

COM POT LA GESTIÓ DE LES DADES DE RECERCA MILLORAR L'ACCÉS AL CONEIXEMENT PER PART DE LA CIUTADANIA?



Photo by Bridget Coila

Clara Llebot Lorente



Oregon State
University

J. Valenciana de Documentació
14, Febrer 2020

Bon dia!

Hi ha dades a la vostra vida?

Com us afecten les “dades” en el vostre dia a dia?



Oregon State
University

Dades per conèixer-vos millor



Dades científiques

- Medicina
- Canvi climàtic i l'efecte en polítiques
- Agricultura, les plantes i animals que mengem
- Tecnologia
- ...

Dades públiques i de governs

- Estadístiques de població
- Estadístiques econòmiques
- Enquestes
- Meteorologia
- Mapes
- ...

En aquesta xerrada vull explicar com, no només les dades afecten les vostres vides, sinó també que, com a professionals de la informació, vosaltres podeu també podeu afectar les dades, i les relacions de les persones amb les dades.

Parlarem de Big Data?



By Jazella



By Eric Jones

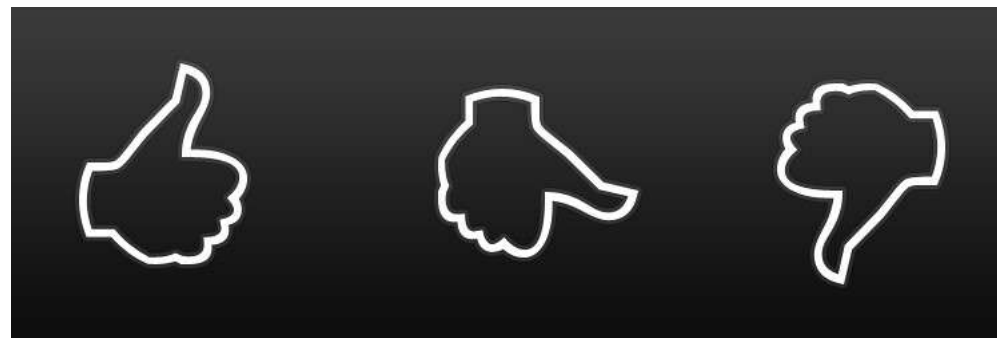
Què són “dades”?

Són els FETS



By Hannah O’Leary

No la interpretació dels fets



By Sean MacEntee, CC-BY

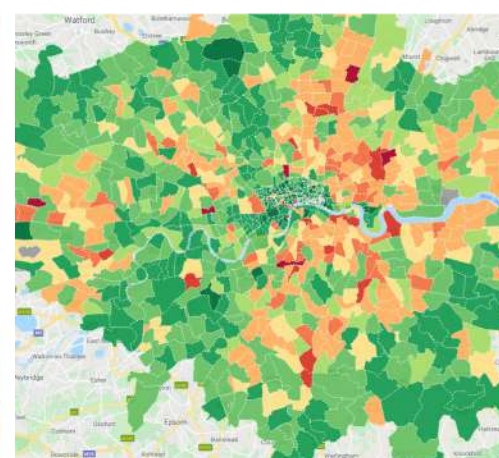
Números, mostres, imatges, mapes, videos, textos, entrevistes, models...



Photo by Chris Khamken



Photo by jsawkins



Plumplot.co.uk



Photo by Oregon State University

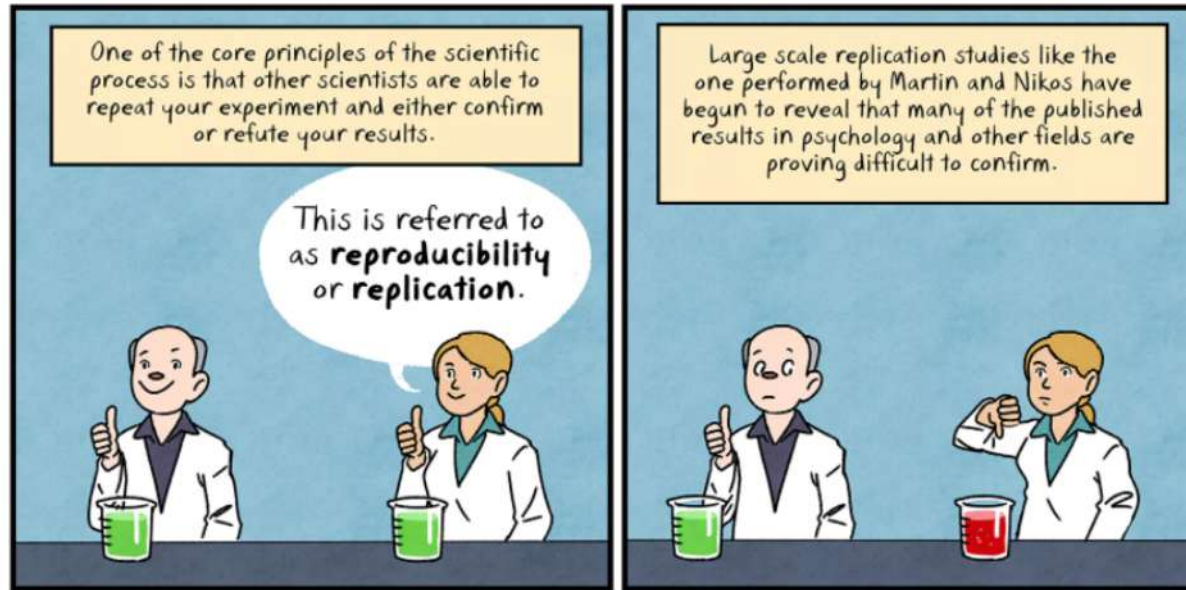
Brian Nosek

- Dr Nosek + 269 co-autors
- Repetició de la feina en 98 articles de psicologia
- Només van poder replicar 39 dels 98 estudis.
- Crisi de Reproductibilitat

Font: [doi:10.1038/nature.2015.18248](https://doi.org/10.1038/nature.2015.18248)



Reproductibilitat i dades obertes



Maki Naro
thenib.com/repeat-after-me

Reproductibilitat: "Obtenir resultats consistents fent servir les mateixes dades; passos computacionals, mètodes, i codi; i condicions d'anàlisi". [National Academies of Sciences, Engineering, and Medicine](#)

Reproductibilitat i dades obertes

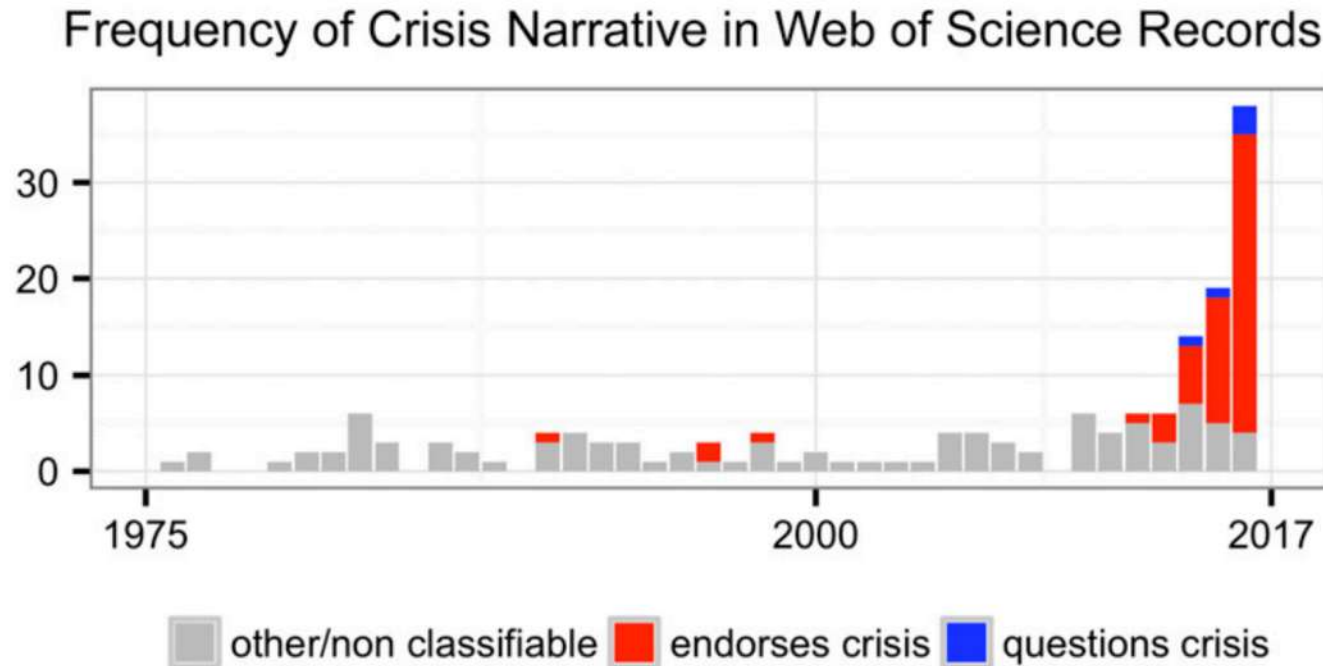


Fig. 1. Number of Web of Science records that in the title, abstract, or keywords contain one of the following phrases: "reproducibility crisis," "scientific crisis," "science in crisis," "crisis in science," "replication crisis," "replicability crisis." Records were classified by the author according to whether, based on title and abstracts, they implicitly or explicitly endorsed the crisis narrative described in the text (red), or alternatively questioned the existence of such a crisis (blue), or discussed "scientific crises" of other kinds or could not be classified due to insufficient information (gray). The complete dataset, which includes all titles and

(Fanelli, 2017)

<https://doi.org/10.1073/pnas.1708272114>

Per què hi ha manca de reproductibilitat?

Errors humans

Ruth Alexander, 2013
<https://www.bbc.com/news/magazine-22223190>

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

By Ruth Alexander
BBC News

🕒 20 April 2013

f 🗨️ 🐦 ✉️ Share

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, are presenting a research paper called Growth in a Time of Debt.



Thomas Herndon

Per què hi ha manca de reproductibilitat?

Fabricació o alteració de dades i plagi

El cas de Diederik Stapel

Nature **479**, 15 (2011) | doi:10.1038/479015a

Report finds massive fraud at Dutch universities

Investigation claims dozens of social-psychology papers contain faked data.

Ewen Callaway

When colleagues called the work of Dutch psychologist Diederik Stapel too good to be true, they meant it as a compliment. But a preliminary investigative report (go.nature.com/tqmp5c) released on 31 October gives literal meaning to the phrase, detailing years of data manipulation and blatant fabrication by the prominent Tilburg University researcher.

"We have some 30 papers in peer-reviewed journals where we are actually sure that they are fake, and there are more to come," says Pim Levelt, chair of



Dutch psychologist
Diederik Stapel.

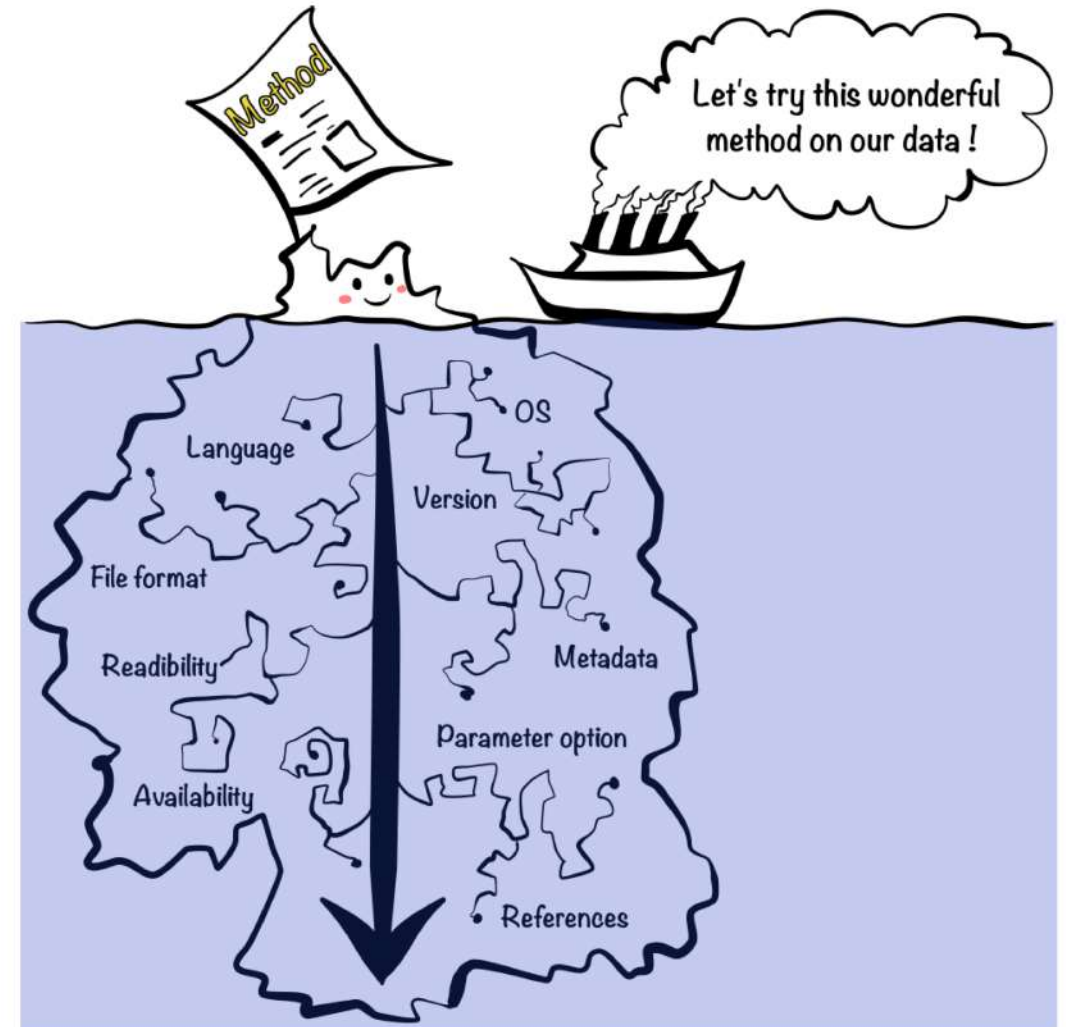
Per què hi ha manca de reproductibilitat?

- Biaix en el sistema de publicacions científiques
 - Els investigadors estan pressionats a publicar molt, i en revistes bones.
 - Les revistes bones tenen moltes submissions, així que publiquen només les més emocionants
 - Els estudis avorrits no es publiquen
 - Els investigadors poden caure en la temptació (conscient o no) de fer més i més anàlisis fins a trobar el resultat desitjat

Què hi podem fer?

- Revistes que prioritzin la metodologia als resultats PLOS ONE
- Pre-registrar investigacions - OSF, clinicaltrials.org
- Posar totes les dades i metodologies a l'abast de tothom perquè puguin ser comprovades

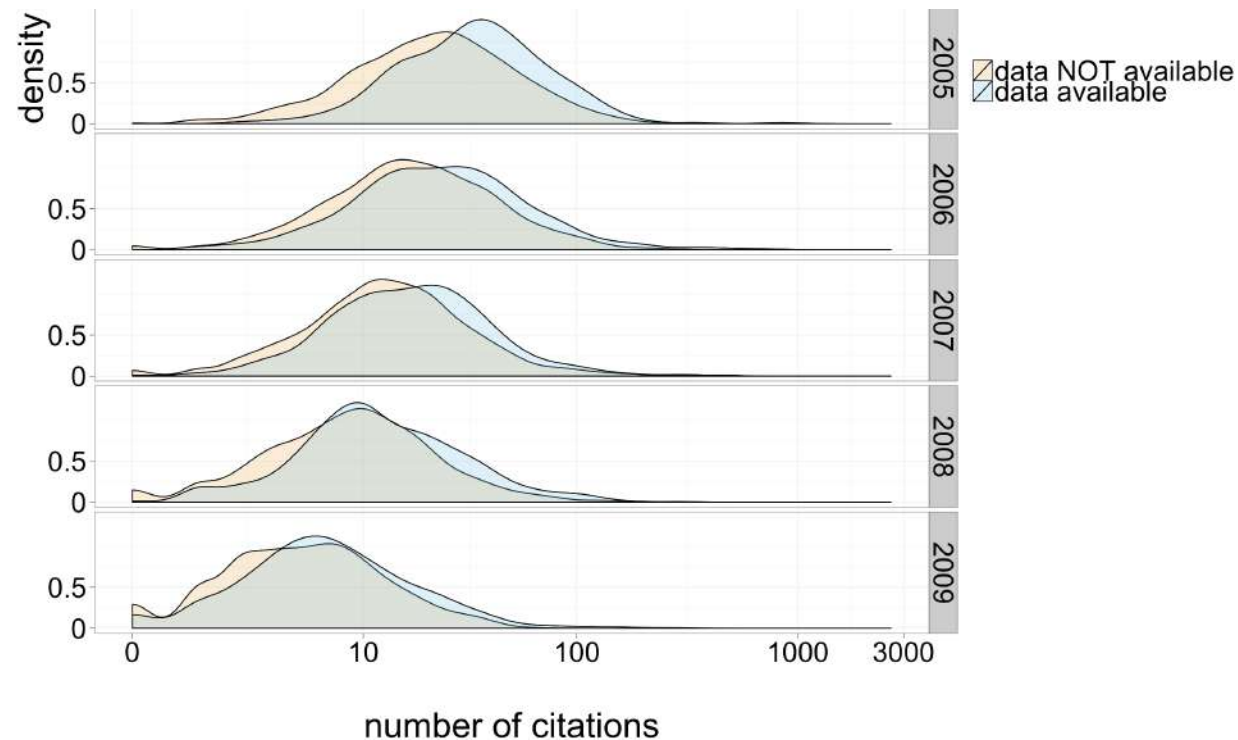
DADES OBERTES



(Kim et al 2017) <https://doi.org/10.1101/143503>

Per què dades obertes?

Perquè és millor per l'investigador



Piwowar & Vision, 2013 peerj.com/articles/175/

Perquè és més just



Per què dades obertes?

- Perquè la ciència avanci més ràpid
- Exemple de com compartir dades pot ajudar a persones amb malalties rares

Nature 534, 435–436 (23 June 2016) doi:10.1038/534435b



The screenshot shows the Nature journal website interface. At the top, the 'nature' logo is displayed in white on a dark red background, with the tagline 'International weekly journal of science' below it. A navigation bar contains links: Home, News & Comment, Research, Careers & Jobs, Current Issue, Archive, Audio & Video, and For Authors. Below this, a secondary navigation bar highlights 'Archive', 'Volume 534', 'Issue 7608', 'Editorial', and 'Article'. The main content area is titled 'NATURE | EDITORIAL' and features the article 'The ups and downs of data sharing in science'. The article's subtitle reads: 'Pooling clinical details helps doctors to diagnose rare diseases — but more sharing is needed.' The publication date is '21 June 2016'. Below the title, there are two buttons: 'PDF' and 'Rights & Permissions'. The article text begins with: 'When doctors in Ottawa saw a child with an unusual developmental disorder last year, they were stumped. Their patient had an abnormally small head and face and had been slow to develop. They sequenced the child's genome hoping to find a genetic explanation, but came up with too many possible candidate genes to pinpoint a likely culprit. This still happens a lot in'.

Per què dades obertes?

- Transparència
- Innovació
- Participació ciutadana



<https://opendataimpactmap.org/>

Per què dades obertes?

Innovació, participació ciutadana, transparència

Ganha:
Farmerline.
Dóna accés a
dades
governamentals
sobre mercats,
meteorologia,
etc. per pagesos.
Accés per
telèfon.

Rússia: pàgina
web RosPil que
publica
contractes del
govern per
evitar corrupció

USA: més de 500
negocis que
depenen en
l'accés a dades
obertes del
govern



Ayuntamiento de
Alicante

Portal integral de información analítica y datos abiertos



Web Datos Abiertos Diputación
Alicante



AJUNTAMENT
DE VALÈNCIA



Universitat d'Alacant
Universidad de Alicante



Open data portal

datos.gob.es
reutiliza la información pública



AJUNTAMENT DE
LA POBLA DE VALLBONA

DADES OBERTES

Portal de **Transparència**
i Dades Obertes



GENERALITAT
VALENCIANA

dades Obertes
PORTAL DE TRANSPARENCIA DE LA GENERALITAT VALENCIANA



Xàtiva
Ajuntament



DADES OBERTES

Totes les dades de l'Ajuntament són teues.



DIPUTACIÓ
D
CASTELLÓ



Dades Obertes



PORTAL OPEN DATA
AJUNTAMENT DE TORRENT

Dades obertes $\neq$ Dades gratuïtes!!!

Dades obertes = Dades gratuïtes + permisos

- Retenció
- Reutilitzar
- Revisar
- Barrejar
- Redistribuir

Què hi podem fer des de les biblioteques?



(From a talk given by Josh Young at Ocean Sciences, 2016)

Dades obertes ≠ Dades gratuïtes!!!

Dades obertes = Dades gratuïtes+permisos

Qualitat

Documentació

Identificadors

Formats

Permanència

Ètica

Procedència

Molt més...

Com han de ser les dades obertes?

Com han de ser les dades obertes?

F_{indable}



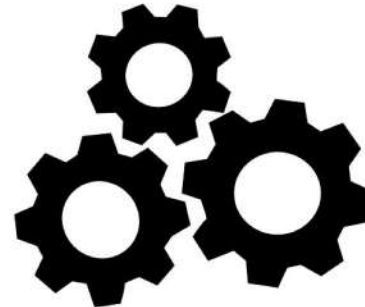
- Metadades trobables
- Identificadors Digitals Permanents

A_{ccessible}



- Protocols oberts
- Format obert
- Software obert

I_{nteroperable}



- Metadades estàndards
- Formats estàndard

R_{eusable}



- Llicències
- Procedència

(Wilkinson et al, 2016)

Com han de ser les dades obertes?

CARE Principles for Indigenous Data Governance



<https://www.gida-global.org/care>

- “Dades tan obertes com sigui possible, tan tancades com sigui necessari”
- Disseny de la recol·lecció de dades per evitar biaixos
- Protecció de dades personals
- Com finançar l’estructura necessària? Preservació de dades a llarg termini.

Boeckhout et al 2018
[10.1038/s41431-018-0160-0](https://doi.org/10.1038/s41431-018-0160-0)

A les biblioteques hi tenim molt a dir!

- Diferents tipus de biblioteques
 - Públiques
 - Escoles
 - Acadèmiques
- Diferents models de Serveis de Dades de Recerca (SDR)
 - Tothom fa SDR
 - Una persona fa SDR
 - Un equip fa SDR



Exemple: docència i divulgació

Information and Data literacy "Alfabetització informacional i de dades"



Quina és la font d'aquestes dades? Me'n refio?
Qui les ha generat? Qui les ha pagat?
Quins mètodes s'han fet servir?
Com estan representades? Com s'entén un gràfic? Hi ha biaixos en la visualització?
Falten dades?

Dades en les mans de ciutadans



“We have analog communities in a digital world” **“Tenim comunitats analògiques en un món digital”** - Valencia Gunder, RDAP 2019

Activista social
Ciutat de Miami

<https://miamigirls.org/miamigirls/valencia-gunder/>

Dades en les mans de ciutadans

Projecte per demostrar la gentrificació que es produïa a la Ciutat de Miami

- Organització dels primers ajuts als barris més pobres de Miami.
- Recollien dades de la feina que feien.
- Cada dia analitzaven les dades del que havien fet, per detectar els llocs on feia falta més ajuda
- Van alimentar 23000 persones

Karen Hao (2019) *What one woman's quest to save her neighbors reveals about climate resilience.* MIT Technology Review



Huracà Irma, 2017 09 08

Example: Oregon State University



Universitat “pública”
~32000 estudiants
(UPV 30317, UA 20424, UV 58390)



Forta component d’investigació
Oceanografia, ciències agràries,
enginyeria

Research Data Services



Clara Llebot

Data Management
Specialist:

50% Gestió de dades
15% Docència
15% Investigació
10% Servei
10% Referència
bibliotecària

Perfil:
Formació d'investigadora.
Experiència amb dades

Exemple: docència

- Generar dades
- Ús i anàlisi de dades



- Tallers d'eines
 - Control de versions
 - Llenguatges de programació
 - Bones pràctiques en fulls de càlcul
 - Programes gestors de referències (Zotero, EndNote, Mendeley...)
 - The Carpentries
- Curs de dos crèdits en gestió de dades
- Classes com a lectora convidada

Exemple: servei de revisió de dades publicades al repositori

1. Dipòsit



- ✓ Descripció
- ✓ Arxius
- ✓ Acceptació del contracte

2. Revisió



- ✓ Reusabilitat

3. DOI i aprovació



Exemple: servei de revisió de dades publicades al repositori

- Dades ordenades
- Dades en formats oberts

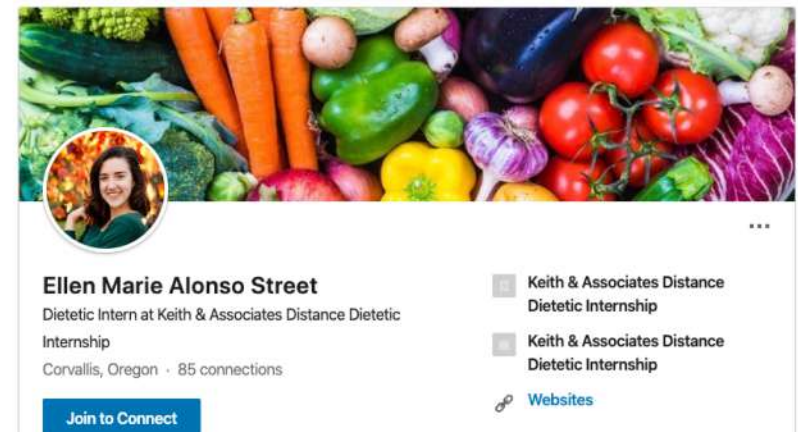


- Documentació i metadades



Readme.txt

- Llicències
- Digital Object Identifier: identificador permanent
- Com citar les dades
- Diferència entre dades-article
- Concepte de preservació a llarg termini

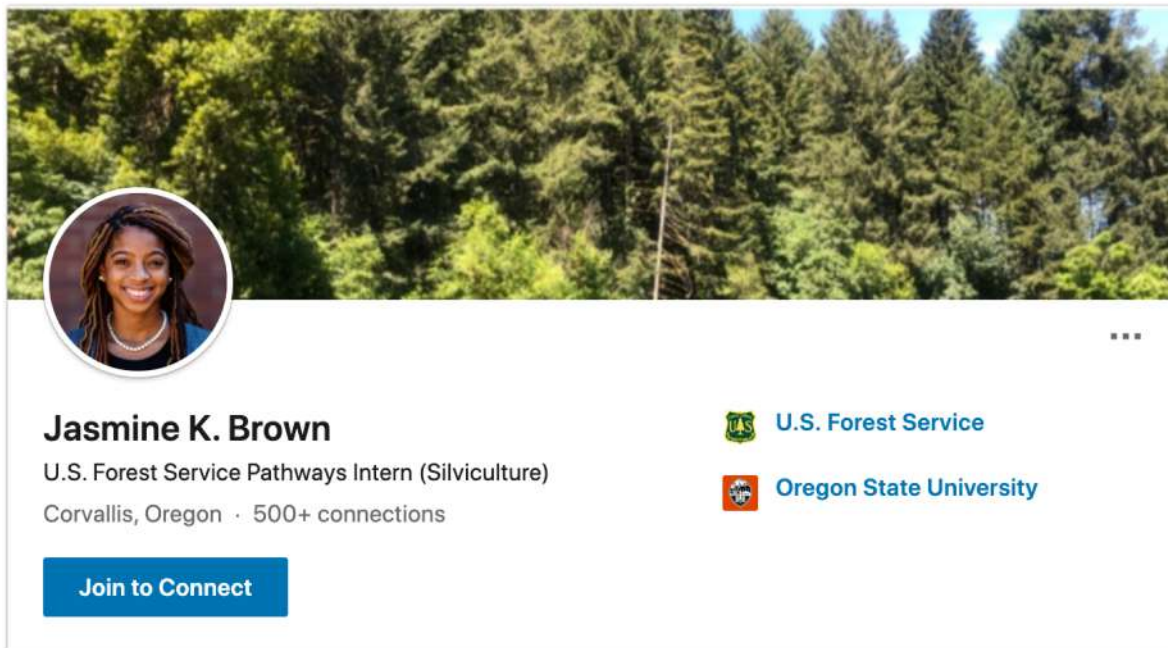


Serveis a la OSU: Plans de Gestió de Dades

- Requeriment per demanar projectes de recerca
- Revisió de DMP:
 - De tots els colleges de la universitat
 - En especial, quan tinc bona relació amb els Research Program Administrators.
- DMPTool <https://dmptool.org/>
- Revisar les normes de l'agència finançadora
- En paper, o en forma de consulta.

Serveis a la OSU: consultes

- Consultes de qualsevol tema relacionat amb la gestió de dades
- M'han preguntat de tot!
- Qui és el propietari de les dades?



David Noakes

View

Publications

Contact



David Noakes

Professor & Senior Scientist, Oregon Hatchery Research Center

david.noakes@oregonstate.edu

Office: 541-737-1953

Social behavior, Early development and behavior, Ethology, Ecology and life history, Evolution and polymorphism

Serveis a la OSU: Política de dades

University of Wisconsin Milwaukee
UWM Digital Commons

UWM Libraries Other Staff Publications

UWM Libraries

2015

Do You Have an Institutional Data Policy? A Review of the Current Landscape of Library Data Services and Institutional Data Policies

Kristin Briney

University of Wisconsin-Milwaukee, kabriney@gmail.com

Abigail Gobin

University of Illinois at Chicago, agoben@uic.edu

Lisa Zilinski

Carnegie Mellon University, ldz@andrew.cmu.edu

<http://doi.org/10.7710/2162-3309.1232>

- Què són dades de recerca
- Qui és el propietari de les dades?
- Responsable de gestionar les dades
- Drets i responsabilitats dels investigadors
- Drets i responsabilitats de les institucions
- Retenció de dades
- Accés a les dades
- Transferència de dades

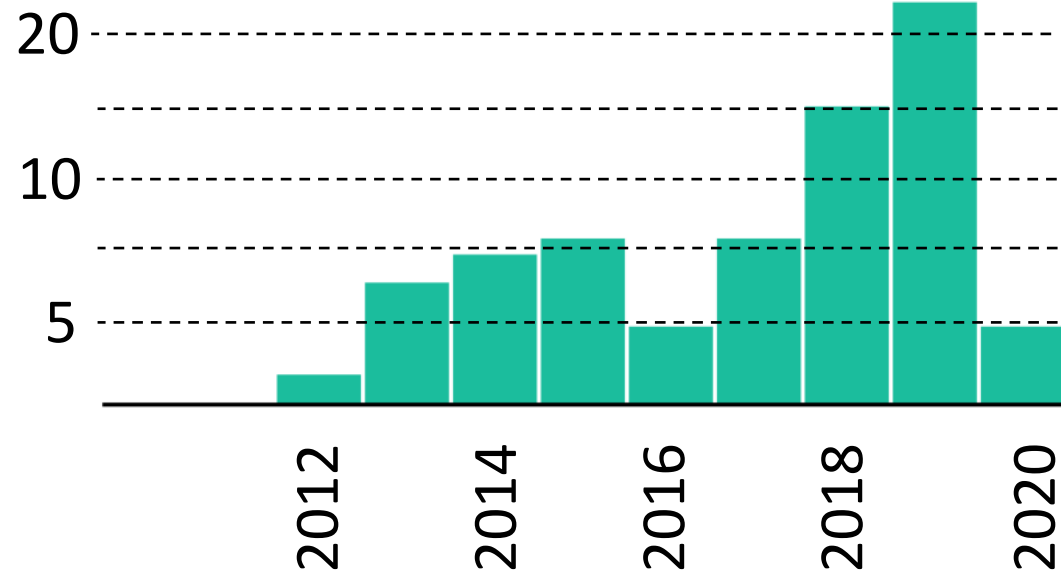
Números a la OSU

En els últims 3 anys:

40 revisions de Plans de Gestió de Dades

53 consultes

~100 conjunts de dades al repositori

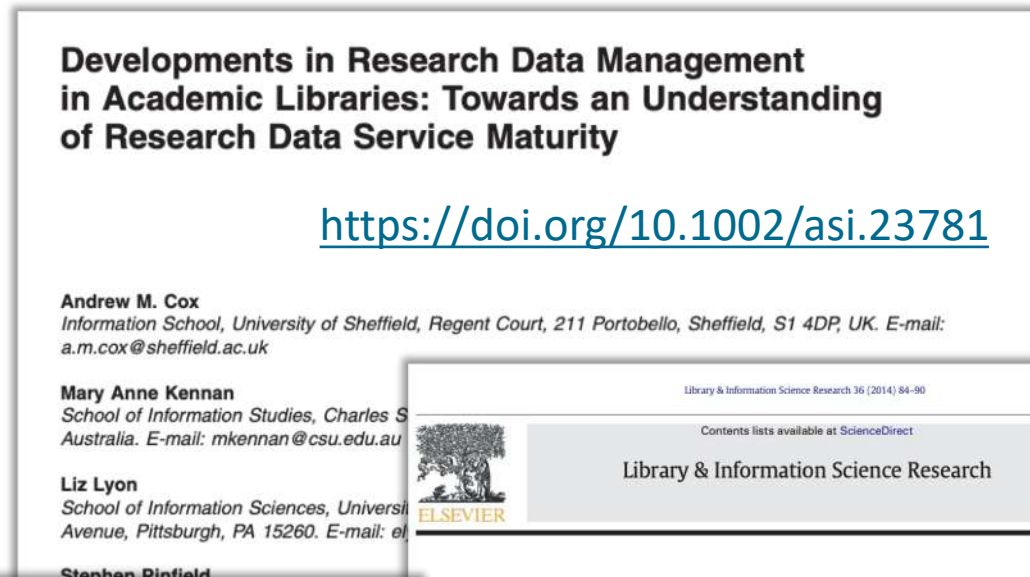


Serveis de Suport a la Gestió de Dades Research Data Services (RDS)

Biblioteques
Centres de suport a la recerca
Informàtica



<http://www.dcc.ac.uk/resources/how-guides/RISE#4>



<https://doi.org/10.1002/asi.23781>



<https://doi.org/10.29085/9781856048750.011>



<https://doi.org/10.7191/jeslib.2015.1085>

Clara Llebot Lorente | Data Management Specialist
clara.llebot@oregonstate.edu

Podeu accedir a aquesta xerrada anant a

<https://beav.es/4oT>

This presentation is licensed under a CC-BY license.