

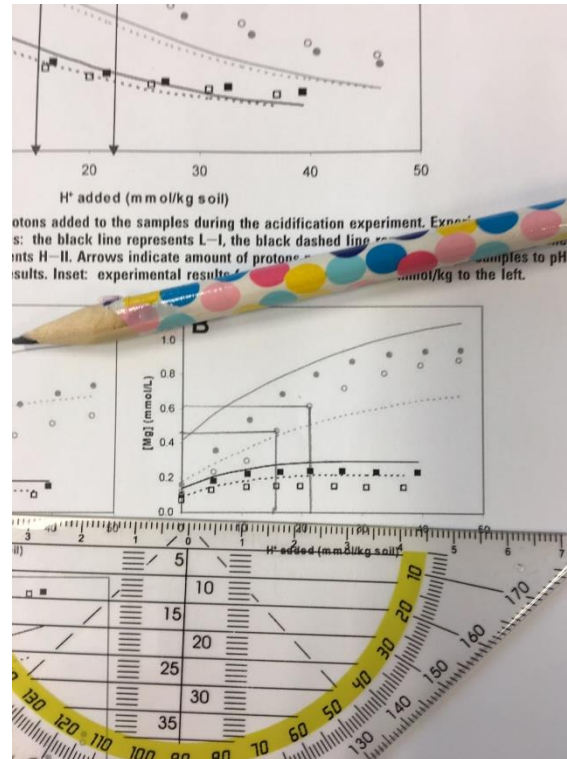
Workshop “Zoeken naar onderzoeksdata”

Hilde van Zeeland & Ellen Fest

VOGIN-IP-LEZING 21 maart 2019

Waarom zoeken naar datasets?

Even terug naar 15 jaar geleden



Proton buffering and metal mobility in Dutch sandy soils
- modeling laboratory and field data -



Ellen P.M.J. Fest

Digitalisering maakt het mogelijk



Zeer veel data via internet beschikbaar



Dataregister van de Nederlandse Overheid

Zoek binnen de [12611](#) beschikbare datasets

Bijvoorbeeld: verkiezingsuitslag 2018

[Zoek open data](#)



OECD iLibrary

Search all content by title

Advanced Search

Browse by Theme ▾

Browse by Country ▾

[Home](#) > [Statistics](#) > [OECD Agriculture Statistics](#)

OECD Agriculture Statistics

The OECD databases on agriculture constitute a unique collection of a quantifying and analysing the agricultural economy. This includes fore agricultural markets and commodities, detailed estimates of policy su performance of agriculture. Data concern both OECD countries and nc



Public Data

A close-up photograph of a hand holding a black magnifying glass over a laptop keyboard. The magnifying glass is positioned over the spacebar and the keys immediately surrounding it. The text 'Focus op onderzoeksdata' is overlaid in white, centered horizontally and partially obscured by the magnifying glass's frame. The background is a blurred laptop keyboard and screen, suggesting a research or data analysis environment.

Focus op onderzoeksdata

In bepaalde vakgebieden al heel gebruikelijk



fotoverantwoording: melkweg: [justin beck](#) via [Unsplash](#), weerstation: Project SCREEN, Center for Climate Change, Universitat Rovira i Virgili, Spain (CC BY 3.0), DNA: [TheDigitalArtist](#) via Pixabay, oceaanonderzoek: U.S. Navy photo (Public domain)

Behoefte aan transparantie vanwege
fraude en reproduceerbaarheid

Measure by measure: Diederik Stapel count rises again, to 54

with 6 comments

[Diederik Stapel](#) is up to 54 retractions.

Here's the [notice](#) from *Self and Identity*:



Schwinghammer, S. A., & Stapel, D. A. (2011). Measure by measure: When implicit and explicit social comparison effects differ. *Self and Identity*, 10, 166-173. DOI: 10.1080/15298861003680117.

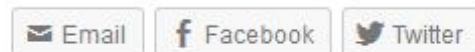
The above named article has been retracted from publication in *Self and Identity* by agreement of the journal Editor, the Publisher and the Co-authors. The retraction has been agreed after the Noort Committee determined fraud in the publication. For further details, please visit the following link <https://www.commissielevelt.nl/>



The paper has been cited just once, according to Thomson Scientific's Web of Knowledge, in a paper published this spring called "[Attractiveness of Limited Edition Artwork for First Generation Newly Affluent Consumers](#)." It's Stapel's [second retraction from *Self and Identity*](#).

Hat tip: Rolf Degen

Share this:



Subscribe to Blog via Email

Join 12,056 other subscribers

Pages

[Help us: Here's some of what we're working on](#)

[How you can support Retraction Watch](#)

[Meet the Retraction Watch staff](#)

[About Adam Marcus](#)

[About Ivan Oransky](#)

[The Center For Scientific Integrity](#)

[Board of Directors](#)

[The Retraction Watch FAQ, including comments policy](#)

[The Retraction Watch](#)

Reinhart, Rogoff... and Herndon: The student who caught out the profs

By Ruth Alexander
BBC News

🕒 20 April 2013 | Magazine

This week, economists have been astonished to find that a famous academic paper often used to make the case for austerity cuts contains major errors. Another surprise is that the mistakes, by two eminent Harvard professors, were spotted by a student doing his homework.

It's 4 January 2010, the Marriott Hotel in Atlanta. At the annual meeting of the American Economic Association, Professor Carmen Reinhart and the former chief economist of the International Monetary Fund, Ken Rogoff,



A Survey on Data Reproducibility in Cancer Research Provides Insights into Our Limited Ability to Translate Findings from the Laboratory to the Clinic

Aaron Mobley, Suzanne K. Linder, Russell Braeuer, Lee M. Ellis , Leonard Zwelling 

Published: May 15, 2013 • <http://dx.doi.org/10.1371/journal.pone.0063221>

83 Saves	38 Citations
13,650 Views	145 Shares

Article	Authors	Metrics	Comments	Related Content
				

Download PDF 
Print
Share

Abstract

- Introduction
- Methods
- Results
- Discussion
- Supporting Information
- Author Contributions
- References

- Reader Comments (2)
- Media Coverage (0)
- Figures

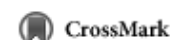
Abstract

Background

The pharmaceutical and biotechnology industries depend on findings from academic investigators prior to initiating programs to develop new diagnostic and therapeutic agents to benefit cancer patients. The success of these programs depends on the validity of published findings. This validity, represented by the reproducibility of published findings, has come into question recently as investigators from companies have raised the issue of poor reproducibility of published results from academic laboratories. Furthermore, retraction rates in high impact journals are climbing.

Methods and Findings

To examine a microcosm of the academic experience with data reproducibility, we surveyed the faculty and trainees at MD Anderson Cancer Center using an anonymous computerized questionnaire; we sought to ascertain the frequency and potential causes of non-reproducible data. We found that ~50% of respondents had experienced at least one episode of the inability



Subject Areas

- Trainees
- Reproducibility
- Cancer drug discovery
- Drug discovery
- Surveys
- Biomarkers
- Research laboratories
- Drug research and d...

Beleid opgesteld door financiers,
universiteiten en tijdschriften



OPEN RESEARCH DATA IN HORIZON 2020

http://ec.europa.eu/research/press/2016/pdf/opendata-infographic_072016.pdf#view=fit&pagemode=none



[contact](#) [kalender](#) [pers](#) [vacatures](#) [English](#)

[Home](#)

[Actueel](#)

[Financiering](#)

[Onderzoek & resultaten](#)

Open (FAIR) data

[Open Science](#)

[Datamanagement](#)
Datamanagementparagraaf

[Contactpersonen](#)

[Plan S](#)

Naast publicaties dienen ook onderzoeksdata die voortkomen uit projecten die NWO financiert, zo veel mogelijk open en voor hergebruik beschikbaar gesteld worden. Daarbij is het adagium: 'open als het kan, beschermd als het moet'. Aspecten als privacy, openbare veiligheid, ethische beperkingen, eigendomsrecht en commerciële belangen kunnen argumenten zijn om af te wijken van deze regel.

<https://www.nwo.nl/beleid/open+science/datamanagement>



<https://www.wur.nl/en/Value-Creation-Cooperation/WDCC/Data-Management-WDCC/Data-policy.htm>
<https://weblog.wur.eu/openscience/new-research-data-policy-at-wur/>



<https://www.organisatiegids.universiteitleiden.nl/reglementen/algemeen/regeling-datamanagement-univeriteit-leiden>

Regeling Datamanagement Universiteit Leiden



**University policy framework for research data
Utrecht University**

<https://www.uu.nl/en/research/research-data-management/guides/policies-codes-of-conduct-and-laws>

“PLOS journals **require** authors to make all data underlying the findings described in their manuscript fully available without restriction at the time of publication. When specific legal or ethical requirements prohibit public sharing of a dataset, authors must indicate how researchers may obtain access to the data.

When submitting a manuscript, authors must provide a Data Availability Statement describing compliance with PLOS's policy. If the article is accepted for publication, the data availability statement will be published as part of the accepted article.”

<https://journals.plos.org/plosone/s/data-availability>

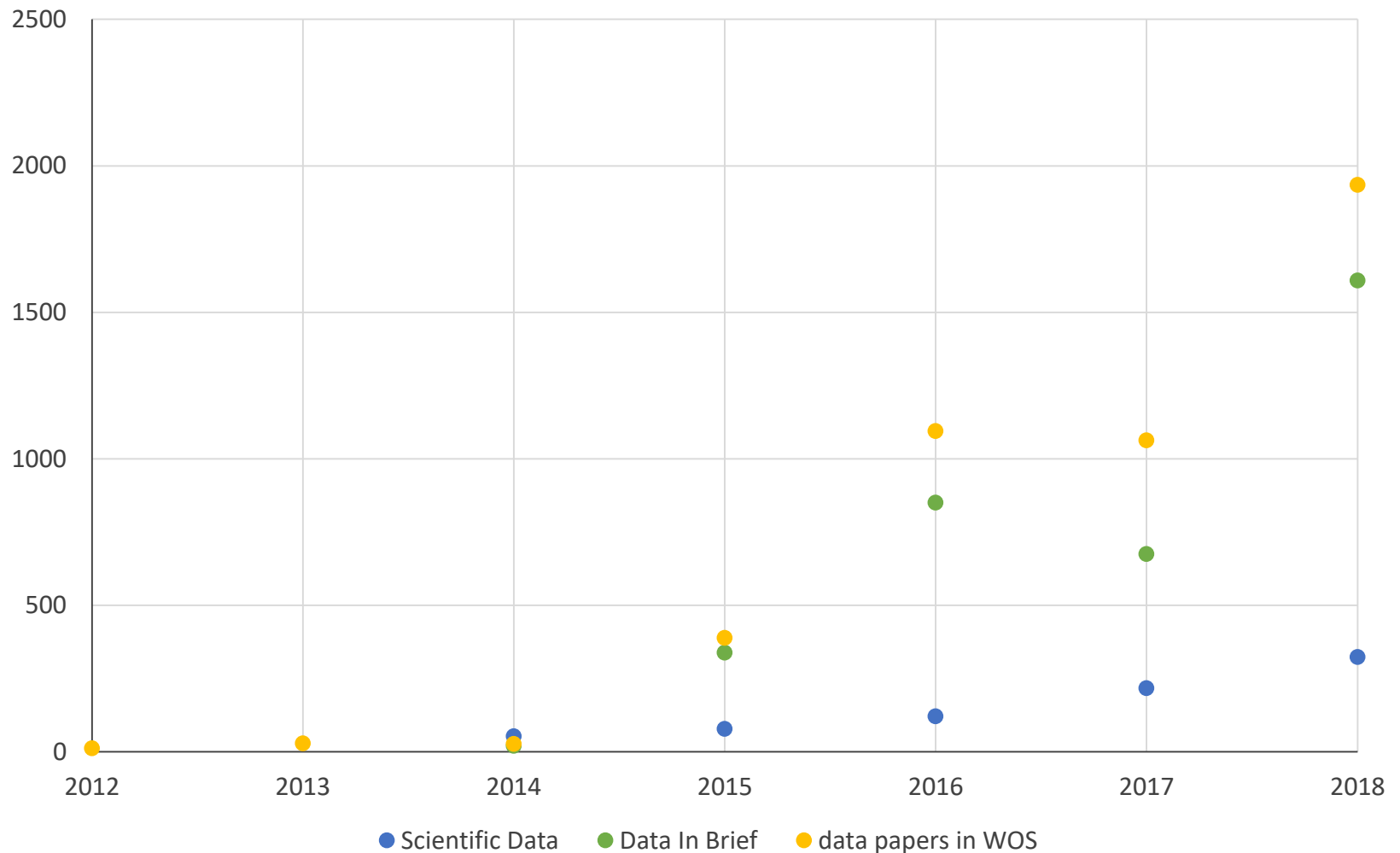
“All manuscripts reporting original research published in Nature journals **must** include a [data availability statement](https://www.nature.com/authors/policies/availability.html). Data availability statements should provide a statement about the availability of the minimal dataset that would be necessary to interpret, replicate and build upon the methods or findings reported in the article.”

<https://www.nature.com/authors/policies/availability.html>

As a researcher, you are increasingly **encouraged**, or even mandated, to make your research data available, accessible, discoverable and usable.

Meer en meer datasets beschikbaar

Growth of data publications



Waar kan je zoeken?



Drie zoekplekken



De literatuur



Data repositories



Zoekmachines

De literatuur



Relatie literatuur-data

1


Data in/bij publicatie

Appendix / supplementary file



Voorbeeld supplementary files

Combined Effects of Ocean Warming and Acidification on Copepod Abundance, Body Size, and Reproductive Success

Jessica Garzke , Thomas Hansen, Stefanie M. H. Isenhardt

Published: May 25, 2016 • <https://doi.org/10.1371/journal.pone.0165888>

Article

Authors



Abstract

Introduction

Materials and Methods

Results

Discussion

Supporting Information

Acknowledgments

Author Contributions

References

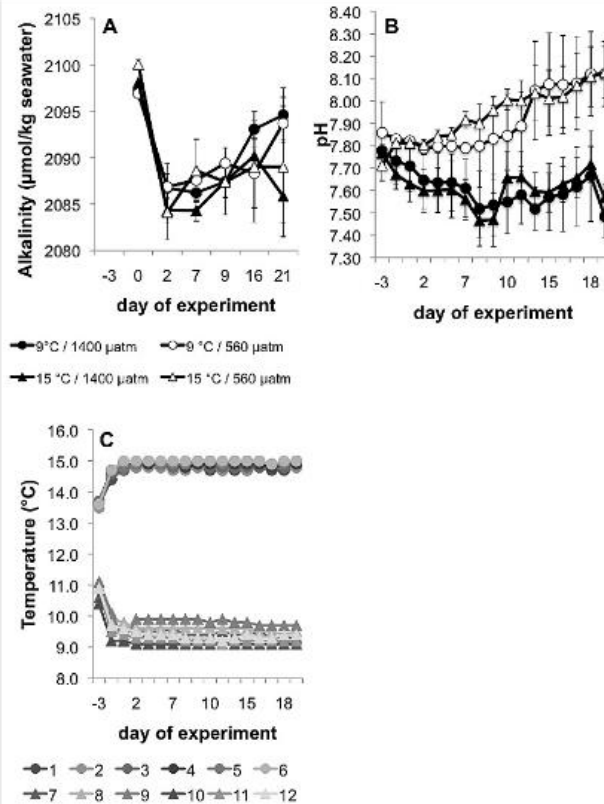
Abstract

Concerns about increasing ocean temperatures and acidification have led to studies on the consequences for marine ecosystems. We present the combined effects of warming and acidification on copepod abundance, body size, and reproductive success.

As a measure of nutritional quality, we used the abundance of copepods and nauplii as consumers in marine ecosystems. Copepod abundance was negatively affected by warming and acidification. The abundance of adult copepods was negatively affected by warming. The

Supporting Information

S1 Fig.TIFF



1 / 7



Download

Relatie literatuur-data

1

Data in/bij publicatie

Appendix / supplementary file



2

Data elders

Losse dataset, vaak wel gelinkt



Voorbeeld dataset elders

MENU ▾

Red clover (*Trifolium pratense* L.) draft genome provides a platform for trait improv...

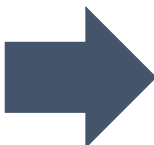
PDF

Tools ▾



Additional Information

Accession codes: All shotgun read libraries ([Supplementary Table 5](#)) and the assembly are deposited in the European Nucleotide Archive (accession PRJEB9186).



Data Availability: The genome assembly and annotation can also be downloaded as individual files (<http://dx.doi.org/10.5281/zenodo.17232>) or browsed online (http://tgac-browser.tgac.ac.uk/trifolium_pratense). Accession numbers of the BAC end sequences are HR235466-298279.

How to cite this article: De Vega, J. J. *et al.* Red clover (*Trifolium pratense* L.) draft genome provides a platform for trait improvement. *Sci. Rep.* **5**, 17394; doi: 10.1038/srep17394 (2015).

Sections

Figures

References

Abstract

Introduction

Results

Discussion

Methods

Additional Information

References

Acknowledgements

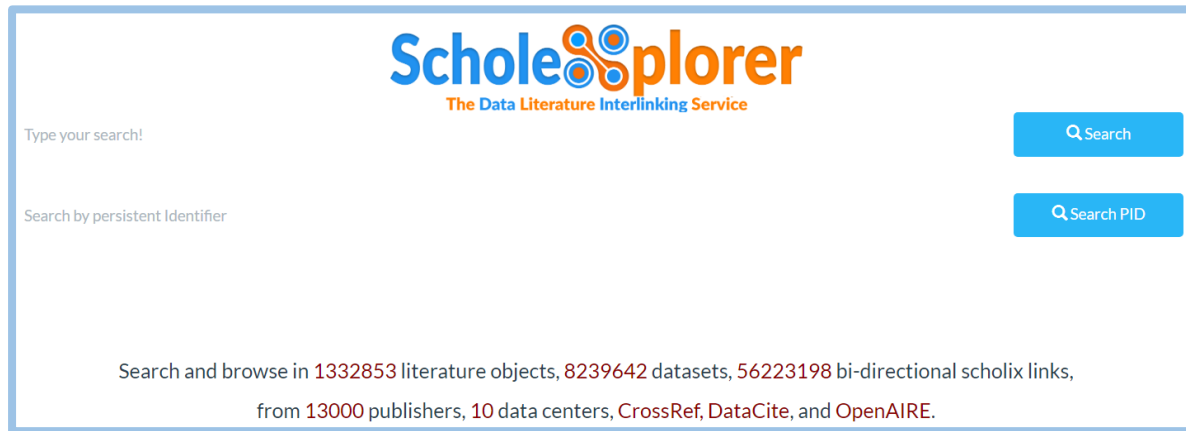
Author information

Supplementary information

Comments

Scholexplorer: data-literatuur links

<http://scholexplorer.openaire.eu>



Voorbeeld Scholexplorer: artikel → data



publication

COFFEE COMMODITY CHAIN

Tine Olsen , Brett Inder

To explain the value added along the coffee commodity chain we propose and estimate a theoretical model of the coffee commodity chain. The theoretical model consists of four markets and five agents in the coffee commodity chain and predicts that prices in the coffee commodity chain move together but are also influenced by income, technology and production. A vector error correction model is used to test the theoretical predictions. In addition to the theoretical conclusions the empirical model confirms the beneficial role of the International Coffee Agreement and the importance of the level of production in determining coffee prices.

[LINK TO DATASETS \(1\)](#)

[LINK TO PUBLICATIONS \(0\)](#)

[OTHER LINKS \(0\)](#)

dataset

[International Financial Statistics](#)

International Monetary Fund,

url : [10.3886/icpsr07629.v2](#)

Sources: OpenAIRE

Scholexplorer vindt niet alle data

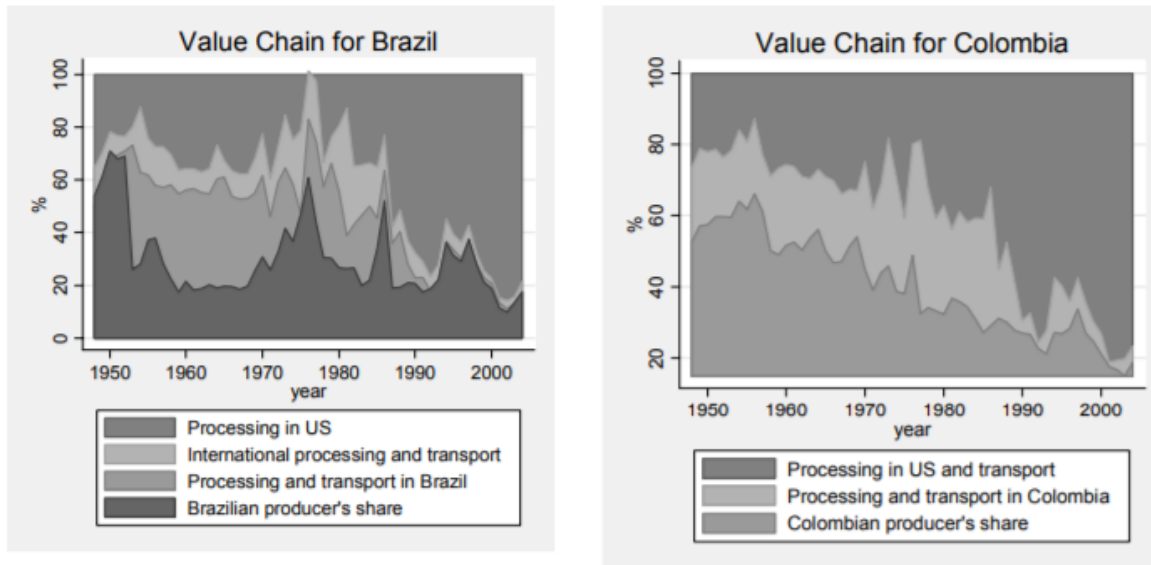


Fig. 1. *Distribution of the Coffee Dollar along the Commodity Chain.*

Sources: Brazilian and Colombian producer prices: FAO (various years), FAOSTAT (2006)

and ICO (2005). Export and import unit values: FAOSTAT online (2006) and U.S.

Department of Commerce: Bureau of the Census (1989). Wholesale prices for Brazil: IFS

(various years). US Retail prices: BLS (2005b).

dataset

International Financial Statistics
International Monetary Fund,
url : [10.3886/icpsr07629.v2](https://www.imf.org/external/data/10.3886/icpsr07629.v2)

Data repositories



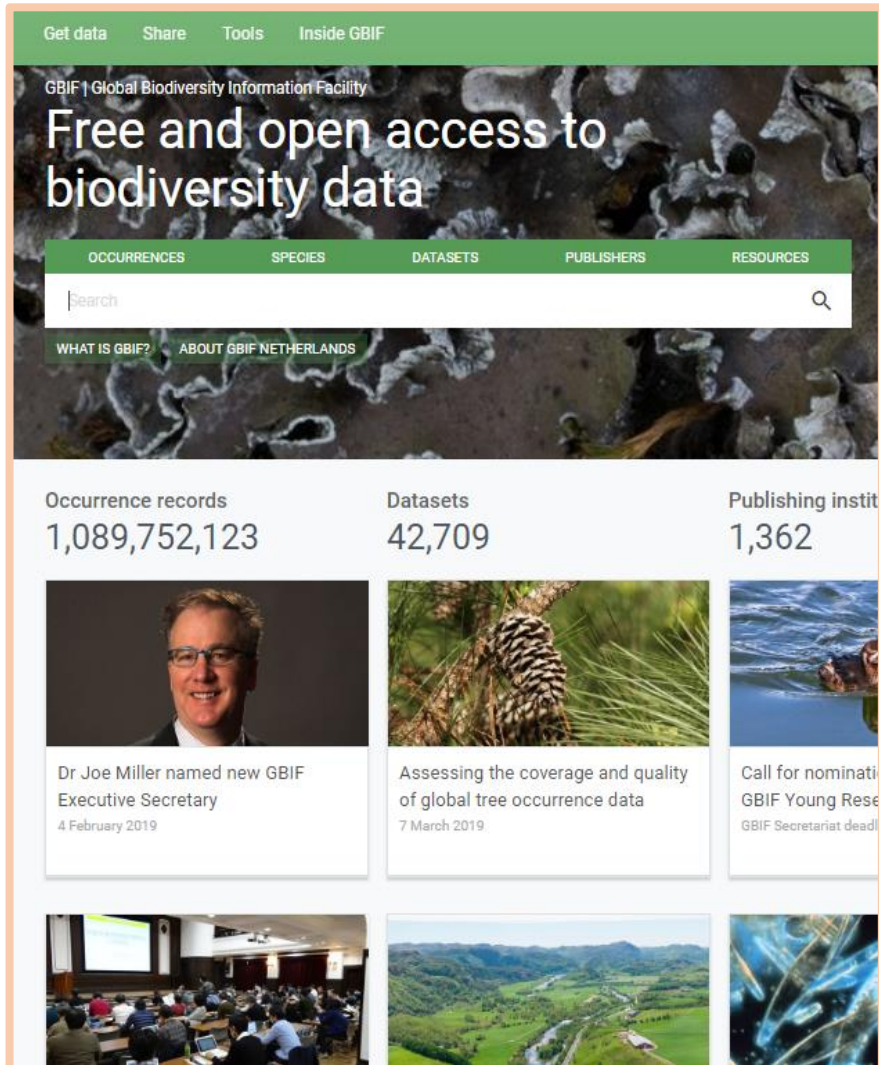
Data repositories

- Diensten voor langetermijnopslag van data
- Doorzoekbaar via websites
- Data open / gesloten / met restricties
- Disciplinair en multidisciplinair



Voorbeelden data repositories

Disciplinair: GBIF



Get data Share Tools Inside GBIF

GBIF | Global Biodiversity Information Facility

Free and open access to biodiversity data

OCCURRENCES SPECIES DATASETS PUBLISHERS RESOURCES

Search

WHAT IS GBIF? ABOUT GBIF NETHERLANDS

Occurrence records
1,089,752,123

Datasets
42,709

Publishing institutions
1,362

Dr Joe Miller named new GBIF Executive Secretary
4 February 2019

Assessing the coverage and quality of global tree occurrence data
7 March 2019

Call for nomination GBIF Young Researcher
GBIF Secretariat deadline

Multidisciplinair: Figshare



figshare Browse Search on figshare... Log in Sign up

Discover research from figshare

FEATURED CATEGORIES

Agricultural and Veterinary Sciences Astronomy, Astrophysics, Space Science Biological Sciences

Built Environment and Design Chemistry Commerce, Management, Tourism and Services

Earth and Environmental Sciences Engineering Health Sciences Humanities

Information and Computing Sciences Language, Communication and Culture Mathematics

Meta Science Physics Psychology Social Science Studies in Creative Arts and Writing

Studies in Human Society Technology Uncategorised

NEW POPULAR CATEGORIES SEARCH


Novel gene regulatory networks identified in response to nitro-conj...
Yanbo Fan 30/09/2025

Model outputs
Zhu Chenguang 10/03/2020


Activated Carbon-Loaded Polydimethylsiloxane Membranes ...
Bruno José Arcanjo Go... 01/12/2019

Hoe vind je repositories?

www.re3data.org



SearchBrowse ▾SuggestResources ▾Contact



Filter

Reset all

Subjects

Content Types

Countries

AID systems

API

Certificates

Data access

Data access restrictions

Database access

Database access restrictions

Database licenses

Data licenses

Data upload

Data upload restrictions

Enhanced publication

Institution responsibility type

Institution type

Keywords

Metadata standards

PID systems

Provider types

Quality management

Repository languages

Software

Syndications

Search...

Search

Toogle short help

← Previous

1

2

3

4

5

6

7

...

15

Next →

Sort by ▾

Found 364 result(s)

Claremont Colleges Digital Library

CCDL

Subject(s)

Humanities

Social and Behavioural Sciences

Social Sciences

Geosciences (including Geography)

Geography

Chemistry

Biology

Mathematics

Computer Science

Humanities and Social Sciences

Natural Sciences

Life Sciences

Computer Science, Electrical and System Engineering

Engineering Sciences

Content type(s)

Audiovisual data

Images

Plain text

other

Country

United States

The Claremont Colleges Digital Library (CCDL) provides access to historical and visual resources collections created both by and for The Claremont Colleges community.

Export Import Data Bank

Tradestat

Subject(s)

Economics

Statistics and Econometrics

Social and Behavioural Sciences

Humanities and Social Sciences

Content type(s)

Standard office documents

Scientific and statistical data formats

Country

India

Export Import Data Bank - Tradestat is an initiative of Department of Commerce, Ministry of Commerce & Industry, Government of India. This databank provides international trade figures for imports, exports and total trade, based on data published by the Directorate General of Commercial Intelligence and

Korte opdracht (~5min):

www.re3data.org

Stel, je wilt graag data over het stemgedrag van Amerikanen.
Zoek in Re3Data een repository met dit soort data.



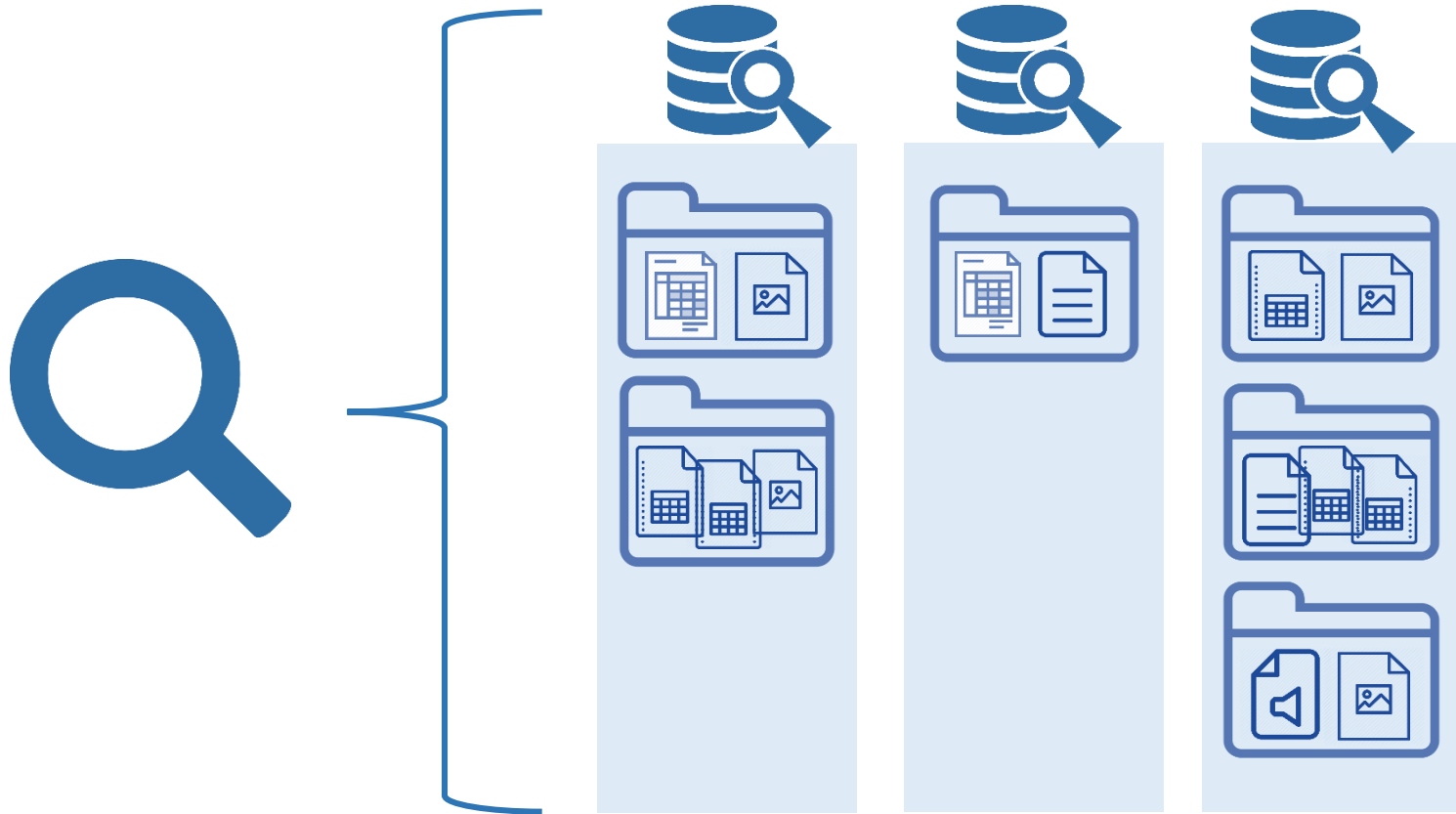
- Je kunt zoeken via de zoekbox of op vakgebied (Browse > Browse by subject > Text)
- Veel resultaten? Gebruik dan de filters aan de linkerkant.

Zoekmachines



Zoekmachines

Bieden niet de data, maar informatie over de data



Beta

DataSearch

late quaternary sediment core

Filter Results

reset

Data File Types

☐ Tabular Data (1480)
 ☐ Image (839)
 ☐ Document (150)
 ☐ Text (95)
 ☐ File Set (92)
 ☐ Geospatial Data (22)
 ☐ Video (2)
 ☐ Unknown (1)
 ☐ Software/Code (1)
 ☐ Slides (1)

Data Source Types

☐ Data Repositories (898)
 ☐ Article Repositories (837)

1735 results for late quaternary sediment core

PANGAEA.

Pollen abundance in Late Quaternary sediment cores off West Africa

Rossignol-Strick, Martine & Duzer, Daniele - 1979-11-15

late Quaternary cores off Senegal shows a juxtaposition of Mediterranean...Late Quaternary sediment cores off West Africa...core 47. Cosmopolitan taxa dominate the assemblage and only a few species...core. Dinoflagellate cysts display a tropical assemblage with mostly estuarine...Late Quaternary pollen and dinoflagellate cysts in marine cores off...

TABULAR DATA 3TEXT

PANGAEA.

Age determination of late Quaternary sediment cores from the South Tasman Rise

Moy, Christopher M, Howard, William R & Gagan, Michael K - 2006-10-29

late Quaternary sediment cores from the South Tasman Rise...Sediment cores MD972106 (45° 09' S, 146° 17' E, 3310 m water depth) and...Quaternary Science, 21(7), 763-777, https://doi.org/10.1002/jqs.1067 ...Late Quaternary palaeoceanography of the Circumpolar Deep Water from the...sediment cores from the South Tasman Rise (STR) to...

FILE SET

ScienceDirect

Dynamics of Late Quaternary North African humid periods documented in the clay mineral record of central Aegean Sea sediments

Werner Ehrmann, Martin Seidel & Gerhard Schmiedl - 2013-08-01

in sediments of core GeoTü SL143, central Aegean Sea, combined with the...sediments of core GeoTü SL143, central Aegean Sea, combined with the occurrences...record in sediment core GeoTü SL123 from the southernmost Aegean Sea (...Late Quaternary...record in sediment core GeoTü SL143 from the central Aegean Sea. AHP: ...late Quaternar...

IMAGE 3TABULAR DATA

late quaternary sediment core



reset

1735 results for late quaternary sediment core

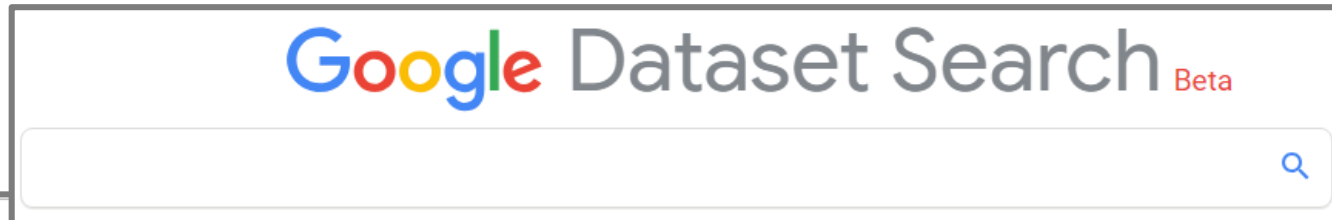
Slides (1)

TEXT

FILE SET

TABULAR DATA

Voorbeeld: Google Dataset Search



Google Dataset Search

Q "climate change" "food production"

54 results found



Effects of Climate Change on
Global Food Production v1

data.world

Updated Sep 12, 2018



Effects of Climate Change on
Global Food Production from...

catalog.data.gov
data.wu.ac.at

Updated Aug 9, 2018



Data from: Climate Change,
Agriculture and Food Securit...

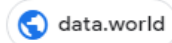
explore.openaire.eu

Created Sep 1, 2012



Dataset of MPI-ESM

Effects of Climate Change on Global Food Production v1



Dataset updated Sep 12, 2018

Dataset published Jun 28, 2017

Dataset provided by

data.world

Available download formats from providers

zip, csv

Description

Source: [data.world](#)

Purpose:

To provide an assessment of potential climate change impacts on world crop production.

Abstract:

Effects of Climate Change on Global Food Production from SRES Emissions and Socioeconomic Scenarios is an update to a major work published in 1997, based on output of HadCM2 model forced with greenhouse gas concentration from the IS95 emission scenario.

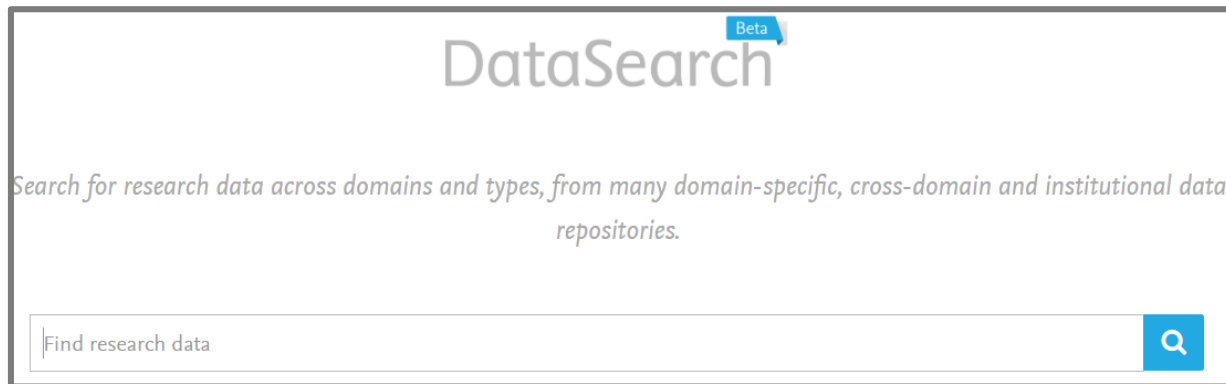
Meer zoekmachines voor data

platform	Link + info
Elsevier DataSearch	datasearch.elsevier.com datasets en data uit artikelen
Google Dataset Search	toolbox.google.com/datasetsearch data die online te vinden is
DataCite Search	search.datacite.org datasets, maar soms ook bijv. artikelen
WorldWideScience	worldwidescience.org diverse outputtypen
Zanran	zanran.com PDFs met numerieke data

Korte opdracht (~5min):

Zoek in Elsevier DataSearch naar tabellen uit artikelen met daarin data over de eetgewoonten van tieners.

<https://datasearch.elsevier.com/#/>

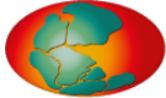



- Speel met de filters aan de linkerkant
- Gebruik aanhalingstekens (bijv. “cat allergy”)


Data gevonden! Wat mag je er nu mee?


- Bekijk de licentie
- De meeste repositories gebruiken de Creative Commons
- Geef altijd een referentie, ook bij CC0



**PANGAEA.**
Data Publisher for Earth & Environmental Science

Citation: **Ferger, Stefan; Schleuning, Matthias; Hemp, Andreas; Howell, Kim; Böhning-Gaese, Katrin (2018):** Various investigations to analyze the effects on species richness of birds during the KiLi (Kilimanjaro) Project. *PANGAEA*,  <https://doi.org/10.1594/PANGAEA.896128>,

Supplement to: Ferger, S et al. (2014): Food resources and vegetation structure mediate climatic effects on species richness of birds. *Global Ecology and Biogeography*, **23(5)**, 541-549,  <https://doi.org/10.1111/geb.12151>

 **Creative Commons Attribution 4.0 International**

Opdrachten (~35 min)

Begin met de door jullie gekozen zoekopdracht!



Discussie opdrachten

1. Wat is jullie opgevallen tijdens het zoeken in:

- Elsevier DataSearch
- DataCite
- WorldWideScience
- Google Dataset Search

Elsevier DS	Google DS	DataCite Search	WorldWideScience
<ul style="list-style-type: none">+ Dekt ook data uit artikelen+ Previews	<ul style="list-style-type: none">+ Aantal citaties slaat niet altijd op de dataset	<ul style="list-style-type: none">+ Toont licenties bij resultaten	<ul style="list-style-type: none">+ Veel filters+ ook resultaten in andere talen
<ul style="list-style-type: none">- Mist data uit DataCite en Figshare	<ul style="list-style-type: none">- Geen filters- Metadata vaak niet compleet- Citaties: niet altijd de data	<ul style="list-style-type: none">- Veel resultaten niet relevant- Weinig metadata in het record	<ul style="list-style-type: none">- Filters dekken niet alle resultaten

Discussie opdrachten

1. Wat is jullie opgevallen tijdens het zoeken in:

- Elsevier DataSearch
- DataCite
- WorldWideScience
- Google Dataset Search

	Elsevier DataSearch	Google Dataset Search	DataCite Search	WorldWideScience
Website	https://datasearch.elsevier.com/	https://toolbox.google.com/datasetsearch	https://search.datacite.org/	https://worldwidescience.org/
Definitie	Zoekmachine voor data, aangeboden door Elsevier	Zoekmachine voor data, aangeboden door Google	Zoekmachine van data met een DataCite-DOI	Zoekmachine voor verschillende wetenschappelijke outputtypen
Dekking	<ul style="list-style-type: none"> • Figuren, tabellen en supplementary files van artikelen (ScienceDirect en arXiv) • Data uit repositories 	Data die Google online vindt en die is beschreven met het schema.org metadata standaard (gebruikt door veel maar niet alle repositories/websites)	Alle objecten die van DataCite een DOI hebben gekregen (miljoenen records)	Wetenschappelijke output, waaronder data, van 100+ databases uit 70+ landen
Voordelen	<ul style="list-style-type: none"> • Indexeert niet alleen datasets maar ook grafieken/tabellen en supplementaire files • Previews van de data files beschikbaar • Zichtbaar welke file types er in een dataset zitten • Resultaten relevant obv zoektermen • Zoektermen zie je vetgedrukt in resultatenlijst • Filter op 'Data file type' • Filter op data uit repositories of uit artikelen 	<ul style="list-style-type: none"> • Indexeert niet alleen datasets maar ook grafieken/tabellen en supplementaire files • Geeft link naar 'Related Article' 	<ul style="list-style-type: none"> • Geeft licenties weer in resultatenlijst (indien bekend) • Filter op 'Resource type' • Genereert automatisch referenties voor de data 	<ul style="list-style-type: none"> • Uitgebreide zoekinterface met veel filters, zoals bestandstype, land, periode. • Bevat naast de tab 'Data/Software' een tab 'Multimedia'; soms zijn dit ook onderzoeksdata. • Verdeelt zoekresultaten in Topics; klikken op een Topic geeft subtopics weer, etc. • Je kunt emailalerts voor zoekopdrachten instellen. • Zoektermen zie je vetgedrukt in resultatenlijst • Zoektermen worden vertaald van én naar 10 andere talen

Zoektips:

- Gebruik “...”
- Maak gebruik van de zoek
- Ben je bewust van dubbele resultaten (vooral vanuit Figshare)
- DataCite en WWW: bekijk de zoekhulp online voor advanced search syntax. Bijvoorbeeld het gebruik van * om woordvormen mee te nemen (robot* geeft robots, robotic, etc.)

Discussie opdrachten

2. Hoe beviel het zoeken naar data? Wat vinden jullie van de data-zoek-mogelijkheden?

Veel onderzoeksdata beschikbaar, maar:

- Data zijn beschikbaar op veel verschillende manieren.
- Metadata is vaak summier, waardoor vindbaarheid wordt beperkt.

Dus:

- De juiste data vinden is lastig.
- Het artikel is vaak nog de beste manier om beschikbare data te vinden.

Bedankt!

