20th National Congress of the Italian Society for Neuroscience
SEPTEMBER 14th – 17th, 2023 | TURIN, CENTRO CONGRESSI LINGOTTO

Neuroscience and Open Science: a winning combination

Valentina Pasquale[1]

[1] Istituto Italiano di Tecnologia

Slides can be found at: https://doi.org/10.5281/zenodo.8349398

ISTITUTO ITALIANO
DI TECNOLOGIA

Setting the context:
why do we talk about Open Science?

My first research article…

Neuroscience 153 (2008) 1354–1369

About 15y ago…

SELF-ORGANIZATION AND NEURONAL AVALANCHES IN NETWORKS OF DISSOCIATED CORTICAL NEURONS

V. PASQUALE,[a] P. MASSOBRIO,[b] L. L. BOLOGNA,[a] M. CHIAPPALONE[a] AND S. MARTINOIA[b]*

[a]*Neuroscience and Brain Technology Department, Italian Institute of Technology, Via Morego 30, 16163 Genova, Italy*

[b]*Neuroengineering and Bio-nano Technology Group, Department of Biophysical and Electronic Engineering, University of Genova, Via all'Opera Pia 11A, 16145 Genova, Italy*

Key words: self-organized criticality, micro-electrode arrays, rat cortical neurons, *in vitro* development, scale-free network model.

Large random cortical networks developing *in vitro* and chronically coupled to micro-electrode arrays (MEAs) represent a valuable experimental model for studying the universal mechanisms governing the formation and con-

Where are today the data supporting my findings?

ISTITUTO ITALIANO DI TECNOLOGIA

- At that time there was no mandatory **data availability statement**... Authors were not required to share data or code supporting their findings: that was the normal!
- **Where are the data today?**
  - Raw data: somewhere in the lab where I did my PhD, stored in CDs or DVDs... Metadata about experiments were tracked on paper-sheets, hopefully still preserved in the lab in ring binders stored in closets. Raw data were saved in the original acquisition format (by MultiChannelSystems).
  - Code: the developed code is backed-up in an old portable hard-drive, which I have at home, in my bookcase.
  - Analyzed data: again, all results (images, figures, statistical analyses, etc.) were backed-up in a portable hard-drive, which is (most likely) at IIT (somewhere in a box where I have my old stuff...). The data format was proprietary Matlab format (.m, .mat).

15 years have passed and original research data are basically not accessible anymore...

Data are fragile

ISTITUTO ITALIANO DI TECNOLOGIA

Scientists losing data at a rapid rate

Decline can mean 80% of data are unavailable after 20 years.

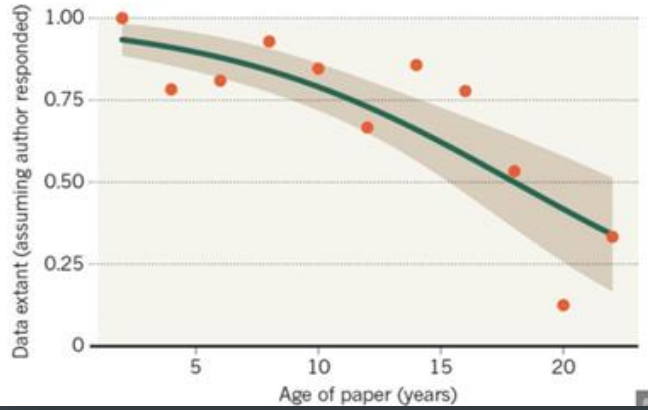
Elizabeth Gibney & Richard Van Noorden

19 December 2013

Rights & Permissions

MISSING DATA

As research articles age, the odds of their raw data being extant drop dramatically.



http://www.nature.com/news/scientists-losing-data-at-a-rapid-rate-1.14416

80% will be lost in 20 years

https://doi.org/10.1016/j.cub.2013.11.014

REPORT | VOLUME 24, ISSUE 1, P94-97, JANUARY 06, 2014

The Availability of Research Data Declines Rapidly with Article Age

Timothy H. Vines • Arianne Y.K. Albert • Rose L. Andrew • ... Jean-Sébastien Moore • Sébastien Renaut • Diana J. Rennison • Show all authors

Open Archive • Published: December 19, 2013 • DOI: https://doi.org/10.1016/j.cub.2013.11.014 •

Adapted from Elena Giglia, https://doi.org/10.5281/zenodo.3519305

https://doi.org/10.1126/science.acx9770

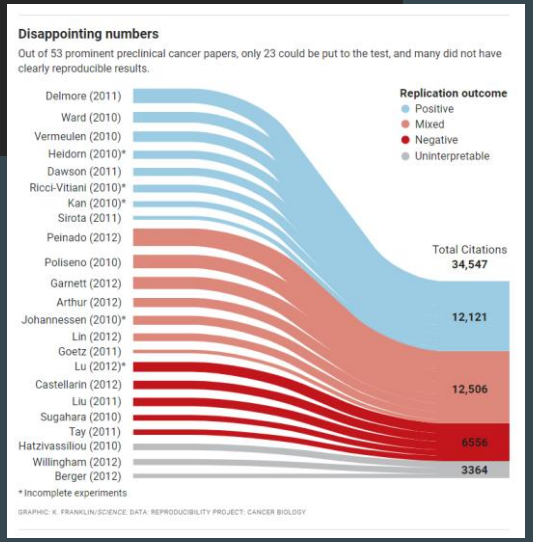
ISTITUTO ITALIANO DI TECNOLOGIA

NEWS | HEALTH

More than half of high-impact cancer lab studies could not be replicated in controversial analysis

Cancer reproducibility project couldn't assess many papers because of uncooperative authors and other challenges

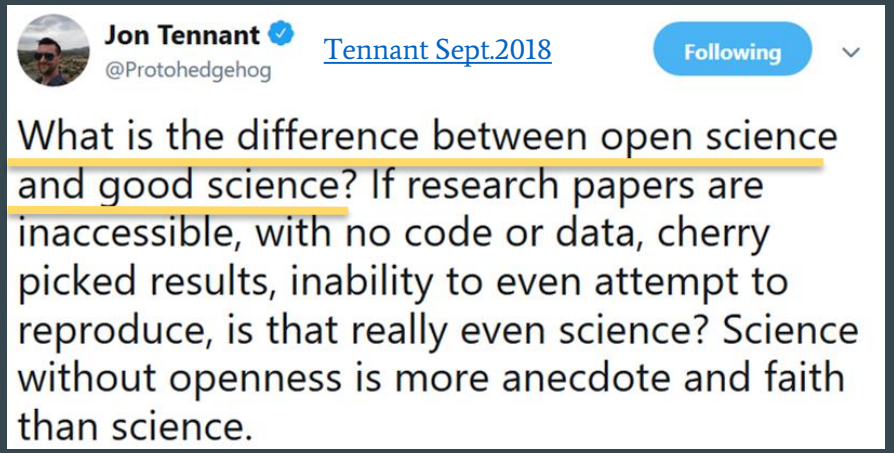
7 DEC 2021 • 8:00 AM • BY JOCELYN KAISER



https://elifesciences.org/collections/9b1e83d1/reproducibility-project-cancer-biology

Is that really the science we need?

Poor reproducibility, selective reporting, no supporting data, … is that really the Science we need?



There is no difference!!

Adapted from slide by E. Giglia, https://doi.org/10.5281/zenodo.3519306

What is Open Science?

A definition of Open Science

"Open science is defined as an inclusive construct that combines various movements and practices aiming to make multilingual scientific knowledge openly available, accessible and reusable for everyone, to increase scientific collaborations and sharing of information for the benefits of science and society, and to open the processes of scientific knowledge creation, evaluation and communication to societal actors beyond the traditional scientific community".
(UNESCO Recommendation on Open Science)

The key pillars of Open Science
Source: UNESCO Recommendation on Open Science

Nov. 26th 2021

Open Science is not just Open Access to publications

Source: UNESCO Open Science brochure.

Open scientific knowledge can be realized not only by opening access to scientific publications, but includes data, source code and software, educational resources, hardware, etc.

1. Dissemination should not be focused only on scientific articles.

2. Data should always be responsibly managed, also when not open.

FAIR data: how to responsibly share data

The FAIR principles were introduced in 2016 as a guideline to correct scientific Data Management.

https://www.go-fair.org/fair-principles

4 fundamental principles
15 sub-principles

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

F indable
A ccessible
I nteroperable
R eusable

For humans and machines

F ully
A (rtificial)
I (ntelligence)
R eady

The FAIR principles emphasize **machine-actionability** (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data.
Reference: GO FAIR website.

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

ISTITUTO ITALIANO
DI TECNOLOGIA

Open Science in Neuroscience: a few examples

The IBL: collaboration is key

Virtual lab unifying 22 neuroscience research groups across the world

Key factors

1. *Interdisciplinarity* | experimentalists & theoreticians
2. *Open data sharing* | all data released within 1y at the latest
3. *Standardization* | standardized methods & documentation
4. *Reproducibility* | open access/open source protocols, tools - hardware & software, including analytical pipelines

Publications

2023

"Pinpoint: trajectory planning for multi-probe electrophysiology and injections in an interactive web-based 3D environment," Daniel Birman, et al., bioRxiv, https://doi.org/10.1101/2023.07.14.548952

"A Brain-Wide Map of Neural Activity during Complex Behaviour," International Brain Laboratory et al, bioRxiv, https://doi.org/10.1101/2023.07.04.547681

"Brain-wide representations of prior information in mouse decision-making," Charles Findling et al., bioRxiv, https://doi.org/10.1101/2023.07.04.547684

"Lightning Pose: improved animal pose estimation via semi-supervised learning, Bayesian ensembling, and cloud-native open-source tools," Dan Biderman et al., bioRxiv, https://doi.org/10.1101/2023.04.28.538703

"A modular architecture for organizing, processing and sharing neurophysiology data," The International Brain Laboratory et al., Nat Methods, https://doi.org/10.1038/s41592-022-01742-6

2022

"Reproducibility of in-vivo electrophysiological measurements in mice," The International Brain Laboratory et al., bioRxiv. https://doi.org/10.1101/2022.05.09.491042

"Mice alternate between discrete strategies during perceptual decision-making," Ashwood et al., Nat Neurosci. https://doi.org/10.1038/s41593-021-01007-z

iit
ISTITUTO ITALIANO DI TECNOLOGIA

Extensive use of bioRxiv for early sharing of results

Many methodological and data articles

Reproducibility and sharing are two keywords

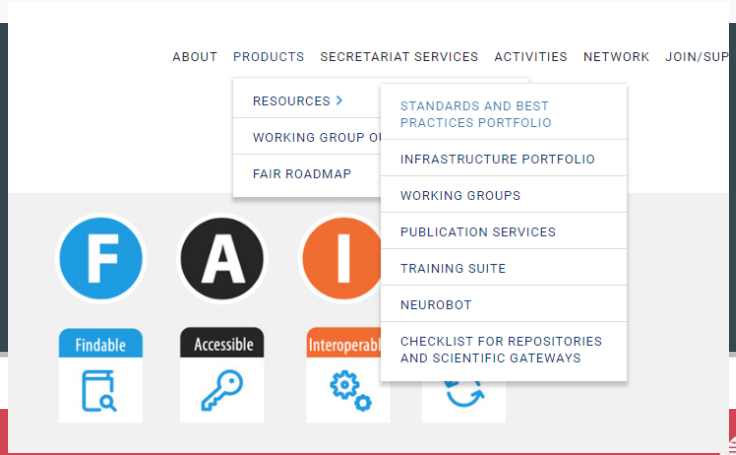
INCF: International Neuroinformatics Coordinating Facility

ISTITUTO ITALIANO DI TECNOLOGIA

incf enabling open and FAIR neuroscience

A collaborative network for open, FAIR, and citable neuroscience

The mission of INCF is to promote the uptake of FAIR data management practices in neuroscience through the development of standards and best practices that support open, FAIR, and citable neuroscience. INCF also provides training on how standards and best practices facilitate reproducibility and enable the sharing of data and code.



INCF supports

- The uptake of standards and best practices for RDM
- The cataloguing and use of shared infrastructures
- Community-building in Open and FAIR neuroscience
- Upskilling and training
- Interoperability of publication and repository services

How to use and participate

incf enabling open and FAIR neuroscience

iit ISTITUTO ITALIANO DI TECNOLOGIA

1. Find and use **open educational resources**

incf Training Space

TrainingSpace (TS) is an online hub that makes multimedia educational content from courses, conference lectures, and laboratory exercises from some of the world's leading neuroscience institutes and societies more accessible to the global neuroscience community. TS provides users with study tracks for self-guided study, tutorials on tools and open science resources for neuroscience research.

incf NeuroStars

Neurostars is an open access question and answer site that serves the INCF network and the global neuroscience community as a forum for knowledge exchange between neuroscience researchers at all levels of expertise, software developers, and infrastructure providers.

Knowledge Space

KnowledgeSpace (KS) is a globally-used, data-driven encyclopedia and search engine for the neuroscience community. Descriptions of neuroscience research concepts, publicly available datasets, publications, and much more can be discovered across multiple resources through KnowledgeSpace.

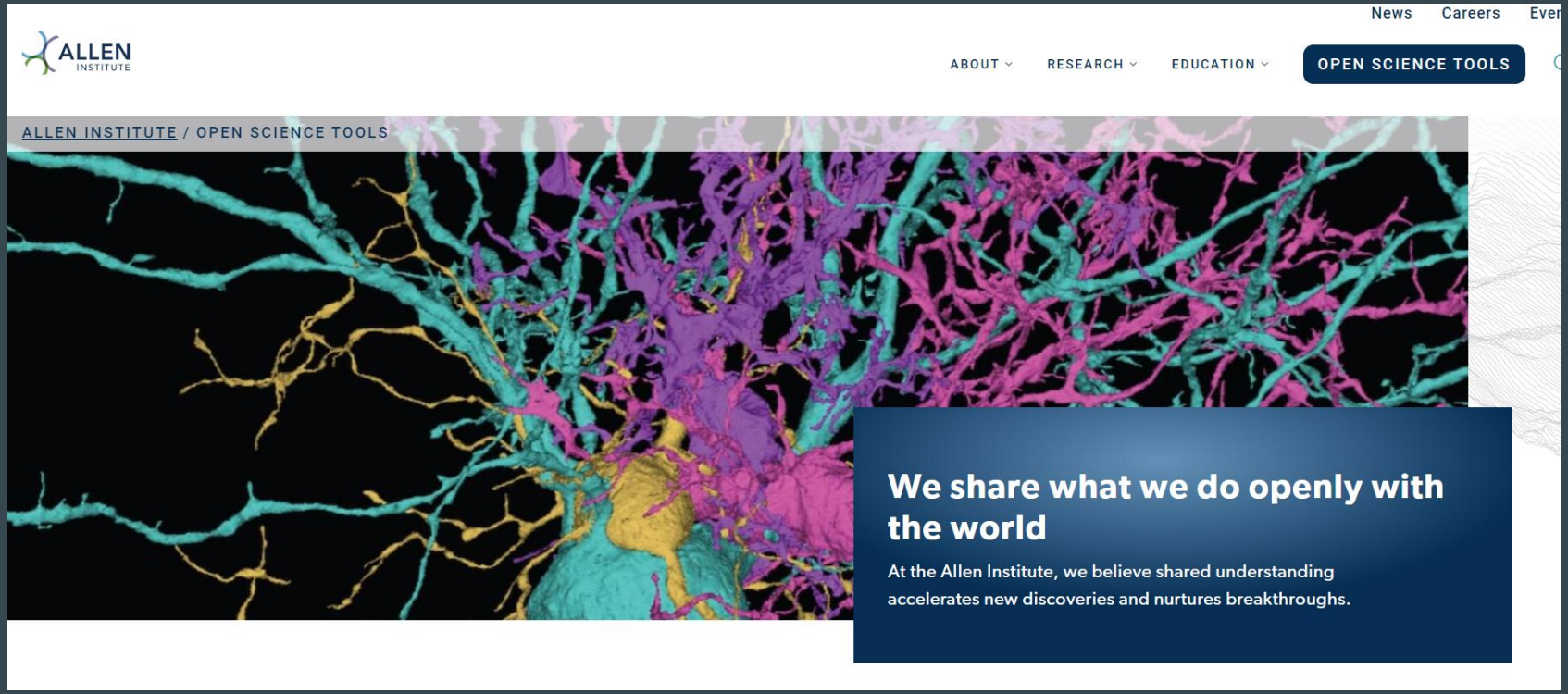
2. Use, review, or submit **standards**

- Submit a Standard
- How standards are endorsed
- Join the Standards reviewer pool
- Recommend a new Standards area

3. Join or propose **working groups**

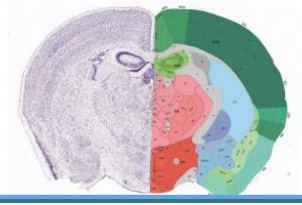
- Join INCF working group
- Requirements for Working Group
- Propose a Working Group

The Allen Institute: shared understanding



Explore our open data portals

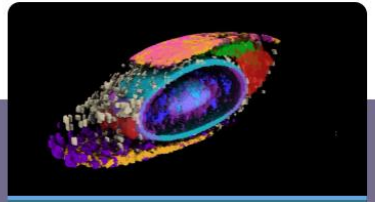
Serving the scientific community is at the center of our mission to expand our understanding of human health and disease.



Allen Brain Map

We publicly release comprehensive, large-scale datasets that investigate the mammalian brain.

VISIT BRAIN MAP ⟶



Allen Cell Explorer

Learn more about our open data, cell lines, software tools, models, plasmids, and more.

VISIT EXPLORER ⟶

Human Immune System Explorer

A portal for exploring data and resources from studies of the human immune

EXPLORE

ALLEN INSTITUTE

iit ISTITUTO ITALIANO DI TECNOLOGIA

Open Data Portals

Education Materials Library

Explore lesson plans and other teaching resources for neuroscience, cell biology, and immunology.

Open educational resources

Where and how to deposit big datasets

ISTITUTO ITALIANO DI TECNOLOGIA

Checklist:

- ✓ If data supports an article, does the journal recommend or require a specific data repository?
- ✓ If not, should I go for a disciplinary repository? Where to find it?

EBRAINS · OpenNEURO · incf enabling open and FAIR neuroscience · re3data.org REGISTRY OF RESEARCH DATA REPOSITORIES

- ✓ If I couldn't find a suitable neuro-repository, is there any generalist or institutional repository I can use?

zenodo · figshare · figshare + · IIT ISTITUTO ITALIANO DI TECNOLOGIA | IIT Dataverse

Important factors:

1. Accessibility level (open, restricted, embargoed)
2. Dataset size
3. Retention period

ISTITUTO ITALIANO DI TECNOLOGIA

Thank you!

We thank Miltenyi Biotec for sponsoring this Symposium