# Open and Transparent: How Research Data Supports Research Integrity

**Dr. Rebecca Grant, Head of Data & Software Publishing, F1000**

F1000Research

# Introduction

**Dr. Rebecca Grant**, Head of Data & Software Publishing, F1000 (part of Taylor & Francis)

- Background in data repositories, libraries and data policy development.
- Qualified Open Data trainer.
- PhD explored the evolution of research data management policy and practice.

# What is research integrity?

Key elements of Research Integrity include:
- **Honesty**
- **Rigour**
- **Transparency** and **open communication**
- The **care** and **respect** of all participants
- **Accountability**

(University of Edinburgh)

Research integrity (RI) is conducting research in a way which allows others to have **trust** and **confidence** in the methods used and the findings in that result.
(Imperial College London)

Research carried out with a high level of integrity upholds values of **honesty**, **rigour**, **transparency** and **open communication**.
(UKRI)

- Intellectual **honesty** in proposing, performing, and reporting research;
- **Accuracy** in representing contributions to research proposals and reports;
- **Fairness** in peer review;
- **Transparency** in conflicts of interest or potential conflicts of interest;
- **Protection** of human subjects in the conduct of research.

(NCBI)

What underpins research integrity?

Trust

Transparency

Openness

Rigour

Honesty


# Open Access publishing

1. Making content **freely available** online to read (OA Libre) and/or making content **reusable** by third parties with little or no restrictions (OA Gratis)

2. **Gold Open Access** – published in fully Open or mixed access (hybrid) journal with an APC (Article Processing Charge)

3. **Green Open Access** – published in “traditional” journal with a copy or pre-print deposited openly elsewhere.

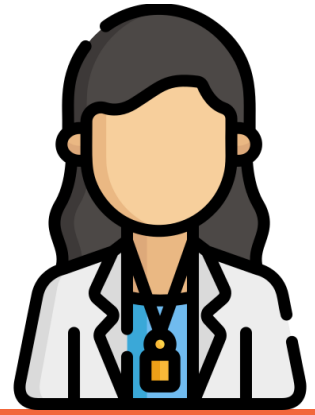
- Open Access offers greater **visibility, transparency** and **impact**.

Articles published Open Access with Taylor & Francis typically receive **32% more citations** and over **6 times as many downloads**.

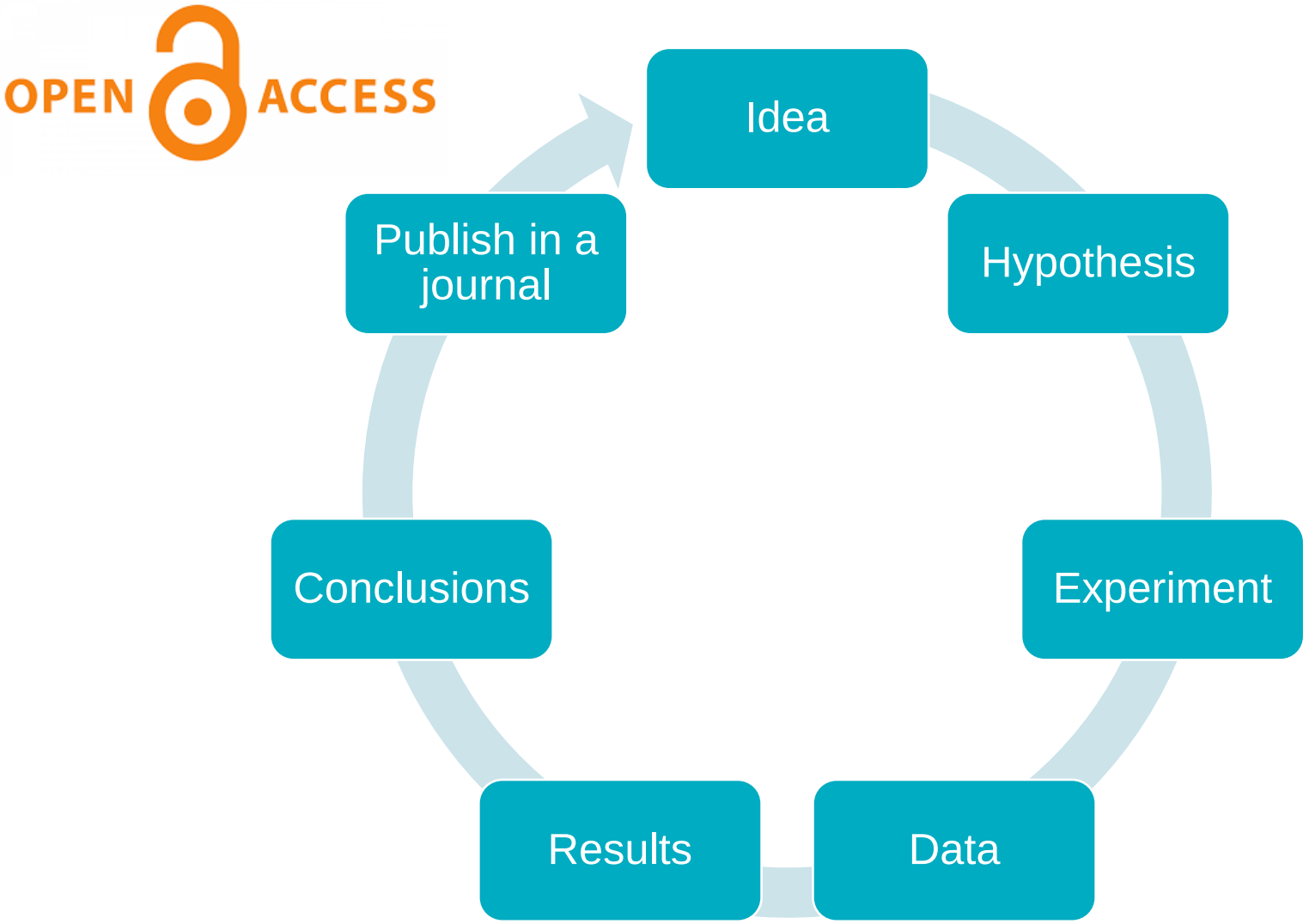
Research integrity

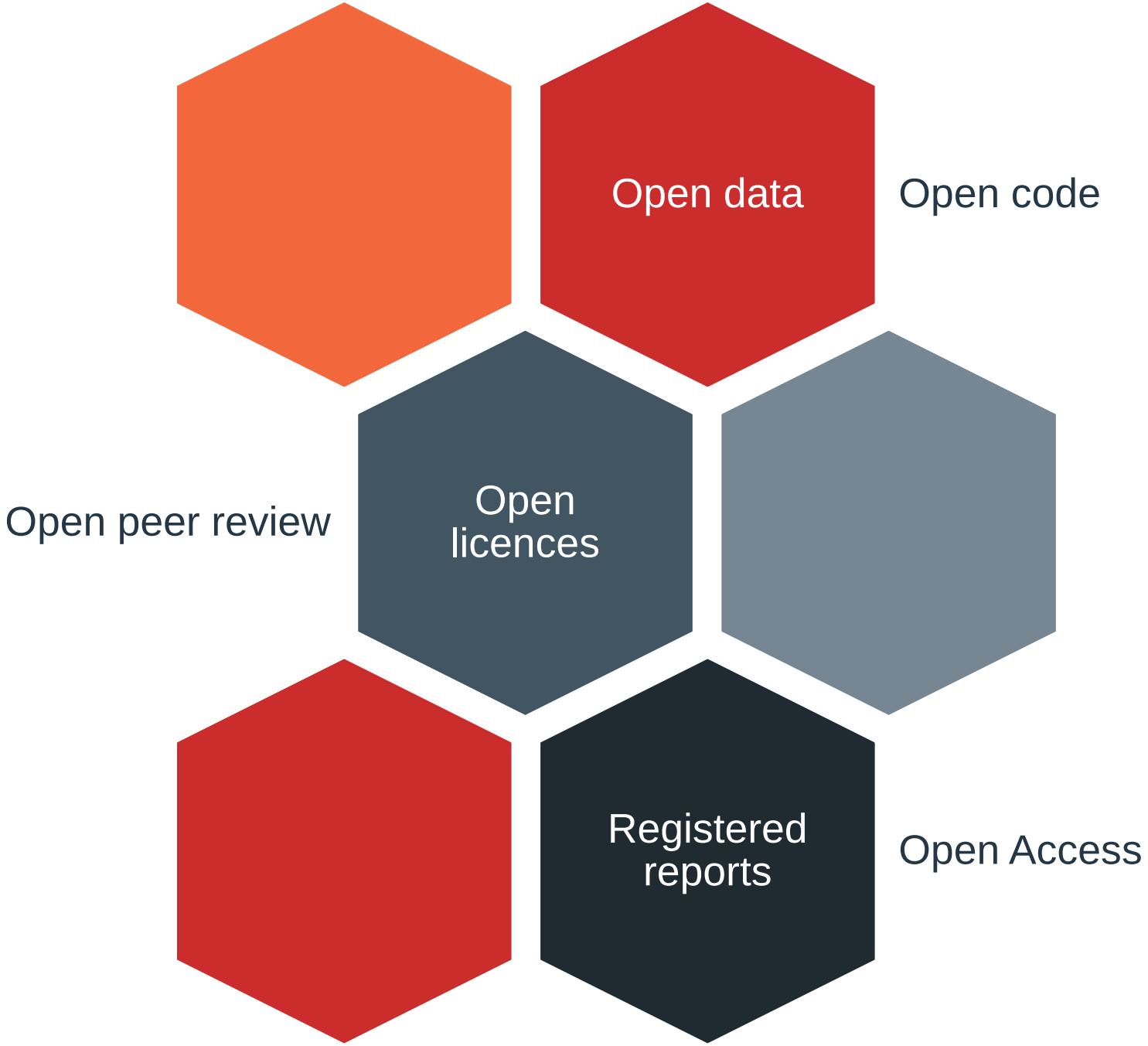
✓ Open Access articles

Anyone can read your published research

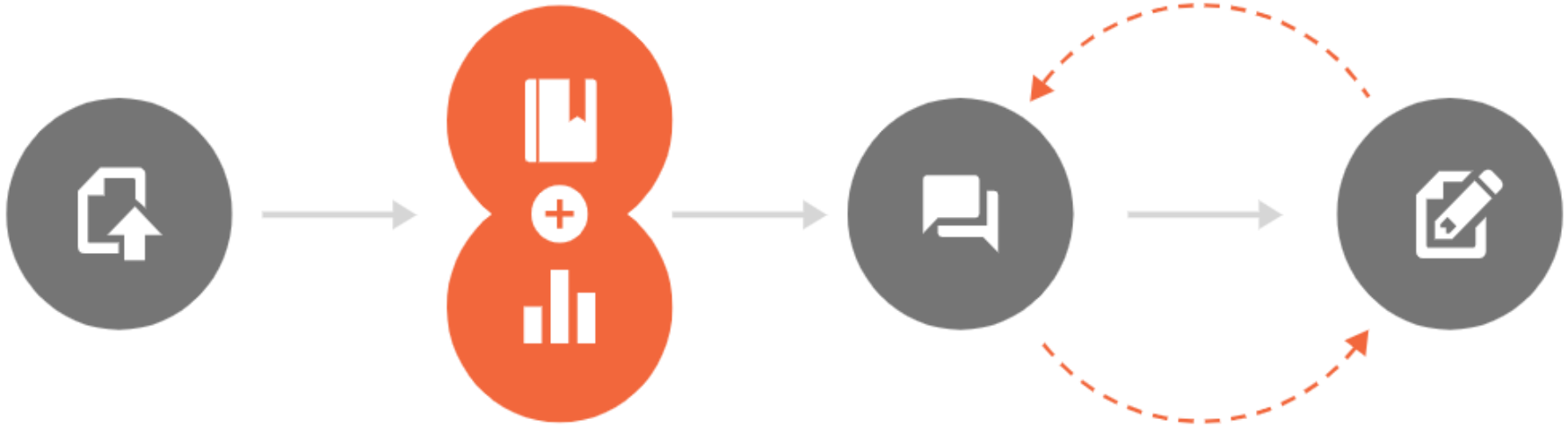


# Open Access to Open Research?




Open data
Open code
Open peer review
Open licences
Registered reports
Open Access

# The F1000 publishing model



**Article Submission**

Submission is via a single-page submission system. The in-house editorial team carries out a comprehensive set of prepublication checks to ensure that all policies and ethical guidelines are adhered to.

**Publication & Data Deposition**

Once the authors have finalised the manuscript, the article is published within a week, enabling immediate viewing and citation.

**Open Peer Review & User Commenting**

Expert reviewers are selected and invited, and their reports and names are published alongside the article, together with the authors' responses and comments from registered users.

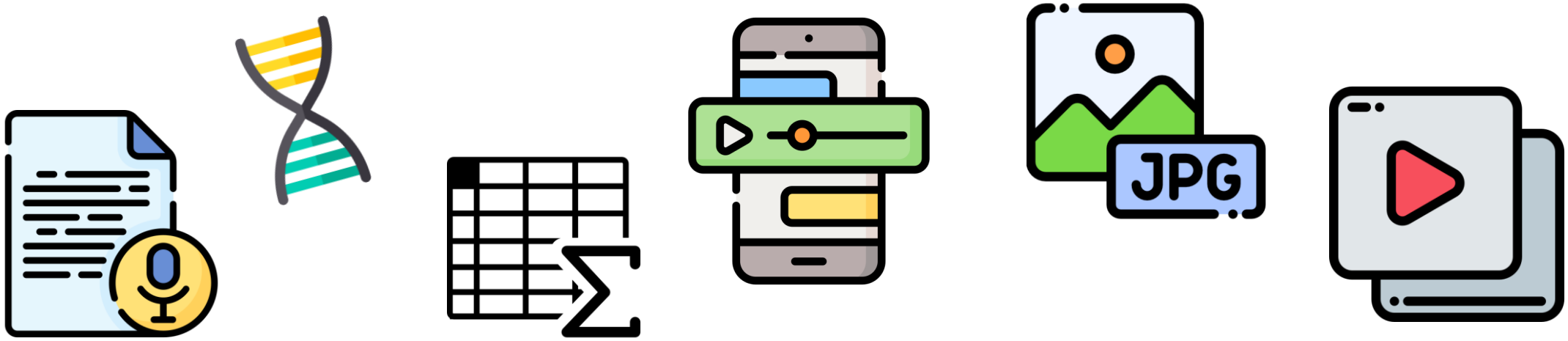
**Article Revision**

Authors are encouraged to publish revised versions of their article. All versions of an article are linked and independently citable. Articles that pass peer review are indexed in external databases such as PubMed, Scopus and Google Scholar.

Open data

# Open Data

1. The research data underpinning your published study, shared openly.
2. “Research data” may be the **input or the output** of your research process, depending on your study design. The format will depend on both your study design and your **research discipline.**



# What is the problem?


**Rejection overruled, retraction ensues when annoyed reviewer does deep dive into data**

**‘We apologize again for the inadvertent mistakes during the assembly of data due to our carelessness’**

**University of New Mexico investigation finds manipulated data and images, prompts retractions**

**Okinawa researcher suspended for faking data denies committing misconduct**

Currently 6300 data-related retractions on the Retraction Watch database

# “No raw data, no science: another possible source of the reproducibility crisis”

Molecular Brain

The Editor in Chief requested that 41 authors supplied data:

- 21 were withdrawn
- 19 provided insufficient data

= 97% could not supply raw data when asked.

**Did the data exist?**

Miyakawa, T. No raw data, no science: another possible source of the reproducibility crisis. *Mol Brain* **13,** 24 (2020). https://doi.org/10.1186/s13041-020-0552-2

# Stakeholder data sharing policies are strengthening

ELSEVIER

F1000Research

SPRINGER NATURE


RDA
RESEARCH DATA ALLIANCE

WILEY

# F1000 Open Data Guidelines

**Accessible data**

- Data availability statement required
- Data must be in a repository

**Reusable data**

- Data must be openly licensed
- Data must be FAIR

**Credit for data**

- Datasets must be cited

Peer reviewers access and assess the data as part of the peer review process

Accessible data

# Writing a data availability statement

- A required section of your manuscript
- It should describe all data underpinning your research and where it can be found.

## Data availability

### Underlying data

Figshare: Underlying data for 'Hematologic and coagulopathy parameter as a survival predictor among moderate to severe COVID-19 patients in non- ICU ward: a single-center study at the main referral hospital in Surabaya, East Java, Indonesia'. https://doi.org/10.6084/m9.figshare.14673060.

The project contains the following underlying data:

- Hema_Coagul_parameter_COVID.xlsx (main data).
- readme.docx (index).

Accessible data

# Depositing data into a repository

- A location on the web for your data to be stored and accessed
- Allows you to provide contextual information so that data can be reused
- Provides a persistent identifier (e.g. a DOI) so that your data can be cited
- Everyone can access your data




# The F1000 recommended repositories list

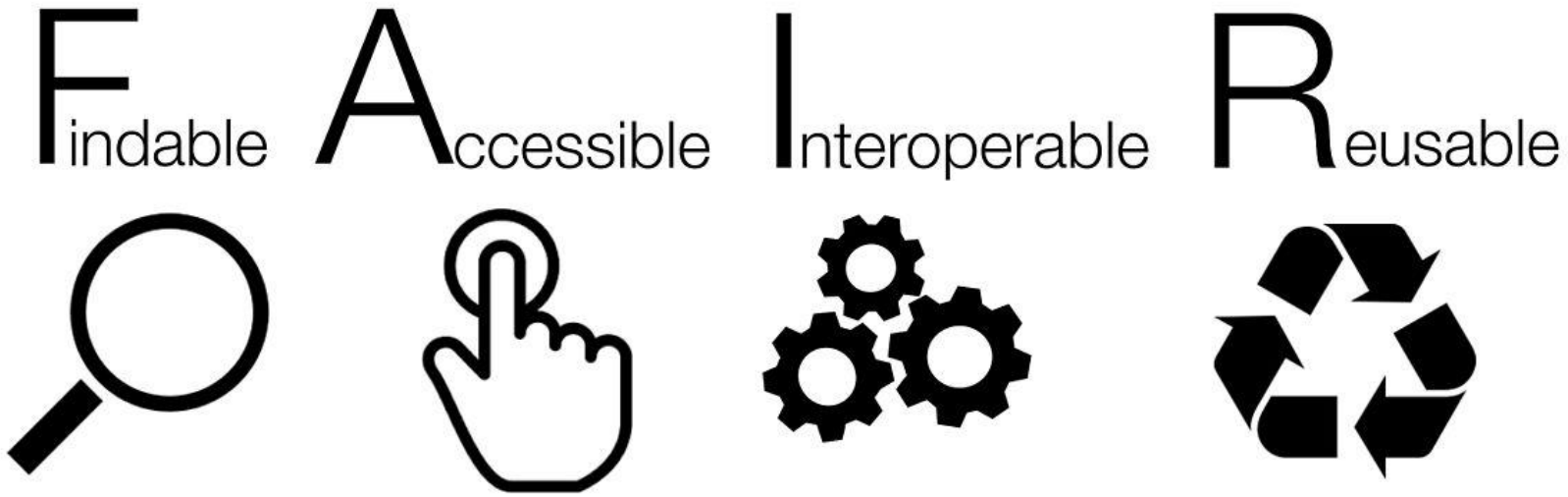
| What type of data | Which repository |
|---|---|
| Any | Dryad |
| Any, but especially data in SAV and POR formats | Dataverse |
| Any | Figshare |
| Any, but especially deposits with mixed data and code | Zenodo |
| Any, but reserved for ISCPR member institutions | Open ICPSR |
| Social and economic data | UK Data Service |

**We find an association between articles that include [data availability] statements that link to data in a repository and up to 25% higher citation impact on average.**

Colavizza G, Hrynaszkiewicz I, Staden I, Whitaker K, McGillivray B (2020) The citation advantage of linking publications to research data. PLoS ONE 15(4): e0230416. https://doi.org/10.1371/journal.pone.0230416

Reusable data

# Standards for high quality, reusable data: FAIR



# Show your data's "FAIR"-ness

1. **Findable**: Include a persistent identifier in your data availability statement, linking to a data repository.

2. **Accessible**: Use a recommended data repository that's accessible on the web.

3. **Interoperable**: Use any applicable reporting standards, vocabularies or ontologies which are common in your discipline.

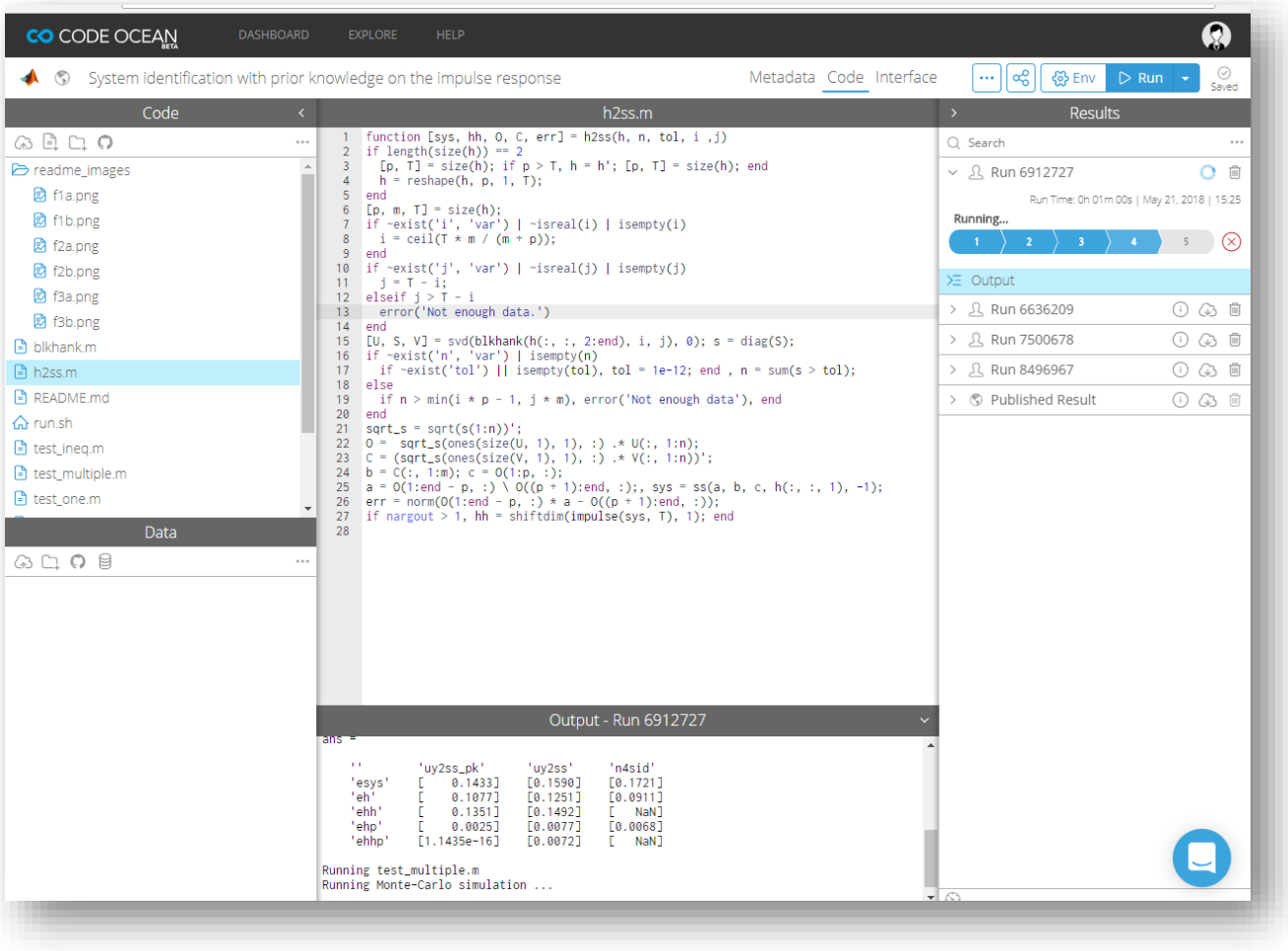
4. **Reusable**: Include a standard licence for your data.

Open code and software

# Open Code and Software

1. Similar to Open Data sharing policies and requirements
2. Can be a key factor in reproducing experiments
3. Allows others to build on existing code and software

# Code Ocean for code sharing

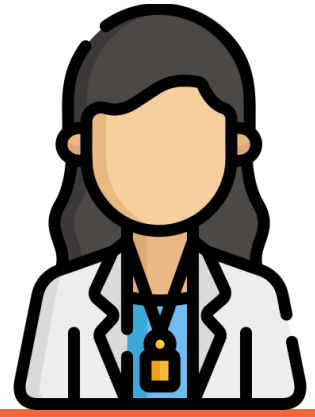



- Code Ocean makes it quick and easy for others to re-run your analysis, and even edit your code to see how results differ by changing parameters.
- Available on F1000 Software tools article and selected T&F journals

# Research integrity

- ✓ Open Access articles
- ✓ Open Data and Open Code

Anyone can access your datasets and code and assess them



Open licences

# Open Licences

1. Licences tell others what can be done with your published data or code.

2. Creative Commons provides a set of standard licences which can be used by any researcher.

# Creative Commons licences

Preferred by publishers

| | | |
|---|---|---|
| **Public Domain Dedication** | CC0 |  |
| **Attribution Only** | CC-BY |  |
| **Attribution-Share Alike** | CC-BY-SA |  |
| **Attribution-Non Commercial** | CC-BY-NC |  |
| **Attribution-Non Commercial-Share Alike** | CC-BY-NC-SA |  |
| **Attribution-No Derivatives** | CC-BY-ND |  |
| **Attribution-Non Commercial-No Derivatives** | CC-BY-NC-ND |  |

More permissive

More restrictive

# Open Licensing in data repositories

zenodo

Early sample tagging and pooling enables simultaneous SARS-CoV-2 detection and variant sequencing at a lower cost than traditional qPCR

**Publication date:**
July 5, 2021

**DOI:**
DOI 10.5281/zenodo.5069979

**Keyword(s):**
apharseq | rna | virus | diagnostics | genotype | variants | sars | corona | covid | covid19 | sars-cov-2

**Related identifiers:**
Supplement to
10.1126/scitranslmed.abj2266 (Journal article)
10.1101/2020.08.08.20170746 (Preprint)

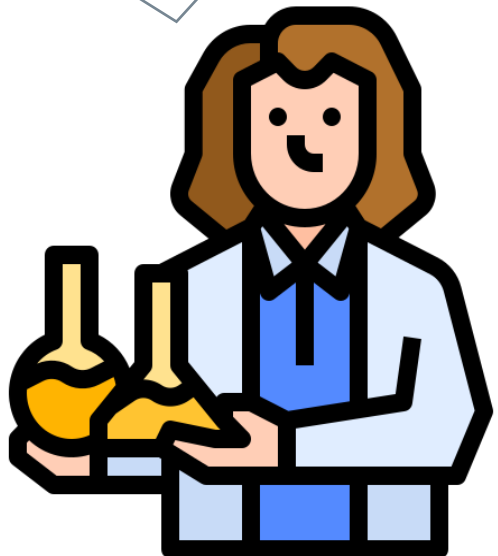
**Communities:**
Coronavirus Disease Research Community - COVID-19

**License (for files):**
Creative Commons Attribution 4.0 International

# Supporting Replicability

I tested the same hypothesis and got the same results as you did

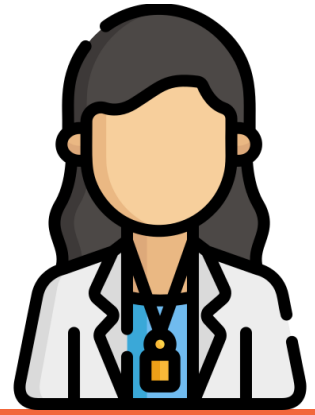
I reused your methods, code and data and got the same results as you did



# Research integrity

- ✓ Open Access articles
- ✓ Open Data and Open Code
- ✓ Open licensing

Anyone can reuse you data and code to reproduce or build on your work




# Registered reports

1. An article type where the hypothesis and study design are submitted to a journal and peer reviewed before the study is conducted.

2. The paper is published regardless of “negative” or null results.

# Key issues in research integrity

"Publish or perish"

Publication bias towards positive results

P-hacking and selective reporting

# Registered Reports

Registered Reports break up the research process to increase transparency and reduce waste.

REVISED

# Stage 1 Registered Report: Effect of deficient phagocytosis on neuronal survival and neurological outcome after temporary middle cerebral artery occlusion (tMCAo) [version 3; peer review: 2 approved]

Julius V. Emmrich [1], Jonas J. Neher[2,3], Philipp Boehm-Sturm[1,4], Matthias Endres[1,5], Ulrich Dirnagl [1,5,6] Christoph Harms[1,6]

## Methods

### Animals, housing and husbandry

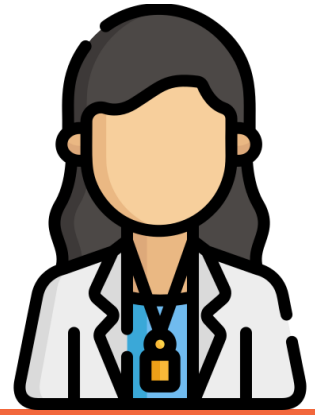
All animal experiments will be performed in accordance with local regulations, and have been approved by the Berlin governmental authorities (Landesamt für Gesundtheit und Soziales, LaGeSo), approval number G057/16.

Male C57BL/6NCrl mice will be derived from Charles River at the age of 8 weeks. Phagocytosis-deficient *Mertk* (Jax: B6;129-*Mertk*$^{tm1Grl}$/J) and *Mfge8* (from C. Théry, INSERM 932, France)[18] knockout mice will be derived from The Jackson Laboratory and Hertie Institute for Clinical Brain Research, respectively, and bred locally. Male homozygous *Mertk* and *Mfge8* knockout mice and their homozygous wildtype littermates will be used in experiments at the age of 10 – 12 weeks. Animals will be group-housed with *ad libitum* access to food and water and cages will be equipped with environmental enrichment tools (red transparent plastic nest box and brown paper towels). Animals will be kept in specific pathogen free (SPF) conditions under a 12 h light/dark cycle (lights on: 8am; lights off: 8pm). Room temperature will be maintained at 22 ± 1°C.

# Research integrity

- ✓ Open Access articles
- ✓ Open Data and Open Code
- ✓ Open Licensing
- ✓ Support publication of null results

Researchers are not incentivised to manipulate data to show positive results



## What underpins research integrity?

Trust

Transparency

Openness

Rigour

Honesty

## How can publishers support research integrity?

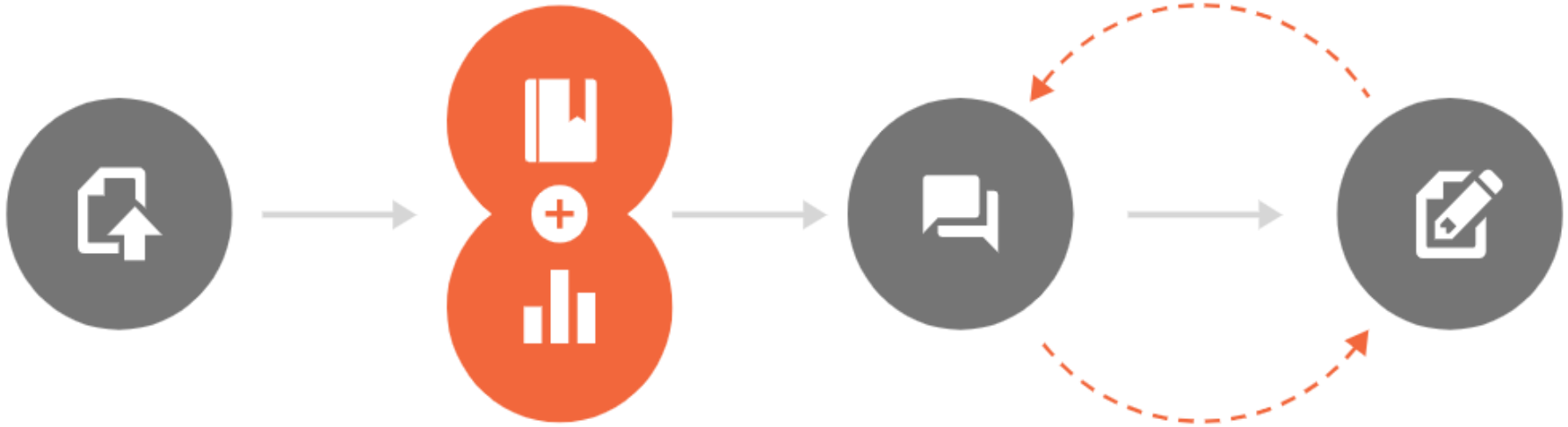
Support a range of accessible outputs

Allow peer review of all outputs

Encourage open licensing and reuse

Address publication bias

# The F1000 publishing model



**Article Submission**

Submission is via a single-page submission system. The in-house editorial team carries out a comprehensive set of prepublication checks to ensure that all policies and ethical guidelines are adhered to.

**Publication & Data Deposition**

Once the authors have finalised the manuscript, the article is published within a week, enabling immediate viewing and citation.

**Open Peer Review & User Commenting**

Expert reviewers are selected and invited, and their reports and names are published alongside the article, together with the authors' responses and comments from registered users.

**Article Revision**

Authors are encouraged to publish revised versions of their article. All versions of an article are linked and independently citable. Articles that pass peer review are indexed in external databases such as PubMed, Scopus and Google Scholar.

# What does this mean for researchers?

## New perspectives?

- Open sharing – visibility of data, code
- Open licensing – expectation of reuse
- Registered reports and data peer review

## New practices?

- Depositing data into repositories before publication
- Writing data availability statements

## New ways of receiving credit?

- Data citations
- Code and software citations

# Any questions?


**Rebecca.grant@f1000.com**

**@beck_grant**

**f1000research.com | think.f1000research.com/open-data**


F1000Research

Image credits:

Freepik, Good Ware and UltimateArm via Flaticon.com