Platform Linked Data Nederland

Innovatie door betekenisvol verbinden

FAIR DATA STEWARDSHIP & LINKED DATA

TOWARDS THE INTERNET OF FAIR DATA AND SERVICES

mobiquity

GO FAIR Foundation

PHORTOS
CONSULTANTS

WHY DO WE NEED FAIR DATA STEWARDSHIP IN THE FIRST PLACE?

The New York Times

Opinion

Artificial Intelligence Hits the Barrier of Meaning

Machine learning algorithms don't yet understand things the way humans do — with sometimes disastrous consequences.
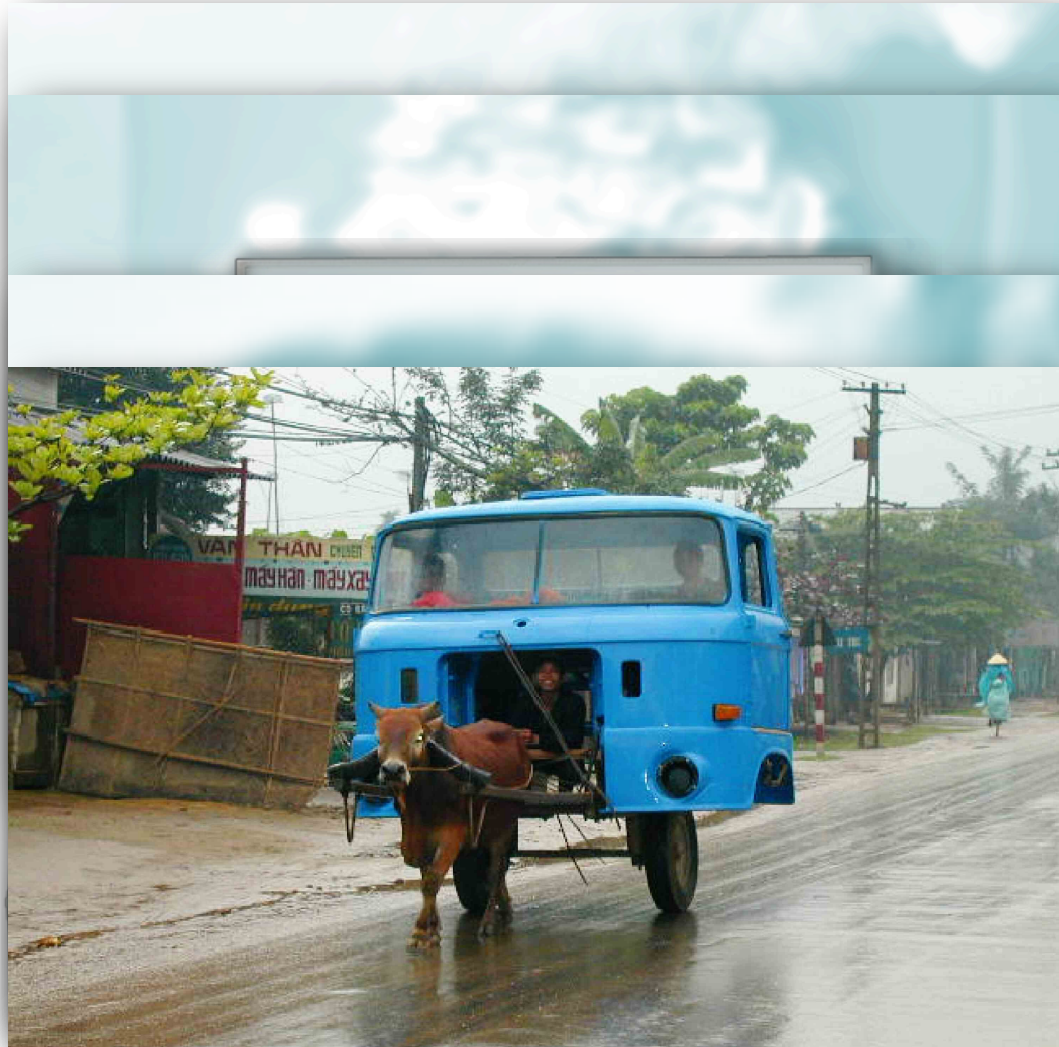
By Melanie Mitchell

Ms. Mitchell is Professor of Computer Science at Portland State University.

Nov. 5, 2018

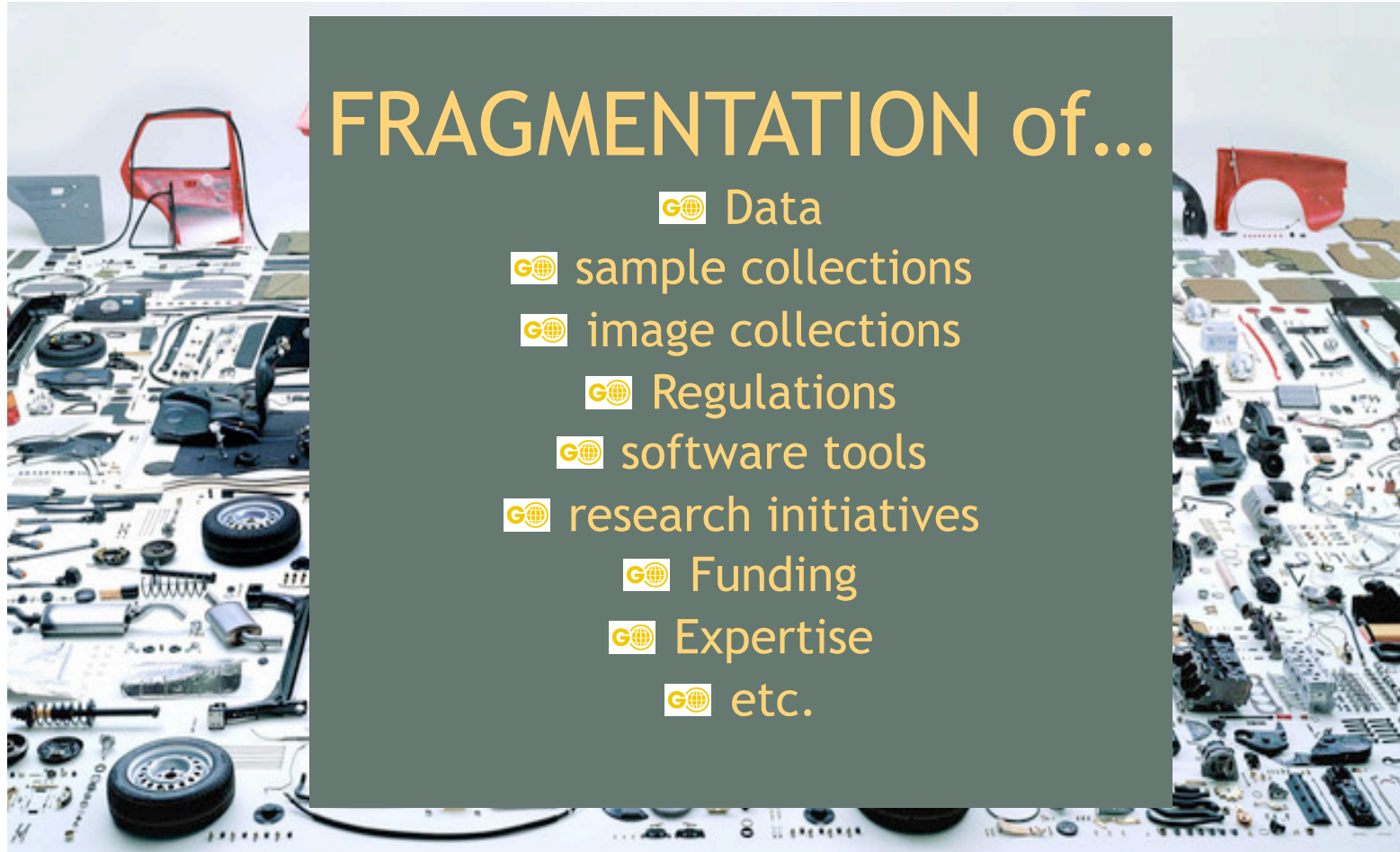**WITHOUT MACHINE ACTIONABLE CONTEXT COMPUTERS HAVE NO CLUE**

The Data Tsunami

Datarrhoeia

Standards

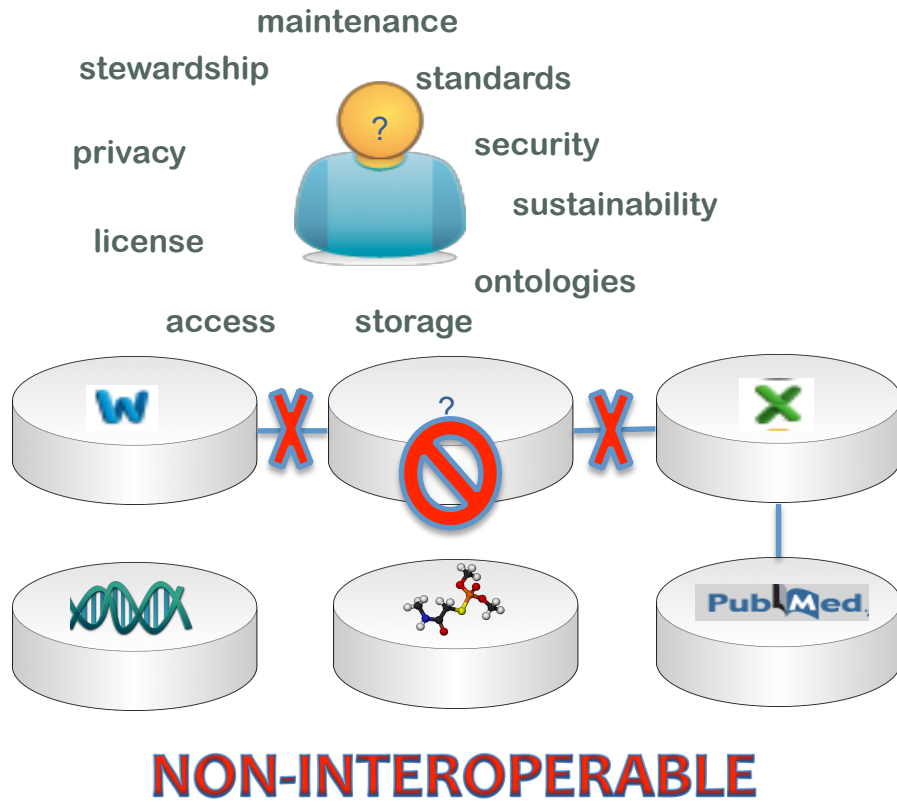Needle Transport

Do It Yourself Data

MOST DATA DON'T 'TALK' TO EACH OTHER

FRAGMENTATION of...

- Data
- sample collections
- image collections
- Regulations
- software tools
- research initiatives
- Funding
- Expertise
- etc.

WE NEED ACTIONABLE DATA!!
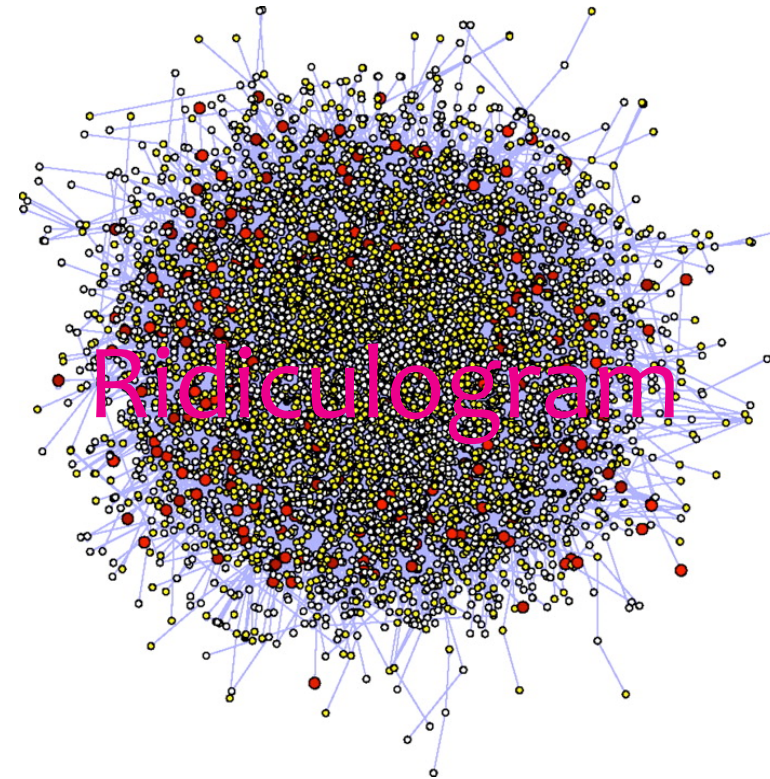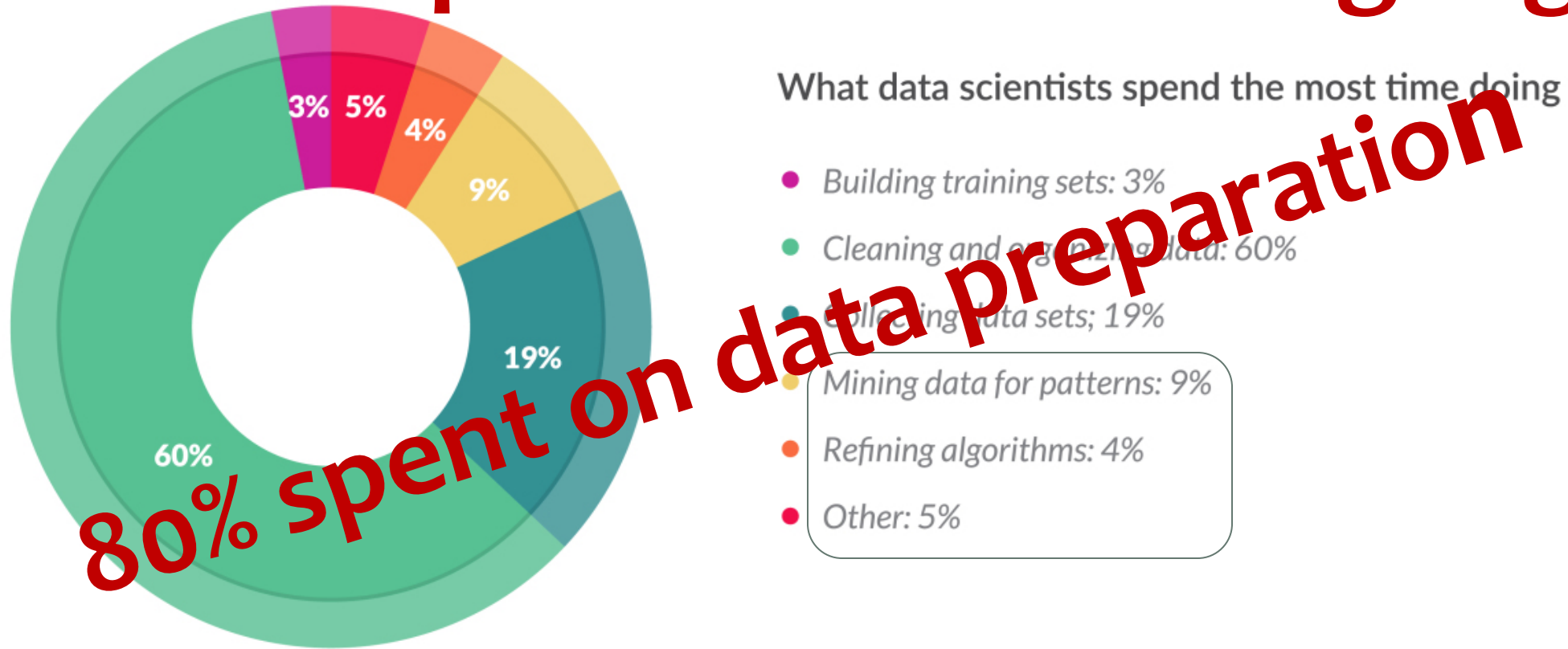
If data are *not* interoperable

maintenance

stewardship   standards

privacy   security

license   sustainability

ontologies

access   storage

NON-INTEROPERABLE

Cross data analytics are *not* instructive

Ridiculogram

DATA EXPERT EFFORT

From 80% spent on data wrangling

80% spent on data preparation

To 80% spent on analytics/research

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%

3% 5% 4% 9% 19% 60%

Research and Development Expenses 2017

50% of €4,5B = €2,25B
Which is 10% of revenue!!
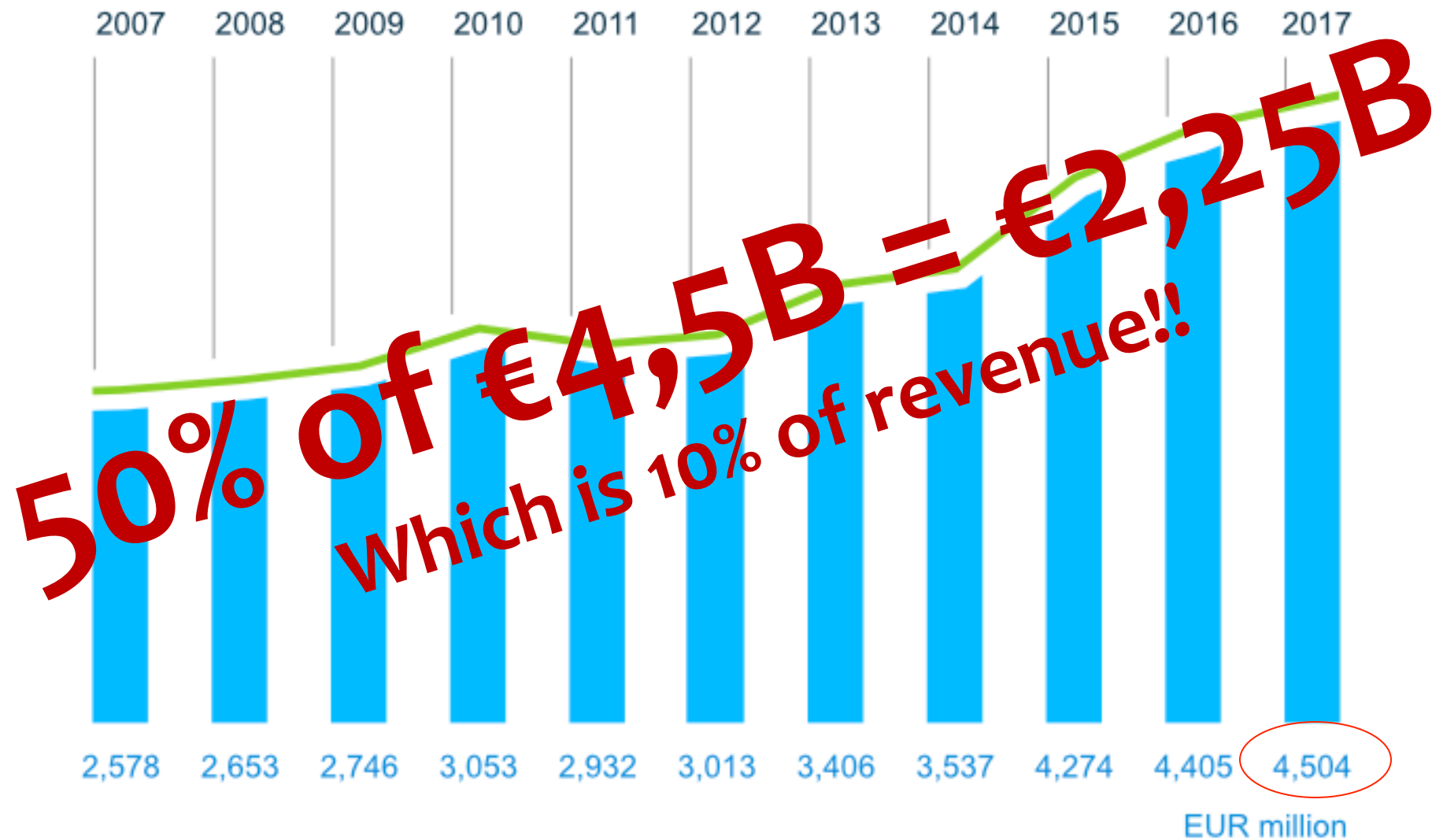
WHAT IS GO FAIR?

- Europe acknowledged the (research) data problem

- Moved for a solution: EOSC

  - Data Stewardship (DS) for better discovery
  - Internet of Data of FAIR Data & Services (IFDS)
  - Training of 500.000 FAIR linked data experts

- Financing

  - €2B for initial phase EOSC
  - 5% of grants for DS
  - DS market €80 billion annually
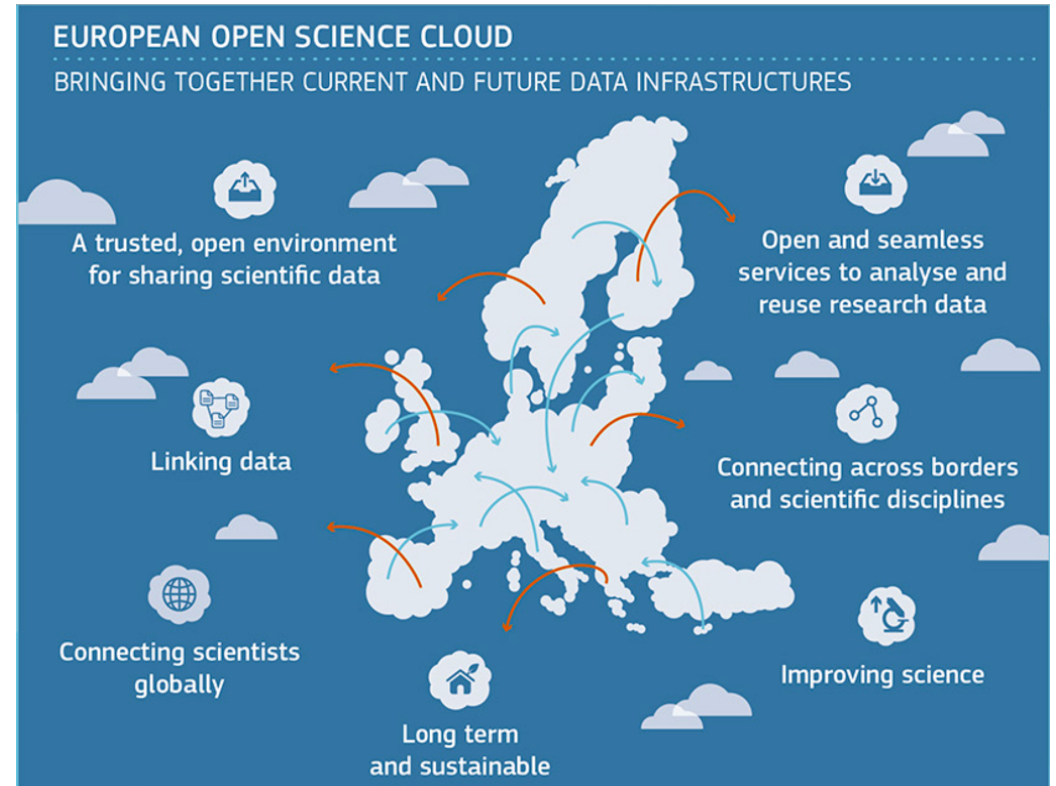
A GRAND VISION BUT NO GUIDELINES FOR IMPLEMENTATION

EUROPEAN OPEN SCIENCE CLOUD
BRINGING TOGETHER CURRENT AND FUTURE DATA INFRASTRUCTURES

A trusted, open environment for sharing scientific data

Open and seamless services to analyse and reuse research data

Linking data

Connecting across borders and scientific disciplines

Connecting scientists globally

Long term and sustainable

Improving science

EUROPEAN COMMISSION
Directorate-General for Research & Innovation

SCIENCE POLICY • TEAM ZBW MEDIATALK • 24.01.2017 • 6 MIN READ

GO-FAIR – A Member States-Up strategy for the EOSC implementation

"GO-FAIR" is a proposal for the practical implementation of the European Open Science Cloud (EOSC) through a federated approach making optimal use of existing initiatives and infrastructures in the participating Member States.

G7 2017 ITALIA
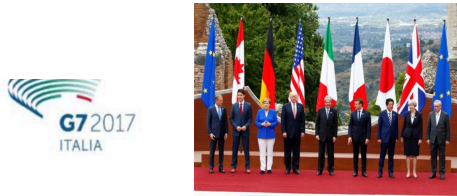
19. We recognize that ICT developments, the digitisation and the vast availability of data, efforts to push the science frontiers, and the need to address complex economic and societal challenges, are transforming the way in which science is performed towards Open Science paradigms. We agree that an 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.

"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

二十国集团领导人杭州峰会
G20 HANGZHOU SUMMIT
中国·杭州 2016年9月4-5日

Let's GO FAIR

NIH National Institutes of Health
Turning Discovery Into Health

The Commons supports biomedical discovery by enabling sharing of digital objects

A set of **Digital Object Compliance** principles that describes the properties of digital objects that enables them to be findable, accessible, interoperable and reproducible (FAIR).

FAIR FAIRSUS GO FAIR

FAIR

- Many interpretations and implementation options

- A lot of self imposed rules and criteria

- Used as a buzz word

- Claimed for almost everything Big Data related

- Source of academic polemics

- Nothing to join

GO FAIR

- A fast-track implementation for the Internet of FAIR Data and Services (IFDS) supporting the EOSC.

- Broad governmental support
    - (EC, NL, DE, FR and many EC Member States and adopted globally)

- Specific implementation choices
    - (Resolvable GUPID, machine-readable metadata, Core metadata elements, RDF, triple based knowledge graphs, Linked data solutions, etc.)

- Providing guidance on metrics and certification towards ACTIONABLE data

- Welcome to join

THE INTERNET OF FAIR DATA AND SERVICES

The Internet solved the problem of the **interoperability of heterogeneous networks**

The **hourglass design** of the Internet system enabled both *interoperability* and *unparalleled flexibility* for extension

Dr. George Strawn, co-founder of the Internet

- A Datanet could "solve" the problem of *interoperability of heterogeneous data*

- An **hourglass design** of a Datanet system would enable both *interoperability* and *unparalleled flexibility for extension*
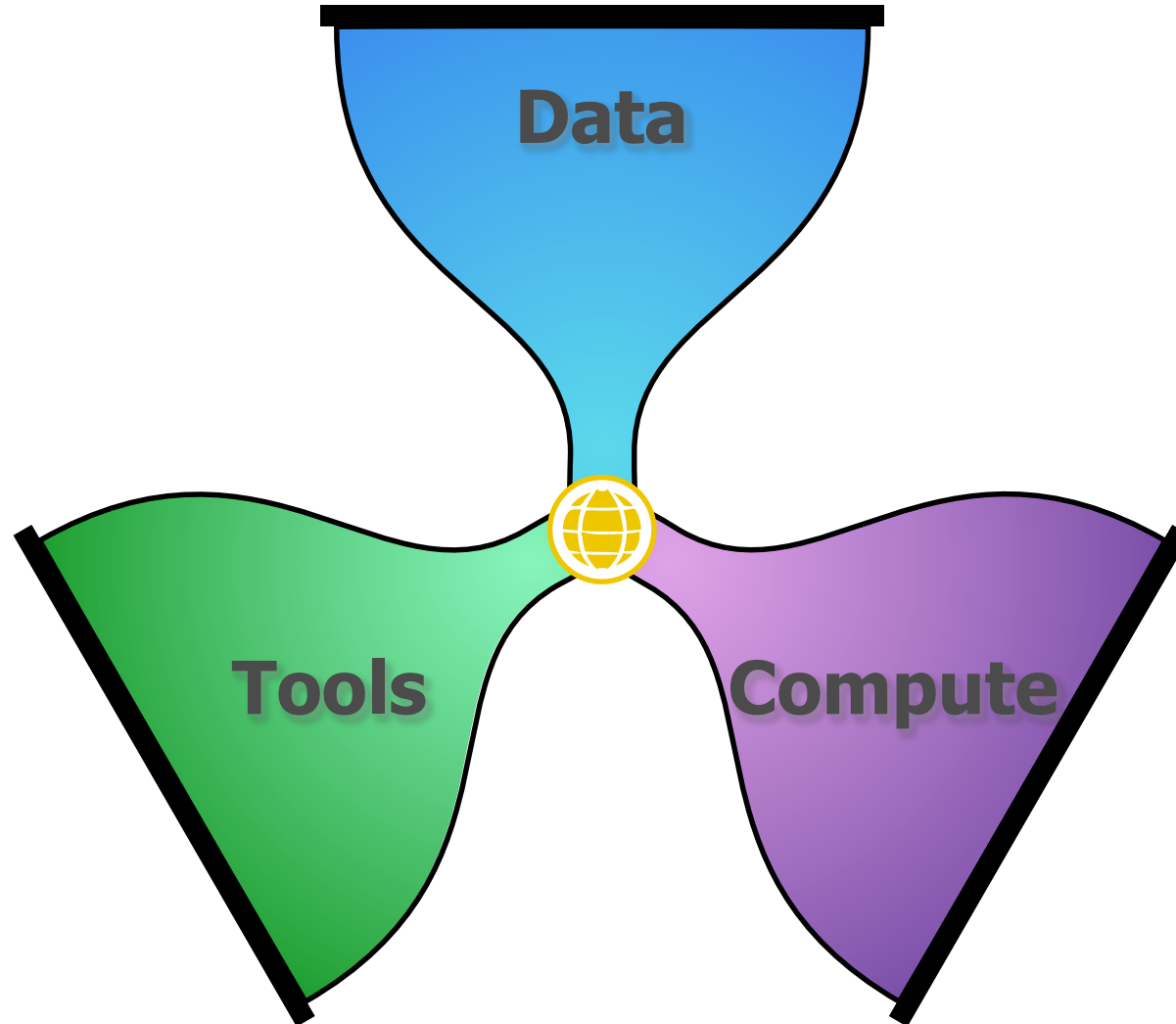
IFDS MAIN ELEMENTS

DATA STATION

- Provides FAIR access to data and metadata
- Allows train to access and interact with data

TRAIN

- Interacts with data (process, integrate, analyze, …)

DATA GATEWAY

- Provides access and control to the data authority regardless of where the data is located/stored
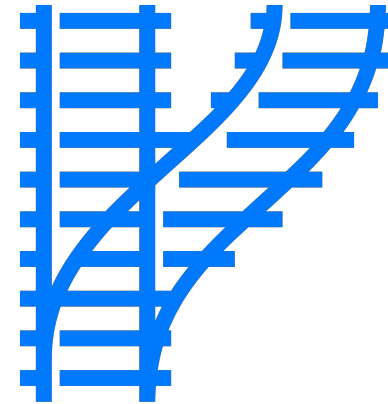
TRACKS

- The routing and transport infrastructure
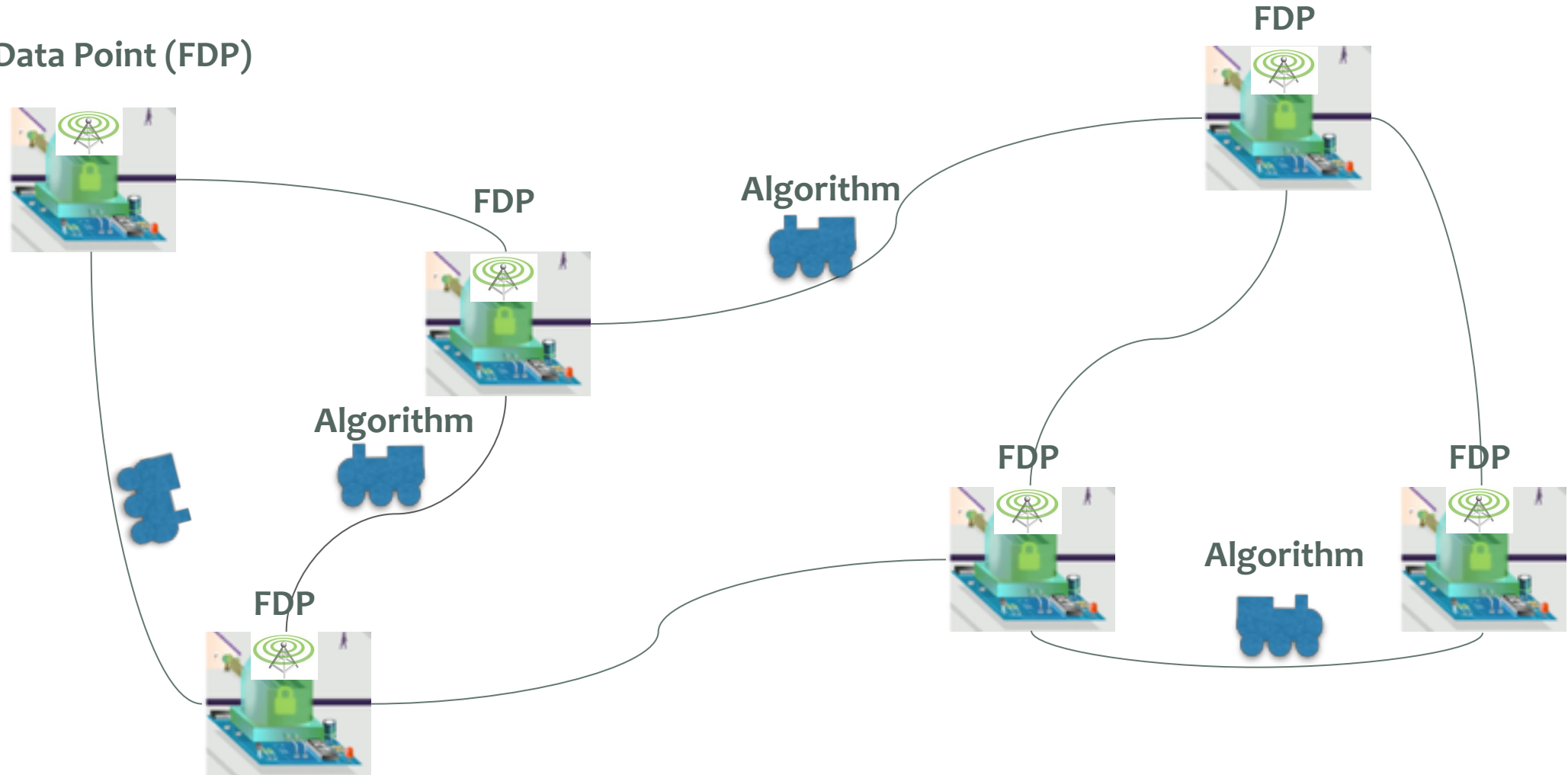
- Many organizations (20+) have participated in FAIR BYODs and trainings

- Several academic institutions and funders have started or are considering GO FAIR Readiness programs

- Several companies have started or are considering the GO FAIR Readiness program

WHAT IS FAIR DATA

AND HOW DO THEY RELATE TO LINKED DATA?

Hollandisch
Dutch
Nederlands

Malaysian

1	
2	
3	Sjalo
4	Knof
5	Gem
6	Pepe
7	Keto
8	Djint
9	Kerri
10	Suike
11	Ketja
12	Trass
13	Kruid
14	Noot
15	Kurk
16	Asem
17	Lauri
18	Citro
19	Citro
20	Bouil
21	Spits
22	Andij
23	Komk
24	Boon
25	Taug
26	Prei
27	Aard
28	Bana
29	Rund

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**;

Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016)

http://fairmetrics.org

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

Acces

A1. (meta) by their identifier using a standardized tocol;

> **A1.1** the protoc niversally implementable;

> **A1.2.** the protocol a ation and authorization procedure,

A2. metadata are accessible, even ta are no longer available;

STAY IN CONTROL of YOUR DATA

THE 7 CANONICAL STEPS OF FAIRIFICATION

Implementation Strategy

0
The Data Stewardship Plan

Preparation
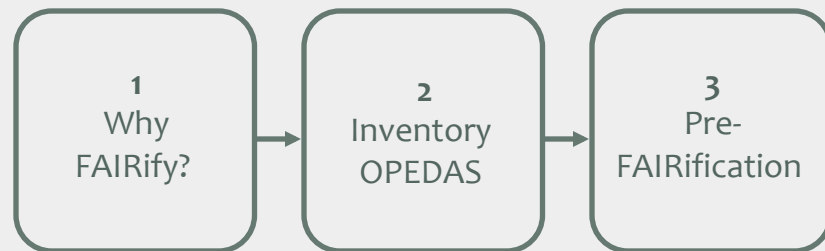
FAIRification

Analysis

1
Why FAIRify?

2
Inventory OPEDAS

3
Pre-FAIRification

4a
Find/Build semantic data model

5a
Transform data records into RDF

4b
Find/Build semantic metadata model

5b
Transform metadata records into RDF

6
Push to FDP & Certified Metrics Score

7
Automatic interoperation & resolution of research question
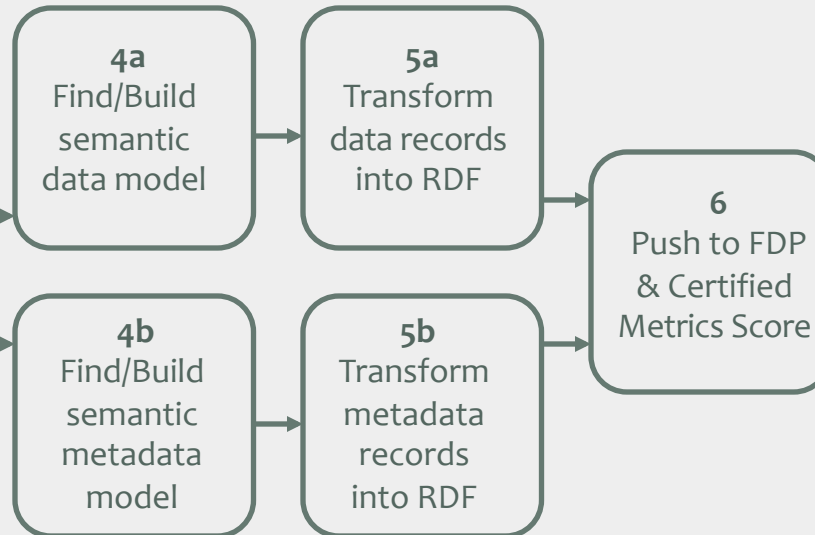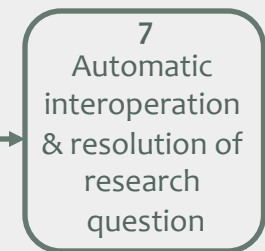
Plan → Create → Publish → Find → Measure

PROBLEM: ALL TOOLS ARE CURRENTLY PROFESSORWARE

FACILITATING STAKEHOLDER DRIVEN IMPLEMENTATION

THE 'PRODUCT' DEVELOPMENT PROCESS

iWIZ	KIK-V	PUZ	RoR
Modelathon	Toolathon	Datathon	
Existing/ New model	Existing/ New tools	Legacy data De Novo data	
Review	Review	Review	
(Re) model	(Re) build	(meta) data model	
Simulate / apply	Use case / apply	FAIRify	
Use for analysis	Test with users	Use for analysis	
Freeze & Deposit	Freeze & Deposit	Publish as FDP	

FACILITATING REPOSITORY

MODELS

- Ontology models
- Semantic
  - metadata models
  - data models

TOOLS

DSW *Data Stewardship Wizard*

FAIRIFIER

FAIR DATA POINT

FAIR DATA
011
001
110
010
1

FAIR Metrics Evaluator
Quantitative | Reproducible | Objective

DATA

FDP FDP
FDP FDP
FDP FDP
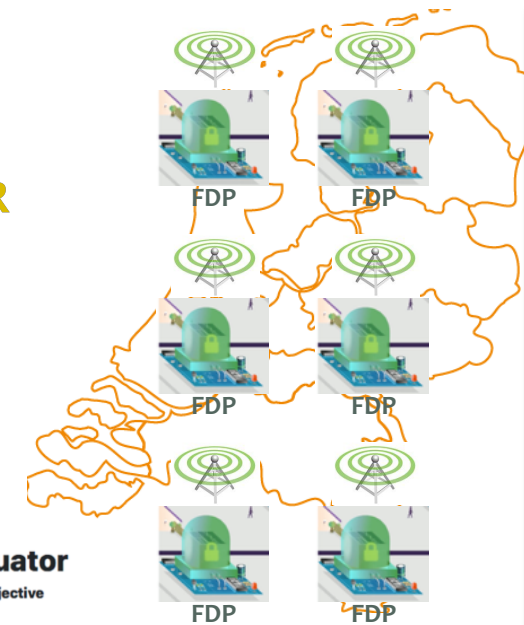
RE-USE ZIN STAKEHOLDERS

AVAILABLE AS ONLINE PROFESSIONAL SERVICES

"Linked Data is a method of publishing structured data using standard Web technologies such as HTTP, RDF and URIs"

Tim Berners-Lee – 2006

Linked Data can facilitate the realization of many of the FAIR principles by providing:

- An inherent machine-actionability approach

- Identifiers for every element

- The support for combining data elements and their semantics

LINKED DATA DIRECT SUPPORT FOR THE FAIR PRINCIPLES

FAIR Principle	Linked Data support
F1. (meta)data are assigned a globally unique and persistent identifier	Unique Resource Identifiers (URIs) are required in Linked Data
F3. metadata clearly and explicitly include the identifier of the data it describes;	Linked Data facilitate making explicit the connection between metadata and the related data. E.g., a triple such as <myMetadataURI> <isMetadataOf> <myDataURI>

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
	Linked Data is based on Web protocols that are open, free, universally implementable and allow for authentication and authorization, if required.

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	RDF is a formal, accessible, shared and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles;	Linked Data facilitate the linking between data items and their semantic annotations using concepts from different vocabularies.
I3. (meta)data include qualified references to other (meta)data;	Linked Data facilitates the interlinking of different data and metadata.

LINKED DATA SUPPORT FOR 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;

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

De titel van deze conferentie is:

Linked Data is FAIR for everyone

Ik stel voor er van te maken:

Linked Data is FAIR made easy for everyone