Digital Objects: **The Science Case**

@DimitrisKoureas

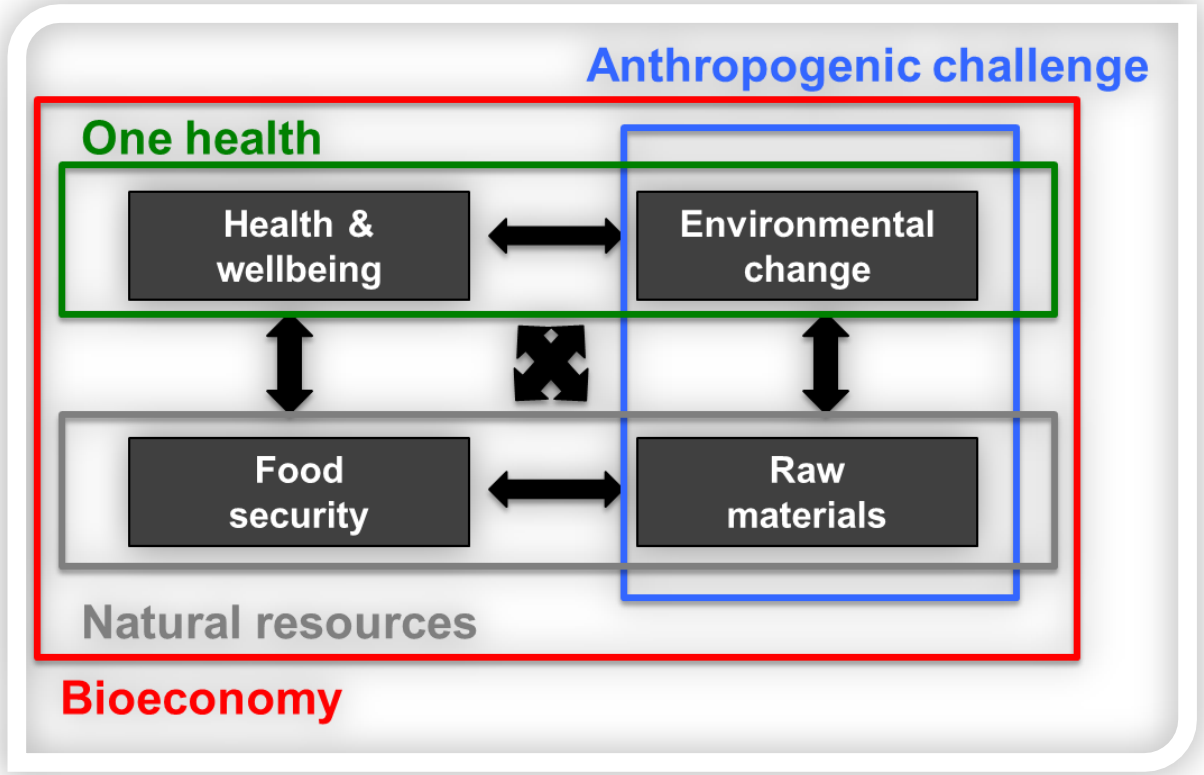
Director, International Biodiversity Infrastructures
Naturalis Biodiversity Center

Our grand challenges require

Data-driven solutions

Need to deliver data at the

Scale, form and precision required



Break Silos | Deliver knowledge

17 PARTNERSHIPS FOR THE GOALS
16 PEACE, JUSTICE AND STRONG INSTITUTIONS
15 LIFE ON LAND
14 LIFE BELOW WATER
13 CLIMATE ACTION
12 RESPONSIBLE CONSUMPTION AND PRODUCTION
11 SUSTAINABLE CITIES AND COMMUNITIES
10 REDUCED INEQUALITIES
9 INDUSTRY, INNOVATION AND INFRASTRUCTURE
8 DECENT WORK AND ECONOMIC GROWTH
7 AFFORDABLE AND CLEAN ENERGY
6 CLEAN WATER AND SANITATION
5 GENDER EQUALITY
4 QUALITY EDUCATION
3 GOOD HEALTH AND WELL-BEING
2 ZERO HUNGER
1 NO POVERTY
Demand
Data
Supply
Science practice

Challenges **global**

- It needs global standards
- Global workflows
- Cooperation of global players

BUT

Science carried out **“locally”**

- By local scientists
- Being part of local infrastructures
- Having local funders

Dark data more important mainly due to their volume[1]

Investigator-focused 'small data'

Locally generated 'invisible data'

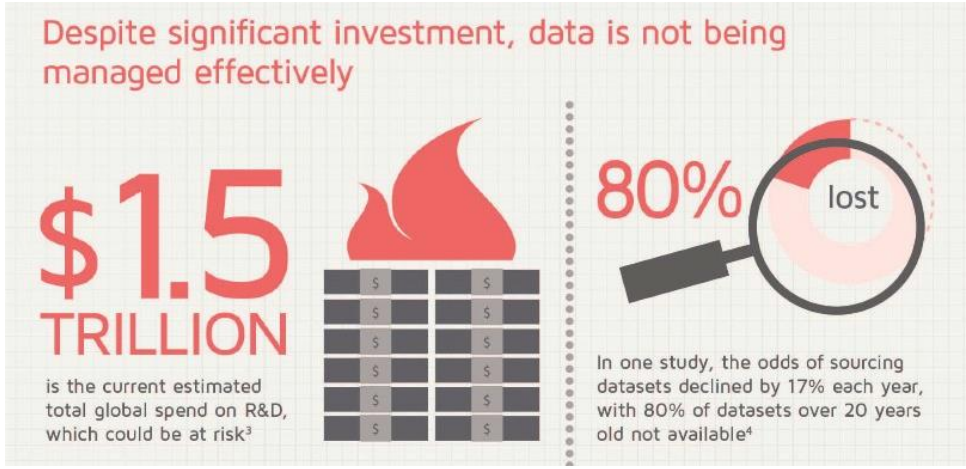
'incidental data'

80%
dark data

Published and discoverable data

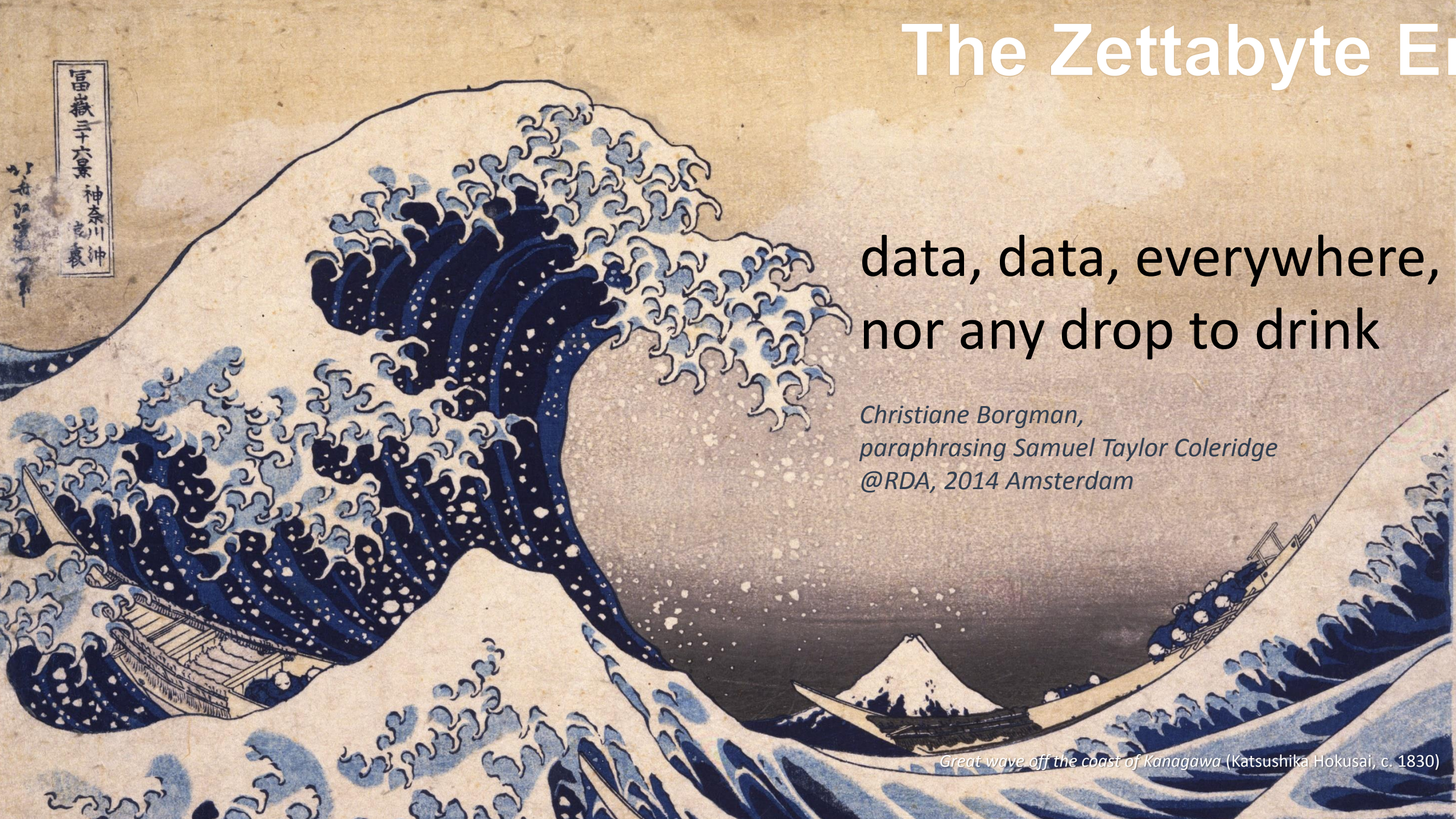
20%

Dark data lost within 20 years



[1]Heidorn PB. *Library Trends* **57:**280-299

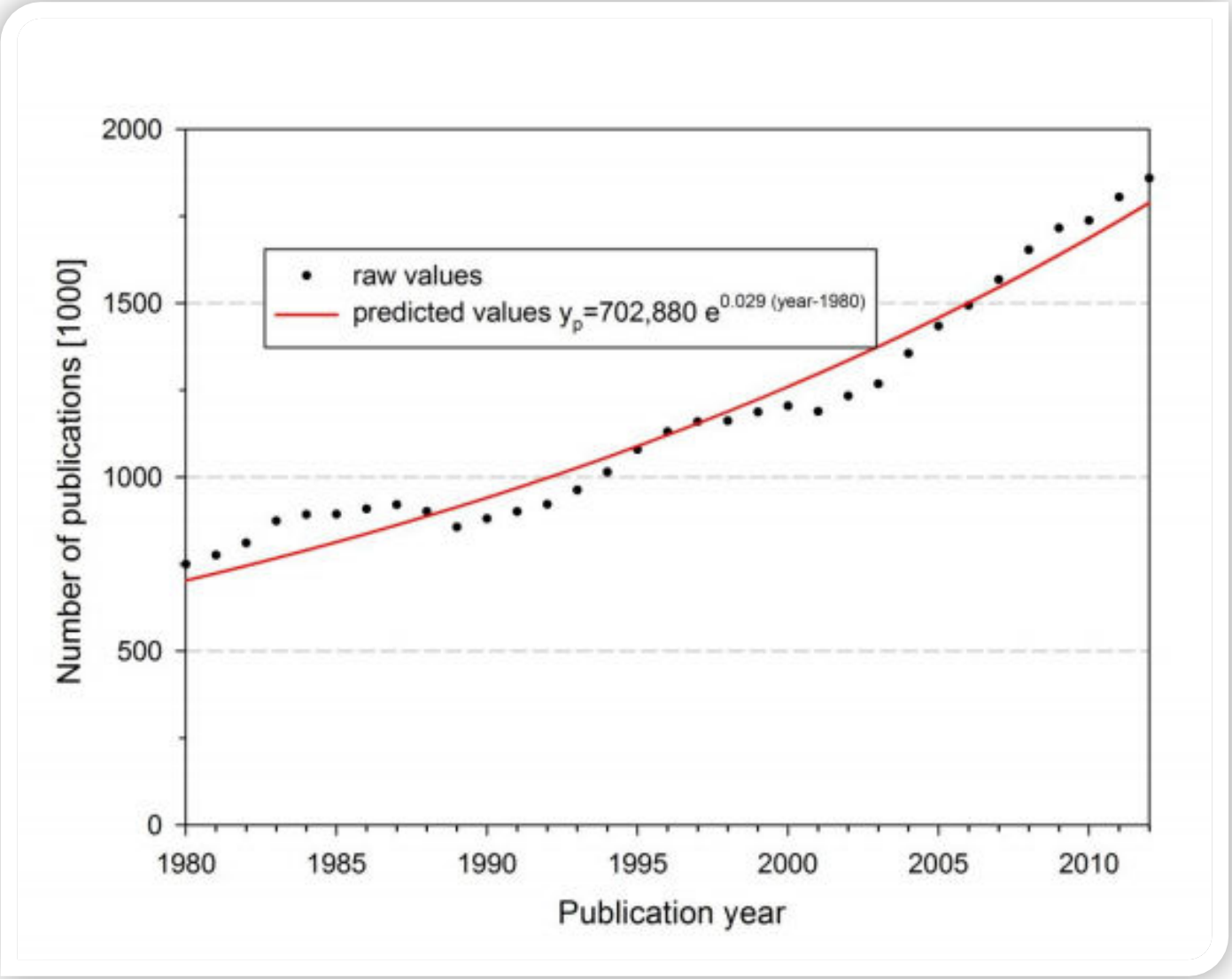
The Zettabyte E



data, data, everywhere, nor any drop to drink

Christiane Borgman,
paraphrasing Samuel Taylor Coleridge
@RDA, 2014 Amsterdam

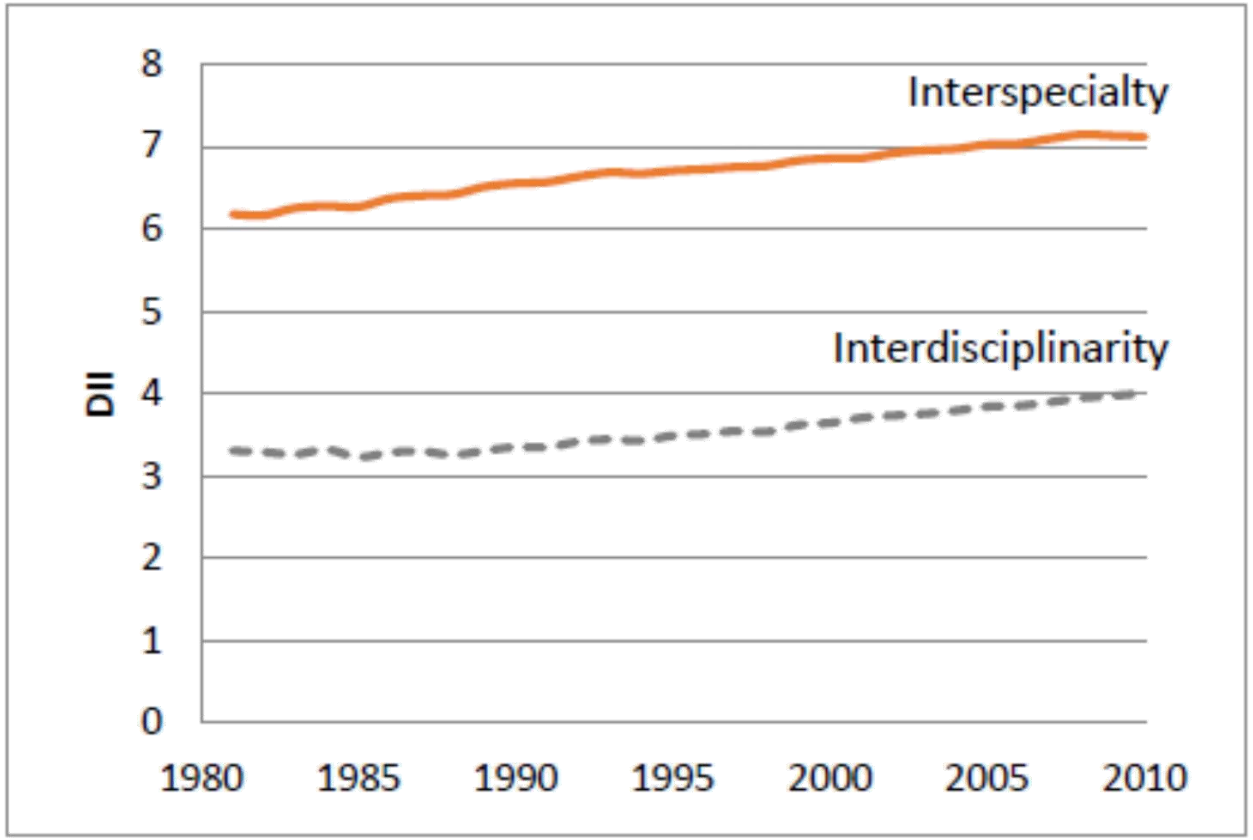
Great wave off the coast of Kanagawa (Katsushika Hokusai, c. 1830)



L. Bornmann & R. Mutz, 2014 **arXiv:1402.4578**

Ever-increasing rate of global scientific products

Does data ‘availability’ affect scientific outputs rate?



Impact Indicator of interdisciplinary research from 1981–2010

Chen, Shiji, et al. "Interdisciplinarity patterns of highly-cited papers: A cross-disciplinary analysis." *Proceedings of the American Society for Information Science and Technology*51.1 (2014): 1-4.

Impact of Interdisciplinary research publications

Challenges **global**

- It needs global standards
- Global workflows
- Cooperation of global players

BUT

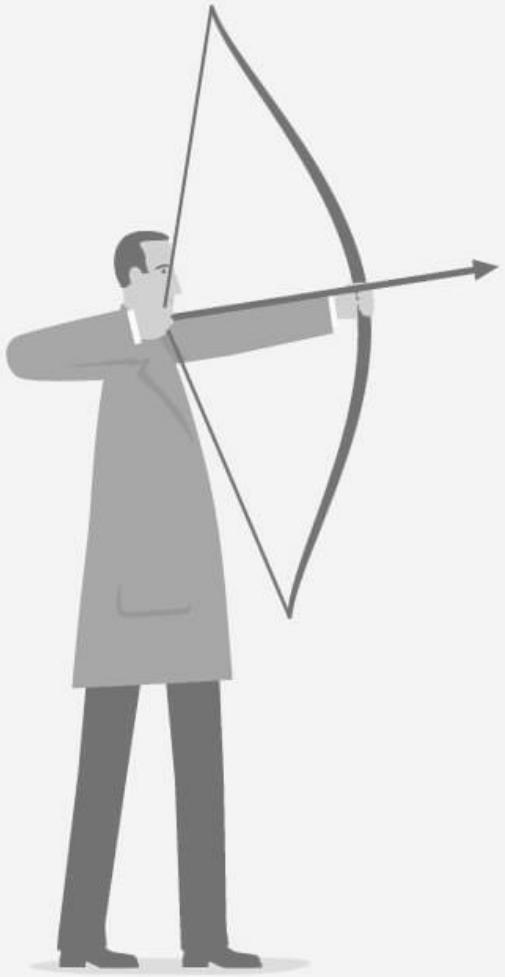
Science carried out **“locally”**

- By local scientists
- Being part of local infrastructures
- Having local funders

Scientists develop and thrive within their respective small communities of practice

TRUST

Practices
People
Data



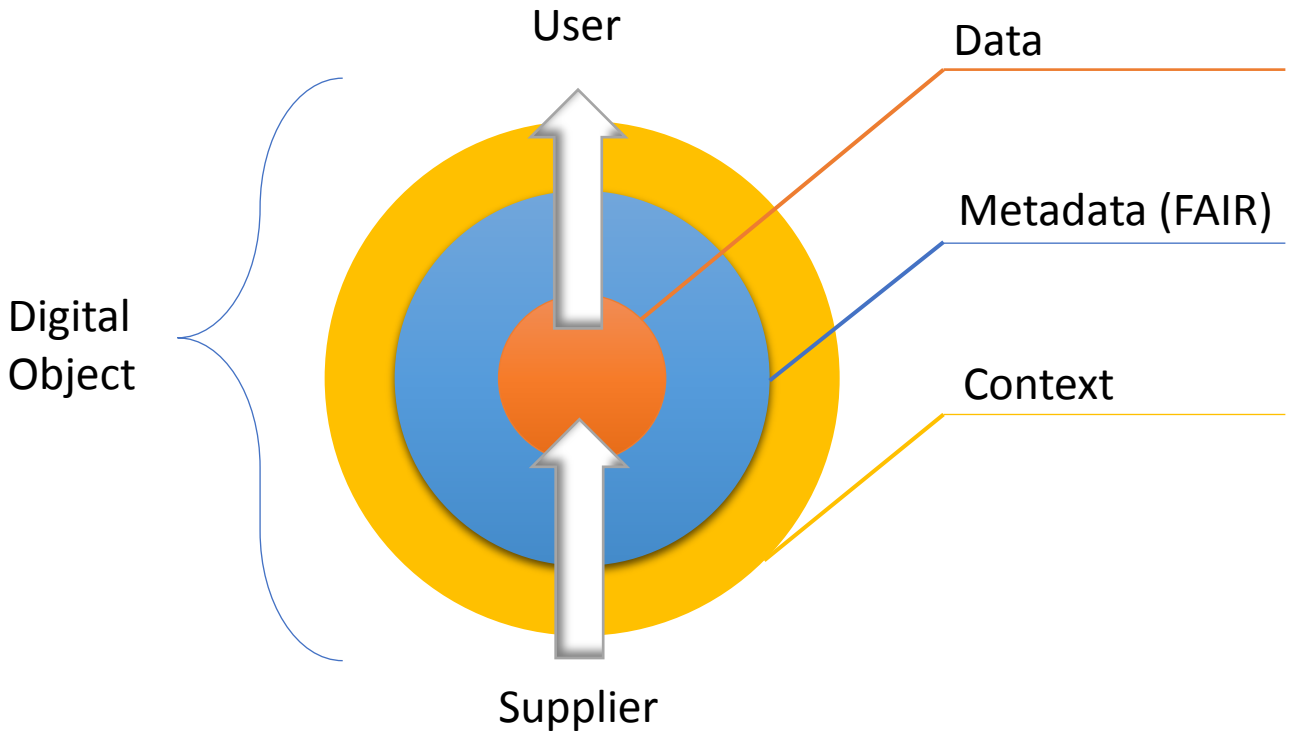
Trust lost when datasets disconnect from:

context in which they were created,
or
communities who created them.

Transforming digital entities to meaningful entities

Need for robust mechanisms, which

- **Link** persistently **data** with its **metadata** (Data in context)
- Enable for data objects to become **self-interpretable**, **actionable** and **trusted** entities
- Enable **curation and annotation** of objects to be performed across repositories and permanently attached to the object



Do DOs help to increase trust?

- DOs can be self-contained and can convey the context in which datasets were generated and allow for future annotations by the community
- it gives each digital (data) entity an identity allowing to prove identity and authenticity even after years
- types of metadata are available even for machine processing (descriptive, system, rights, provenance, etc.)
- transactions can be verified
- Respect the domain-specific specificities

What builds

TRUST

in data?

Relevance

Provenance

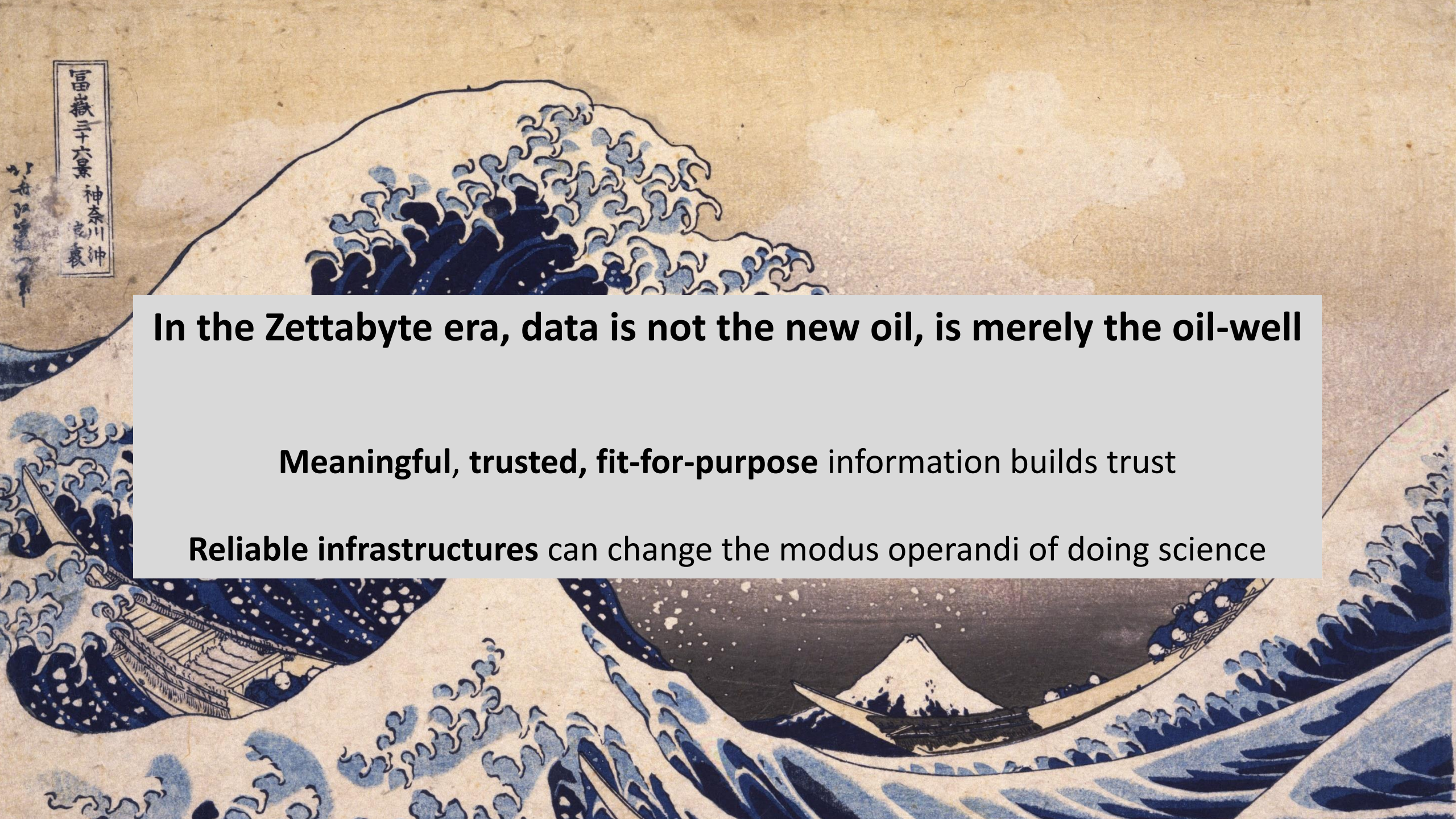
Attribution

Completeness

Fitness-for-purpose

Agility

Branding (Datatyping)



In the Zettabyte era, data is not the new oil, is merely the oil-well

Meaningful, **trusted, fit-for-purpose** information builds trust

Reliable infrastructures can change the modus operandi of doing science

Science is a ‘light’s better’ endeavour in that research effort is **not directed at areas where the work is technically infeasible**. Research is directed where real, interpretable results may be obtained.

We do, in fact, conduct research where the light’s better.

But, when the light changes, so does science.

With better illumination, we look in new areas.

We find new things…

Thank you
@DimitrisKoureas