(GIGA)^n DB
Revolutionizing data dissemination, organization, and use
Home | About | Contact | Terms of use
Help
Login
Create account
Search
Proteomic
Contact Submitter

Data released on August 16, 2017

Supporting data for "Proteomic landscape of the primary somatosensory cortex upon sensory deprivation"

Baltissen, M; Celikel, T; Jansen, P; Kole, K; Lindeboom, R; Tiesinga, P; Vermeulen, M (2017): Supporting data for "Proteomic landscape of the primary somatosensory cortex upon sensory deprivation" GigaScience Database. http://dx.doi.org/10.5524/100336 RIS BIBTEX TEXT

Experience-dependent plasticity (EDP) powerfully shapes neural circuits by inducing long-lasting molecular changes in the brain. Molecular mechanisms of EDP have been traditionally studied by identifying single or small subsets of targets along the biochemical pathways that link synaptic receptors to nuclear processes. Recent technological advances in large-scale analysis of gene transcription and translation now allow systematic observation of thousands of molecules simultaneously. Here we employed label-free quantitative mass spectrometry to address experience-dependent changes in the proteome after sensory deprivation of the primary somatosensory cortex.
Cortical column- and layer-specific tissue samples were collected from control animals, with all whiskers intact, and animals whose C-row whiskers were bilaterally plucked for 11-14 days. 33 samples from cortical layers (L) 2/3 and L4 spanning across control, deprived, 1st and 2nd order spared columns yielded at least 10,000 peptides mapping to ~5000 protein groups. Of these, 4,676 were identified with high confidence and >3000 are found in all samples. This comprehensive database provides a snapshot of the proteome after whisker deprivation, a protocol that has been widely used to unravel the synaptic, cellular and network mechanisms of EDP. Complementing the recently made available transcriptome for identical experimental conditions (see DOI:10.5524/100296), the database can be used to (1) mine novel targets whose translation is modulated by sensory organ use, (2) cross-validate experimental protocols from the same developmental time point, and (3) statistically map the molecular pathways of cortical plasticity at a columnar and laminar resolution.

Read the peer-reviewed publication(s):
Kole, K., Lindeboom, R. G. H., Baltissen, M. P. A., Jansen, P. W. T. C., Vermeulen, M., Tiesinga, P., & Celikel, T. (2017). Proteomic landscape of the primary somatosensory cortex upon sensory deprivation. GigaScience, 6(10), 1–10. doi:10.1093/gigascience/gix082

Related datasets:
doi:10.5524/100336 Cites doi:10.5524/100296

Accessions (data included in GigaDB):
PX: PXD005971

Funding:
Funding body - Radboud University Nijmegen Faculty of Science
Award ID - 626830 – 6200821
Awardee - Mr. Paul Tiesinga

Funding body - NWO – ALW Open Programme
Award ID - 824.14.022
Awardee - Mr. Tansu Celikel

Findability Agenda

1. Persistent Identifiers (PIDs)
   - Which ones are there? How should they be used?
   - DOIs for research data → minimum criteria for a good data repo
2. Choosing FAIR repositories
3. Setup help
4. Welcome reception

PIDs are everywhere:

Researcher IDs

ORCID

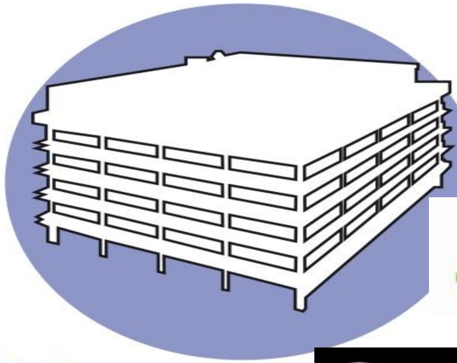
Scopus®

isni

RESEARCHERID

THOMSON REUTERS

Organisation IDs, Funder IDs



fundref

Ringgold

Identify

GRID

Resource IDs (articles, data, software, …)

doi®

ePIC

Persistent Identifiers for eResearch

Handle.Net®

URN-SERVICE

ARK (Archival Resource Key)

PICHE – Persistent Identifiers for Cultural Heritage Entities

And even more new PIDs…

- Projects IDs
- Instrument IDs
- Ship cruises IDs
- Physical sample IDs,
- DMP IDs…

Answer: Do researchers need to care about PIDs? → YES!

But what do they need to KNOW about PIDs?

→ Remember: For a scientist, it is about the project, equipment, DMP, researcher, funder, resource …

→ It is not about the PID. PIDs are infrastructure.

→ In order to use PIDs, scientists do not need to know all about their whereabouts. A basic knowledge should be enough.

TIB Survey 2017

1400 scientists in the natural sciences & engineering (across Germany)

→ 70% of the researchers are using DOIs for journal publications

→ less than 10% use DOIs for research data
(have a look: data available at doi.org/10.22000/54)

Why?

- 56% answered that they don't know about the option to use DOIs for other publications (datasets, conference papers etc.)
- 57% stated no need for DOI counselling services
- 40% of the questioned researchers need more information
- 30% cannot see a benefit from a DOI

With the new digital age: Possibilities & struggles!

Have a look: https://www.re3data.org/search
Out of more than 2115 repository systems listed in re3data.org in July 2018, only 809 (less than 39 %!) state to provide a PID service, with 524 of them using the DOI system

A PID is

- Provenance
- Metadata
- Policies & Guarantees
- Machine readability
- Metrics

Researchers should know that…

Provenance means validation & credibility – a researcher should comply to good scientific practices and be sure about what should get a PID (and what not).

Metadata is central to visibility and citability – metadata behind a PID should be provided with consideration.

Policies behind a PID system ensure persistence in the WWW - point. At least metadata will be available for a long time.

Machine readability will be an essential part of future discoverability – resources should be checked and formats should be adjusted (as far possible).

Metrics (e.g. altmetrics) are supported by PID systems.

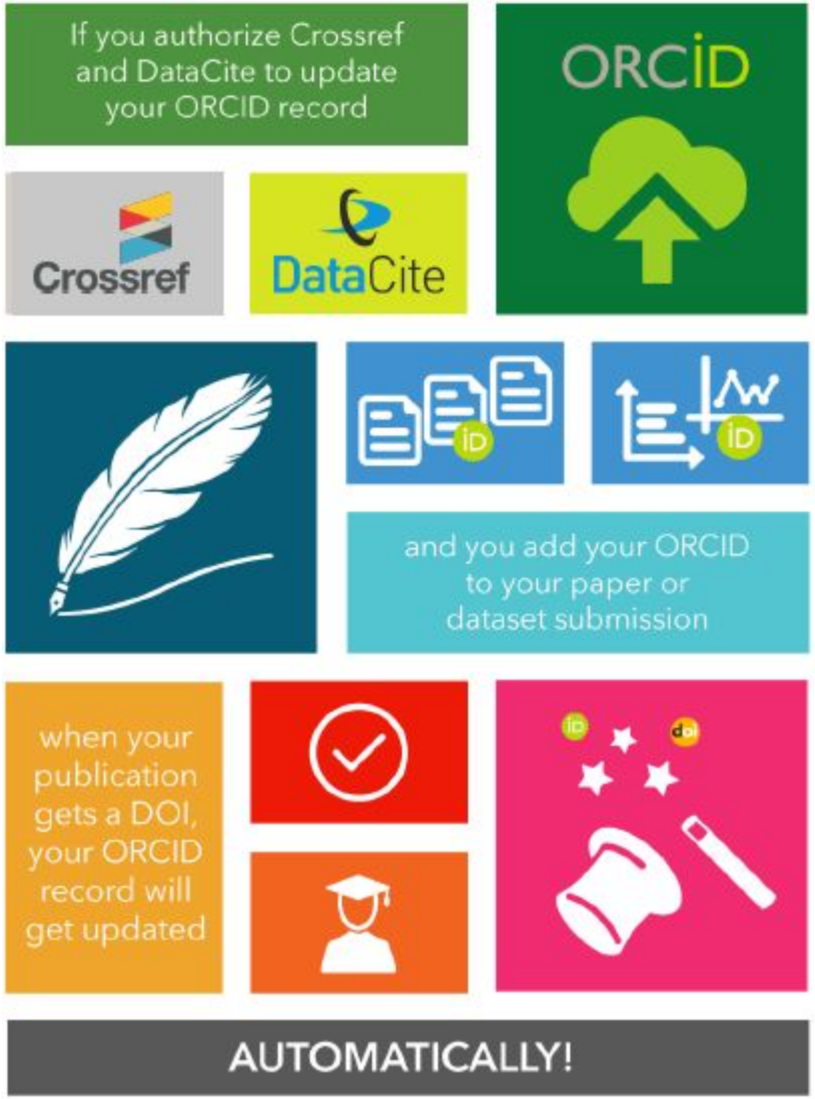
PIDs provide interoperable Metadata

- Example:

→ Automatic ORCID profile update when DOI is minted

DataCite – CrossRef – ORCID collaboration

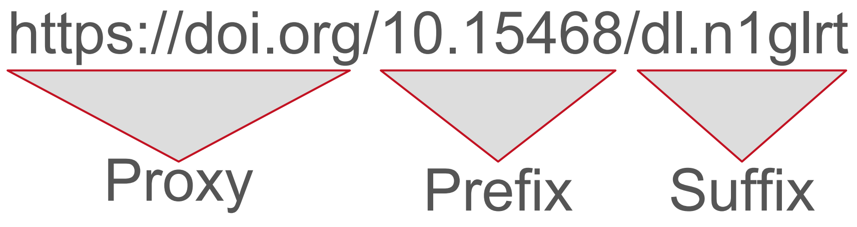
→ PID of choice for RDM:
Here: The Digital Object Identifier (DOI)



If you authorize Crossref and DataCite to update your ORCID record
ORCID
Crossref
DataCite
and you add your ORCID to your paper or dataset submission
when your publication gets a DOI, your ORCID record will get updated
AUTOMATICALLY!

Digital Object Identifier (DOI)

- Persistent and unique identifier for objects in the digital environment
- DOIs refer to the objects not the location → remain valid
- DOIs are minted for research data, software and code, physical objects, grey literature
- DOI-System is an internationally recognised and supported standard

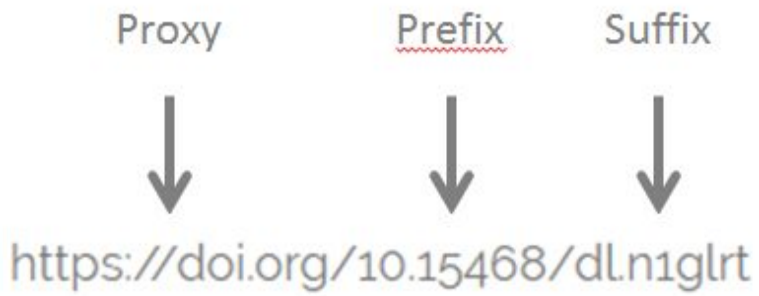


https://doi.org/10.15468/dl.n1glrt

Proxy

Prefix

Suffix



The DOI Service

- National mandate to support German academic institutions and the research sector publishing and citing research output, including:
    - research data
    - software
    - videos
    - images
    - 3D models
    - grey literature

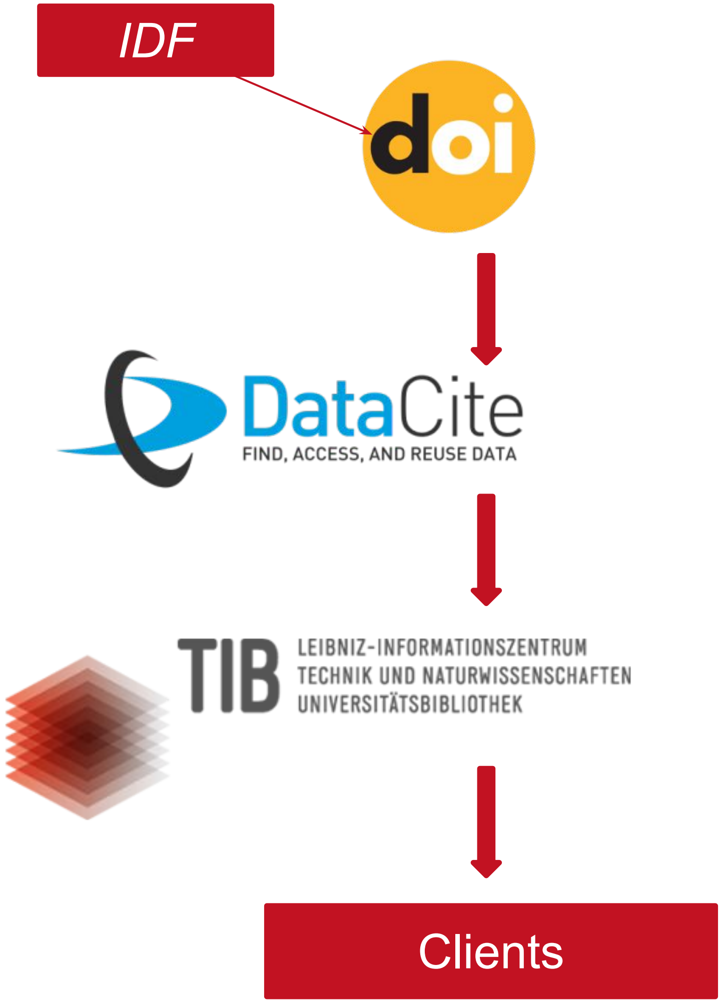
- With focus on Science and Technology
- TIB is founding member of DataCite, a global DOI agency for research data

TIB Services

- Managing prefix assignment
- First-Level-Support
- Training, counseling and hand-outs
- Providing access to the registration plattform (Metadata Store and DOI Fabrica)
- Support in all aspects of good research data management practice

Service Structure

- International DOI Foundation (IDF) manages DOI-System
- DataCite e.V. maintains and operates the infrastructure for DOI registration
- TIB as DOI Provider grants access to the infrastructure and provides services
- Client registers DOIs via TIB



IDF

doi
DataCite
FIND, ACCESS, AND REUSE DATA
TIB
LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK

Clients

DOI at TIB: Facts

Registered DOI

- **Total 1,293,389 DOIs** (June 2018)
  - 65 % Research data
  - 25 % Grey literature
  - 10 % Images
  - 0,4 % AV media

External clients of the DOI-Service

- **Total 165 data centers**
  - Major research centers i.e. Pangaea, WDCC and ESO
  - 68 universities/university libraries
  - 12 Leibniz Institute /11 Helmholtz Zentren

TIB user of the DOI-Service

- catalogue: Team DTF, retro-digitisation
- TIB-Portal: AWI reports, AV-Media

PID 101 for Researchers (*or: Resolving some PID myths*)

1. A PID is a „long lasting reference to a digital resource“
2. There are different sorts of PIDs & different uses,
   (e.g. for articles, data, persons, organizations, …)
3. PIDs are offered by organizations - Ask your institute/library
4. You do NOT have to pay for PIDs (by yourself)!
5. PIDs are mostly used for (persistent) citation – All published resources should have one
6. A correct citation always includes a PID → look in your citation manager
7. Metadata behind a PID are most important – please take care when providing them
8. PIDs are not perfect (they are issued by organizations, aka humans!)
9. PIDs are really useful & fun – they make yourself & your work more visible!

And what researchers do NOT need to know…
(*although some may want to know*)

- Total number of PIDs registered
- Names of the agencies
- Names & schemes of identifiers
- How persistence works
- How PID providers struggle
- How (and why) PID providers fight each other
- How perfect a PID (system) is (it is certainly not)

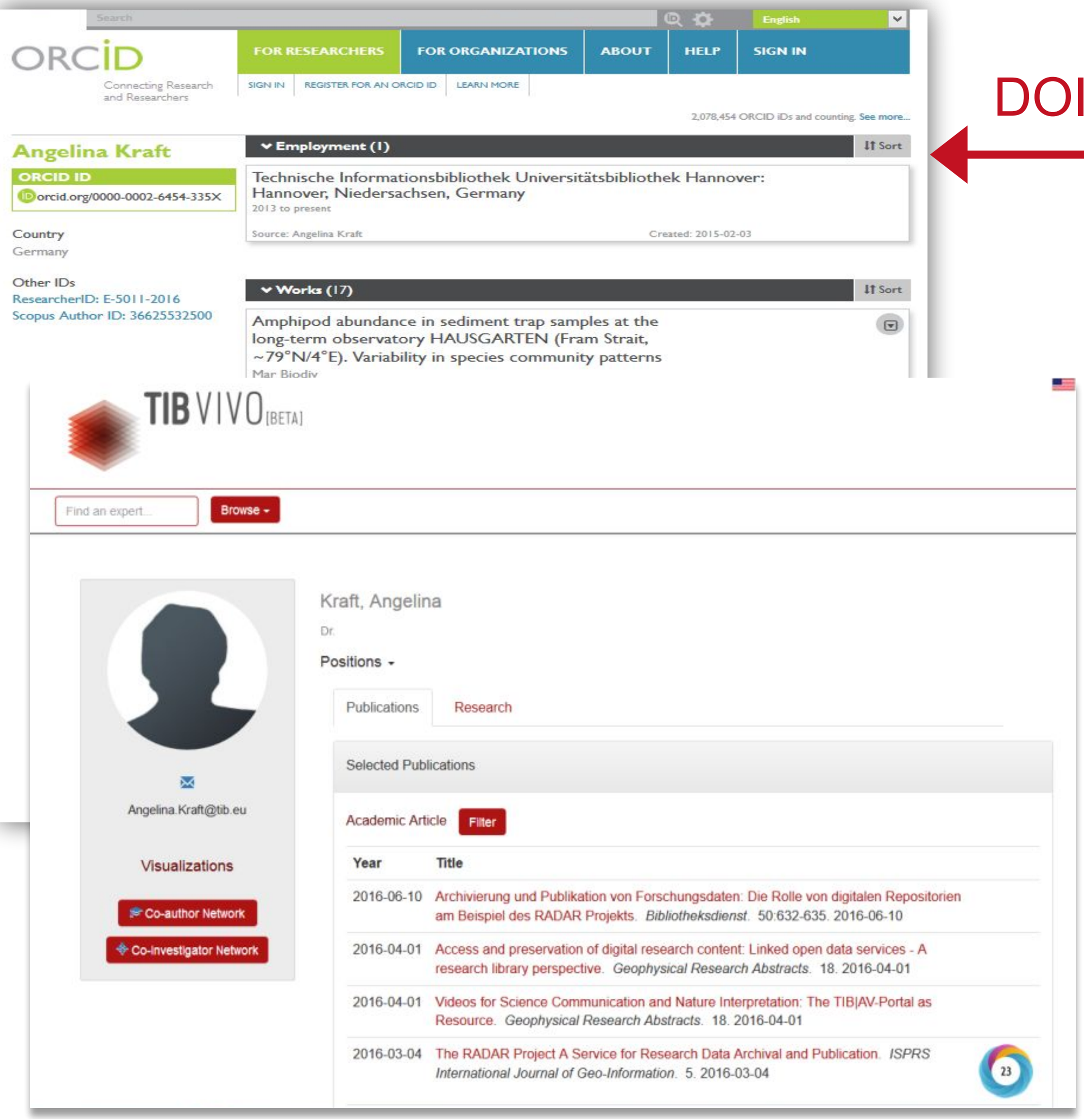
→ Researchers care about their research (= their passion)

→ As long as their not in information science themselves, PID providers should focus on communicating the practical points

→ Citeability & visibility; the benefit for the researcher should be crystal clear

Summary: PIDs (DOIs) = the glue!

Digital CV / CRIS



2,078,454 ORCID iDs and counting. See more...

Angelina Kraft

ORCID ID
orcid.org/0000-0002-6454-335X

Country
Germany

Other IDs
ResearcherID: E-5011-2016
Scopus Author ID: 36625532500

Employment (1) Sort

Technische Informationsbibliothek Universitätsbibliothek Hannover: Hannover, Niedersachsen, Germany
2013 to present
Source: Angelina Kraft Created: 2015-02-03

Works (17) Sort

Amphipod abundance in sediment trap samples at the long-term observatory HAUSGARTEN (Fram Strait, ~79°N/4°E). Variability in species community patterns
Mar Biodiv

Kraft, Angelina
Dr.
Positions

Angelina.Kraft@tib.eu

Year	Title
2016-06-10	Archivierung und Publikation von Forschungsdaten: Die Rolle von digitalen Repositorien am Beispiel des RADAR Projekts. Bibliotheksdienst. 50:632-635. 2016-06-10
2016-04-01	Access and preservation of digital research content: Linked open data services - A research library perspective. Geophysical Research Abstracts. 18. 2016-04-01
2016-04-01	Videos for Science Communication and Nature Interpretation: The TIB|AV-Portal as Resource. Geophysical Research Abstracts. 18. 2016-04-01
2016-03-04	The RADAR Project A Service for Research Data Archival and Publication. ISPRS International Journal of Geo-Information. 5. 2016-03-04

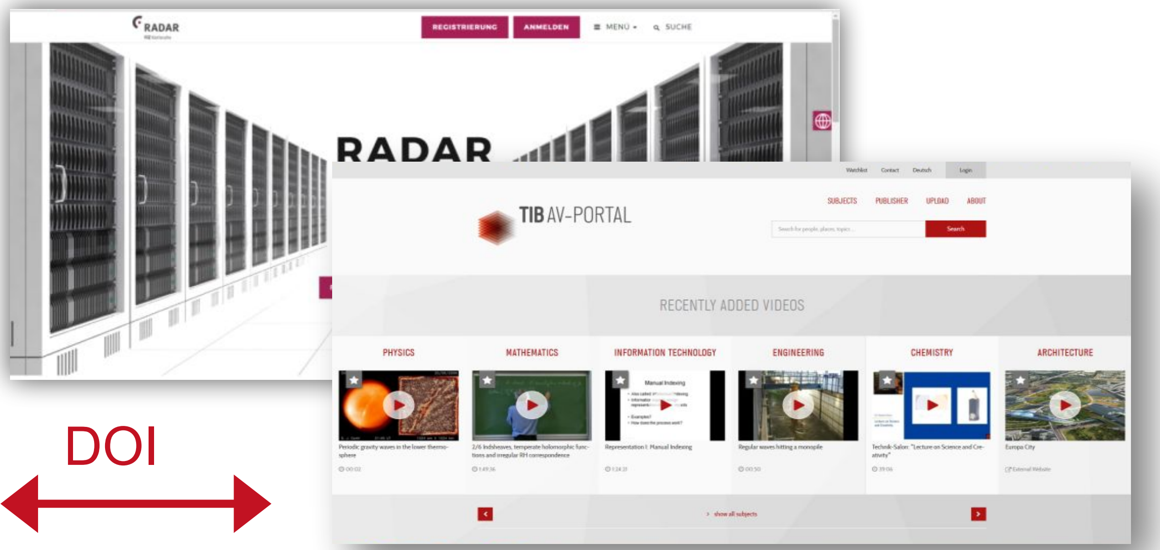
DOI

DOI

DOI

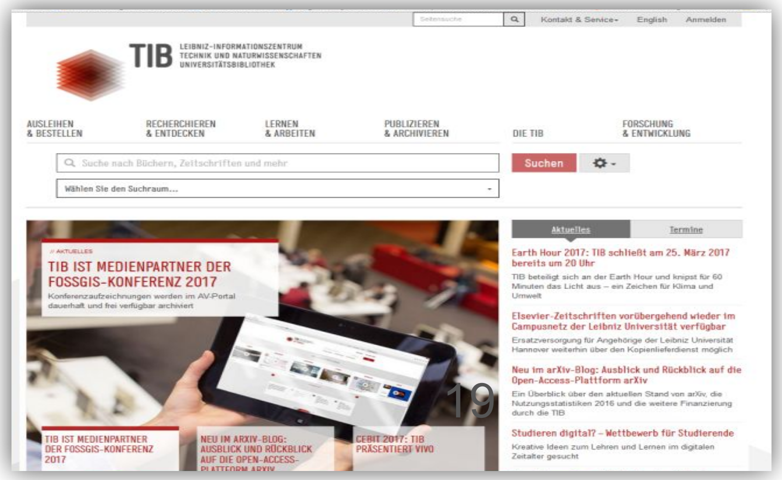
DOI

Data & AV Repository



RADAR
REGISTRIERUNG
ANMELDEN
MENÜ
SUCHE
RADAR
Watchlist
Contact
Deutsch
Login
TIB AV-PORTAL
SUBJECTS
PUBLISHER
UPLOAD
ABOUT
Search for people, places, topics...
Search
RECENTLY ADDED VIDEOS
PHYSICS
MATHEMATICS
INFORMATION TECHNOLOGY
ENGINEERING
CHEMISTRY
ARCHITECTURE
Periodic gravity waves in the lower thermosphere
2/6 Indsheaves, temperate holomorphic functions and irregular RH correspondence
Representation I: Manual Indexing
Regular waves hitting a monopile
Technik-Salon: "Lecture on Science and Creativity"
Europa City
External Website
show all subjects

Portal



Seitensuche
Kontakt & Service
English
Anmelden
TIB LEIBNIZ-INFORMATIONSZENTRUM TECHNIK UND NATURWISSENSCHAFTEN UNIVERSITÄTSBIBLIOTHEK
AUSLEIHEN & BESTELLEN
RECHERCHIEREN & ENTDECKEN
LERNEN & ARBEITEN
PUBLIZIEREN & ARCHIVIEREN
DIE TIB
FORSCHUNG & ENTWICKLUNG
Suche nach Büchern, Zeitschriften und mehr
Suchen
Wählen Sie den Suchraum...
Aktuelles
Termine
// AKTUELLES

TIB IST MEDIENPARTNER DER FOSSGIS-KONFERENZ 2017
Konferenzaufzeichnungen werden im AV-Portal dauerhaft und frei verfügbar archiviert

Earth Hour 2017: TIB schließt am 25. März 2017 bereits um 20 Uhr
TIB beteiligt sich an der Earth Hour und knipst für 60 Minuten das Licht aus – ein Zeichen für Klima und Umwelt

Elsevier-Zeitschriften vorübergehend wieder im Campusnetz der Leibniz Universität verfügbar
Ersatzversorgung für Angehörige der Leibniz Universität Hannover weiterhin über den Kopienlieferdienst möglich

Neu im arXiv-Blog: Ausblick und Rückblick auf die Open-Access-Plattform arXiv
Ein Überblick über den aktuellen Stand von arXiv, die Nutzungsstatistiken 2016 und die weitere Finanzierung durch die TIB

Studieren digital? – Wettbewerb für Studierende
Kreative Ideen zum Lehren und Lernen im digitalen Zeitalter gesucht

TIB IST MEDIENPARTNER DER FOSSGIS-KONFERENZ 2017

NEU IM ARXIV-BLOG: AUSBLICK UND RÜCKBLICK AUF DIE OPEN-ACCESS-PLATTFORM ARXIV

CEBIT 2017: TIB PRÄSENTIERT VIVO

Data Journals

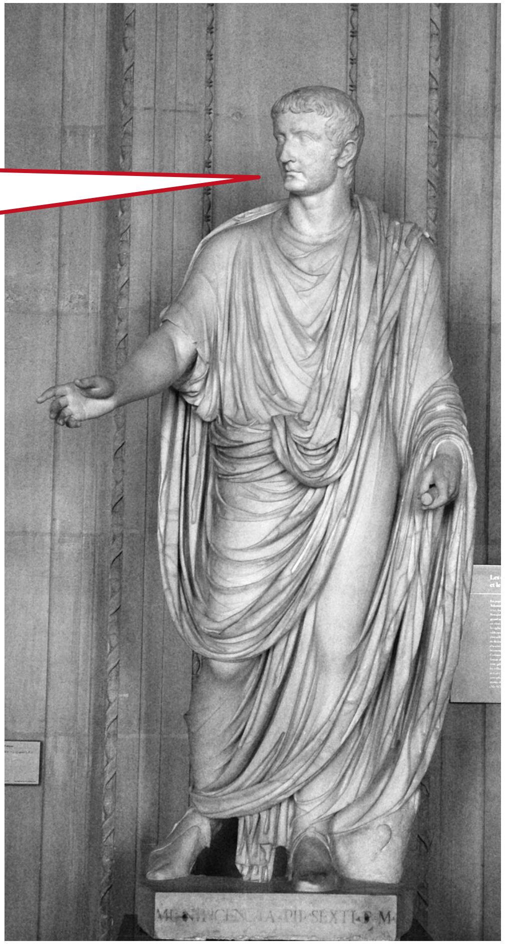
Articles in Data Journals

- describing research data
- Containing information about data creation and collection, data qualities, functionalities
- Showing possibilities of re-using data
- Recommendation or provision of data deposit
- Discipline specific or generic data journals
- Examples
  - Nature Scientific data
  - Biomedical Data Journal (BMDJ)

GitHub + Zenodo = DOI (guides.GitHub.com/activities/citable-code)

- official integration thanks to science.Mozilla.org/projects/codemeta
- intrinsic IDs (Git’s SHA1 hashes) vs. “minted” PIDs
  - technical vs. procedural persistence
- Zenodo: file backup & persistent landing page for each release version
  - SoftwareHeritage.org ingests automatically, but not in real-time (source)
- Read more Software.ac.uk/tags/software-preservation
- demo: GitHub.com/TIBHannover/BacDiveR/issues/14

When in Rome, dress like the Romans.



Public Domain by Marie-Lan Nguyen, via
commons.Wikimedia.org/w/index.php?curid=549920

Findability Agenda

abstraction layers that direct you to the actual object locations

1. Persistent Identifiers (PIDs)
   - Which ones are there? How should they be used?
   - DOIs for research data → minimum criteria for a good data repo
2. Choosing FAIR repositories
3. Setup help
4. Welcome reception

The right repository?

There is no right or wrong

Decision for or against a specific repository depends on various criteria, e.g.

- Data quality
- Discipline
- Institutional requirements
- Reputation (researcher and/or repository)
- Visibility of research
- Legal terms and conditions
- Data value (FAIR Principles)
- Exit strategy (tested?)
- Certificate (based only on documents?)

Decision has to be taken by the researcher him-/herself

Help: Checklist – where to keep research data (DCC)

Benefits of data sharing/publication in (good) data repositories

- data are kept safe in a secure environment
- data are regularly backed up and preserved (long-term) for future use
- data can be easily discovered by search engines and included in online catalogues
- intellectual property rights and licencing of data are managed
- access to data can be administered and usage monitored
- the visibility of data can be enhanced
- enables more use and citation
- citation of data increases researchers scientific reputation

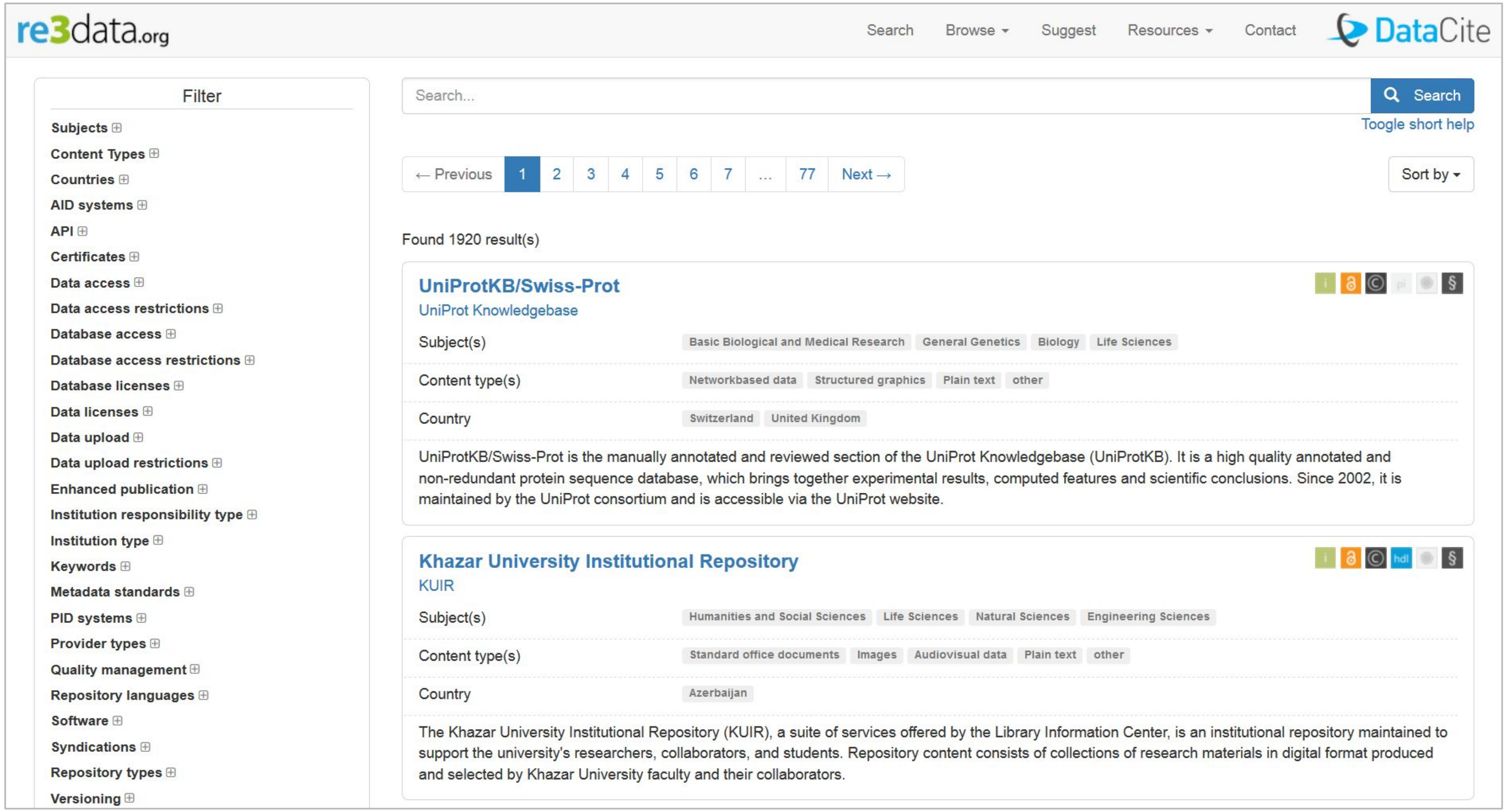
Some recommendations:
→ look for the usage of PIDs
→ look for the usage of standards (DataCite, Dublin Core, discipline-specific metadata
→ look for licences offered
→ look for certifications (DSA / Core Trust Seal, DINI/nestor, WDS, …)

FAIRsharing.org vs. re3data.org (June 2018)

	FAIRsharing.org standards, databases, policies	re3data.org REGISTRY OF RESEARCH DATA REPOSITORIES
Mode of use, operator	Online platform University of Oxford	Online platform DataCite (KIT)
Type of data	Metadata records of “*Standards, databases, policies, and Collections /Recommendations*“	Metadata records of *data repositories*
No. of records	Standards: 1169 Databases (=Repositories): 1064 Polices: 112 Collections/Recommendations: 44 Identifier schema: 7	Metadata standards: 76 Repositories: 2096 Identifier schema (PID Systems): 5 +others
Findability	URL; recommendations (EU H2020, journal guidelines)	URL; recommendations (EU H2020, journal guidelines)
Accessibility (timeline)	Since 2016 (started with biosharing.org)	Since 2012 (merged with Databib 2015)
Interoperability	Mark up with schema.org & BioSchema project planned, API	XML templates, API
Re-useability	API (read-only, pre-contact required) Content licensed via CC-BY-SA 4.0	API Content licensed via CC-0 1.0
Data policy (who can upload information)	Registered users who are maintainers, ORCiD preferred, “sanity checks” by curators, approval/refusal	Repository operators, application form, analyzes of the repository website by curators, metadata checked, approval/refusal

How to find a repository

- Ask your colleagues & collaborators
- Look for institutional repository at your own institution
- Search at re3data.org – Registry of Research Data Repositories



Found 1920 result(s)

UniProtKB/Swiss-Prot

UniProt Knowledgebase

Subject(s): Basic Biological and Medical Research, General Genetics, Biology, Life Sciences

Content type(s): Networkbased data, Structured graphics, Plain text, other

Country: Switzerland, United Kingdom

UniProtKB/Swiss-Prot is the manually annotated and reviewed section of the UniProt Knowledgebase (UniProtKB). It is a high quality annotated and non-redundant protein sequence database, which brings together experimental results, computed features and scientific conclusions. Since 2002, it is maintained by the UniProt consortium and is accessible via the UniProt website.

Khazar University Institutional Repository

KUIR

Subject(s): Humanities and Social Sciences, Life Sciences, Natural Sciences, Engineering Sciences

Content type(s): Standard office documents, Images, Audiovisual data, Plain text, other

Country: Azerbaijan

The Khazar University Institutional Repository (KUIR), a suite of services offered by the Library Information Center, is an institutional repository maintained to support the university's researchers, collaborators, and students. Repository content consists of collections of research materials in digital format produced and selected by Khazar University faculty and their collaborators.

Findability Agenda

abstraction layers that direct you to the actual object locations

1. Persistent Identifiers (PIDs)
   - Which ones are there? How should they be used?
   - DOIs for research data → minimum criteria for a good data repo

2. Choosing FAIR repositories

3. **Setup help needed? Red stickies, please!**

4. Welcome reception at Waterloostraße 1; directions: TIBHannover.GitHub.io/2018-07-09-FAIR-Data-and-Software/#schedule

LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY

TIB

Which questions do you have for us?

Contact information:
Katrin.Leinweber@TIB.eu & Angelina.Kraft@TIB.eu
T +49 511 762-14693 & -14238

Leibniz
Association