

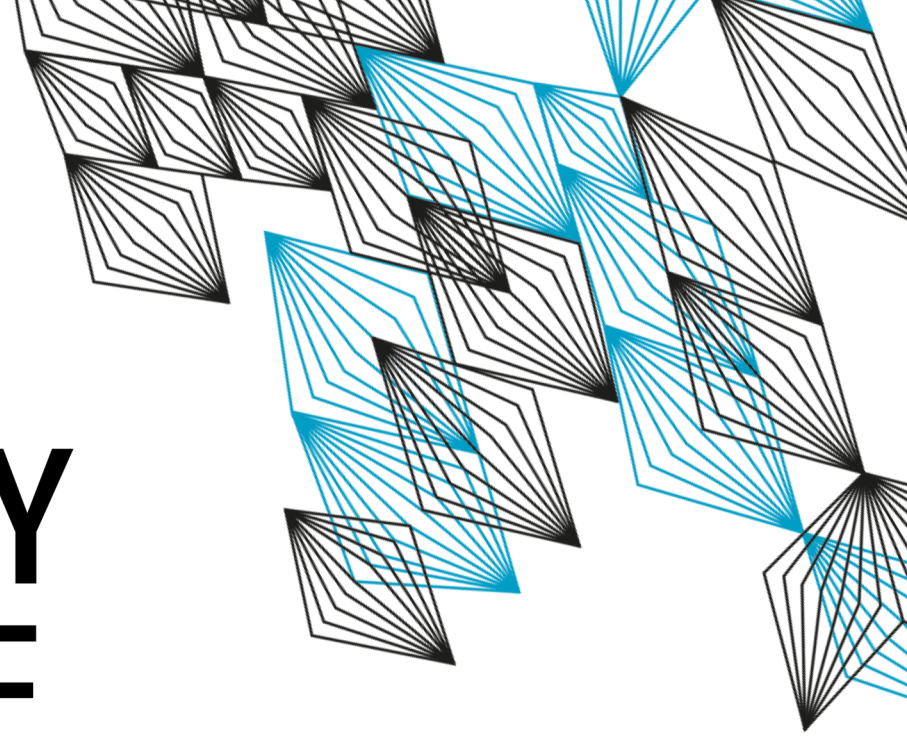
UNIVERSITY OF TWENTE.

THE FAIR PRINCIPLES

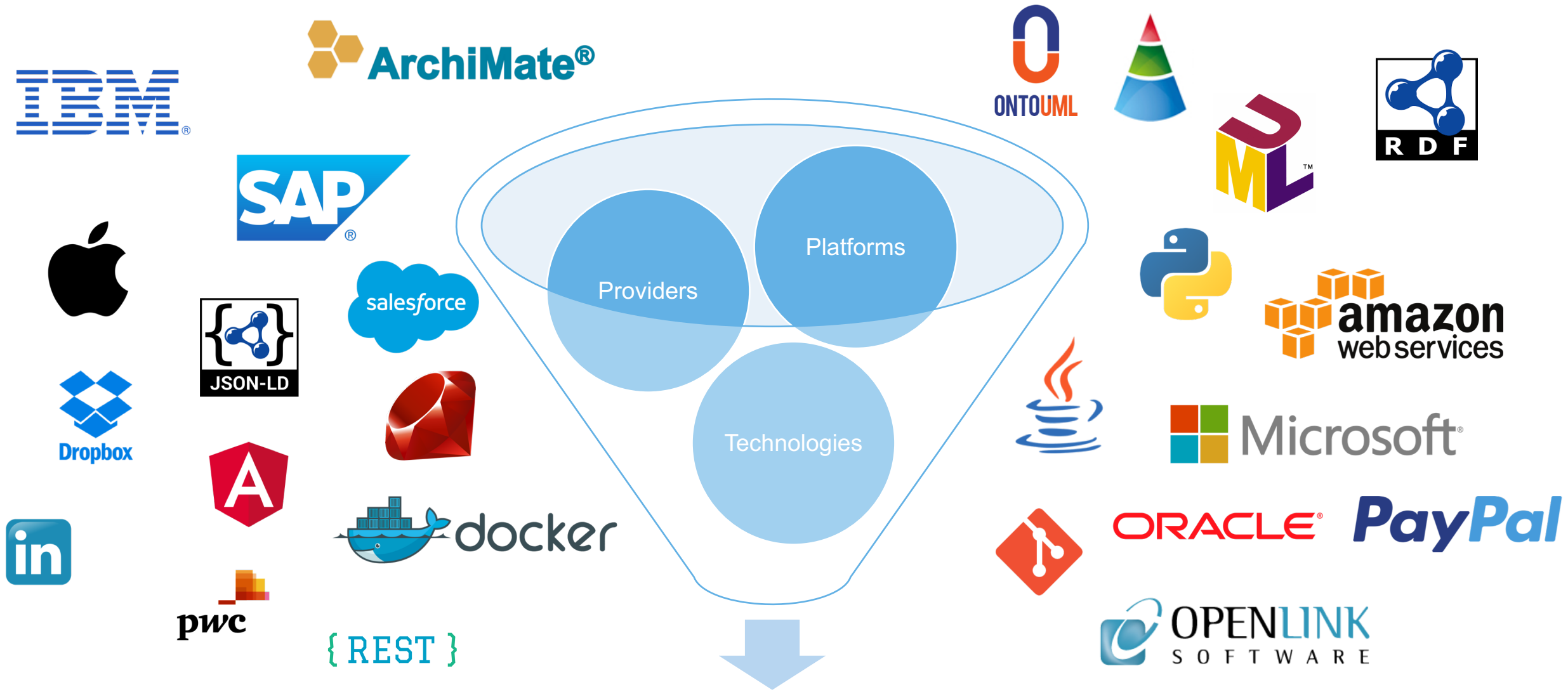
DCC MONTHLY THEMATIC SESSION

JUNE 28, 2021

LUIZ BONINO

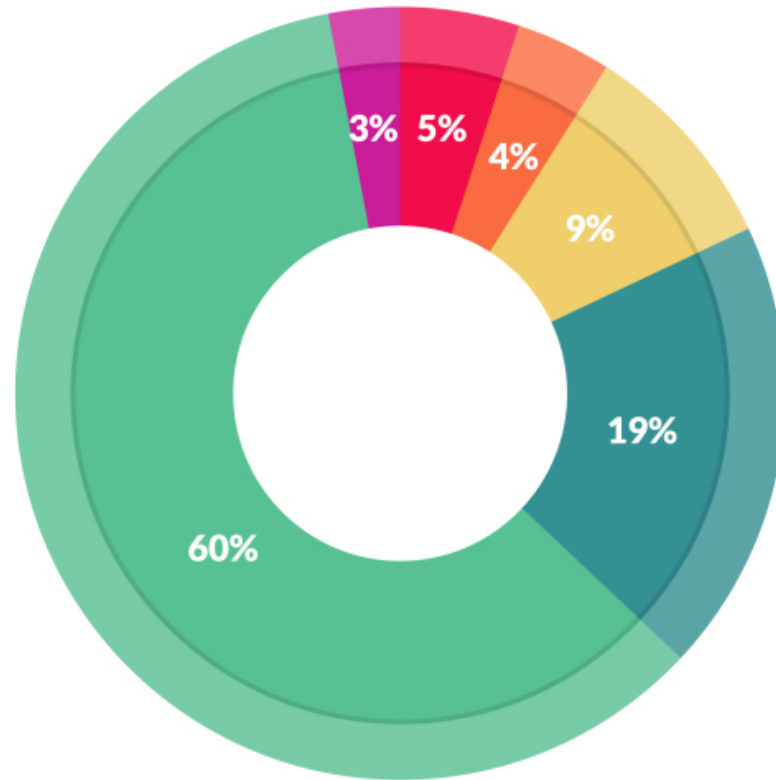


THE HETEROGENEOUS REALITY



Integrated Organization

DATA EXPERT EFFORT

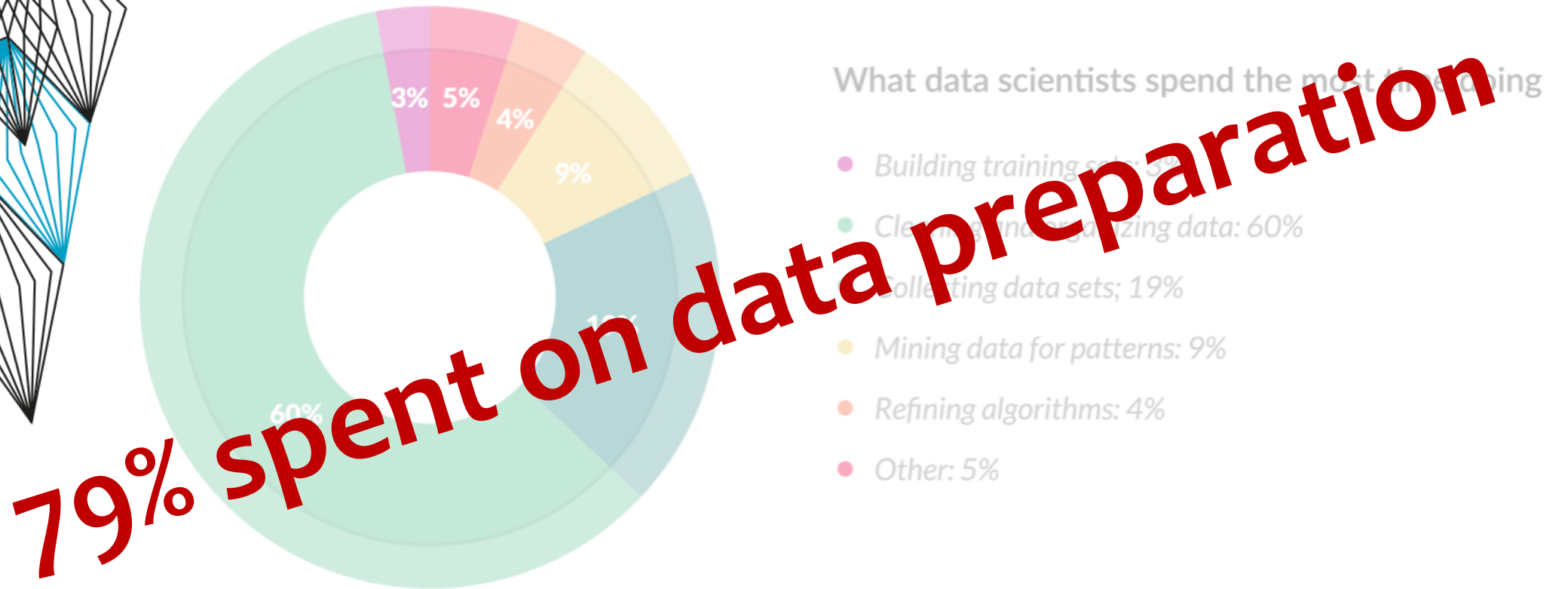


What data scientists spend the most time doing

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

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

DATA EXPERT EFFORT



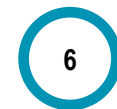
Source: Data Science Report 2016, CrowdFlower, 2016: http://visit.crowdfunder.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf

**F
I
N
D
A
B
L
E**



**F
I
N
D
A
B
L
E**

**A
C
C
E
S
S
I
B
L
E**



**F
I
N
D
A
B
L
E**

**A
C
C
E
S
S
I
B
L
E**

**I
N
T
E
R
O
P
E
R
A
B
L
E**



**F
I
N
D
A
B
L
E**

**A
C
C
E
S
S
I
B
L
E**

**I
N
T
E
R
O
P
E
R
A
B
L
E**

**R
E
U
S
A
B
L
E**

FAIR

I
N
D
A
B
L
E

C
C
E
S
S
I
B
L
E

N
T
E
R
O
P
E
R
A
B
L
E

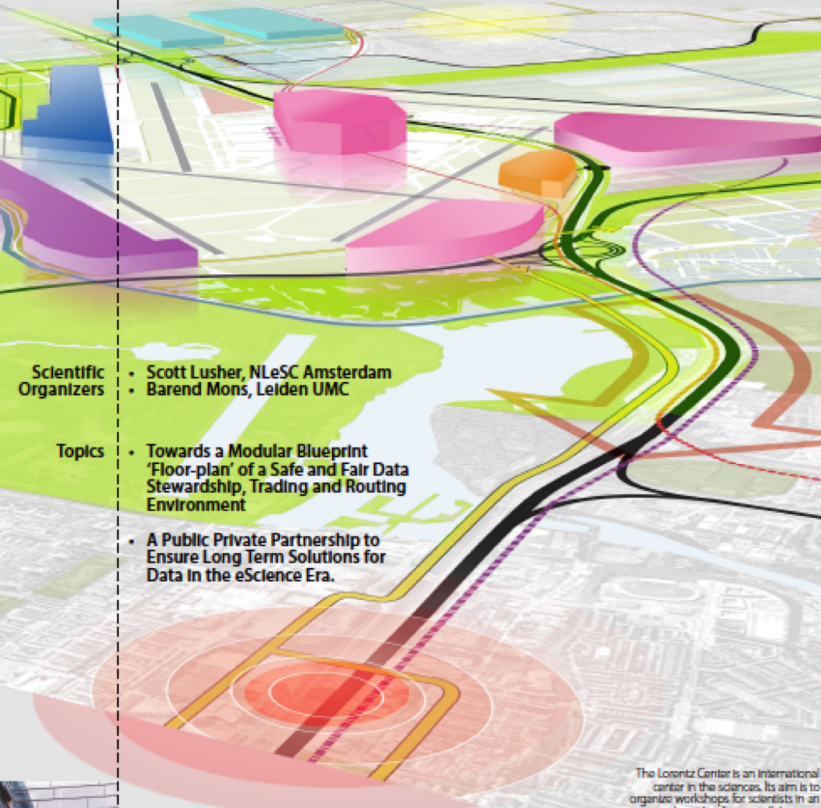
E
U
S
A
B
L
E

FAIR HISTORY – JANUARY 2014

Lorentz center

Jointly Designing a Data FAIRPORT

Workshop: 13 – 16 January 2014, Leiden, the Netherlands



Scientific Organizers

- Scott Lusher, NLeSC Amsterdam
- Barend Mons, Leiden UMC

Topics


- Towards a Modular Blueprint 'Floor-plan' of a Safe and Fair Data Stewardship, Trading and Routing Environment
- A Public Private Partnership to Ensure Long Term Solutions for Data in the eScience Era.

The Lorentz Center is an international center in the sciences. Its aim is to organize workshops for scientists in an atmosphere that fosters collaborative work, discussions and interactions. For registration see: www.lorentzcenter.nl

Image: Structure Plan 'Schiphol Airport' by ICAP Architectuur/Urban, Poster design: Superflex/Studio, NL

Lorentz center

www.lorentzcenter.nl



Organized by:



netherlands **eScience** 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 Boume, 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

The FAIR Guiding Principles for scientific data management and stewardship

MD Wilkinson, M Dumontier, IJJ Aalbersberg... - Scientific data, 2016

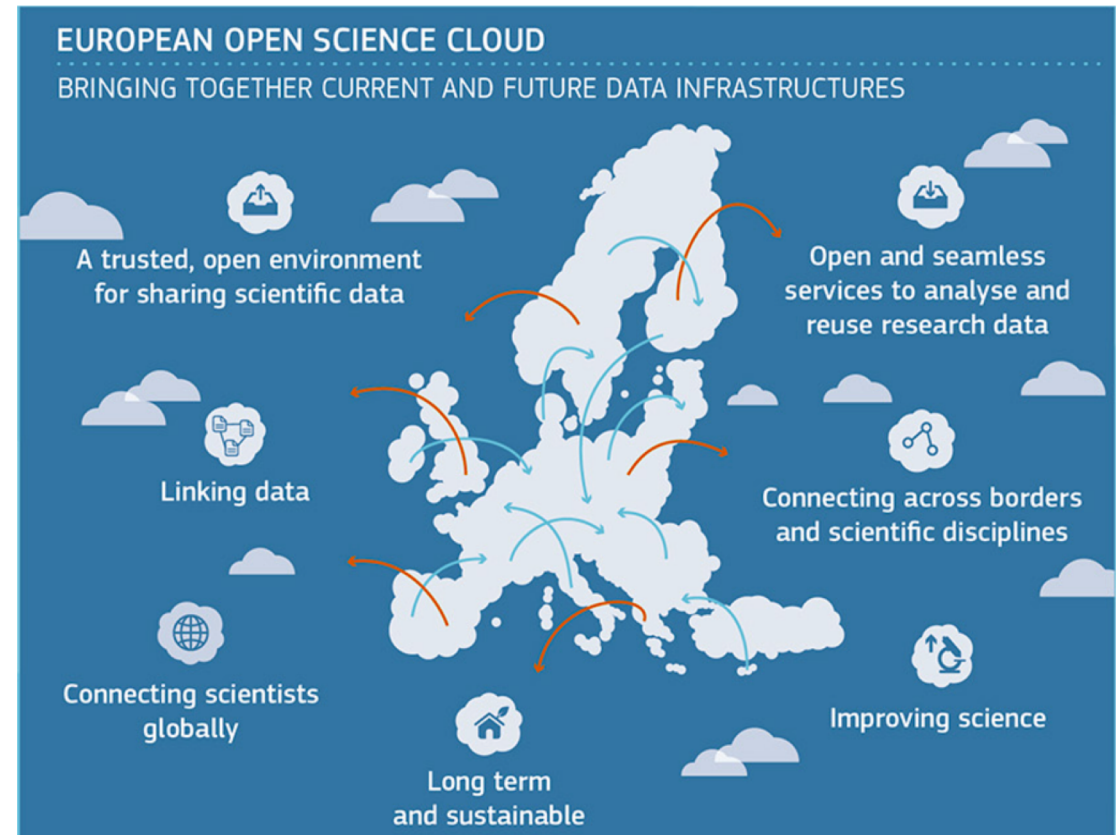
Cited by 4770 Related articles All 44 versions

Open
data
is about
MORE
THAN
DISCLOSURE
it must be
Fair

- Findable
- Accessible
- Interoperable
- Reusable

EC TAKES ACTION: EUROPEAN OPEN SCIENCE CLOUD

- Based on the FAIR principles
- Moved for a solution: EOSC
 - Data Stewardship (DS) for better discovery
 - Internet of Data of FAIR Data & Services
 - Training of 500.000 data experts
- Financing
 - €2B already invested in the initial phase EOSC
 - DS market \$85B annually
- Projects
 - FAIRsFAIR, EOSC Life, FAIRplus, FAIR4Health, ENVRI-FAIR, EERAdata, FAIR4Fusion, EJP Rare Diseases, Personal Genetics Locker, DiSSCo, C4yourself, ...



THE G20: SEPTEMBER 2016

“We support appropriate efforts to promote open science and facilitate appropriate access to publicly funded research results on findable, accessible, interoperable and reusable (FAIR) principles.” (Statement 12)

http://europa.eu/rapid/press-release_STATEMENT-16-2967_en.htm



G7 - 2017



19. We recognize that ICT developments, the digitisation and the vast availability of data, efforts to push the science frontiers, and the need to address complex economic and societal challenges, are transforming the way in which science is performed towards Open Science paradigms. We agree that an international approach can help the speed and coherence of this transition, and that it should target in particular two aspects. First, the incentives for the openness of the research ecosystem: the evaluation of research careers should better recognize and reward Open Science activities. Secondly, the infrastructures for an optimal use of research data: all researchers should be able to deposit, access and analyse scientific data across disciplines and at the global scale, and research data should adhere to the FAIR principles of being findable, accessible, interoperable, and reusable.

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;

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;

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;

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;

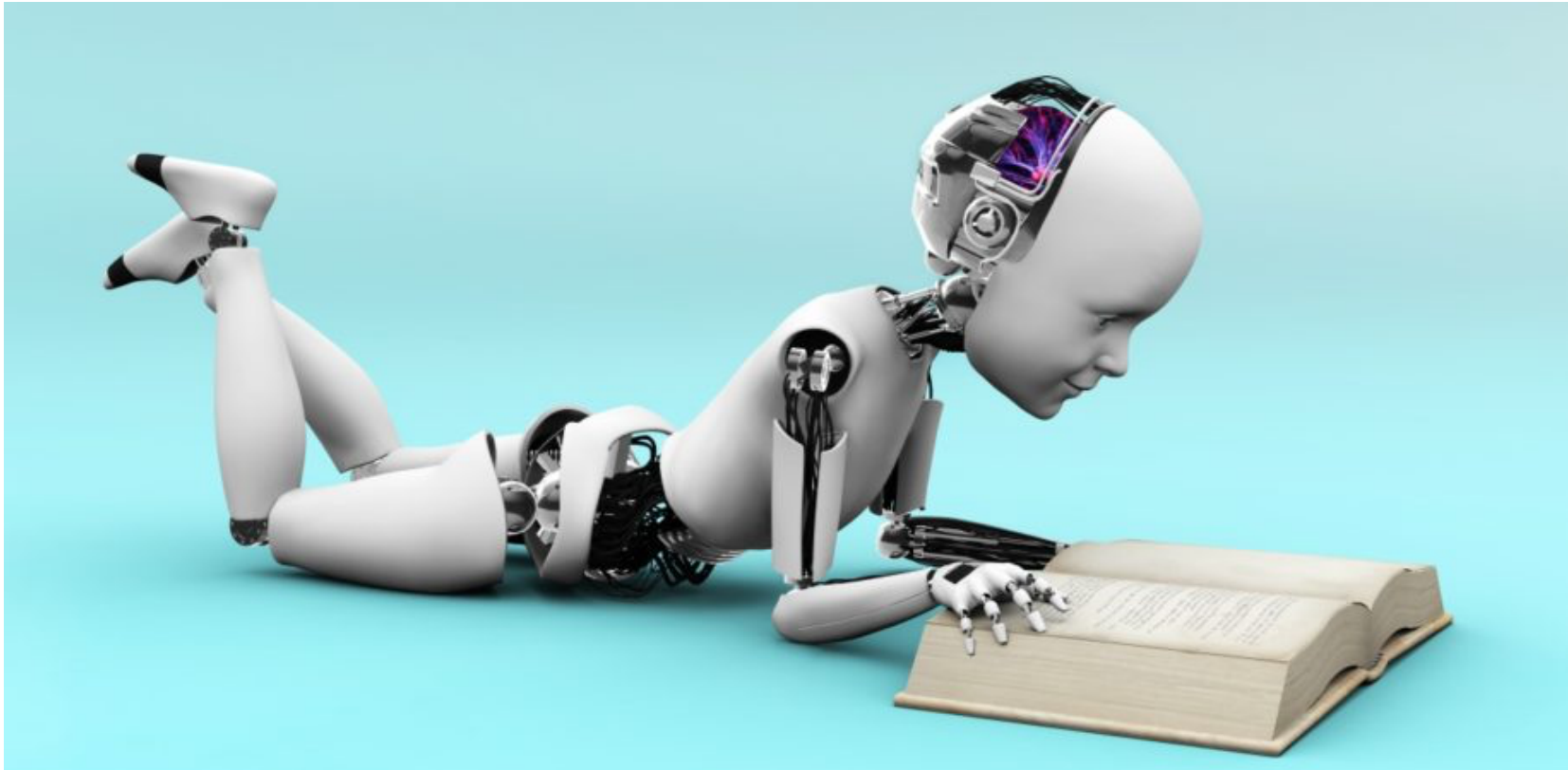
THE PRINCIPLES TARGET PRIMARILY MACHINES



THEN THE MACHINES CAN HELP US



**BUT FIRST THEY NEED TO
“UNDERSTAND WHAT WE MEAN”**



CURRENT STATUS ON FAIR

- FAIR infrastructure (e.g., FAIR Digital Object Framework, Data Stations, evaluators, certification (GO FAIR Foundation), FAIRification efforts, ...);
- Projects: FAIRsFAIR, EOSC Life, FAIRplus, FAIR4Health, ENVRI-FAIR, EERAdata, FAIR4Fusion, EJP Rare Diseases, Personal Genetics Locker, DiSSCo, C4yourself, ...
- Global reach: e.g., G7, G20, Open Science Cloud Executives Roundtable (OSCER);
- GO FAIR's FAIR Festival (June 21-23) > 600 attendees;
- Requirement in H2020, Horizon Europe, national funding (e.g., NWO(NL), NIH(US), HRB(IE), ...);
- Government adoption: e.g. in NL → VWS, MIW, Economic Affairs, Justice, Dutch Police;
- and much more ...

UNIVERSITY OF TWENTE.



Luiz Bonino

Associate Services and Cybersecurity – U. Twente
Associate Professor BioSemantics – LUMC
International Technology Coordinator – GO FAIR

E-mail: l.o.boninodasilvasantos@utwente.nl