



Implementing a Flexible Infrastructure for a Digital Object Based Data Management

Peter Wittenburg, Dimitris Koureas, Larry Lannom & Robert Quick

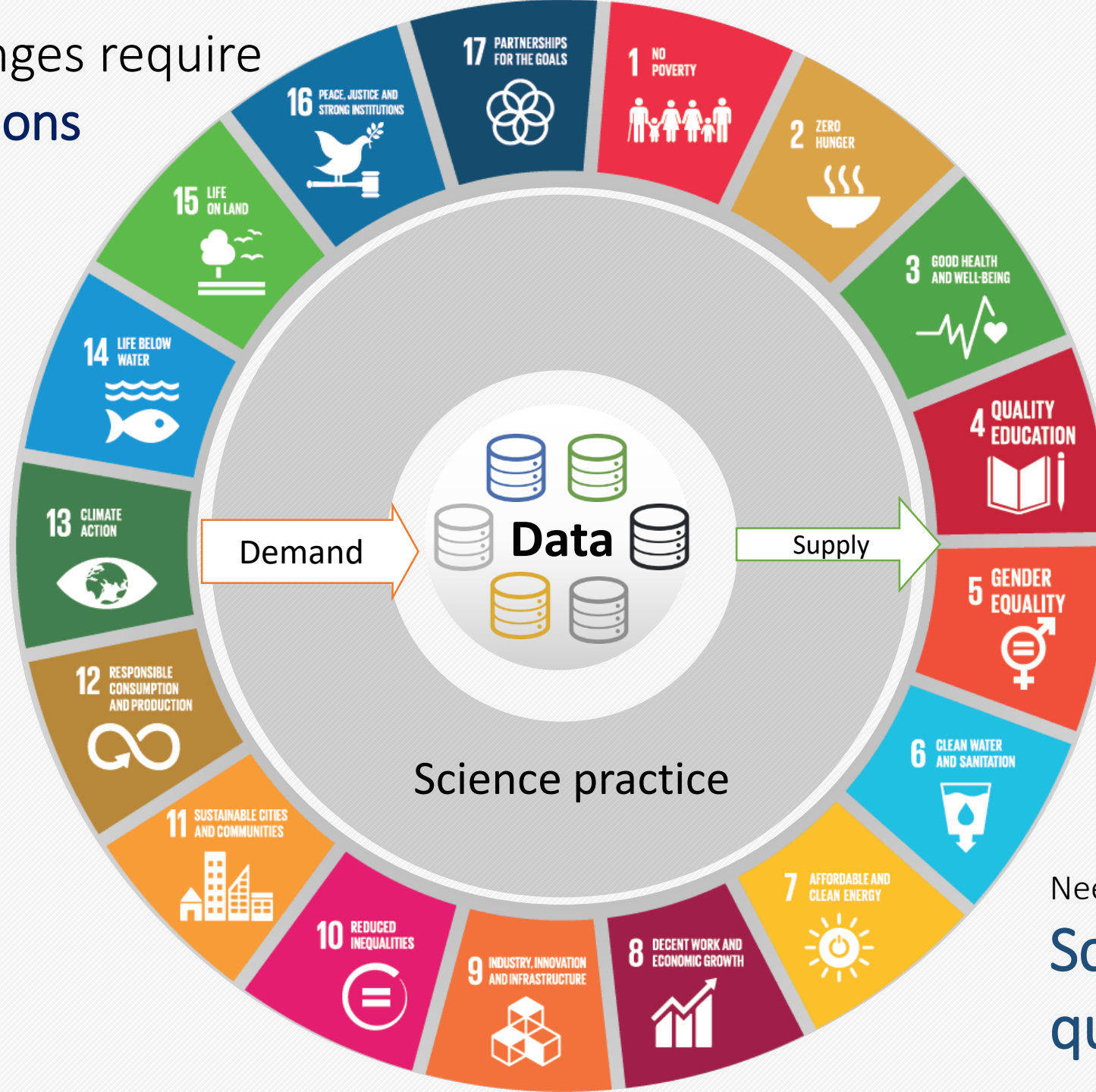
@DimitrisKoureas

Director, **International Biodiversity Infrastructures, Naturalis Biodiversity Center**

Chair, **Biodiversity Information Standards Organisation (TDWG)**

Member, **RDA Technical Advisory Board**

Our grand challenges require
Data-driven solutions

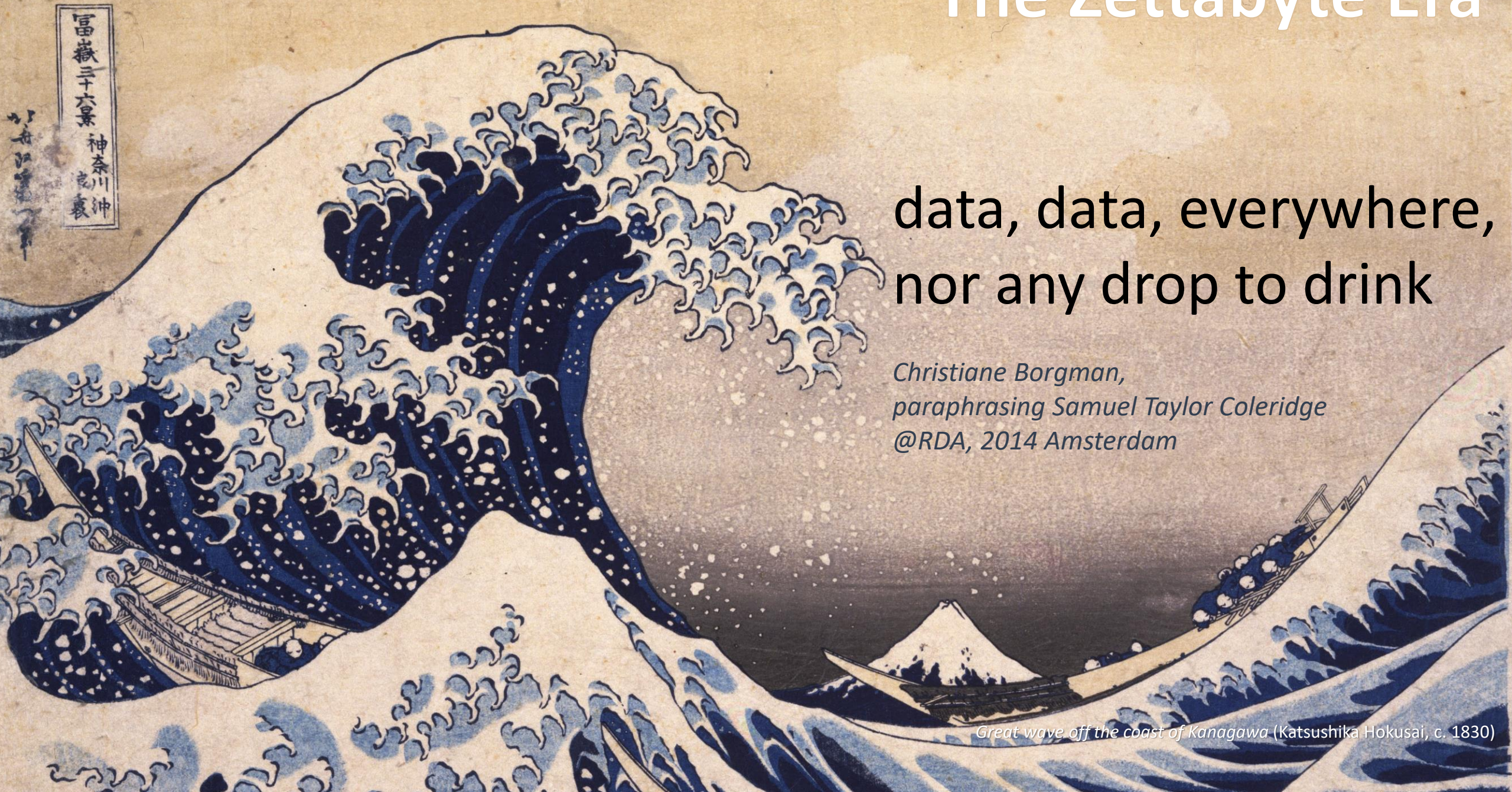


Need to deliver data at the
**Scale, form and
quality required**

The Zettabyte Era

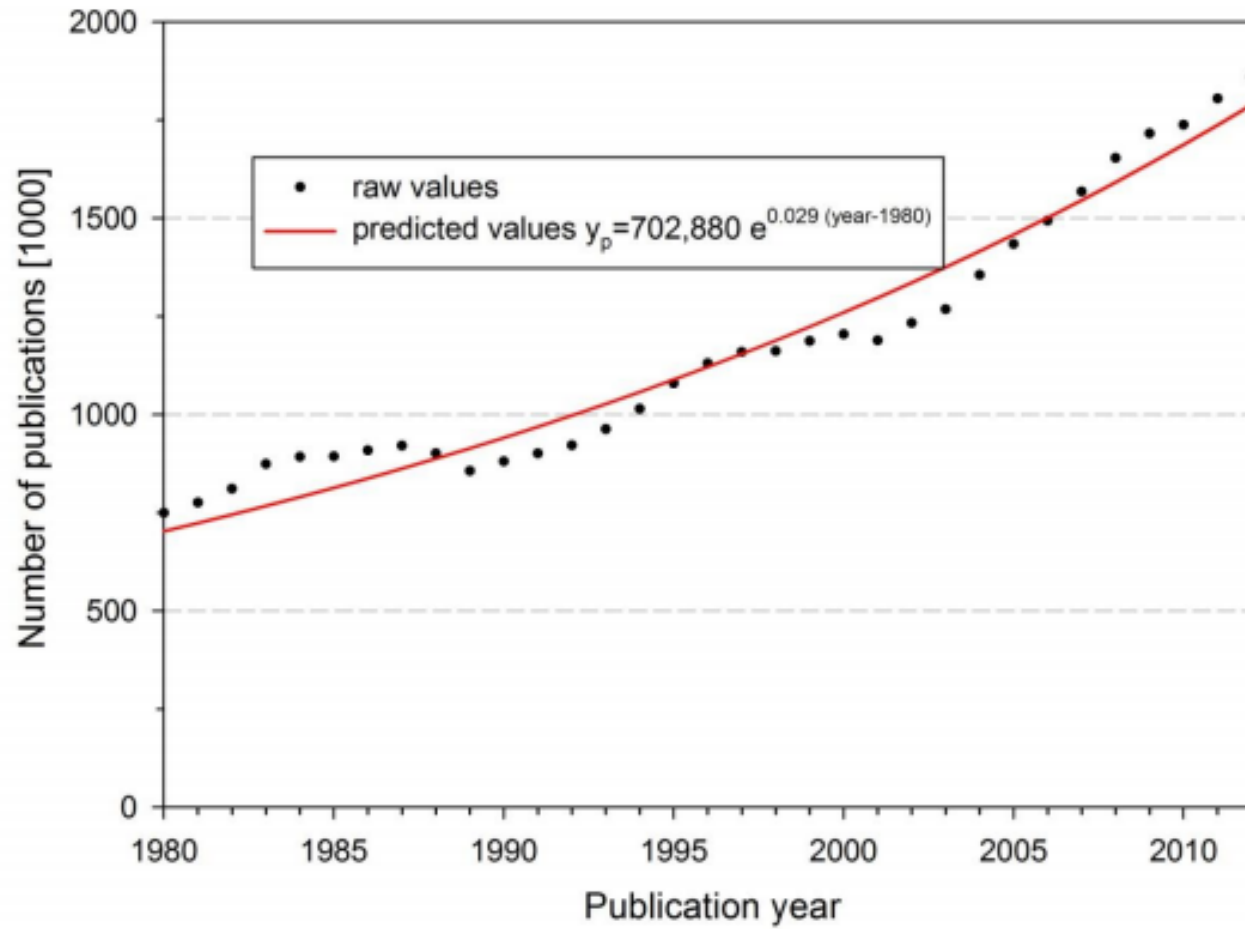
data, data, everywhere,
nor any drop to drink

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



富嶽三十六景
神奈川
波

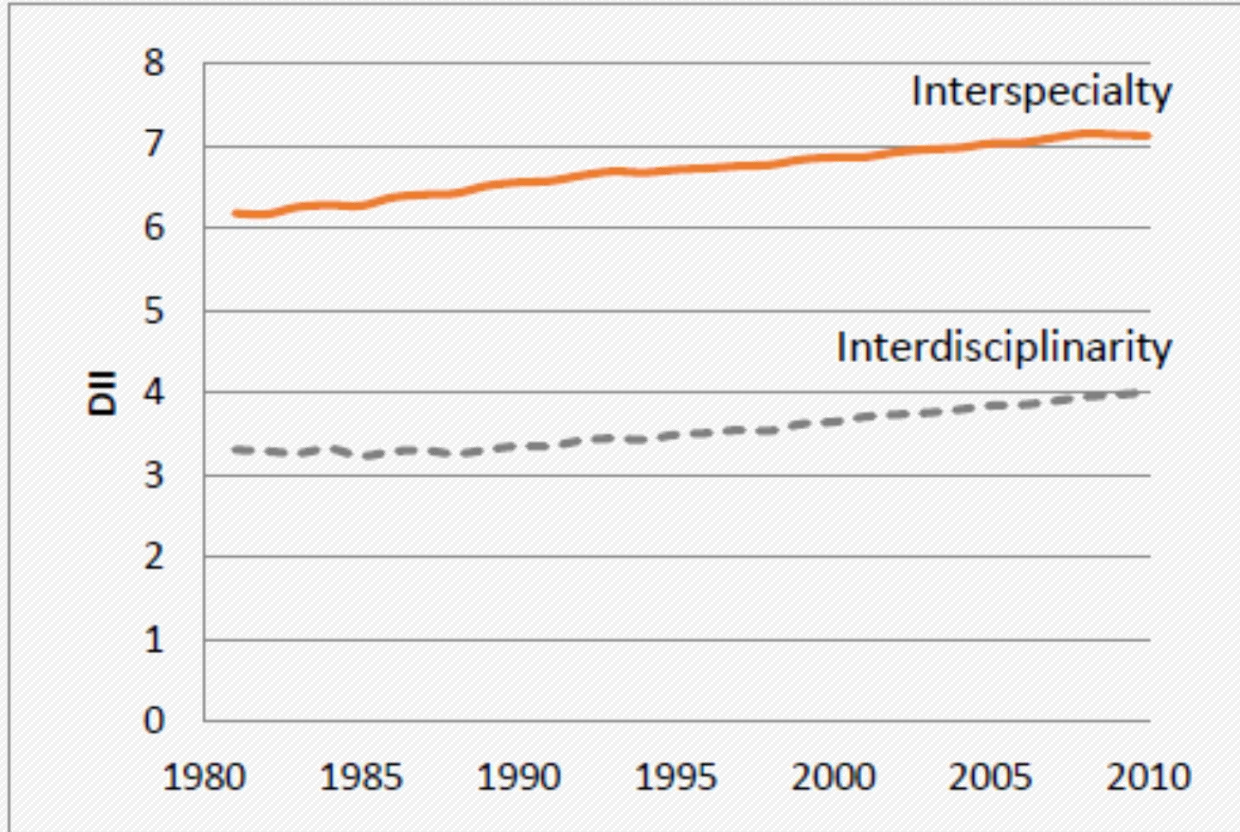
Great wave off the coast of Kanagawa (Katsushika Hokusai, c. 1830)



Ever-increasing rate of global scientific products

Does data 'availability' affect scientific outputs rate?

L. Bornmann & R. Mutz, 2014 [arXiv:1402.4578](https://arxiv.org/abs/1402.4578)



Impact of Interdisciplinary research publications

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.



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



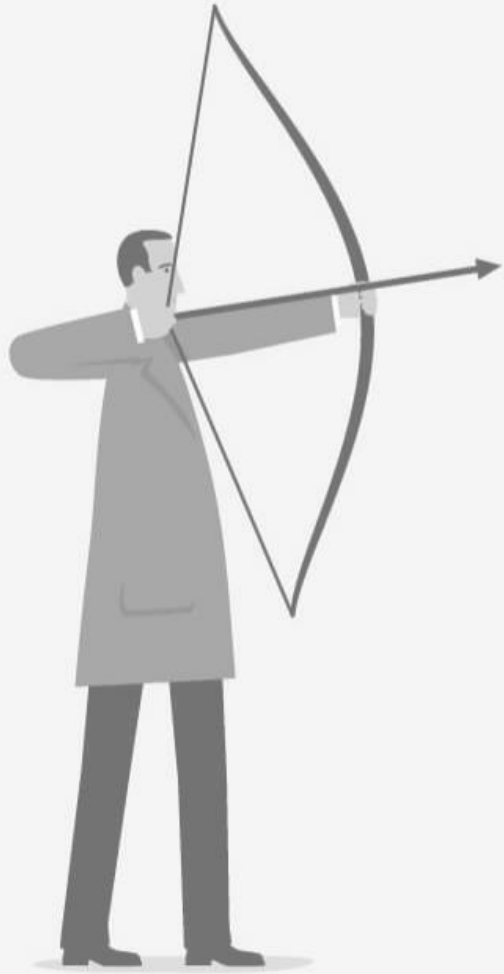
**Scientists develop and thrive
within their respective small
communities of practice**



TRUST

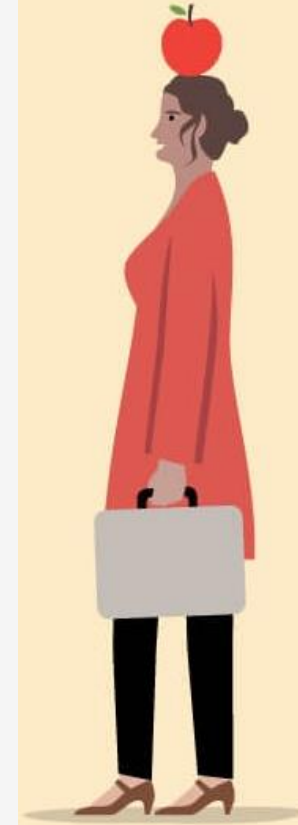
Practices
People
Data





Trust lost when datasets
disconnect from:

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



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

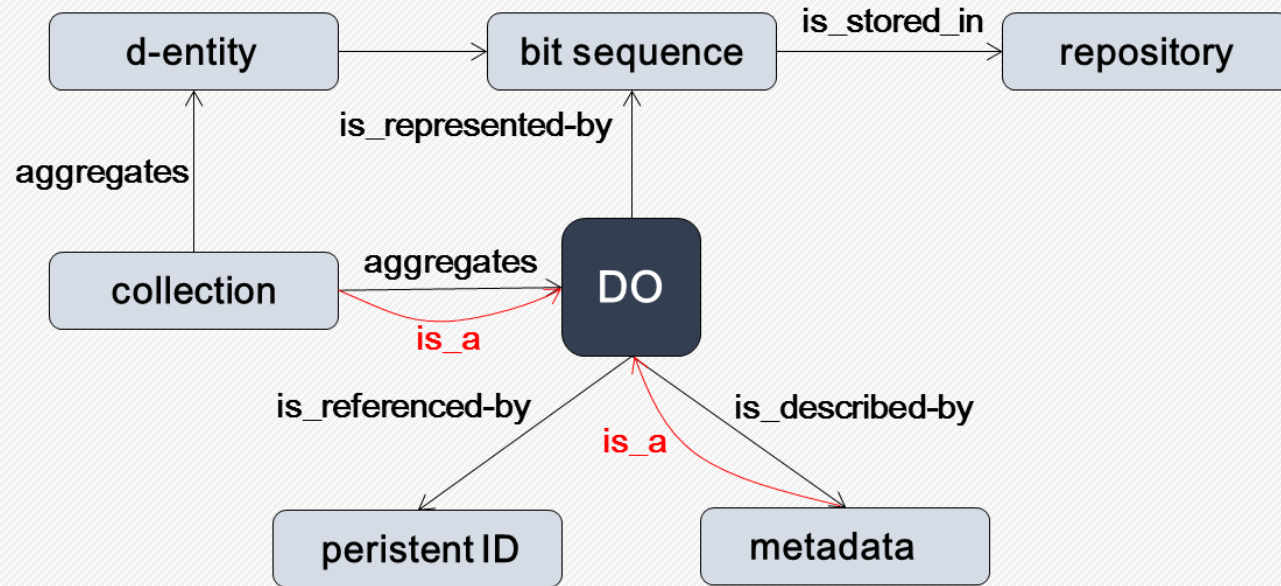
Completeness

Fitness-for-purpose

Agility

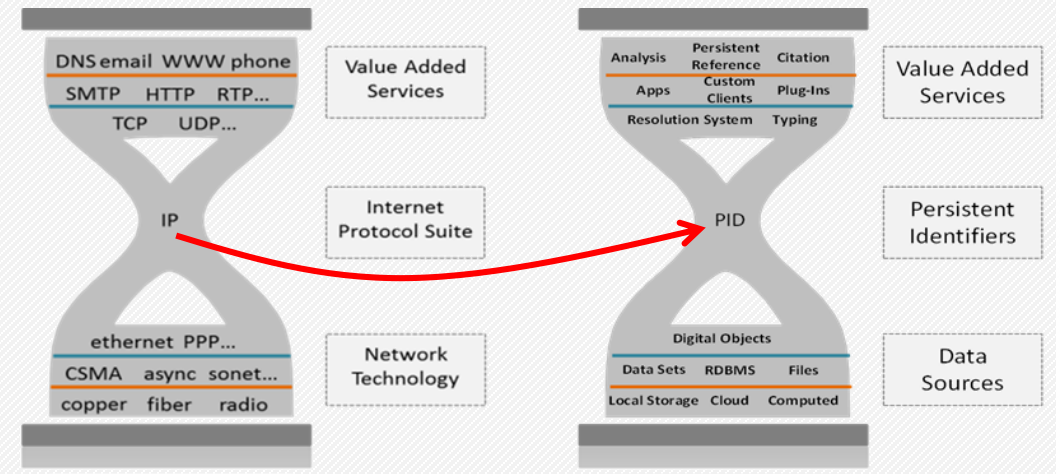
Branding (Datatyping)

DOs: self described, sharable & actionable entities



- RDA standardisation since 2013 (DFT Core Model)
- all based on >20 use cases from various disciplines
- accepted now as EU ICT Specification

Getting dependent on functioning
PID resolution system



Global Digital Object Infrastructure

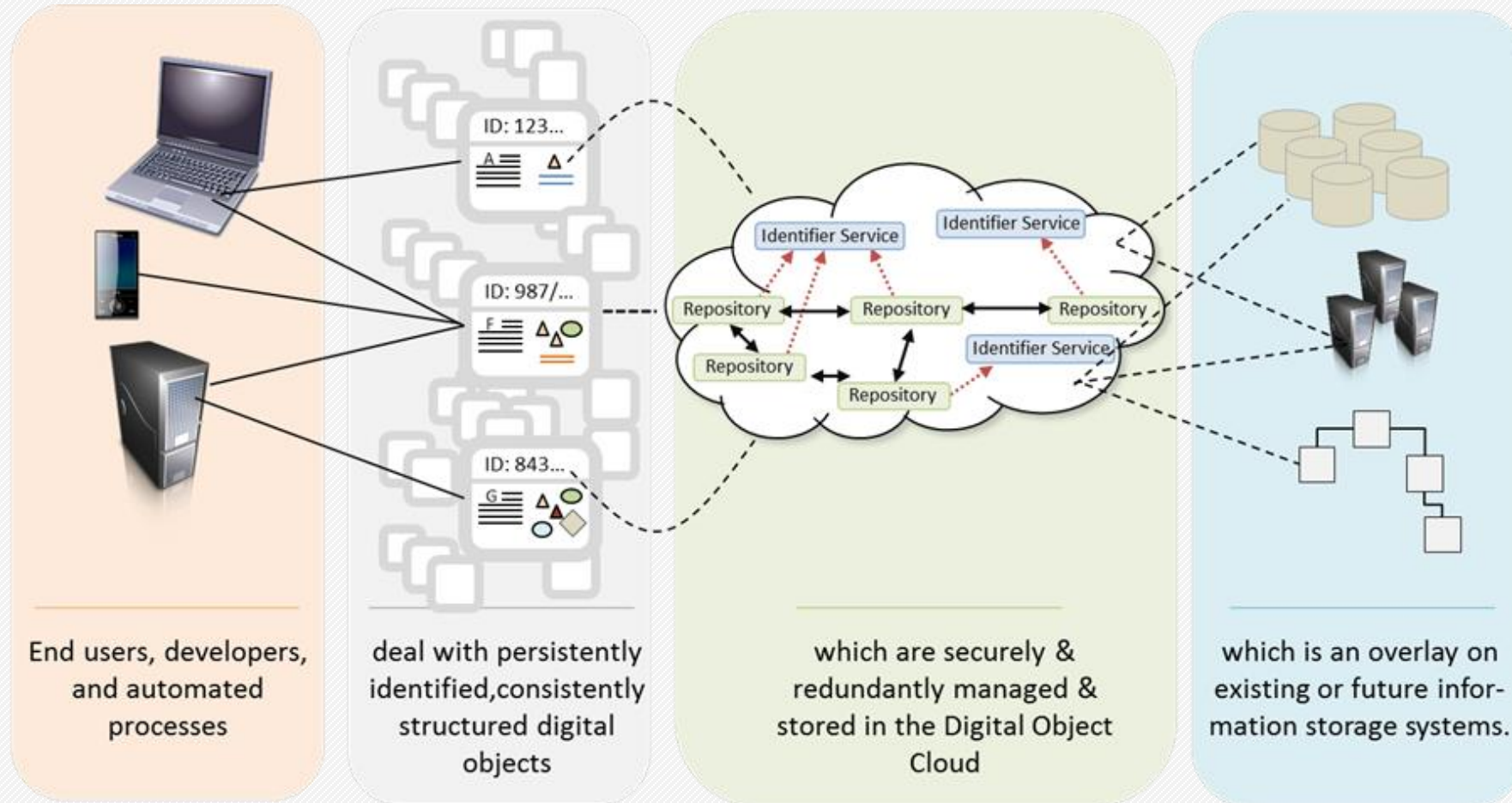
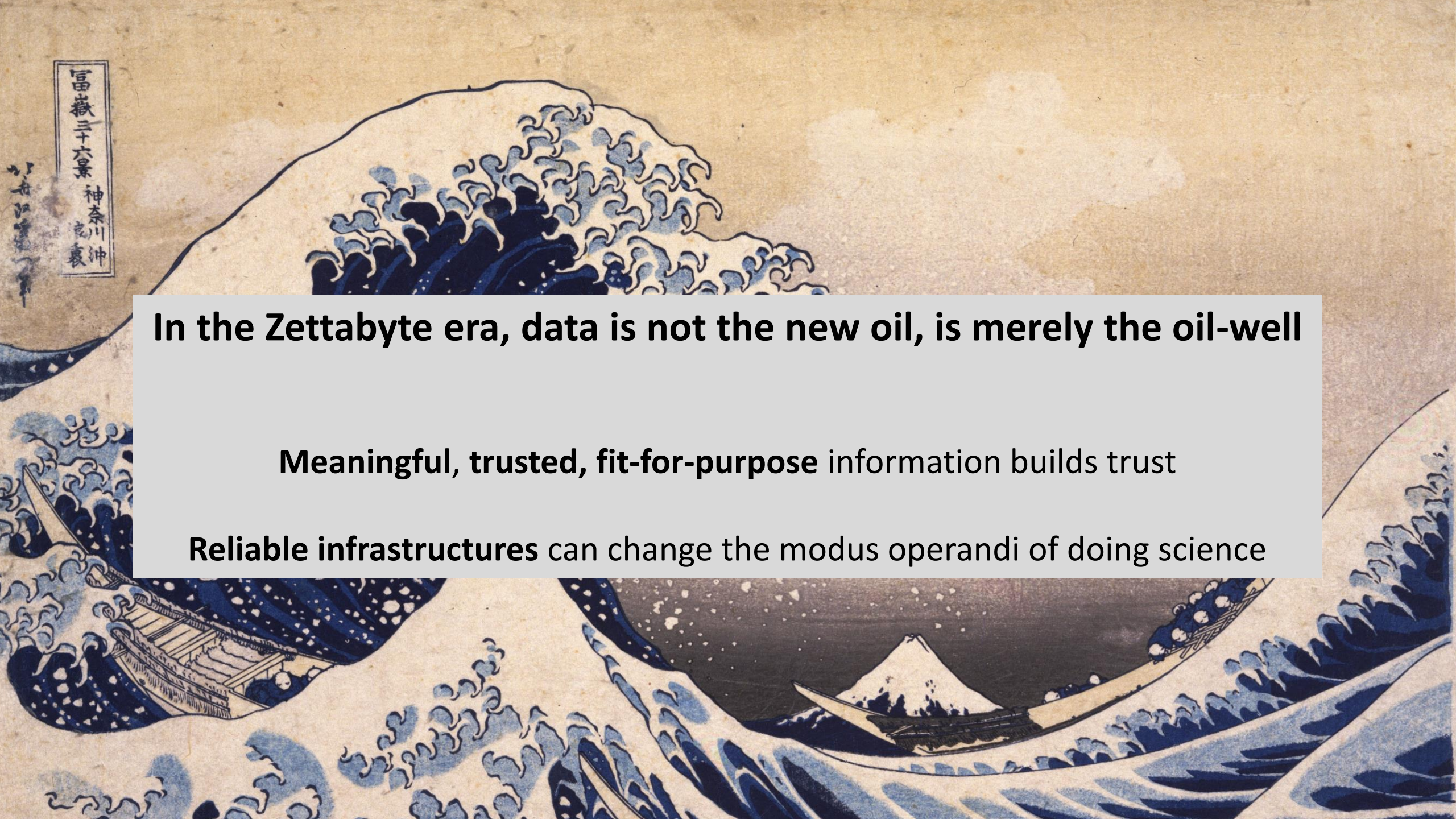


diagram from
Larry Lannom, CNRI

- users only deal with Digital Objects (Metadata and **PIDs**) - virtualisation
- components already being implementing, Clouds based on virtualisation
- working now on global testbed project (C2CAMP): DO operators, workflows, etc.

What Exactly are we Proposing to Do?

- Implement a prototype distributed environment based on the digital object model
 - Everything in the environment is a digital object
 - For basic information management tasks every object can be treated the same, regardless of information content
 - Every object has a globally unique and actionable identifier
 - Every object is typed
 - Every object has tightly associated metadata
 - Every object has a query-able set of operations that can be performed on it
- Start with the minimal set of components and services that enable the DO model
 - Identifiers + Resolution System
 - Types + Type Registries
 - DO Repositories, including repositories of metadata, aka registries
 - Mapping/brokering software & services to map existing data storage and management systems to DOs
 - Digital Object Interface Protocol, implemented by DO Repositories
- Open the environment to as many use cases as possible to hone the core infrastructural pieces



富嶽三十六景
神奈川
波
神

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*

How can DO-based architecture help build TRUST?

Relevance

Provenance

Attribution

Completeness

Fitness-for-purpose

Agility

Branding (Datatyping)

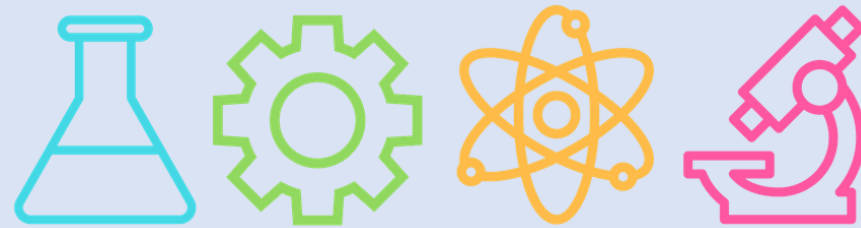
DOI Implementation could benefit from

Respect domain specific needs in terms of data

Operate within a trusted framework (a marketplace)

Deliver clear added value to existing practices

Develop in a future-proof way



Strong science cases need to drive implementation
Research Infrastructures need to be at the forefront

FAIR