4eu+

RESEARCH DATA MANAGEMENT - INTRODUCTION TO FAIR AND OPEN DATA

Jochen APEL (Heidelberg University)
Paola GALIMBERTI (University of Milan)
Milan JANÍČEK (Charles University, Prague)

Open for you! An introduction series to open science | 7 March 2022

CHARLES UNIVERSITY

UNIVERSITÄT HEIDELBERG
ZUKUNFT
SEIT 1386

SORBONNE UNIVERSITÉ

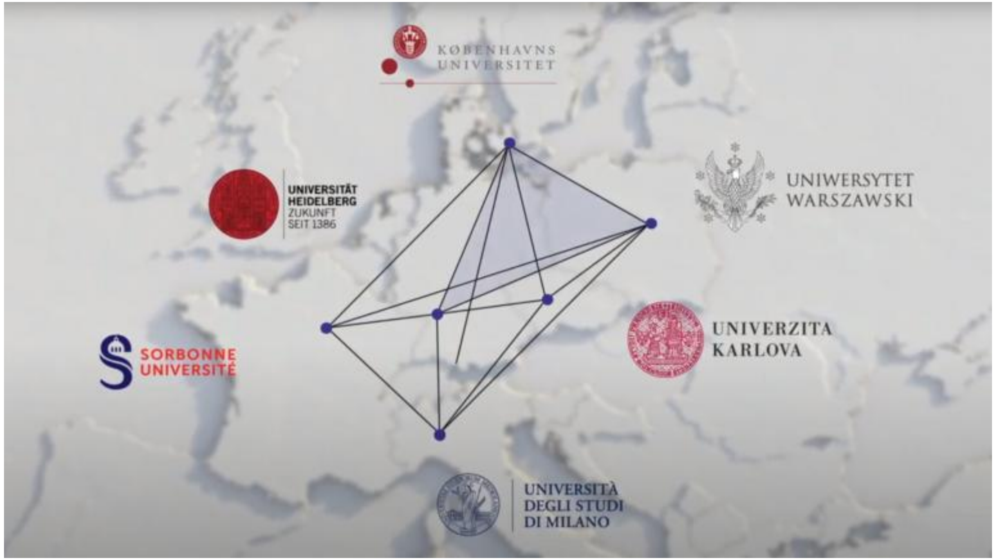
UNIVERSITY OF COPENHAGEN

UNIVERSITÀ DEGLI STUDI DI MILANO

UNIVERSITY OF WARSAW

4EU+ Alliance and Open Science

- 4EU+ is a transnational strategic university association.
  - Aim: Strengthen the European vision of deepened cooperation and mutual enrichment in research and teaching
  - Open Science is an integral part of this.
  - Two 4EU+ projects currently work on Open Science.
  - Open for you – an Introduction Series to Open Science“ – 14 session on OS topics!



Agenda

- Introduction
- Why bother with FAIR Data?
- Why bother with Open Data?
- Good practices: a closer look on various data repositories
  - Subject-specific repositories
  - Generic Repositories
  - Institutional Repositories

What we will talk about?

“An article […] in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship.”

Buckheit, Jonathan B., and David L. Donoho. “WaveLab and Reproducible Research.” In *Wavelets and Statistics*, edited by Anestis Antoniadis and Georges Oppenheim, 55–81. Lecture Notes in Statistics. New York, NY: Springer, 1995. https://doi.org/10.1007/978-1-4612-2544-75.

Publications

Interviews

Notes

Annotations

Methods

Bibliographic references

Databases

Images

VIsualizations

Video

Corpora

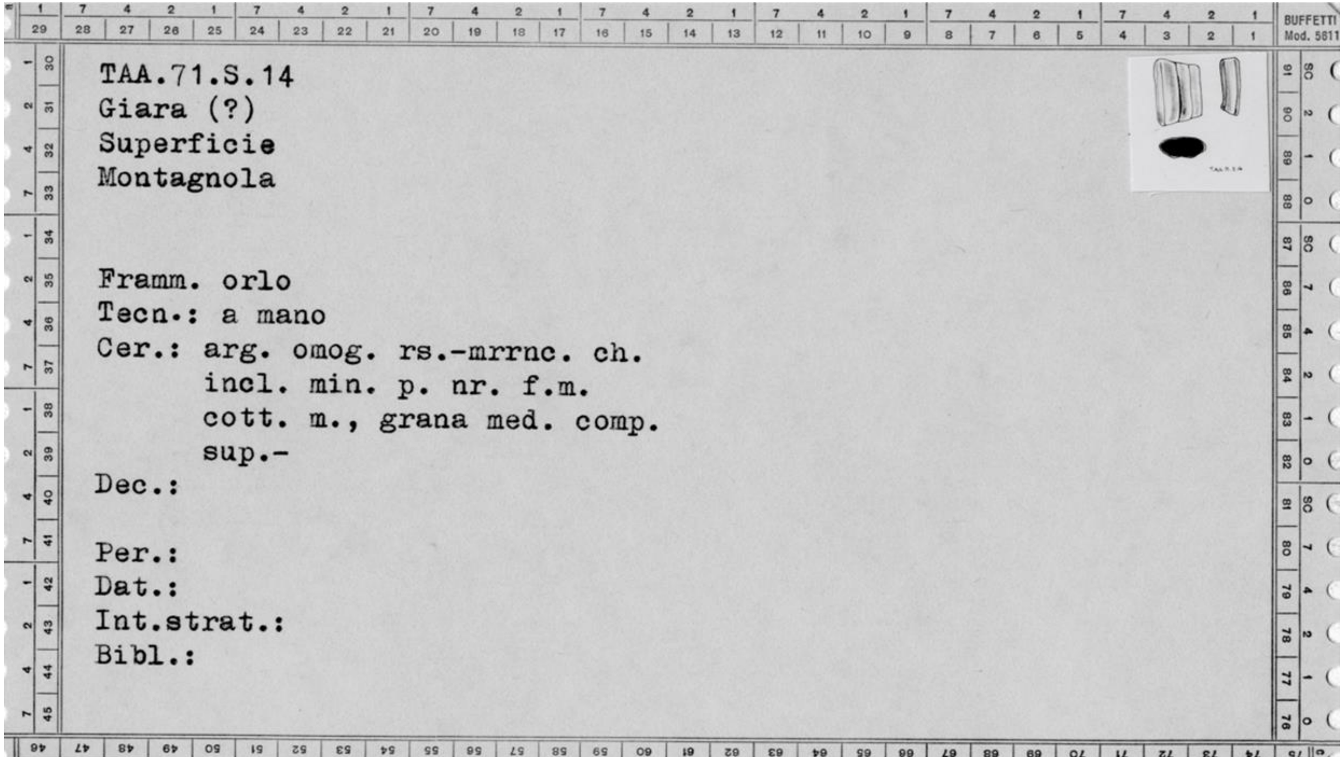
Feedback

Experimental data

The tip of the iceberg

What is research data?

Is this data?



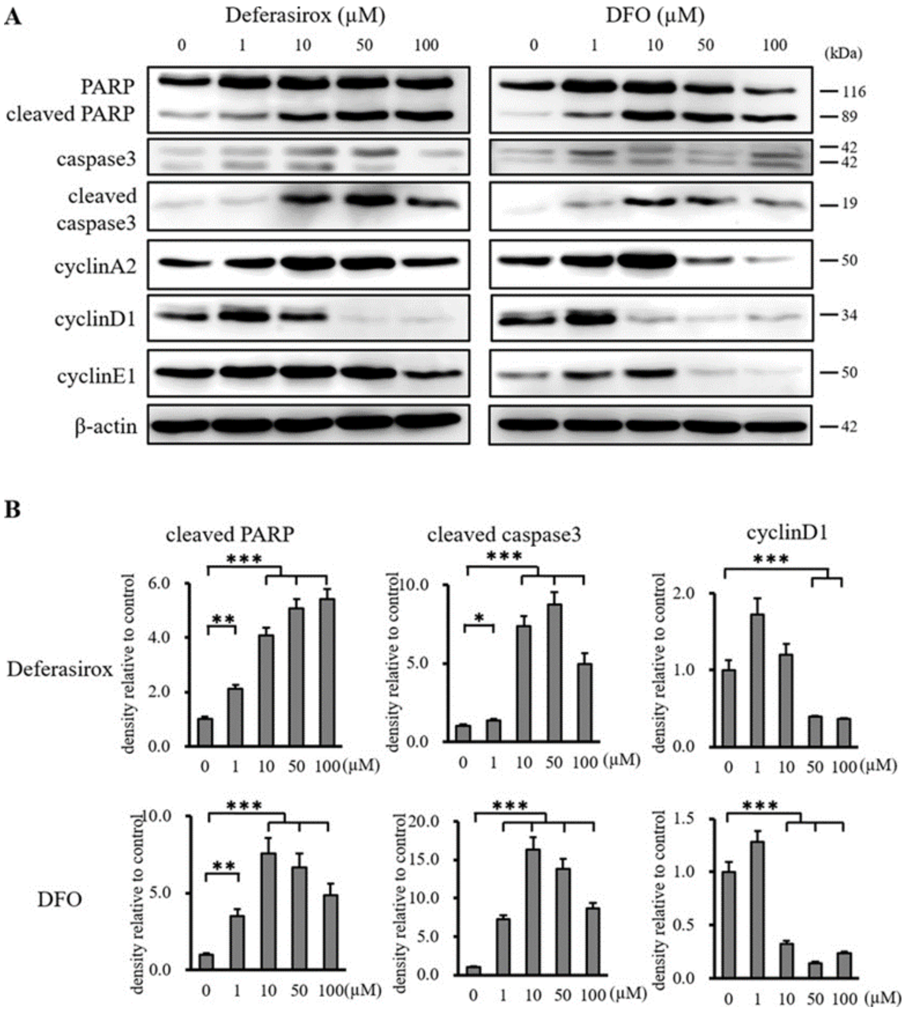
BUFFETTI
Mod. 5611

TAA.71.S.14
Giara (?)
Superficie
Montagnola

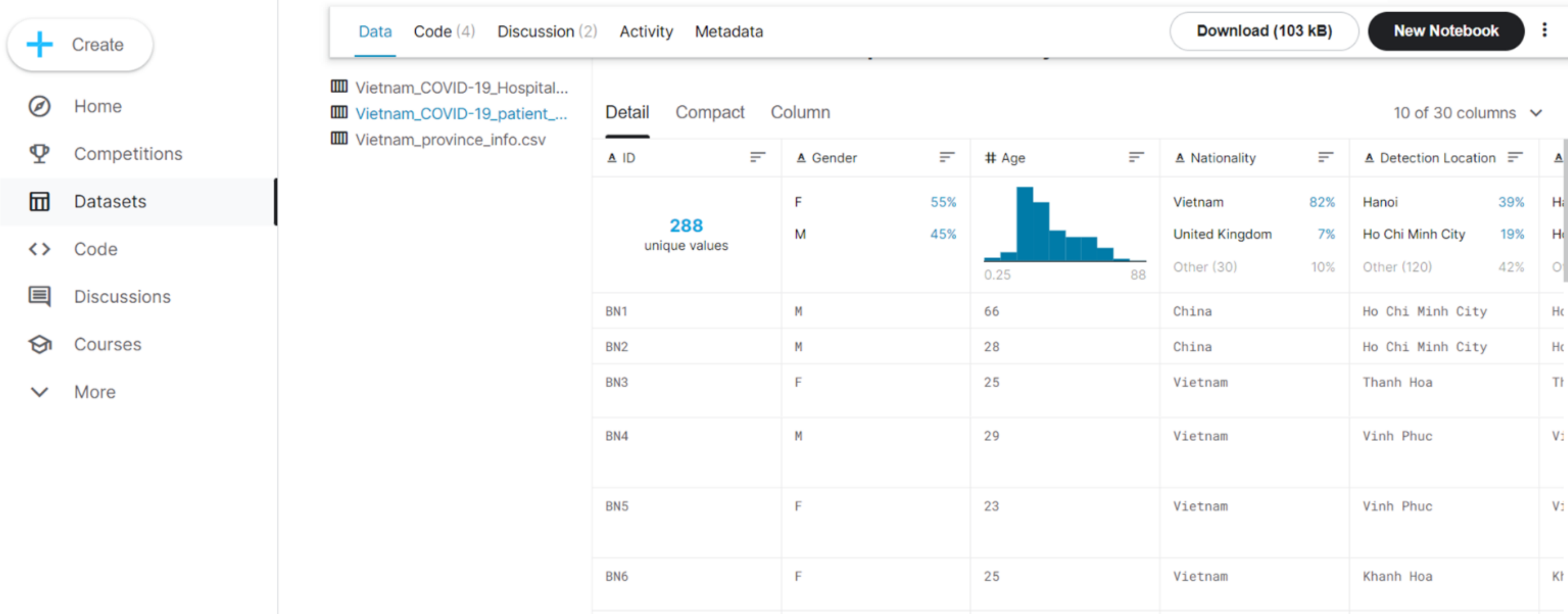
Framm. orlo
Tecn.: a mano
Cer.: arg. omog. rs.-mrrnc. ch.
incl. min. p. nr. f.m.
cott. m., grana med. comp.
sup.-
Dec.:

Per.:
Dat.:
Int.strat.:
Bibl.:

Is this data?


A
Deferasirox (μM)
DFO (μM)
0
1
10
50
100
(kDa)
PARP
cleaved PARP
caspase3
cleaved caspase3
cyclinA2
cyclinD1
cyclinE1
β-actin
116
89
42
42
19
50
34
50
42
B
cleaved PARP
cleaved caspase3
cyclinD1
Deferasirox
DFO
density relative to control
***
**
*
6.0
4.0
2.0
0.0
10.0
5.0
20.0
15.0
1.5
1.0
0.5
0 1 10 50 100 (μM)

Is this data?



ID	Gender	Age	Nationality	Detection Location
288 unique values	F 55% M 45%		Vietnam 82% United Kingdom 7% Other (30) 10%	Hanoi 39% Ho Chi Minh City 19% Other (120) 42%
BN1	M	66	China	Ho Chi Minh City
BN2	M	28	China	Ho Chi Minh City
BN3	F	25	Vietnam	Thanh Hoa
BN4	M	29	Vietnam	Vinh Phuc
BN5	F	23	Vietnam	Vinh Phuc
BN6	F	25	Vietnam	Khanh Hoa

What is research data?

Any information that has been

collected,

observed,

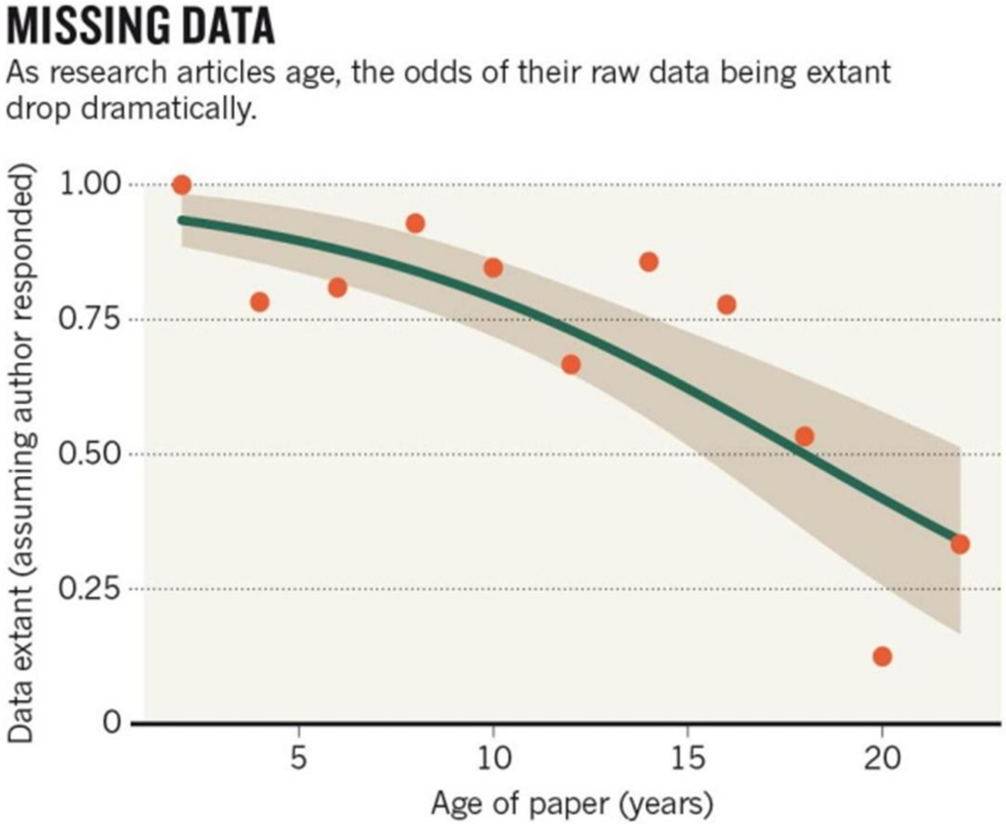
generated,

created

to validate original research findings

Research data is a very important resource, so we must ensure that it can be reused, first and foremost by ourselves.

Data can be missed



Data can get lost

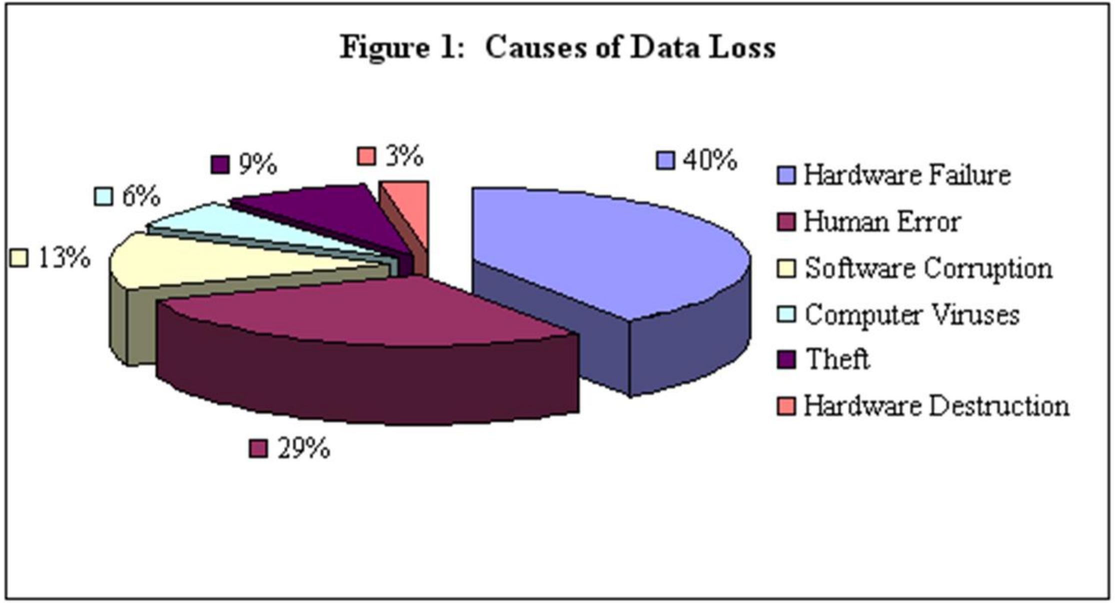
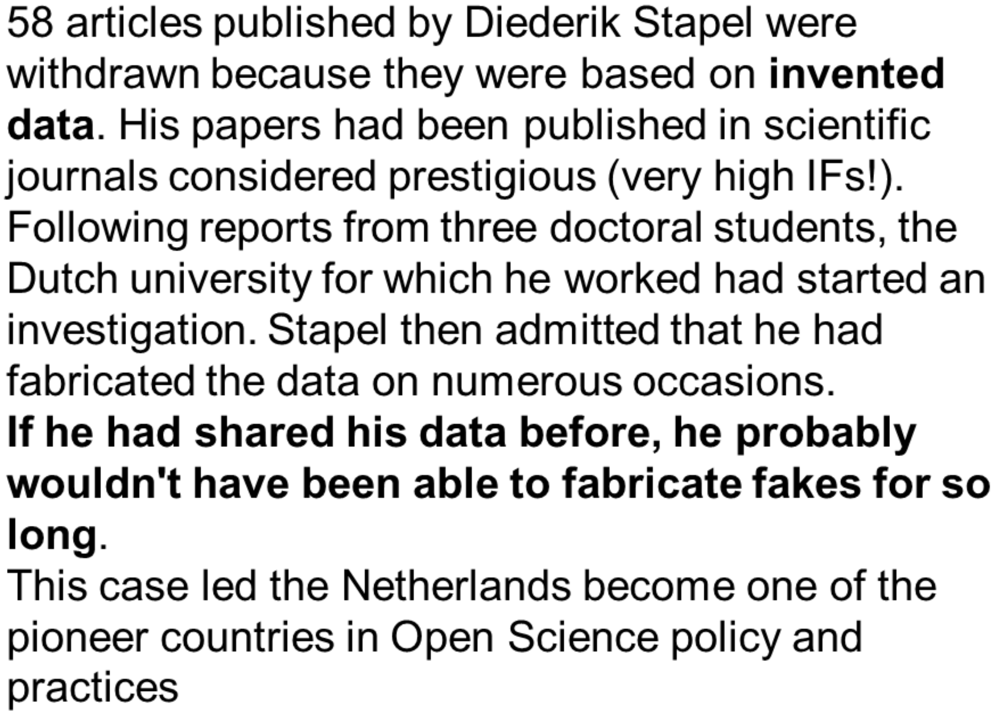


Figure 1: Causes of Data Loss

Source: Author's estimates based on data from Safeware, The Insurance Agency, Inc., "2000 Safeware Loss Study," 2001; and ONTRACK Data International, Inc., "Understanding Data Loss," 2003.

Data can be manipulated or altered

Paul Jump, A Star's Collapse.Dutch begin documenting and trying to explain top psychologist's massive fraud.
Times Higher Education, November 28, 2011



58 articles published by Diederik Stapel were withdrawn because they were based on **invented data**. His papers had been published in scientific journals considered prestigious (very high IFs!).
Following reports from three doctoral students, the Dutch university for which he worked had started an investigation. Stapel then admitted that he had fabricated the data on numerous occasions.
If he had shared his data before, he probably wouldn't have been able to fabricate fakes for so long.
This case led the Netherlands become one of the pioneer countries in Open Science policy and practices

These problems can be partially solved through responsible research data management

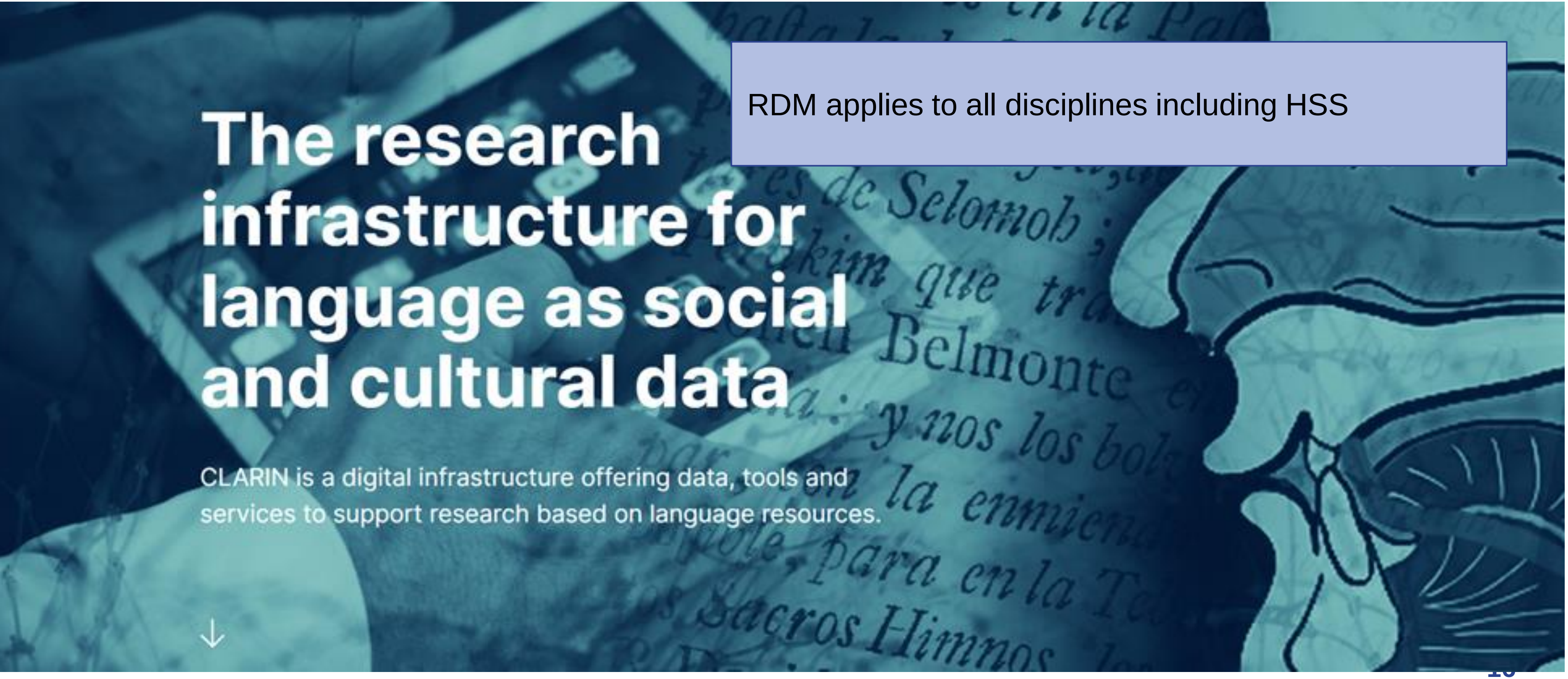
Research Data Management to the resc

"Research data management concerns the organisation of data, from its entry to the research cycle through to the dissemination and archiving of valuable results. It aims to ensure reliable verification of results, and permits new and innovative research built on existing information."

(Whyte, A., Tedds, J. (2011). ['Making the Case for Research Data Management'](). DCC Briefing Papers. Edinburgh: Digital Curation Centre.)

https://library.sydney.edu.au/research/data-management/research-data-management.html

RDM applies to all disciplines including HSS



The research infrastructure for language as social and cultural data

CLARIN is a digital infrastructure offering data, tools and services to support research based on language resources.

Funders are pushing RDM & Open Data - European level

Horizon 2020 & Horizon Europe: FAIR Data Management

- Participating projects will be required to develop a **Data Management Plan** (DMP)
- Participating projects are **required to deposit research data**, preferably into a research data repository
- “[…]as far as possible, projects must then **take measures to enable for third parties to access**, mine, exploit, reproduce and disseminate (free of charge for any user) this research data.“

Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020 | Guidelines on Data Management in Horizon 2020

Funders are pushing RDM & Open Data - example for a national funder

Deutsche Forschungsgemeinschaft

DFG Guidelines on the Handling of Research Data

“[…] Assuming that the publication of research data from a DFG-funded project does not conflict with the rights of third parties (in particular data protection or copyright), **research data should be made available as soon as possible.**

[…]

Applicants may request funding for project-specific costs that arise in connection with a scientific project, for the preparation of research data for subsequent reuse and/or the transfer of research data to existing infrastructures as part of a proposal to the DFG. […]”

4eu+

The National Recovery and Resilience Plan

Piano Nazionale
di Ripresa e Resilienza
#NEXTGENERATIONITALIA

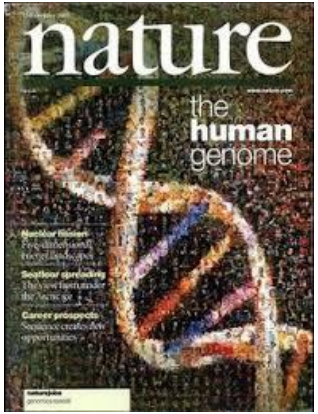
Coerentemente con le finalità del presente Avviso, ai risultati del Programma di ricerca e ai relativi dati (ad esempio, le pubblicazioni di risultati originali della ricerca scientifica, i dati grezzi e i metadati, le fonti, le rappresentazioni digitali grafiche e di immagini e i materiali multimediali scientifici) deve essere garantito un accesso aperto al pubblico nel minor tempo e con il minor numero di limitazioni possibile, **secondo i principi “Open science” e “FAIR Data”.**

Journals: Nature

An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims. Therefore, a condition of publication in a Nature journal is that authors are required to make materials, data and associated protocols promptly available to readers without undue qualifications.

[…]The preferred way to share large data sets is via public repositories.

http://www.nature.com/authors/policies/availability.html



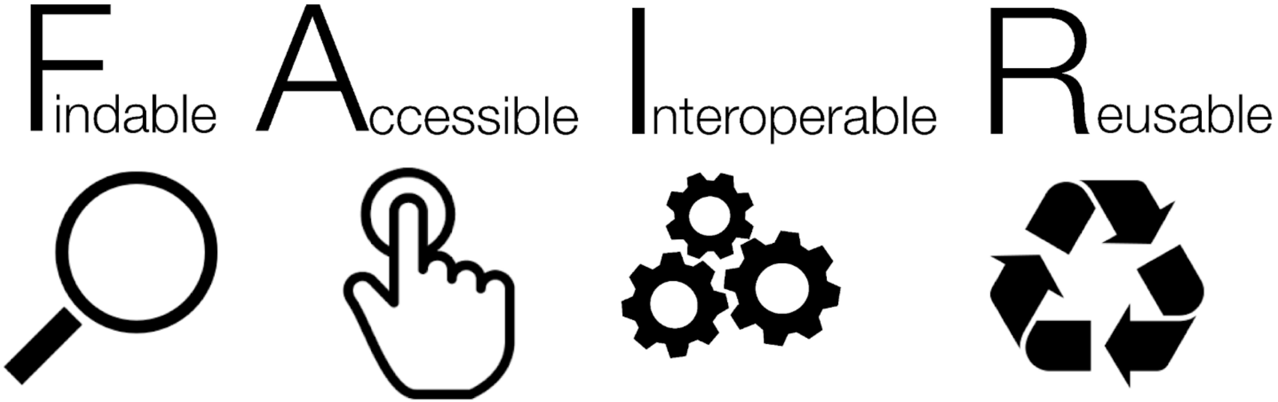
Journals: PLOS

Data Availability

PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception.
[…]
Refusal to share data and related metadata and methods in accordance with this policy will be grounds for rejection. […]

http://www.plosone.org/static/policies.action#sharing

FAIR Data Principles



- FAIR Data Principles
- Wilkinson et al. (2016), The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data 3, doi:10.1038/sdata.2016.18
- SNF: Explanation of the FAIR Data Principles

By SangyaPundir [CC BY-SA 4.0], from Wikimedia Commons

F - Findable

others can discover your data

- described using rich metadata
- using persistent identifiers
- indexed in catalogues or databases

F - Findable

imagine looking for a book in a library

- you can use the catalogue
  - you need the book to be described (**metadata**)
    - e.g. Nordic crime series; encyclopedia of modern airplanes

- you will find the location

- you will need some kind of very specific (unified) identification
    - “6th floor, shelf 6A/276, HF5718 .22 .F85 2014”

F - Findable – how to?

in libraries librarians create the description (metadata) for you

in online world, the author has to describe uploaded content himself (!)

repository can help you with identification – assign persistent identifiers (eg. DOI)

DOI:

DOI 10.5281/zenodo.3978090

Keyword(s):

pine wood nematode | Bursaphelechus xylophilus | quarantine pest | survey | Finland

Subject(s):

Bursaphelenchus | surveillance | data

Related identifiers:

Supplement to

10.3897/neobiota.58.38313 (Journal article)

10.5281/zenodo.3842358 (Software)

Metadata

- metadata describe data
  - e.g. title, author, keywords, subject headings, identifiers...

- metadata help other people find your data!

- author of datasets usualy chooses metadata when uploading to repository – do not underestimate this!
  - your data can be lost or found... it depends on you!

- libraries can help
  - librarians love metadata ;-)

Persistent identifiers (PID)

- sequence of letters/numbers in specified format, managed by some service provider
- **unique** – different things have different identifiers
    - note: name is not unique identifier (e.g. Carl Johansson)
- **persistent** – should work ‘forever’
- **resolvable** – there is a way how to get representation of identified object (e.g. links - see below)

Persistent identifiers (PID)

- provided by some service
  - DOI – digital object identifier ***10.5281/zenodo.5665754***
    - articles
    - data
  - ORCID – Open Researcher and Contributor ID -- ***0000-0002-8271-3674***
  - ROR.org – Research Organization Registry - ***024d6js02***

- note: URL is NOT identifier - https://zenodo.org/record/5665755#.Yh6O1-iZOUk

- ..but most PIDs are resolvable (you can use them in URL):
- https://doi.org/10.5281/zenodo.5665754, https://orcid.org/0000-0002-8271-3674, https://ror.org/024d6js02

A - Accessible

your data can be made available to others

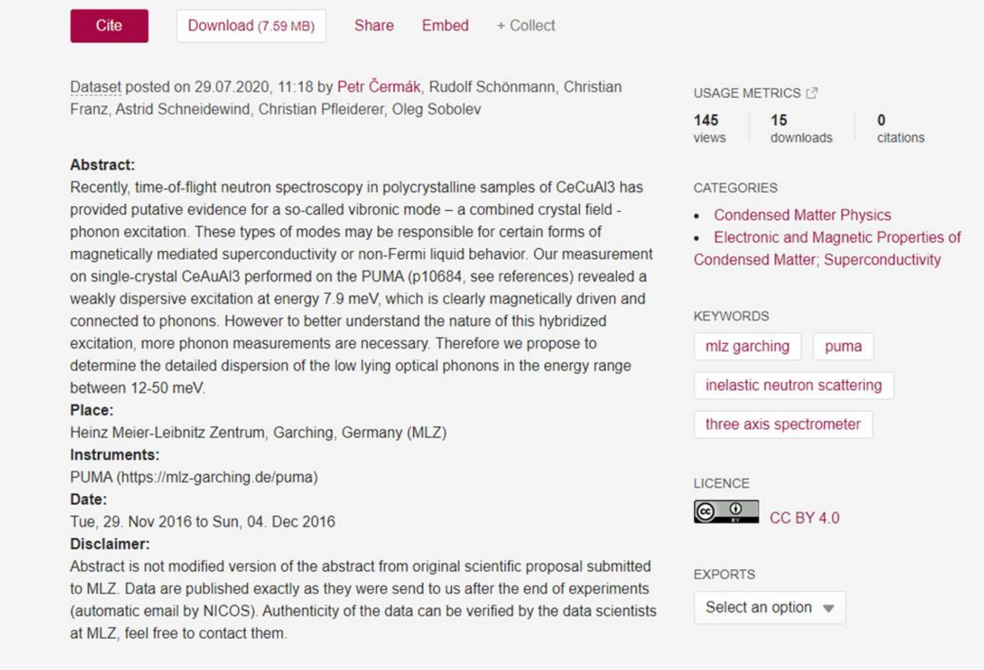
- using some „simple protocol“ – e.g. web browser
- or using some more sophisticated way (asking for permission etc.)

- knowing persistent identifier should always help you
  - you should always get at least metadata

- not everything has to be shared without limitations
  - “As open as possible, as closed as necessary.”

A - Accessible – how to?

- publish/share in a **repository**
- publish at least **metadata**



Dataset posted on 29.07.2020, 11:18 by Petr Čermák, Rudolf Schönmann, Christian Franz, Astrid Schneidewind, Christian Pfleiderer, Oleg Sobolev

Abstract:
Recently, time-of-flight neutron spectroscopy in polycrystalline samples of CeCuAl3 has provided putative evidence for a so-called vibronic mode – a combined crystal field - phonon excitation. These types of modes may be responsible for certain forms of magnetically mediated superconductivity or non-Fermi liquid behavior. Our measurement on single-crystal CeAuAl3 performed on the PUMA (p10684, see references) revealed a weakly dispersive excitation at energy 7.9 meV, which is clearly magnetically driven and connected to phonons. However to better understand the nature of this hybridized excitation, more phonon measurements are necessary. Therefore we propose to determine the detailed dispersion of the low lying optical phonons in the energy range between 12-50 meV.
Place:
Heinz Meier-Leibnitz Zentrum, Garching, Germany (MLZ)
Instruments:
PUMA (https://mlz-garching.de/puma)
Date:
Tue, 29. Nov 2016 to Sun, 04. Dec 2016
Disclaimer:
Abstract is not modified version of the abstract from original scientific proposal submitted to MLZ. Data are published exactly as they were send to us after the end of experiments (automatic email by NICOS). Authenticity of the data can be verified by the data scientists at MLZ, feel free to contact them.

USAGE METRICS

145	15	0
views	downloads	citations

CATEGORIES
- Condensed Matter Physics
- Electronic and Magnetic Properties of Condensed Matter; Superconductivity

KEYWORDS
mlz garching
puma
inelastic neutron scattering
three axis spectrometer

LICENCE
CC BY 4.0

EXPORTS
Select an option

I - Interoperable

other people should be able to use your data (with their own data)

- it should be possible to understand the data
  - format should be readable
  - there should be helpful documentation
  - well known terminology should be used (domain standards)

I - Interoperable – how to?

- use open formats
  (plaintext rules!)

- use the right language!

2019-nCoV Vaccine mRNA-1273 MeSH Descriptor
Data 2022

- Amino Acids, Peptides, and Proteins [D12]
  - Proteins [D12.776]
    - Recombinant Proteins [D12.776.828]
      - Vaccines, Synthetic [D12.776.828.868]
        - Nucleic Acid-Based Vaccines [D12.776.828.868.228]
          - mRNA Vaccines [D12.776.828.868.228.500]
            - **2019-nCoV Vaccine mRNA-1273 [D12.776.828.868.228.500.500]**
            - BNT162 Vaccine [D12.776.828.868.228.500.750]

Interoperable

- open format is not enough by itself -
    - people should be able to understand your data

- example: temperature
    - 36,6
    - 36,1
    - 37,8
    - 41,5
- what is the context?
    - patient data or hot summer?

- **describe** your data to make it interoperable (and reusable)!

R - Reusable

others should be allowed to (confidently) reuse your data

- it should be possible to trust the data
  - provenance
- it should be clear what are is allowed to do with the data
  - licensing conditions

R - Reusable – how to?

- describe the origin of the data sufficiently
  - how did you get them?
  - how were the data modified/processed?

- use license describing what is possible/legal to do with your data
  - share it freely?

Publication date:
April 4, 2021

DOI:
DOI 10.5281/zenodo.4661737

Keyword(s):
COVID-19 | Air Traffic | flight tracks | fuel consumption | Black Carbon emission | contrail | CoCiP

License (for files):
Creative Commons Attribution 4.0 International

Why Open Data?

Illustration by Ainsley Seago. From Roche, et. al. *Troubleshooting Public Data Archiving: Suggestions to Increase Participation. PLoS Biology* 12(1): e1001779. https://doi.org/10.1371/journal.pbio.1001779. Licensed under Creative Commons Attribution 4.0 International License.

“Currencies“ of science

- Peer-reviewed articles in “high-impact“ journals
- Publishing books with well-respected publishers
- Citations of your own work through other authors
- Conference contributions (talks, posters)

- At the end of the day, that’s what you get grants, funding, professorships,… for.

The scientific incentive system and the distribution of scientific results

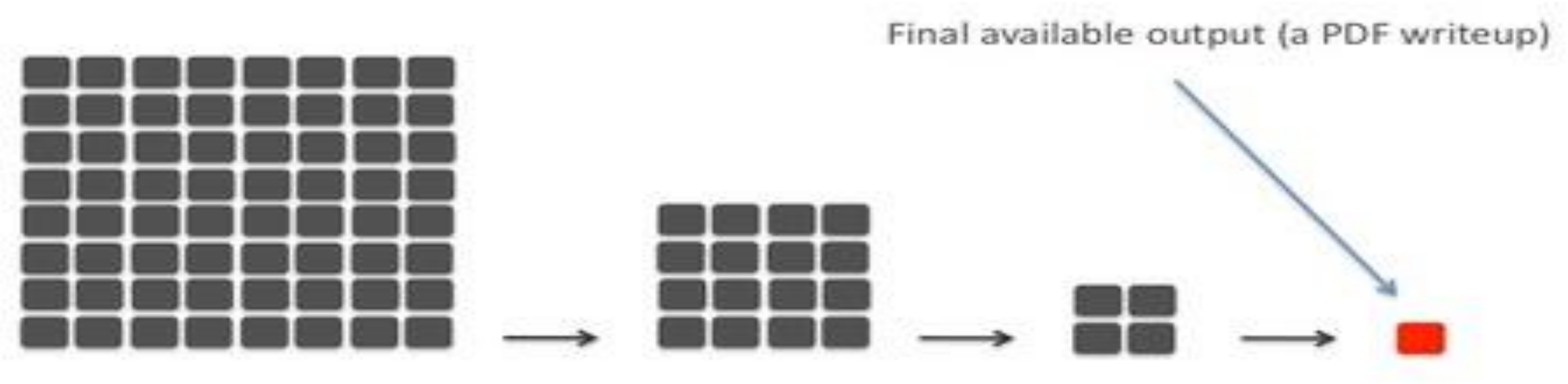
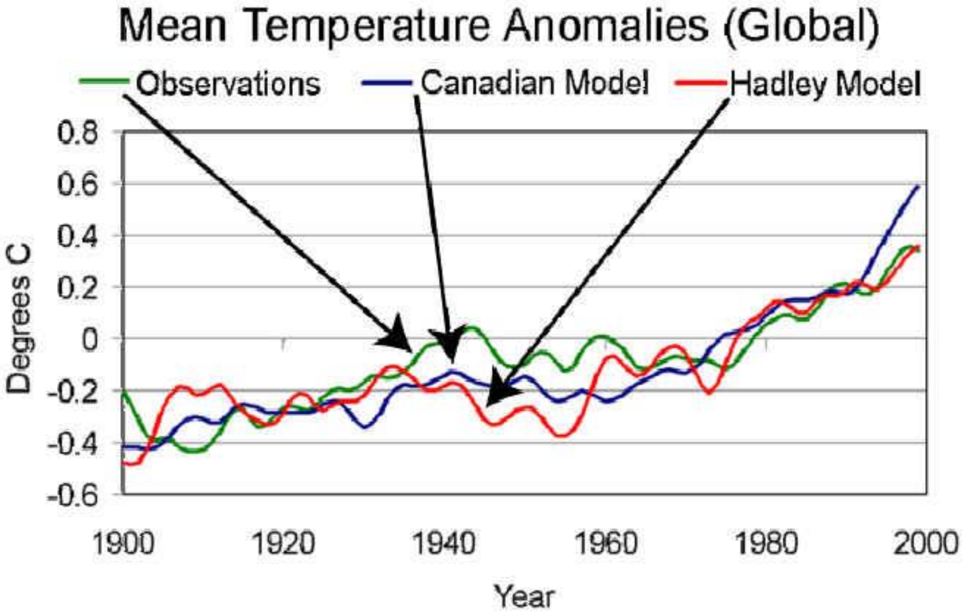


Illustration from http://blog.peerj.com/post/65345738206/changing-the-currency-of-science-to-solve-our-greatest

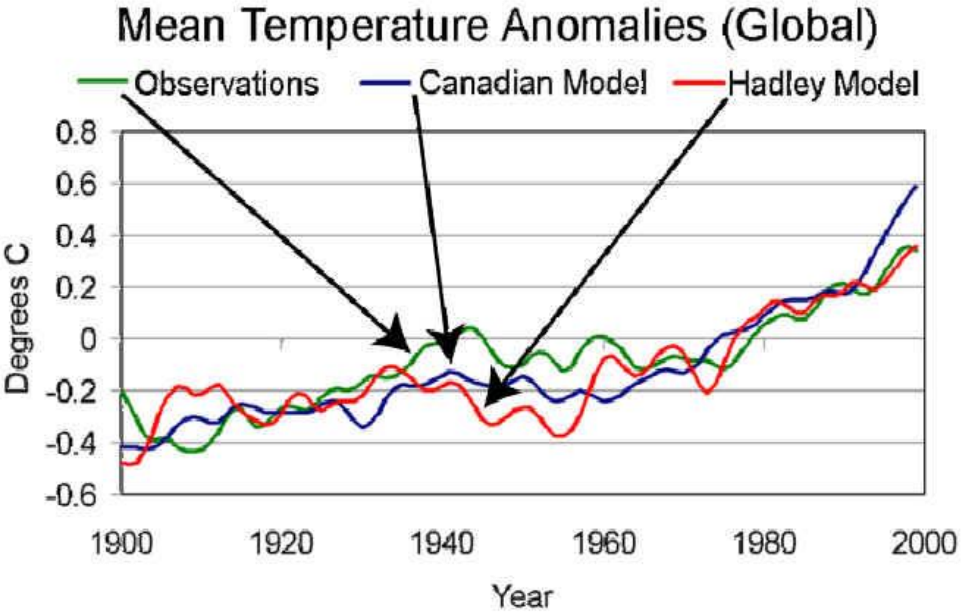
But open data improve science!

Example Climate Science



But open data improve science!

Example Climate Science



Captain Cook's log books help scientists chart climate change

Weather logs kept by Captain James Cook and other 18th and 19th century explorers are being used by scientists to predict the change in climate.

Log books from Captain Cook's Endeavour are being used to chart climate change Photo: AP

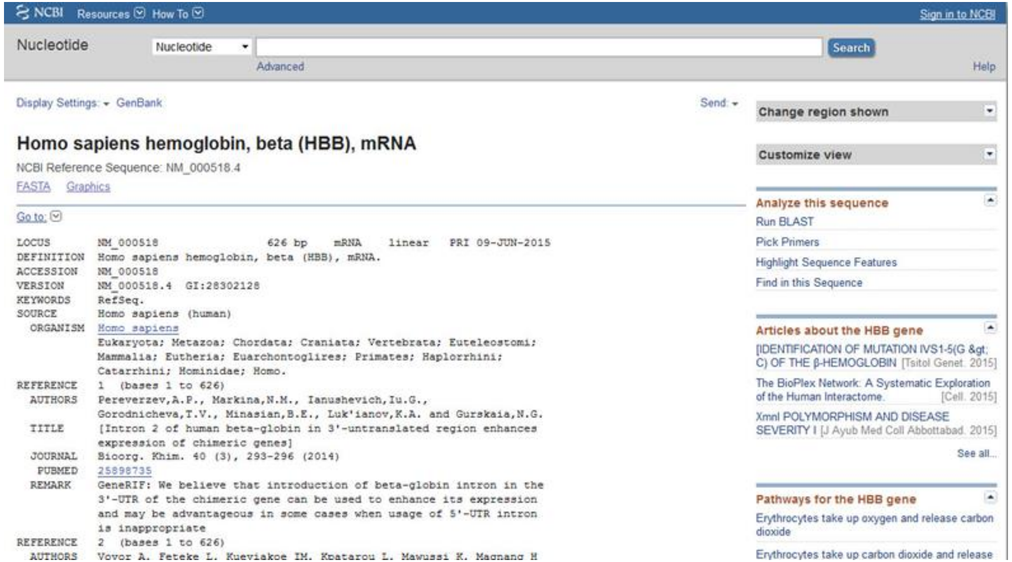
7:57AM BST 06 Oct 2009

Captain Cook's weather reports from Discovery and Resolution, made at noon each day on his voyages to unknown lands, William Bligh's Bounty

Bolivia's second largest lake dries up

China's dirty old town

But open data improve science!
Example DNA Sequencing Data



Homo sapiens hemoglobin, beta (HBB), mRNA

NCBI Reference Sequence: NM_000518.4

```
LOCUS       NM_000518                626 bp    mRNA    linear   PRI 09-JUN-2015
DEFINITION  Homo sapiens hemoglobin, beta (HBB), mRNA.
ACCESSION   NM_000518
VERSION     NM_000518.4  GI:28302128
KEYWORDS    RefSeq.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 626)
  AUTHORS   Pereverzev,A.P., Markina,N.M., Ianushevich,Iu.G.,
            Gorodnicheva,T.V., Minasian,B.E., Luk'ianov,K.A. and Gurskaia,N.G.
  TITLE     [Intron 2 of human beta-globin in 3'-untranslated region enhances
            expression of chimeric genes]
  JOURNAL   Bioorg. Khim. 40 (3), 293-296 (2014)
   PUBMED   25898735
  REMARK    GeneRIF: We believe that introduction of beta-globin intron in the
            3'-UTR of the chimeric gene can be used to enhance its expression
            and may be advantageous in some cases when usage of 5'-UTR intron
            is inappropriate
REFERENCE   2  (bases 1 to 626)
  AUTHORS   Vovor A, Feteke L, Kueviakoe IM, Kpatarou L, Mawussi K, Magnang H
```

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Articles about the HBB gene

[IDENTIFICATION OF MUTATION IVS1-5(G > C) OF THE β-HEMOGLOBIN [Tsitol Genet. 2015]

The BioPlex Network: A Systematic Exploration of the Human Interactome. [Cell. 2015]

XmnI POLYMORPHISM AND DISEASE SEVERITY I [J Ayub Med Coll Abbottabad. 2015]

See all...

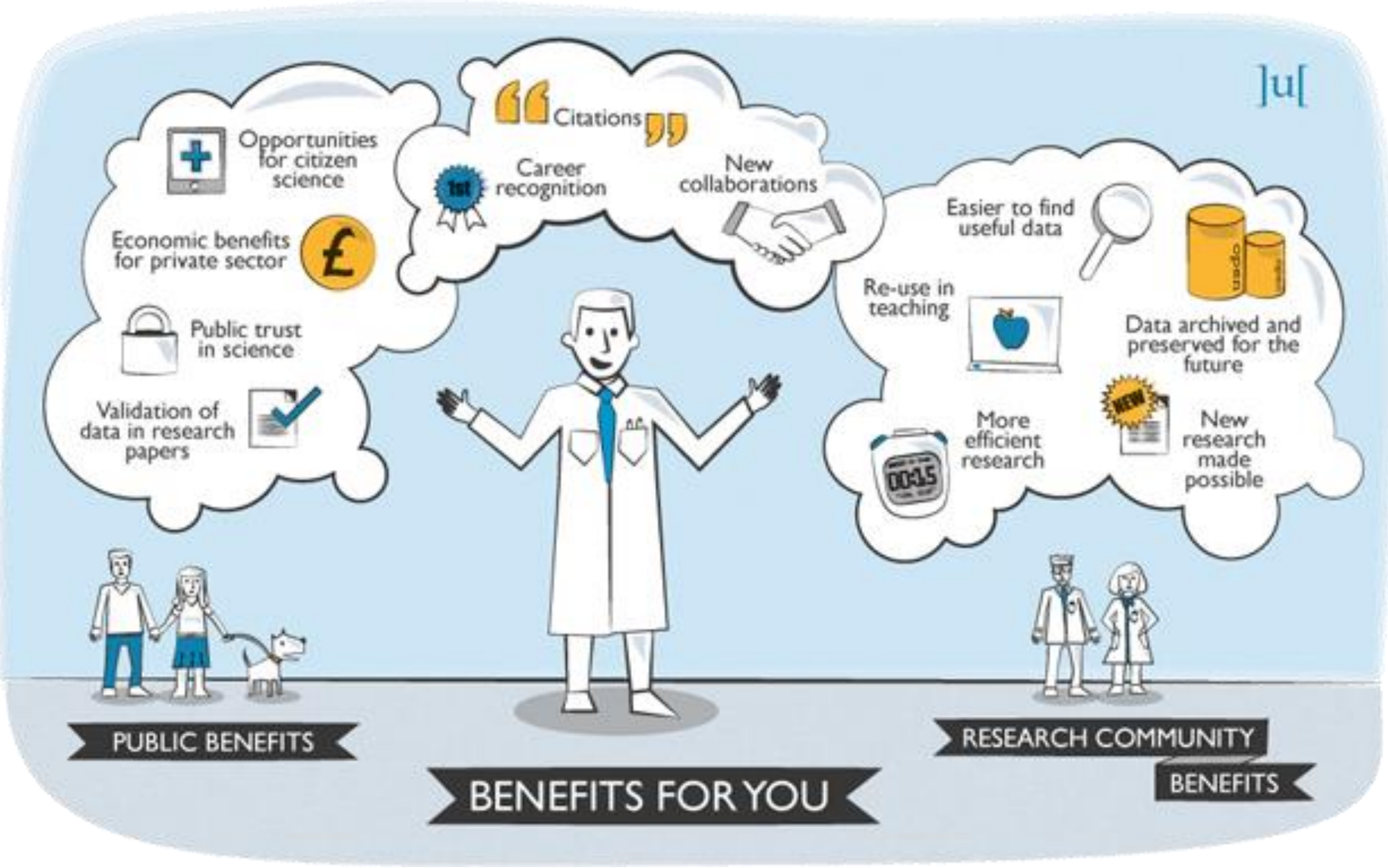
Pathways for the HBB gene

Erythrocytes take up oxygen and release carbon dioxide

Erythrocytes take up carbon dioxide and release

Human Genome Project: Sequencing and mapping of the human genome, completed 2003

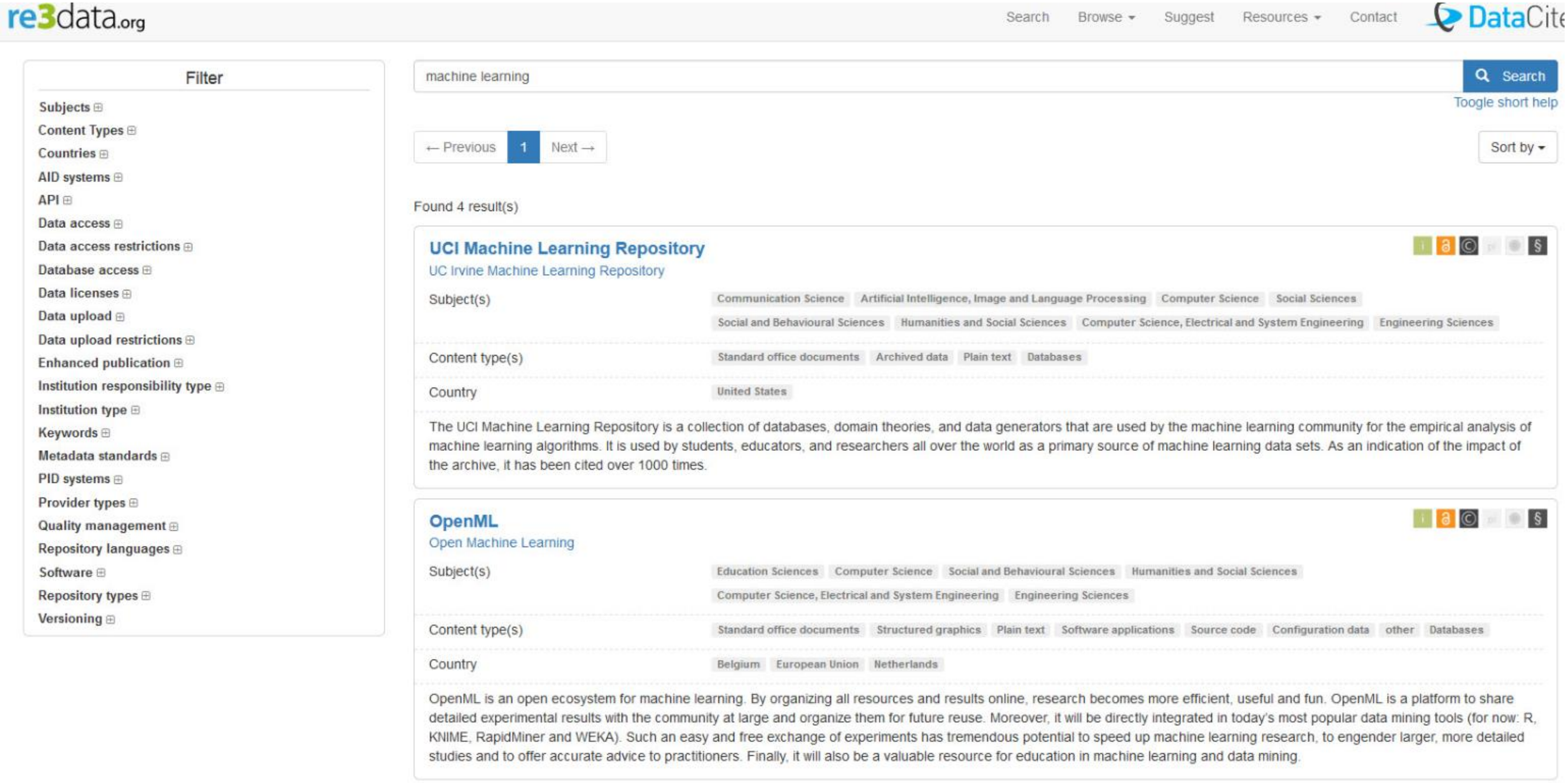
"The success of the genome project is in no small part due to the fact that the world's entire library of published DNA sequences has been an open access public source for the past 20 years. If sequences could be obtained only in the way that traditionally published work can be obtained - there would be no genome project“
(Prof. Patrick Brown, Biochemist at Stanford University, 2004)



Picture:
Hole, B. (2015). Open Science: A New publisher Perspective. Ubiquity press. (CC BY 4.0 International)

What kind of repositories are around for my field of research?

4eu+



machine learning

Found 4 result(s)

UCI Machine Learning Repository

UC Irvine Machine Learning Repository

Subject(s): Communication Science · Artificial Intelligence, Image and Language Processing · Computer Science · Social Sciences · Social and Behavioural Sciences · Humanities and Social Sciences · Computer Science, Electrical and System Engineering · Engineering Sciences

Content type(s): Standard office documents · Archived data · Plain text · Databases

Country: United States

The UCI Machine Learning Repository is a collection of databases, domain theories, and data generators that are used by the machine learning community for the empirical analysis of machine learning algorithms. It is used by students, educators, and researchers all over the world as a primary source of machine learning data sets. As an indication of the impact of the archive, it has been cited over 1000 times.

OpenML

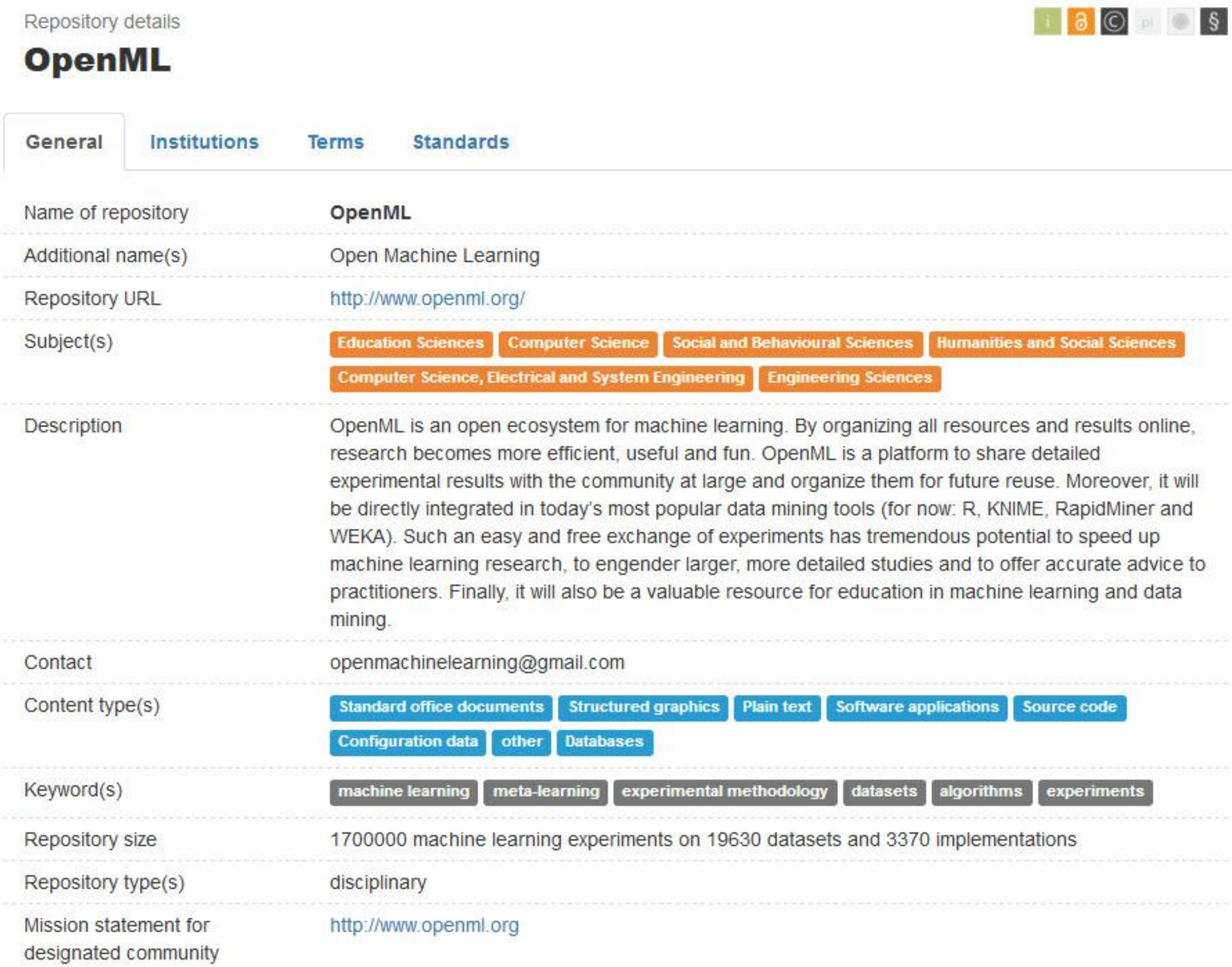
Open Machine Learning

Subject(s): Education Sciences · Computer Science · Social and Behavioural Sciences · Humanities and Social Sciences · Computer Science, Electrical and System Engineering · Engineering Sciences

Content type(s): Standard office documents · Structured graphics · Plain text · Software applications · Source code · Configuration data · other · Databases

Country: Belgium · European Union · Netherlands

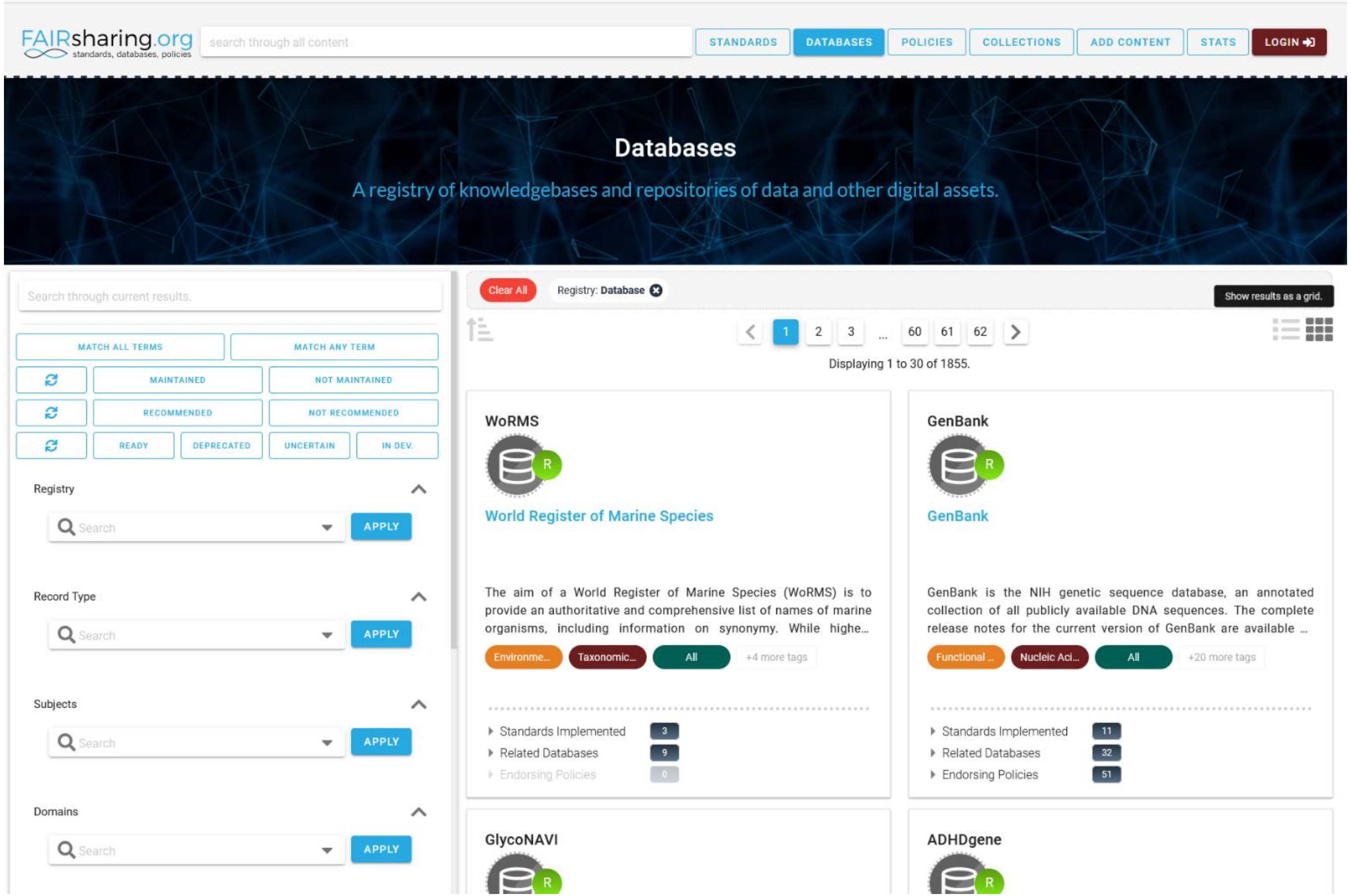
OpenML is an open ecosystem for machine learning. By organizing all resources and results online, research becomes more efficient, useful and fun. OpenML is a platform to share detailed experimental results with the community at large and organize them for future reuse. Moreover, it will be directly integrated in today's most popular data mining tools (for now: R, KNIME, RapidMiner and WEKA). Such an easy and free exchange of experiments has tremendous potential to speed up machine learning research, to engender larger, more detailed studies and to offer accurate advice to practitioners. Finally, it will also be a valuable resource for education in machine learning and data mining.



Repository details

OpenML

Name of repository	**OpenML**
Additional name(s)	Open Machine Learning
Repository URL	http://www.openml.org/
Subject(s)	Education Sciences Computer Science Social and Behavioural Sciences Humanities and Social Sciences Computer Science, Electrical and System Engineering Engineering Sciences
Description	OpenML is an open ecosystem for machine learning. By organizing all resources and results online, research becomes more efficient, useful and fun. OpenML is a platform to share detailed experimental results with the community at large and organize them for future reuse. Moreover, it will be directly integrated in today's most popular data mining tools (for now: R, KNIME, RapidMiner and WEKA). Such an easy and free exchange of experiments has tremendous potential to speed up machine learning research, to engender larger, more detailed studies and to offer accurate advice to practitioners. Finally, it will also be a valuable resource for education in machine learning and data mining.
Contact	openmachinelearning@gmail.com
Content type(s)	Standard office documents Structured graphics Plain text Software applications Source code Configuration data other Databases
Keyword(s)	machine learning meta-learning experimental methodology datasets algorithms experiments
Repository size	1700000 machine learning experiments on 19630 datasets and 3370 implementations
Repository type(s)	disciplinary
Mission statement for designated community	http://www.openml.org



Databases

A registry of knowledgebases and repositories of data and other digital assets.

Displaying 1 to 30 of 1855.

WoRMS

World Register of Marine Species

The aim of a World Register of Marine Species (WoRMS) is to provide an authoritative and comprehensive list of names of marine organisms, including information on synonymy. While highe...

Standards Implemented	3
Related Databases	9
Endorsing Policies	0

GenBank

GenBank

GenBank is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences. The complete release notes for the current version of GenBank are available ...

Standards Implemented	11
Related Databases	32
Endorsing Policies	51

GlycoNAVI

ADHDgene

NCBI

GEO Gene Expression Omnibus

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: https://www.coronavirus.gov .
Get the latest research from NIH: https://www.nih.gov/coronavirus.
Find NCBI SARS-CoV-2 literature, sequence, and clinical content: https://www.ncbi.nlm.nih.gov
cov-2/.

NCBI › GEO › **Accession Display** ⓘ Not logge

GEO help: Mouse over screen elements for information.

Scope: Self Format: HTML Amount: Quick GEO accession: GSE164332

Series GSE164332 Query DataSets for GSE164332

Status Public on Mar 09, 2021
Title Frontal Cortex whole transcriptome sequencing in COVID-19 old patients
Organism Homo sapiens
Experiment type Expression profiling by high throughput sequencing
Summary SARS-Cov-2 infection is frequently associated with Nervous System (NS) manifestations. However, it is not clear how SARS-CoV-2 can cause neurological dysfunctions and which molecular processes affects. In this work, we examined the frontal cortex tissue of patients who died of COVID-19 for the presence of SARS-CoV-2, comparing qRT-PCR with ddPCR. We also investigated the transcriptomic profile of frontal cortex from COVID-19 patients and matched controls by RNA-seq analysis to characterize the transcriptional signature.

Overall design Frontal c
healthy c
PCR expe

Contributor(s) Gagliardi
Visonà S,

Citation missing Has this s
Submission date Jan 06, 2
Last update date Mar 10, 2

PANGAEA.

Data Publisher for Earth & Environmental Science

Welcome to PANGAEA® Data Publisher

Our services are generally open for archiving, publishing, and re-usage of data. The World Data Center PANGAEA is member of the World Data System.

Search for measurement type, author name, project, taxa...

Published 3 days ago by Jérôme Klein

Chemotion Repository for molecules, reactions and research data

Visibility and Impact

- Publish your structures, attach your characterization data, and make them citable via DOI
- Automated registration at various scientific data providers
- Long-term archival - from scientists for scientists
- NEW *Now, publish your Reactions*

published
2567
Samples
450 under review
545 under embargo

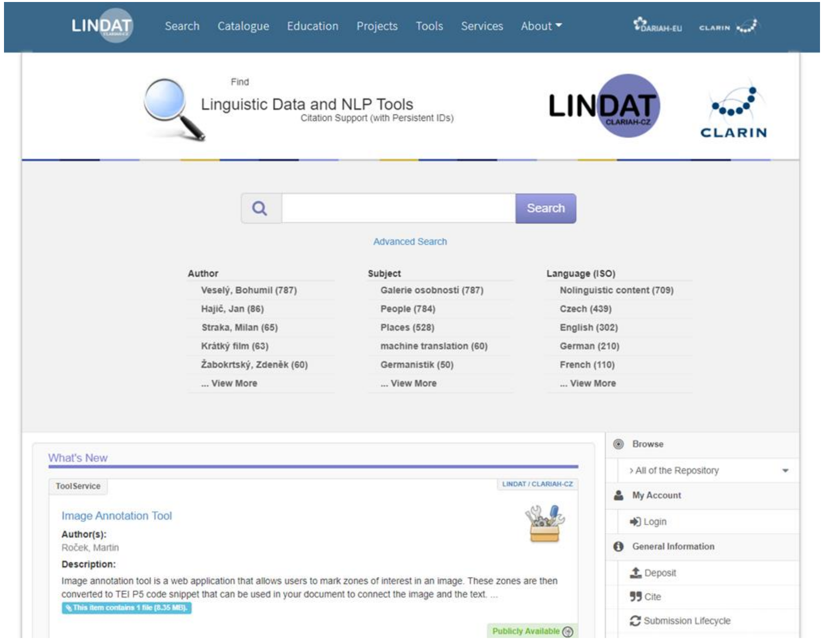
published
887
Reactions
468 under review
584 under embargo

8017
Analyses
Top 3: 3252 NMR
2105 IR
1613 Mass

A rough typology of data repositories:

- Controlled access repositories
- Subject- or domain-specific repositories
- Institutional repositories
- Generic repositories

Domain repositories



Author	Subject	Language (ISO)
Veselý, Bohumil (787)	Galerie osobností (787)	Nolinguistic content (709)
Hajič, Jan (86)	People (784)	Czech (439)
Straka, Milan (65)	Places (528)	English (302)
Krátký film (63)	machine translation (60)	German (210)
Žabokrtský, Zdeněk (60)	Germanistik (50)	French (110)
... View More	... View More	... View More

What's New

ToolService

LINDAT / CLARIAH-CZ

Image Annotation Tool

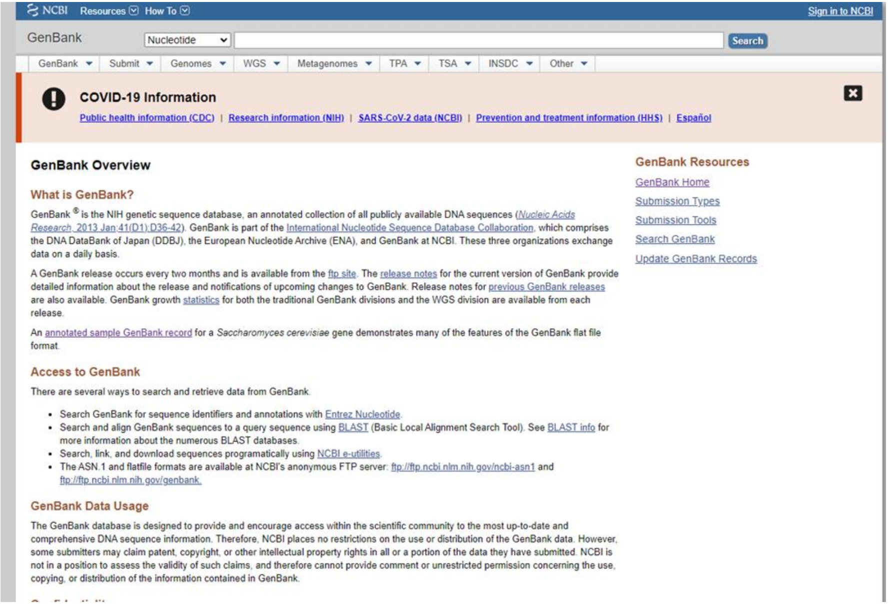
Author(s):
Roček, Martin

Description:
Image annotation tool is a web application that allows users to mark zones of interest in an image. These zones are then converted to TEI P5 code snippet that can be used in your document to connect the image and the text. ...

This item contains 1 file (8.35 MB).

Publicly Available

https://lindat.mff.cuni.cz/repository/



COVID-19 Information

Public health information (CDC) | Research information (NIH) | SARS-CoV-2 data (NCBI) | Prevention and treatment information (HHS) | Español

GenBank Overview

What is GenBank?

GenBank ® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences (*Nucleic Acids Research*, 2013 Jan;41(D1):D36-42). GenBank is part of the International Nucleotide Sequence Database Collaboration, which comprises the DNA DataBank of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank at NCBI. These three organizations exchange data on a daily basis.

A GenBank release occurs every two months and is available from the ftp site. The release notes for the current version of GenBank provide detailed information about the release and notifications of upcoming changes to GenBank. Release notes for previous GenBank releases are also available. GenBank growth statistics for both the traditional GenBank divisions and the WGS division are available from each release.

An annotated sample GenBank record for a *Saccharomyces cerevisiae* gene demonstrates many of the features of the GenBank flat file format.

Access to GenBank

There are several ways to search and retrieve data from GenBank.

- Search GenBank for sequence identifiers and annotations with Entrez Nucleotide.
- Search and align GenBank sequences to a query sequence using BLAST (Basic Local Alignment Search Tool). See BLAST info for more information about the numerous BLAST databases.
- Search, link, and download sequences programatically using NCBI e-utilities.
- The ASN.1 and flatfile formats are available at NCBI's anonymous FTP server: ftp://ftp.ncbi.nlm.nih.gov/ncbi-asn1 and ftp://ftp.ncbi.nlm.nih.gov/genbank.

GenBank Data Usage

The GenBank database is designed to provide and encourage access within the scientific community to the most up-to-date and comprehensive DNA sequence information. Therefore, NCBI places no restrictions on the use or distribution of the GenBank data. However, some submitters may claim patent, copyright, or other intellectual property rights in all or a portion of the data they have submitted. NCBI is not in a position to assess the validity of such claims, and therefore cannot provide comment or unrestricted permission concerning the use, copying, or distribution of the information contained in GenBank.

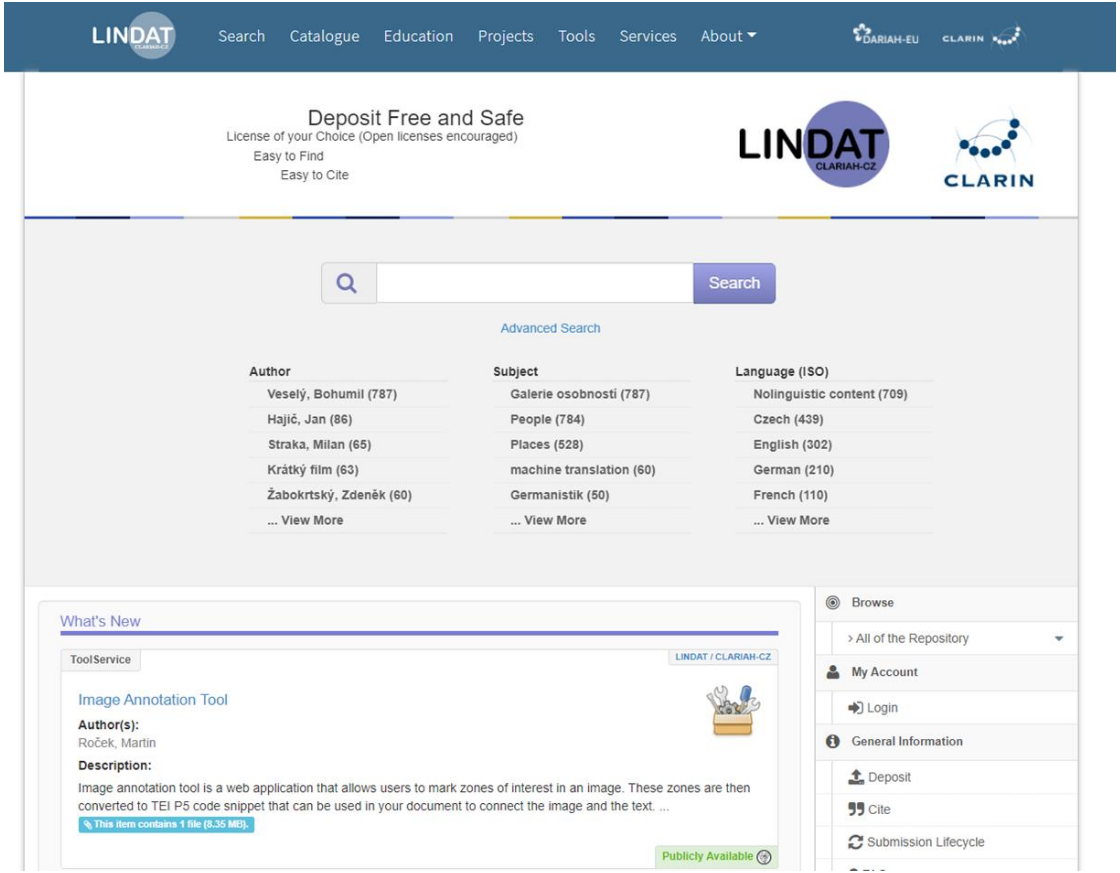
GenBank Resources

- GenBank Home
- Submission Types
- Submission Tools
- Search GenBank
- Update GenBank Records

https://www.ncbi.nlm.nih.gov/genbank/

LINDAT/CLARIAH-CZ

- linguistics / digital humanities repository
- repository + services



Deposit Free and Safe
License of your Choice (Open licenses encouraged)
Easy to Find
Easy to Cite

Author	Subject	Language (ISO)
Veselý, Bohumil (787)	Galerie osobností (787)	Nolinguistic content (709)
Hajič, Jan (86)	People (784)	Czech (439)
Straka, Milan (65)	Places (528)	English (302)
Krátký film (63)	machine translation (60)	German (210)
Žabokrtský, Zdeněk (60)	Germanistik (50)	French (110)
... View More	... View More	... View More

What's New

ToolService

LINDAT / CLARIAH-CZ

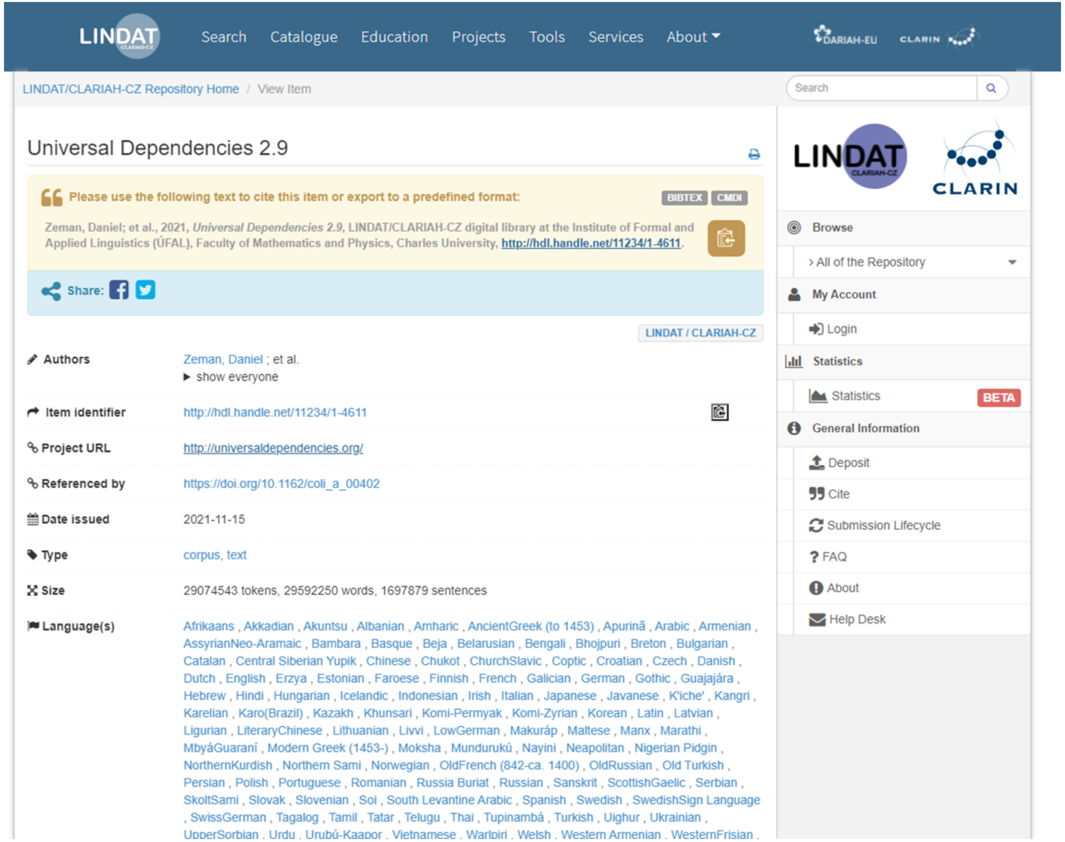
Image Annotation Tool

Author(s):
Roček, Martin

Description:
Image annotation tool is a web application that allows users to mark zones of interest in an image. These zones are then converted to TEI P5 code snippet that can be used in your document to connect the image and the text. ...

This item contains 1 file (8.35 MB).

Publicly Available



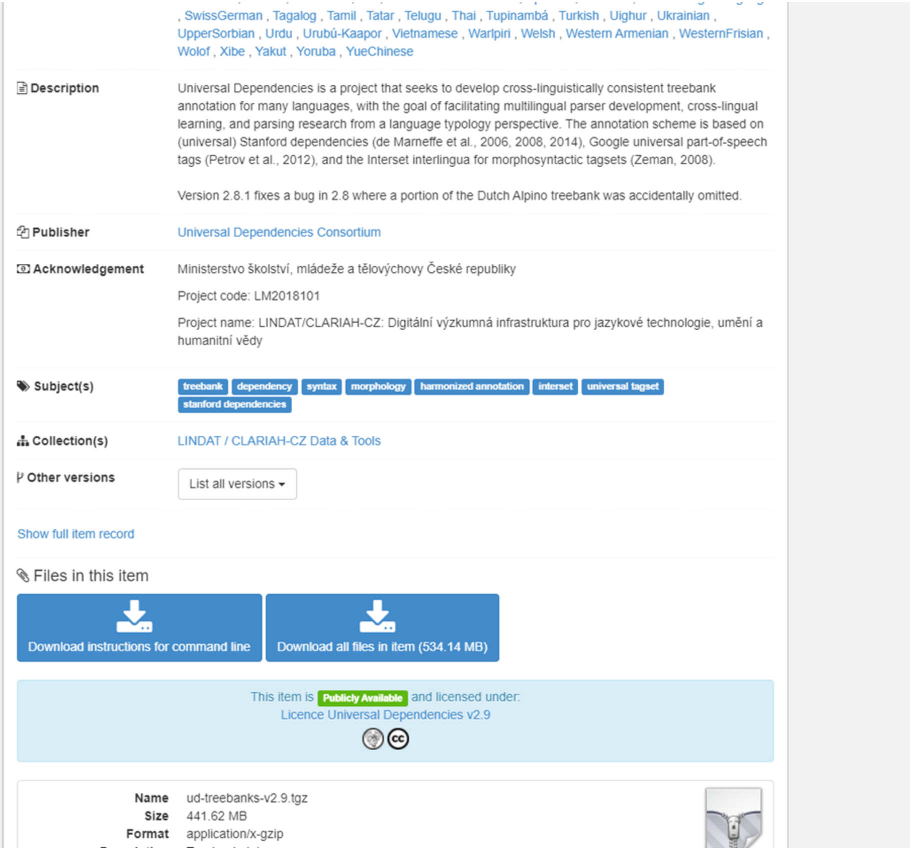
Universal Dependencies 2.9

Please use the following text to cite this item or export to a predefined format:

Zeman, Daniel; et al., 2021, *Universal Dependencies 2.9*, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, http://hdl.handle.net/11234/1-4611.

Authors	Zeman, Daniel ; et al.
Item identifier	http://hdl.handle.net/11234/1-4611
Project URL	http://universaldependencies.org/
Referenced by	https://doi.org/10.1162/coli_a_00402
Date issued	2021-11-15
Type	corpus, text
Size	29074543 tokens, 29592250 words, 1697879 sentences
Language(s)	Afrikaans , Akkadian , Akuntsu , Albanian , Amharic , AncientGreek (to 1453) , Apurinã , Arabic , Armenian , AssyrianNeo-Aramaic , Bambara , Basque , Beja , Belarusian , Bengali , Bhojpuri , Breton , Bulgarian , Catalan , Central Siberian Yupik , Chinese , Chukot , ChurchSlavic , Coptic , Croatian , Czech , Danish , Dutch , English , Erzya , Estonian , Faroese , Finnish , French , Galician , German , Gothic , Guajajára , Hebrew , Hindi , Hungarian , Icelandic , Indonesian , Irish , Italian , Japanese , Javanese , K'iche' , Kangri , Karelian , Karo(Brazil) , Kazakh , Khunsari , Komi-Permyak , Komi-Zyrian , Korean , Latin , Latvian , Ligurian , LiteraryChinese , Lithuanian , Livvi , LowGerman , Makuráp , Maltese , Manx , Marathi , MbyáGuaraní , Modern Greek (1453-) , Moksha , Mundurukú , Nayini , Neapolitan , Nigerian Pidgin , NorthernKurdish , Northern Sami , Norwegian , OldFrench (842-ca. 1400) , OldRussian , Old Turkish , Persian , Polish , Portuguese , Romanian , Russia Buriat , Russian , Sanskrit , ScottishGaelic , Serbian , SkoltSami , Slovak , Slovenian , Soi , South Levantine Arabic , Spanish , Swedish , SwedishSign Language , SwissGerman , Tagalog , Tamil , Tatar , Telugu , Thai , Tupinambá , Turkish , Uighur , Ukrainian , UpperSorbian , Urdu , Urubú-Kaapor , Vietnamese , Warlpiri , Welsh , Western Armenian , WesternFrisian ,

- **reference to the article**
- **languages are important**
- **PID is handle (not DOI)**


, SwissGerman , Tagalog , Tamil , Tatar , Telugu , Thai , Tupinambá , Turkish , Uighur , Ukrainian , UpperSorbian , Urdu , Urubú-Kaapor , Vietnamese , Warlpiri , Welsh , Western Armenian , WesternFrisian , Wolof , Xibe , Yakut , Yoruba , YueChinese

Description Universal Dependencies is a project that seeks to develop cross-linguistically consistent treebank annotation for many languages, with the goal of facilitating multilingual parser development, cross-lingual learning, and parsing research from a language typology perspective. The annotation scheme is based on (universal) Stanford dependencies (de Marneffe et al., 2006, 2008, 2014), Google universal part-of-speech tags (Petrov et al., 2012), and the Interset interlingua for morphosyntactic tagsets (Zeman, 2008).

Version 2.8.1 fixes a bug in 2.8 where a portion of the Dutch Alpino treebank was accidentally omitted.

Publisher Universal Dependencies Consortium

Acknowledgement Ministerstvo školství, mládeže a tělovýchovy České republiky

Project code: LM2018101

Project name: LINDAT/CLARIAH-CZ: Digitální výzkumná infrastruktura pro jazykové technologie, umění a humanitní vědy

Subject(s) treebank dependency syntax morphology harmonized annotation interset universal tagset stanford dependencies

Collection(s) LINDAT / CLARIAH-CZ Data & Tools

Other versions List all versions ▾

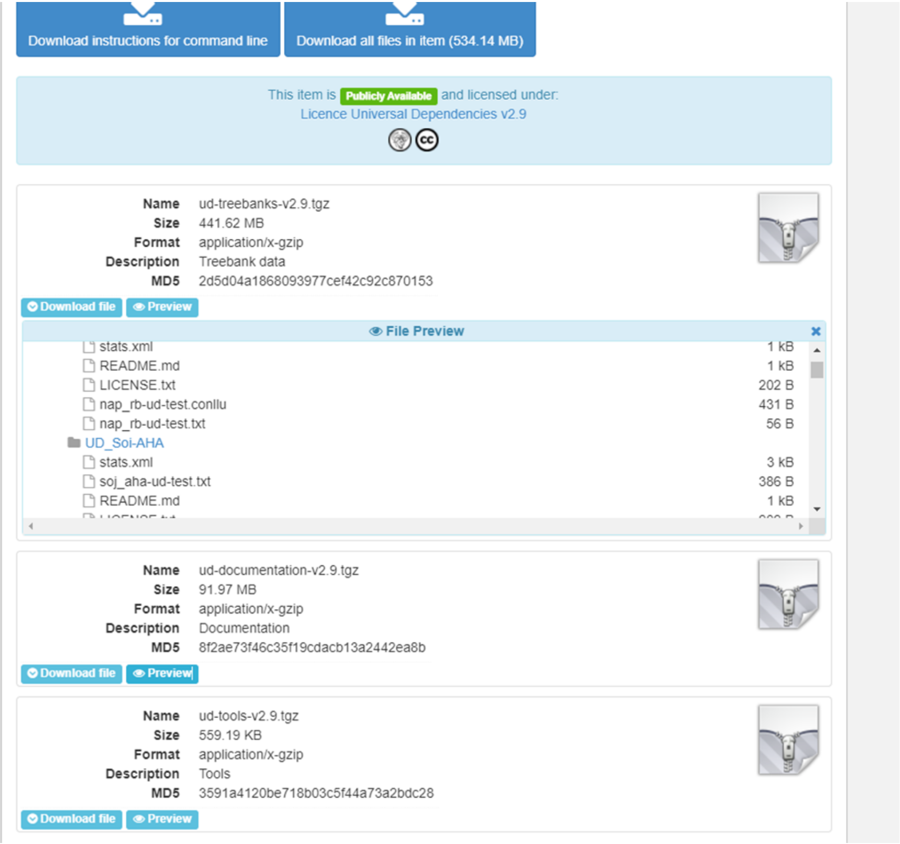
Show full item record

Files in this item

This item is Publicly Available and licensed under:
Licence Universal Dependencies v2.9

Name ud-treebanks-v2.9.tgz
Size 441.62 MB
Format application/x-gzip

- description
- subjects (keywords)
- different versions of the dataset
- license



This item is Publicly Available and licensed under:
Licence Universal Dependencies v2.9

Name	ud-treebanks-v2.9.tgz
Size	441.62 MB
Format	application/x-gzip
Description	Treebank data
MD5	2d5d04a1868093977cef42c92c870153

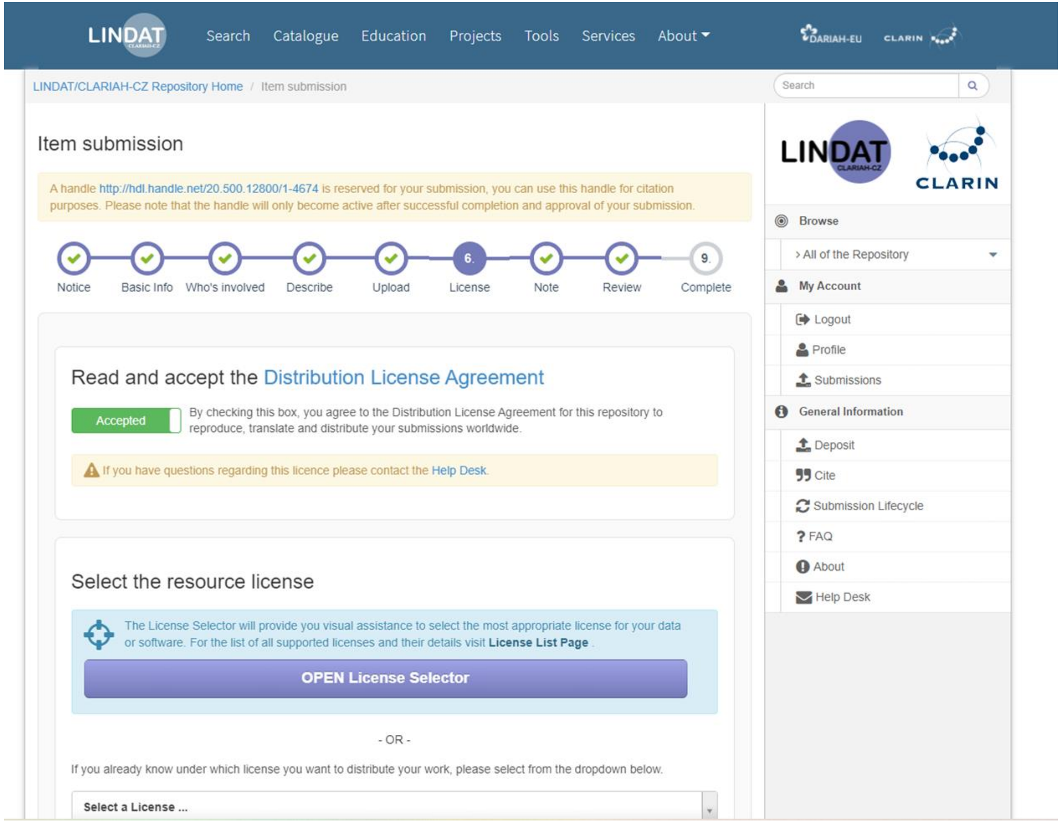
stats.xml	1 kB
README.md	1 kB
LICENSE.txt	202 B
nap_rb-ud-test.conllu	431 B
nap_rb-ud-test.txt	56 B
UD_Soi-AHA	
stats.xml	3 kB
soj_aha-ud-test.txt	386 B
README.md	1 kB

Name	ud-documentation-v2.9.tgz
Size	91.97 MB
Format	application/x-gzip
Description	Documentation
MD5	8f2ae73f46c35f19cdacb13a2442ea8b

Name	ud-tools-v2.9.tgz
Size	559.19 KB
Format	application/x-gzip
Description	Tools
MD5	3591a4120be718b03c5f44a73a2bdc28

- **file preview**
- **this dataset = data + documentation + tools**

4eu+



Item submission

A handle http://hdl.handle.net/20.500.12800/1-4674 is reserved for your submission, you can use this handle for citation purposes. Please note that the handle will only become active after successful completion and approval of your submission.

Read and accept the Distribution License Agreement

By checking this box, you agree to the Distribution License Agreement for this repository to reproduce, translate and distribute your submissions worldwide.

If you have questions regarding this licence please contact the Help Desk.

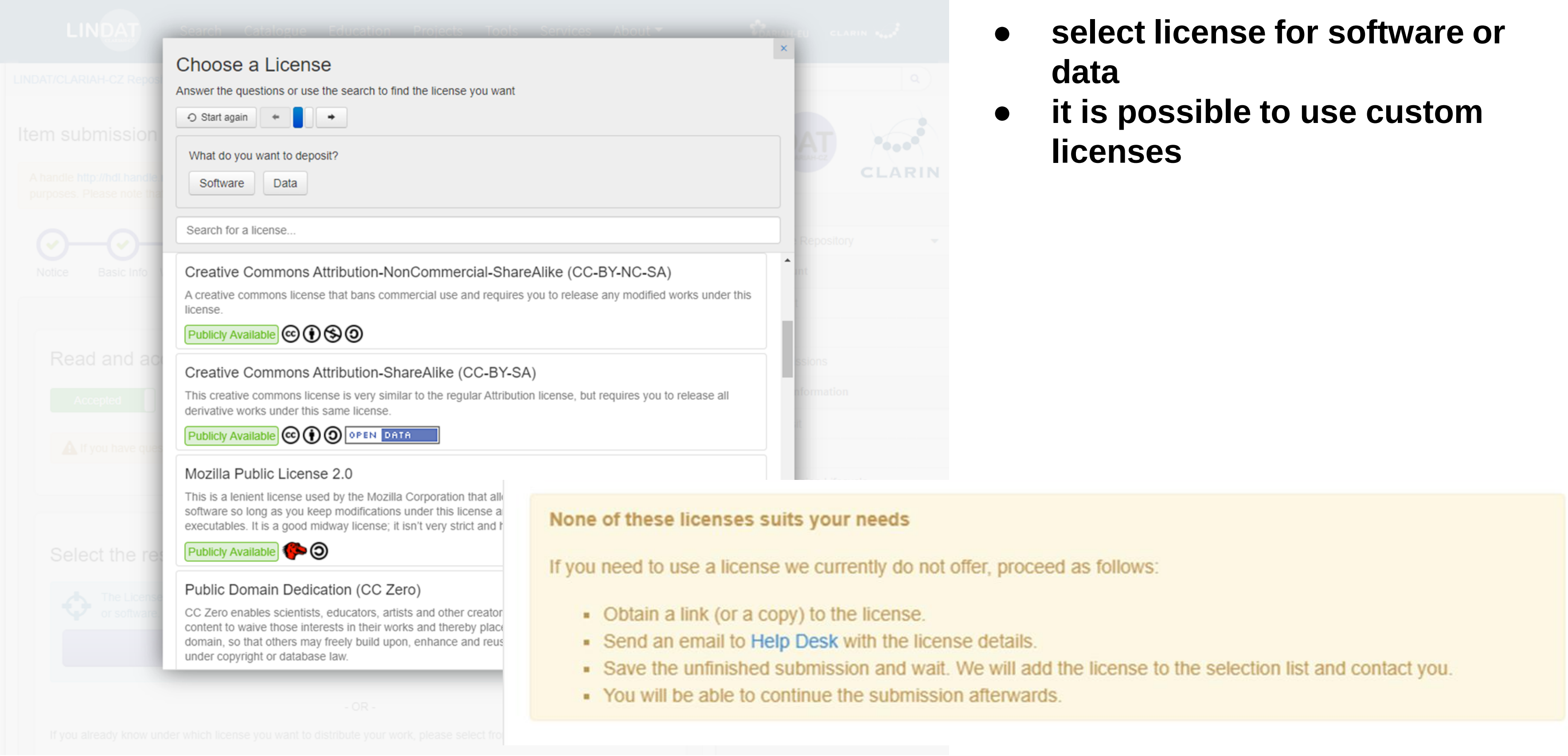
Select the resource license

The License Selector will provide you visual assistance to select the most appropriate license for your data or software. For the list of all supported licenses and their details visit **License List Page** .

- OR -

If you already know under which license you want to distribute your work, please select from the dropdown below.

- **login with your institutional credentials**
- **steps of new submission**
- **license is important →…**



Choose a License

Answer the questions or use the search to find the license you want

What do you want to deposit?

Search for a license...

Creative Commons Attribution-NonCommercial-ShareAlike (CC-BY-NC-SA)

A creative commons license that bans commercial use and requires you to release any modified works under this license.

Creative Commons Attribution-ShareAlike (CC-BY-SA)

This creative commons license is very similar to the regular Attribution license, but requires you to release all derivative works under this same license.

Mozilla Public License 2.0

This is a lenient license used by the Mozilla Corporation that all software so long as you keep modifications under this license a executables. It is a good midway license; it isn't very strict and

Public Domain Dedication (CC Zero)

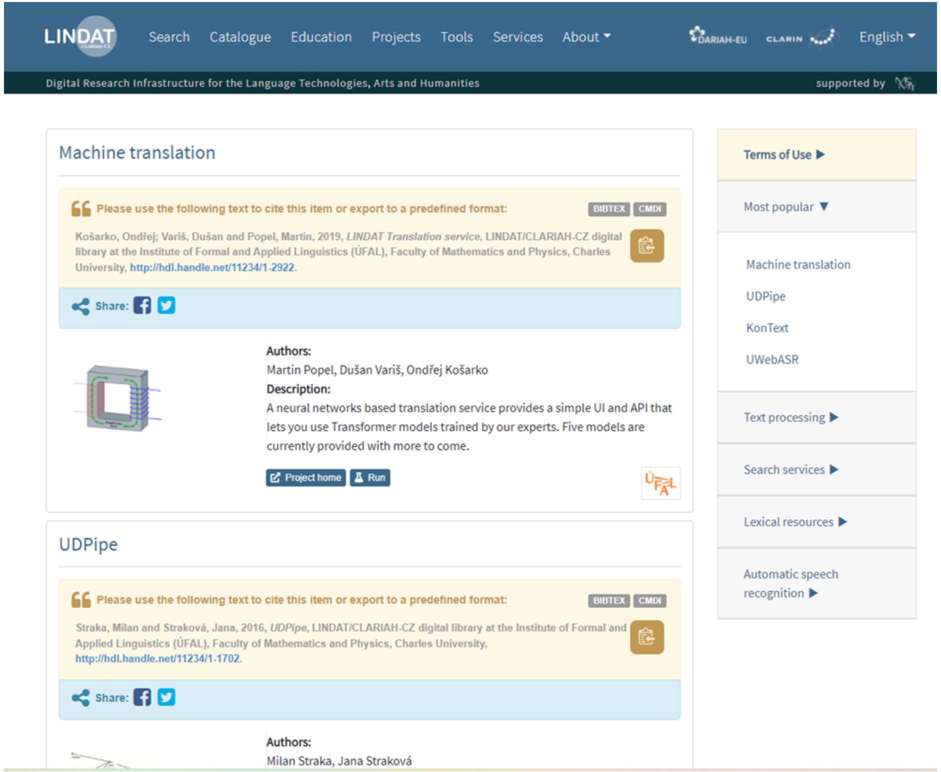
CC Zero enables scientists, educators, artists and other creator content to waive those interests in their works and thereby plac domain, so that others may freely build upon, enhance and reus under copyright or database law.

None of these licenses suits your needs

If you need to use a license we currently do not offer, proceed as follows:

- Obtain a link (or a copy) to the license.
- Send an email to Help Desk with the license details.
- Save the unfinished submission and wait. We will add the license to the selection list and contact you.
- You will be able to continue the submission afterwards.

- **select license for software or data**
- **it is possible to use custom licenses**



Digital Research Infrastructure for the Language Technologies, Arts and Humanities

Machine translation

Please use the following text to cite this item or export to a predefined format:

Košarko, Ondřej; Variš, Dušan and Popel, Martin, 2019, *LINDAT Translation service*, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, http://hdl.handle.net/11234/1-2922.

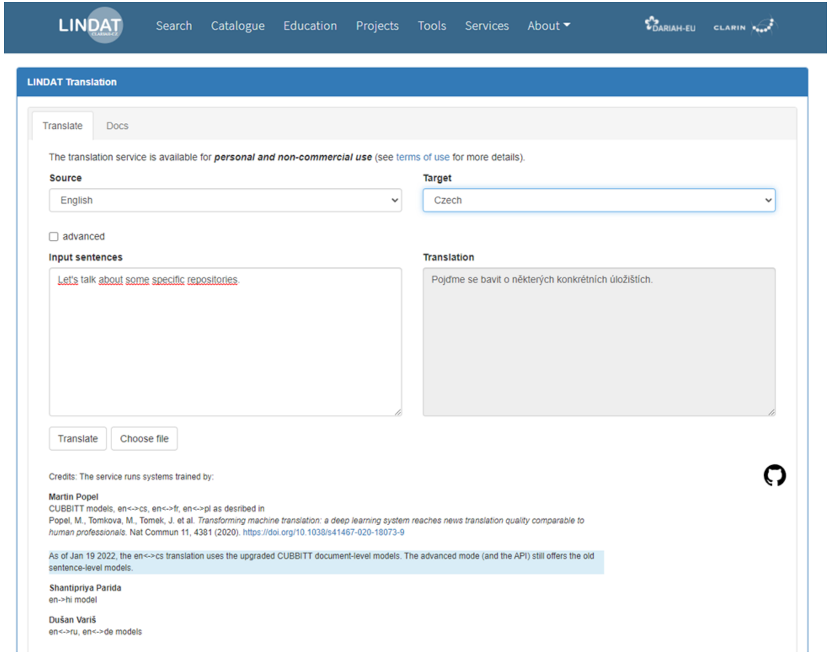
Authors:
Martin Popel, Dušan Variš, Ondřej Košarko
Description:
A neural networks based translation service provides a simple UI and API that lets you use Transformer models trained by our experts. Five models are currently provided with more to come.

UDPipe

Please use the following text to cite this item or export to a predefined format:

Straka, Milan and Straková, Jana, 2016, *UDPipe*, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, http://hdl.handle.net/11234/1-1702.

Authors:
Milan Straka, Jana Straková



LINDAT Translation

The translation service is available for ***personal and non-commercial use*** (see terms of use for more details).

Source: English
Target: Czech

☐ advanced

Input sentences
Let's talk about some specific repositories.

Translation
Pojďme se bavit o některých konkrétních úložištích.

Credits: The service runs systems trained by:

Martin Popel
CUBBITT models, en<->cs, en<->fr, en<->pl as desribed in
Popel, M., Tomkova, M., Tomek, J. et al. *Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals.* Nat Commun 11, 4381 (2020). https://doi.org/10.1038/s41467-020-18073-9

As of Jan 19 2022, the en<->cs translation uses the upgraded CUBBITT document-level models. The advanced mode (and the API) still offers the old sentence-level models.

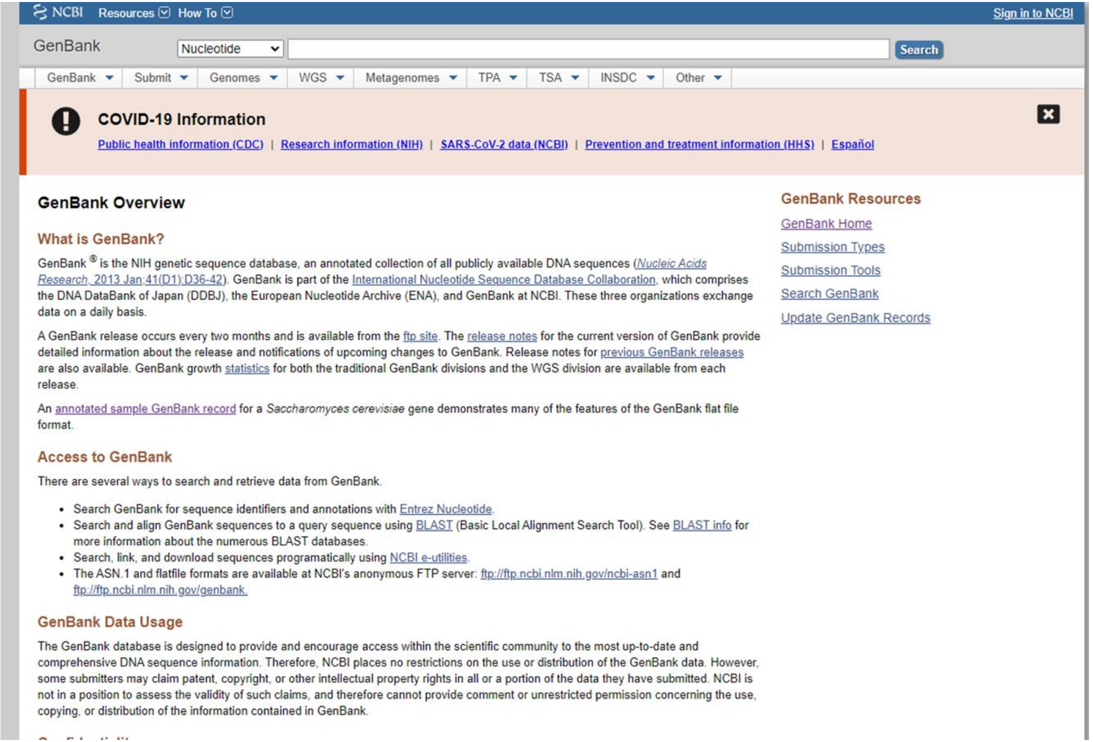
Shantipriya Parida
en->hi model

Dušan Variš
en<->ru, en<->de models

- **repositories can be connected to infrastructures and provide services**

GenBank

“GenBank is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences”



COVID-19 Information

Public health information (CDC) | Research information (NIH) | SARS-CoV-2 data (NCBI) | Prevention and treatment information (HHS) | Español

GenBank Overview

What is GenBank?

GenBank ® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences (*Nucleic Acids Research*, 2013 Jan;41(D1):D36-42). GenBank is part of the International Nucleotide Sequence Database Collaboration, which comprises the DNA DataBank of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank at NCBI. These three organizations exchange data on a daily basis.

A GenBank release occurs every two months and is available from the ftp site. The release notes for the current version of GenBank provide detailed information about the release and notifications of upcoming changes to GenBank. Release notes for previous GenBank releases are also available. GenBank growth statistics for both the traditional GenBank divisions and the WGS division are available from each release.

An annotated sample GenBank record for a *Saccharomyces cerevisiae* gene demonstrates many of the features of the GenBank flat file format.

Access to GenBank

There are several ways to search and retrieve data from GenBank.

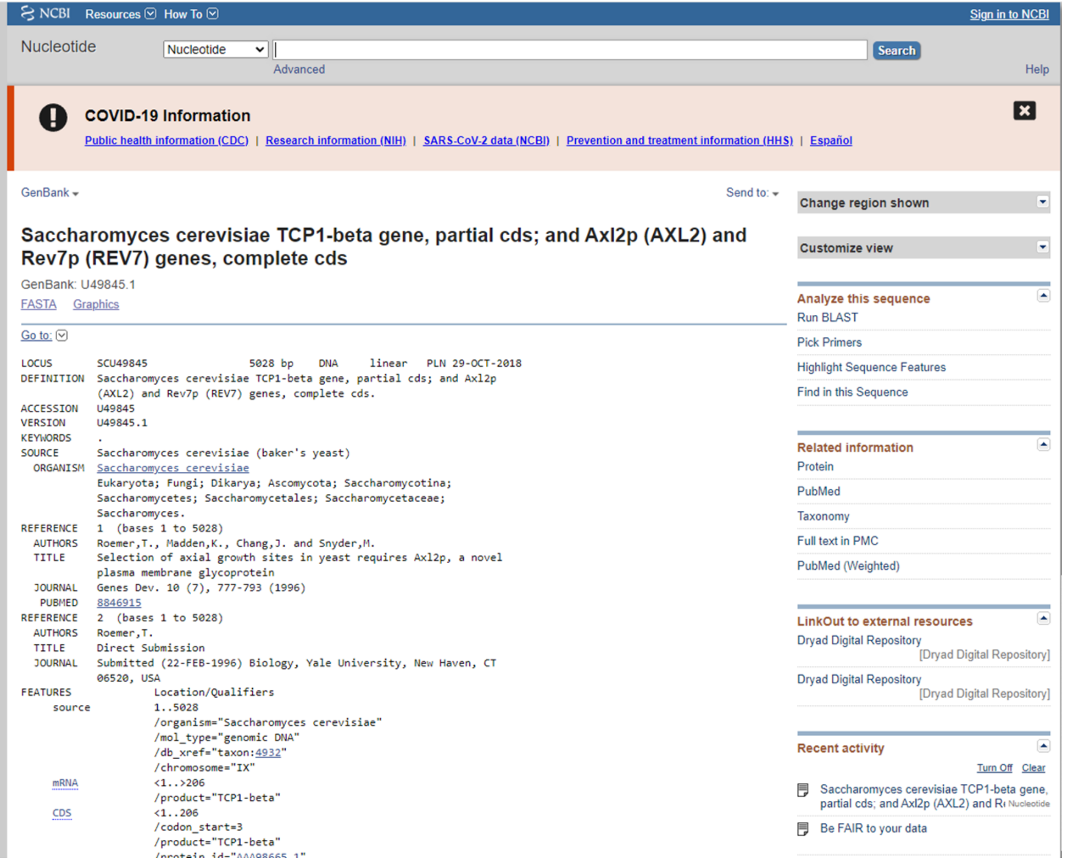
- Search GenBank for sequence identifiers and annotations with Entrez Nucleotide.
- Search and align GenBank sequences to a query sequence using BLAST (Basic Local Alignment Search Tool). See BLAST info for more information about the numerous BLAST databases.
- Search, link, and download sequences programatically using NCBI e-utilities.
- The ASN.1 and flatfile formats are available at NCBI's anonymous FTP server: ftp://ftp.ncbi.nlm.nih.gov/ncbi-asn1 and ftp://ftp.ncbi.nlm.nih.gov/genbank.

GenBank Data Usage

The GenBank database is designed to provide and encourage access within the scientific community to the most up-to-date and comprehensive DNA sequence information. Therefore, NCBI places no restrictions on the use or distribution of the GenBank data. However, some submitters may claim patent, copyright, or other intellectual property rights in all or a portion of the data they have submitted. NCBI is not in a position to assess the validity of such claims, and therefore cannot provide comment or unrestricted permission concerning the use, copying, or distribution of the information contained in GenBank.

GenBank Resources

- GenBank Home
- Submission Types
- Submission Tools
- Search GenBank
- Update GenBank Records



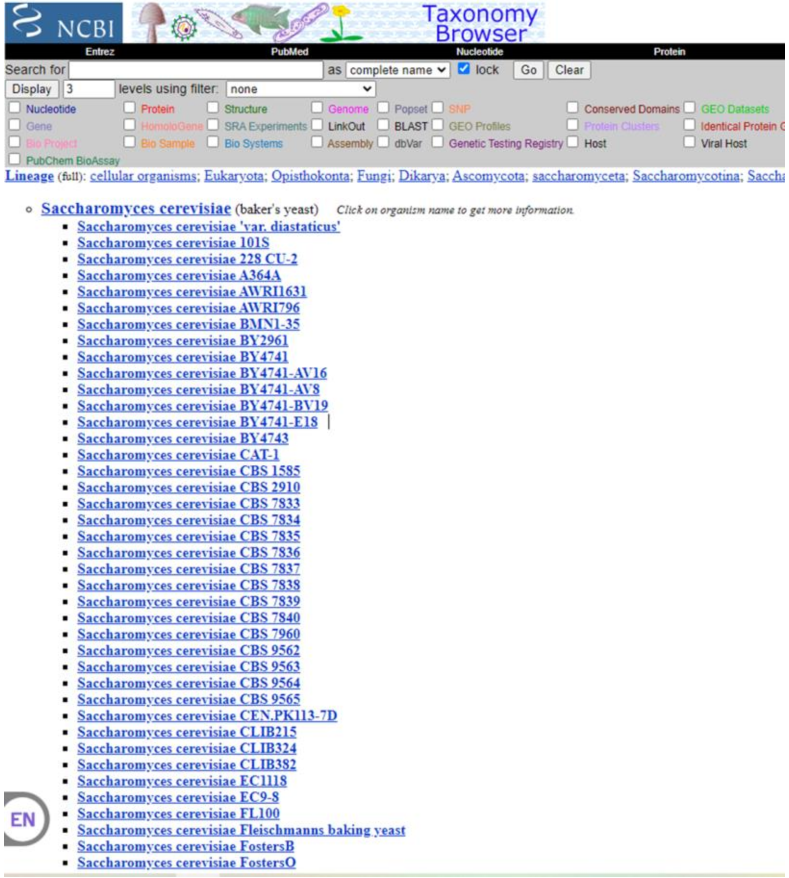
Saccharomyces cerevisiae TCP1-beta gene, partial cds; and Axl2p (AXL2) and Rev7p (REV7) genes, complete cds

GenBank: U49845.1

```
LOCUS       SCU49845                5028 bp    DNA     linear   PLN 29-OCT-2018
DEFINITION  Saccharomyces cerevisiae TCP1-beta gene, partial cds; and Axl2p
            (AXL2) and Rev7p (REV7) genes, complete cds.
ACCESSION   U49845
VERSION     U49845.1
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Dikarya; Ascomycota; Saccharomycotina;
            Saccharomycetes; Saccharomycetales; Saccharomycetaceae;
            Saccharomyces.
REFERENCE   1  (bases 1 to 5028)
  AUTHORS   Roemer,T., Madden,K., Chang,J. and Snyder,M.
  TITLE     Selection of axial growth sites in yeast requires Axl2p, a novel
            plasma membrane glycoprotein
  JOURNAL   Genes Dev. 10 (7), 777-793 (1996)
   PUBMED   8846915
REFERENCE   2  (bases 1 to 5028)
  AUTHORS   Roemer,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1996) Biology, Yale University, New Haven, CT
            06520, USA
FEATURES             Location/Qualifiers
     source          1..5028
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4932"
                     /chromosome="IX"
     mRNA            <1..>206
                     /product="TCP1-beta"
     CDS             <1..206
                     /codon_start=3
                     /product="TCP1-beta"
```

- GenBank identifier (U49845)
- reference to article is important
- specific metadata fields
- ORGANISM from taxonomy

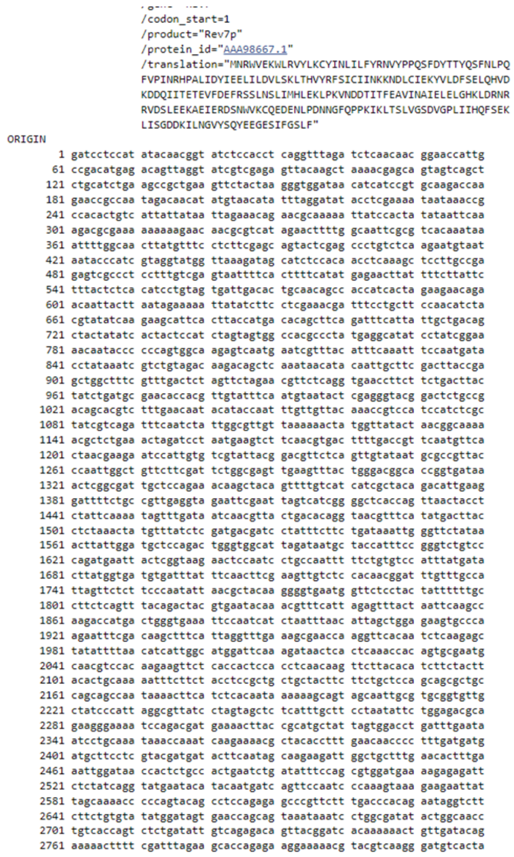
→ ...



Lineage (full): cellular organisms; Eukaryota; Opisthokonta; Fungi; Dikarya; Ascomycota; saccharomyceta; Saccharomycotina; Sacch

- **Saccharomyces cerevisiae** (baker's yeast) *Click on organism name to get more information.*
  - Saccharomyces cerevisiae 'var. diastaticus'
  - Saccharomyces cerevisiae 101S
  - Saccharomyces cerevisiae 228 CU-2
  - Saccharomyces cerevisiae A364A
  - Saccharomyces cerevisiae AWRI1631
  - Saccharomyces cerevisiae AWRI796
  - Saccharomyces cerevisiae BMN1-35
  - Saccharomyces cerevisiae BY2961
  - Saccharomyces cerevisiae BY4741
  - Saccharomyces cerevisiae BY4741-AV16
  - Saccharomyces cerevisiae BY4741-AV8
  - Saccharomyces cerevisiae BY4741-BV19
  - Saccharomyces cerevisiae BY4741-E18
  - Saccharomyces cerevisiae BY4743
  - Saccharomyces cerevisiae CAT-1
  - Saccharomyces cerevisiae CBS 1585
  - Saccharomyces cerevisiae CBS 2910
  - Saccharomyces cerevisiae CBS 7833
  - Saccharomyces cerevisiae CBS 7834
  - Saccharomyces cerevisiae CBS 7835
  - Saccharomyces cerevisiae CBS 7836
  - Saccharomyces cerevisiae CBS 7837
  - Saccharomyces cerevisiae CBS 7838
  - Saccharomyces cerevisiae CBS 7839
  - Saccharomyces cerevisiae CBS 7840
  - Saccharomyces cerevisiae CBS 7960
  - Saccharomyces cerevisiae CBS 9562
  - Saccharomyces cerevisiae CBS 9563
  - Saccharomyces cerevisiae CBS 9564
  - Saccharomyces cerevisiae CBS 9565
  - Saccharomyces cerevisiae CEN.PK113-7D
  - Saccharomyces cerevisiae CLIB215
  - Saccharomyces cerevisiae CLIB324
  - Saccharomyces cerevisiae CLIB382
  - Saccharomyces cerevisiae EC1118
  - Saccharomyces cerevisiae EC9-8
  - Saccharomyces cerevisiae FL100
  - Saccharomyces cerevisiae Fleischmanns baking yeast
  - Saccharomyces cerevisiae FostersB
  - Saccharomyces cerevisiae FostersO

- **ORGANISM from taxonomy → Saccharomyces cerevisiae**



```
                     /codon_start=1
                     /product="Rev7p"
                     /protein_id="AAA98667.1"
                     /translation="MNRWVEKWLRVYLKCYINLILFYRNVYPPQSFDYTTYQSFNLPQ
                     FVPINRHPALIDYIEELILDVLSKLTHVYRFSICIINKKNDLCIEKYVLDFSELQHVD
                     KDDQIITETEVFDEFRSSLNSLIMHLEKLPKVNDDTITFEAVINAIELELGHKLDRNR
                     RVDSLEEKAEIERDSNWVKCQEDENLPDNNGFQPPKIKLTSLVGSDVGPLIIHQFSEK
                     LISGDDKILNGVYSQYEEGESIFGSLF"
ORIGIN
        1 gatcctccat atacaacggt atctccacct caggtttaga tctcaacaac ggaaccattg
       61 ccgacatgag acagttaggt atcgtcgaga gttacaagct aaaacgagca gtagtcagct
      121 ctgcatctga agccgctgaa gttctactaa gggtggataa catcatccgt gcaagaccaa
      181 gaaccgccaa tagacaacat atgtaacata tttaggatat acctcgaaaa taataaaccg
      241 ccacactgtc attattataa ttagaaacag aacgcaaaaa ttatccacta tataattcaa
      301 agacgcgaaa aaaaaagaac aacgcgtcat agaacttttg gcaattcgcg tcacaaataa
      361 attttggcaa cttatgtttc ctcttcgagc agtactcgag ccctgtctca agaatgtaat
      421 aatacccatc gtaggtatgg ttaaagatag catctccaca acctcaaagc tccttgccga
      481 gagtcgccct cctttgtcga gtaattttca cttttcatat gagaacttat tttcttattc
      541 tttactctca catcctgtag tgattgacac tgcaacagcc accatcacta gaagaacaga
      601 acaattactt aatagaaaaa ttatatcttc ctcgaaacga tttcctgctt ccaacatcta
      661 cgtatatcaa gaagcattca cttaccatga cacagcttca gatttcatta ttgctgacag
      721 ctactatatc actactccat ctagtagtgg ccacgcccta tgaggcatat cctatcggaa
      781 aacaataccc cccagtggca agagtcaatg aatcgtttac atttcaaatt tccaatgata
      841 cctataaatc gtctgtagac aagacagctc aaataacata caattgcttc gacttaccga
      901 gctggctttc gtttgactct agttctagaa cgttctcagg tgaaccttct tctgacttac
      961 tatctgatgc gaacaccacg ttgtatttca atgtaatact cgagggtacg gactctgccg
     1021 acagcacgtc tttgaacaat acataccaat ttgttgttac aaaccgtcca tccatctcgc
     1081 tatcgtcaga tttcaatcta ttggcgttgt taaaaaacta tggttatact aacggcaaaa
     1141 acgctctgaa actagatcct aatgaagtct tcaacgtgac ttttgaccgt tcaatgttca
     1201 ctaacgaaga atccattgtg tcgtattacg gacgttctca gttgtataat gcgccgttac
     1261 ccaattggct gttcttcgat tctggcgagt tgaagtttac tgggacggca ccggtgataa
     1321 actcggcgat tgctccagaa acaagctaca gttttgtcat catcgctaca gacattgaag
     1381 gattttctgc cgttgaggta gaattcgaat tagtcatcgg ggctcaccag ttaactacct
     1441 ctattcaaaa tagtttgata atcaacgtta ctgacacagg taacgtttca tatgacttac
     1501 ctctaaacta tgtttatctc gatgacgatc ctatttcttc tgataaattg ggttctataa
     1561 acttattgga tgctccagac tgggtggcat tagataatgc taccatttcc gggtctgtcc
     1621 cagatgaatt actcggtaag aactccaatc ctgccaattt ttctgtgtcc atttatgata
     1681 cttatggtga tgtgatttat ttcaacttcg aagttgtctc cacaacggat ttgtttgcca
     1741 ttagttctct tcccaatatt aacgctacaa ggggtgaatg gttctcctac tattttttgc
     1801 cttctcagtt tacagactac gtgaatacaa acgtttcatt agagtttact aattcaagcc
     1861 aagaccatga ctgggtgaaa ttccaatcat ctaatttaac attagctgga gaagtgccca
     1921 agaatttcga caagctttca ttaggtttga aagcgaacca aggttcacaa tctcaagagc
     1981 tatattttaa catcattggc atggattcaa agataactca ctcaaaccac agtgcgaatg
     2041 caacgtccac aagaagttct caccactcca cctcaacaag ttcttacaca tcttctactt
     2101 acactgcaaa aatttcttct acctccgctg ctgctacttc ttctgctcca gcagcgctgc
     2161 cagcagccaa taaaacttca tctcacaata aaaaagcagt agcaattgcg tgcggtgttg
     2221 ctatcccatt aggcgttatc ctagtagctc tcatttgctt cctaatattc tggagacgca
     2281 gaagggaaaa tccagacgat gaaaacttac cgcatgctat tagtggacct gatttgaata
     2341 atcctgcaaa taaaccaaat caagaaaacg ctacaccttt gaacaacccc tttgatgatg
     2401 atgcttcctc gtacgatgat acttcaatag caagaagatt ggctgctttg aacactttga
     2461 aattggataa ccactctgcc actgaatctg atatttccag cgtggatgaa aagagagatt
     2521 ctctatcagg tatgaataca tacaatgatc agttccaatc ccaaagtaaa gaagaattat
     2581 tagcaaaacc cccagtacag cctccagaga gcccgttctt tgacccacag aataggtctt
     2641 cttctgtgta tatggatagt gaaccagcag taaataaatc ctggcgatat actggcaacc
     2701 tgtcaccagt ctctgatatt gtcagagaca gttacggatc acaaaaaact gttgatacag
     2761 aaaaactttt cgatttagaa gcaccagaga aggaaaaacg tacgtcaagg gatgtcacta
```

- **data have specified, well known format**

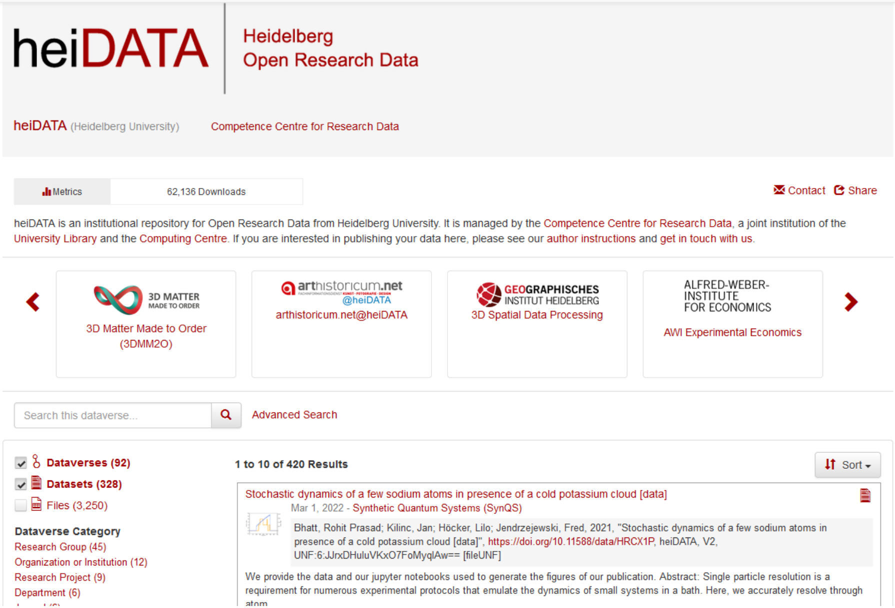
```
 421 aatacccatc gtaggtatgg ttaaagatag catctccaca acctcaaagc tccttgccga
 481 gagtcgccct cctttgtcga gtaattttca cttttcatat gagaacttat tttcttattc
 541 tttactctca catcctgtag tgattgacac tgcaacagcc accatcacta gaagaacaga
 601 acaattactt aatagaaaaa ttatatcttc ctcgaaacga tttcctgctt ccaacatcta
 661 cgtatatcaa gaagcattca cttaccatga cacagcttca gatttcatta ttgctgacag
 721 ctactatatc actactccat ctagtagtgg ccacgcccta tgaggcatat cctatcggaa
 781 aacaataccc cccagtggca agagtcaatg aatcgtttac atttcaaatt tccaatgata
 841 cctataaatc gtctgtagac aagacagctc aaataacata caattgcttc gacttaccga
 901 gctggctttc gtttgactct agttctagaa cgttctcagg tgaaccttct tctgacttac
 961 tatctgatgc gaacaccacg ttgtatttca atgtaatact cgagggtacg gactctgccg
1021 acagcacgtc tttgaacaat acataccaat ttgttgttac aaaccgtcca tccatctcgc
1081 tatcgtcaga tttcaatcta ttggcgttgt taaaaaacta tggttatact aacggcaaaa
1141 acgctctgaa actagatcct aatgaagtct tcaacgtgac ttttgaccgt tcaatgttca
1201 ctaacgaaga atccattgtg tcgtattacg gacgttctca gttgtataat gcgccgttac
1261 ccaattggct gttcttcgat tctggcgagt tgaagtttac tgggacggca ccggtgataa
1321 actcggcgat tgctccagaa acaagctaca gttttgtcat catcgctaca gacattgaag
1381 gattttctgc cgttgaggta gaattcgaat tagtcatcgg ggctcaccag ttaactacct
1441 ctattcaaaa tagtttgata atcaacgtta ctgacacagg taacgtttca tatgacttac
1501 ctctaaacta tgtttatctc gatgacgatc ctatttcttc tgataaattg ggttctataa
1561 acttattgga tgctccagac tgggtggcat tagataatgc taccatttcc gggtctgtcc
1621 cagatgaatt actcggtaag aactccaatc ctgccaattt ttctgtgtcc atttatgata
1681 cttatggtga tgtgatttat ttcaacttcg aagttgtctc cacaacggat ttgtttgcca
1741 ttagttctct tcccaatatt aacgctacaa ggggtgaatg gttctcctac tatttttgc
1801 cttctcagtt tacagactac gtgaatacaa acgtttcatt agagtttact aattcaagcc
1861 aagaccatga ctgggtgaaa ttccaatcat ctaatttaac attagctgga gaagtgccca
1921 agaatttcga caagctttca ttaggtttga aagcgaacca aggttcacaa tctcaagagc
1981 tatattttaa catcattggc atggattcaa agataactca ctcaaaccac agtgcgaatg
2041 caacgtccac aagaagttct caccactcca cctcaacaag ttcttacaca tcttctactt
2101 acactgcaaa aatttcttct acctccgctg ctgctacttc ttctgctcca gcagcgctgc
2161 cagcagccaa taaaacttca tctcacaata aaaaagcagt agcaattgcg tgcggtgttg
2221 ctatcccatt aggcgttatc ctagtagctc tcatttgctt cctaatattc tggagacgca
2281 gaagggaaaa tccagacgat gaaaacttac cgcatgctat tagtggacct gatttgaata
2341 atcctgcaaa taaaccaaat caagaaaacg ctacaccttt gaacaacccc tttgatgatg
2401 atgcttcctc gtacgatgat acttcaatag caagaagatt ggctgctttg aacactttga
2461 aattggataa ccactctgcc actgaatctg atatttccag cgtggatgaa aagagagatt
2521 ctctatcagg tatgaataca tacaatgatc agttccaatc ccaaagtaaa gaagaattat
2581 tagcaaaacc cccagtacag cctccagaga gcccgttctt tgacccacag aataggtctt
2641 cttctgtgta tatggatagt gaaccagcag taaataaatc ctggcgatat actggcaacc
2701 tgtcaccagt ctctgatatt gtcagagaca gttacggatc acaaaaaact gttgatacag
2761 aaaaactttt cgatttagaa gcaccagaga aggaaaaacg tacgtcaagg gatgtcacta
2821 tgtcttcact ggacccttgg aacagcaata ttagcccttc tcccgtaaga aaatcagtaa
2881 caccatcacc atataacgta acgaagcatc gtaaccgcca cttacaaaat attcaagact
2941 ctcaaagcgg taaaaacgga atcactccca caacaatgtc aacttcatct tctgacgatt
3001 ttgttccggt taaagatggt gaaaattttt gctgggtcca tagcatggaa ccagacagaa
3061 gaccaagtaa gaaaaggtta gtagattttt caaataagag taatgtcaat gttggtcaag
3121 ttaaggacat tcacggacgc atcccagaaa tgctgtgatt atacgcaacg atattttgct
3181 taattttatt ttcctgtttt atttttatt agtggtttac agataccta tattttattt
3241 agtttttata cttagagaca tttaatttta attccattct tcaaatttca tttttgcact
3301 taaaacaaag atccaaaaat gctctcgccc tcttcatatt gagaatacac tccattcaaa
3361 attttgtcgt caccgctgat taattttca ctaaactgat gaataatcaa aggccccacg
```

- data have specified, well known format

  →

- repository can visualize specific segments of data

Institutional repositories



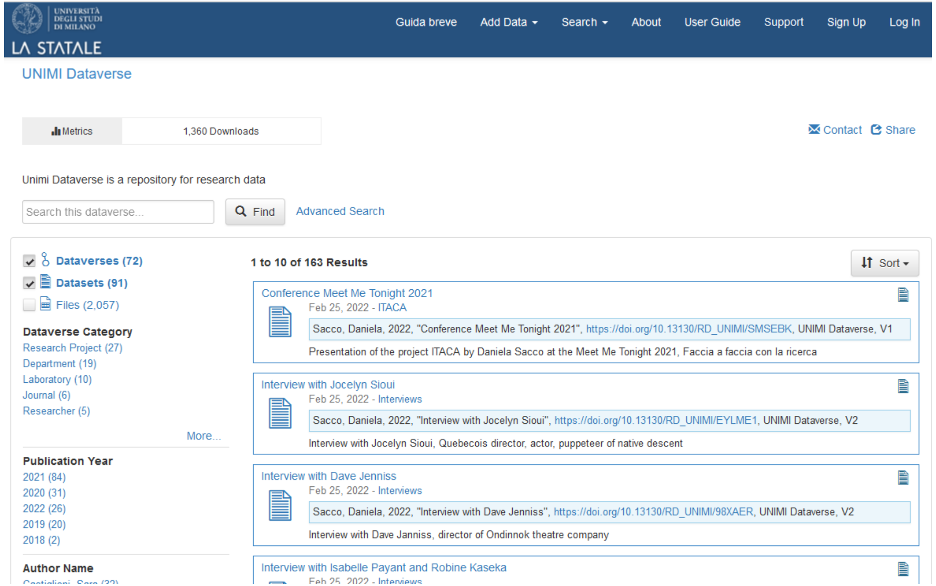
heiDATA is an institutional repository for Open Research Data from Heidelberg University. It is managed by the Competence Centre for Research Data, a joint institution of the University Library and the Computing Centre. If you are interested in publishing your data here, please see our author instructions and get in touch with us.

☑ Dataverses (92)
☑ Datasets (328)
☐ Files (3,250)

Dataverse Category
Research Group (45)
Organization or Institution (12)
Research Project (9)
Department (6)

1 to 10 of 420 Results

Stochastic dynamics of a few sodium atoms in presence of a cold potassium cloud [data]
Mar 1, 2022 - Synthetic Quantum Systems (SynQS)
Bhatt, Rohit Prasad; Kilinc, Jan; Höcker, Lilo; Jendrzejewski, Fred, 2021, "Stochastic dynamics of a few sodium atoms in presence of a cold potassium cloud [data]", https://doi.org/10.11588/data/HRCX1P, heiDATA, V2, UNF:6:JJrxDHuluVKxO7FoMyqlAw== [fileUNF]
We provide the data and our jupyter notebooks used to generate the figures of our publication. Abstract: Single particle resolution is a requirement for numerous experimental protocols that emulate the dynamics of small systems in a bath. Here, we accurately resolve through atom



Unimi Dataverse is a repository for research data

☑ Dataverses (72)
☑ Datasets (91)
☐ Files (2,057)

Dataverse Category
Research Project (27)
Department (19)
Laboratory (10)
Journal (6)
Researcher (5)

Publication Year
2021 (84)
2020 (31)
2022 (26)
2019 (20)
2018 (2)

Author Name

1 to 10 of 163 Results

Conference Meet Me Tonight 2021
Feb 25, 2022 - ITACA
Sacco, Daniela, 2022, "Conference Meet Me Tonight 2021", https://doi.org/10.13130/RD_UNIMI/SMSEBK, UNIMI Dataverse, V1
Presentation of the project ITACA by Daniela Sacco at the Meet Me Tonight 2021, Faccia a faccia con la ricerca

Interview with Jocelyn Sioui
Feb 25, 2022 - Interviews
Sacco, Daniela, 2022, "Interview with Jocelyn Sioui", https://doi.org/10.13130/RD_UNIMI/EYLME1, UNIMI Dataverse, V2
Interview with Jocelyn Sioui, Quebecois director, actor, puppeteer of native descent

Interview with Dave Jenniss
Feb 25, 2022 - Interviews
Sacco, Daniela, 2022, "Interview with Dave Jenniss", https://doi.org/10.13130/RD_UNIMI/98XAER, UNIMI Dataverse, V2
Interview with Dave Janniss, director of Ondinnok theatre company

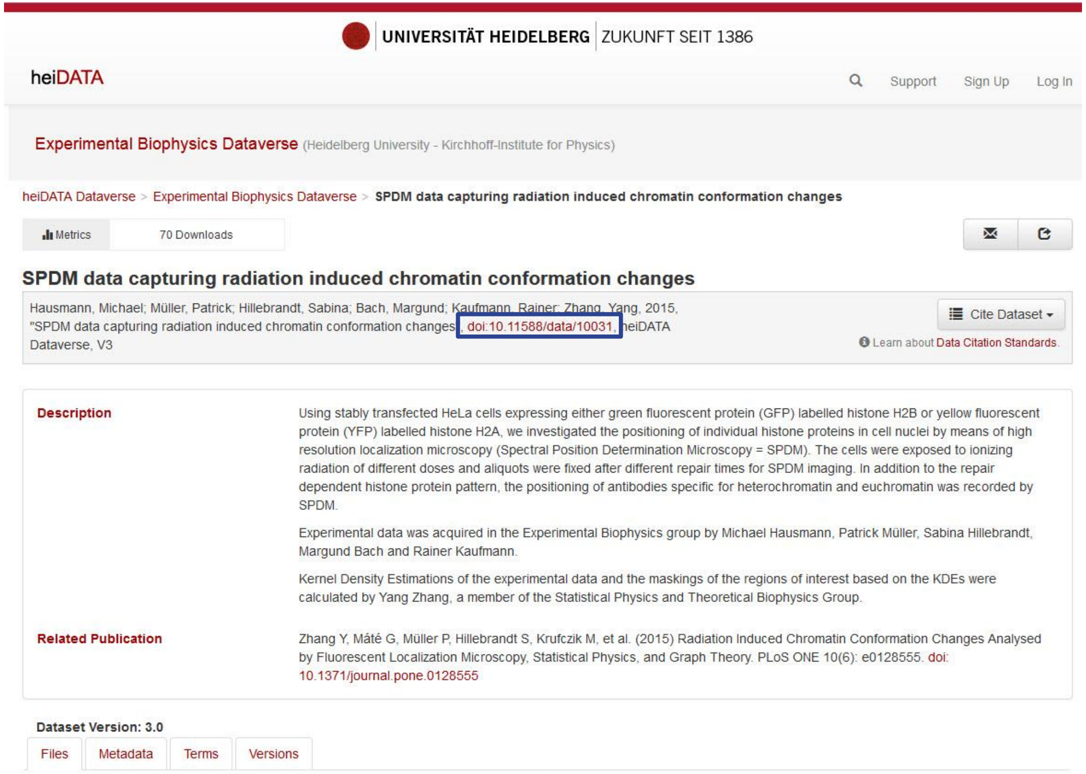
Interview with Isabelle Payant and Robine Kaseka

https://heidata.uni-heidelberg.de

https://dataverse.unimi.it/

Findability

- DOI's



Experimental Biophysics Dataverse (Heidelberg University - Kirchhoff-Institute for Physics)

heiDATA Dataverse > Experimental Biophysics Dataverse > SPDM data capturing radiation induced chromatin conformation changes

SPDM data capturing radiation induced chromatin conformation changes

Hausmann, Michael; Müller, Patrick; Hillebrandt, Sabina; Bach, Margund; Kaufmann, Rainer; Zhang, Yang, 2015, "SPDM data capturing radiation induced chromatin conformation changes", doi:10.11588/data/10031, heiDATA Dataverse, V3

Learn about Data Citation Standards.

Description

Using stably transfected HeLa cells expressing either green fluorescent protein (GFP) labelled histone H2B or yellow fluorescent protein (YFP) labelled histone H2A, we investigated the positioning of individual histone proteins in cell nuclei by means of high resolution localization microscopy (Spectral Position Determination Microscopy = SPDM). The cells were exposed to ionizing radiation of different doses and aliquots were fixed after different repair times for SPDM imaging. In addition to the repair dependent histone protein pattern, the positioning of antibodies specific for heterochromatin and euchromatin was recorded by SPDM.

Experimental data was acquired in the Experimental Biophysics group by Michael Hausmann, Patrick Müller, Sabina Hillebrandt, Margund Bach and Rainer Kaufmann.

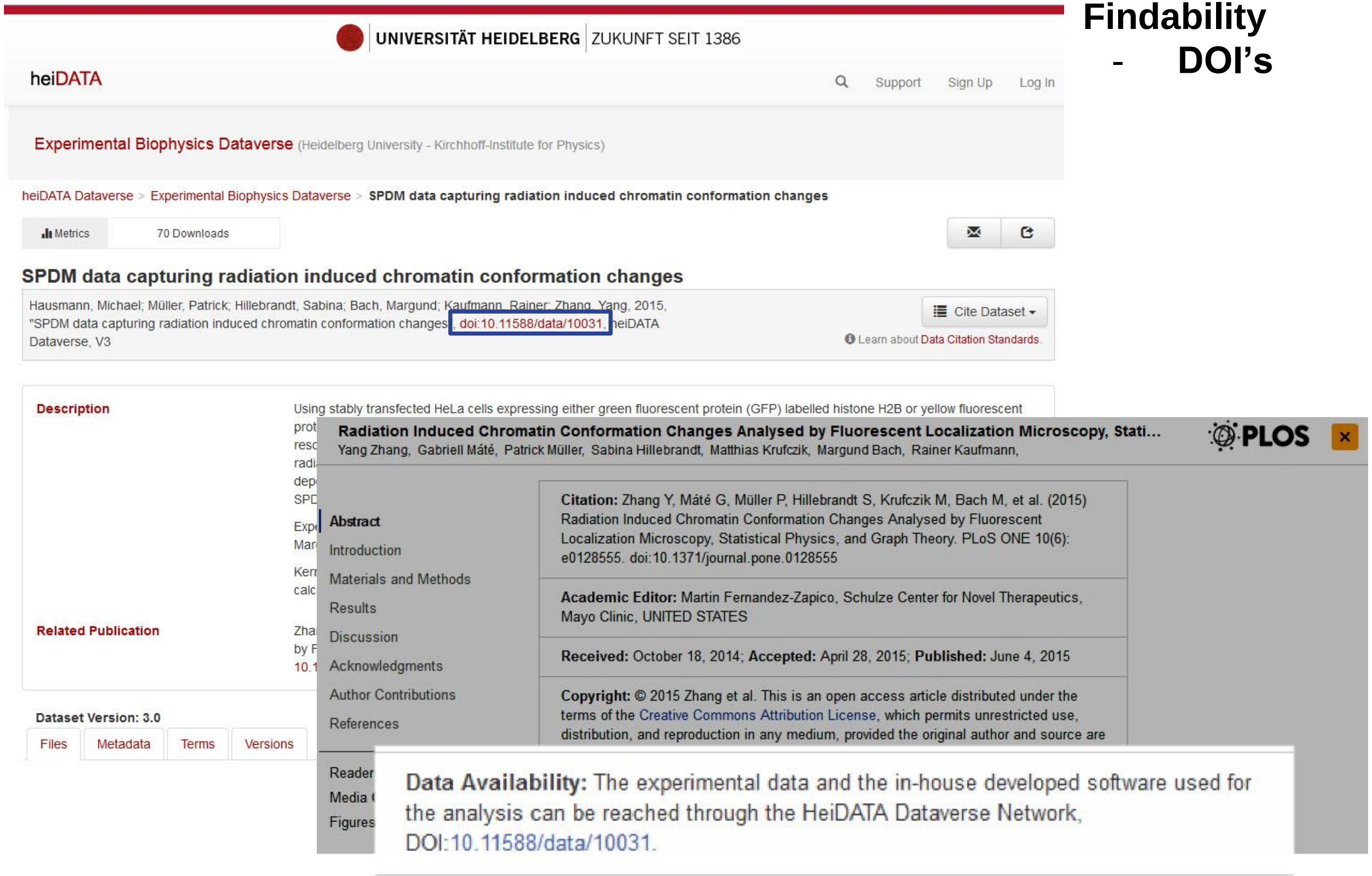
Kernel Density Estimations of the experimental data and the maskings of the regions of interest based on the KDEs were calculated by Yang Zhang, a member of the Statistical Physics and Theoretical Biophysics Group.

Related Publication

Zhang Y, Máté G, Müller P, Hillebrandt S, Krufczik M, et al. (2015) Radiation Induced Chromatin Conformation Changes Analysed by Fluorescent Localization Microscopy, Statistical Physics, and Graph Theory. PLoS ONE 10(6): e0128555. doi: 10.1371/journal.pone.0128555

Dataset Version: 3.0

Findability
- DOI's



Experimental Biophysics Dataverse (Heidelberg University - Kirchhoff-Institute for Physics)

heiDATA Dataverse > Experimental Biophysics Dataverse > SPDM data capturing radiation induced chromatin conformation changes

SPDM data capturing radiation induced chromatin conformation changes

Hausmann, Michael; Müller, Patrick; Hillebrandt, Sabina; Bach, Margund; Kaufmann, Rainer; Zhang, Yang, 2015, "SPDM data capturing radiation induced chromatin conformation changes", doi:10.11588/data/10031, heiDATA Dataverse, V3

Learn about Data Citation Standards.

Description Using stably transfected HeLa cells expressing either green fluorescent protein (GFP) labelled histone H2B or yellow fluorescent

Related Publication

Dataset Version: 3.0

Radiation Induced Chromatin Conformation Changes Analysed by Fluorescent Localization Microscopy, Stati...

Yang Zhang, Gabriell Máté, Patrick Müller, Sabina Hillebrandt, Matthias Krufczik, Margund Bach, Rainer Kaufmann,

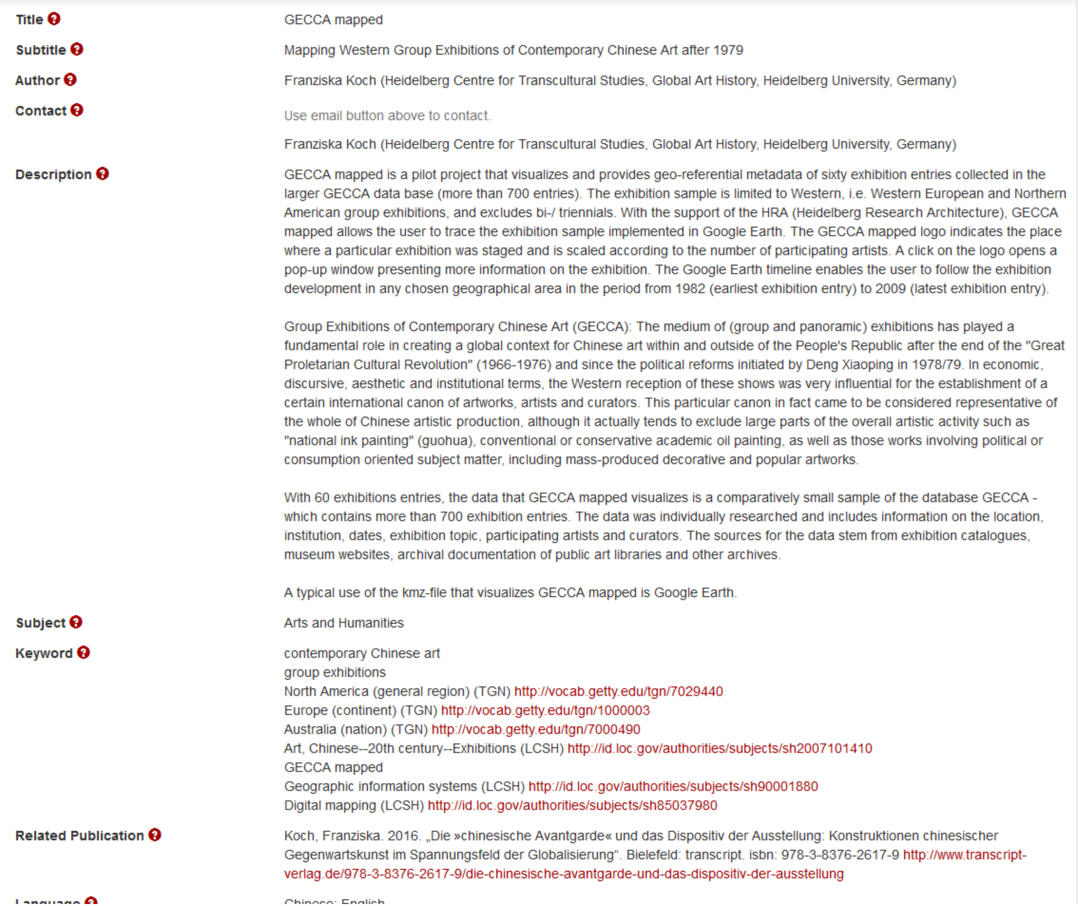
Citation: Zhang Y, Máté G, Müller P, Hillebrandt S, Krufczik M, Bach M, et al. (2015) Radiation Induced Chromatin Conformation Changes Analysed by Fluorescent Localization Microscopy, Statistical Physics, and Graph Theory. PLoS ONE 10(6): e0128555. doi:10.1371/journal.pone.0128555

Academic Editor: Martin Fernandez-Zapico, Schulze Center for Novel Therapeutics, Mayo Clinic, UNITED STATES

Received: October 18, 2014; **Accepted:** April 28, 2015; **Published:** June 4, 2015

Copyright: © 2015 Zhang et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are

Data Availability: The experimental data and the in-house developed software used for the analysis can be reached through the HeiDATA Dataverse Network, DOI:10.11588/data/10031.



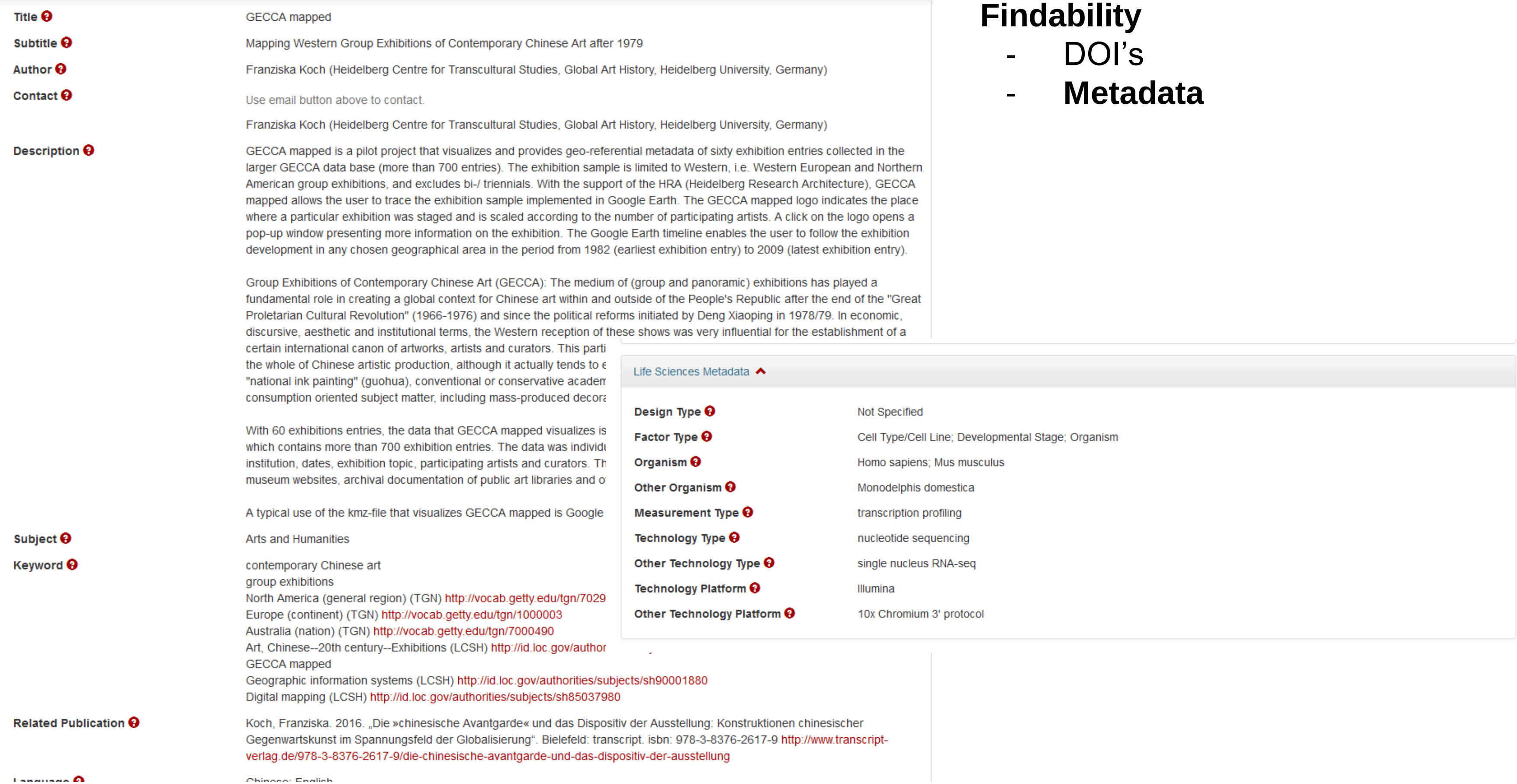
Title	GECCA mapped
Subtitle	Mapping Western Group Exhibitions of Contemporary Chinese Art after 1979
Author	Franziska Koch (Heidelberg Centre for Transcultural Studies, Global Art History, Heidelberg University, Germany)
Contact	Use email button above to contact. Franziska Koch (Heidelberg Centre for Transcultural Studies, Global Art History, Heidelberg University, Germany)
Description	GECCA mapped is a pilot project that visualizes and provides geo-referential metadata of sixty exhibition entries collected in the larger GECCA data base (more than 700 entries). The exhibition sample is limited to Western, i.e. Western European and Northern American group exhibitions, and excludes bi-/ triennials. With the support of the HRA (Heidelberg Research Architecture), GECCA mapped allows the user to trace the exhibition sample implemented in Google Earth. The GECCA mapped logo indicates the place where a particular exhibition was staged and is scaled according to the number of participating artists. A click on the logo opens a pop-up window presenting more information on the exhibition. The Google Earth timeline enables the user to follow the exhibition development in any chosen geographical area in the period from 1982 (earliest exhibition entry) to 2009 (latest exhibition entry). Group Exhibitions of Contemporary Chinese Art (GECCA): The medium of (group and panoramic) exhibitions has played a fundamental role in creating a global context for Chinese art within and outside of the People's Republic after the end of the "Great Proletarian Cultural Revolution" (1966-1976) and since the political reforms initiated by Deng Xiaoping in 1978/79. In economic, discursive, aesthetic and institutional terms, the Western reception of these shows was very influential for the establishment of a certain international canon of artworks, artists and curators. This particular canon in fact came to be considered representative of the whole of Chinese artistic production, although it actually tends to exclude large parts of the overall artistic activity such as "national ink painting" (guohua), conventional or conservative academic oil painting, as well as those works involving political or consumption oriented subject matter, including mass-produced decorative and popular artworks. With 60 exhibitions entries, the data that GECCA mapped visualizes is a comparatively small sample of the database GECCA - which contains more than 700 exhibition entries. The data was individually researched and includes information on the location, institution, dates, exhibition topic, participating artists and curators. The sources for the data stem from exhibition catalogues, museum websites, archival documentation of public art libraries and other archives. A typical use of the kmz-file that visualizes GECCA mapped is Google Earth.
Subject	Arts and Humanities
Keyword	contemporary Chinese art group exhibitions North America (general region) (TGN) http://vocab.getty.edu/tgn/7029440 Europe (continent) (TGN) http://vocab.getty.edu/tgn/1000003 Australia (nation) (TGN) http://vocab.getty.edu/tgn/7000490 Art, Chinese--20th century--Exhibitions (LCSH) http://id.loc.gov/authorities/subjects/sh2007101410 GECCA mapped Geographic information systems (LCSH) http://id.loc.gov/authorities/subjects/sh90001880 Digital mapping (LCSH) http://id.loc.gov/authorities/subjects/sh85037980
Related Publication	Koch, Franziska. 2016. „Die »chinesische Avantgarde« und das Dispositiv der Ausstellung: Konstruktionen chinesischer Gegenwartskunst im Spannungsfeld der Globalisierung". Bielefeld: transcript. isbn: 978-3-8376-2617-9 http://www.transcript-verlag.de/978-3-8376-2617-9/die-chinesische-avantgarde-und-das-dispositiv-der-ausstellung
Language	Chinese; English

Findability

- DOI's
- **Metadata**

Findability

- DOI’s
- **Metadata**



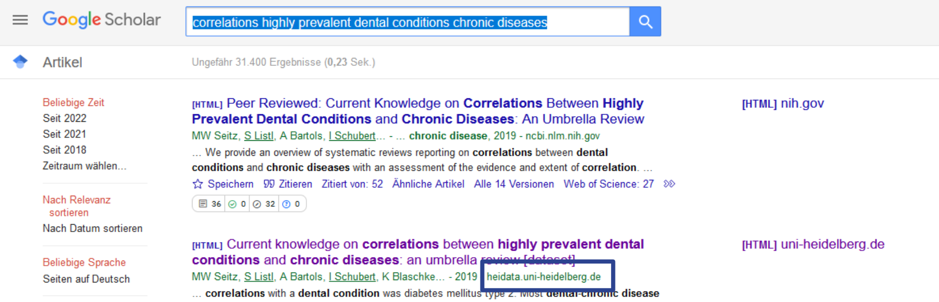
Title ❓	GECCA mapped
Subtitle ❓	Mapping Western Group Exhibitions of Contemporary Chinese Art after 1979
Author ❓	Franziska Koch (Heidelberg Centre for Transcultural Studies, Global Art History, Heidelberg University, Germany)
Contact ❓	Use email button above to contact. Franziska Koch (Heidelberg Centre for Transcultural Studies, Global Art History, Heidelberg University, Germany)
Description ❓	GECCA mapped is a pilot project that visualizes and provides geo-referential metadata of sixty exhibition entries collected in the larger GECCA data base (more than 700 entries). The exhibition sample is limited to Western, i.e. Western European and Northern American group exhibitions, and excludes bi-/ triennials. With the support of the HRA (Heidelberg Research Architecture), GECCA mapped allows the user to trace the exhibition sample implemented in Google Earth. The GECCA mapped logo indicates the place where a particular exhibition was staged and is scaled according to the number of participating artists. A click on the logo opens a pop-up window presenting more information on the exhibition. The Google Earth timeline enables the user to follow the exhibition development in any chosen geographical area in the period from 1982 (earliest exhibition entry) to 2009 (latest exhibition entry). Group Exhibitions of Contemporary Chinese Art (GECCA): The medium of (group and panoramic) exhibitions has played a fundamental role in creating a global context for Chinese art within and outside of the People's Republic after the end of the "Great Proletarian Cultural Revolution" (1966-1976) and since the political reforms initiated by Deng Xiaoping in 1978/79. In economic, discursive, aesthetic and institutional terms, the Western reception of these shows was very influential for the establishment of a certain international canon of artworks, artists and curators. This parti the whole of Chinese artistic production, although it actually tends to e "national ink painting" (guohua), conventional or conservative acaden consumption oriented subject matter, including mass-produced decora With 60 exhibitions entries, the data that GECCA mapped visualizes is which contains more than 700 exhibition entries. The data was individu institution, dates, exhibition topic, participating artists and curators. Th museum websites, archival documentation of public art libraries and o A typical use of the kmz-file that visualizes GECCA mapped is Google
Subject ❓	Arts and Humanities
Keyword ❓	contemporary Chinese art group exhibitions North America (general region) (TGN) http://vocab.getty.edu/tgn/7029 Europe (continent) (TGN) http://vocab.getty.edu/tgn/1000003 Australia (nation) (TGN) http://vocab.getty.edu/tgn/7000490 Art, Chinese--20th century--Exhibitions (LCSH) http://id.loc.gov/author GECCA mapped Geographic information systems (LCSH) http://id.loc.gov/authorities/subjects/sh90001880 Digital mapping (LCSH) http://id.loc.gov/authorities/subjects/sh85037980
Related Publication ❓	Koch, Franziska. 2016. „Die »chinesische Avantgarde« und das Dispositiv der Ausstellung: Konstruktionen chinesischer Gegenwartskunst im Spannungsfeld der Globalisierung“. Bielefeld: transcript. isbn: 978-3-8376-2617-9 http://www.transcript-verlag.de/978-3-8376-2617-9/die-chinesische-avantgarde-und-das-dispositiv-der-ausstellung
Language ❓	Chinese; English

Life Sciences Metadata ^

Design Type ❓	Not Specified
Factor Type ❓	Cell Type/Cell Line; Developmental Stage; Organism
Organism ❓	Homo sapiens; Mus musculus
Other Organism ❓	Monodelphis domestica
Measurement Type ❓	transcription profiling
Technology Type ❓	nucleotide sequencing
Other Technology Type ❓	single nucleus RNA-seq
Technology Platform ❓	Illumina
Other Technology Platform ❓	10x Chromium 3' protocol

Findability

- DOI’s
- Metadata
- **Indexing in catalogs and databases (enabling automatic harvesting of metadata)**



correlations highly prevalent dental conditions chronic diseases

Ungefähr 31.400 Ergebnisse (0,23 Sek.)

[HTML] Peer Reviewed: Current Knowledge on **Correlations** Between **Highly Prevalent Dental Conditions** and **Chronic Diseases**: An Umbrella Review

MW Seitz, S Listl, A Bartols, I Schubert... - ... **chronic disease**, 2019 - ncbi.nlm.nih.gov

... We provide an overview of systematic reviews reporting on **correlations** between **dental conditions** and **chronic diseases** with an assessment of the evidence and extent of **correlation**. ...

Zitiert von: 52 Ähnliche Artikel Alle 14 Versionen Web of Science: 27

36 0 32 0

[HTML] nih.gov

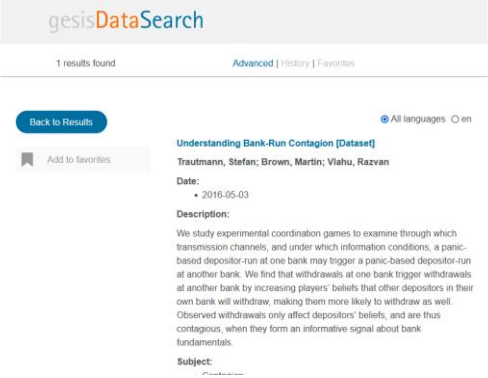
[HTML] Current knowledge on **correlations** between **highly prevalent dental conditions** and **chronic diseases**: an umbrella review [dataset]

MW Seitz, S Listl, A Bartols, I Schubert, K Blaschke... - 2019 - heidata.uni-heidelberg.de

... **correlations** with a **dental condition** was diabetes mellitus type 2. Most **dental-chronic disease correlations** were ... type 2 and periodontitis and cardiovascular **disease**. Freque

Zitiert von: 1 Ähnliche Artikel Alle 2 Versionen

[HTML] uni-heidelberg.de



1 results found

● All languages ○ en

Understanding Bank-Run Contagion [Dataset]

Trautmann, Stefan; Brown, Martin; Vlahu, Razvan

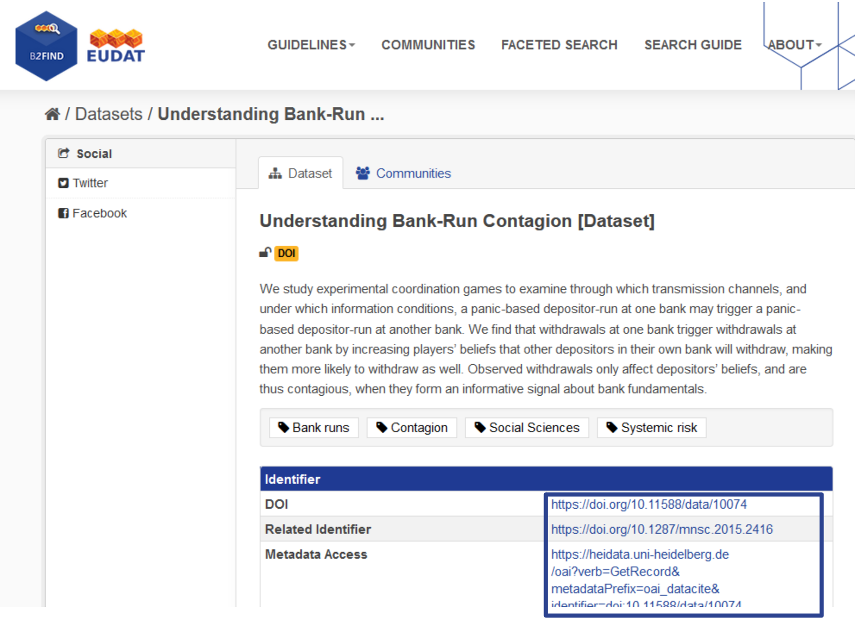
Date:

- 2016-05-03

Description:

We study experimental coordination games to examine through which transmission channels, and under which information conditions, a panic-based depositor-run at one bank may trigger a panic-based depositor-run at another bank. We find that withdrawals at one bank trigger withdrawals at another bank by increasing players’ beliefs that other depositors in their own bank will withdraw, making them more likely to withdraw as well. Observed withdrawals only affect depositors’ beliefs, and are thus contagious, when they form an informative signal about bank fundamentals.

Subject:



/ Datasets / **Understanding Bank-Run ...**

Understanding Bank-Run Contagion [Dataset]

DOI

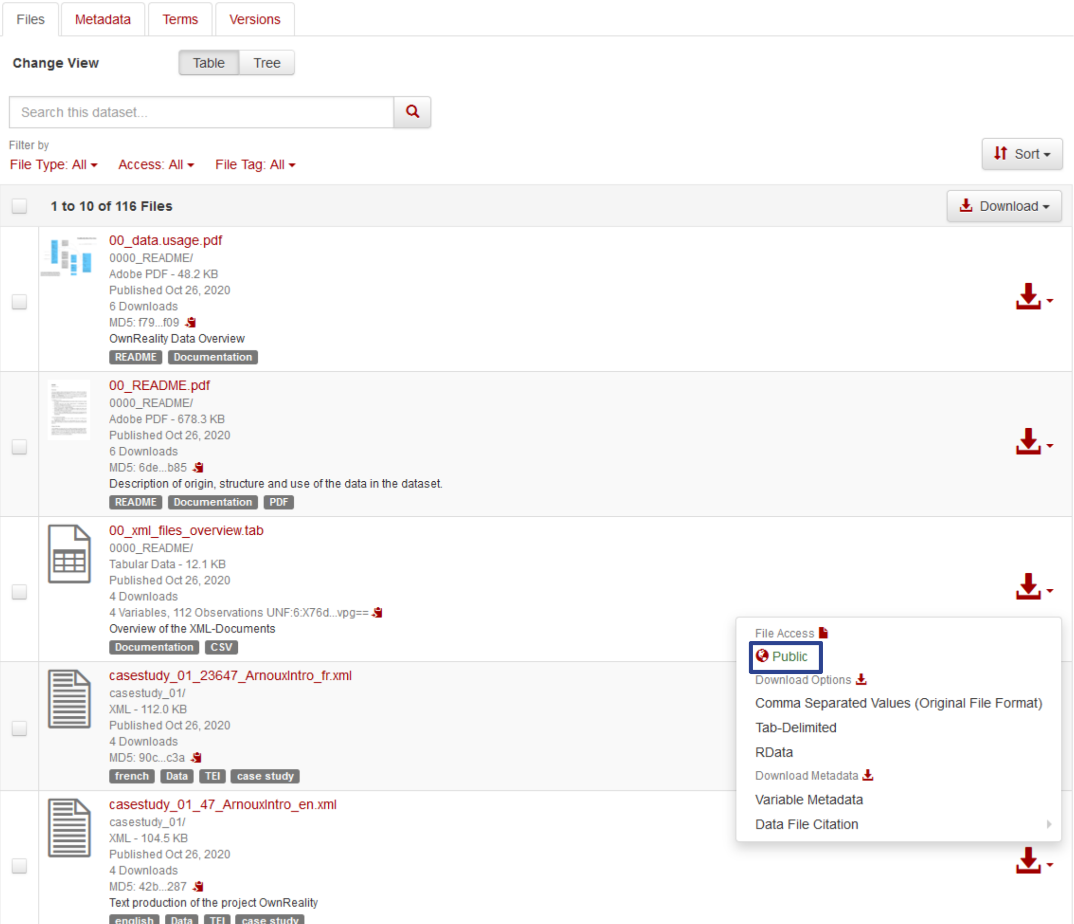
We study experimental coordination games to examine through which transmission channels, and under which information conditions, a panic-based depositor-run at one bank may trigger a panic-based depositor-run at another bank. We find that withdrawals at one bank trigger withdrawals at another bank by increasing players’ beliefs that other depositors in their own bank will withdraw, making them more likely to withdraw as well. Observed withdrawals only affect depositors’ beliefs, and are thus contagious, when they form an informative signal about bank fundamentals.

Bank runs | Contagion | Social Sciences | Systemic risk

Identifier	
DOI	https://doi.org/10.11588/data/10074
Related Identifier	https://doi.org/10.1287/mnsc.2015.2416
Metadata Access	https://heidata.uni-heidelberg.de/oai?verb=GetRecord&metadataPrefix=oai_datacite&

Accessibility

- Download of public files via browser or via API



1 to 10 of 116 Files

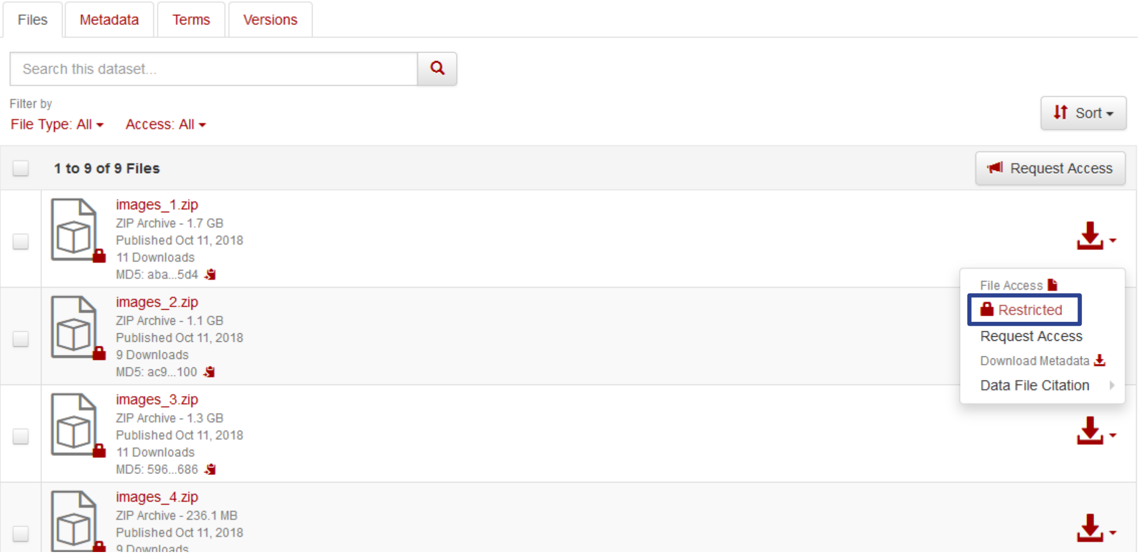
00_data.usage.pdf
0000_README/
Adobe PDF - 48.2 KB
Published Oct 26, 2020
6 Downloads
MD5: f79...f09
OwnReality Data Overview
README | Documentation

00_README.pdf
0000_README/
Adobe PDF - 678.3 KB
Published Oct 26, 2020
6 Downloads
MD5: 6de...b85
Description of origin, structure and use of the data in the dataset.
README | Documentation | PDF

00_xml_files_overview.tab
0000_README/
Tabular Data - 12.1 KB
Published Oct 26, 2020
4 Downloads
4 Variables, 112 Observations UNF:6:X76d...vpg==
Overview of the XML-Documents
Documentation | CSV

casestudy_01_23647_ArnouxIntro_fr.xml
casestudy_01/
XML - 112.0 KB
Published Oct 26, 2020
4 Downloads
MD5: 90c...c3a
french | Data | TEI | case study

casestudy_01_47_ArnouxIntro_en.xml
casestudy_01/
XML - 104.5 KB
Published Oct 26, 2020
4 Downloads
MD5: 42b...287
Text production of the project OwnReality
english | Data | TEI | case study



1 to 9 of 9 Files

images_1.zip
ZIP Archive - 1.7 GB
Published Oct 11, 2018
11 Downloads
MD5: aba...5d4

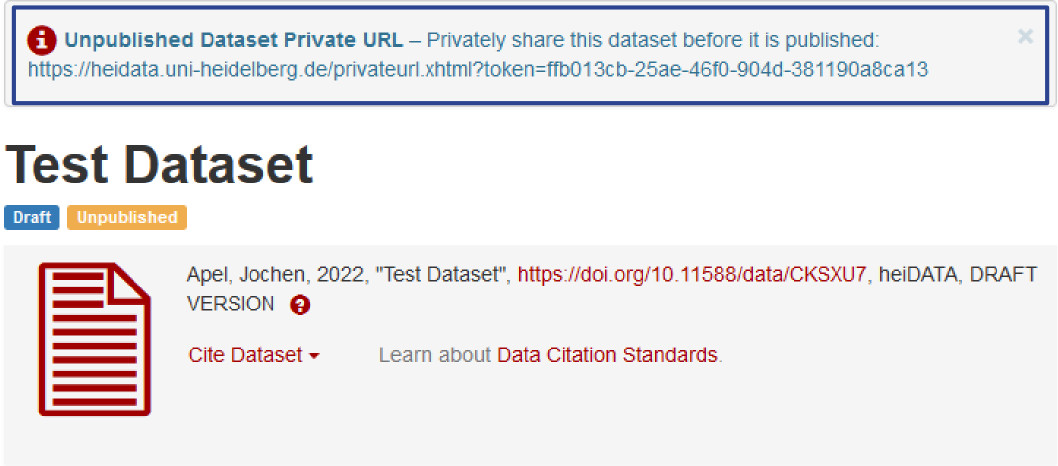
images_2.zip
ZIP Archive - 1.1 GB
Published Oct 11, 2018
9 Downloads
MD5: ac9...100

images_3.zip
ZIP Archive - 1.3 GB
Published Oct 11, 2018
11 Downloads
MD5: 596...686

images_4.zip
ZIP Archive - 236.1 MB
Published Oct 11, 2018
9 Downloads

Accessibility

- Download of public files via browser or via API
- **“As open as possible, but as closed as necessary”**



Unpublished Dataset Private URL – Privately share this dataset before it is published: https://heidata.uni-heidelberg.de/privateurl.xhtml?token=ffb013cb-25ae-46f0-904d-381190a8ca13

Test Dataset

Apel, Jochen, 2022, "Test Dataset", https://doi.org/10.11588/data/CKSXU7, heiDATA, DRAFT VERSION

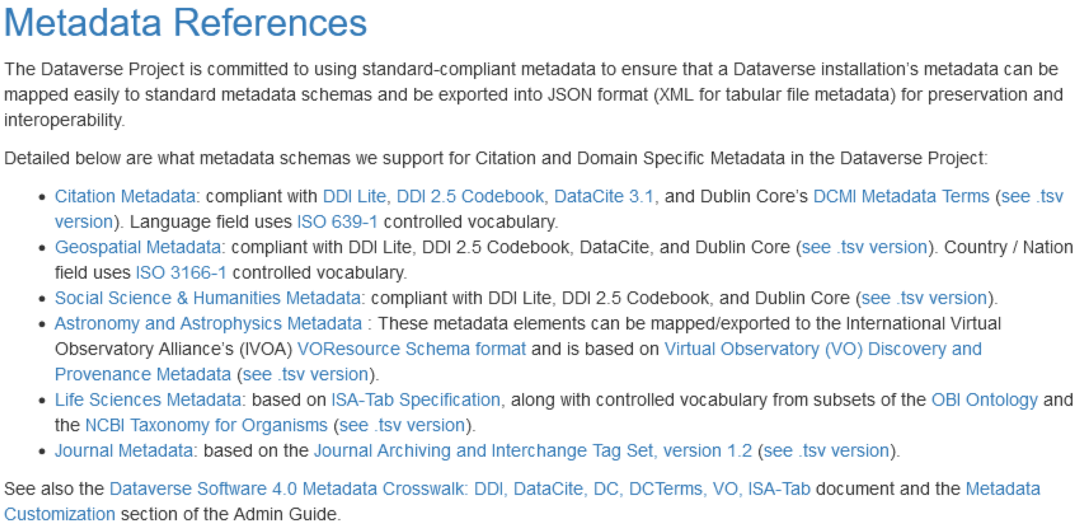
Learn about Data Citation Standards.

Accessibility

- Download of public files via browser or via API
- “As open as possible, but as closed as necessary”
- **Private URLs for pre-publication access (e.g. for reviewers)**

Interoperability

- Metadata standards



Metadata References

The Dataverse Project is committed to using standard-compliant metadata to ensure that a Dataverse installation's metadata can be mapped easily to standard metadata schemas and be exported into JSON format (XML for tabular file metadata) for preservation and interoperability.

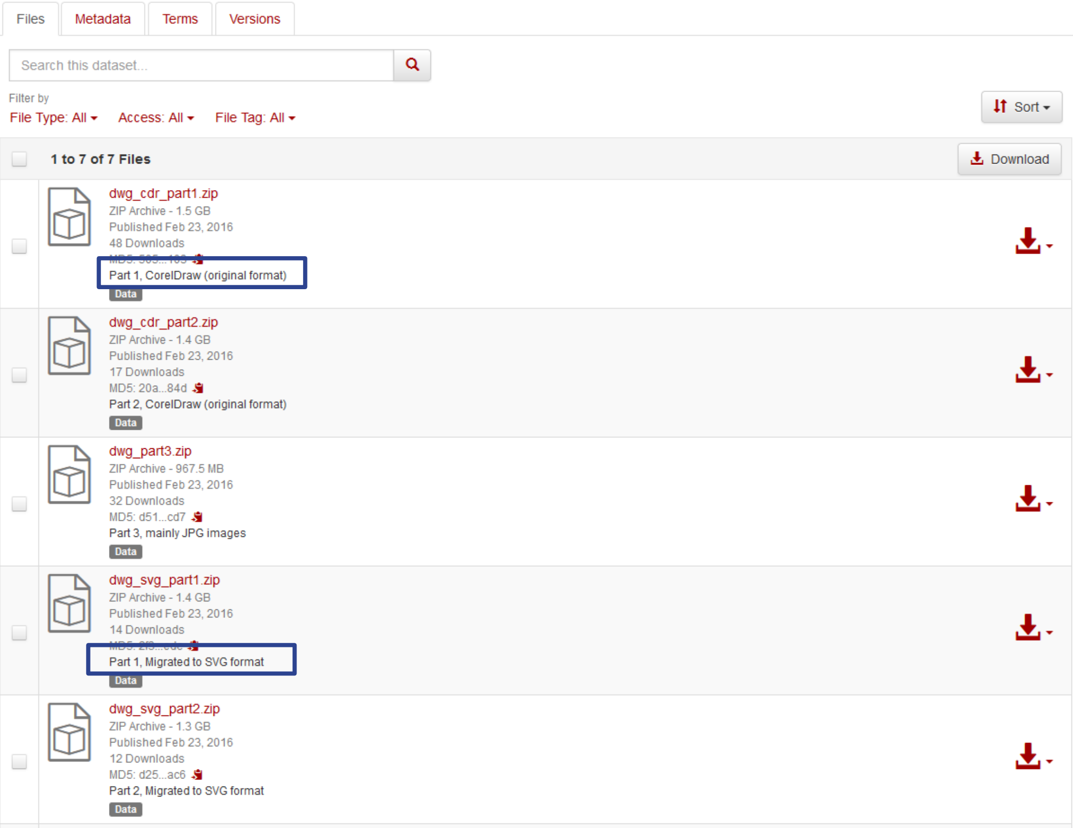
Detailed below are what metadata schemas we support for Citation and Domain Specific Metadata in the Dataverse Project:

- Citation Metadata: compliant with DDI Lite, DDI 2.5 Codebook, DataCite 3.1, and Dublin Core's DCMI Metadata Terms (see .tsv version). Language field uses ISO 639-1 controlled vocabulary.
- Geospatial Metadata: compliant with DDI Lite, DDI 2.5 Codebook, DataCite, and Dublin Core (see .tsv version). Country / Nation field uses ISO 3166-1 controlled vocabulary.
- Social Science & Humanities Metadata: compliant with DDI Lite, DDI 2.5 Codebook, and Dublin Core (see .tsv version).
- Astronomy and Astrophysics Metadata : These metadata elements can be mapped/exported to the International Virtual Observatory Alliance's (IVOA) VOResource Schema format and is based on Virtual Observatory (VO) Discovery and Provenance Metadata (see .tsv version).
- Life Sciences Metadata: based on ISA-Tab Specification, along with controlled vocabulary from subsets of the OBI Ontology and the NCBI Taxonomy for Organisms (see .tsv version).
- Journal Metadata: based on the Journal Archiving and Interchange Tag Set, version 1.2 (see .tsv version).

See also the Dataverse Software 4.0 Metadata Crosswalk: DDI, DataCite, DC, DCTerms, VO, ISA-Tab document and the Metadata Customization section of the Admin Guide.

Interoperability

- Metadata standards
- **Advice on suitable file formats, support with format conversion**
- **Technical validity checks**



1 to 7 of 7 Files

dwg_cdr_part1.zip
ZIP Archive - 1.5 GB
Published Feb 23, 2016
48 Downloads
MD5: 505...103
Part 1, CorelDraw (original format)
Data

dwg_cdr_part2.zip
ZIP Archive - 1.4 GB
Published Feb 23, 2016
17 Downloads
MD5: 20a...84d
Part 2, CorelDraw (original format)
Data

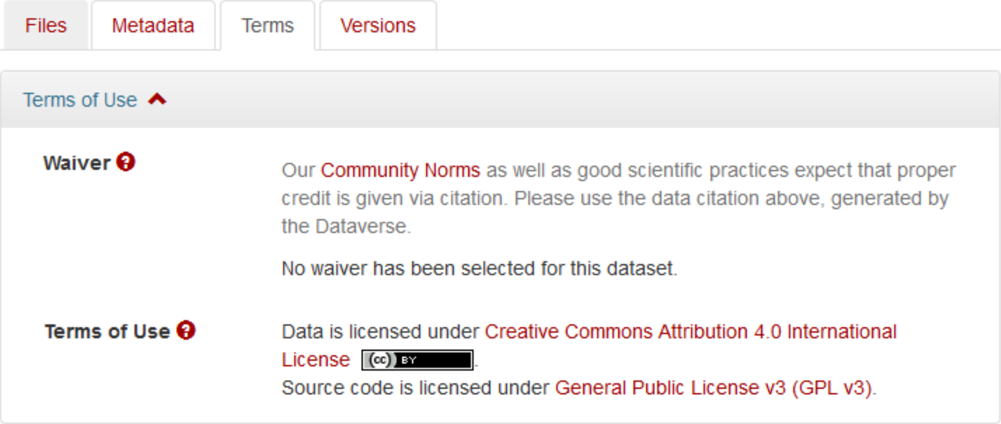
dwg_part3.zip
ZIP Archive - 967.5 MB
Published Feb 23, 2016
32 Downloads
MD5: d51...cd7
Part 3, mainly JPG images
Data

dwg_svg_part1.zip
ZIP Archive - 1.4 GB
Published Feb 23, 2016
14 Downloads
MD5: 2f3...cdc
Part 1, Migrated to SVG format
Data

dwg_svg_part2.zip
ZIP Archive - 1.3 GB
Published Feb 23, 2016
12 Downloads
MD5: d25...ac6
Part 2, Migrated to SVG format
Data

Re-Usability

- Open content licenses



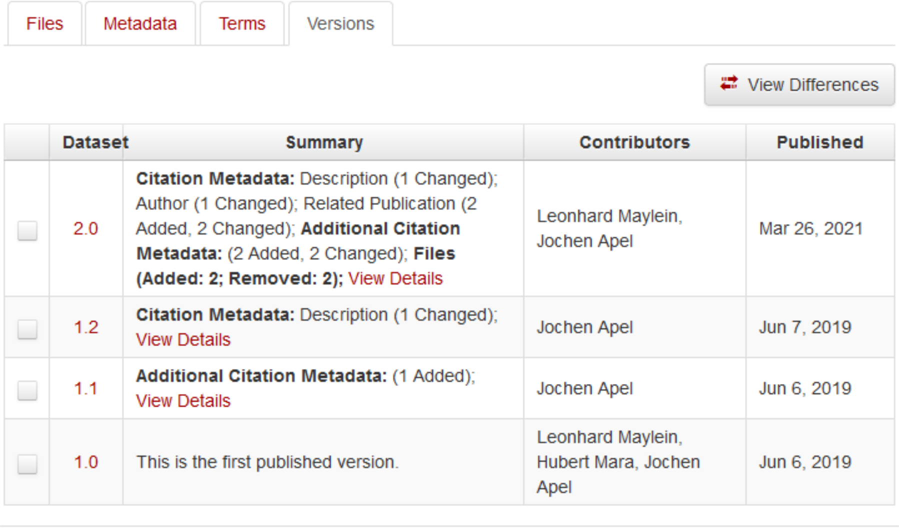
Waiver

Our Community Norms as well as good scientific practices expect that proper credit is given via citation. Please use the data citation above, generated by the Dataverse.

No waiver has been selected for this dataset.

Terms of Use

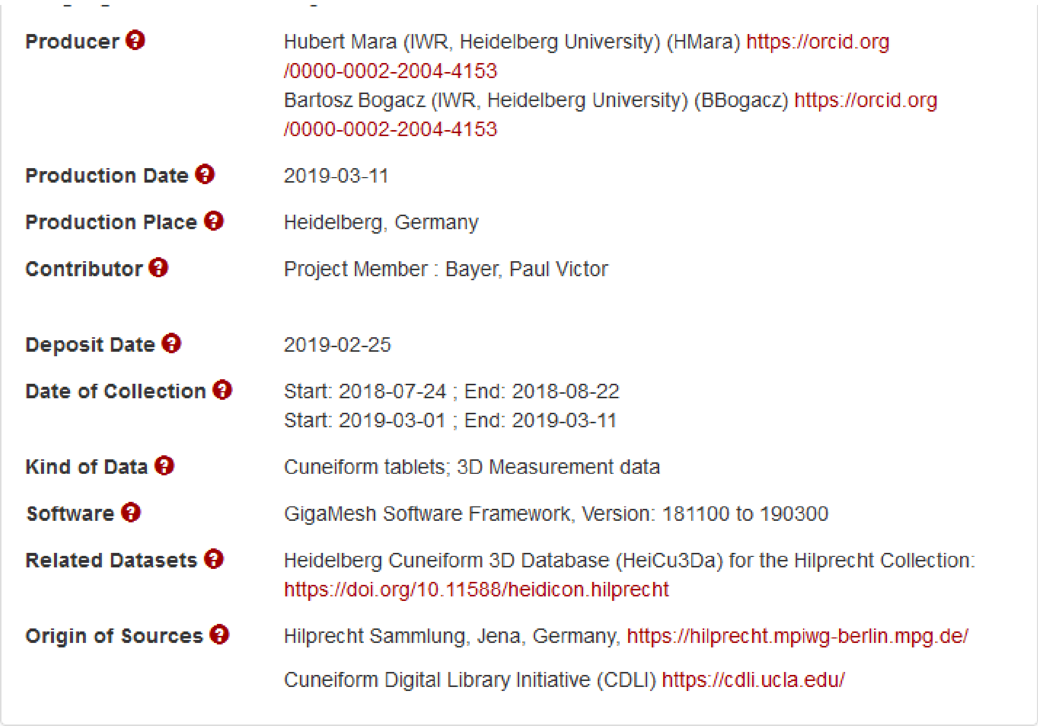
Data is licensed under Creative Commons Attribution 4.0 International License (cc) BY.
Source code is licensed under General Public License v3 (GPL v3).



	Dataset	Summary	Contributors	Published
☐	2.0	**Citation Metadata:** Description (1 Changed); Author (1 Changed); Related Publication (2 Added, 2 Changed); **Additional Citation Metadata:** (2 Added, 2 Changed); **Files (Added: 2; Removed: 2);** View Details	Leonhard Maylein, Jochen Apel	Mar 26, 2021
☐	1.2	**Citation Metadata:** Description (1 Changed); View Details	Jochen Apel	Jun 7, 2019
☐	1.1	**Additional Citation Metadata:** (1 Added); View Details	Jochen Apel	Jun 6, 2019
☐	1.0	This is the first published version.	Leonhard Maylein, Hubert Mara, Jochen Apel	Jun 6, 2019

Re-Usability

- Open content licenses
- **Transparent versioning**



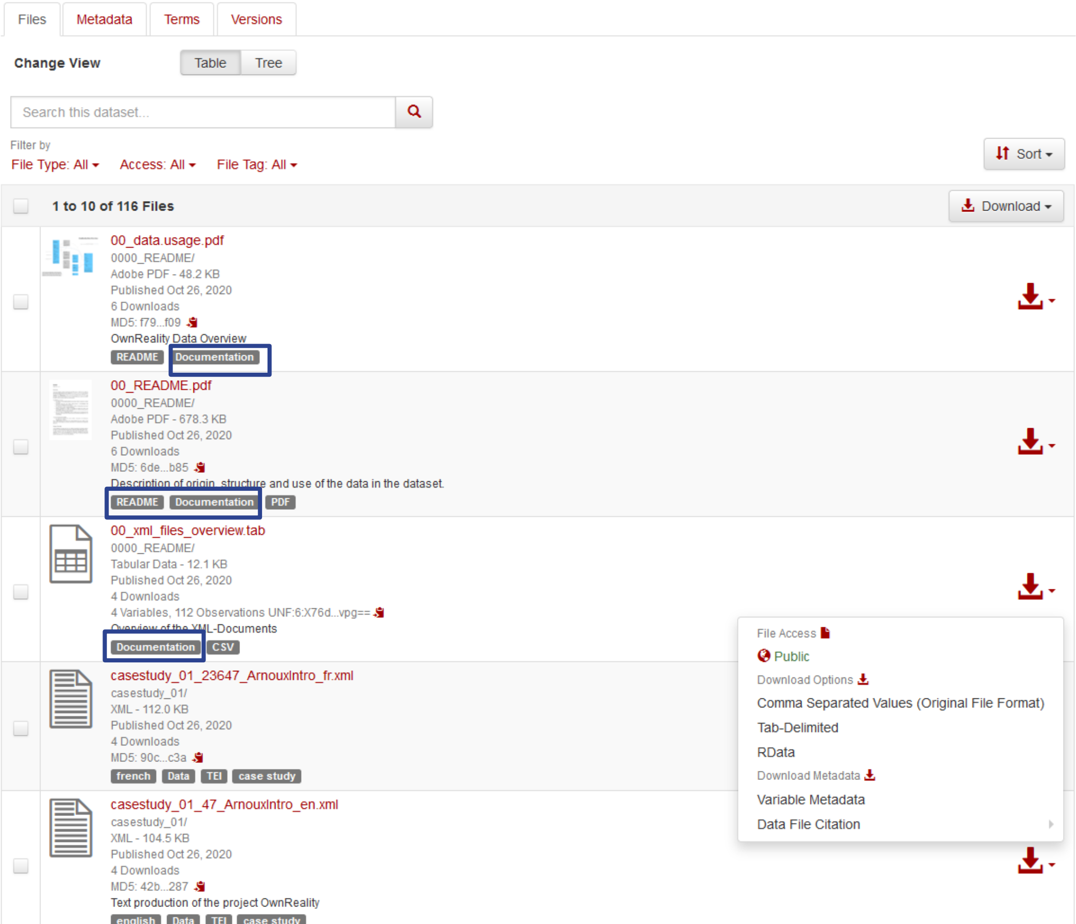
Producer	Hubert Mara (IWR, Heidelberg University) (HMara) https://orcid.org/0000-0002-2004-4153 Bartosz Bogacz (IWR, Heidelberg University) (BBogacz) https://orcid.org/0000-0002-2004-4153
Production Date	2019-03-11
Production Place	Heidelberg, Germany
Contributor	Project Member : Bayer, Paul Victor
Deposit Date	2019-02-25
Date of Collection	Start: 2018-07-24 ; End: 2018-08-22 Start: 2019-03-01 ; End: 2019-03-11
Kind of Data	Cuneiform tablets; 3D Measurement data
Software	GigaMesh Software Framework, Version: 181100 to 190300
Related Datasets	Heidelberg Cuneiform 3D Database (HeiCu3Da) for the Hilprecht Collection: https://doi.org/10.11588/heidicon.hilprecht
Origin of Sources	Hilprecht Sammlung, Jena, Germany, https://hilprecht.mpiwg-berlin.mpg.de/ Cuneiform Digital Library Initiative (CDLI) https://cdli.ucla.edu/

Re-Usability

- Open content licenses
- Transparent versioning
- **Provenance information**

Re-Usability

- Open content licenses
- transparent versioning
- Provenance information
- **Documentation files**



☐ 1 to 10 of 116 Files

☐ 00_data.usage.pdf
0000_README/
Adobe PDF - 48.2 KB
Published Oct 26, 2020
6 Downloads
MD5: f79...f09
OwnReality Data Overview
README Documentation

☐ 00_README.pdf
0000_README/
Adobe PDF - 678.3 KB
Published Oct 26, 2020
6 Downloads
MD5: 6de...b85
Description of origin, structure and use of the data in the dataset.
README Documentation PDF

☐ 00_xml_files_overview.tab
0000_README/
Tabular Data - 12.1 KB
Published Oct 26, 2020
4 Downloads
4 Variables, 112 Observations UNF:6:X76d...vpg==
Overview of the XML-Documents
Documentation CSV

☐ casestudy_01_23647_ArnouxIntro_fr.xml
casestudy_01/
XML - 112.0 KB
Published Oct 26, 2020
4 Downloads
MD5: 90c...c3a
french Data TEI case study

☐ casestudy_01_47_ArnouxIntro_en.xml
casestudy_01/
XML - 104.5 KB
Published Oct 26, 2020
4 Downloads
MD5: 42b...287
Text production of the project OwnReality
english Data TEI case study

Generic repositories

Featured communities

Chicago COVID-19 Response

Recent uploads

A large-scale COVID-19 Twitter chatter dataset for open scientific research - an international collaboration

Need help?

Contact us

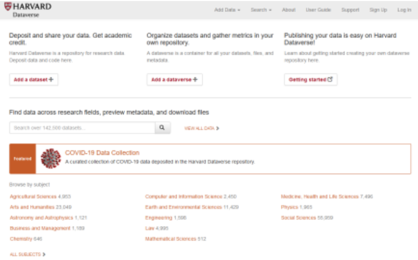
BIP4COVID19: Impact metrics and indicators for coronavirus related

The place to share your research

OSF is a free, open platform to support your research and enable collaboration.

Discover public research

Discover projects, data, materials, and collaborators on OSF that might be helpful to your own research.



Deposit and share your data. Get academic credit.

Harvard Dataverse is a repository for research data. Deposit data and code here.

Organize datasets and gather metrics in your own repository.

A dataverse is a container for all your datasets, files, and metadata.

Publishing your data is easy on Harvard Dataverse!

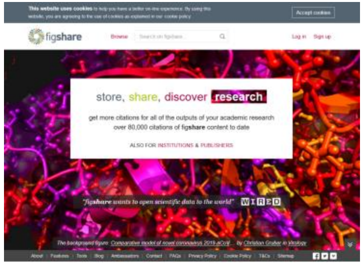
Learn about getting started creating your own dataverse repository here.

Find data across research fields, preview metadata, and download files

COVID-19 Data Collection

A curated collection of COVID-19 data deposited in the Harvard Dataverse repository.

Browse by subject



store, share, discover research

get more citations for all of the outputs of your academic research
over 80,000 citations of figshare content to date

ALSO FOR INSTITUTIONS & PUBLISHERS

"figshare wants to open scientific data to the world"

https://zenodo.org/

https://osf.io/

https://dataverse.harvard.edu/

https://figshare.com/

- Providing suitable technology for FAIR Data, partly even very elaborate functionality
- No or only basic domain-specific functionalities
- Only basic individual counselling
- Non-commercial and commercial providers

Wrap up

Managing data properly allows for reuse - by you or your community!

Managing data costs less than not managing it!

Manage data from the beginning of your research!

Manage data according to FAIR principles!

Use trustworthy data repositories!

Register for our upcoming events: https://4euplus.eu/4EU-273.html

Next sessions:

- “Open Research Software” | 21 March 2022, tba
- “Open science and the role of rights management” | 4 April, 2022, 12:30 - 14:00
- 4 further sessions after April 2022

4eu+

Thank you!

Jochen APEL
Paola GALIMBERTI
Milan JANÍČEK

Open for you! An introduction series to open science | 7 March 2022

CHARLES UNIVERSITY

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

SORBONNE UNIVERSITÉ

UNIVERSITY OF COPENHAGEN

UNIVERSITÀ DEGLI STUDI DI MILANO

UNIVERSITY OF WARSAW