Nov 12th, 11:30 AM - 12:00 PM

Pre-mortem: UnF.A.I.R - the Death of Open Research

Min-Yen KAN
School of Computing, National University of Singapore (NUS)

Pre-mortem: UnF.A.I.R The Death of Open Research

Kan Min Yen
NUS School of Computing
Web IR / NLP Group (WING)
http://wing.comp.nus.edu.sg

NUS | Computing
National University of Singapore

What is a pre-mortem?

Pre-mortem

NUS National University of Singapore | Computing

An exercise used by teams to identify potential problems in a decision *before* they occur, as a way to prevent failure.

A project hypothetically fails.

What went wrong?

Usually deployed in a team setting and also documented, but for today we'll skip these steps.

Dall-E 3 Multi-prompted generated artwork

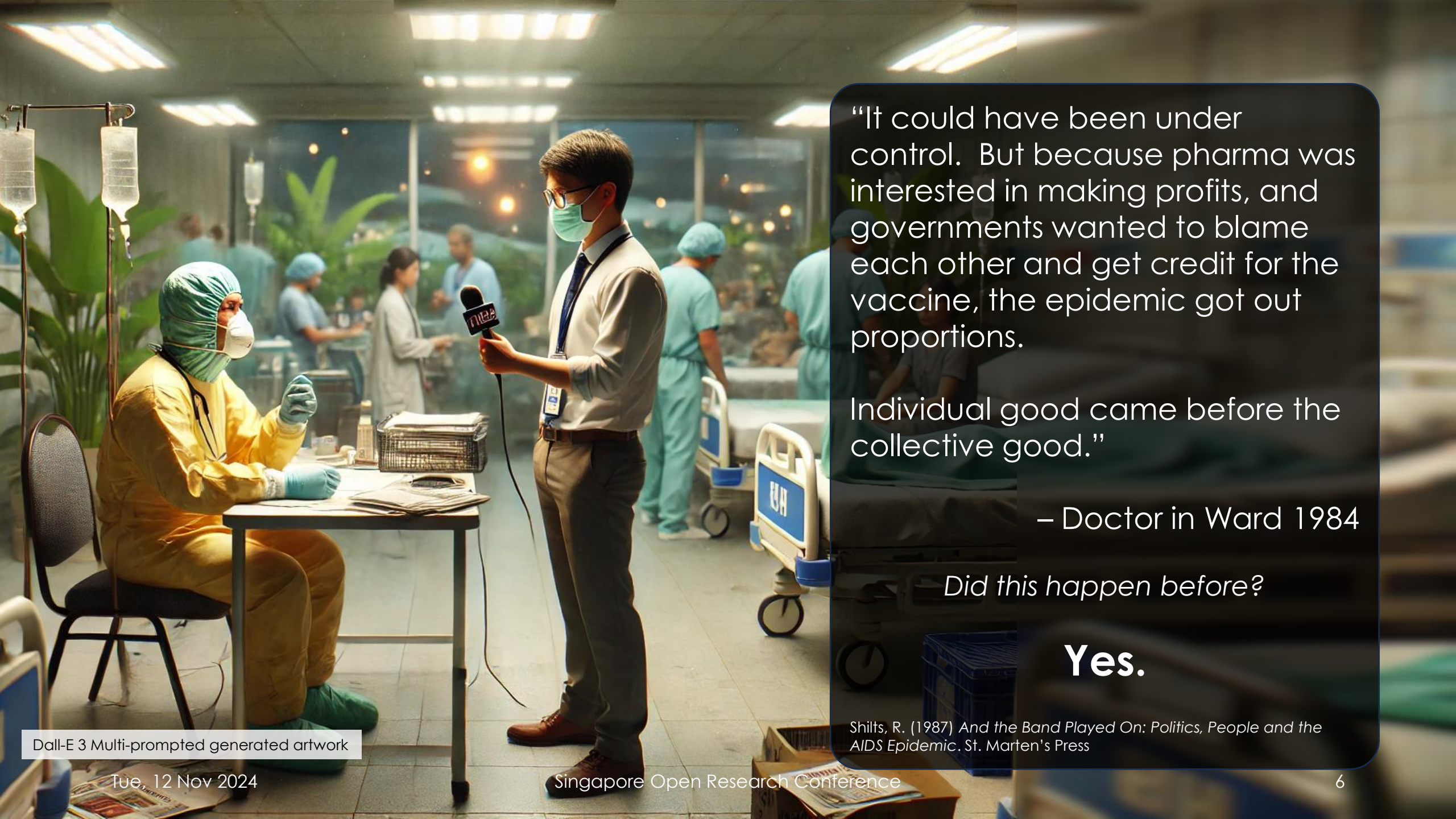
Asia Channel News – 12 Nov 2024 (Singapore)

Families deported their foreign domestic helpers and ran to supermarkets to stock up bleach and toilet paper to ensure their bathrooms were clean and free of the virus.

Ms Tan (not her real name) and her husband of ten years stated while their children enjoyed the company of their helper, safety concerns came first.

Disclaimer: Fictitious Story

Dall-E 3 Multi-prompted generated artwork



"It could have been under control. But because pharma was interested in making profits, and governments wanted to blame each other and get credit for the vaccine, the epidemic got out proportions.

Individual good came before the collective good."

– Doctor in Ward 1984

Did this happen before?

Yes.

Shilts, R. (1987) *And the Band Played On: Politics, People and the AIDS Epidemic*. St. Marten's Press

Dall-E 3 Multi-prompted generated artwork

Could it happen again?

Yes.

How do we stop it from happening again?

By being F.A.I.R.

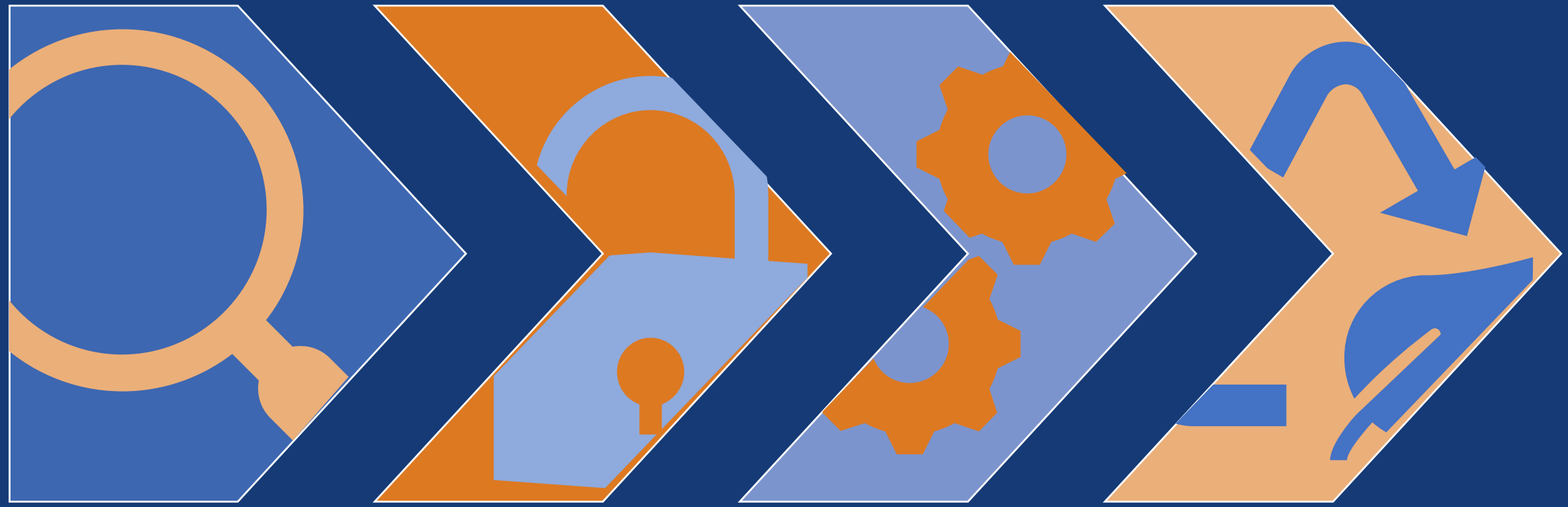
F indable
A ccessible
I nteroperable
R eusable

Dall-E 3 Multi-prompted generated artwork

Outline

- The Death of Open Research
- **What is Open Research, anyways?**
  - F.A.I.R. Principles
- Why does Open Research Matter
- Implications of AI on F.A.I.R.ness

Open Research = F.A.I.R.



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

F.A.I.R. Principles

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

Findable

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

What's your permanent address?

Metadata inextricably linked to data

Expansive notion of data: algorithms, tools, ancillary artefacts

Accessible

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

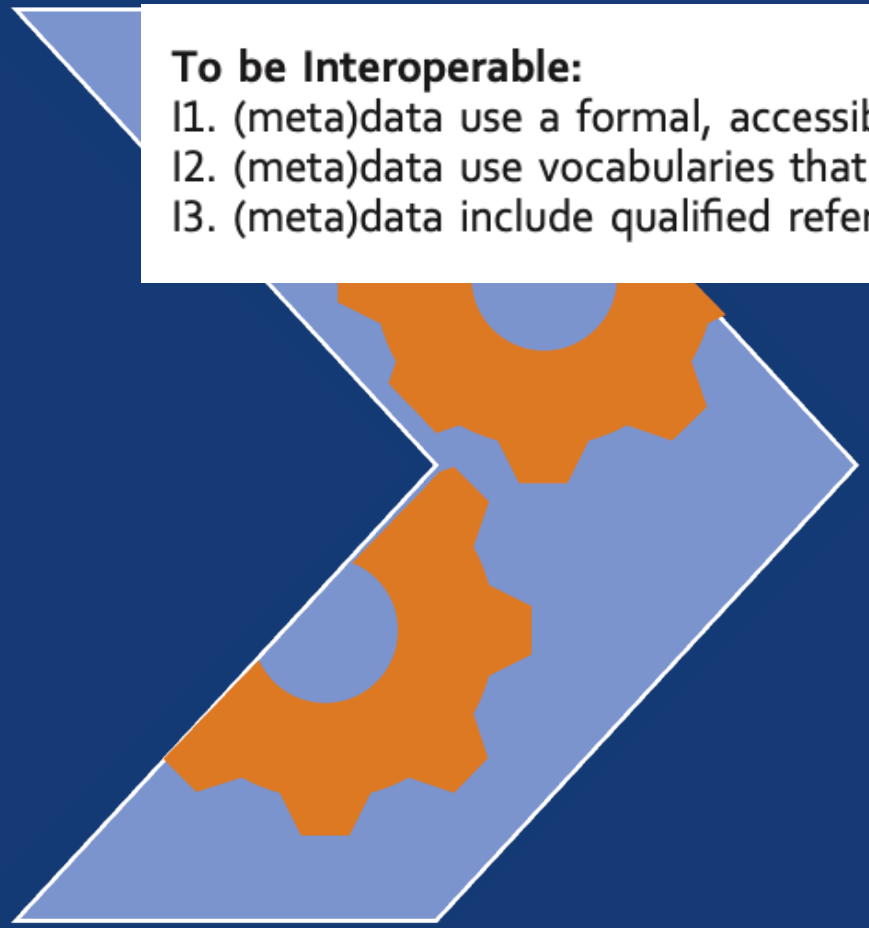
… as in Open Access

Includes authorization as a first-class citizen (e.g., Intellectual Property)

Metadata as distinct from data

Interoperable

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



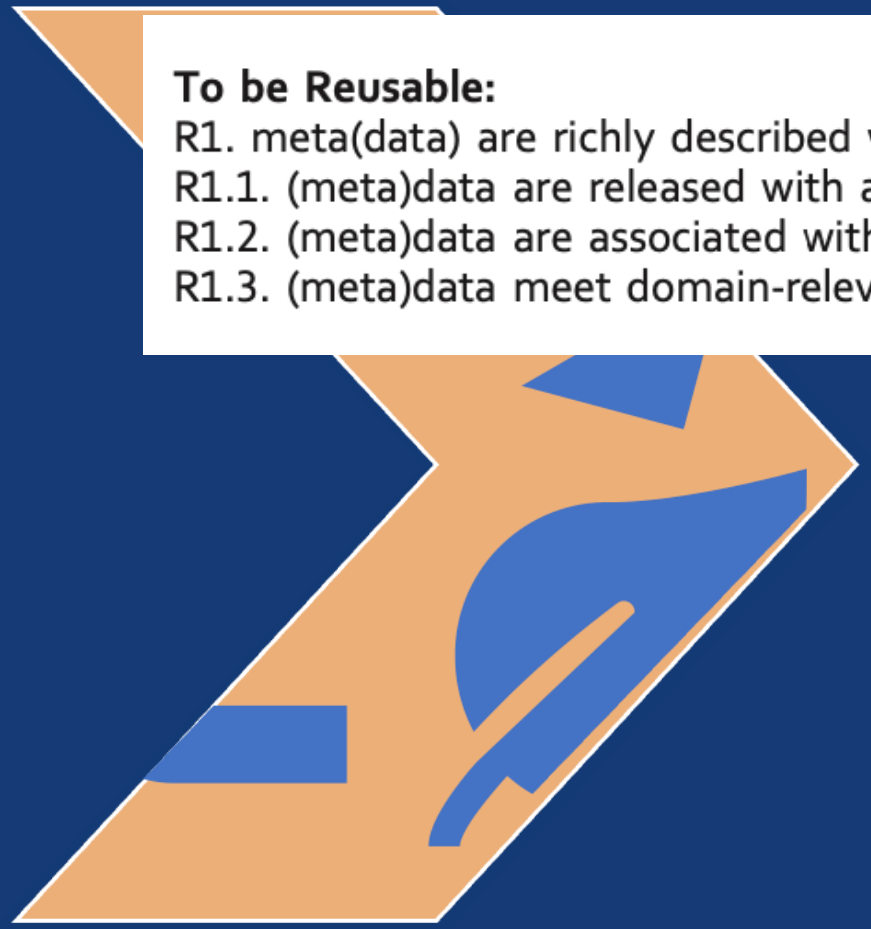
Let's agree on our signs and symbols

Extensibility from first principles (least common denominator; Dublin Core)

Promotes machine (AI) readability and manipulation

Reusable

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



Stand on the shoulders of giants

A citation network reproduced in metadata

Granular data licensing

Outline

- The Death of Open Research
- What is Open Research, anyways?

- **Why does Open Research Matter**
  - When is Closed Research appropriate

- Implications of AI on F.A.I.R.ness

UnF.A.I.R.

NUS | Computing
National University of Singapore

Redundant research

- Corollary: Biased discovery software generate echo chambers

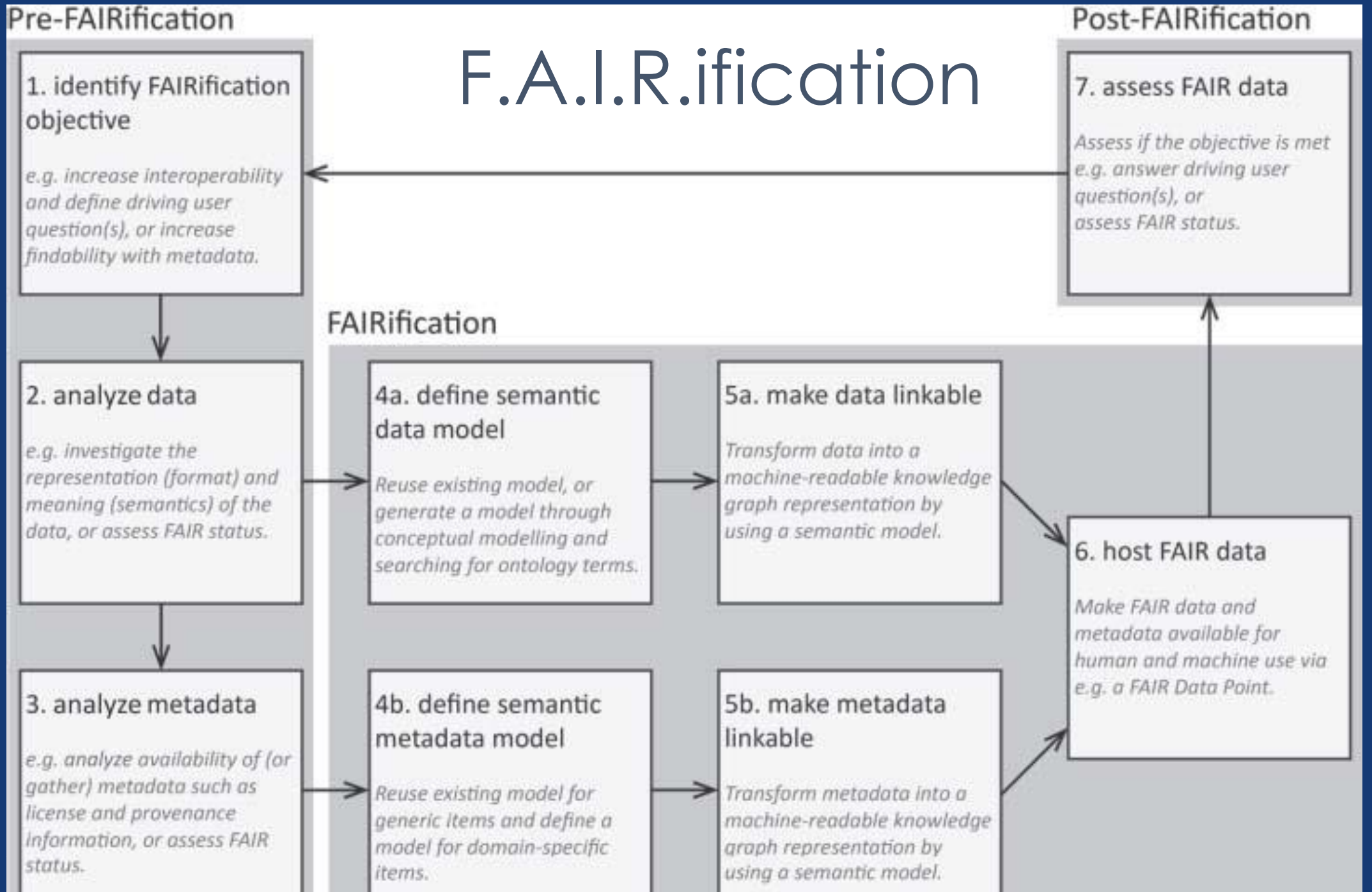
Pass the increased costs to researchers and institutions

- Slower and more expensive scientific progress

Inequity in research access

- Self-Reinforcement Loop

F.A.I.R.ification



Jacobsen, A., et al. (2020). "A generic workflow for the data FAIRification process." *Data Intelligence*, 2(1-2), 56-65. https://doi.org/10.1162/dint_a_00028

Tue, 12 Nov

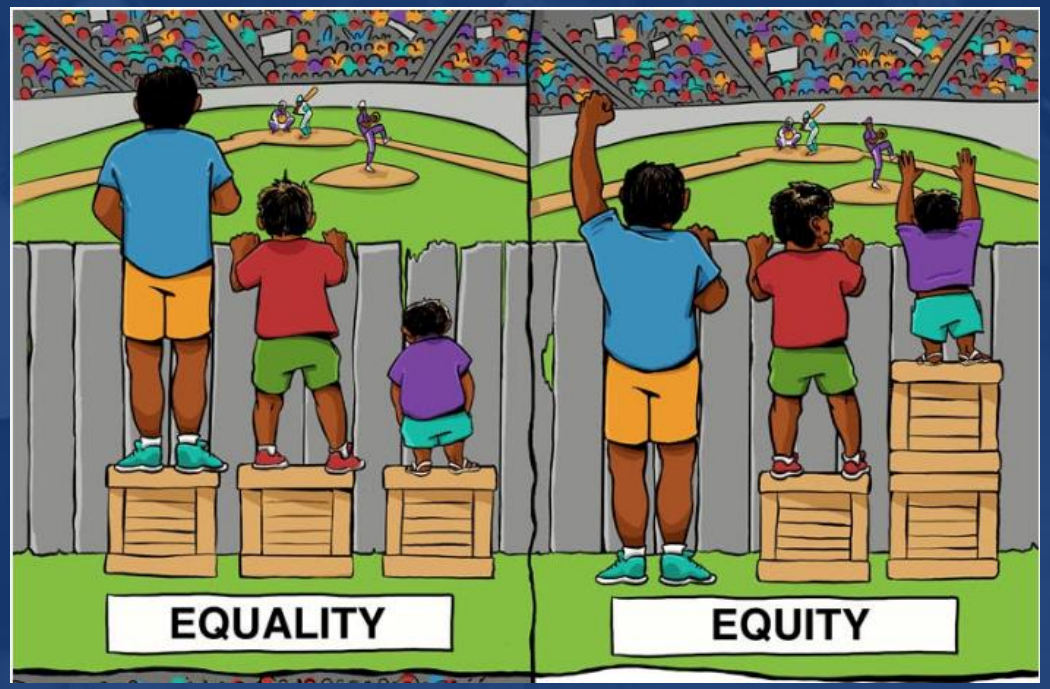
F.A.I.R.ness is tough

Wicked problem requiring regulatory and financial commitment from stakeholders

- People (co-authors, publishers, institutions) are involved, and that makes it complicated.

F.A.I.R.ness does not affect all parts of the ecosystem uniformly

- Who bears the cost?
- Who reaps the benefit?



Maintain incentives, but also guarantee opportunity access

Original artwork credit: interactioninstitute.org and madewithangus.com, respectively.

Closed Research

When is closed research an appropriate model to follow?

- **National Imperative** – US Manhattan Project
- **Competitive Advantage** – Drug Development
- **Vertical Integration** – Apple Ecosystem

NUS | Computing
National University of Singapore

Warnings of an UnF.A.I.R Future

Warning signs that open research might be at risk:

- Digital divide in the global south deepens; open access initiatives get defunded
- Increase in restrictive publishing practices; "dark knowledge" grows; provenance is hard to trace.

Why Open Research Matters

Reinforce the benefits of open research:

- Accelerates innovation, such as specialised discovery services atop common metadata
- Enhances global collaboration and increases trust and transparency

Dall-E generated artwork and stock image from Microsoft Office

Pre-Mortem: Preventing UnF.A.I.R Outcomes

Encourage foundational, long-term funding for open access platforms for both data and literature

Advocate user education in discovery tool use, collaboration, transparency and trust in research

Dall-E 3 Multi-prompted generated artwork

Outline

- The Death of Open Research
- What is Open Research, anyways?
- Why does Open Research Matter
- Steps to Prevent UnF.A.I.R Outcomes

- **Implications of AI on F.A.I.R.ness**

NUS | Computing
National University of Singapore

You can't spell F.A.I.R. without **A.I.**

Dall-E 3 Multi-prompted generated artwork

“Genie Wish” from FORCE11 2021

Brigitte Mathiak: Sustainable funding for open discovery infrastructures and recognition of the long-term value of accessible knowledge.

Suzanne Dumonchel: Improved metadata quality through better-trained data providers.

Nancy Kwangwa: An inclusive, comprehensive discovery platform that integrates various forms of research outputs, from articles to data, across all disciplines and regions.

F.A.I.R.ness supports AI

Emphasis on conventionalised metadata, accessibility, provenance and interoperability afford A.I., **by design.**

Findability is a related, separate component (echo chambers)

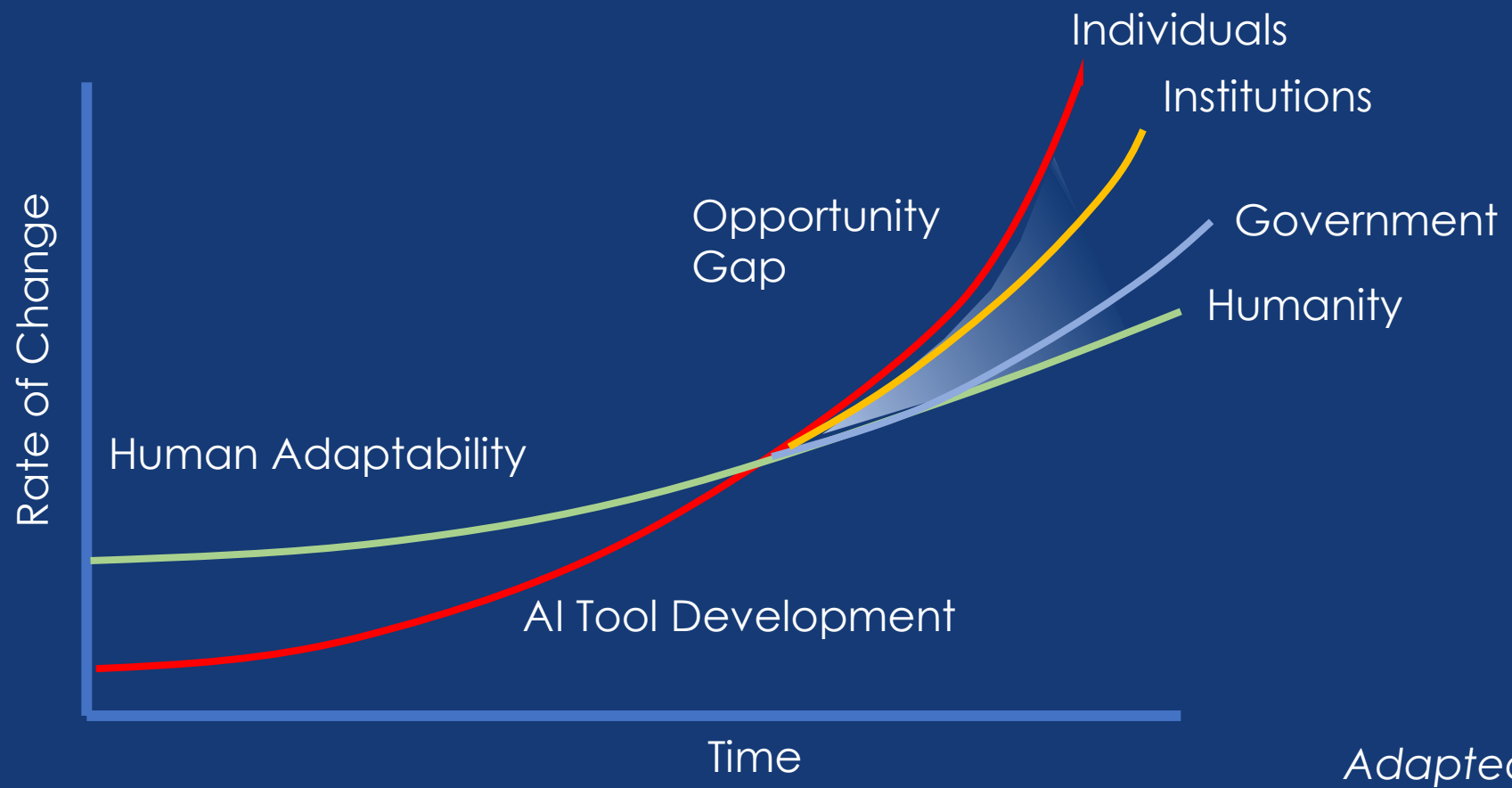
Then let's look at the remainder: A.I.Rification

A.I.R.ification

Using A.I. to address the discrepancy in metadata engineering and quality.

- Automated metadata extraction, assignment by Large Language Models (LLMs)
- Automated provenance and licensing insertion (e.g., Wikibot)
- Automated quality checking (e.g., LLM Arena)

Acceleration has it out for us



Adapted from Astro Teller's graph

Last Words – Weapons of Math Destruction

NUS National University of Singapore | Computing

A WMD is

- Massive
- Opaque
- No feedback loop

The Class Break

Then: what are the WMDs in FAIR research?

Thank you for your attention!

Slides:
https://soc-n.us/241112-unfair