

# Kako iskoristiti otvorenu znanost u svom istraživanju

Dr. Antica Čulina

@antica\_c

## TEME:

1. Kako/zašto sam počela s otvorenom znanosti
2. Otvorena znanost
3. Korištenje otvorenih podataka
4. FAIR podatci
5. Kompjuterski programi & kodovi
6. Zašto bi nas trebalo biti briga?
7. Kako se uključiti?

I) KAKO/ZAŠTO SAM POČELA S OTVORENOM  
ZNANOSTI

## Trading up: the fitness consequences of divorce in monogamous birds

Antica Culina\*, Reinder Radersma and Ben C. Sheldon

*Edward Grey Institute, Department of Zoology, University of Oxford, Oxford, OX1 3PS, U.K.*

**Meta-analysis:** Kombinira rezultate svih studija o određenom istraživačkom pitanju kako bi se procijenila veličina i smjer efekta

**Systematic review:** Proces kojim se indentificira sve studije koje su rađene na određenoj temi

### Systematic review

- Identify all literature on a particular topic
- Evaluate risk of bias
- Abstract relevant data

1. Vrijednosti se ne podudaraju
2. Nedostaju informacije
3. Nejasne metode

.....  
↓  
**PROVJERI PODATKE**  
↓





# Meta-analysis meets Open Science

Open Science is making more and more datasets available to incorporate into meta-analysis, either to verify (or supplement) the results of primary studies, or to use them on their own, as data-points in the analysis.

Our project seeks to:

# ŠTO MI JE POMOGLO NA POČETKU

- Konferencije, seminari.... vezani za otvorenu znanost
- Twitter
- Puno čitanja 😊

# PODATCI 'NESTAJU' + OSTALI PROBLEMI

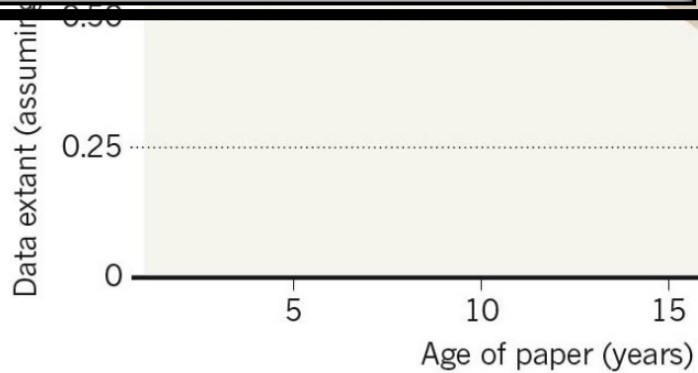
## MISSING DATA



**A survey of publication bias within evolutionary ecology**

**Dissemination biases in ecology: effect sizes matter more than quality**

E. Kathryn Barto and Matthias C. Rillig



RESEARCH ARTICLE

**Questionable research practices in ecology and evolution**


Vines et al. 2014, Curr Biol



## 2) ŠTO JE OTVORENA ZNANOST

Comment | [Open Access](#)

# When will 'open science' become simply 'science'?

Mick Watson 

*Genome Biology* 2015 **16**:101

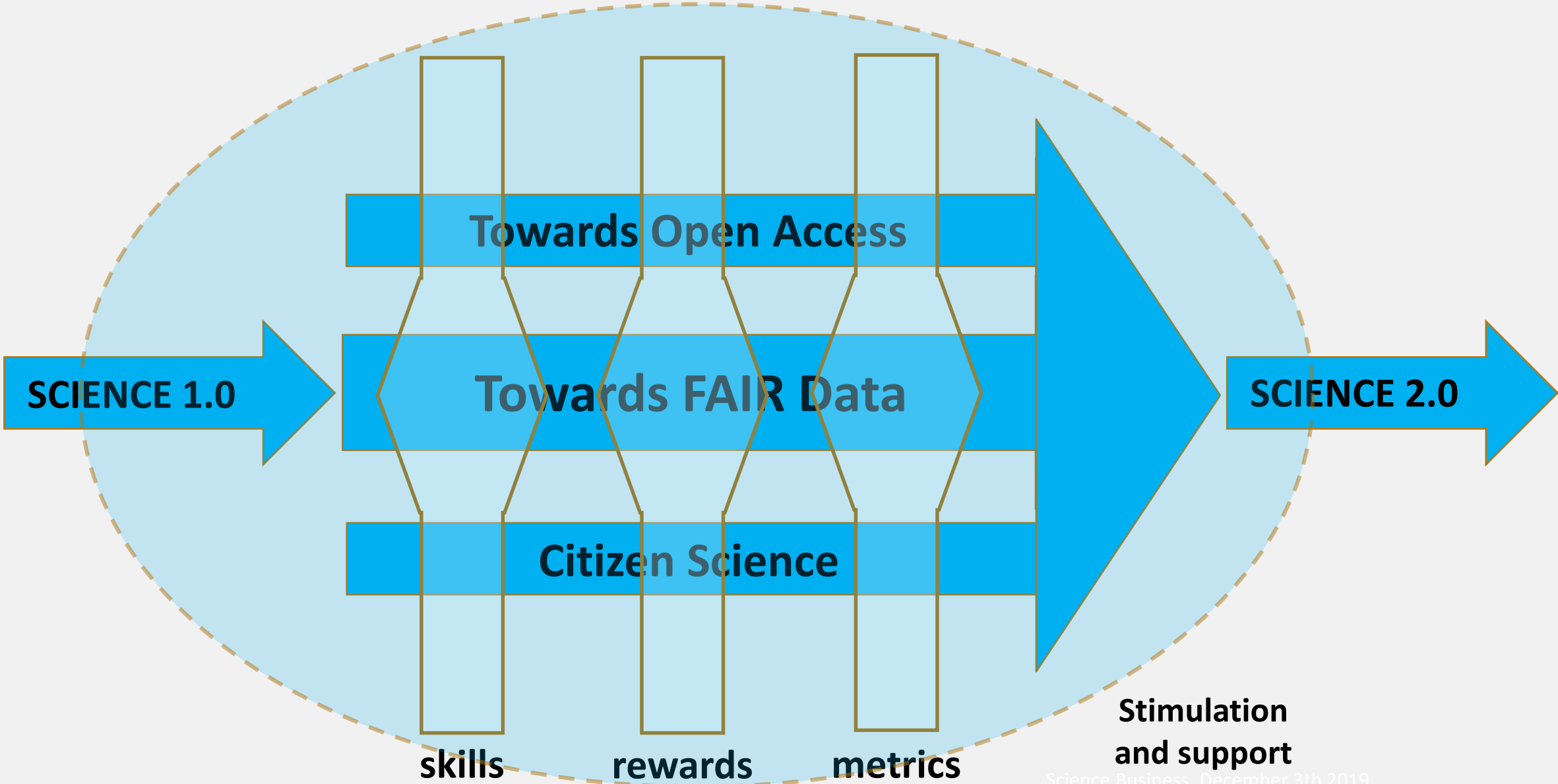
<https://doi.org/10.1186/s13059-015-0669-2> | © Watson; licensee BioMed Central. 2015

**Published:** 19 May 2015

## Abstract

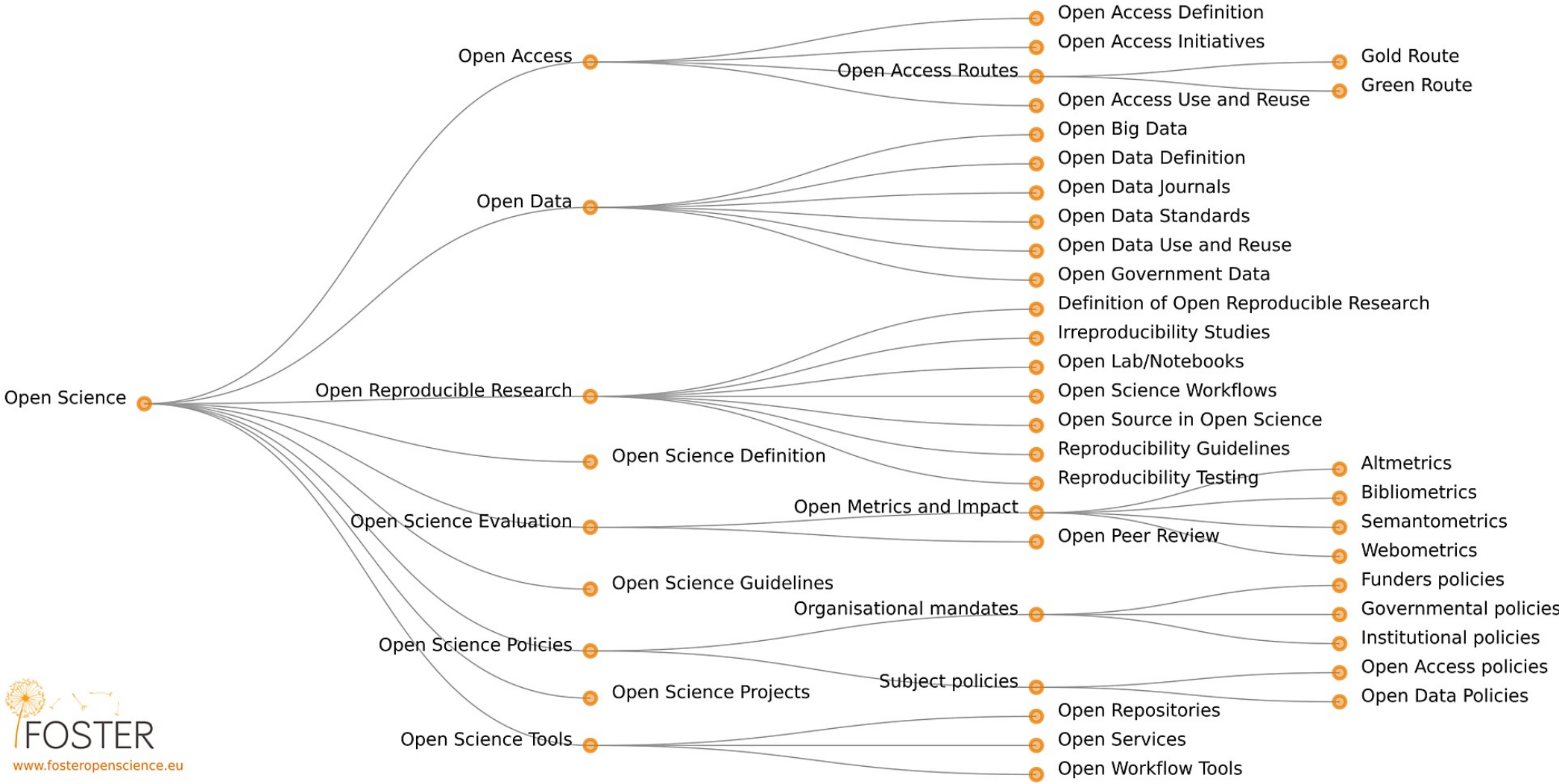
Open science describes the practice of carrying out scientific research in a completely transparent manner, and making the results of that research available to everyone. Isn't that just 'science'?

# OPEN SCIENCE

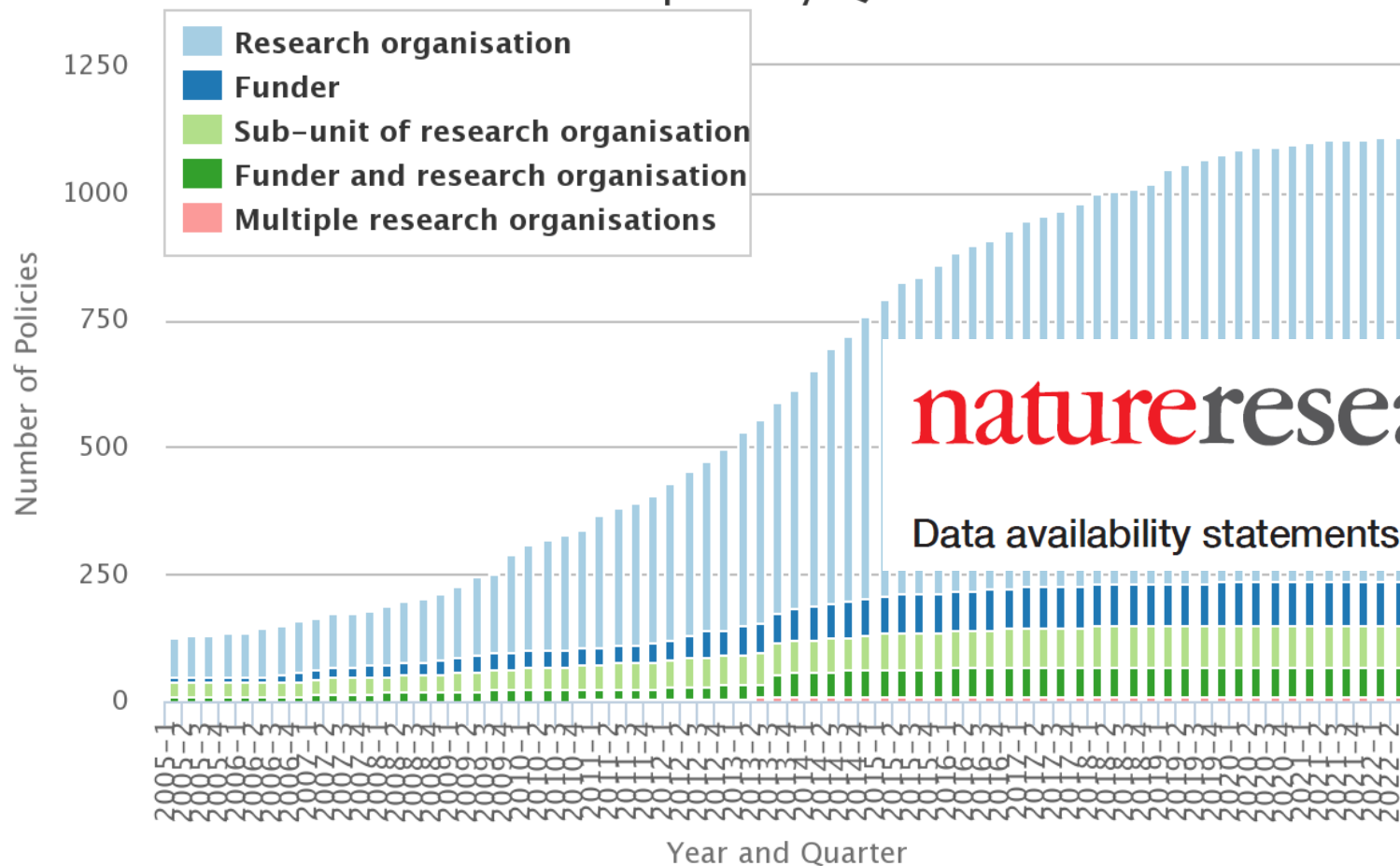


Science Business, December 3th 2019, Brussels

# Open Science Taxonomy



## Policies Adopted by Quarter



**nature**research

Data availability statements and data citations policy: guidance for authors



Ekolozi skupljaju informacije I daju rješenja koja su:

- specifična za određeni problem, mjesto, period
- prediktivno, preskriptivno i skalabilno

Otvorena znanost



Nove tehnologije za prikupljanje podataka



Integrativni pristupi



# PUT NIJE LAK

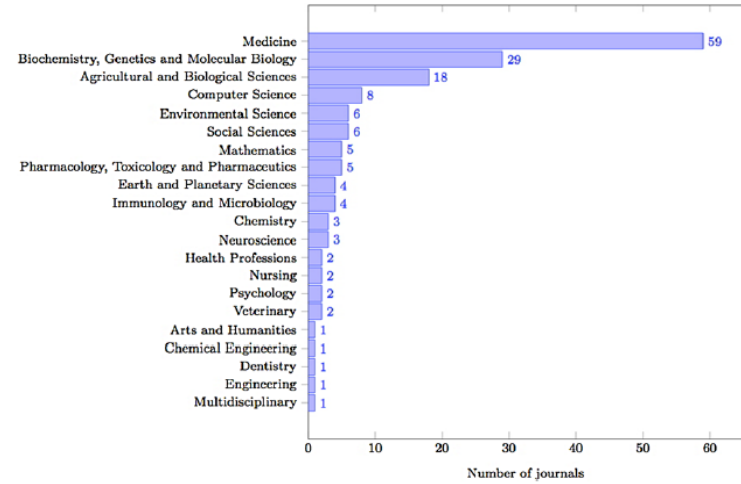
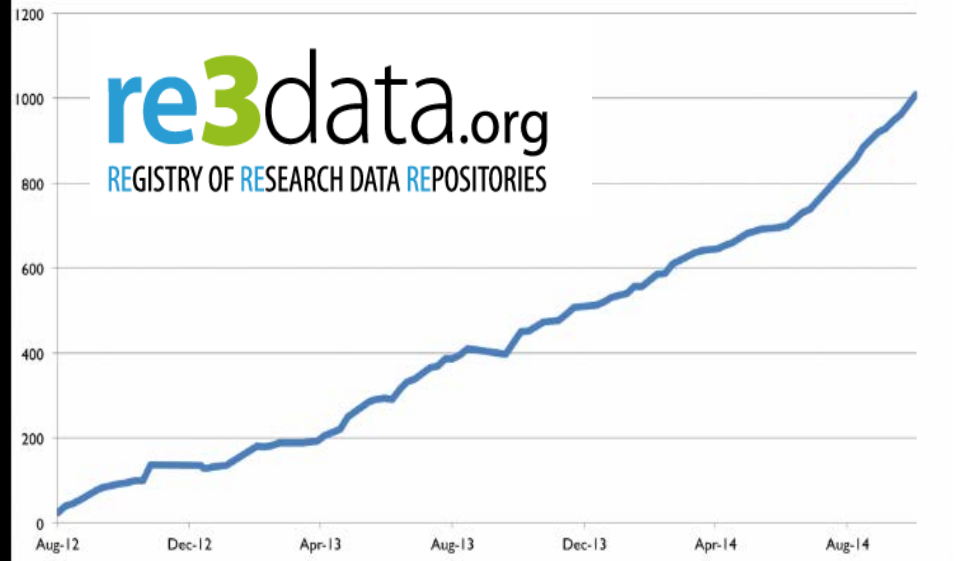
## The Internet's Own Boy: The Story of Aaron Swartz (2014)



### 3) KORIŠTENJE OTVORENIH PODATAKA

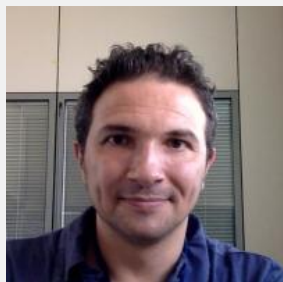


Growth of re3data.org



JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY, \*\*(\*\*):\*\*--\*\*, 2015

# OMOGUĆAVANJE KORIŠTENJA OTVORENIH PODATAKA U EKOLOGIJI



**Paolo Manghi**



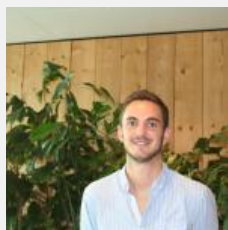
**Miriam Baglioni**



**Saskia  
Woutersen-Windhower**



**Marcel Visser**



**Tom Crowther**



Culina et al. 2018a, Nature Eco Evo

Podatci su obično u:

- Repozitoriju
- Data journal
- Supplement
- Osobne web stranice





# ResourceCatalogue

## Welcome to the EcoEvo Catalogue!

It provides a list of best online places (sources) that can be used to conduct a search for ecological and evolutionary datasets.

The catalogue is an evolving list of these sources, and can be amended by the community members. As the amount of EcoEvo datasets that are freely available online is rapidly increasing, we need a comprehensive list of places where these datasets can be searched for in one interface, speeding up the search process and making it more comprehensive..

Each data source is described with features characterizing the data source itself (such as name and organization) and with features of EcoEvo datasets that the source hosts or collects information on (e.g. available metadata formats, content reuse policies).

This catalogue accompanies the paper Culina et al. 2017

## Items Search

[See All](#)

## EcoEvo Catalogue statistics

**23**

items

**1**

organisation

**4**

types

## What data I need?

Are there community specific primary data sources?

Identify using ROAR (ROAR <http://roar.eprints.org/cgi/search/advanced>) or Re3Data.org <http://www.re3data.org/>

- **Yes:** Search the data within the community specific data sources
- **No:** Use *data discovery sources* to search for data across many primary data sources <https://ckan-ecoevo.d4science.org/>



## I have obtained the datasets

Am I allowed do use data (legal considerations)?

1) Copyright and database rights attached

- *CC-BY licence = share and adapt the data, attribution (e.g. citation) required.*
- *CC0 = data in the public domain (no requirements attached)*
- *no permission has been granted = ask permission of reuse from the data owner*

2) Research data protected by terms of use = comply with these

Avoid misinterpretation and biases:

- contact the authors, think about potential spatial, temporal, taxonomic etc biases

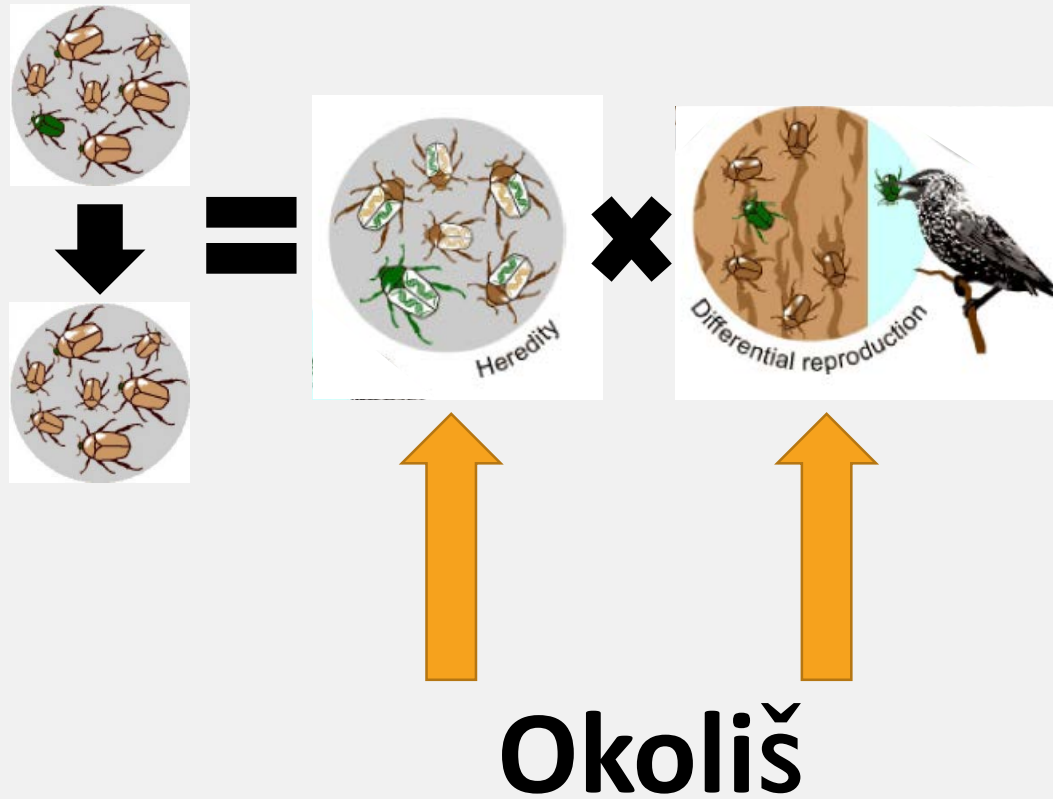
Credit the authors

- many datasets have DOI and are accompanied with the information on how to cite them

# A CASE STUDY

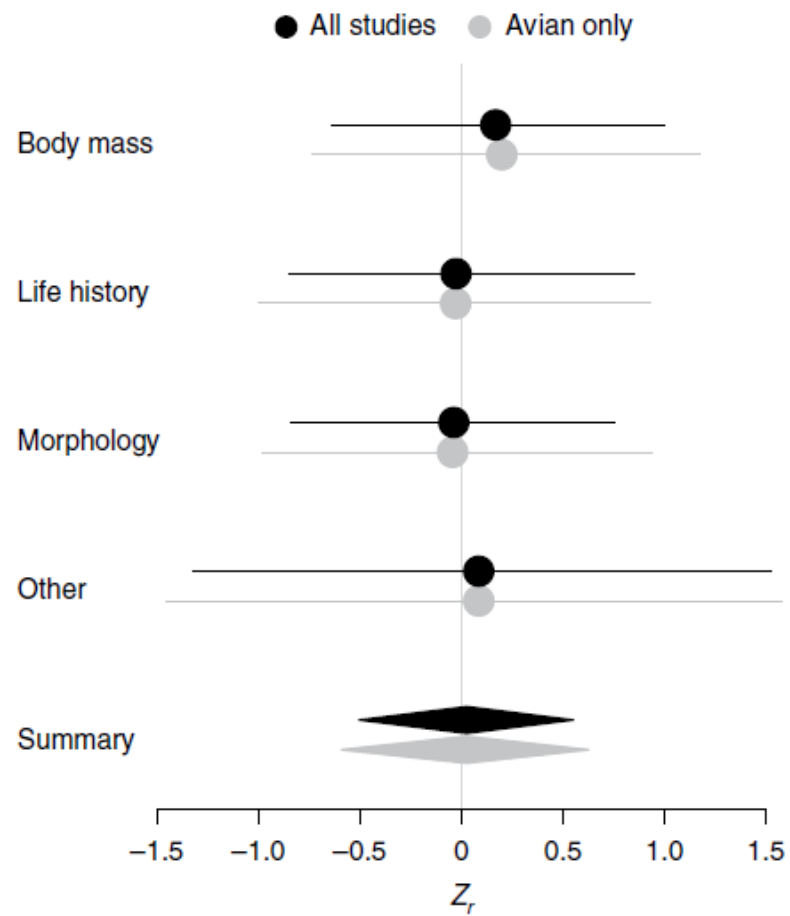


## Environmental coupling of heritability & selection



Stopa genetske promjene ovisi o heritabilnosti i selekciji

Ramakers et al. 2018, Nature Eco Evo



**Fig. 2 | Meta-analysis on the heritability-selection correlation coefficients.** Coefficients  $r$  were standardized using Fisher's  $Z$

## nature ecology & evolution

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [nature ecology & evolution](#) > [comment](#) > [article](#)

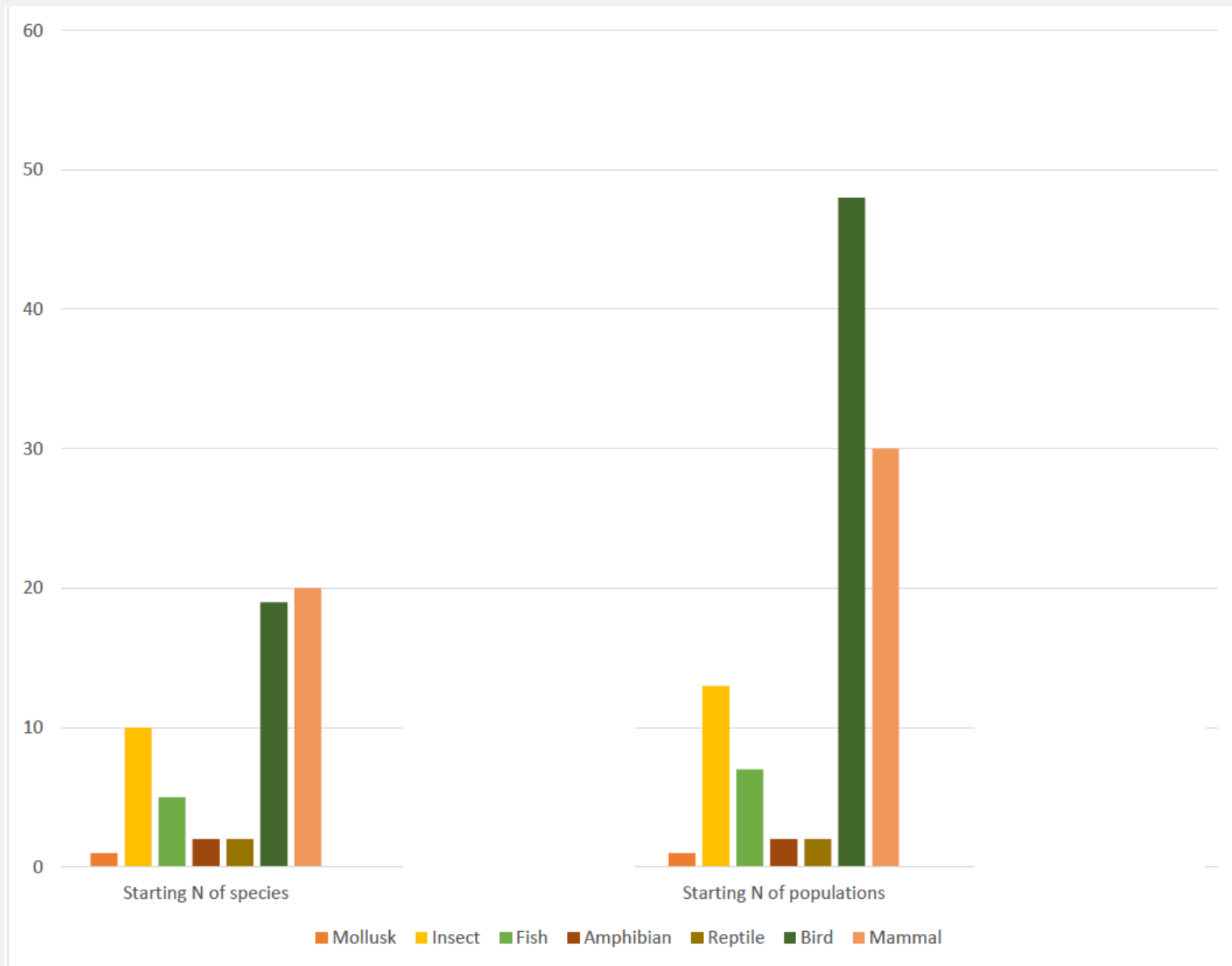
Comment | [Published: 18 June 2018](#)

### How to do meta-analysis of open datasets

[Antica Culina](#) , [Thomas W. Crowther](#), [Jip J. C. Ramakers](#), [Phillip Gienapp](#) & [Marcel E. Visser](#)

JESU LI PODATCI ISKORISTIVI?

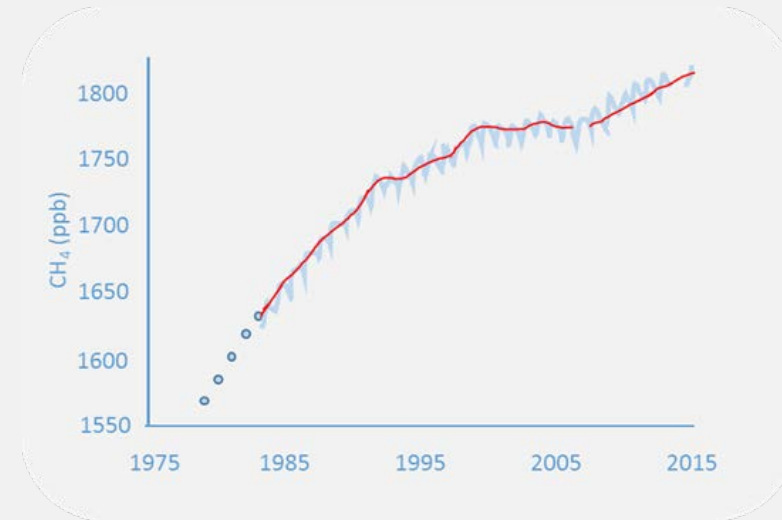
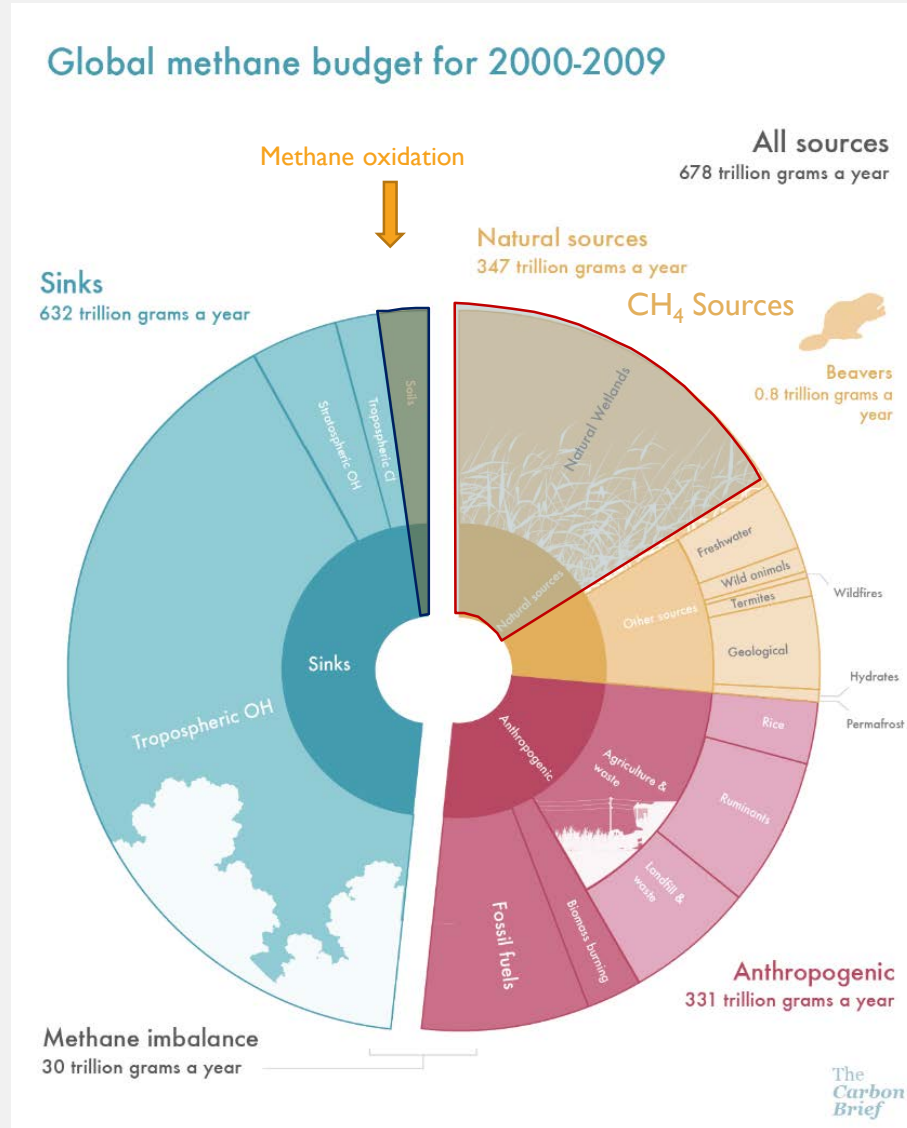




Do you want more data?

You were not even born when I started to collect the data

# Kako su zajednica metanotrofa strukturirane okolisem ?



Open metagenome data  
Community structure

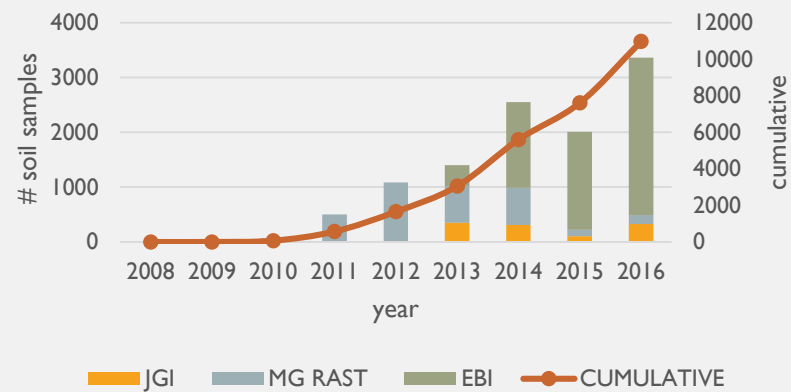


pH, Temp

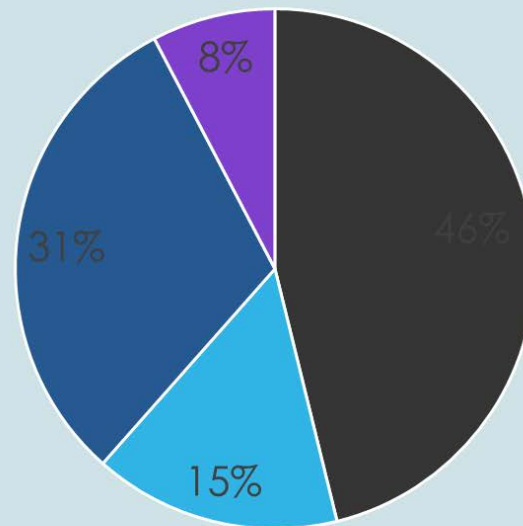
C, N



soil samples that became publicly available



no units % mg/kg NO3 mg/L



## 4) FAIR PODATCI

ID	Age	FRL	Sex
A	1	30.2	F
B	4	35.7	M
C	6	40.5	F
D	4	40.0	F
E	7	39.8	M
F	9	41.6	F
E	3	39.2	M

NOT  
FAIR

## META-DATA

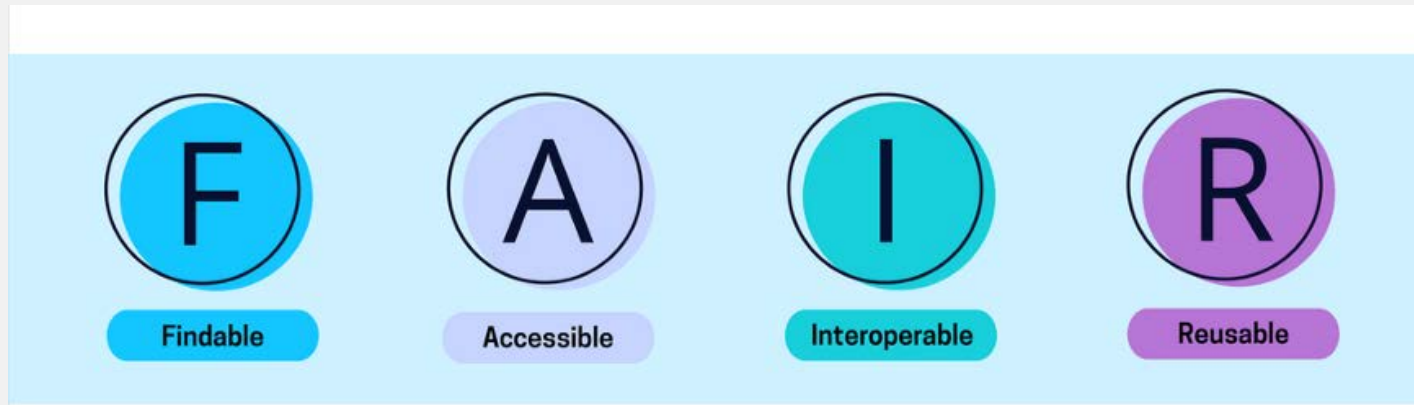
Site: Uviraljka cave  
Species: Plecotus auritus  
Date: 23/06/2020  
Capture method: Nett  
Etc

ID = Unique identification number of the animal

Age = age at capture, in years

FRL = forearm length, in mm

Sex = sex of the animal, determined by....



FAIR poboljšava sposobnost strojeva da  
pronađu, pristupe i koriste podatke

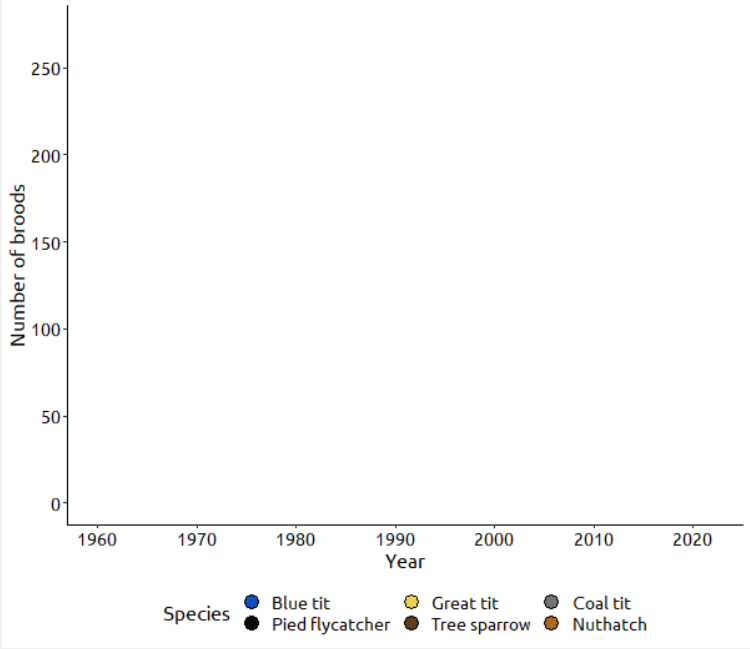
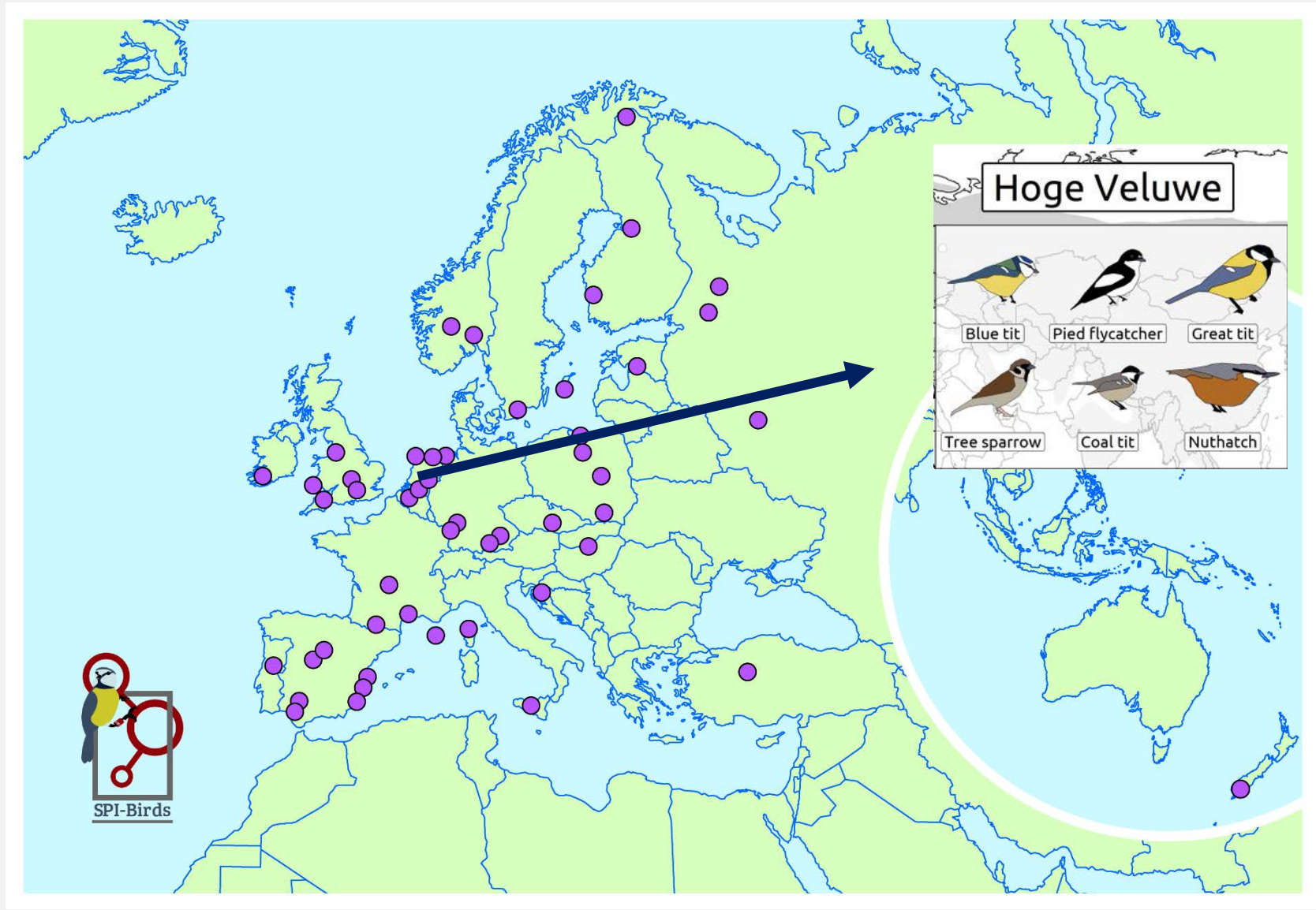
= MACHINE ACTIONABLE

# FAIR DATA PRIMJER

## SPI-Birds Network & Database

@spibirds



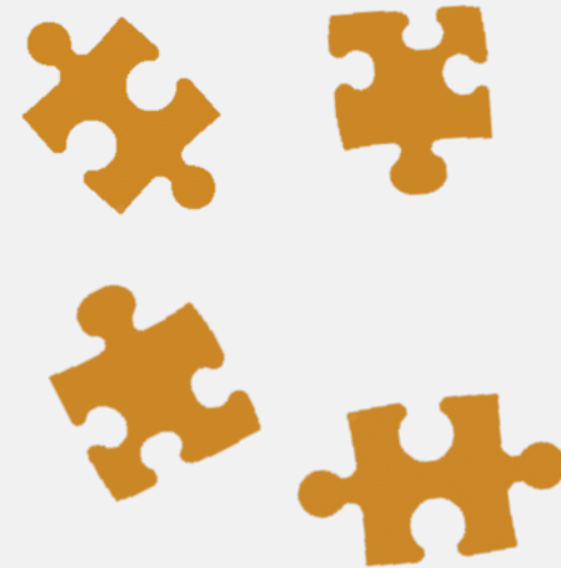




# PROBLEMI

Različite strategije upravljanja podacima

Kako su variable nazvane?



Kako su podatci spremeljni?

Kontrola kvalitete?

Kombinacija podataka je vrlo teska/spora/nemoguca

## CONSIDERATIONS

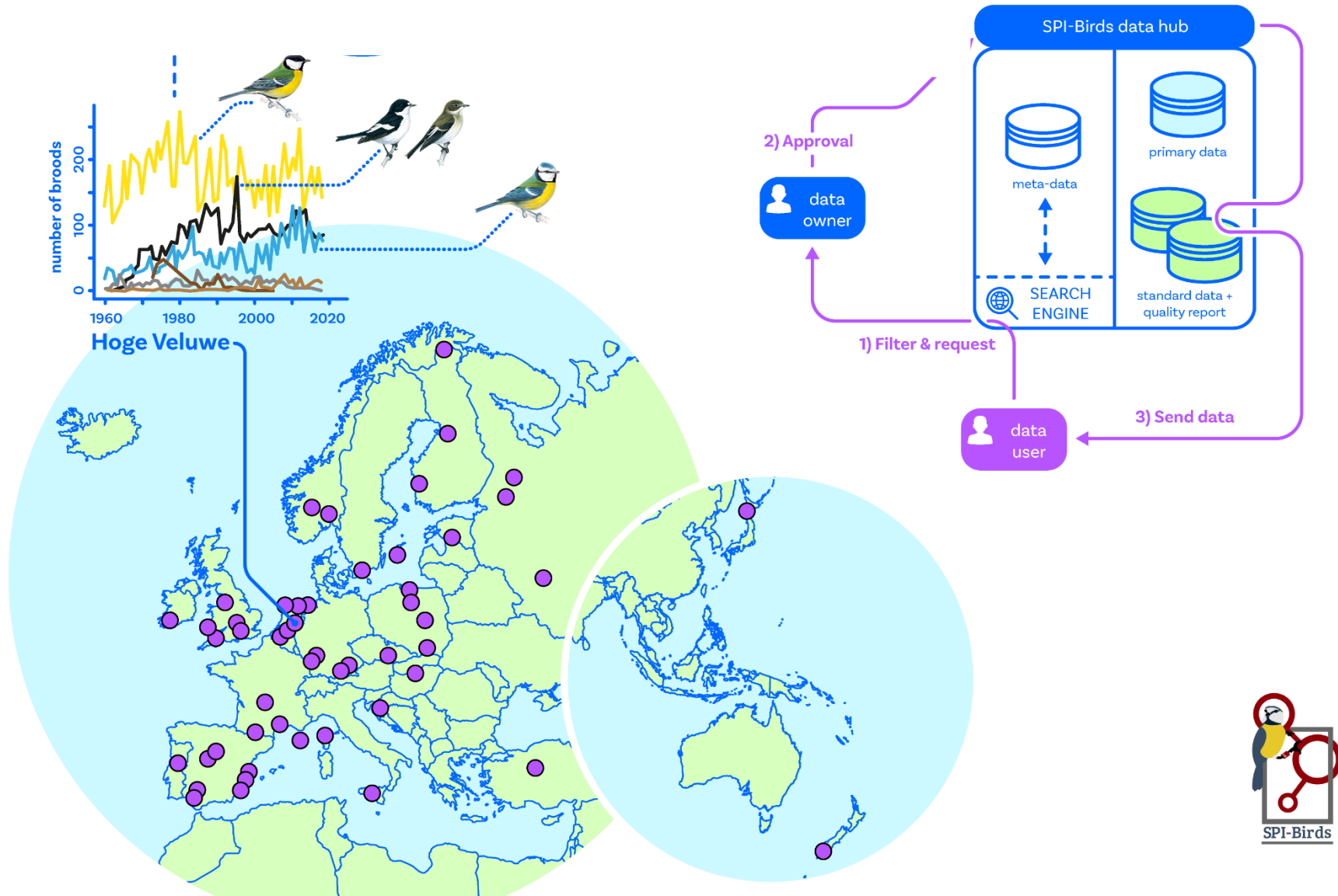
I don't want my  
long-collected data  
to be **OPEN**

I want my data to  
be used, and use  
others data, but  
under my control



I see the benefits of  
a common data  
format but do not  
have capacity to  
change my data  
format

# Workflow



# IDE NAM SUPER

**>120 members**

**>110 datasets**

**28 species**

**~2.0 mill individuals**



Which one is more important for scientific progress (especially in Ecology/Evolution)? 1/2 #OpenData #FAIR #sciencetwitter



129 votes · Final results

10:06 AM · Nov 17, 2020 · Twitter Web App

Which one is easier to achieve (especially in Ecology/Evolution)? 2/2 #OpenData #FAIR #sciencetwitter



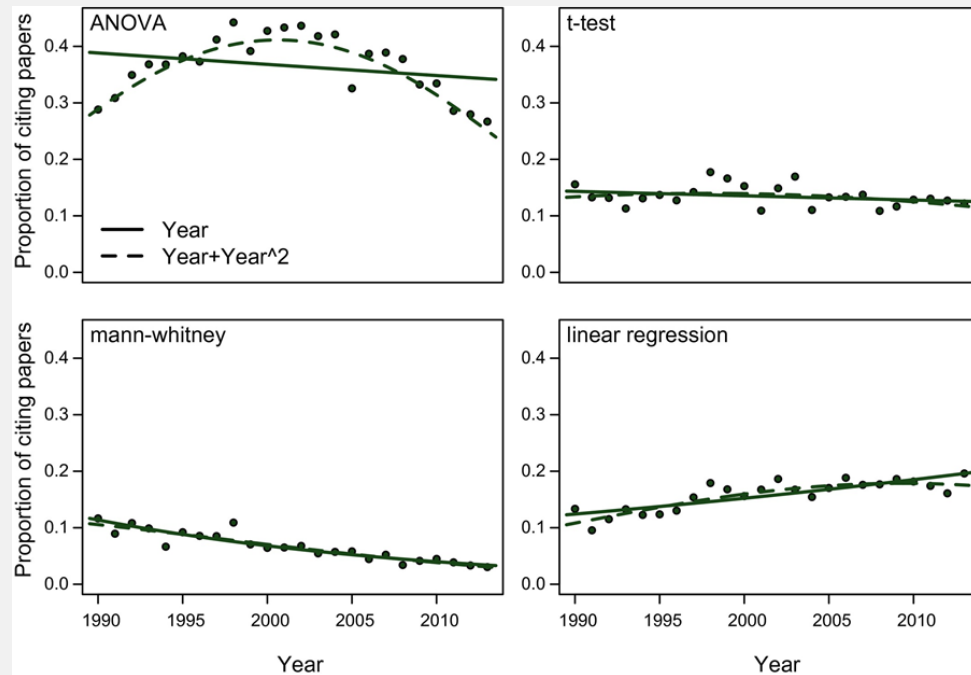
52 votes · Final results



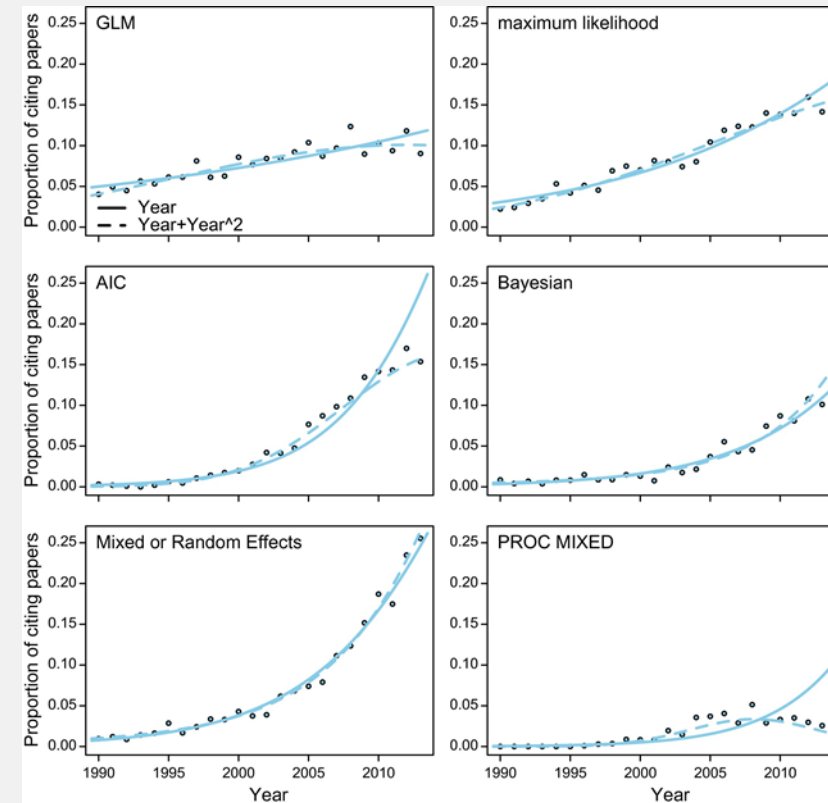
## 5) PROGRAMI/KODOVI

# Danas skoro svaki rad ima nekakav kod u pozadini

## Jednostavni



## Kompleksniji

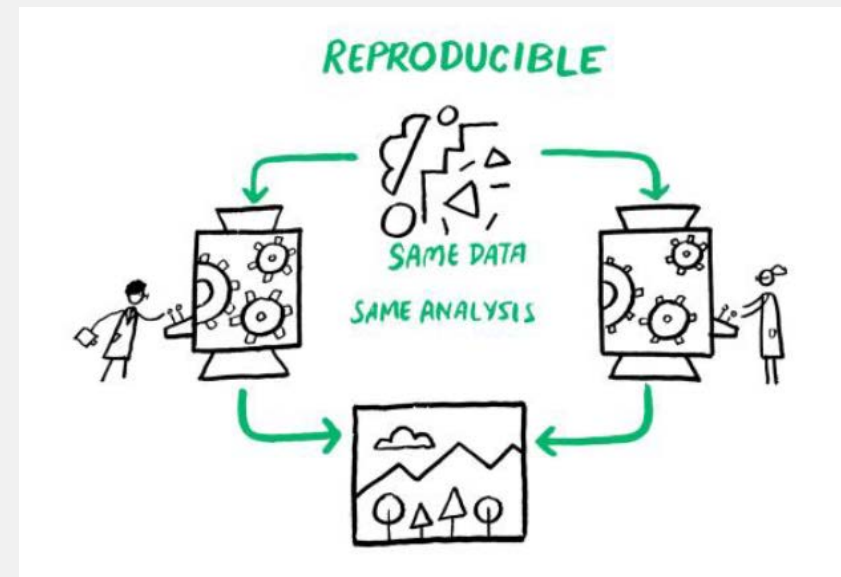
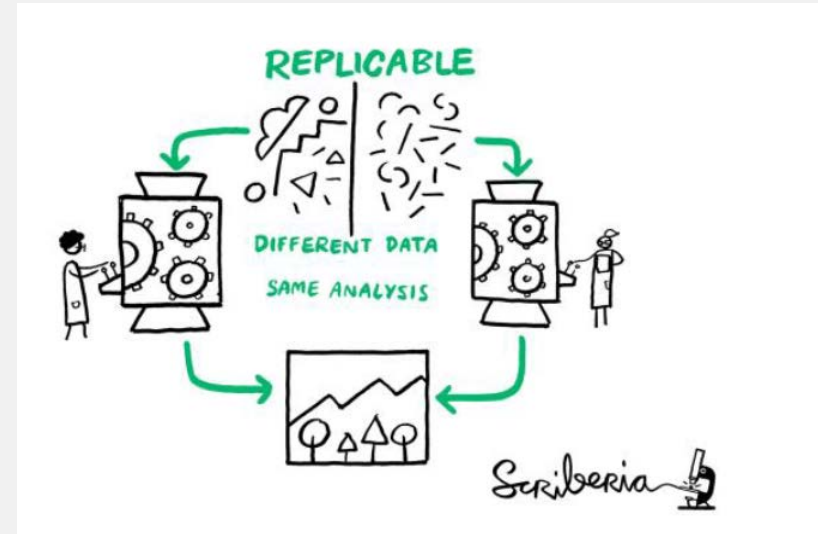


[Touchon & McCoy 2016](#)

<https://doi.org/10.1002/ecs2.1394>

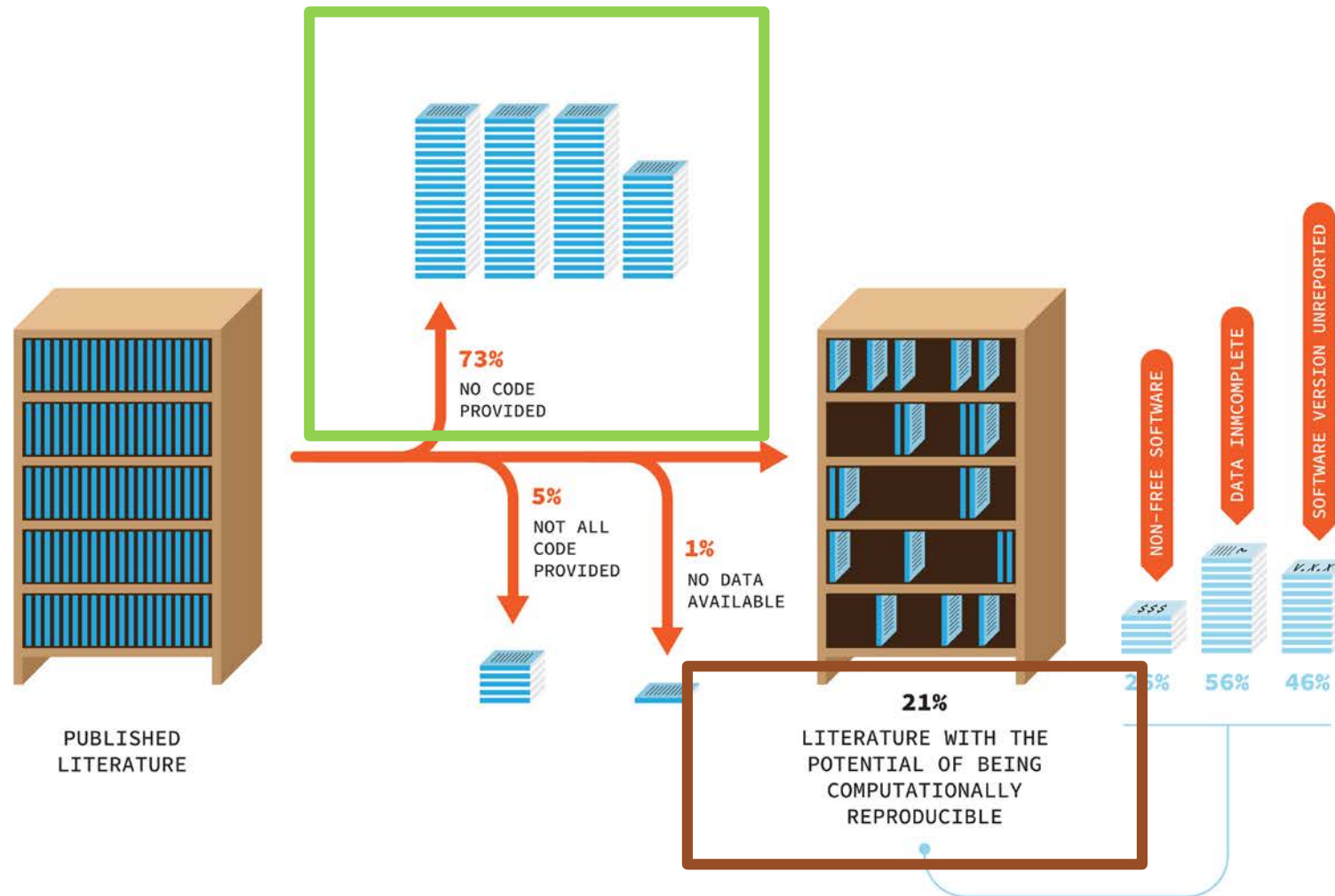
# Kod pomaže:

- 1) Da razumijemo analizu
- 2) Evaluiramo rezultate
- 3) Iskoristimo ga
- 4) Povećava povjerenje u znanost:
  - reproducibility & replicability

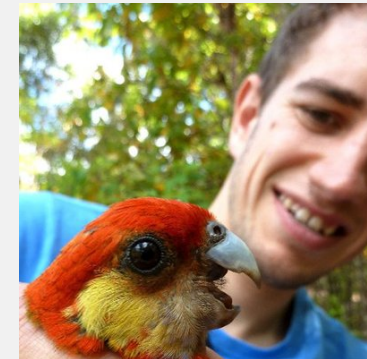




# REPRODUCIBILNOST U EKOLOGIJI & EVOLUCIJI



Alfredo Sánchez-Tójar



Simon Evans

Ilona van den Berg

- ‘One single solitary line of code, a “I –“ that does not belong, and you have a positive result where it should be a negative result.’
- ‘why don’t we review R code (or other custom software) as part of the peer-review process?’



*Evolutionary Ecology Research*, 2009, 11: 1217–1233

Morphological and dietary differences between individuals are ~~weakly but positively~~ <sup>NOT</sup> correlated within a population of threespine stickleback

Daniel I. Bolnick and Jeffrey S. Paull

*Section of Integrative Biology, University of Texas at Austin, Austin, Texas, USA*

**ABSTRACT**

**Background:** Many theoretical models of speciation and niche evolution assume that the ecological similarity between conspecific individuals depends on their phenotypic similarity. Thus, competition between individuals is expected to depend on their phenotypic similarity. Theoretical models often assume that this intraspecific competition function is Gaussian.

**Questions:** Are morphological similarity and diet similarity positively correlated? If so, is this relationship non-linear?

**Data:** Stomach contents, stable isotope ratios ( $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$ ), and trophic morphology (standard length, gape width, body width, gill raker number, and gill raker length) for 265 threespine sticklebacks (*Gasterosteus aculeatus*) from a single population from a lake in British Columbia.

**Analysis:** We calculated the diet similarity and morphological similarity between all pairs of individuals in our sample. We examined the correlation between diet and morphological similarity, and tested whether the relationship exhibits any non-linearity.

**Conclusions:** Similarity in trophic morphology is correlated with dietary similarity between individuals. ~~However,~~ <sup>↑</sup> both body size and trophic morphology explained ~~remarkably small~~ <sup>NONE</sup> percentages of the variance in diet overlap. Also, we found no evidence of curvature in the intraspecific competition function. ~~NOT~~

<http://ecoevoeco.blogspot.com/2016/12/wrong-lot.html>

source is mentioned in our blog: [www.unbelievablefactsblog.com](http://www.unbelievablefactsblog.com)



In 1962, a programmer omitted a single hyphen in the code for the Mariner I rocket causing it to explode shortly after take off. This typo cost NASA the today's equivalent of \$630 million dollars.

## Mariner I

Arthur C. Clarke who opined that Mariner I was :

**"wrecked by the most expensive hyphen in history,"**

6) ZAŠTO BI NAS TREBALO BITI BRIGA?

# RAVNOPRAVNOST, RASTA ZNANJA

- <https://youtu.be/9vz06QO3UkQ?t=6032>

**View**

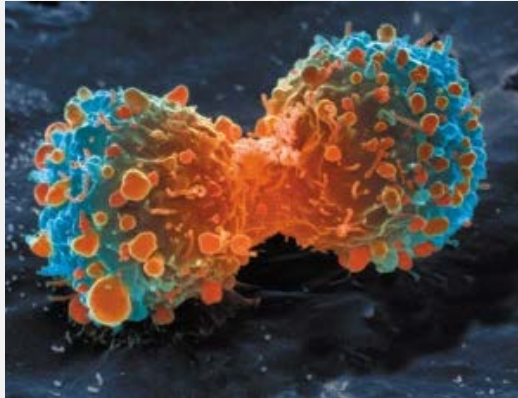
**Workflow**

## How Aaron Swartz paved way for Jack Andraka's revolutionary cancer test

Jack Andraka's breakthrough pancreatic cancer test would have never come about were it not for access to online journals -- what Internet guru Aaron Swartz was promoting before his death. Andraka "religiously" used free online academic journals in the research because "in most online databases, articles cost about \$35, and there are only about 10 pages."

Massoud Hayoun | Posted: Jan 29th, 2013





Findings confirmed in only 6 (11%) cases

Bagley & Elis 2012, Nature



Replication largely failed

97% of original studies had significant results ( $P < .05$ );

36% percent of replications had significant results;

OSC, 2015, Nature

The  
Economist

OCTOBER 19TH-25TH 2013

Economist.com

Washington's lawyer surplus

How to do a nuclear deal with Iran

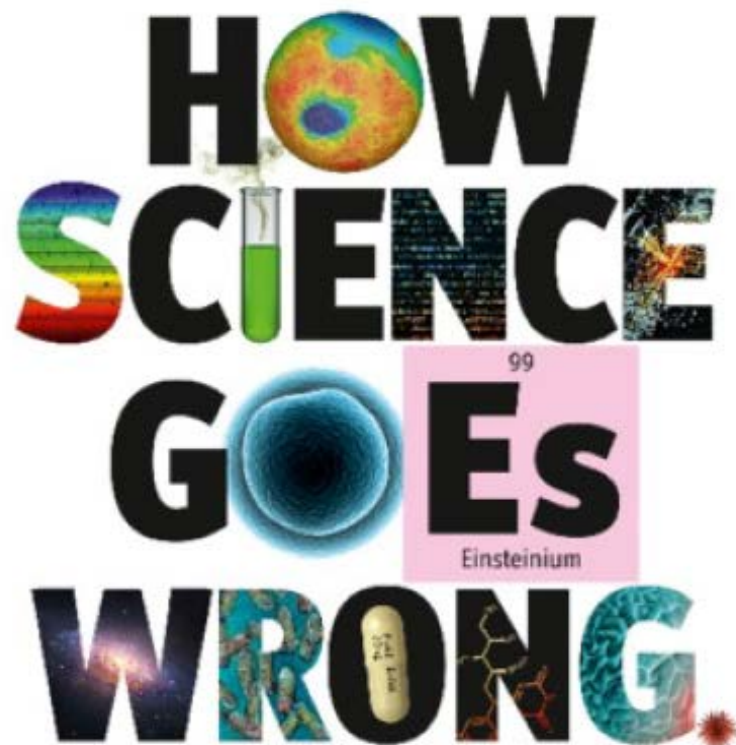
Investment tips from Nobel economists

Junk bonds are back

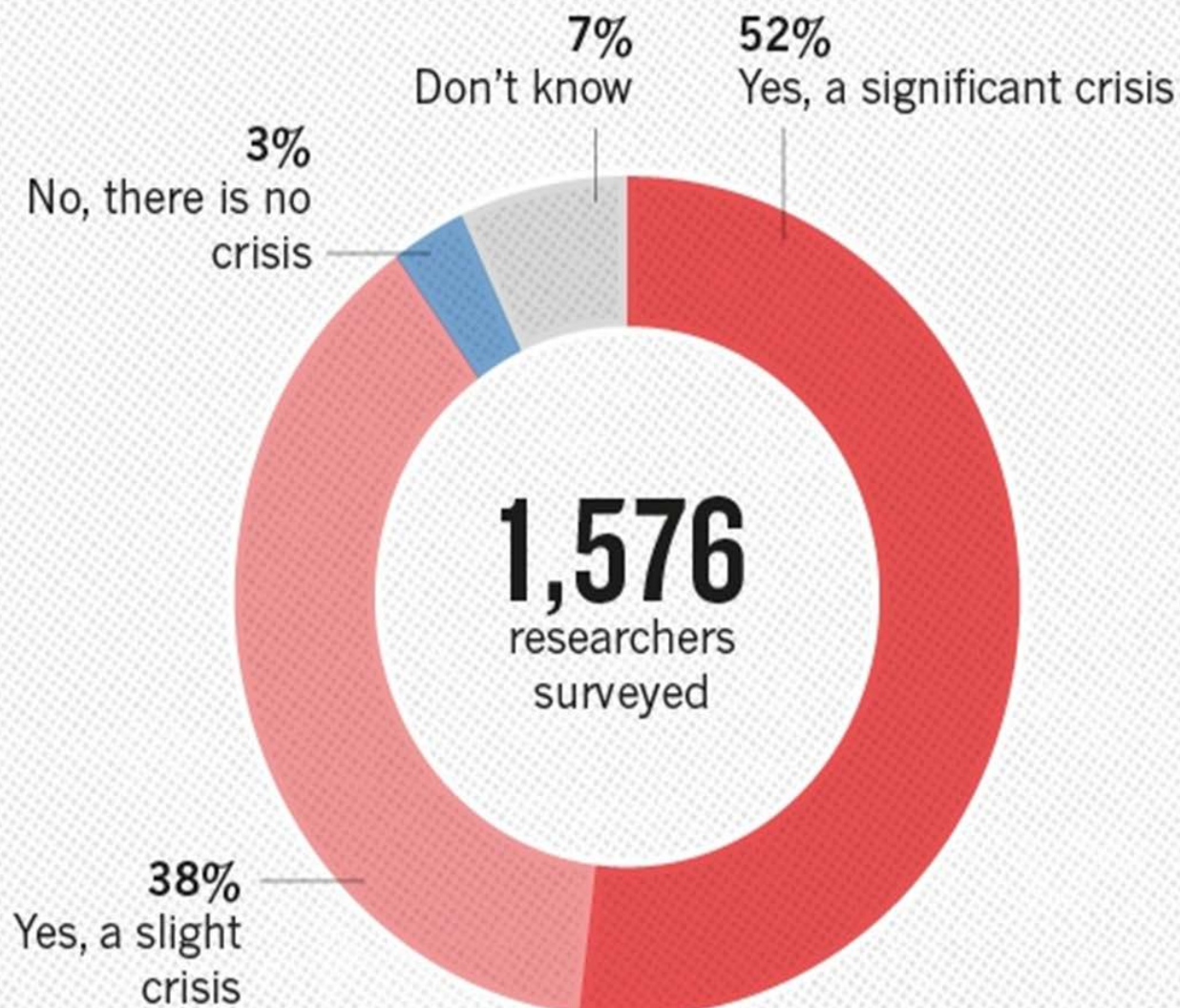
The meaning of Sachin Tendulkar

HOW  
SCIENCE  
GOES  
WRONG

99  
Einsteinium



## IS THERE A REPRODUCIBILITY CRISIS?





MRKVA I BATINA



**Table 1** A list, with brief explanations, of each of the eight existing Transparency and Openness Promotion (TOP) guidelines (<https://cos.io/top/>)

## TOP Guideline

1. Citation standards (citation of data sets, etc.)
2. Data transparency (data archiving)
3. Analytic methods (code) transparency (code archiving)
4. Research materials transparency (materials archiving)
5. Design and analysis transparency (reporting of details of methods and results)
6. Pre-registration of studies (registering study prior to initiation)
7. Pre-registration of analysis plans (registering analysis plan prior to study initiation)
8. Replication (a study designed to replicate a previously published study)

<https://www.cos.io/initiatives/top-guidelines>

nature portfolio

View all journals

Search

nature > nature portfolio > editorial policies > reporting standards and availability of data, materials, code and protocols

Editorial policies

Authorship

Competing interests

Confidentiality

Plagiarism and duplicate publication

Preprints & Conference Proceedings

Image integrity and standards

Peer Review

## Reporting standards and availability of data, materials, code and protocols

An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims. A condition of publication in a Nature Portfolio journal is that **authors are required to make materials, data, code, and associated protocols promptly available to readers without undue qualifications**. Any restrictions on the availability of materials or information must be disclosed to the editors at the time of submission. Any restrictions must also be disclosed in the submitted manuscript.

T. H. Parker,<sup>1\*</sup> S. Nakagawa,<sup>2</sup> J. Gurevitch,<sup>3</sup> and IIEE (Improving Inference in Evolutionary Biology and Ecology) workshop participants\*

<sup>1</sup>Department of Biology Whitman College Walla Walla, USA, <sup>2</sup>School of Biological Earth and Environmental Sciences University of New South Wales Sydney, Australia and <sup>3</sup>Department of Ecology and Evolution Stony Brook University USA

## Conservation Biology

Editorial

## Promoting transparency in conservation science



PLOS BIOLOGY PREREGISTERED RESEARCH ARTICLE

## GUIDELINES FOR AUTHORS

**Preregistered Research Articles** (aka Registered Reports) are a form of empirical article offered at *PLOS Biology* in which study rationale, methods and proposed analyses are reviewed prior to research being conducted. High quality protocols are reviewed for technical soundness of the proposed methodology, and provisionally accepted for publication before data collection commences. Refer to the [Center for Open Science](#) for more details.

# Journal retracts controversial paper on dangers of microplastics to fish

Researchers behind study, which may have helped cement case for [banning microbeads](#), found guilty of scientific misconduct

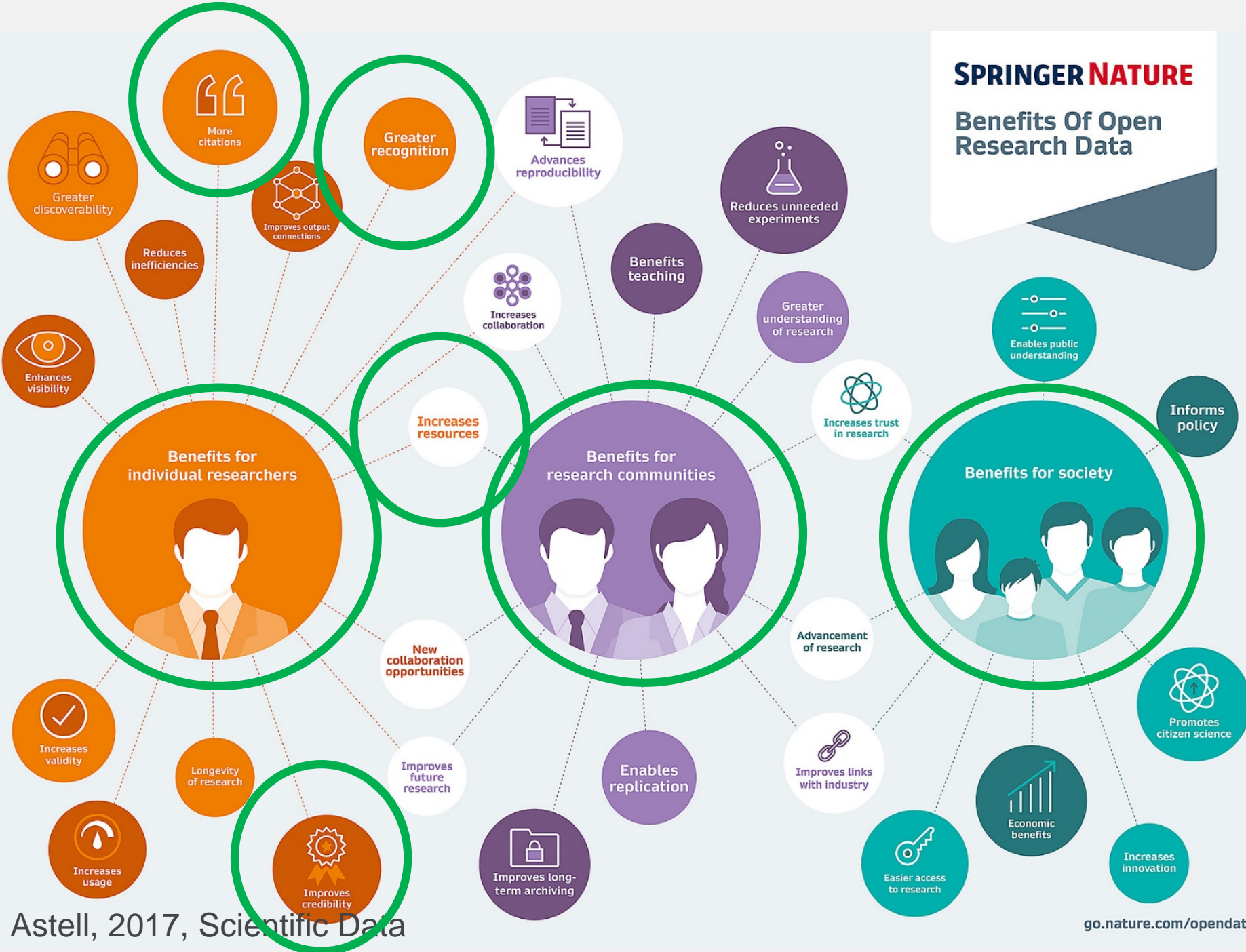


the guardian

Scientists say there is still compelling evidence to back legislation banning tiny plastic beads that are widely



**Benefits Of Open Research Data**



# Responding to needs and crisis



towardsdatascience.com/scientific-data-analysis-pipeline-and-reproducibility-75f9d5b4c5

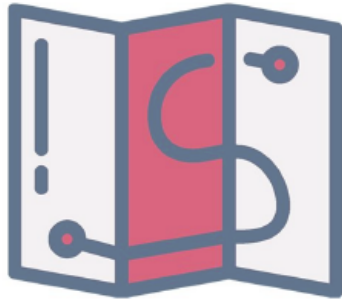
## Reproducibility of published studies is still problematic

- 21% pharmacology data ([doi.org/10.1038/nrd3439-c1](https://doi.org/10.1038/nrd3439-c1))
- 11% cancer data ([doi.org/10.1038/483531a](https://doi.org/10.1038/483531a))
- unsatisfactory in ML ([openreview.net/pdf?id=By4I2PbQ-](https://openreview.net/pdf?id=By4I2PbQ-))



theconversation.com/how-robots-can-help-us-embrace-a-more-human-view-of-disability-76815

## Human-machine collaboration is the future



**Data-related mandates by funders and institutions are growing**



**The scholarly publishing ecosystem is changing**



**Researchers need recognition and credit**

Otvoreni pristup je budućnost  
& timski sport

**ALI, VIŠE NIJE IZBORAN**



7) KAKO SE UKLJUČITI/'OKORISTITI'?

# Potražite literaturu o temama relevantnim za vaše istraživanje

Postoje mnoge studije o temama kao što su:

- Otvoreni podatci
- Pre-registracija
- Statističke pogreške
- Pseudoreplikacija
- .....

**Upotrijebite podatke iz ovih studija da poboljšate svoje istraživanje i objasnite drugima zašto je važno usvojiti bolje prakse**

Potražite i pridružite se postojećim grupama ili pokrenite svoju



## Welcome to ReproducibiliTea

We are a grassroots [journal club initiative](#) that helps researchers create local Open Science journal clubs at their universities to discuss diverse issues, papers and ideas about improving science, reproducibility and the Open Science movement. Started in early 2018 at the University of Oxford, ReproducibiliTea has now spread to 139 institutions in 27 different countries. We are completely volunteer run, and provide a unique and supportive community for our members, who are predominantly Early Career Researchers.

# MATERIJALI



- <https://www.cos.io/events>
- <https://www.fosteropenscience.eu/events>
- <https://www.openaire.eu/support>

- <https://aimos.community/>
- <https://www.sortee.org/>
- Oxford/ Berlin summer school



# ŠTO ZNANSTVENICIMA TREBA



- razumjeti zašto je otvorenost korisna i bitna
- jednostavna i dostatna rješenja
- pomoć u implementaciji
- mogućnost učenja
- prepoznavanje otvorenih praksi

- Kontaktirajte mene ili kolege iz CZI

# HVALA

## **Kolaboratori & studenti:**

M. Purgar

T. Klanjscek

A. Sanchez-Tojar

S. Evans

M. E .Visser

S. Nakagawa

J. Ramarkes

A.Veraart

A.Aldas Verga

I.Van den Berg

T. Crowhter

M. Baglioni

P. Manghi

S.Woutersen-Windhouver

T. Parker

A.Bezine

S.Vriend

L. Bailey

SPI-Birds Network

SORTEE