

# Open Science and FAIR data

NI4OS-Europe Dissemination Event

Prof. Vladimir Trajkovic

November 6, 2020



# Who am I?

Vladimir Trajkovikj ([trvlado@finki.ukim.mk](mailto:trvlado@finki.ukim.mk))

Faculty of Computer Science and Engineering, R. N. Macedonia



## Special thanks for content sources to:

René van Horik & Cees Hof, Data Archiving and Networked Services (DANS), The Netherlands

Prof. Nikola Stikov, Polytechnique Montréal, Canada

Prof. Roberto Barbera – University of Catania – Italy



## Special thanks for infrastructure and NI4OS data to:

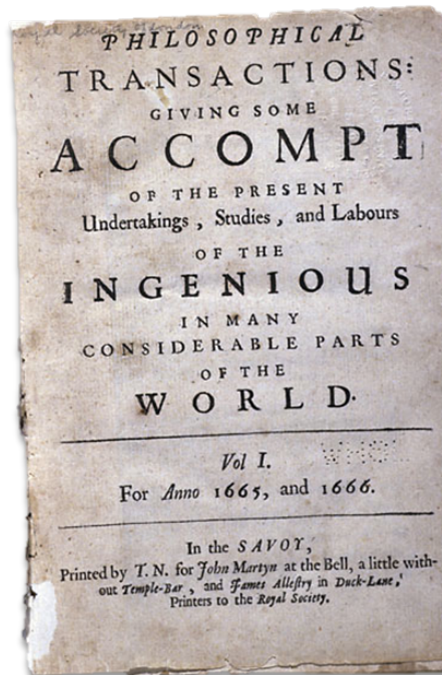
Prof. Atanas Mishev, FCSE, R.N. Macedonia

Prof Sonja Filiposka, FCSE, R.N. Macedonia



- ❑ Scientific challenges today
  
- ❑ Open Science
  - ❑ FAIR data and principles
  - ❑ Next steps
  
- ❑ Conclusion & Discussion

# Scientific output: Challenge 1



arXiv.org > hep-ex > arXiv:1207.7214

### High Energy Physics - Experiment

#### Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC

The ATLAS Collaboration

(Submitted on 31 Jul 2012 (v1), last revised 31 Aug 2012 (this version, v2))

A search for the Standard Model Higgs boson in proton-proton collisions with the ATLAS detector at the LHC is presented. The datasets used correspond to integrated luminosities of approximately 4.8 fb<sup>-1</sup> collected at sqrt(s) = 7 TeV in 2011 and 5.8 fb<sup>-1</sup> at sqrt(s) = 8 TeV in 2012. Individual searches in the channels H->ZZ(\*)->llll, H->gamma gamma and H->WW->e nu mu nu in the 8 TeV data are combined with previously published results of searches for H->ZZ(\*)->WW(\*)-l, bbbar and tau+>tau- in the 7 TeV data and results from improved analyses of the H->ZZ(\*)->llll and H->gamma gamma channels in the 7 TeV data. Clear evidence for the production of a neutral boson with a measured mass of 126.0 +/- 0.4(stat) +/- 0.4(sys) GeV is presented. This observation, which has a significance of 5.9 standard deviations, corresponding to a background fluctuation probability of 1.7x10<sup>-9</sup>, is compatible with the production and decay of the Standard Model Higgs boson.

Comments: 24 pages plus author list (38 pages total), 12 figures, 7 tables, revised author list, matches version to appear in Physics Letters B

Subjects: High Energy Physics - Experiment (hep-ex)

Journal reference: Phys.Lett. B716 (2012) 1-29

DOI: 10.1016/j.physletb.2012.08.020

Report number: CERN-PH-EP-2012-218

Cite as: arXiv:1207.7214 [hep-ex] (or arXiv:1207.7214v2 [hep-ex] for this version)

Submission history

From: ATLAS Publications [view email]

[v1] Tue, 31 Jul 2012 11:59:59 GMT (33448)

[v2] Fri, 31 Aug 2012 19:29:54 GMT (33448)

Which authors of this paper are endorsers? | Disable MathJax (What is MathJax?)

Link back to: arXiv, form interface, contact.



Physics Letters B 916 (2012) 1-29

Contents lists available at ScienceDirect

### Physics Letters B

www.elsevier.com/locate/physletb

#### Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC<sup>a</sup>

ATLAS Collaboration<sup>b</sup>

This paper is dedicated to the memory of our ATLAS colleagues who did not live to see the full impact and significance of their contributions to the experiment.

ARTICLE INFO ABSTRACT

Received 11 July 2012  
Revised 16 August 2012  
Accepted 11 August 2012  
Available online 14 August 2012  
Editor: W. Schuler

A search for the Standard Model Higgs boson in proton-proton collisions with the ATLAS detector at the LHC is presented. The datasets used correspond to integrated luminosities of approximately 4.8 fb<sup>-1</sup> collected at sqrt(s) = 7 TeV in 2011 and 5.8 fb<sup>-1</sup> at sqrt(s) = 8 TeV in 2012. Individual searches in the channels H->ZZ(\*)->ll, H->gamma gamma and H->WW->e nu mu nu in the 8 TeV data are combined with previously published results of searches for H->ZZ(\*)->WW(\*)-l, bbbar and tau+>tau- in the 7 TeV data and results from improved analyses of the H->ZZ(\*)->llll and H->gamma gamma channels in the 7 TeV data. Clear evidence for the production of a neutral boson with a measured mass of 126.0 +/- 0.4(stat) +/- 0.4(sys) GeV is presented. This observation, which has a significance of 5.9 standard deviations, corresponding to a background fluctuation probability of 1.7 x 10<sup>-9</sup>, is compatible with the production and decay of the Standard Model Higgs boson.

© 2012 CERN. Published by Elsevier B.V. All rights reserved.

1. Introduction

The Standard Model (SM) of particle physics [1-4] has been tested by many experiments over the last four decades and has been shown to successfully describe high energy particle interactions. However, the mechanism that breaks electroweak symmetry in the SM has not been verified experimentally. This mechanism [5-16], which gives mass to massive elementary particles, implies the existence of a scalar particle, the SM Higgs boson. The search for the Higgs boson, the only elementary particle in the SM that has not yet been observed, is one of the highlights of the Large Hadron Collider (LHC) physics programme.

Indirect limits on the SM Higgs boson mass of m<sub>H</sub> < 158 GeV at 95% confidence level (CL) have been set using global fits to precision electroweak results [17]. Direct searches at LEP [18], the Tevatron [14,16] and the LHC [11,8] have previously excluded, at 95% CL, a SM Higgs boson with mass below 600 GeV, apart from some mass regions between 116 GeV and 127 GeV.

Both the ATLAS and CMS Collaborations reported excesses of events in their 2011 datasets of proton-proton (pp) collisions at centre-of-mass energy sqrt(s) = 7 TeV at the LHC, which were compatible with SM Higgs boson production and decay in the mass region 124-126 GeV, with significance of 2.9 and 3.1 standard deviations (σ), respectively [17,18]. The CD0 and D0 experiments at the Tevatron have also recently reported a broad excess in the mass region 120-135 GeV using the existing LHC constraints, the observed local significance for m<sub>H</sub> = 125 GeV are 2.7σ for CD0 [14], 1.1σ for D0 [15] and 2.8σ for their combination [16].

The previous ATLAS searches in 4.6-8.8 fb<sup>-1</sup> of data at sqrt(s) = 7 TeV are combined here with new searches for H->ZZ(\*)->ll, H->gamma gamma and H->WW(\*)->tau+>tau- in the 8.8-5.8 fb<sup>-1</sup> of pp collision data taken at sqrt(s) = 8 TeV between April and June 2012.

The data were recorded with instantaneous luminosities up to 6.8 x 10<sup>31</sup> cm<sup>-2</sup> s<sup>-1</sup>; they are therefore affected by multiple pp collisions occurring in the same or neighbouring bunch crossings (pile-up). In the 7 TeV data, the average number of interactions per bunch crossing was approximately 10; the average increased to approximately 20 in the 8 TeV data. The reconstruction, identification and isolation criteria used for electrons and photons in the 8 TeV data are improved, making the H->ZZ(\*)->ll and H->gamma gamma searches more robust against the increased pile-up. These analyses were re-optimised with simulation and frozen before looking at the 8 TeV data.

In the H->WW(\*)->gamma gamma channel, the increased pile-up deteriorates the event missing transverse momentum, E<sub>miss</sub>, resolution, which results in significantly larger Drell-Yan background in the same-massive final states. Since the gamma channel provides most of the sensitivity of the search, only this final state is used in the analysis of the 8 TeV data. The kinematic region in which a SM Higgs boson with a mass between 110 GeV and 140 GeV is

<sup>a</sup> © CERN for the benefit of the ATLAS Collaboration.  
<sup>b</sup> E-mail address: atlas@atlas.cern.ch.

0550-3213/\$ - see front matter © 2012 CERN. Published by Elsevier B.V. All rights reserved.  
http://dx.doi.org/10.1016/j.physletb.2012.08.020

It is the same since almost 4 centuries!



## Repeatability

- ❑ The closeness of agreement between independent results obtained with the **same method on identical test material, under the same conditions** (same operator, same apparatus, same laboratory and after short intervals of time)
- ❑ Affected by random errors

## Reproducibility

- ❑ The closeness of agreement between independent results obtained with the **same method on identical test material but under different conditions** (different operators, different apparatus, different laboratories and/or after different intervals of time)
- ❑ Affected by systematic errors

## Repeatability

- The closeness of agreement between results obtained with the same method on identical test material by the same operator, same apparatus, under the same conditions (same laboratory, after short intervals of time)

- Affected by systematic errors

**Is science really reproducible?**

- The closeness of agreement between independent results obtained with the same method on identical test material but under different conditions (different operators, different apparatus, different laboratories and/or after different intervals of time)

- Affected by systematic errors

# Challenge 2: Scientific Reproducibility



**nature** International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors

Archive > Specials & supplements archive > Challenges in irreproducible research

Take part in Nature Publishing Group's annual reader survey here for the chance to win a Macbook Air. Find out more

### SPECIAL

See all speci

#### CHALLENGES IN IRREPRODUCIBLE RESEARCH

No research paper can ever be considered to be the final word, and the replication and corroboration of research results is key to the scientific process. In studying complex entities, especially animals and human beings, the complexity of the system and of the techniques can all too easily lead to results that seem robust in the lab, and valid to editors and referees of journals, but which do not stand the test of further studies. *Nature* has published a series of articles about the worrying extent to which research results have been found wanting in this respect. The editors of *Nature* and the *Nature* life sciences research journals have also taken substantive steps to put our own houses in order, in improving the transparency and robustness of what we publish. Journals, research laboratories and institutions and funders all have an interest in tackling issues of irreproducibility. We hope that the articles contained in this collection will help.

Free full access

- Editorial
- Features
- News and analysis
- Comment
- Perspectives and reviews

#### EDITORIAL

**Journals unite for reproducibility**  
Consensus on reporting principles aims to improve biomedical research.  
*Nature* 515, 7 ( 8 November 2014 )

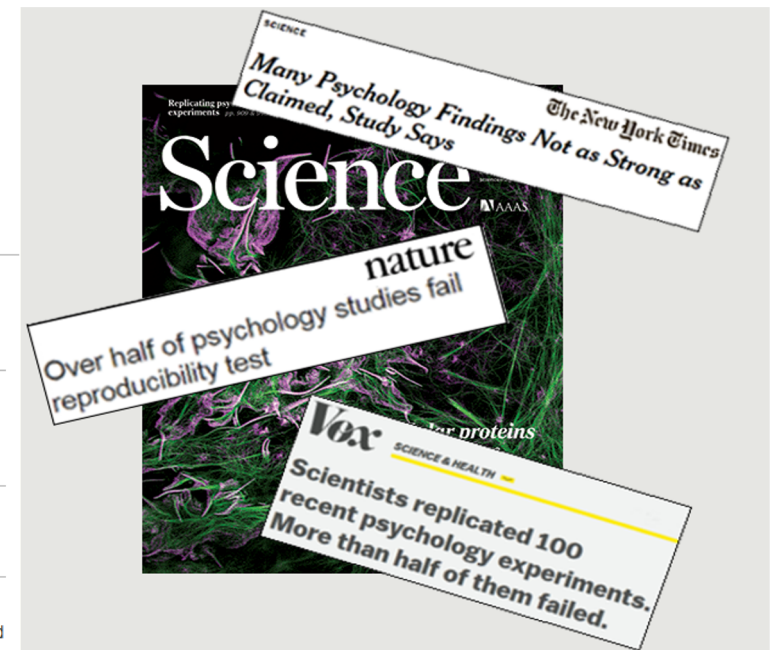
**Code share**  
Papers in *Nature* journals should make computer code accessible where possible.  
*Nature* 514, 538 ( 29 October 2014 )

**Reducing our irreproducibility**  
*Nature* 496, 398 ( 25 April 2013 )

**Further confirmation needed**  
A new mechanism for independently replicating research findings is one of several changes required to improve the quality of the biomedical literature.  
*Nature Biotechnology* 30, 806 ( 10 September 2012 )

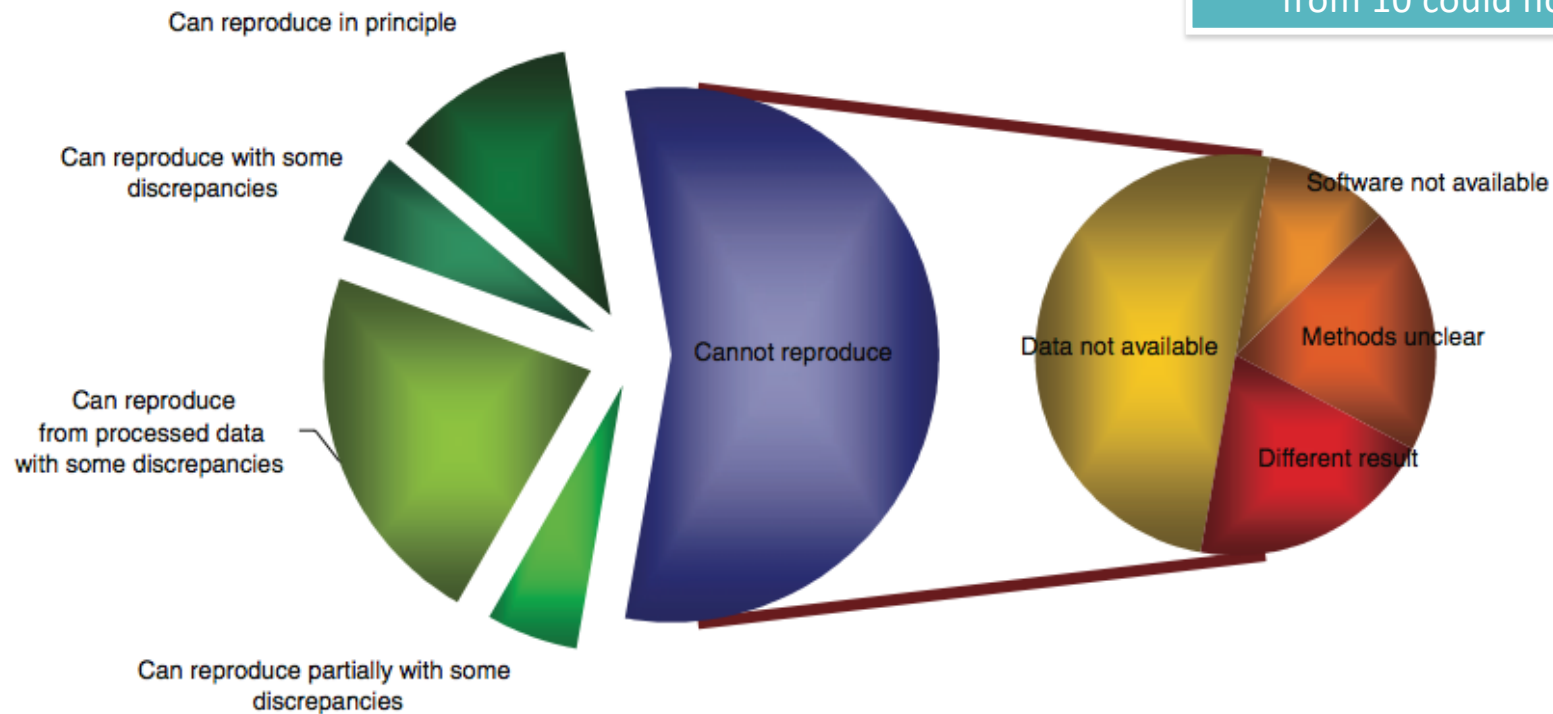
**Error prone**  
Biologists must realize the pitfalls of work on massive amounts of data.  
*Nature* 487, 406 ( 26 July 2012 )

**Must try harder**  
Too many sloppy mistakes are creeping into scientific papers. Lab heads must look more rigorously at the data — and at themselves.  
*Nature* 483, 509 ( 29 March 2012 )



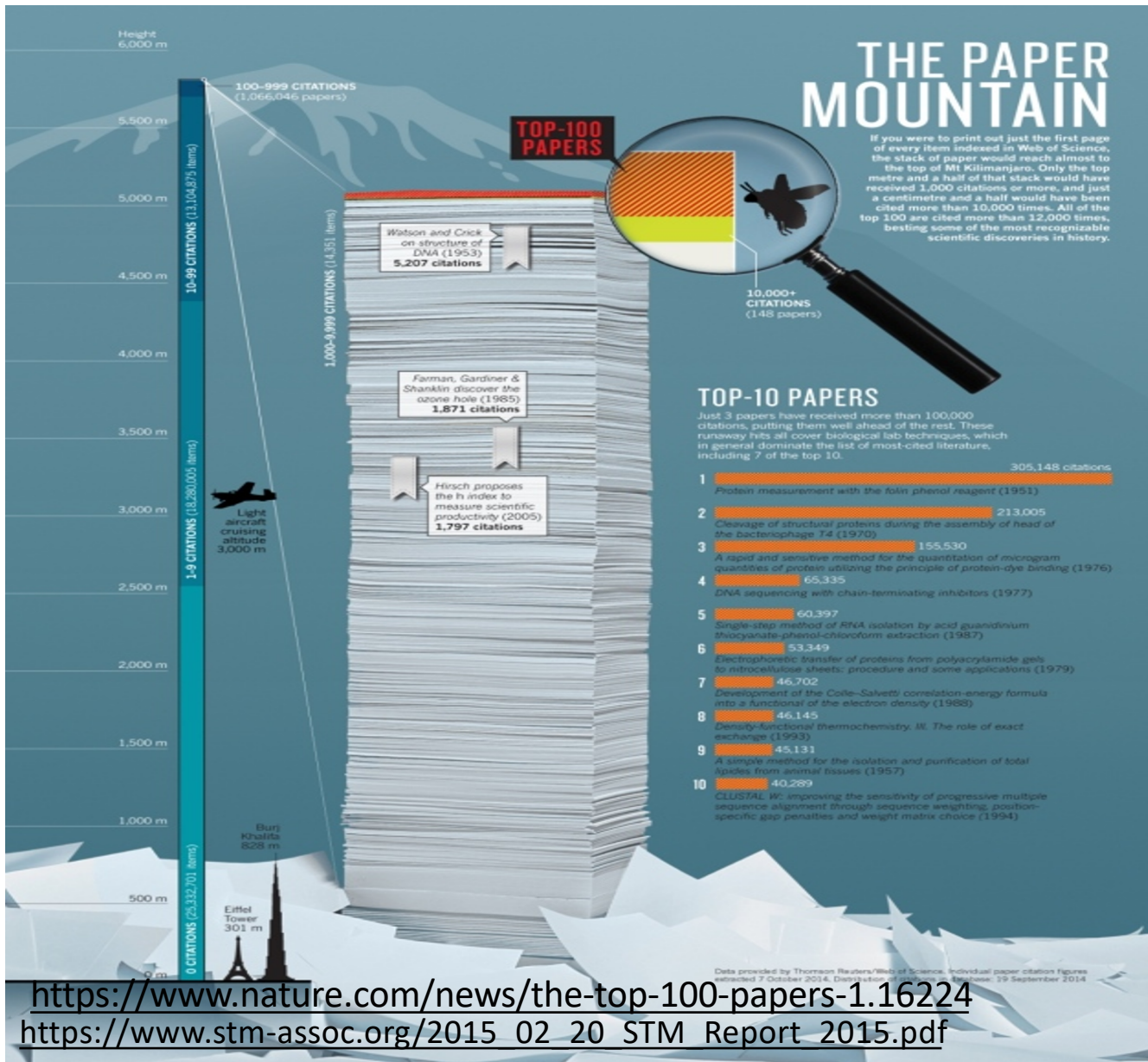
# Challenge 2: Scientific Reproducibility

Out of 18 microarray papers, results from 10 could not be reproduced

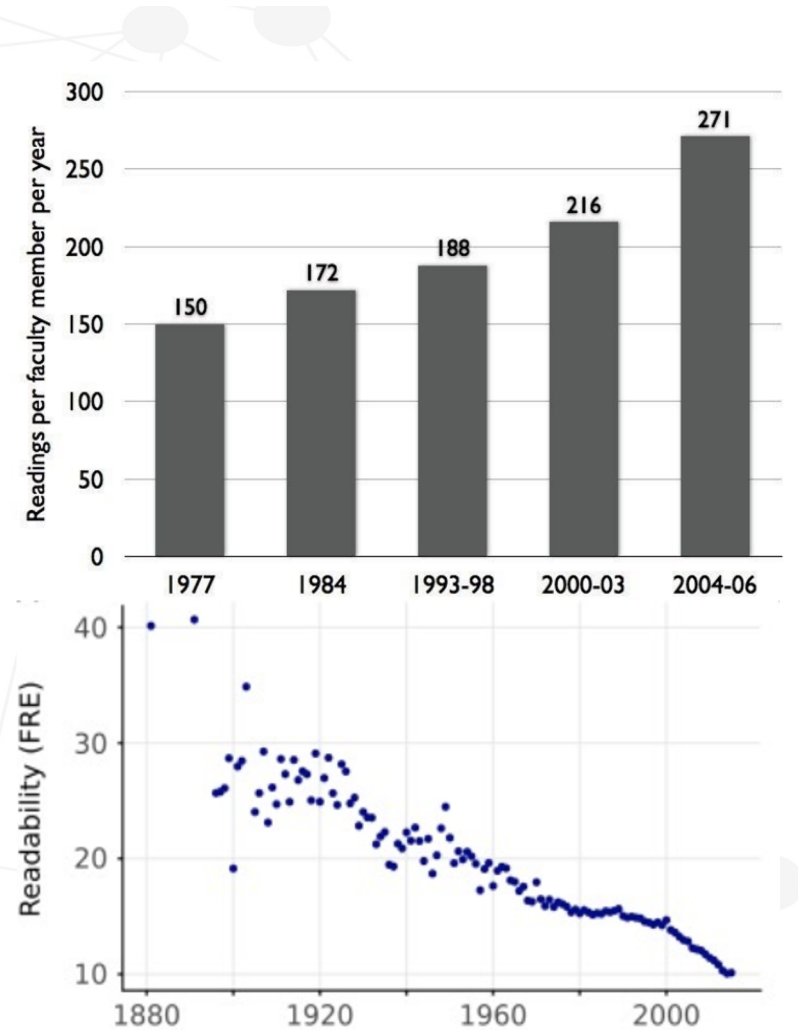


1. Ioannidis et al., 2009. Repeatability of published microarray gene expression analyses. *Nature Genetics* 41: 14
2. Science publishing: The trouble with retractions <http://www.nature.com/news/2011/111005/full/478026a.html>
3. Bjorn Brembs: Open Access and the looming crisis in science <https://theconversation.com/open-access-and-the-looming-crisis-in-science-14950>



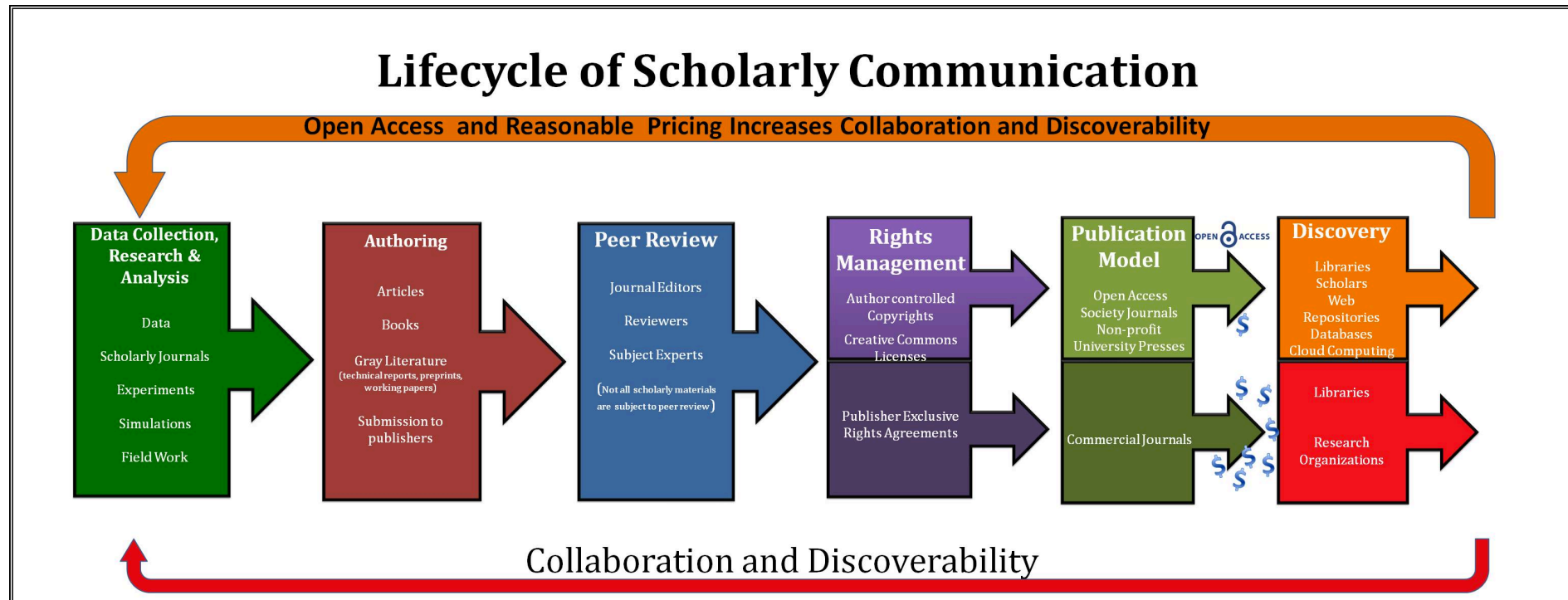


<https://www.nature.com/news/the-top-100-papers-1.16224>  
[https://www.stm-assoc.org/2015\\_02\\_20\\_STM\\_Report\\_2015.pdf](https://www.stm-assoc.org/2015_02_20_STM_Report_2015.pdf)

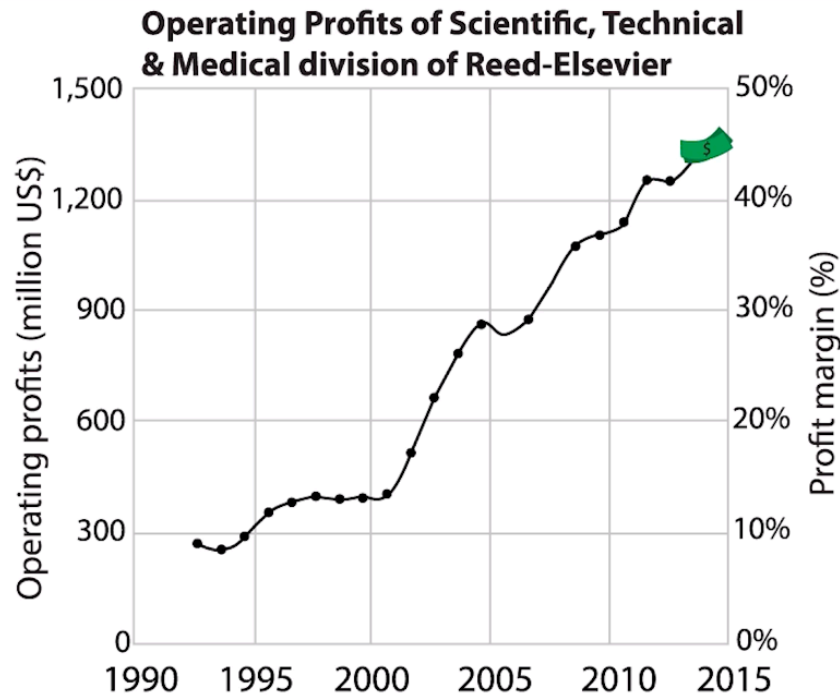


<https://elifesciences.org/articles/27725>

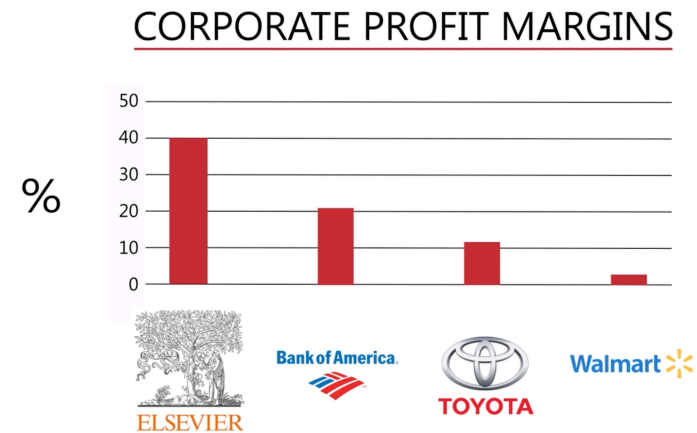
# Challenge 3: Scientific Communication



# Challenge 3: Scientific Communication



Source: Larivière, V., Haustein, S., Mongeon, P. (June, 2015). "The Oligopoly of Academic Publishers in the Digital Era," PLOS.

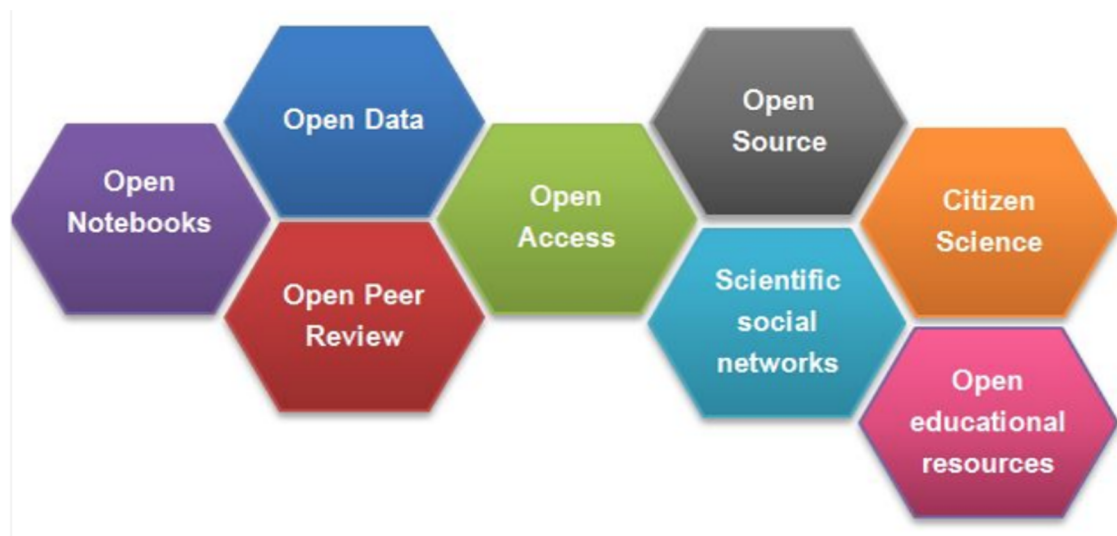


**What does this mean for small universities with no significant funding?**

# OPEN SCIENCE INTRO

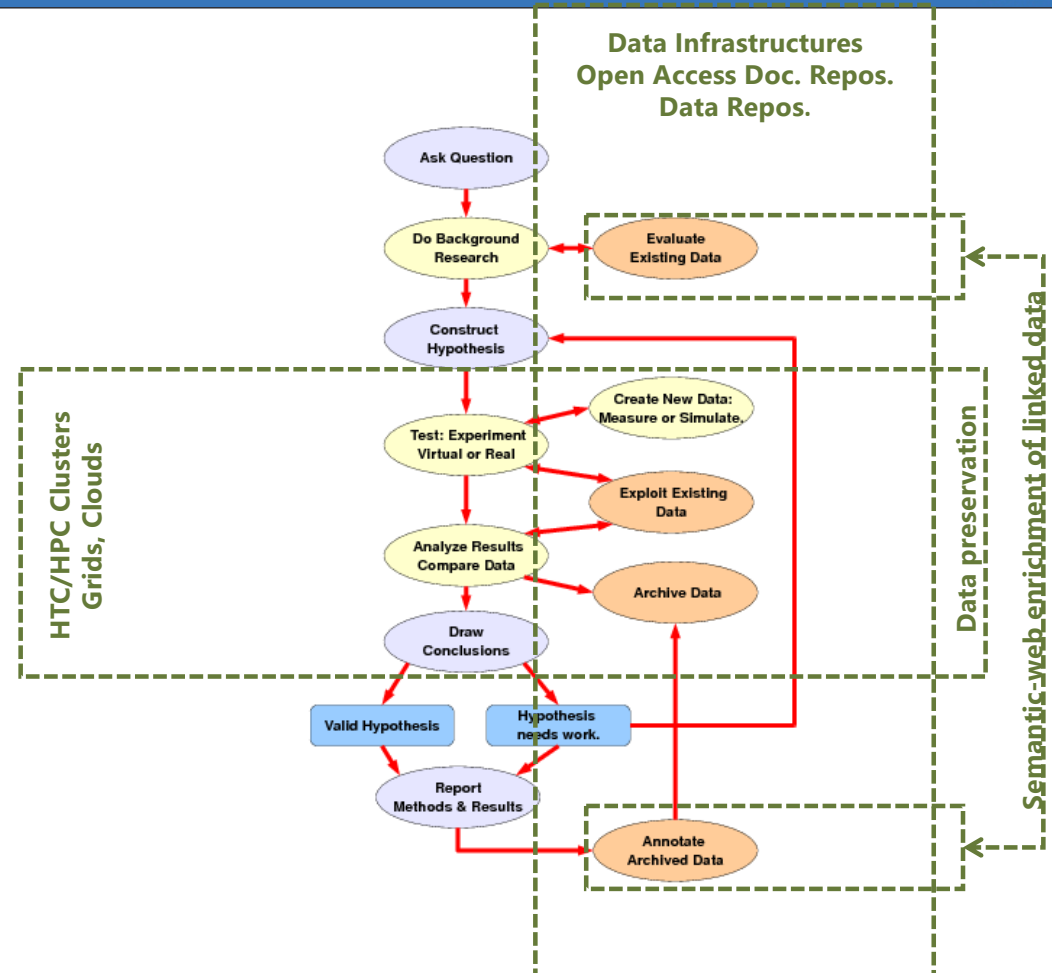


Open Science is the movement to make scientific research (publications, software, data) and its dissemination accessible to all levels of society. Open Science is transparent and accessible knowledge that is shared and developed through collaborative networks.



<https://www.fosteropenscience.eu/node/1420>

# How to fit open science in research data lifecycle



<https://www.ukdataservice.ac.uk/manage-data/lifecycle>

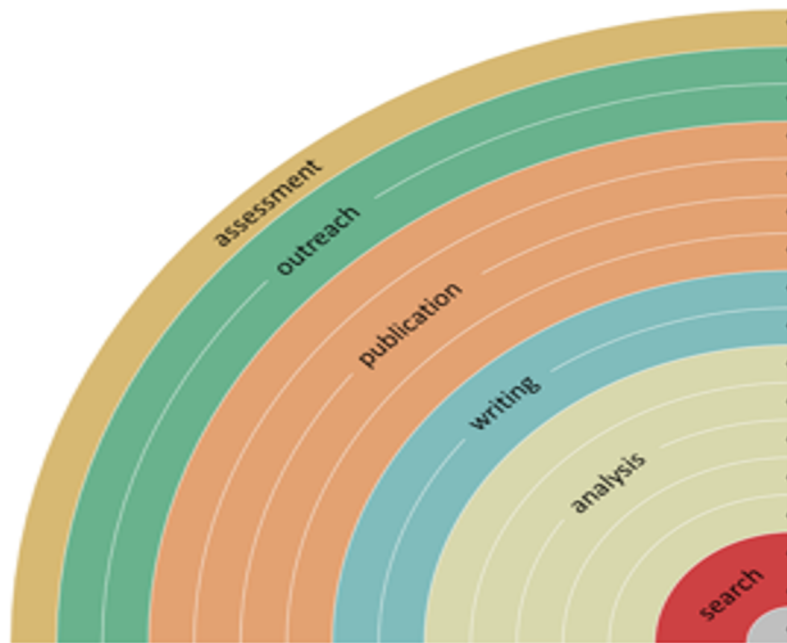
# Open science approach example



<https://slides.com/agahkarakuzu/qmrlabjn1>

# More possibilities?

You can make your workflow more open by ...



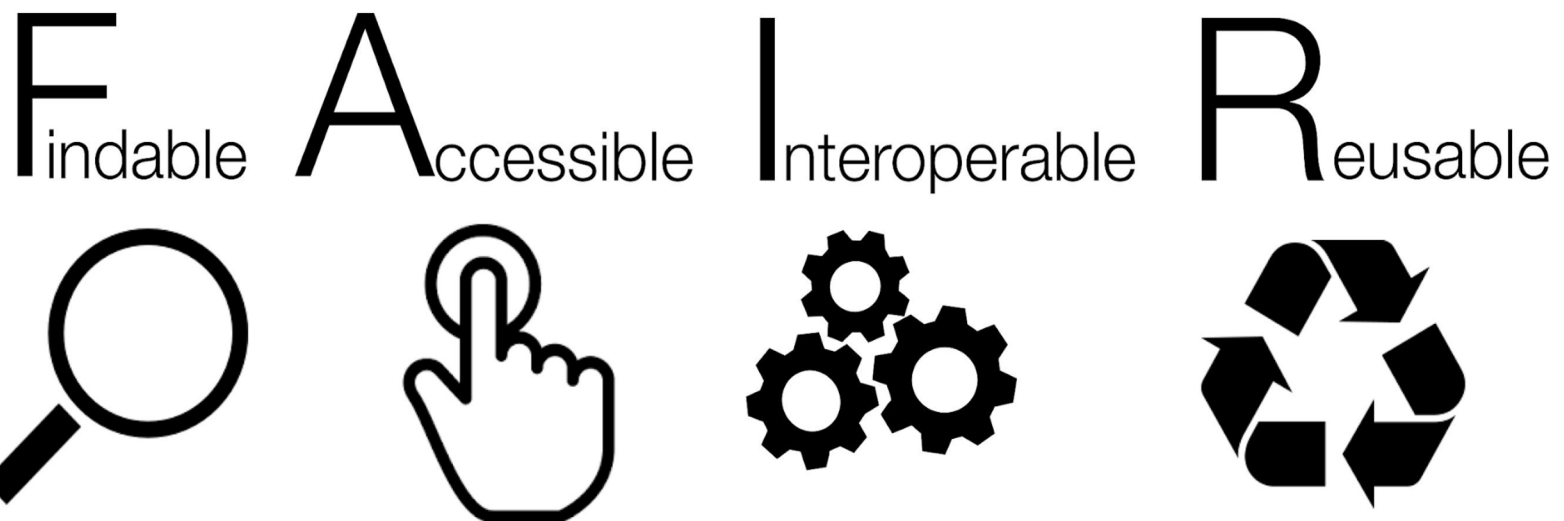
- adding alternative evaluation, e.g. with altmetrics
- communicating through social media, e.g. Twitter
- sharing posters & presentations, e.g. at FigShare
- using open licenses, e.g. CC0 or CC-BY
- publishing open access, 'green' or 'gold'
- using open peer review, e.g. at journals or PubPeer
- sharing preprints, e.g. at OSF, arXiv or bioRxiv
- using actionable formats, e.g. with Jupyter or CoCalc
- open XML-drafting, e.g. at Overleaf or Authorea
- sharing protocols & workfl., e.g. at Protocols.io
- sharing notebooks, e.g. at OpenNotebookScience
- sharing code, e.g. at GitHub with GNU/MIT license
- sharing data, e.g. at Dryad, Zenodo or Dataverse
- pre-registering, e.g. at OSF or AsPredicted
- commenting openly, e.g. with Hypothes.is
- using shared reference libraries, e.g. with Zotero
- sharing (grant) proposals, e.g. at RIO



Bianca Kramer & Jeroen Bosman <https://101innovations.wordpress.com>

DOI: 10.5281/zenodo.1147025

# FAIR data principles



[https://en.wikipedia.org/wiki/FAIR\\_data#/media/File:FAIR\\_data\\_principles.jpg](https://en.wikipedia.org/wiki/FAIR_data#/media/File:FAIR_data_principles.jpg)

# FAIR data principles

- ❑ **F**indable – Easy to find by both humans and computer systems and based on mandatory description of the metadata that allow the discovery of interesting datasets;
- ❑ **A**ccessible – Stored for long term such that they can be easily accessed and/or downloaded with well-defined license and access conditions (Open Access when possible), whether at the level of metadata, or at the level of the actual data content;
- ❑ **I**nteroperable – Ready to be combined with other datasets by humans as well as computer systems;
- ❑ **R**e-usable – Ready to be used for future research and to be processed further using computational methods

<http://www.nature.com/articles/sdata201618>  
[www.force11.org/group/fairgroup/fairprinciples](http://www.force11.org/group/fairgroup/fairprinciples)



# FAIR data principles: What need to be done?

## To be Findable:

- assigned a globally unique and eternally persistent identifier
- described with rich metadata
- registered or indexed in a searchable resource
- metadata specify the data identifier

## To be Accessible:

- retrievable by an identifier in a standardized communications protocol
- protocol is open, free, and universally implementable
- protocol allows for an authentication and authorization procedure
- metadata are accessible, even when the data are no longer available

# FAIR data principles: What need to be done?

## **To be Interoperable:**

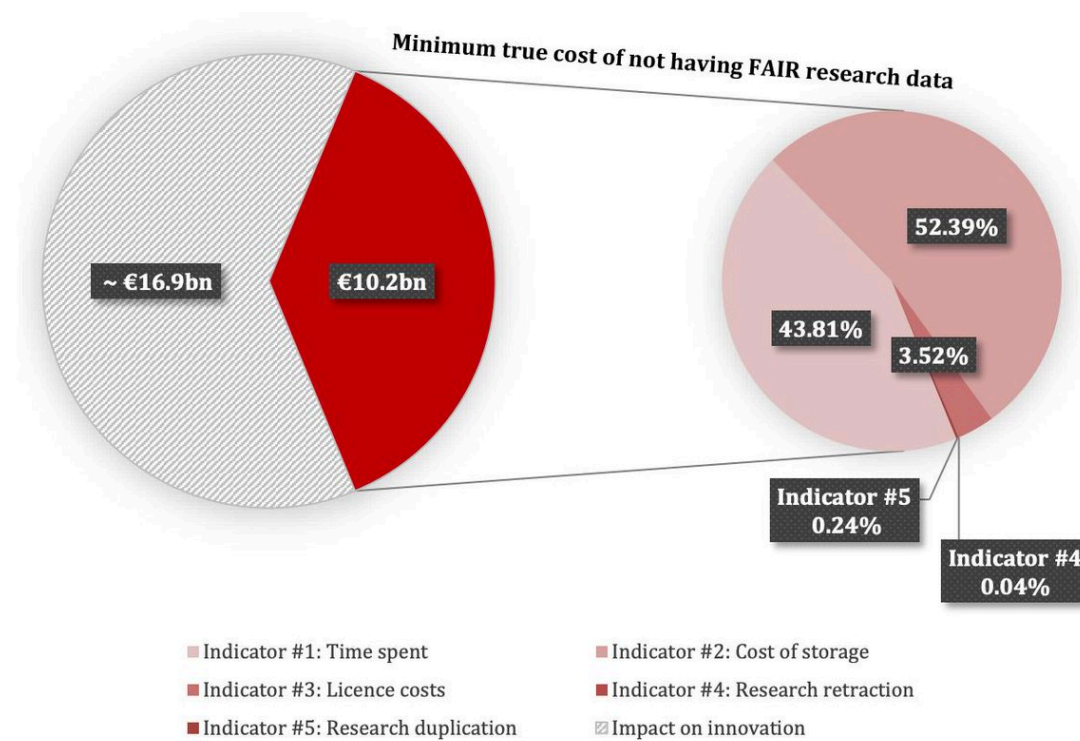
- formal, accessible, shared, broadly applicable language for knowledge representation
- use vocabularies that follow FAIR principles
- include qualified references to other (meta)data

## **To be Re-usable:**

- plurality of accurate and relevant attributes
- released with a clear and accessible data usage license
- associated with their provenance
- meet domain-relevant community standards



## Likely cost of not having FAIR research data



<https://publications.europa.eu/en/publication-detail/-/publication/d375368c-1a0a-11e9-8d04-01aa75ed71a1/language-en>

# Why it is needed?



⇒ **Virtual Research Communities**

⇒ **e-Infrastructure (NI4OS)**

⇒ **Applications**

⇒ **Data**

⇒ **Instruments/sensors**

# NEXT STEPS

# Initiatives concerning FAIR

FAIRsFAIR (Fostering FAIR data practices) - [www.fairsfair.eu](http://www.fairsfair.eu)

(aims to supply practical solutions for the use of the FAIR data principles throughout the research data life cycle)

GO FAIR initiative - [www.go-fair.org](http://www.go-fair.org)

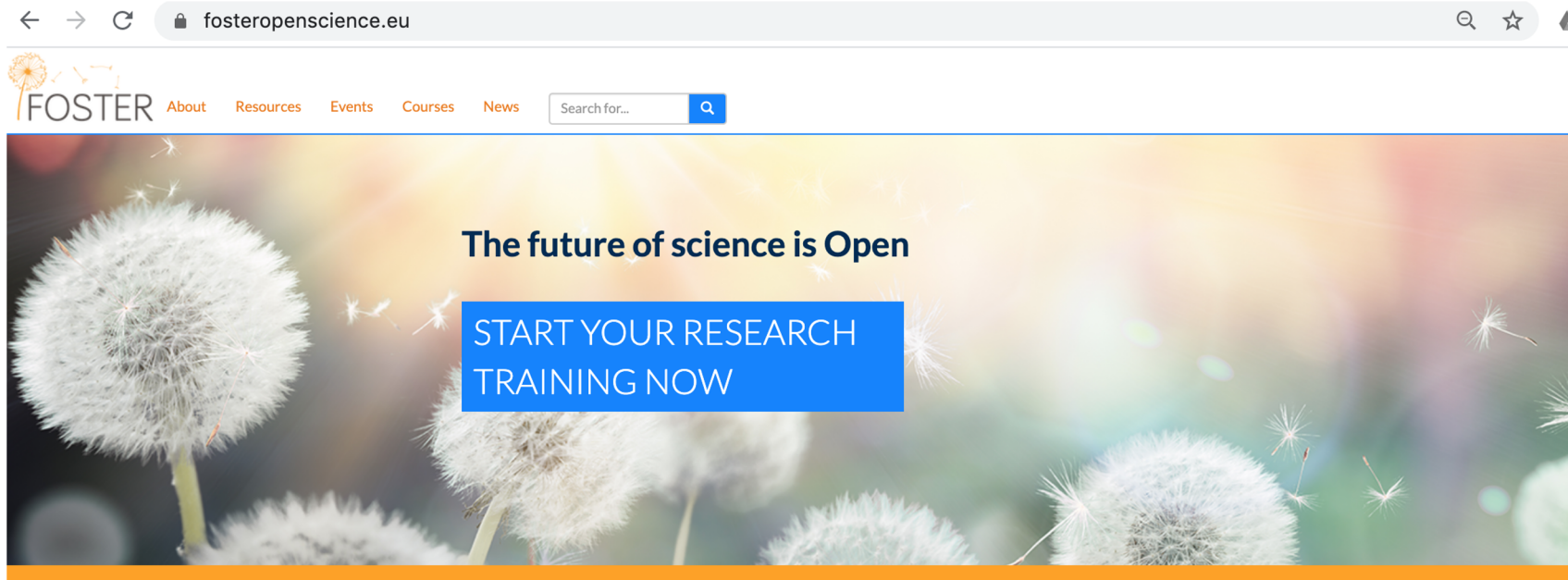
(aims to implement the FAIR data principles)

Detailed description of FAIR principles <https://www.go-fair.org/fair-principles/>  
FAIR resources <https://www.go-fair.org/resources/>

FAIRsharing - [fairsharing.org](http://fairsharing.org)

(A curated , informative and educational resource on data and metadata standards inter-related to databases and data policies)

# FOSTER: e-learning platform on Open Science



## USE FOSTER TO:



### Access Courses

Put Open Science into practice with our **Open Science training toolkit**. Our **courses** are authored by experts and experienced educators.



### Earn Badges

Get recognised for taking **our courses** and follow our **learning paths** to specialisations.



### Participate

Join our community of trainers and access our **Trainers' Corner**.

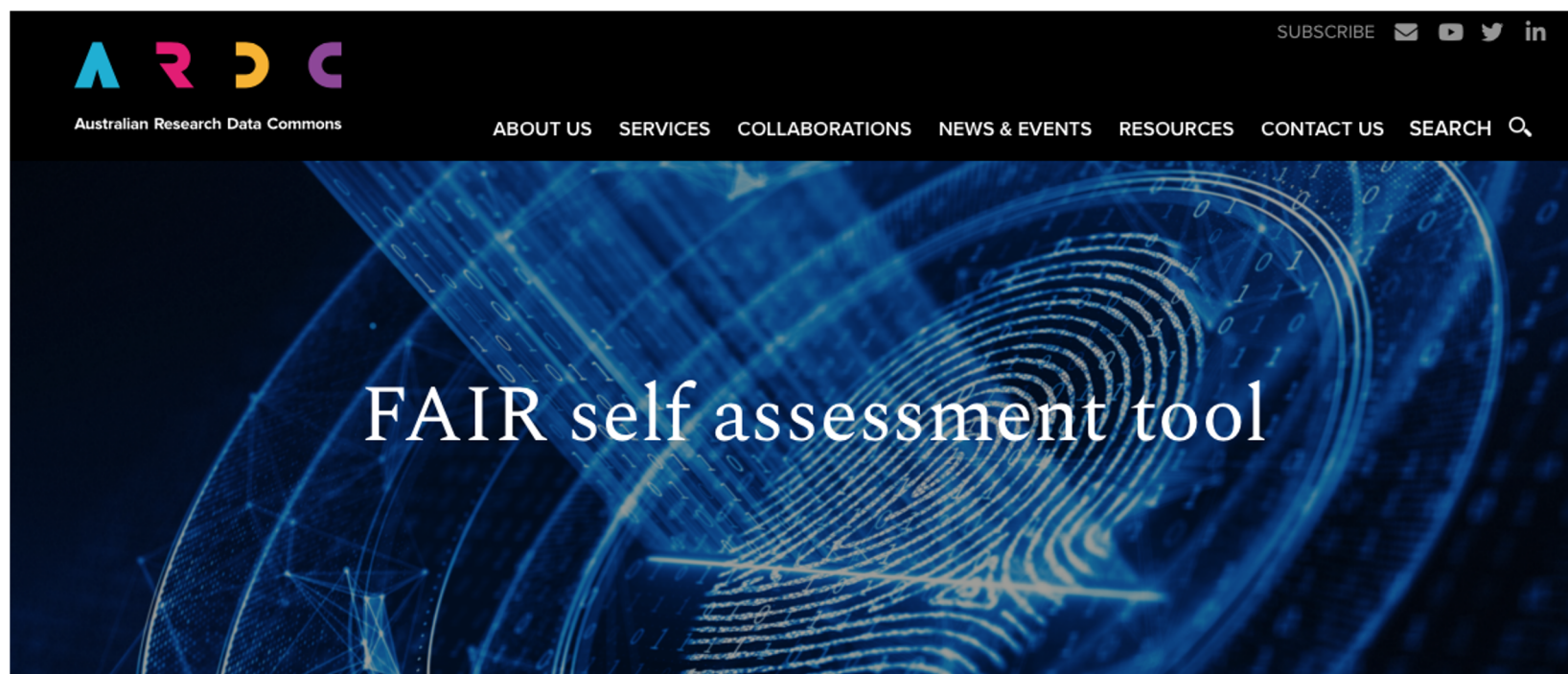


### Promote Open Science

Use the **Open Science training handbook**. In a variety of formats and languages.

# FAIR services for existing data (After research)

## FAIR data assessment services:



<https://ardc.edu.au/resources/working-with-data/fair-data/fair-self-assessment-tool/>



## Training on data platforms and repositories

Choosing the right platform / repository very much determines the FAIRness of the stored data!

Exercises / training around re3data.org are often clarifying the issues that require attention (metadata, certification, etc.) but **not readily available...**

<https://www.re3data.org>



re3data.org  
REGISTRY OF RESEARCH DATA REPOSITORIES

# Thanks!



Vladimir Trajkovikj ([trvlado@finki.ukim.mk](mailto:trvlado@finki.ukim.mk))

 [@NI4OS\\_eu](https://twitter.com/NI4OS_eu)

 [@NI4OS](https://facebook.com/NI4OS)