FACULTY OF ELECTRICAL ENGINEERING, MATHEMATICS AND COMPUTER SCIENCE

SEMANTICS, CYBERSECURITY AND SERVICES GROUP

UNIVERSITY OF TWENTE.

Leiden University Medical Center

FAIR Principles and Open Science

LUIZ BONINO

CC BY

OPEN SCIENCE

Leiden University Medical Center
UNIVERSITY OF TWENTE.

OPEN SCIENCE

"Open Science is defined as an inclusive construct that combines various movements and practices aiming to ***make*** *multilingual* ***scientific knowledge*** *openly* ***available****,* ***accessible*** *and* ***reusable*** *for everyone, to increase scientific collaborations and sharing of information for the benefits of science and society, and to* ***open*** *the* ***processes*** *of scientific* ***knowledge creation****,* ***evaluation*** *and* ***communication*** *to societal actors beyond the traditional scientific community."*

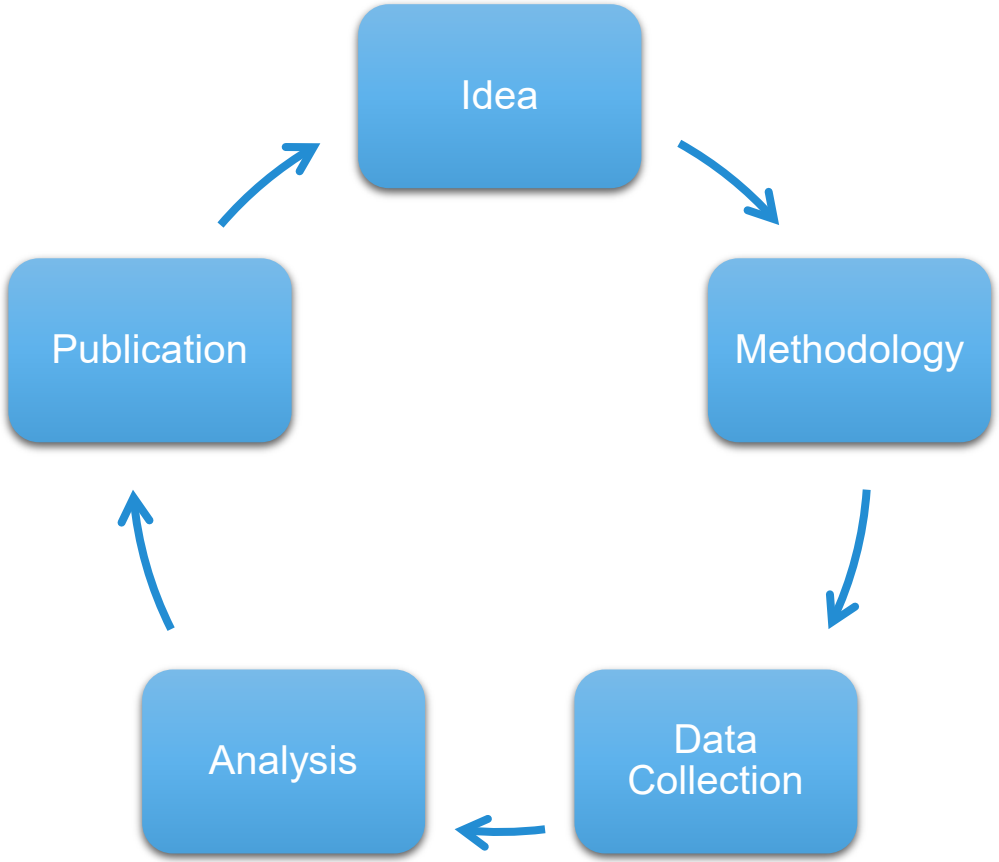
UNESCO Recommendation on Open Science, 2021

https://unesdoc.unesco.org/ark:/48223/pf0000379949

Leiden University Medical Center

UNIVERSITY OF TWENTE.

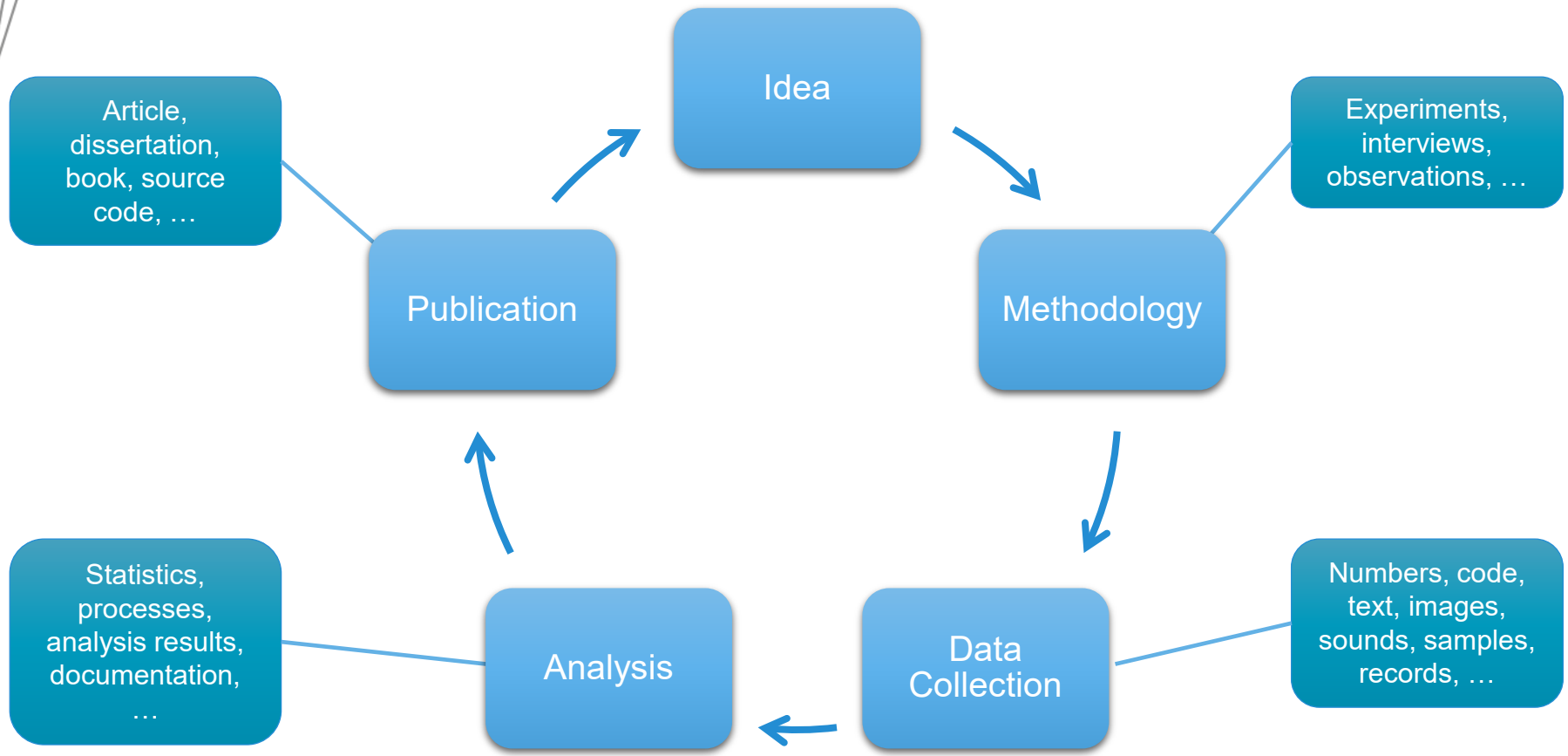
RESEARCH LIFECYCLE



Leiden University Medical Center

UNIVERSITY OF TWENTE.

RESEARCH LIFECYCLE



Leiden University Medical Center

UNIVERSITY OF TWENTE.

BENEFITS OF OPEN SCIENCE

- Increases **research efficiency**
- Promotes scholarly rigour and enhances **research quality**
- Enhances **visibility** and engagement
- Enables the creation of **new research questions**
- Enhances **collaboration** and community building

Leiden University Medical Center

UNIVERSITY OF TWENTE.

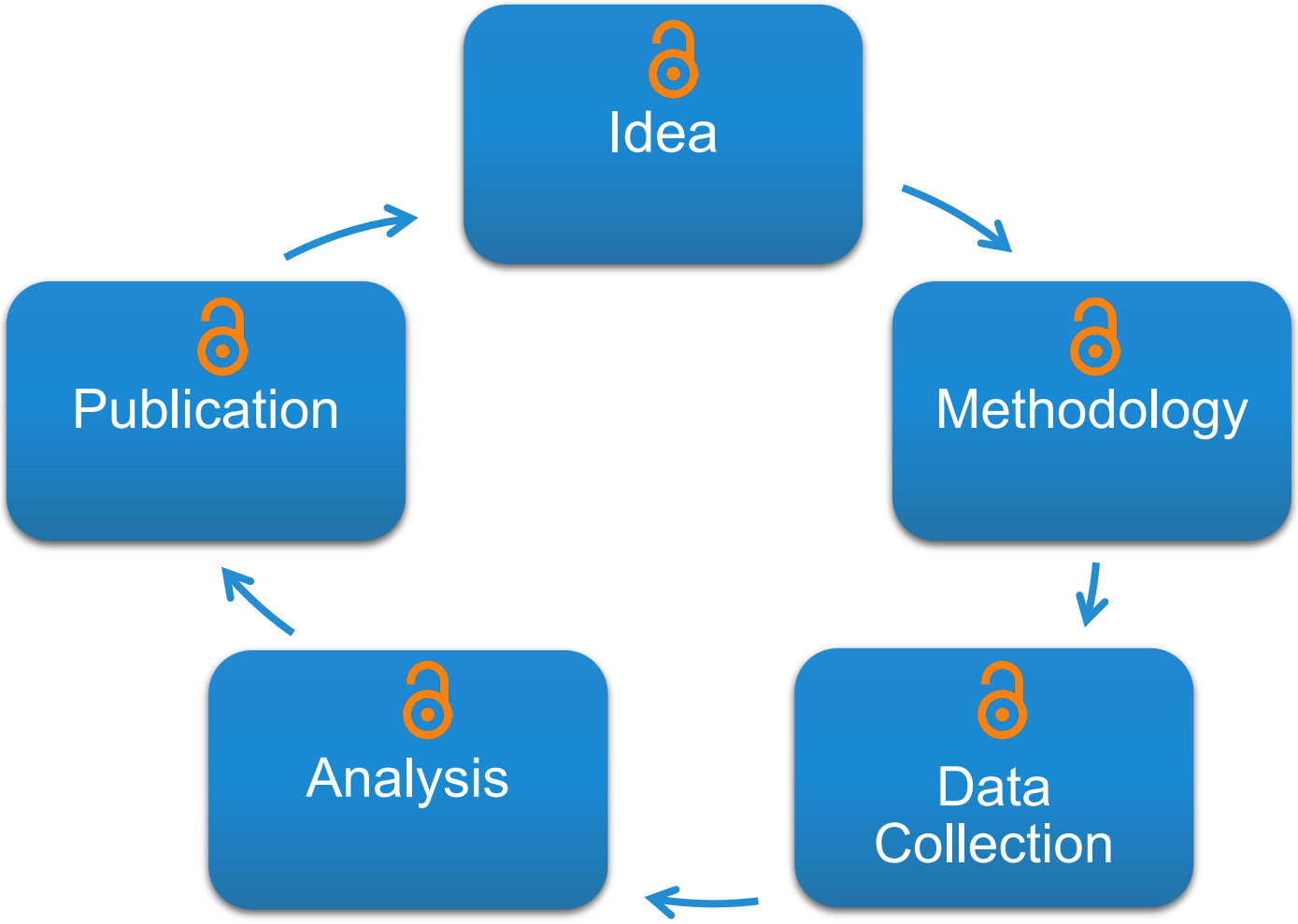
RESEARCH LIFECYCLE

"… open the processes of scientific knowledge creation, evaluation and communication …" – UNESCO 2021

Leiden University Medical Center

UNIVERSITY OF TWENTE.

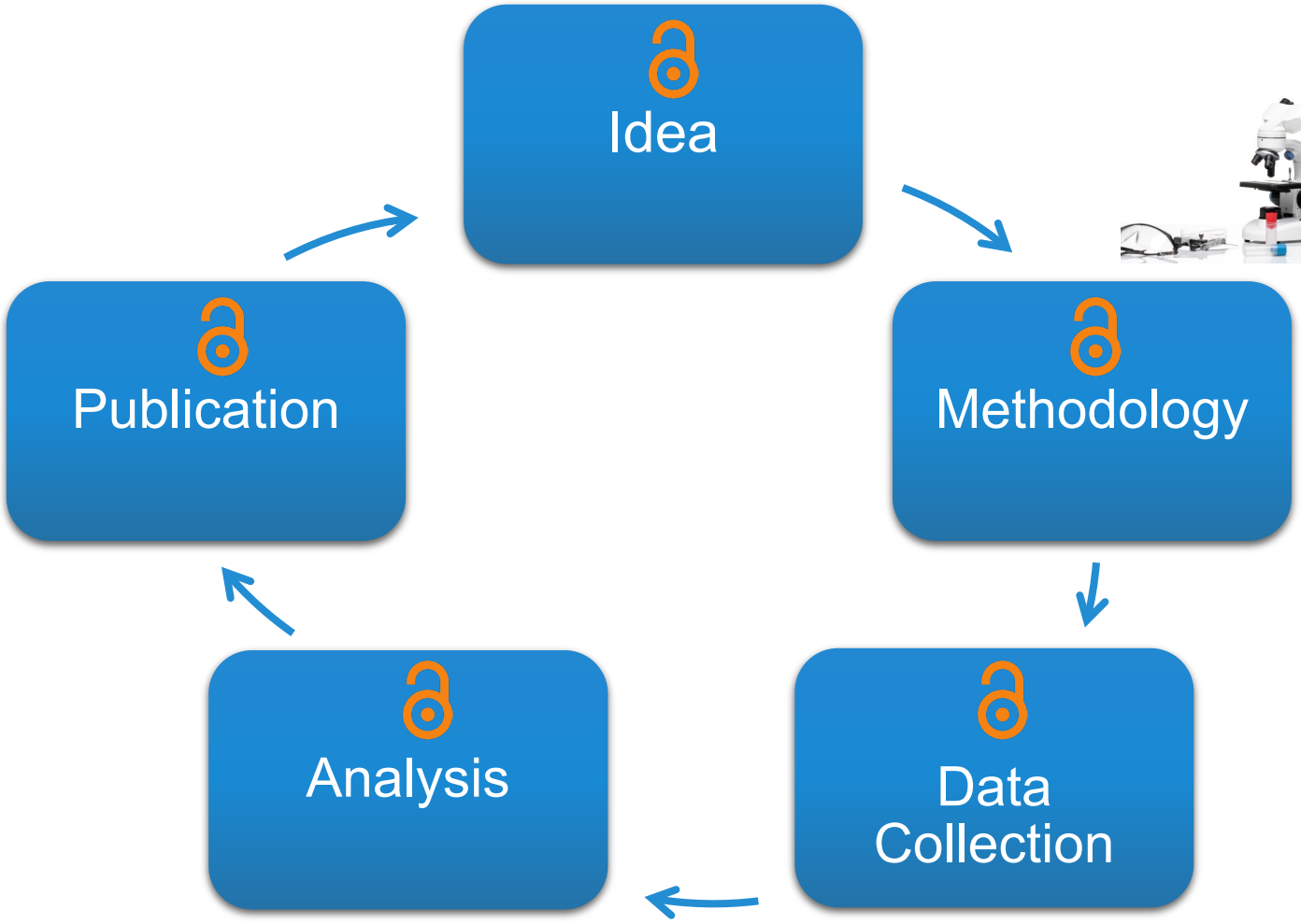
IMPROVE RESEARCH LIFECYCLE



Leiden University Medical Center

UNIVERSITY OF TWENTE.

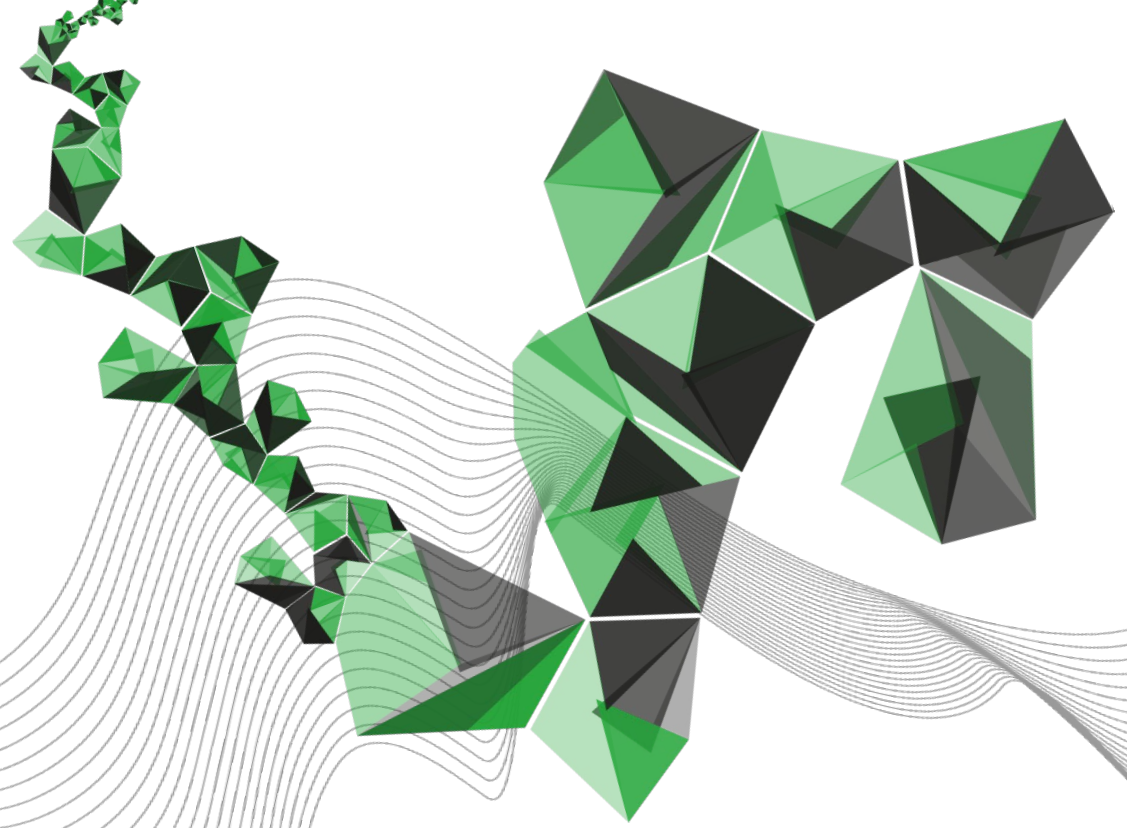
IMPROVE RESEARCH LIFECYCLE



Leiden University Medical Center

UNIVERSITY OF TWENTE.

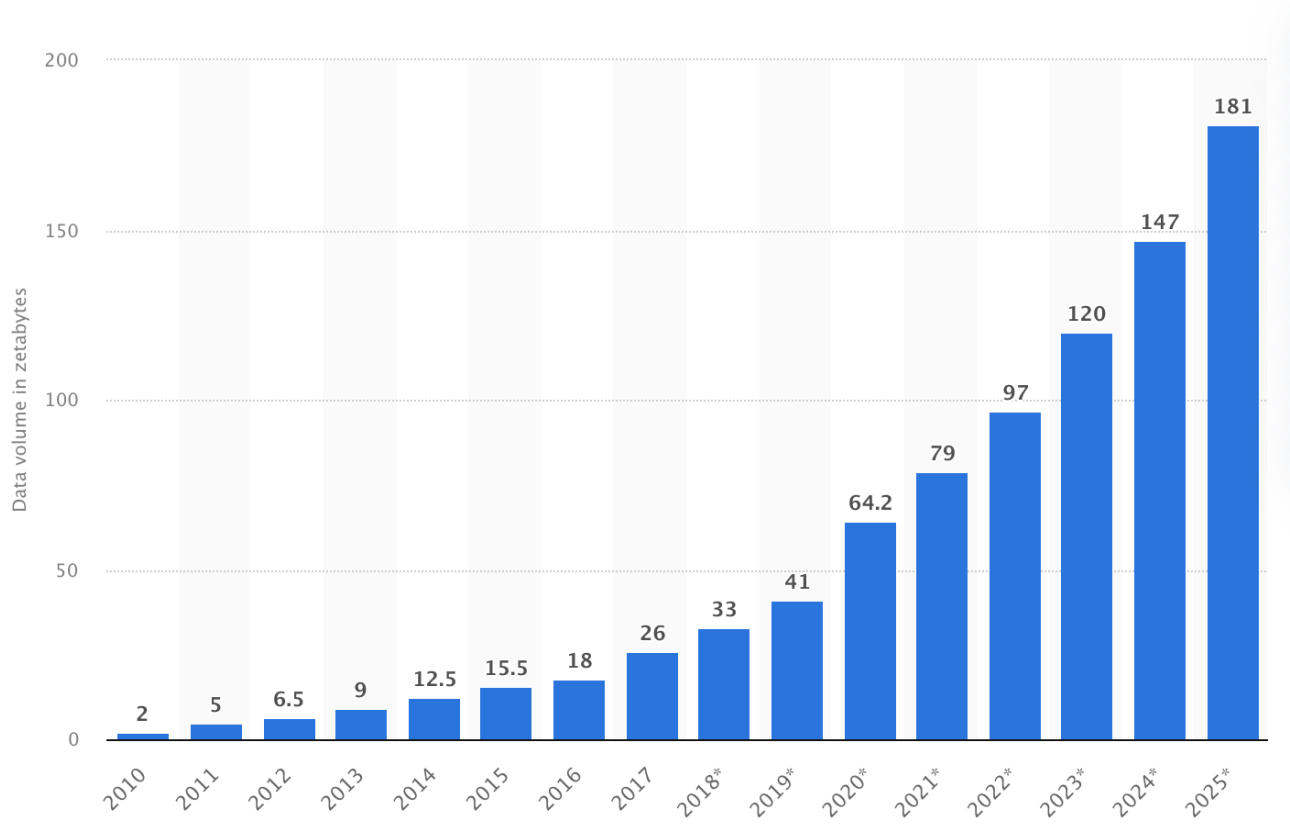
THE DATA SITUATION



Leiden University Medical Center

UNIVERSITY OF TWENTE.

STAGGERING DATA GROWTH



Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2025 *(in zettabytes)*

Source: Statista

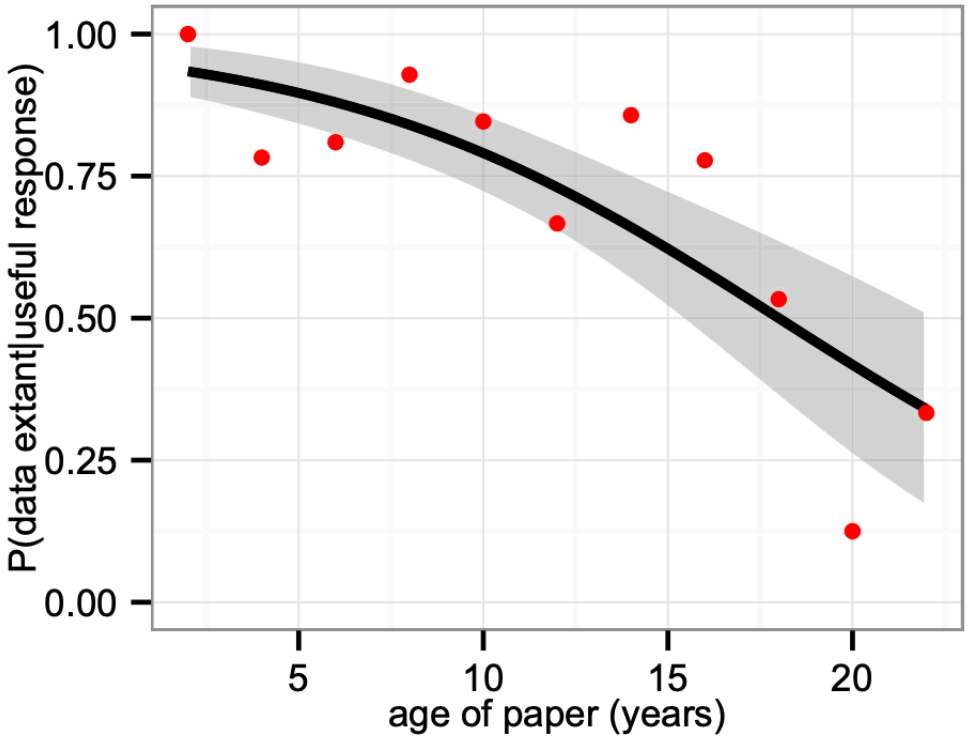
https://www.statista.com/statistics/871513/worldwide-data-created/

Leiden University Medical Center

UNIVERSITY OF TWENTE.

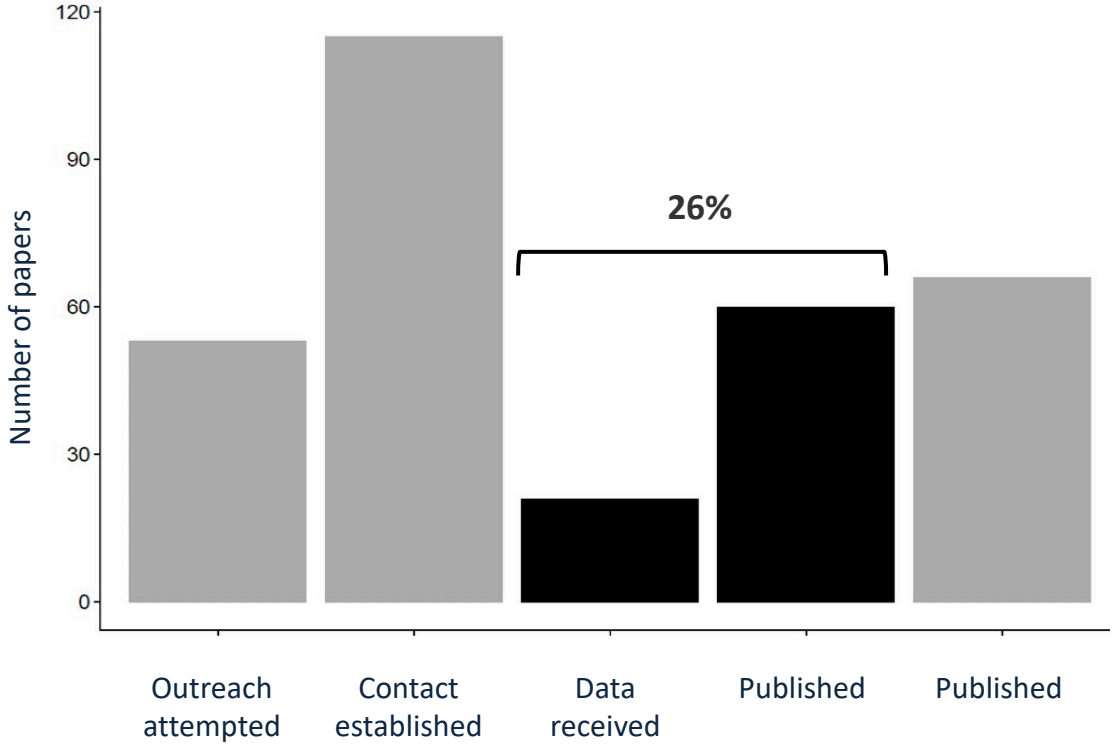
BUT CAN WE ACCESS THESE DATA?

..the odds of a **dataset being extant fell by 17% per year**. In addition, the odds that we could find a working e-mail address for the first, last, or **corresponding author fell by 7% per year**.



The Availability of Research Data Declines Rapidly with Article Age. December 19, 2013. Current Biology. https://doi.org/10.1016/j.cub.2013.11.014

"..we tested the ability to recover data collected under a particular **funder-imposed requirement of public availability**. Overall the majority of data were not recovered **(26% recovery of 315 data projects…**"



A funder-imposed data publication requirement seldom inspired data sharing. July 2018, PLoS ONE. http://dx.doi.org/10.1371/journal.pone.0199789

Leiden University Medical Center

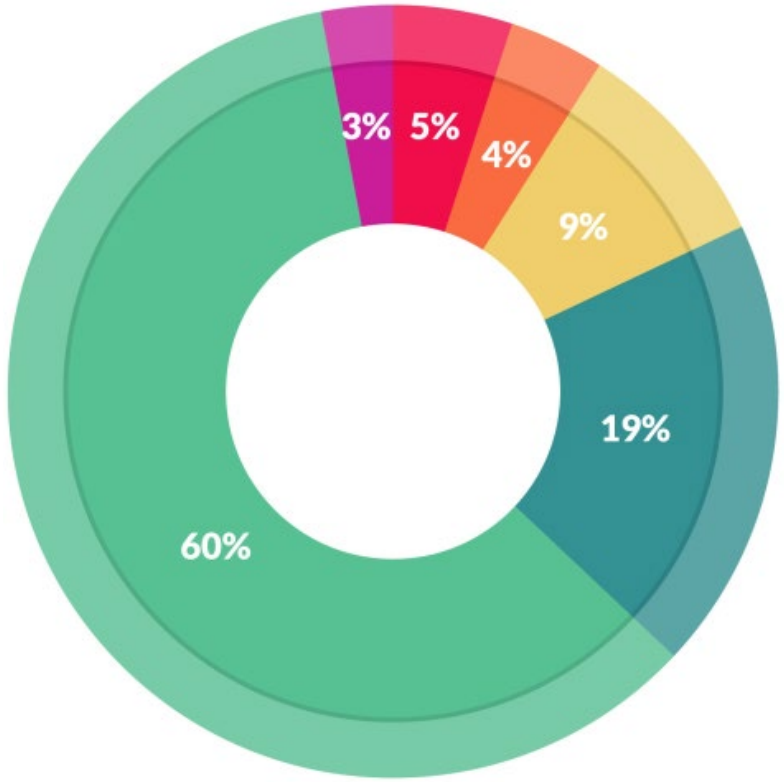
UNIVERSITY OF TWENTE.

AND WHEN WE FIND THEM, THEY ARE SILOED

Leiden University Medical Center

UNIVERSITY OF TWENTE.

DATA EXPERT EFFORT



What data scientists spend the most time doing

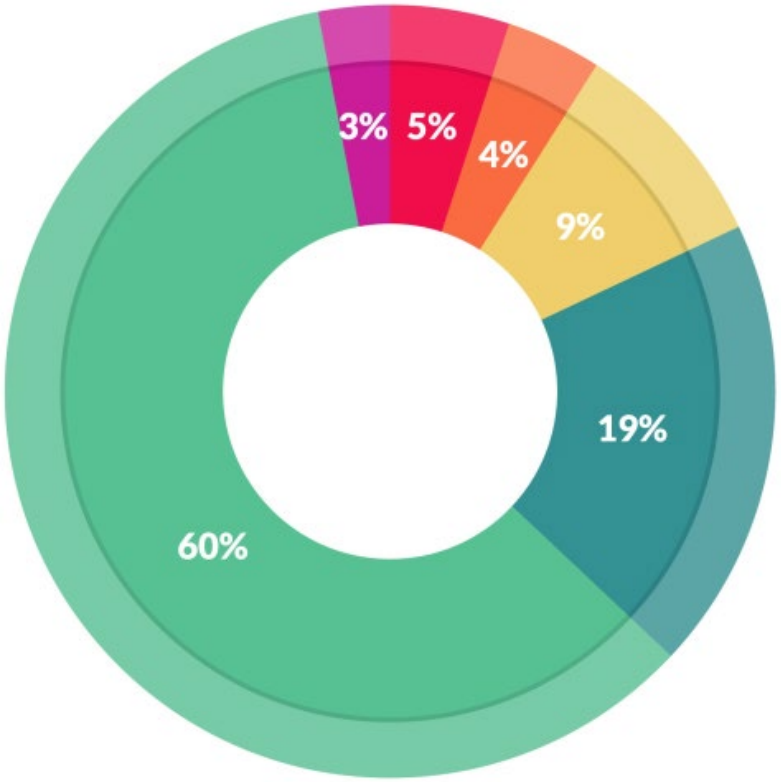
- Collecting data sets: 19%
- Cleaning and organizing data: 60%
- Building training sets: 3%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%

Source: Data Science Report 2016, CrowdFlower, 2016: http://visit.crowdflower.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf

Leiden University Medical Center

UNIVERSITY OF TWENTE.

INEFFICIENT USE OF DATA EXPERTISE AND RESOURCES



What data scientists spend the most time doing

- Collecting data sets: 19%
- Cleaning and organizing data: 60%

79%

- Building training sets: 3%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%

21%

Source: Data Science Report 2016, CrowdFlower, 2016: http://visit.crowdflower.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf

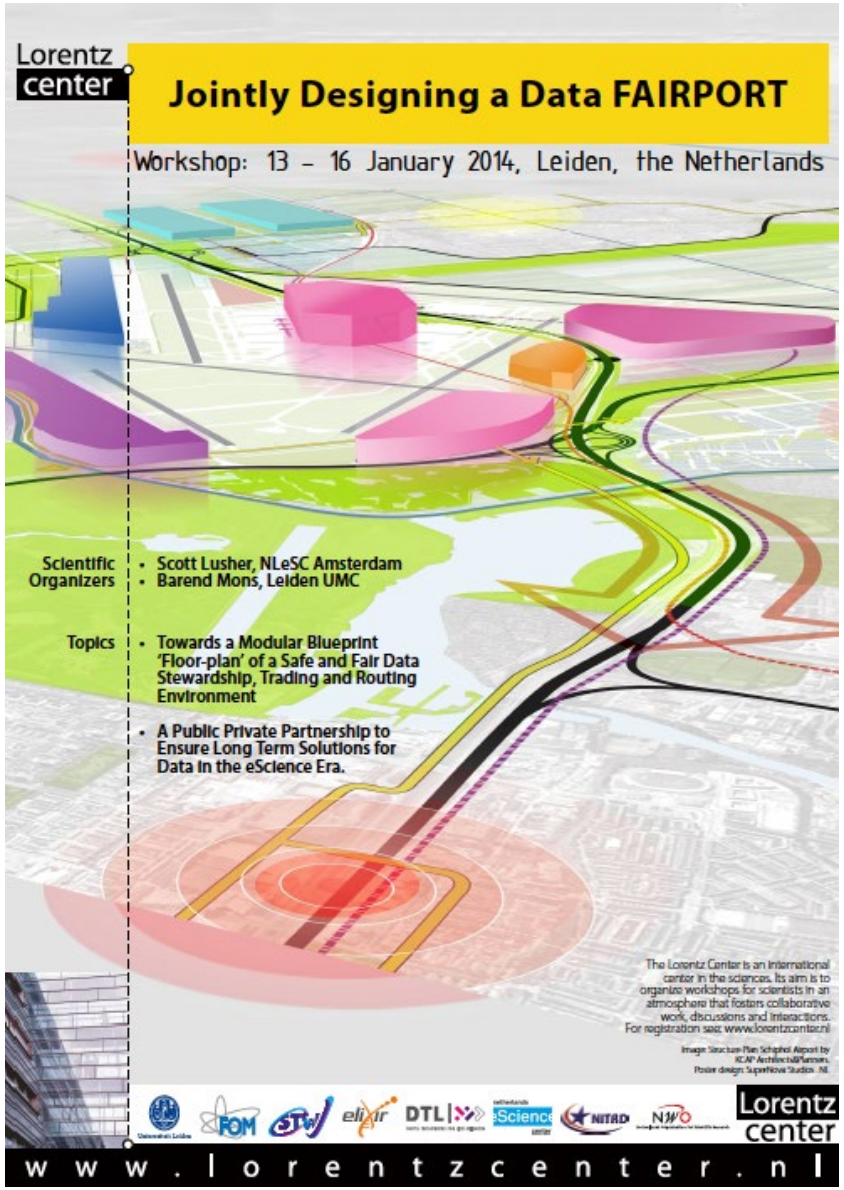
Leiden University Medical Center

UNIVERSITY OF TWENTE.

THE FAIR PRINCIPLES

Leiden University Medical Center
UNIVERSITY OF TWENTE.

FAIR PRINCIPLES GENESIS – JANUARY 2014



Organized by:

DTL
DUTCH TECHCENTRE FOR LIFE SCIENCES

netherlands eScience center

17

Leiden University Medical Center

UNIVERSITY OF TWENTE.

FAIR HISTORY – MARCH 2016

SCIENTIFIC DATA

IN PRESS

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, Jildau Bouwman, Anthony J Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J G Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C. 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons

http://www.nature.com/sdata/ nature publishing group npg

The FAIR Guiding Principles for scientific data management and stewardship
MD Wilkinson, M Dumontier, IJJ Aalbersberg… - Scientific data, 2016
Cited by 7392 Related articles All 44 versions

Leiden University Medical Center

UNIVERSITY OF TWENTE.

THE FAIR PRINCIPLES

Findable:

F1. (meta)data are assigned a globally unique and persistent identifier;

F2. data are described with rich metadata;

F3. metadata clearly and explicitly include the identifier of the data it describes;

F4. (meta)data are registered or indexed in a searchable resource;

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;

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;

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;

https://www.nature.com/articles/sdata201618

FAIR principles: data (or other digital objects), metadata & enablers

Findable:

F1. (meta)data are assigned a globally unique and persistent identifier;

F2. data are described with rich metadata;

F3. metadata clearly and explicitly include the identifier of the data it describes;

F4. (meta)data are registered or indexed in a searchable resource;

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;

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;

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;

THE FAIR PRINCIPLES AND THEIR MULTIPLE FACETS

Findable:

F1. (meta)data are assigned a globally unique and persistent identifier;

F2. data are described with rich metadata;

F3. metadata clearly and explicitly include the identifier of the data it describes;

F4. (meta)data are registered or indexed in a searchable resource;

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;

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;

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;

Leiden University Medical Center

UNIVERSITY OF TWENTE.

THE PRINCIPLES TARGET PRIMARILY MACHINES

Leiden University Medical Center

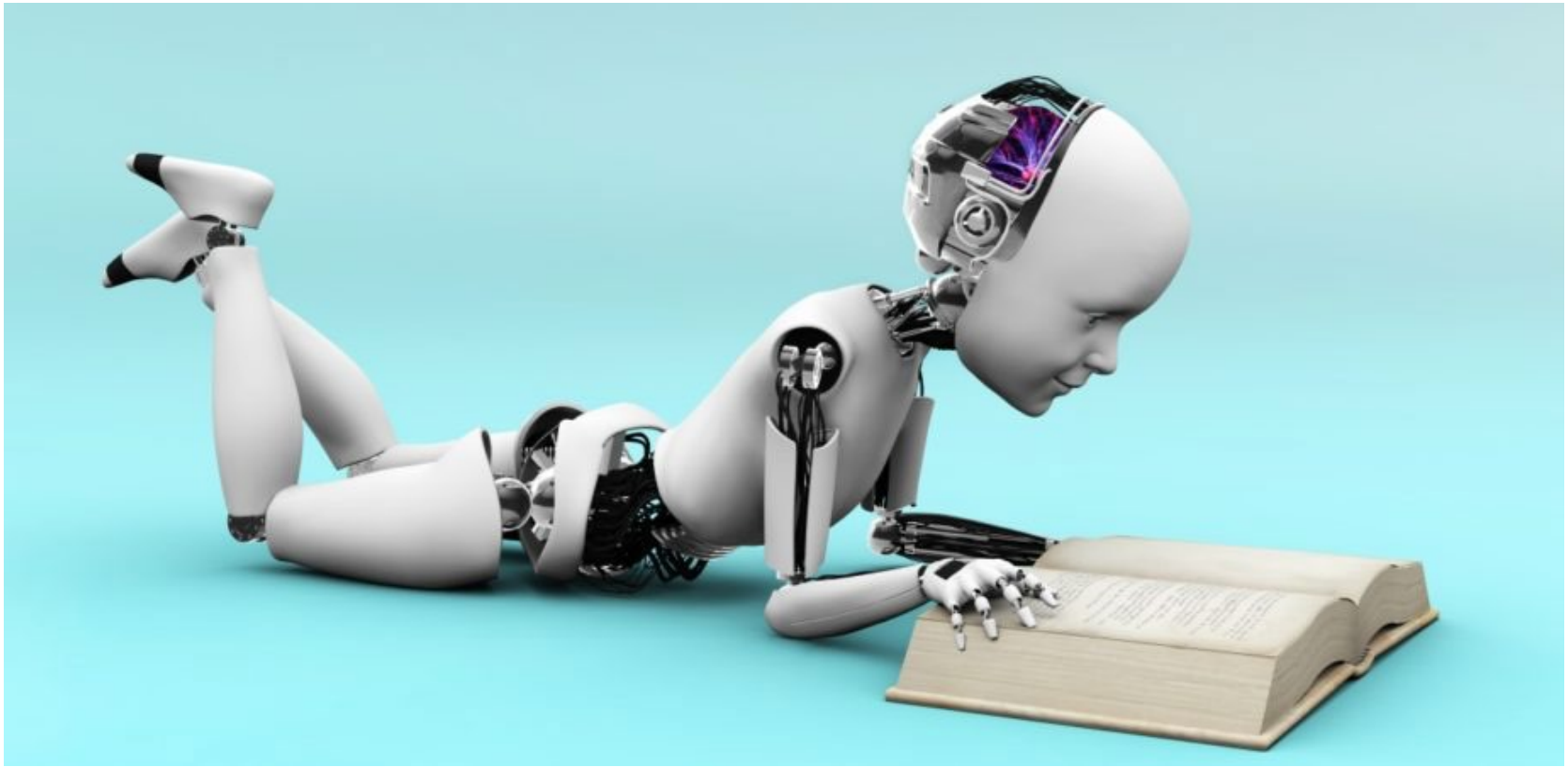
UNIVERSITY OF TWENTE.

THEN THE MACHINES CAN HELP US

Leiden University Medical Center

UNIVERSITY OF TWENTE.

BUT FIRST THEY (AND US) NEED TO "UNDERSTAND WHAT WE MEAN"



Leiden University Medical Center

UNIVERSITY OF TWENTE.

CONTACT INFO

UNIVERSITY OF TWENTE.

Leiden University Medical Center

Luiz Bonino

Associate Professor – SCS – U. Twente
Associate Professor – BioSemantics – LUMC

E-mail: **l.o.boninodasilvasantos@utwente.nl**
Skype: **luizolavobonino**