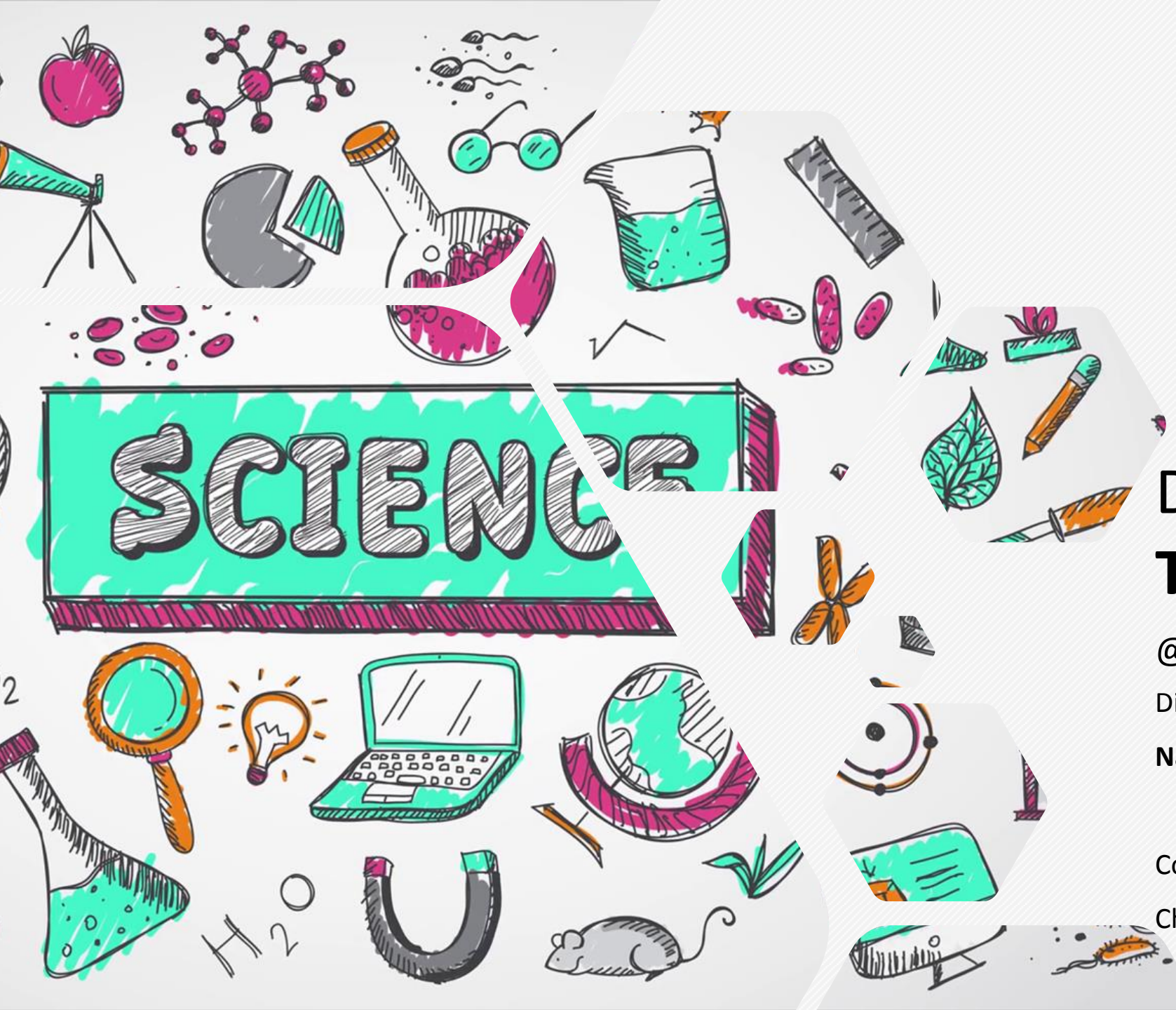


Digital Objects:
The Science Case

@DimitrisKoureas

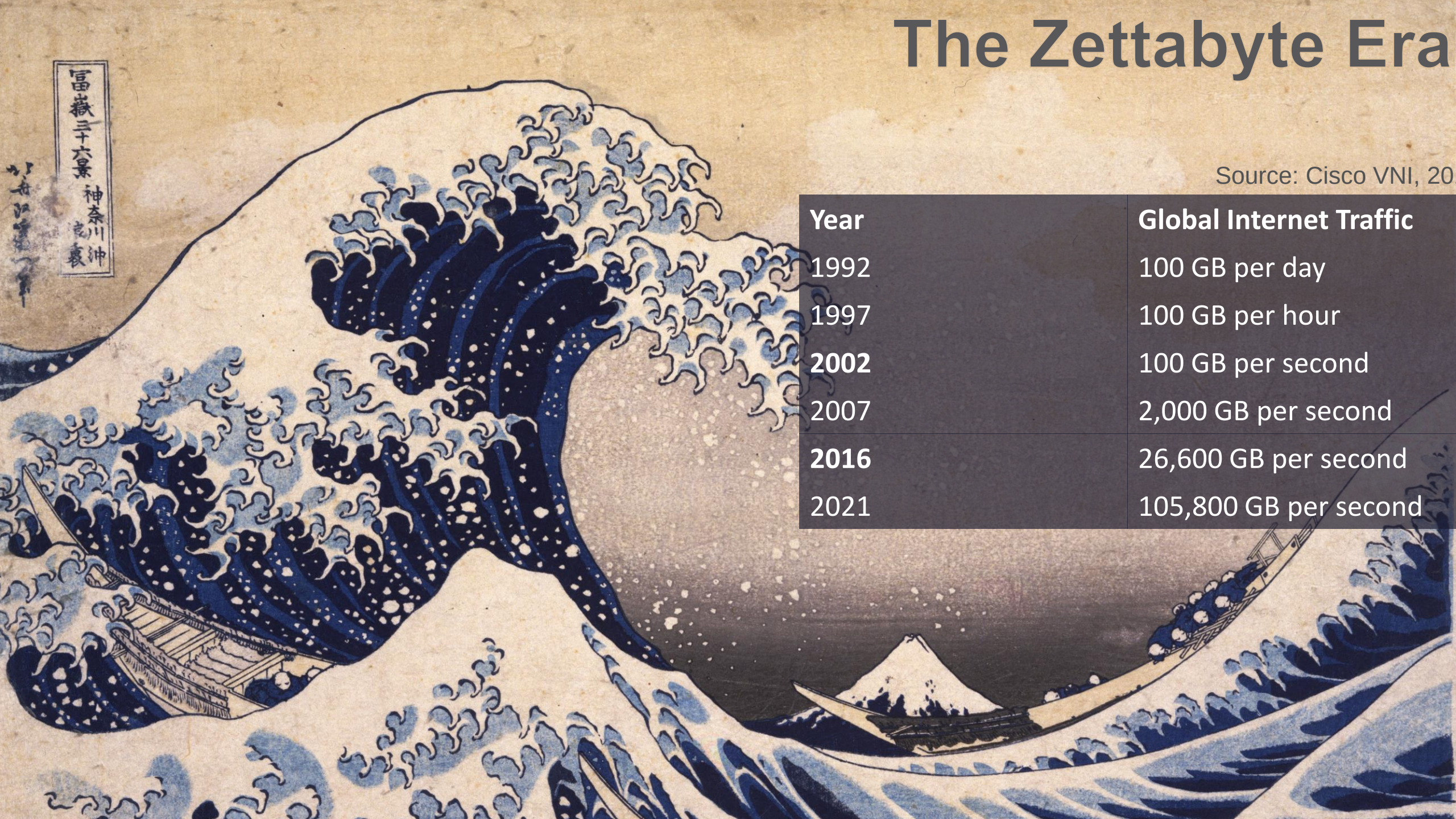
Director, International Biodiversity Infrastructures

Naturalis Biodiversity Center

Coordinator, **DiSSCo RI**

Chair, **Biodiversity Information Standards - TDWG**

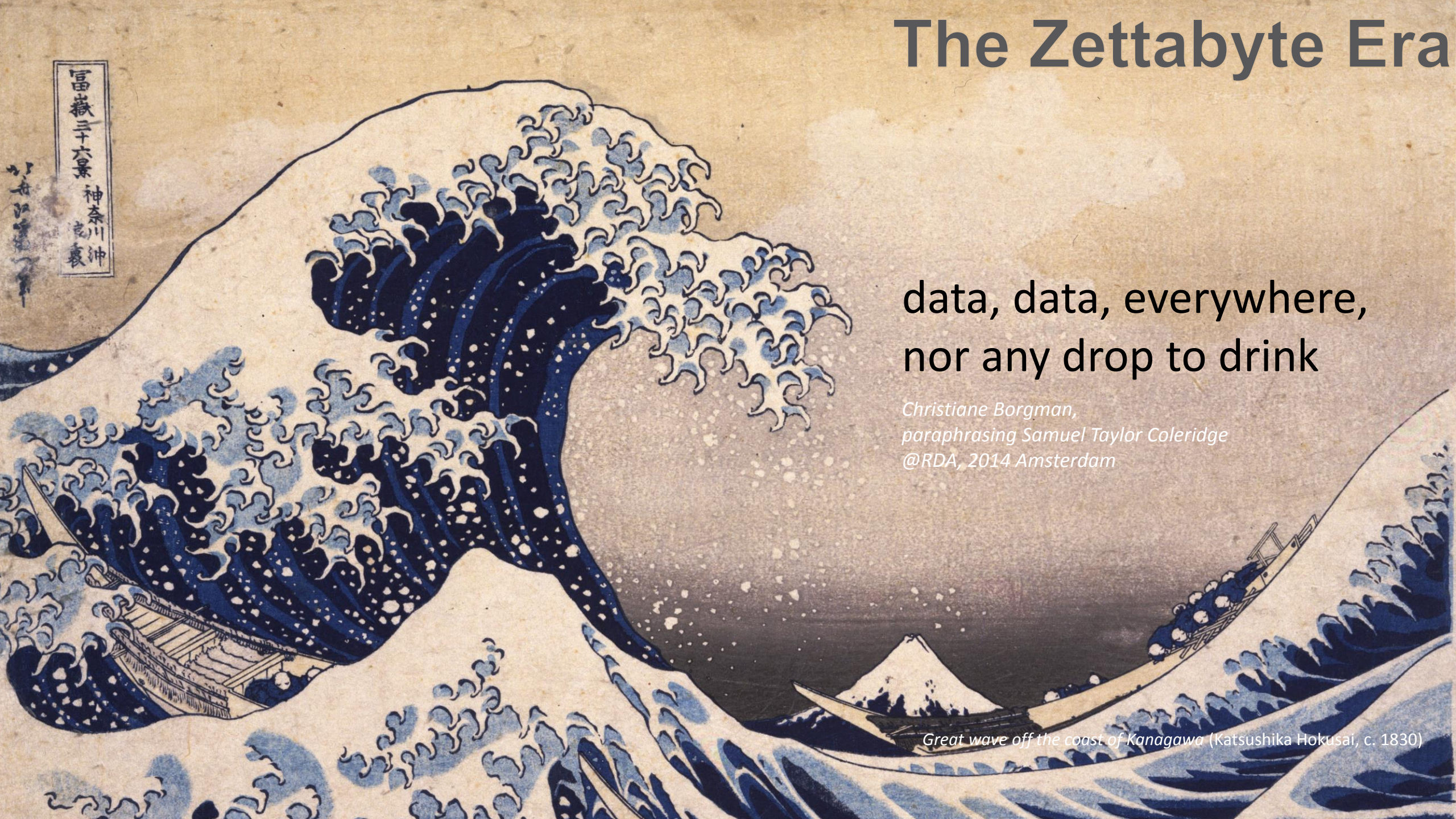
The Zettabyte Era



Source: Cisco VNI, 20

Year	Global Internet Traffic
1992	100 GB per day
1997	100 GB per hour
2002	100 GB per second
2007	2,000 GB per second
2016	26,600 GB per second
2021	105,800 GB per second

The Zettabyte Era



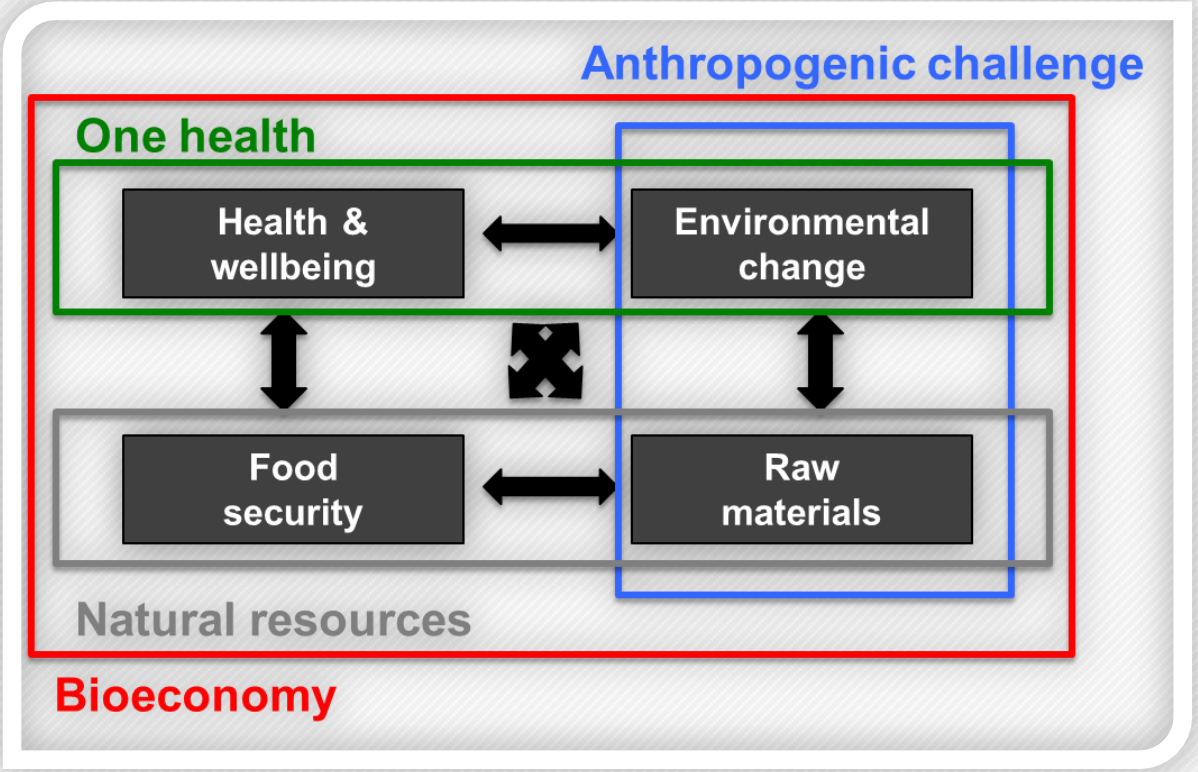
data, data, everywhere,
nor any drop to drink

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

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

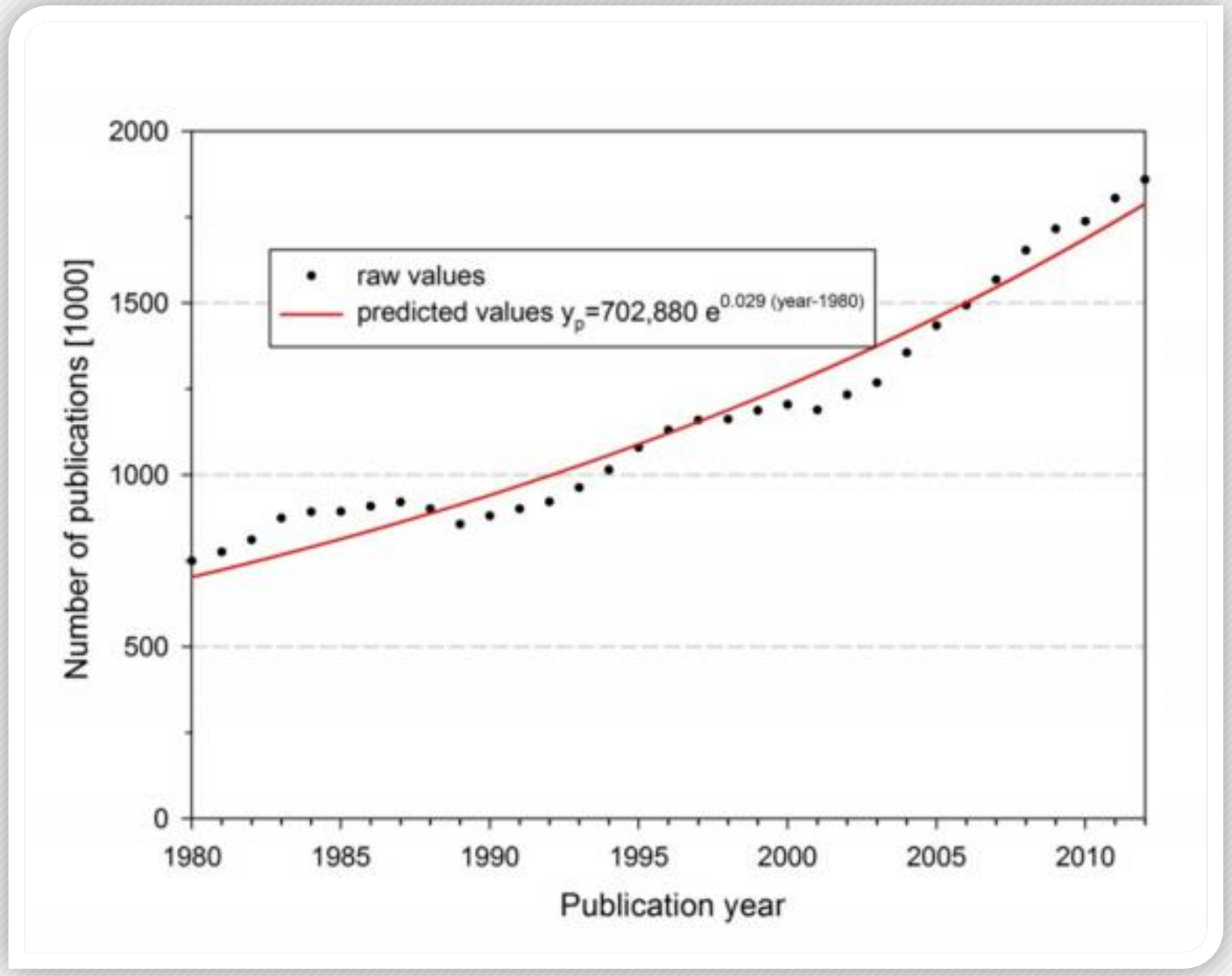
Our grand challenges require

Data-driven solutions



Need to deliver data at the

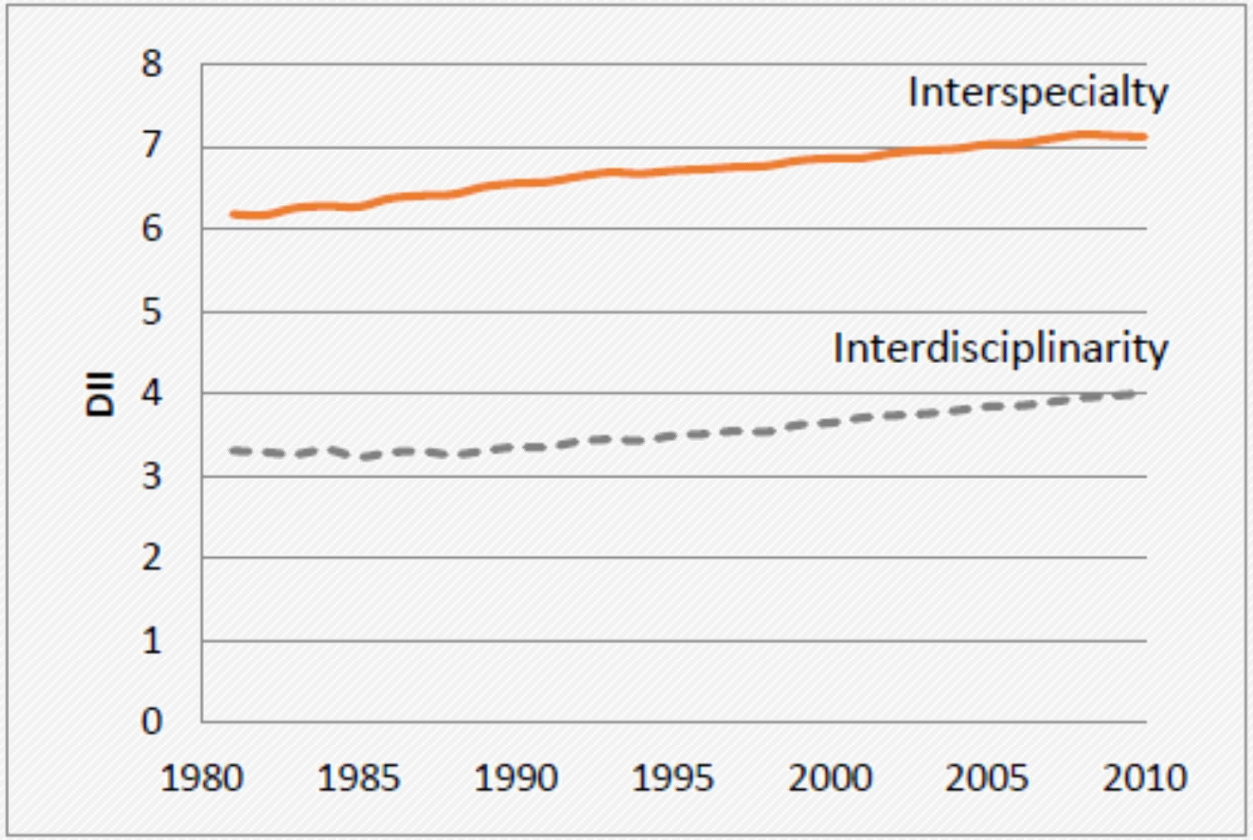
Scale, form and precision required



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

Ever-increasing rate of global scientific products

Does data ‘availability’ affect scientific output 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

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

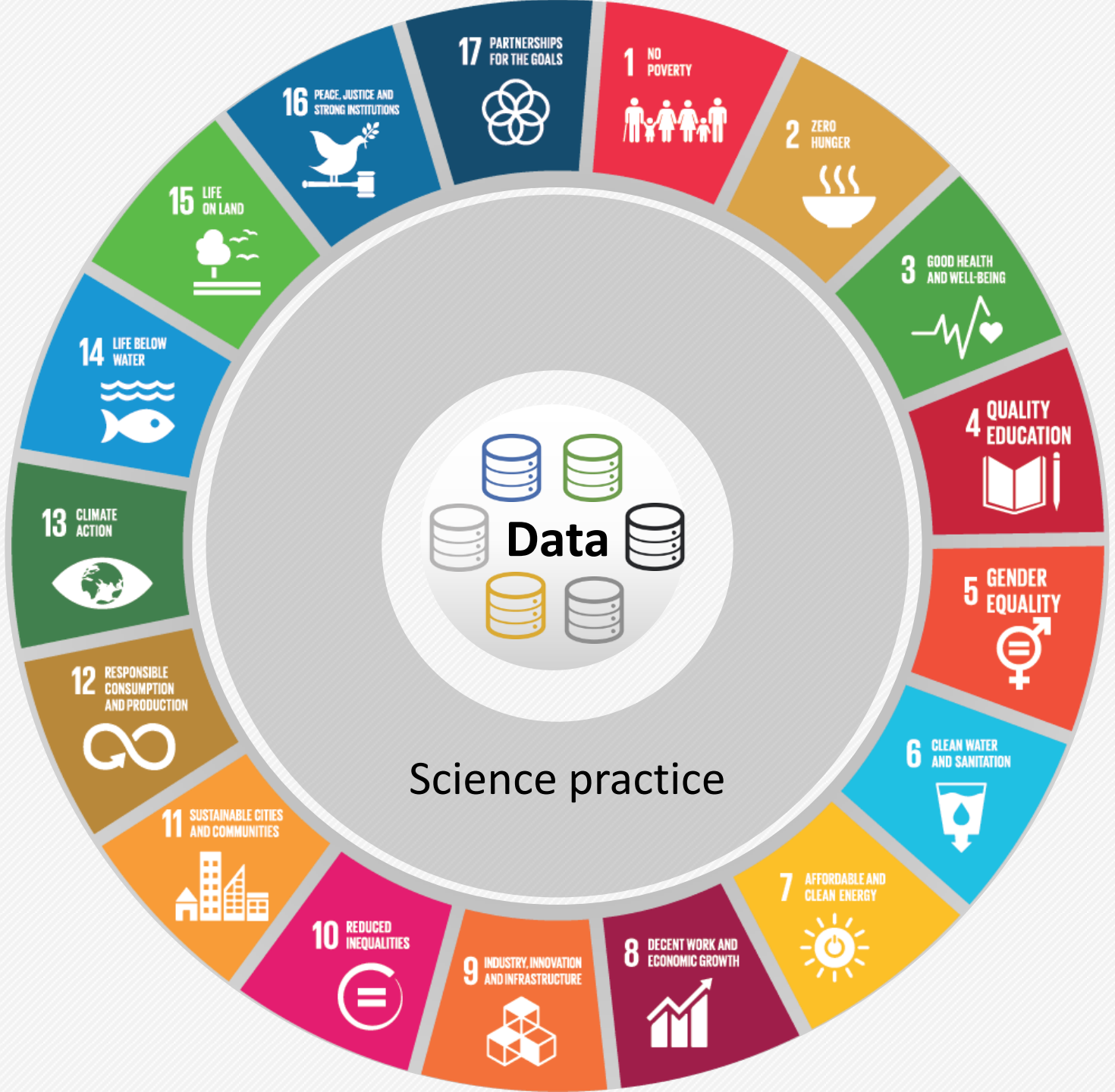
Challenges **global**

- 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


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

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

TRUST

Practices
People
Data

Disconnecting data from the context in which they were produced is one of the problems with the way we handle big-data for science

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

Meaningful & fit-for-purpose information builds trust and **reliable services** can change the modus operandi of doing science

How can DO-based architecture help build TRUST?

Relevance

Provenance

Attribution

Completeness

Fitness-for-purpose

Agility

Branding (Datatyping)

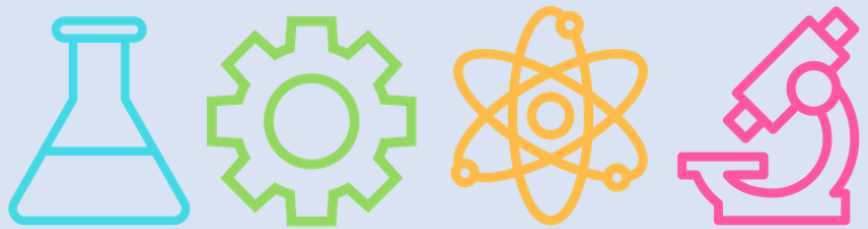
DOA Implementation could benefit from

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

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…

Scientific applications of DO architectures

Digitals Objects as direct input into the CLARIN Language Resource Switchboard	Twan Gosen, Dieter van Uytvanck (CLARIN)
How a Digital Object Architecture could help ICOS streamline data service provisioning	Margareta Hellström (ICOS)
DiSSCo Digital Specimens- Widening access to natural science collections	Alex Hardisty (DISSCO)
The DO Case in Virtual Atomic and Molecular Data Centre	Carlo Maria Zwölf (VAMDC)
Digital Object Management for ENES: Challenges and opportunities	Tobias Weigel (ENES)

Researchers have been making data FAIR throughout modern scientific practice

Discovering
Interpreting
Linking
Sharing

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

Investigator-focused 'small data'

Locally generated 'invisible data'

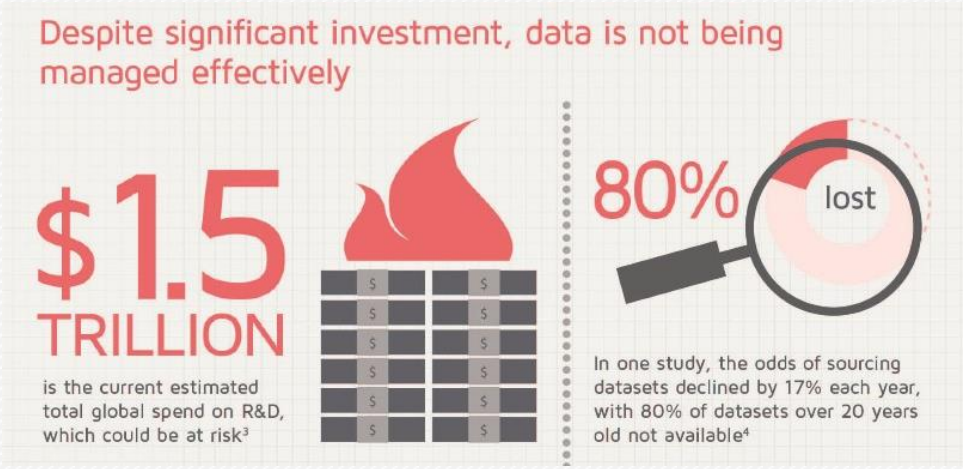
'incidental data'

80%
dark data

Published and discoverable data

20%

Dark data lost within 20 years



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