Mohamed bin Zayed University of Artificial Intelligence

Fall 2026, MBZUAI

CBIO7501 - On the Art of Reproducible & Open Science

Open Science – Lecture 2

Sharing papers, data, code and methods so that others can check, reuse and build build

Aziz Khan, PhD
Assistant Professor of Computational Biology, MBZUAI - khanlab.bio
aziz.khan@mbzuai.ac.ae

Mohamed bin Zayed University of Artificial Intelligence

Week 1 learning goals

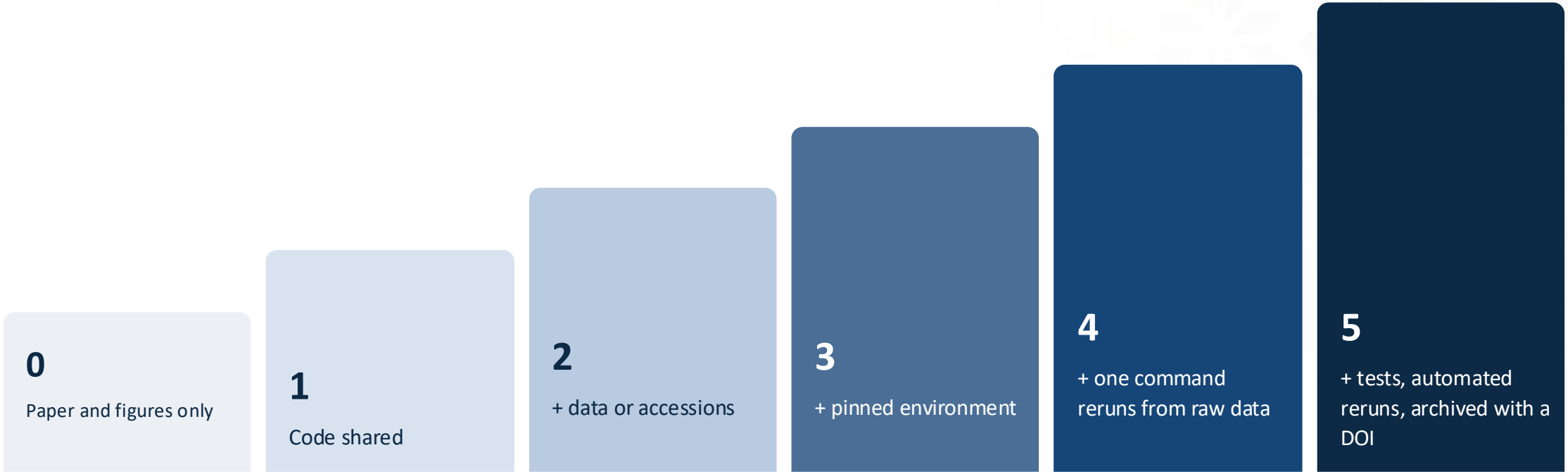
Monday · reproducibility

1. Explain what reproducibility is and how it differs from replicability
2. Explain the factors and reasons behind irreproducible research
3. List and define tools and techniques that support each level of reproducibility

Tuesday · open science

4. Understand the motivation for open science
5. Explain the core FAIR data principles and open licensing
6. Use checklists and resources to make your own work more open

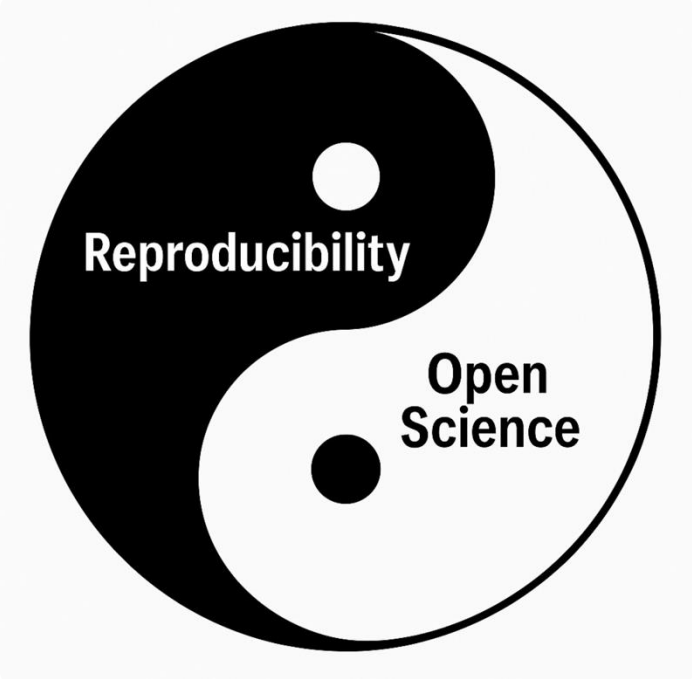
Where did your papers land on the ladder?



What stopped you going one rung higher? That barrier is today’s agenda.

Mohamed bin Zayed University of Artificial Intelligence

Reproducibility and open science: the yin and yang of trustworthy science



Reproducibility (yin): inward-facing, about rigour, verification and the integrity of methods and results.

Open science (yang): outward-facing, about access, sharing and engaging the wider community.

Reproducibility without openness

Results might be verifiable, but hidden and inaccessible.

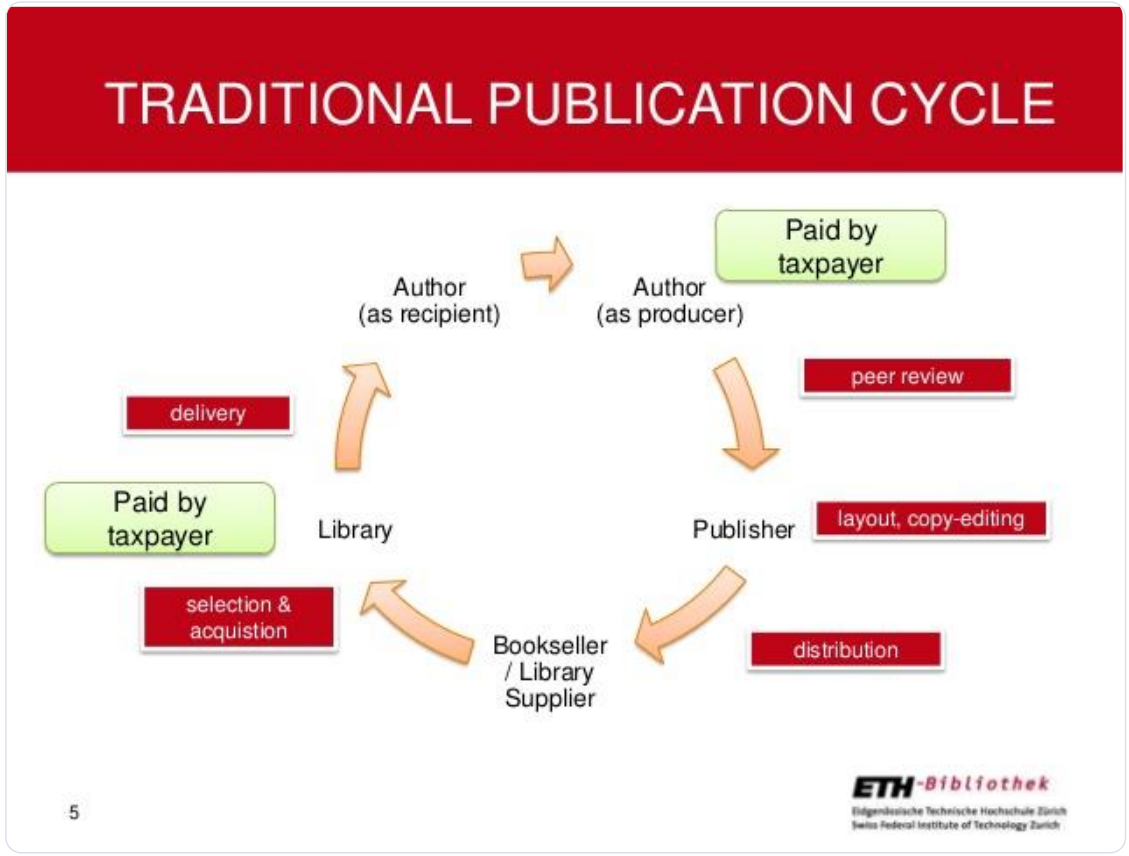
Openness without reproducibility

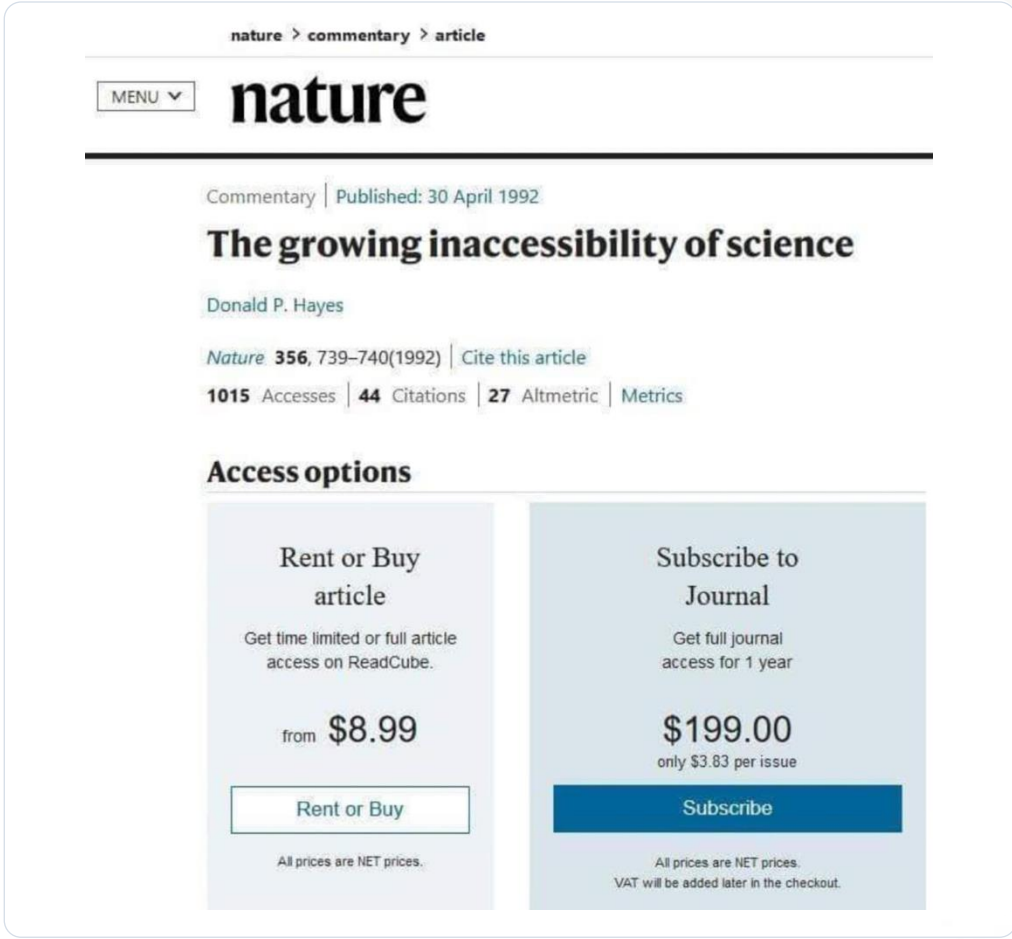
Research is shared, but possibly unreliable.

Distinct but complementary: together they make science trustworthy and useful.

Mohamed bin Zayed University of Artificial Intelligence

Traditional academic publishing is broken





Mohamed bin Zayed University of Artificial Intelligence

A global definition, adopted by 193 countries in 2021

IN SHORT

Making scientific knowledge openly available, accessible and reusable for everyone, and opening the processes that create, evaluate and communicate it to society.

Paraphrased from the UNESCO Recommendation on Open Science

unesco

Four pillars

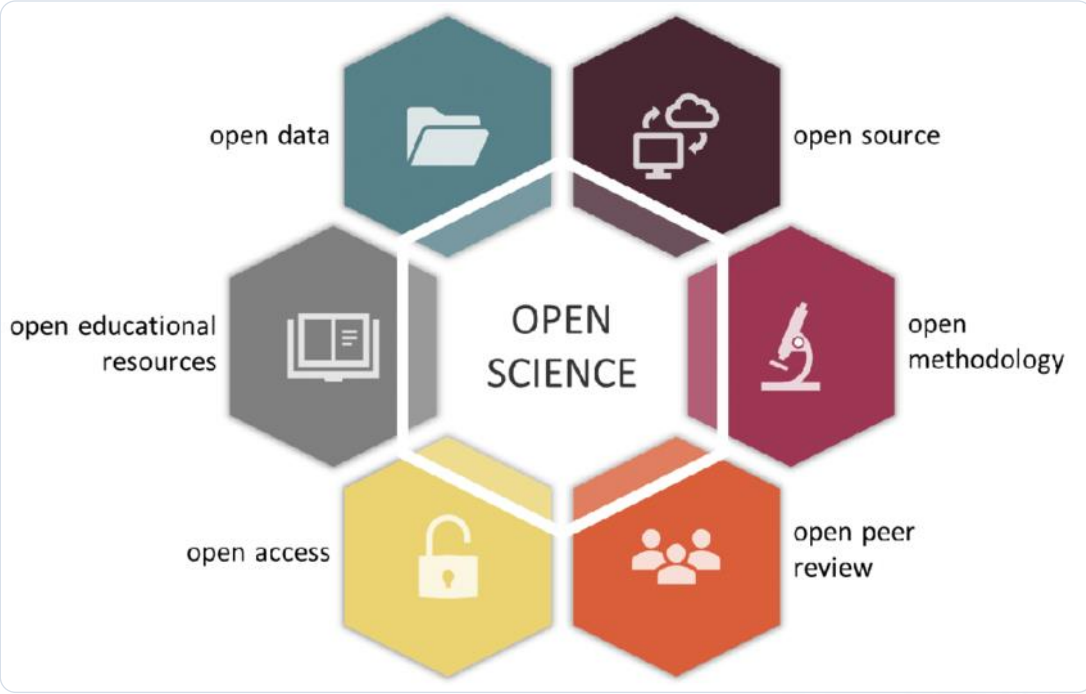
- Open scientific knowledge
- Open science infrastructures
- Open engagement of societal actors
- Open dialogue with other knowledge systems

Mohamed bin Zayed University of Artificial Intelligence

What is open science?

> *"Open science involves opening up the processes of creating, evaluating, sharing, exploring and storing scientific knowledge, practices and perspectives."*

A global movement to make scientific research and its outcomes freely accessible to everyone, through practices such as data sharing and preregistration.



Mohamed bin Zayed University of Artificial Intelligence

Each open practice fills a gap we saw last week

Open access

Anyone can read the claim and its methods

Open data

Others can rerun and re-analyse

Open code

The analysis is executable, not only described

Open methods

Plans and protocols on record before results

Open evaluation

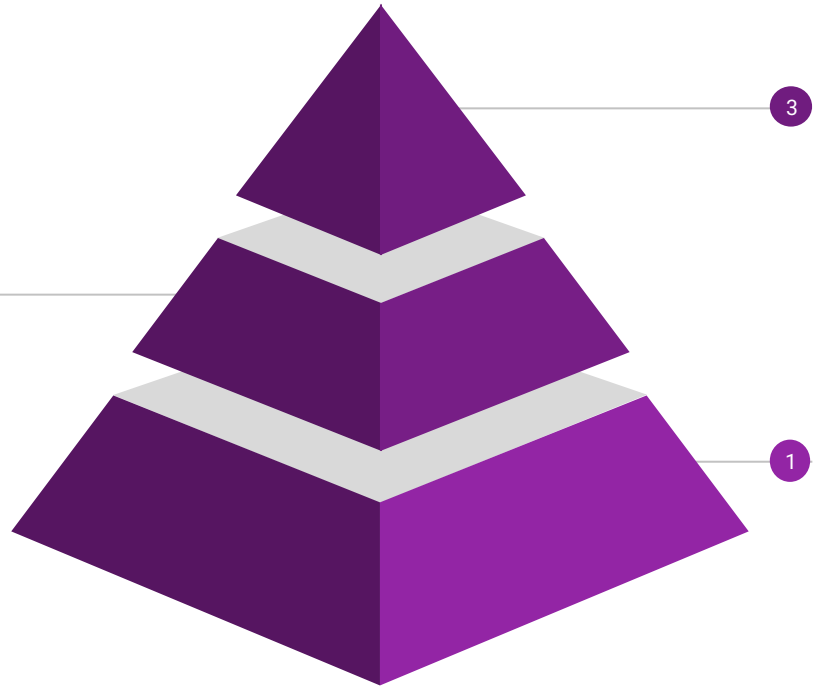
Reviews published; scrutiny continues after publication

Open education

Teaching materials others can reuse and adapt

Mohamed bin Zayed University of Artificial Intelligence

Three key components of openly shared research



Open Results
Open access to research results is good for science and society. An open paper without code/data is simply an advertisement

Open Code
Code and methods are essential for open and reproducible research

FAIR Data
FAIR data sharing is the foundation for open science and reproducibility

Mohamed bin Zayed University of Artificial Intelligence

Benefits of open science



1 · Good for science

Outputs are open to all, not stuck behind paywalls, levelling the field and field and speeding up discovery.

2 · Good for society

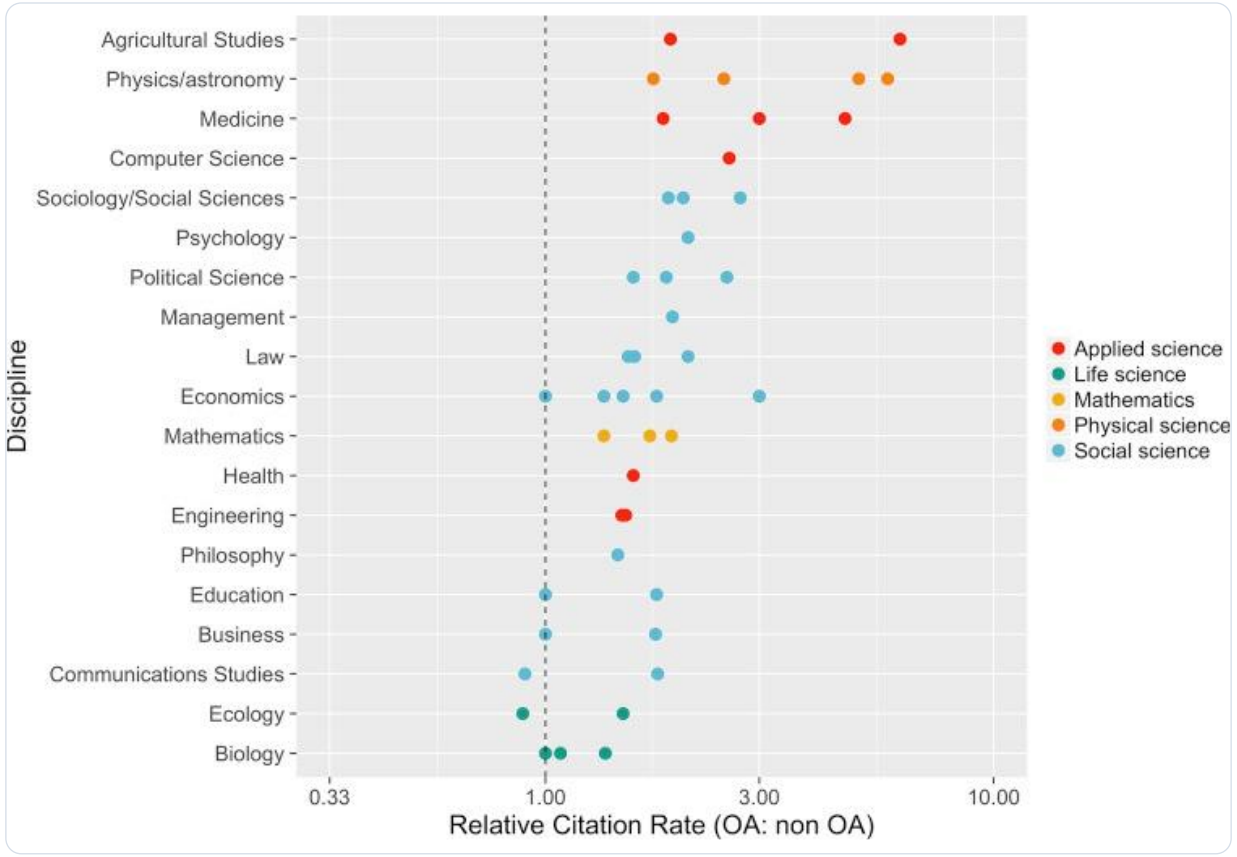
A better return on public investment in research, and faster answers to answers to shared problems.

3 · Good for scientists

Data and code become citable outputs; work is more visible, more cited more cited and opens new collaborations.

Mohamed bin Zayed University of Artificial Intelligence

Open access articles are cited more



Across 19 fields, open access articles are cited more on average than paywalled ones.

Each point is the ratio of mean citations for OA to non-OA articles from one study; values above 1 favour open access.

Correlation, not proof of cause, but consistent across disciplines.

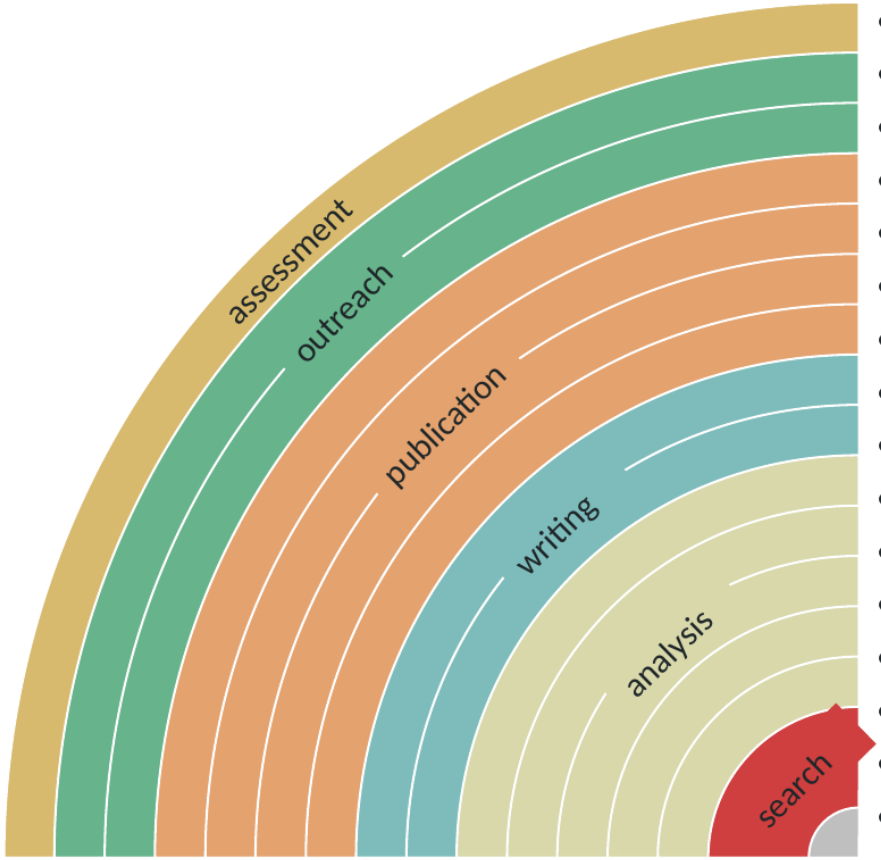
Mohamed bin Zayed University of Artificial Intelligence

As open as possible, as closed as necessary.

Openness is the default, and every restriction needs a stated reason: consent, privacy, security or law.

Mohamed bin Zayed University of Artificial Intelligence

You can make every stage of your workflow more open



- adding alternative evaluation, e.g. with altmetrics
- communicating through social media, e.g. BlueSky
- 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 Peerage of Science
- sharing preprints, e.g. at arXiv, bioRxiv or OSF
- using actionable formats, e.g. with Jupyter, Quarto
- open XML-drafting e.g. at Overleaf or Authorea
- sharing protocols & workfl. e.g. at MyExperiment
- sharing notebooks e.g. at OpenNotebookScience
- sharing code e.g. at GitHub with GNU/Open license
- sharing data/models, e.g. Zenodo, HuggingFace
- 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

Mohamed bin Zayed University of Artificial Intelligence

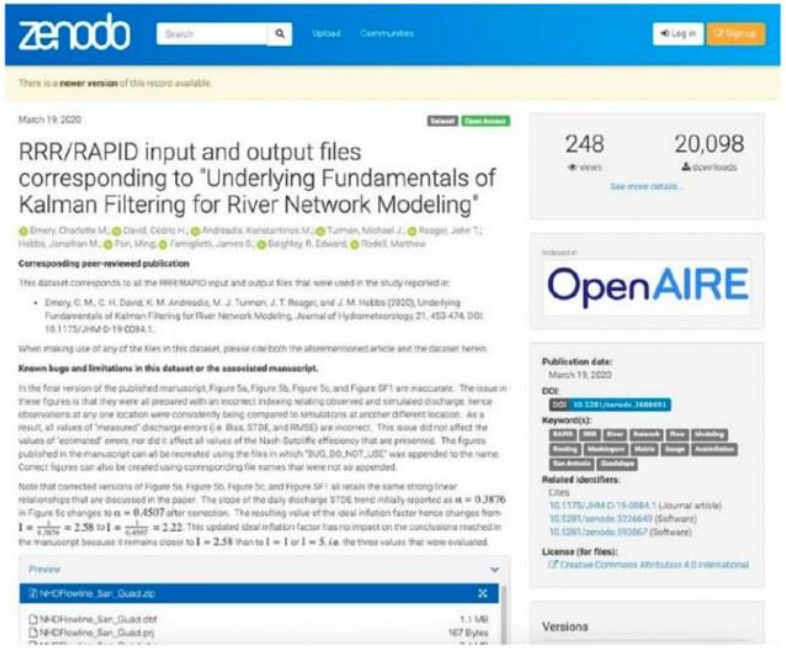
Hypothetical research workflows



Mohamed bin Zayed University of Artificial Intelligence

Sharing every research component: an example

Data



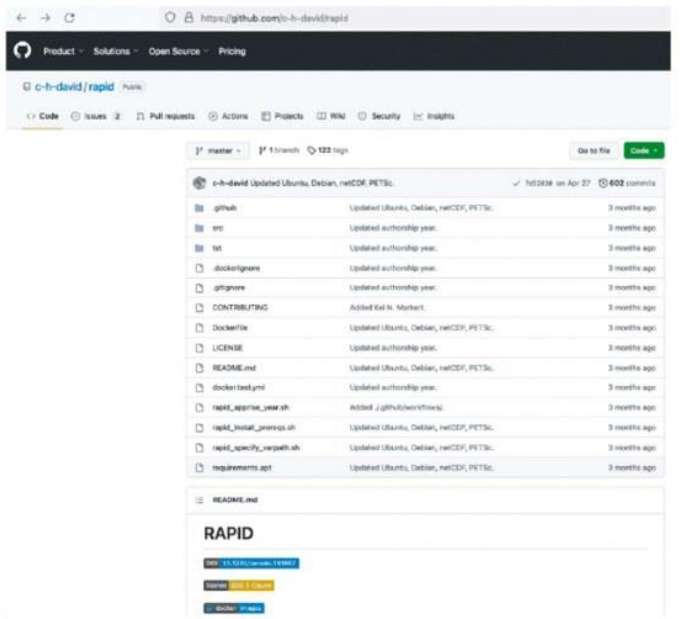
Zenodo Example Screenshot. (2023). Zenodo https://doi.org/10.5281/zenodo.3688690

Paper



Publication Example Screenshot. (2023). https://doi.org/10.1175/JHM-D-19-0084.1

Software



GitHub Example Screenshot. (2023). Github. This version: https://github.com/c-h-david/rapid

Mohamed bin Zayed University of Artificial Intelligence

Another example: the James Webb Space Telescope



Planning for Open Science

The JWST Transiting Exoplanet Community Early Release Science Program Publication Policy

Version 1.0, 5 March 2018

All ERS Team members of the Transiting ERS Program (ERS 1366) are constrained by this policy. The policy covers publication of the ERS core activities. The policies are meant to enable four main core values and requirements of the program:

Core Values and Requirements:

The JWST Transiting Exoplanet Community Early Release Science Program Publication Policy [Screenshot]. (2023). GitHub. https://ers-transit.github.io/documents/publication-policy-v1.0.pdf

Open Source Software

JOSS
The Journal of Open Source Software

Eureka!: An End-to-End Pipeline for JWST Time-Series Observations

Taylor J. Bell[1], Eva-Maria Ahrer[2], Jonathan Brande[3], Aarynn L. Carter[4], Adina D. Feinstein[5], Giannina Guzman Caloca[6], Megan Mansfield[7,8], Sebastian Zieba[9], Caroline Piaulet[10], Björn Benneke[10], Joseph Filippazzo[11], Erin M. May[12], Pierre-Alexis Roy[10], Laura Kreidberg[9], and Kevin B. Stevenson[12]

Eureka!: An End-to-End Pipeline for JWST Time-Series Observations [Screenshot]. (2023). The Journal of Open Source Software. https://joss.theoj.org/papers/10.21105/joss.04503

Open Access to Results



The JWST Transiting Exoplanet Community Early Release Science Program [Screenshot]. (2023). Zenodo https://zenodo.org/communities/ers-transit

Mohamed bin Zayed University of Artificial Intelligence

Five colours of open access

Route	How it works	Who pays	Reuse licence
Gold	Fully open-access journal	Usually the author's fee (APC)	Open, often CC BY
Diamond	Open-access journal with no author fees	Institutions, societies, funders	Open, often CC BY
Hybrid	Subscription journal; pay to open one article	Author fee on top of subscriptions	Open for that article
Green	Accepted manuscript deposited in a repository	No one directly	Depends; may follow an embargo
Bronze	Free to read on the publisher's site	Publisher's choice	None: can be withdrawn

Mohamed bin Zayed University of Artificial Intelligence

Preprints: share first, review in the open

1991

arXiv: physics, maths, then computer science and ML

2013

bioRxiv: the life sciences

2019

medRxiv: health sciences, with extra screening

Why post one

Results out months earlier; a dated DOI that establishes priority; feedback before the version of record

What to watch

Not yet peer reviewed; check the target journal's policy; post data and code with it, not later

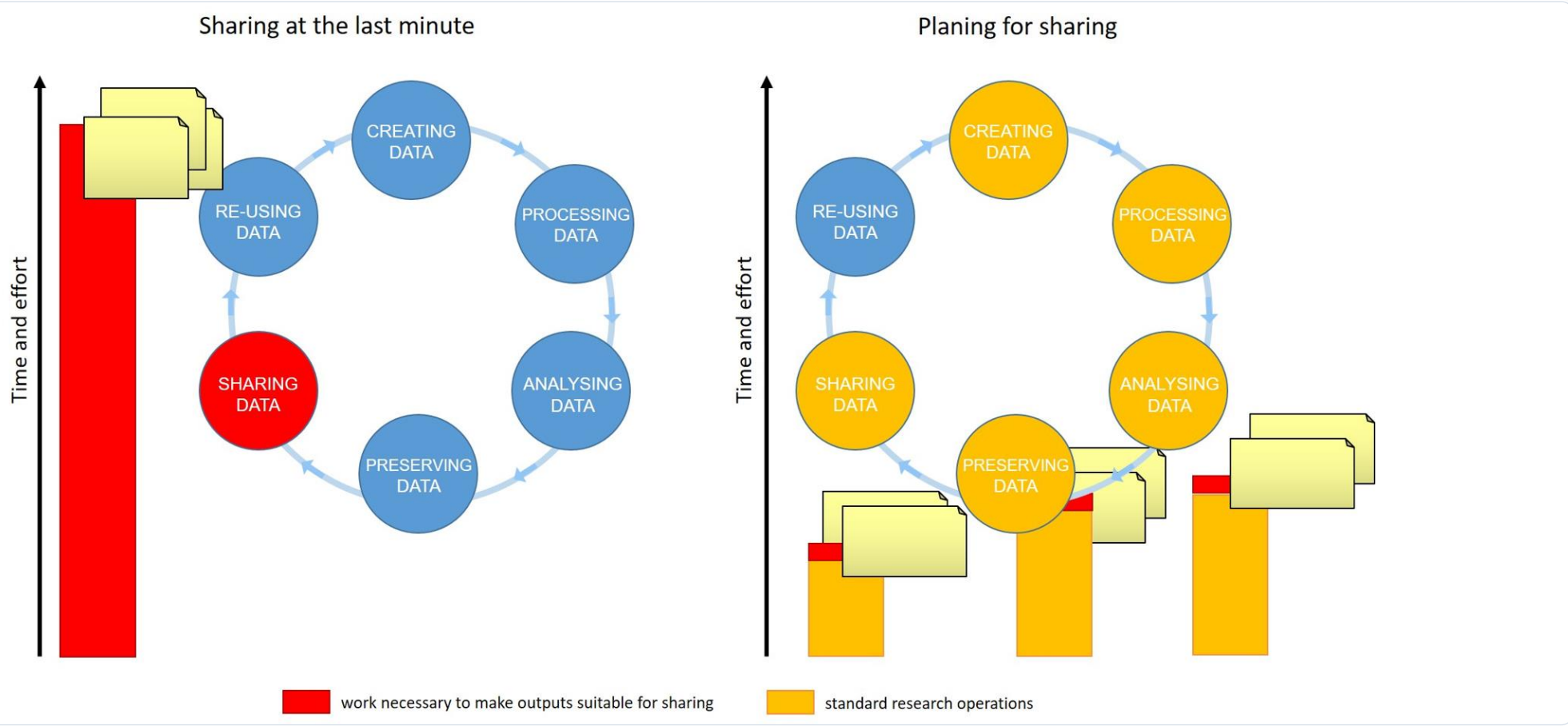
Mohamed bin Zayed University of Artificial Intelligence

Funders increasingly require openness, and the rules move

Year	Policy	In brief
2008	NIH Public Access Policy	Papers deposited in PubMed Central; up to 12-month embargo
2018	Plan S launched (cOAlition S)	Mainly European funders; immediate open access, in force from 2021
2021	UNESCO Recommendation	First global standard-setting text on open science
2022	US OSTP “ Nelson memo”	Asked all federal funders for zero-embargo access to papers and data
2023	NIH Data Management and Sharing Policy	A data-sharing plan for all NIH-funded research
2025	NIH zero-embargo policy in force	Accepted manuscripts public in PMC on publication day, from 1 July
2026	Nelson memo under review	Administration weighing repeal; agency policies such as NIH's still apply

Mohamed bin Zayed University of Artificial Intelligence

Data management is a continuous process: plan early



Mohamed bin Zayed University of Artificial Intelligence

FAIR data: usable by people and by machines

F

Findable

A persistent identifier and rich metadata, indexed where people search

A

Accessible

Retrievable by that identifier through a standard protocol; metadata outlives the data

I

Interoperable

Community formats and vocabularies: BED, VCF, BAM, ontology terms

R

Reusable

A clear licence, documented provenance, and community standards met

Mohamed bin Zayed University of Artificial Intelligence

FAIR in practice

Findable

Enough metadata, a unique persistent identifier, and an entry in a searchable resource.

Example: the dataset has a DOI from Zenodo.

Accessible

Metadata and data readable by humans and machines, held in a trusted repository.

Example: downloadable in a browser or by script.

Interoperable

A common structure and recognised vocabularies.

Do: open formats such as .csv or .tsv; APIs where possible. **Never** share data tables as Word or PDF.

Reusable

Clear licence, clear provenance, community standards.

Do: add a licence, a README, and descriptive column headers.

Mohamed bin Zayed University of Artificial Intelligence

FAIR is not the same as open

Open	Registered	Controlled	Closed
Anyone downloads, no sign-in	Sign in and accept terms of use	A data access committee reviews each request	Not shared; metadata may still be public
Cell lines, model organisms, consented public cohorts	Some summary-level human data	Individual-level human genomes, clinical data	No consent for sharing; legal limits

Controlled-access data can still be FAIR: findable metadata, a clear access route and a licence for what is allowed.

Mohamed bin Zayed University of Artificial Intelligence

Use a domain repository first, a general one second

What you have	Where it usually goes
Raw sequencing reads	SRA, ENA or DDBJ (they mirror each other)
Functional genomics experiments	GEO or ArrayExpress: raw reads plus processed files
Human data with consent limits	EGA or dbGaP, under controlled access
Variants	dbSNP, ClinVar, the European Variation Archive
Proteomics; structures	PRIDE; PDB
Everything else, plus code and models	Zenodo, figshare or Dryad, each with a DOI

Not sure? FAIRsharing.org lists repositories and standards by data type, and many journals publish their own lists.

Mohamed bin Zayed University of Artificial Intelligence

Data and code availability: be like this, not like this

Be like this

Data availability

Metadata and cellranger outputs are available at Zenodo (https://doi.org/10.5281/zenodo.6401895). Expressed cellular barcodes (ECB) sequencing data are available at bioProject ID (PRJNA838456). Genomic sequencing and scRNA-seq data are available at dbGAP under accession no. phs003249.v1. Source data are provided with this paper.

Code availability

The computational methods, procedures and analyses summarized above are implemented in custom R and python, and bash scripts are available via the Curtis Lab: https://github.com/cancersysbio/gastric_organoid_evolution.

Not like this



Juan Nunez-Iglesias
@jnuneziglesias

sigh why is this still a thing in 2021?

Data availability

All data that support the findings of this study are available from the corresponding author upon reasonable request. Source data are provided with this paper.

Code availability

The code used to analyze the data is available from the corresponding author upon reasonable request.

Mohamed bin Zayed University of Artificial Intelligence

Making the gapminder data more FAIR

r-novice-gapminder-files / data / **gapminder-FiveYearData.csv**



sritchie73 moved data into data directory

Preview | Code | Blame | 1705 lines (1705 loc) · 80 KB

Search this file

1	country	year	pop	continent	lifeExp	gdpPercap
2	Afghanistan	1952	8425333	Asia	28.801	779.4453145
3	Afghanistan	1957	9240934	Asia	30.332	820.8530296
4	Afghanistan	1962	10267083	Asia	31.997	853.10071

The dataset includes variables as follows and some are renamed to using FAIR principles:

Column name	Description
`country`	Name of the country where the data was collected.
`yeay`	Year when the observation was recorded.
`population`	Total population of the country in the given year.
`continent`	Continent where the country is located.
`life_expectancy`	Average life expectancy at birth (in years) for the given country and year.
`gdp_per_capita`	GDP per capita (Gross Domestic Product per person, in USD).

Mohamed bin Zayed University of Artificial Intelligence

“Available on request” rarely is, and sharing tends to pay

6.8%

of 1,792 papers whose authors said data were available on request actually provided them when asked

Gabelica, Bojčić & Puljak, J Clin Epidemiol (2022)

+69%

citations for cancer microarray trials with public data (85 trials, adjusted for journal and date)

Piwowar, Day & Fridsma, PLoS ONE 2:e308 (2007)

~25%

more citations when the paper links to its data in a repository (≈532,000 PLOS and BMC articles)

Colavizza et al., PLoS ONE 15:e0230416 (2020)

Citation studies show association, not cause; depositing at publication is what matters.

Mohamed bin Zayed University of Artificial Intelligence

Code ensures the reproducibility of results

- Use version control (Git), documentation, and tests with GitHub Actions for continuous integration
- Add a **LICENSE** file: MIT, GPL or another open licence
- Add a **CITATION.cff** so others cite it correctly
- Add **CONTRIBUTING.md** and **CODE_OF_CONDUCT.md**

Example repository

```
khan-lab/ENCODEfetch
├── LICENSE
├── CITATION.cff
├── CONTRIBUTING.md
├── CODE_OF_CONDUCT.md
├── README.md
└── .github/workflows/
```

github.com/khan-lab/ENCODEfetch

Mohamed bin Zayed University of Artificial Intelligence

No licence means no one may legally reuse it

For	Common choice	What it allows
Code	MIT, BSD-3, Apache-2.0	Permissive: any use, including commercial, with attribution
Code	GPL-3.0	Copyleft: derived software must stay open under the same terms
Text, figures	CC BY 4.0	Any reuse with credit; NC or ND variants block much reuse
Data	CC0 or CC BY 4.0	CC0 waives rights entirely, which makes combining datasets simplest

Put a LICENSE file in the repository root on day one. Check your institution's IP policy before choosing.

Mohamed bin Zayed University of Artificial Intelligence

From permissive to restrictive licences

Permissive

License	Type of use	Symbols
Public Domain CC0	The copyright holder has voluntarily waived their rights and the work can be freely used.	PUBLIC DOMAIN
Attribution (BY)	You must credit the creator, the title and the license the work is under.	BY
Share Alike (SA)	Any new material produced using the work must be made available under the same license as the original work.	BY SA
Non Commercial (NC)	The work cannot be used for commercial purposes.	BY NC
No Derivatives (ND)	The work can only be used exactly how it is. The work cannot be adapted or modified in any way.	BY ND

Restrictive

Image Credit: NASA; License to share: How the Creative Commons licensing system encourages the remixing and reuse of published materials.

A licence states the rights of the author and the user of a paper, dataset or piece of software.

- It ensures authors receive credit and control
- It allows others to reuse the work legally

When practising open science, choose the licence with the fewest restrictions that still meets your needs.

Mohamed bin Zayed University of Artificial Intelligence

Make software citable, and cite the software you use

Archive each release

Link GitHub to Zenodo: every tagged release gets its own DOI. Software Heritage keeps public source code long term.

Tell people how to cite it

A CITATION.cff file makes GitHub show a “Cite this repository” button.

Cite tools with versions

Name the version and cite the paper or DOI for every tool, so its developers get credit.

CITATION.CFF

```
cff-version: 1.2.0
title: peakmerge
version: 1.0.2
doi: [10.5281/zenodo.___]
authors:
  - family-names: [Surname]
    given-names: [Name]
```

Mohamed bin Zayed University of Artificial Intelligence

Preregistration and Registered Reports move review earlier

USUAL ROUTE

Design	Data	Analysis	Write-up	**Review**	Publish?

REGISTERED REPORT

Design	**Review 1**	Data	Analysis	**Review 2**	Publish

Preregistration

A time-stamped plan on OSF or a trial registry. Exploratory work stays welcome, just labelled.

Registered Report

Accepted in principle before results exist, so null findings get published too.

Mohamed bin Zayed University of Artificial Intelligence

Reporting checklists do the remembering for you

Standard	Use it for
MIAME / MINSEQE	Microarray and sequencing experiments: the minimum metadata to interpret them
ARRIVE 2.0	Animal studies, including xenograft models
CONSORT / STROBE	Clinical trials; observational studies
TRIPOD+ AI	Clinical prediction models, including machine learning
Model cards, datasheets	What a model or dataset is for, how it was built, and its limits
protocols.io	Step-by-step wet-lab and computational protocols, versioned, with a DOI

Mohamed bin Zayed University of Artificial Intelligence

Peer review is opening up, and continuing after publication

Transparent review

Reviewer reports and author responses published with the paper, named or anonymous

Reviewed preprints

Review attached to a public preprint; eLife has published an assessment instead of an accept/reject decision since 2023

Post-publication review

Comments on PubPeer and elsewhere; corrections and retractions as scrutiny continues

Lower submission friction

Format-free first submission, so time goes into science rather than reformatting

Mohamed bin Zayed University of Artificial Intelligence

Concerns about open science, and what answers them

Concern	The worry	What answers it
Being scooped	Others publish faster using my shared data or code	DOIs and timestamps establish priority
Loss of credit	Datasets, software or ideas reused without attribution	CITATION.cff files and licences ensure credit
Data misuse	Shared data misinterpreted or used unethically	Clear metadata, licences and community norms
Legal and privacy	Patient and clinical data cannot always be fully open	Controlled-access repositories and de-identification
Extra workload	FAIR data, documentation and repositories take time	Templates, automation and FAIR-aligned tools
Career risk	Open outputs undervalued in hiring and promotion	Funders and institutions increasingly reward them

Mohamed bin Zayed University of Artificial Intelligence

Concerns about open science, and what answers them

01	**Being scooped**	What others may publish faster using openly shared data/code?	Open sharing with DOIs and timestamps establishes priority, not risk.
02	**Loss of credit**	Datasets, software, or ideas may be reused without proper attribution	Citation files (CITATION.cff) and licenses ensure proper attribution.
03	**Data misuse**	Shared data may be misinterpreted or used unethically	Clear metadata, licenses, and community norms reduce misuse & misinterpretation.
04	**Legal and privacy**	Sensitive data (patients, clinical) cannot always be fully open	Controlled access repositories and de-identification protect sensitive data.
05	**Extra workload**	Preparing FAIR data, docs, and repos takes time/resources	Using templates, automation, and FAIR-aligned tools reduces the burden.
06	**Career risk**	Open outputs (data, code) often undervalued in hiring/promotion	Increasingly, funders and institutions reward open outputs in evaluations.

Mohamed bin Zayed University of Artificial Intelligence

“Anonymous” genomes are not anonymous for long

2008 · POOLED DATA

Homer and colleagues showed a person's presence in a study study could be detected from published allele frequencies alone. frequencies alone.

Aggregate GWAS data then moved behind controlled access.

2013 · SURNAME INFERENCE

Gymrek and colleagues matched Y-chromosome markers to markers to public genealogy databases and recovered surnames surnames of de-identified research participants.

Combined with age and location, this re-identified individuals. individuals.

Consent written today must anticipate the sharing and the re-identification risks of tomorrow.

Mohamed bin Zayed University of Artificial Intelligence

Open for whom?

~80%

of GWAS participants have been of European ancestry, so open data still under-serves most of the world

Ju D et al. Annu Rev Biomed Data Sci 5:321–339 (2022)

Who pays to publish

Author fees move the paywall from readers to writers, who may be in low-income settings

Who controls the data

The CARE principles: Collective benefit, Authority to control, Responsibility, Ethics

Carroll SR et al. Data Sci J 19:43 (2020)

Mohamed bin Zayed University of Artificial Intelligence

Openness audit of your homework paper

1

Alone · 6 min

Score your paper on the six-row scorecard, 0 to 2 per row. Total out of 12.

2

Groups of 3 · 8 min

Compare cards. For each paper's lowest row, decide: closed for a stated reason, or avoidable?

3

Groups · 4 min

Pick one paper. Rewrite its data and code availability statement as it should have been, in 60 words or fewer.

Mohamed bin Zayed University of Artificial Intelligence

Openness scorecard

Row	0	1	2
Paper	Paywalled	Free to read, no licence	Open licence, or preprint
Data	None, or " on request"	Processed or partial only	Raw data with accessions
Code	None	Linked, no version	Archived release with a DOI
Environment	Not stated	Tool names only	Versions, lockfile or container
Methods	Vague	Detailed in the text	Protocol, checklist or preregistration
Licences	None on data or code	Unclear or partial	Clear for data and code

Total: ____ / 12 · Lowest row: _____ · Closed for a reason, or avoidable?

Mohamed bin Zayed University of Artificial Intelligence

What a good availability statement looks like

TEMPLATE

Raw ChIP-seq and ATAC-seq reads: GEO [GSE_____]. Peaks and source data for every figure: Zenodo [DOI], CC BY 4.0. Code: archived release v1.0, Zenodo [DOI], MIT licence; environment in envs/. Human genotypes: EGA [EGAS_____], via the data access committee.

Every item resolvable

An accession or DOI, not a promise

Every item licensed

So reuse is legal, not just possible

Restrictions explained

Closed parts say why, and how to apply

Mohamed bin Zayed University of Artificial Intelligence

Hands-on checklist for your next project

Before you start

Choose an open licence (MIT, CC BY, GPL)
Create a repository template with README, CONTRIBUTING and .gitignore
Draft a data management plan: storage, access, sharing, ethics

During the project

Track code and data with Git
Write and run unit and integration tests
Maintain environment files: requirements.txt, environment.yml, Dockerfile
Capture logs and metadata; build workflows

At submission

Archive data, code and models in trusted repositories (Zenodo, Hugging Face)
Get DOIs for software, datasets and protocols
Write documentation and examples
Add CITATION.cff; post a preprint

After publication

Respond to issues and user feedback
Maintain code and data: fixes and security patches
Plan a roadmap for new versions

Mohamed bin Zayed University of Artificial Intelligence

Five things to remember

1. Open science is what lets people other than you reproduce your work.
2. Open by default; closed only for a reason you can state.
3. FAIR is about how data are described and reached, not whether anyone can download them.
4. A promise to share is not sharing: accessions, DOIs and licences are.
5. Privacy and equity are design questions, to answer before the data exist.

Mohamed bin Zayed University of Artificial Intelligence

Takeaways from the two lectures

1. Reproducibility is good for you first, then for science and society
2. Plan ahead and start with an organised project structure from day one
3. Record and freeze software versions and dependencies (conda, pixi)
4. Use Git for version control, GitHub to collaborate, and an open licence
5. Containerise compute environments (Docker, Singularity)
6. Use Jupyter, R Markdown or Quarto notebooks for literate programming
7. Deposit code, data and models in persistent repositories (Zenodo, Hugging Face) and make data FAIR

Mohamed bin Zayed University of Artificial Intelligence

“Reproducibility is like brushing your teeth. Once you learn it, it becomes a habit.”

Irakli Loladze, Bryan College of Health Sciences, Lincoln, Nebraska

Mohamed bin Zayed University of Artificial Intelligence

The future of science should be not only reproducible and open, but also **inclusive, equitable and environmentally sustainable**.

“If you want to go fast, go alone. If you want to go far, go together.”

African proverb

Mohamed bin Zayed University of Artificial Intelligence

Reading list

Core

UNESCO. *Recommendation on Open Science* (2021)

Wilkinson MD et al. The FAIR Guiding Principles. *Sci Data* 3:160018 (2016)

McKiernan EC et al. How open science helps researchers succeed. *eLife* 5:e16800 (2016)

Carroll SR et al. The CARE Principles for Indigenous Data Governance. *Data Sci J* 19:43 (2020)

Chambers CD & Tzavella L. Registered Reports. *Nat Hum Behav* 6:29–42 (2022)

Further

Gabelica M, Bojčić R & Puljak L. *J Clin Epidemiol* (2022)

Piwowar HA et al. *PLoS ONE* 2:e308 (2007); Colavizza G et al. *PLoS ONE* 15:e0230416 (2020)

Smith AM et al. Software citation principles. *PeerJ Comput Sci* 2:e86 (2016)

Homer N et al. *PLoS Genet* 4:e1000167 (2008); Gymrek M et al. *Science* 339:321 (2013)

Ju D et al. *Annu Rev Biomed Data Sci* 5:321–339 (2022)

Kramer B & Bosman J. 101 Innovations in Scholarly Communication; NASA *Open Science Essentials*

Mohamed bin Zayed University of Artificial Intelligence

Tools and communities worth bookmarking

To…	Start with
Learn the practices	The Turing Way handbook; Software Carpentry and Data Carpentry lessons
Share outputs	OSF for projects and preregistrations; Zenodo for DOIs; protocols.io
Find standards	FAIRsharing.org for data; EQUATOR Network for reporting guidelines
Check policies	Sherpa Romeo for journal self-archiving; your funder's own pages
Get credit	ORCID iD; CITATION.cff; DORA for fairer research assessment
Find peers	ReproducibiliTea journal clubs; national reproducibility networks

Mohamed bin Zayed University of Artificial Intelligence

Reproducible for you, open for everyone else

Exit ticket · 2 minutes, on paper

Write one step from the starter kit you will take this month, and one question about open science you still have.