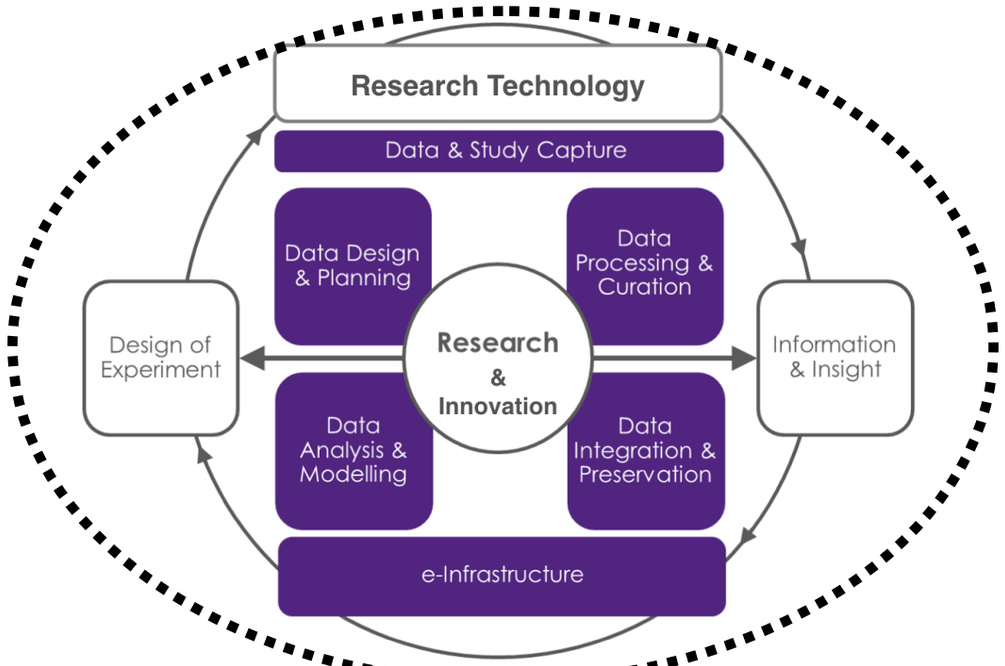
Open Science
The Internet for Social Machines
The end of data sharing as we know it

FAIR > The Machine Knows what I Mean

Barend Mons
Gdansk, October 2019

The Future belongs to Data Stewards ??



Technical infrastructure
Social agreements/contracts (domain-relevant content / third-party services)

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

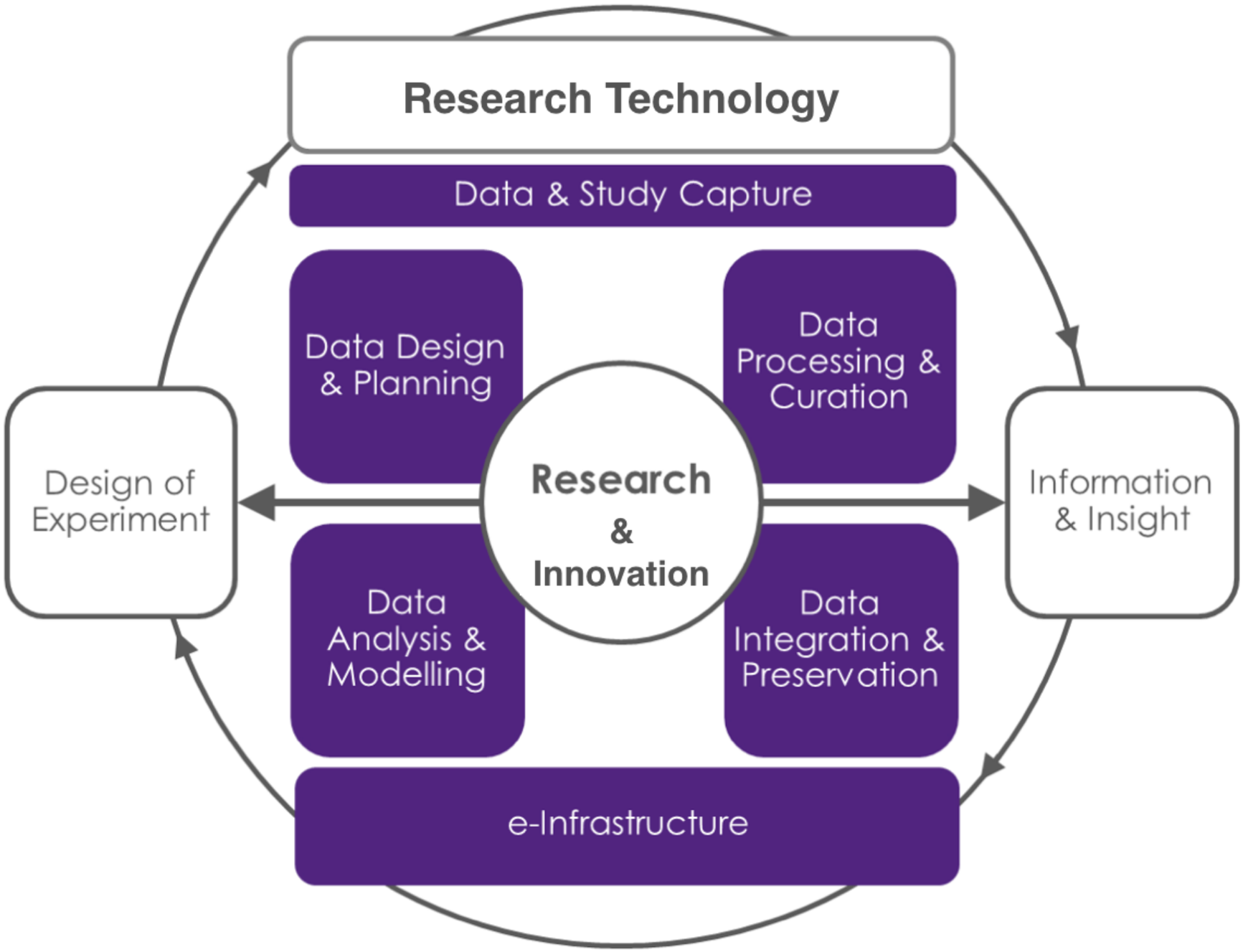
To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be 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

To be 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

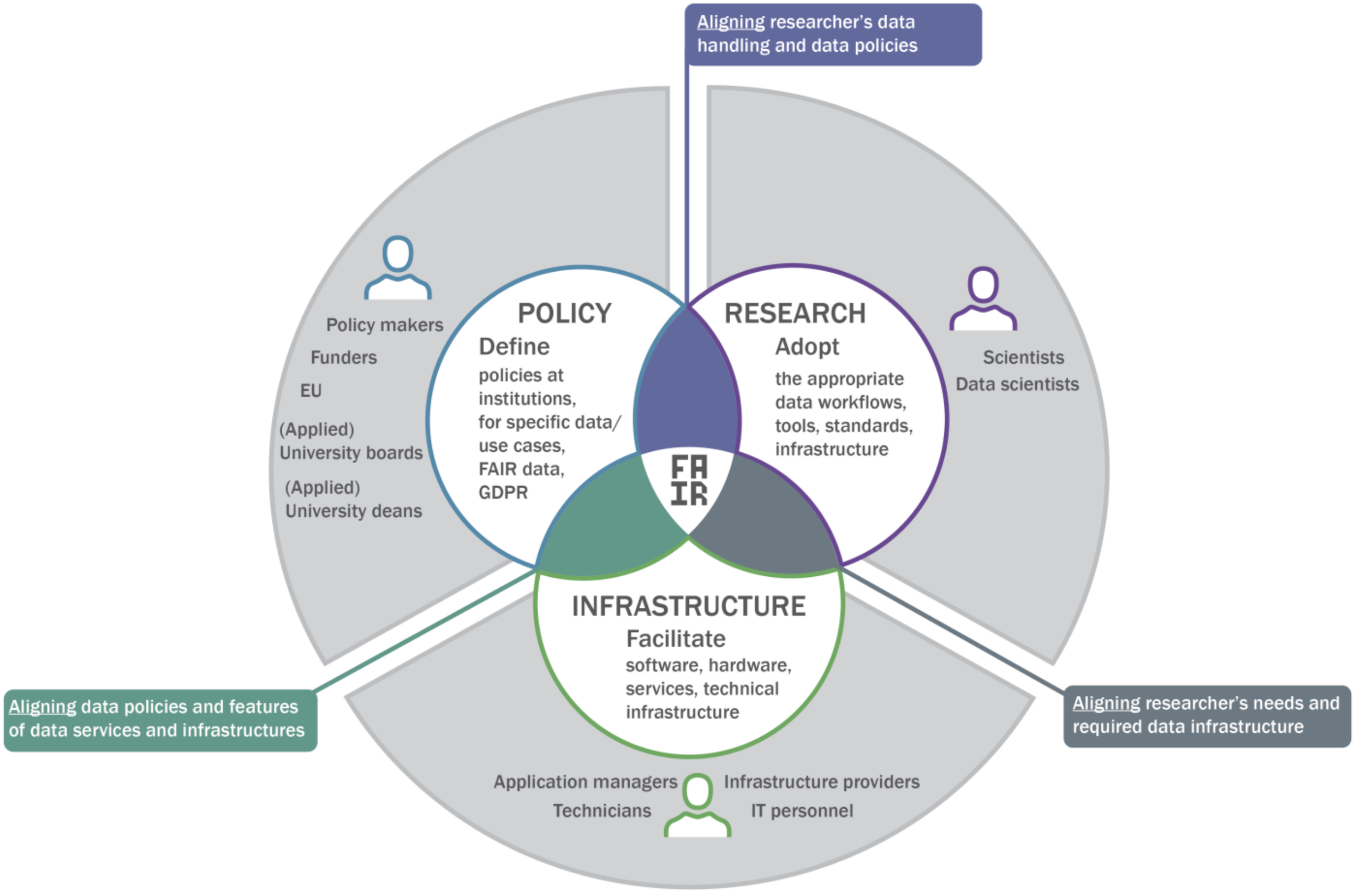
DOI: 10.1038/sdata.2016.18

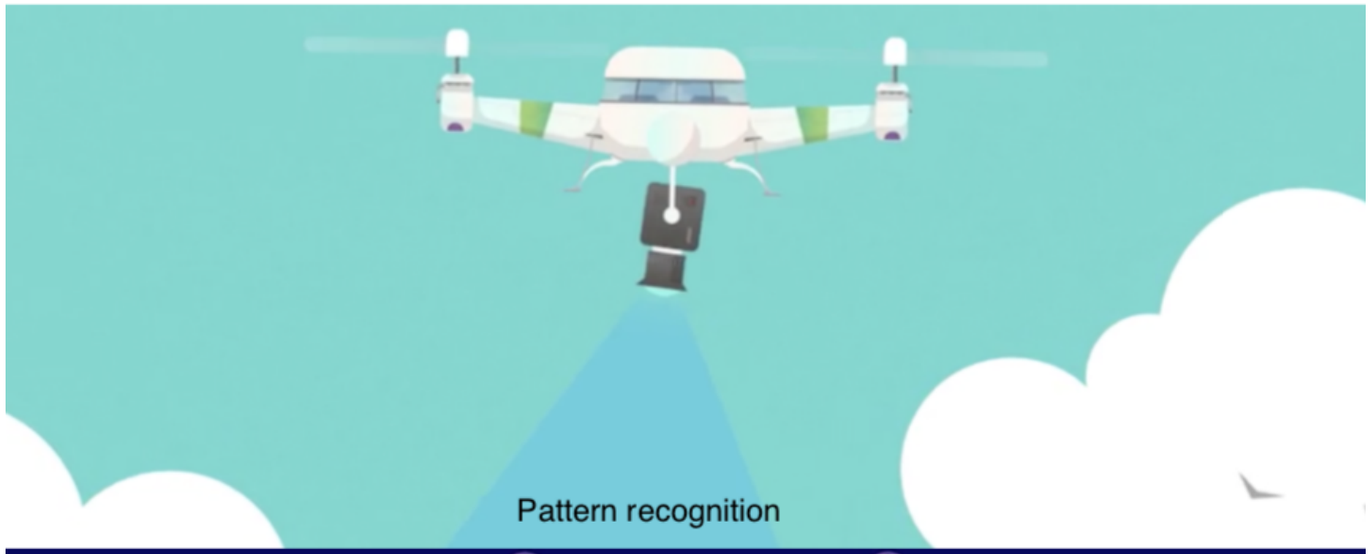
5%

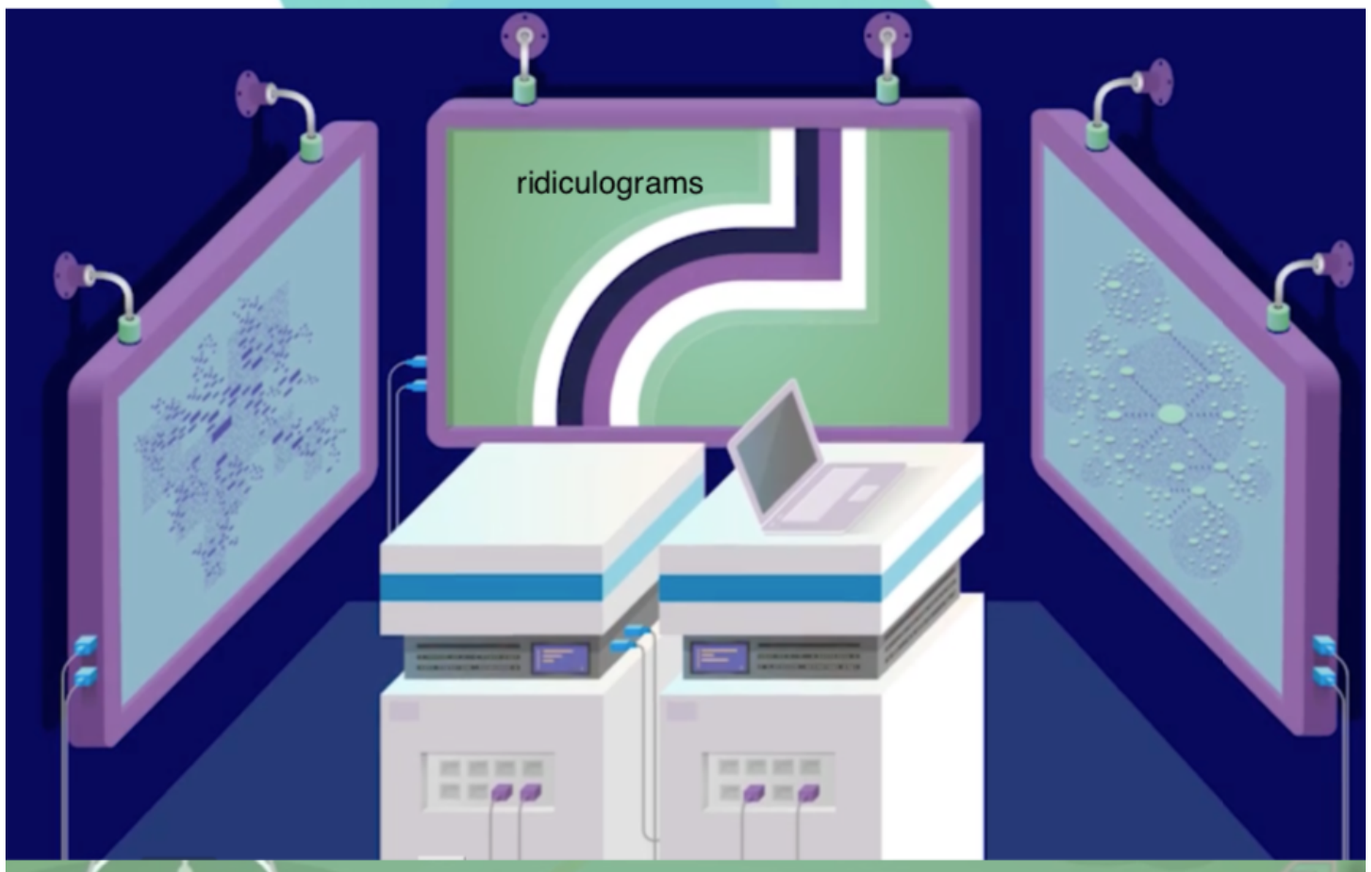

Research Technology
Data & Study Capture
Data Design & Planning
Data Processing & Curation
Design of Experiment
Research & Innovation
Information & Insight
Data Analysis & Modelling
Data Integration & Preservation
e-Infrastructure

10 B

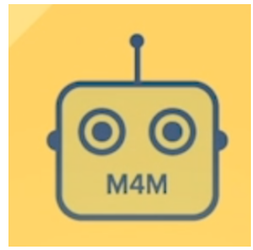
Implementation areas for data stewardship




Pattern recognition


ridiculograms


actionable knowledge


M4M

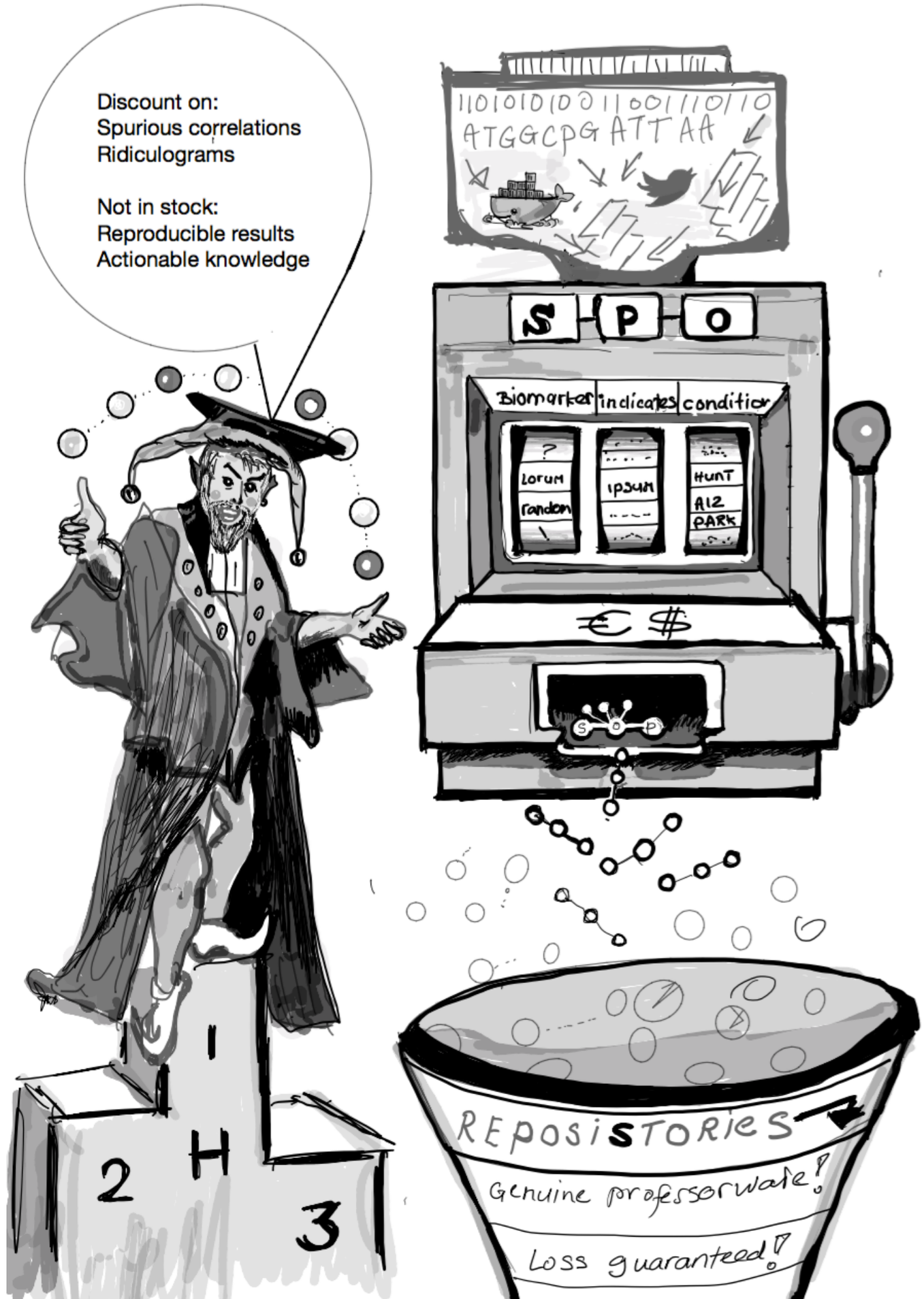
The seven capital sins of Open Science

1: **Age factor….Reward only** narrative.com

2: Ignore complexity and existing data

3: Disrespect other disciplines

4: publish data without a supplementary paper



5: create a nightmare for machines

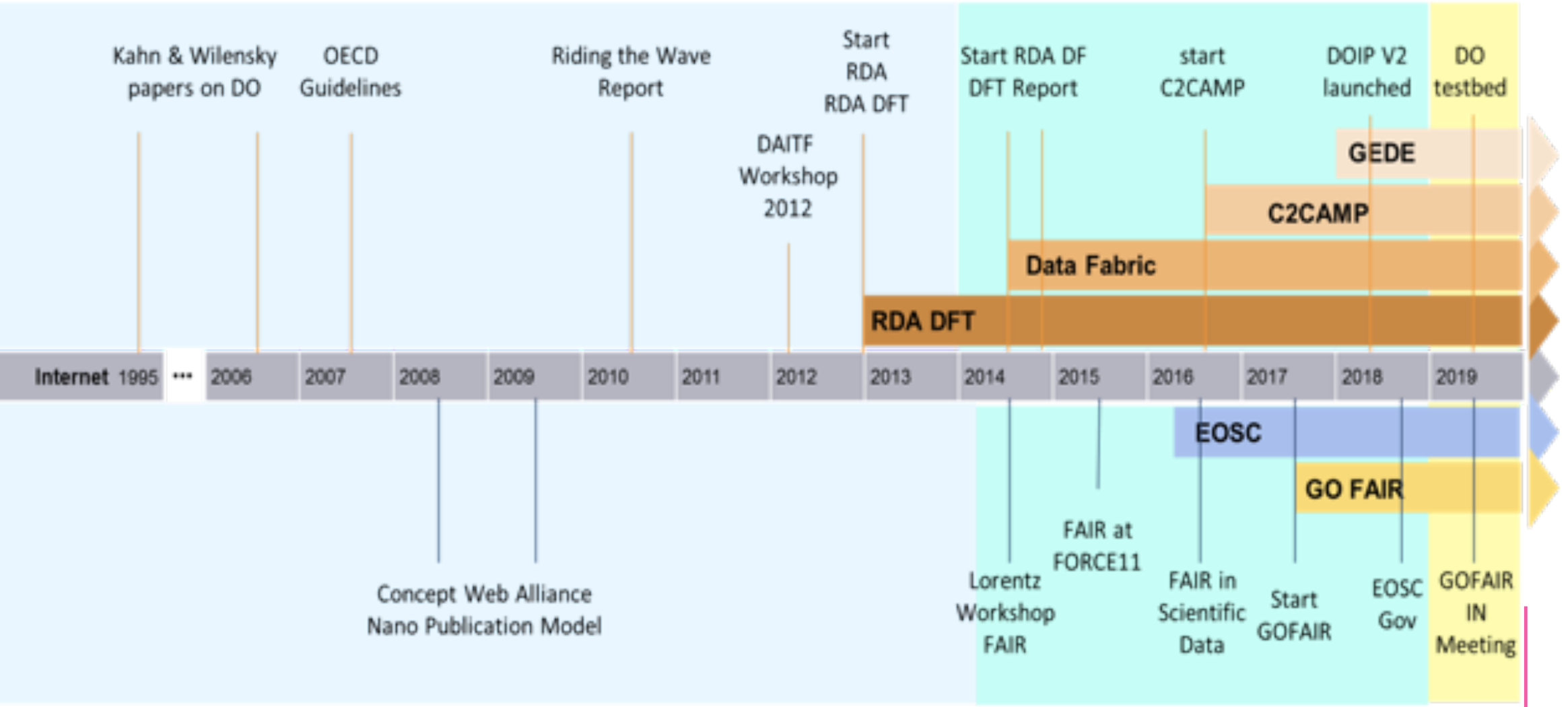
6: refuse to invest in research -infrastructure

7 : Create Data without a Data Stewardship plan

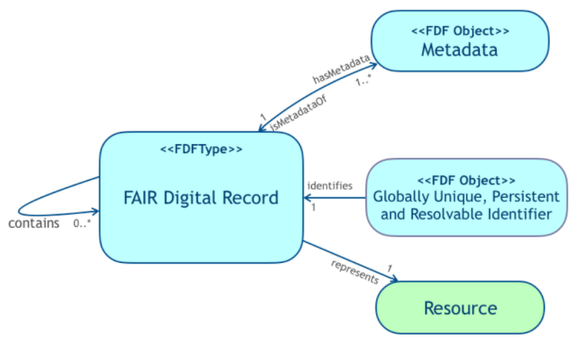
Machines ‘know’ what it means

FAIR

Advances in Digital Objects



Advances in FAIR Principles



FAIR and GO FAIR

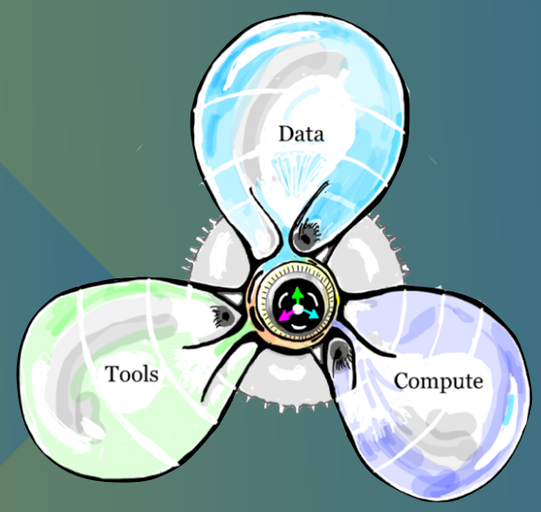
Lorentz

EOSC

FAIRdICT

GO FAIR

IFDS



EOSC

Birth	Infancy		Adolescence		Maturity
2014	2015	2016	2017	2018…	

The Road to FAIRness

The Road to FAIRness

FAIR

Machines know what it means

2

what operations are possible?

1

what (type) is it?

3

what operations are allowed?

Digital/physical Resource/Object /Construct

Semantic descriptions
Types/operations

<<FDF Object>>
Metadata

hasMetadata
1..*
1
isMetadataOf

<<FDFType>>
FAIR Digital Record

contains 0..*

identifies
1

<<FDF Object>>
Globally Unique, Persistent and Resolvable Identifier

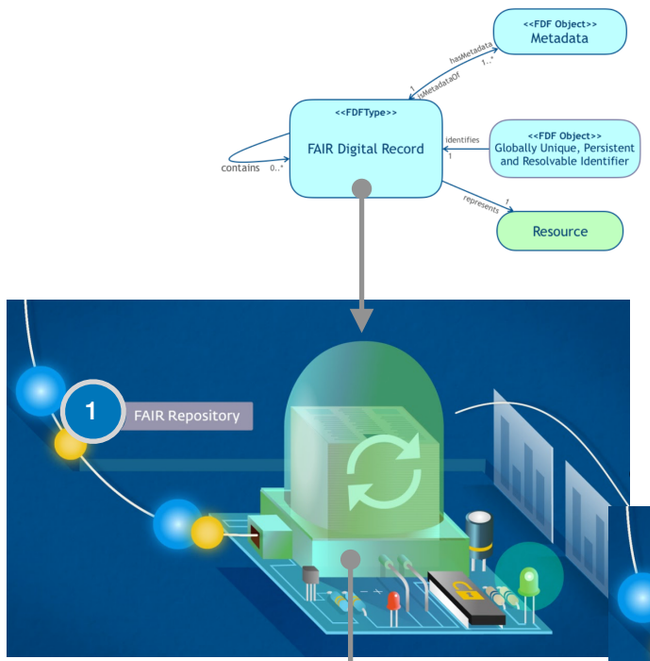
represents 1

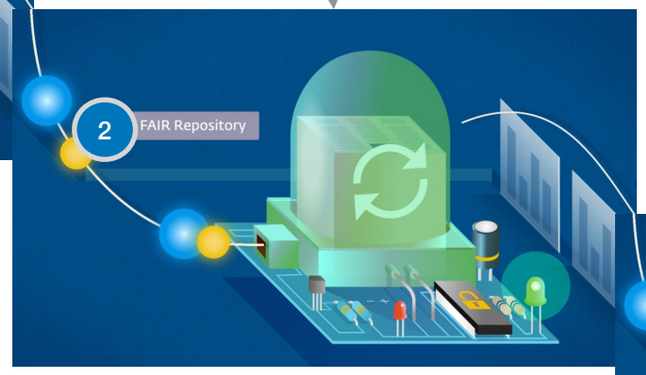
Resource

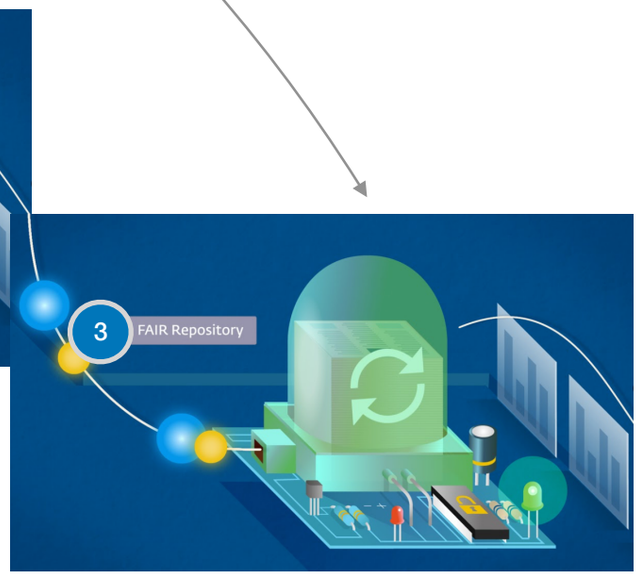
4

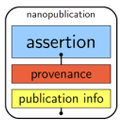
operation

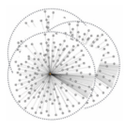
Many different DO's of the type 'metadata' can be stored in many Repositories and point to the same DO

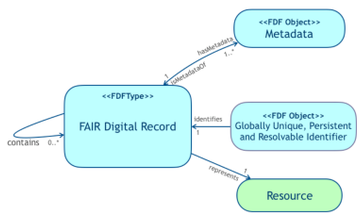




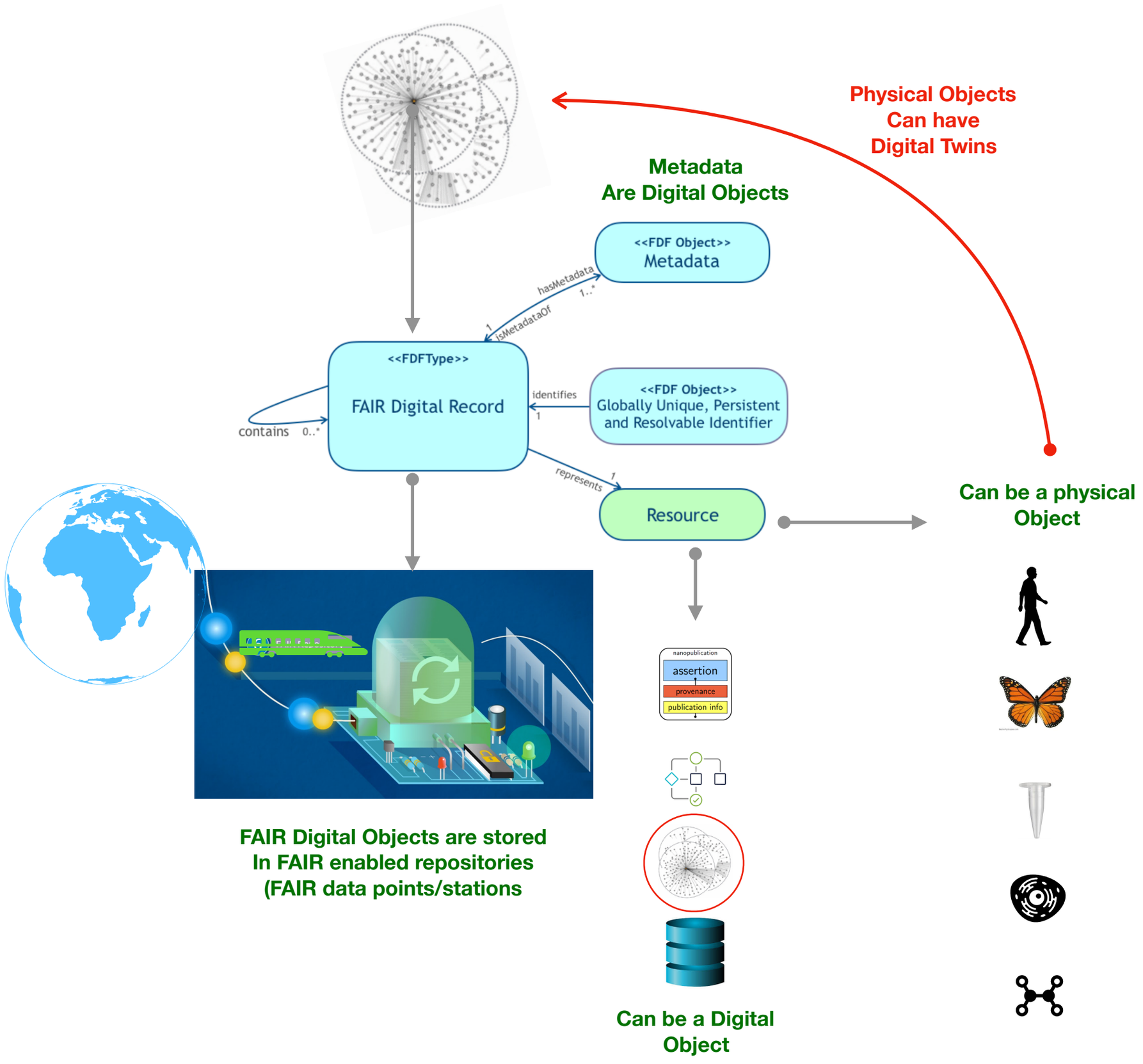








Digital Twins are a type of Digital Objects



Intrinsic

Rights

Reuse

Citations

C

B

<<FDF Object>>
Metadata

D

hasMetadata
1..*
1
isMetadataOf

<<FDFType>>

<<FDF Object>>
Metadata

identifies
1

<<FDF Object>>
Globally Unique, Persistent and Resolvable Identifier

DOI

hasMetadata
1..*
1
isMetadataOf

represents
1

<<FDFType>>

FAIR Digital Record

contains
0..*

identifies
1

<<FDF Object>>
Globally Unique, Persistent and Resolvable Identifier

Resource

represents
1

Resource

Supplementary (traditional) Article about RO (with nano pub 'abstract')

A

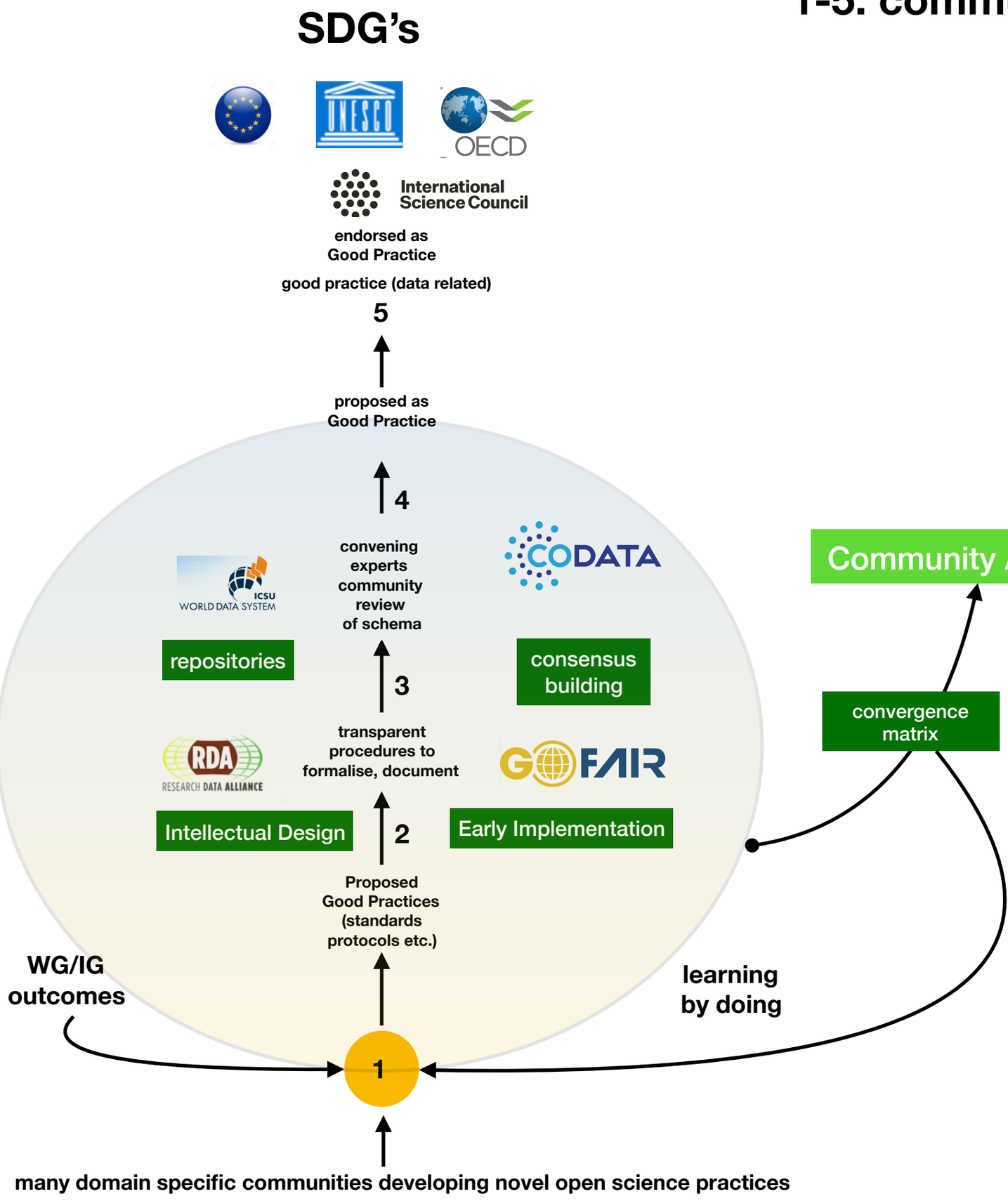
E

Primary publication of (FAIR Digital Twins of) Research Objects (RO)

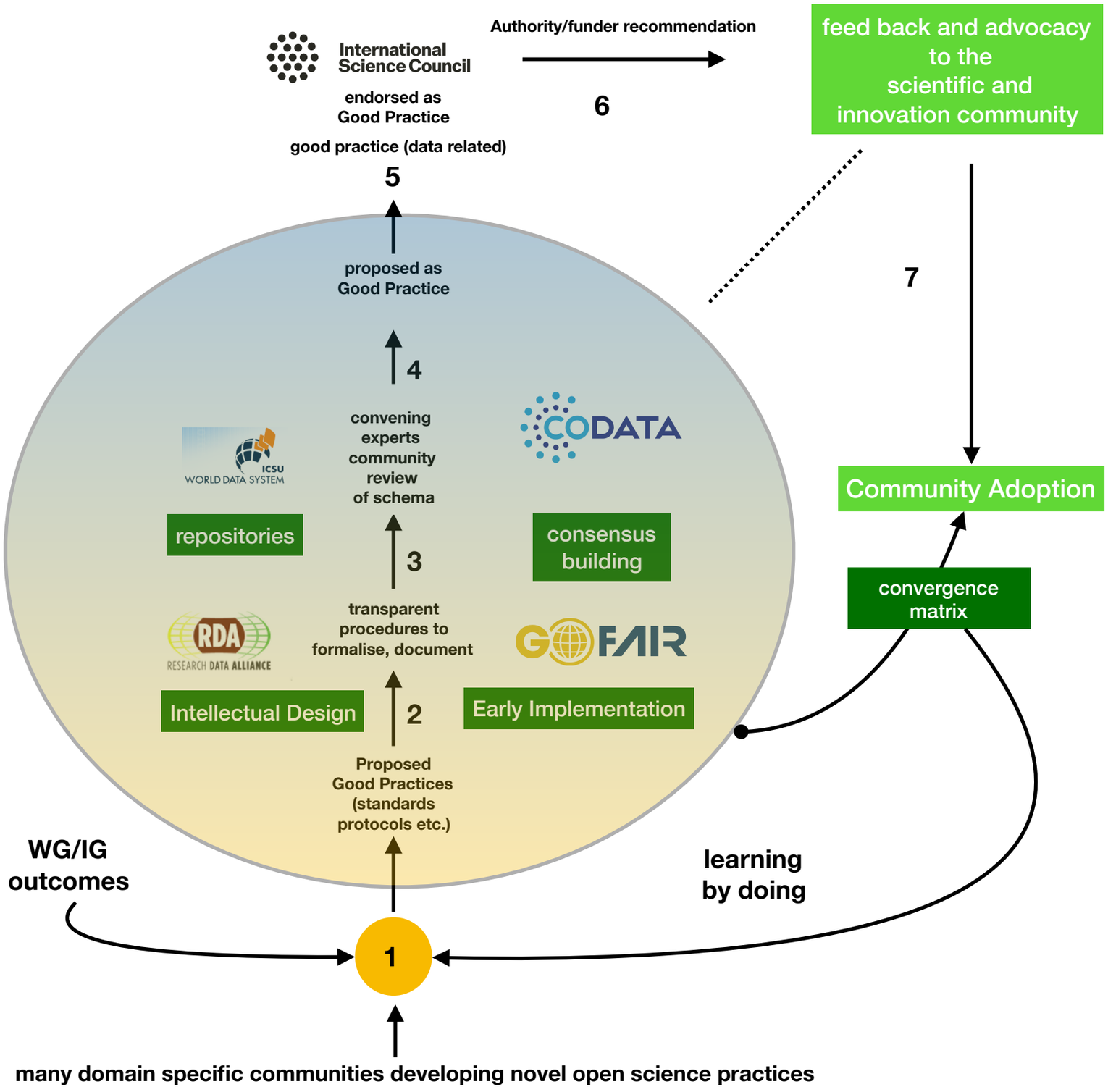
Independently Citable/awardable

All metadata also linked to article

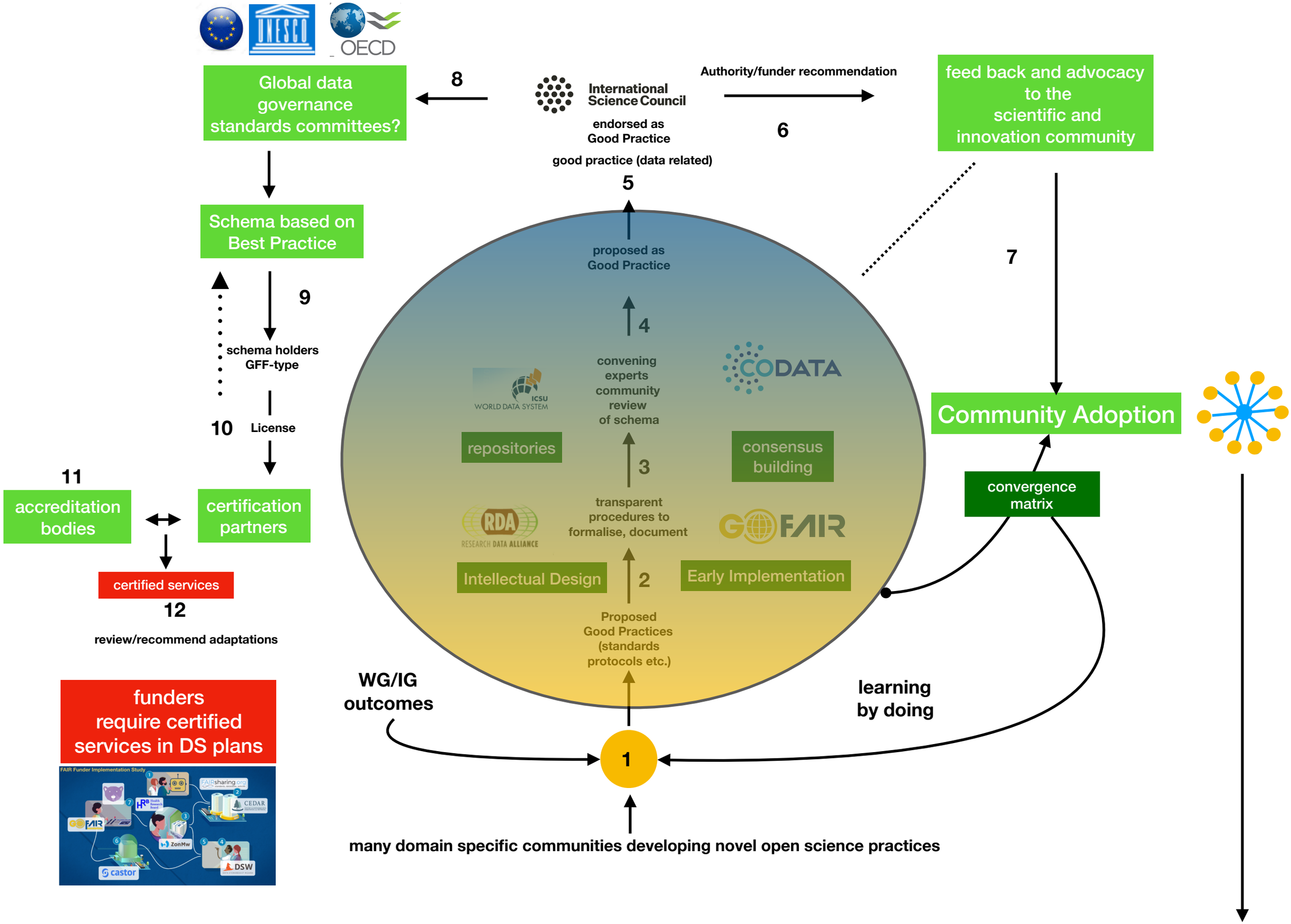
1-5: community driven

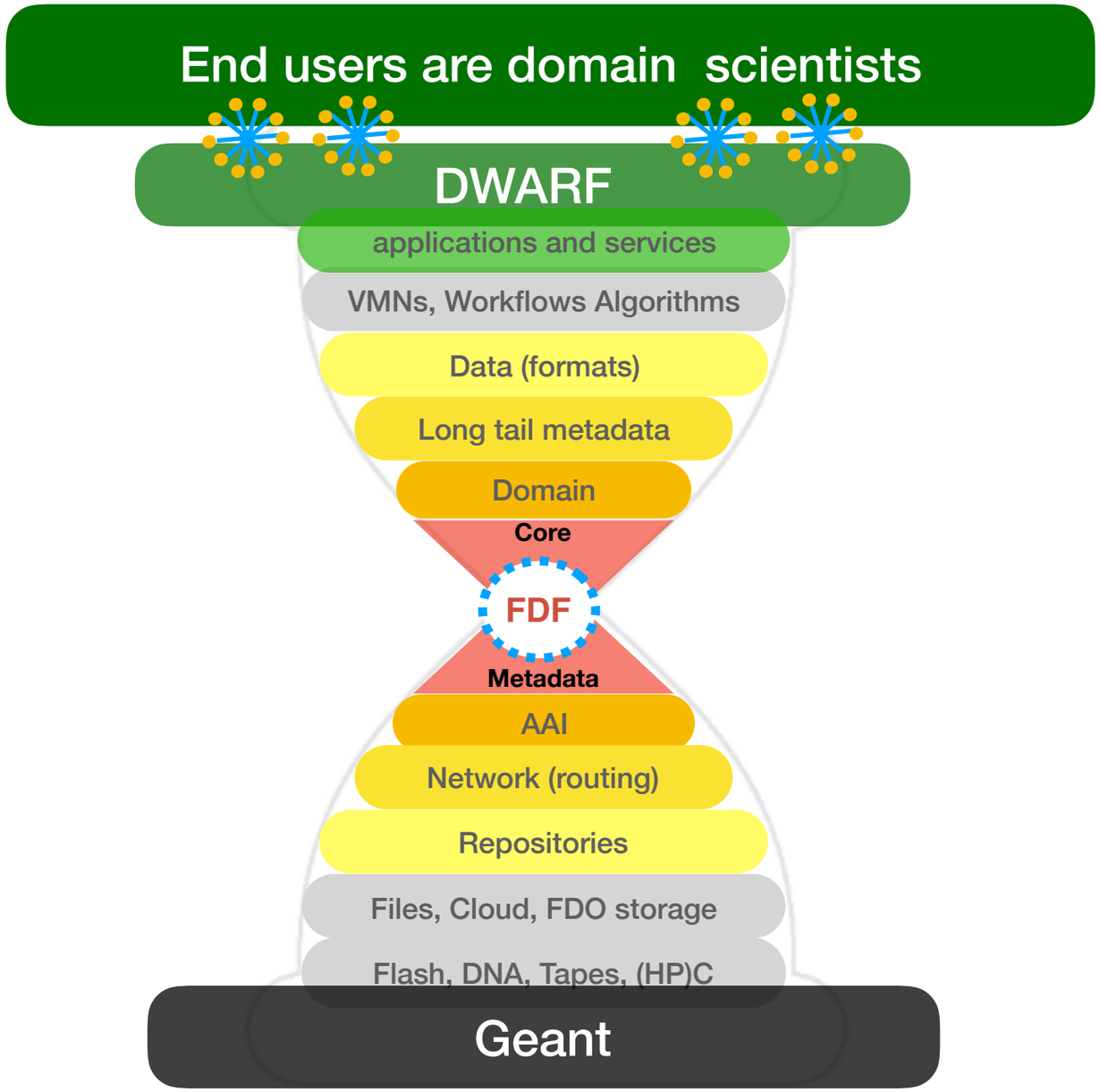


6-7: voluntary community adoption



8-12: Funder requested adoption




End users are domain scientists
DWARF
applications and services
VMNs, Workflows Algorithms
Data (formats)
Long tail metadata
Domain
Core
FDF
Metadata
AAI
Network (routing)
Repositories
Files, Cloud, FDO storage
Flash, DNA, Tapes, (HP)C
Geant

End users are domain scientists

DWARF

applications and services

VMNs, Workflows Algorithms

Data (formats)

Long tail metadata

Domain

Core

FDF

Metadata

AAI

Network (routing)

Repositories

Files, Cloud, FDO storage

Flash, DNA, Tapes, (HP)C

Geant

GO FAIR

RDA RESEARCH DATA ALLIANCE

UNESCO

CODATA

International Science Council

ICSU WORLD DATA SYSTEM

e-IRG e-Infrastructure Reflection Group

> 10 DSCC's

Example NL (Health)

End users are domain scientists

NPOS+

applications and services

VMNs, Workflows Algorithms

Data (formats)

Long tail metadata

Domain

Core

Metadata

AAI

Network (routing)

Repositories

Files, Cloud, FDO storage

Flash, DNA, Tapes, (HP)C

Geant

FDF

GO FAIR

DSCC IN

RDA

RESEARCH DATA ALLIANCE

CODATA

GO FAIR

CODATA

LUMC

PERSONAL HEALTH TRAIN

partner of Health-RI

data driven health RI

DTL

DUTCH TECHCENTRE FOR LIFE SCIENCES

UNESCO

International Science Council

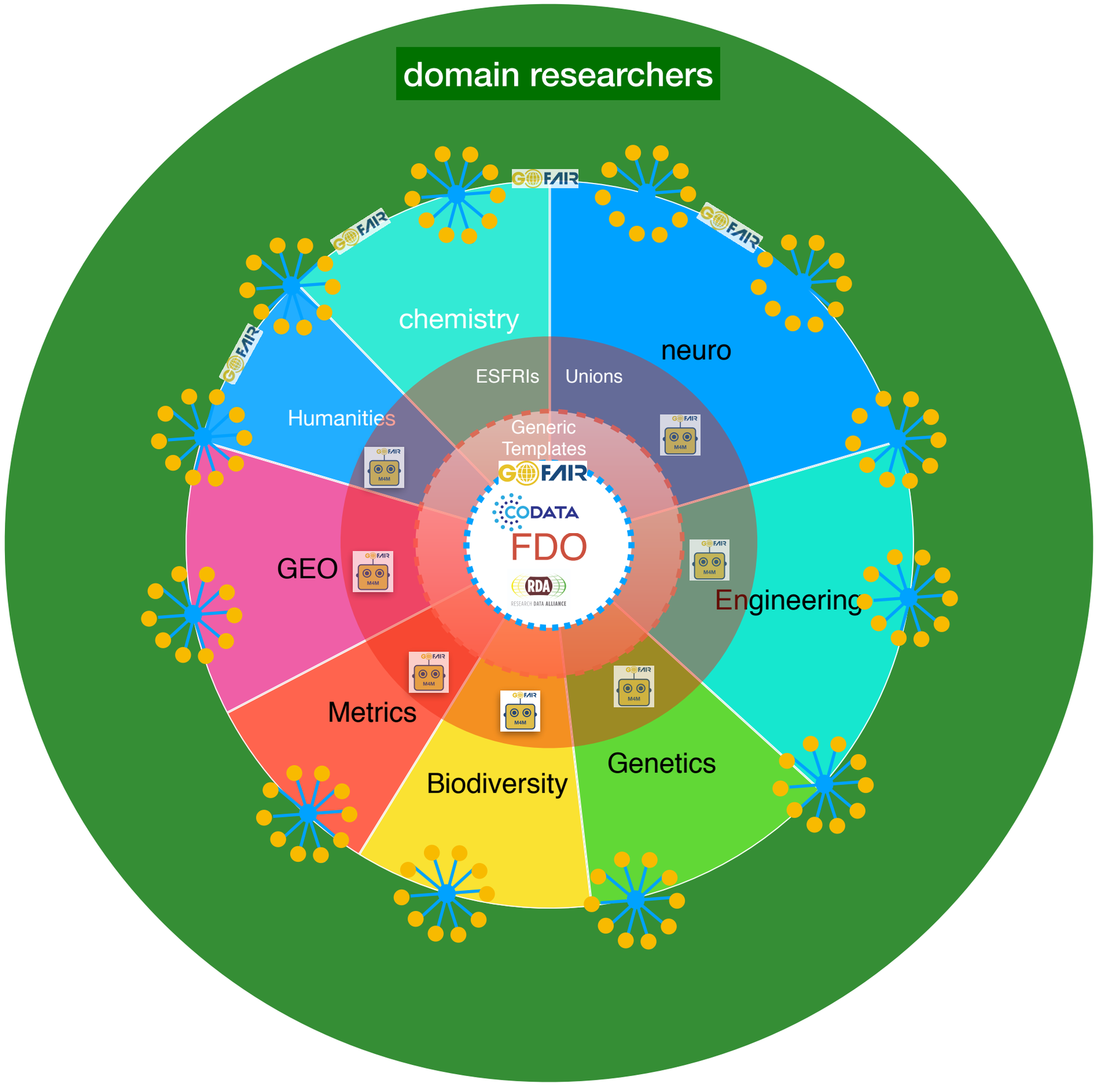
ICSU

WORLD DATA SYSTEM

e-IRG

e-Infrastructure Reflection Group

SURF


domain researchers
chemistry
neuro
Humanities
ESFRIs
Unions
Generic
Templates
GO FAIR
CODATA
FDO
RDA
RESEARCH DATA ALLIANCE
GEO
Engineering
Metrics
Biodiversity
Genetics
M4M

domain researchers

GO FAIR

Humanities

neuro

GEO

ESFRIs

Unions

Generic Templates

GO FAIR

CODATA

FDO

RDA

Metrics

Engineering

Biodiversity

Genetics

Global Alliance for Genomics & Health

elixir

DTL DUTCH TECHCENTRE FOR LIFE SCIENCES

THE HUMAN VARIOME PROJECT an NGO official partner of UNESCO

GODAN Global Open Data for Agriculture & Nutrition

End users

LUMC

Metadata = FDO

Generic templates

Basic genomics

Bioinformatics

Federated Human Data

Develops long-term strategies for managing and accessing sensitive human data.

Human Copy Number Variation

Aims to make it easier to detect, annotate and interpret human Copy Number Variations (hCNVs).

Marine Metagenomics

Develops a sustainable metagenomics infrastructure to nurture research and innovation in the marine domain.

Microbial Biotechnology

Helps the development of tailor-made microbes and biological systems.

Proteomics

Develops and maintains sustainable proteomics tools and data resources.

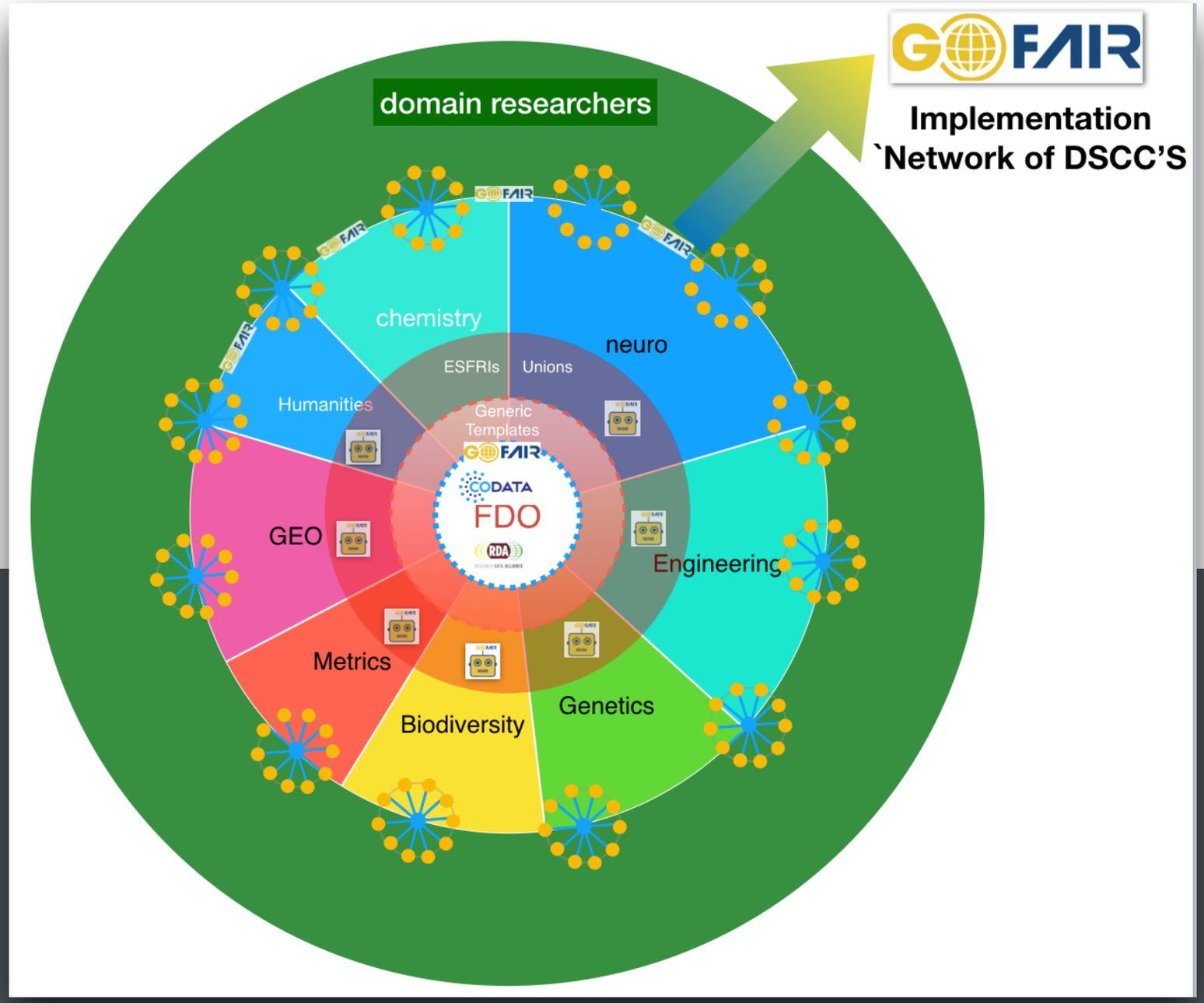
Rare Diseases

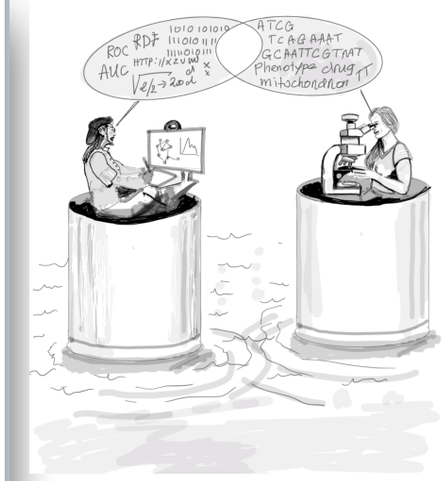
Supports the development of new therapies for rare diseases.

Plant Sciences

Develops an infrastructure to facilitate genotype-phenotype analysis for crop and tree species.

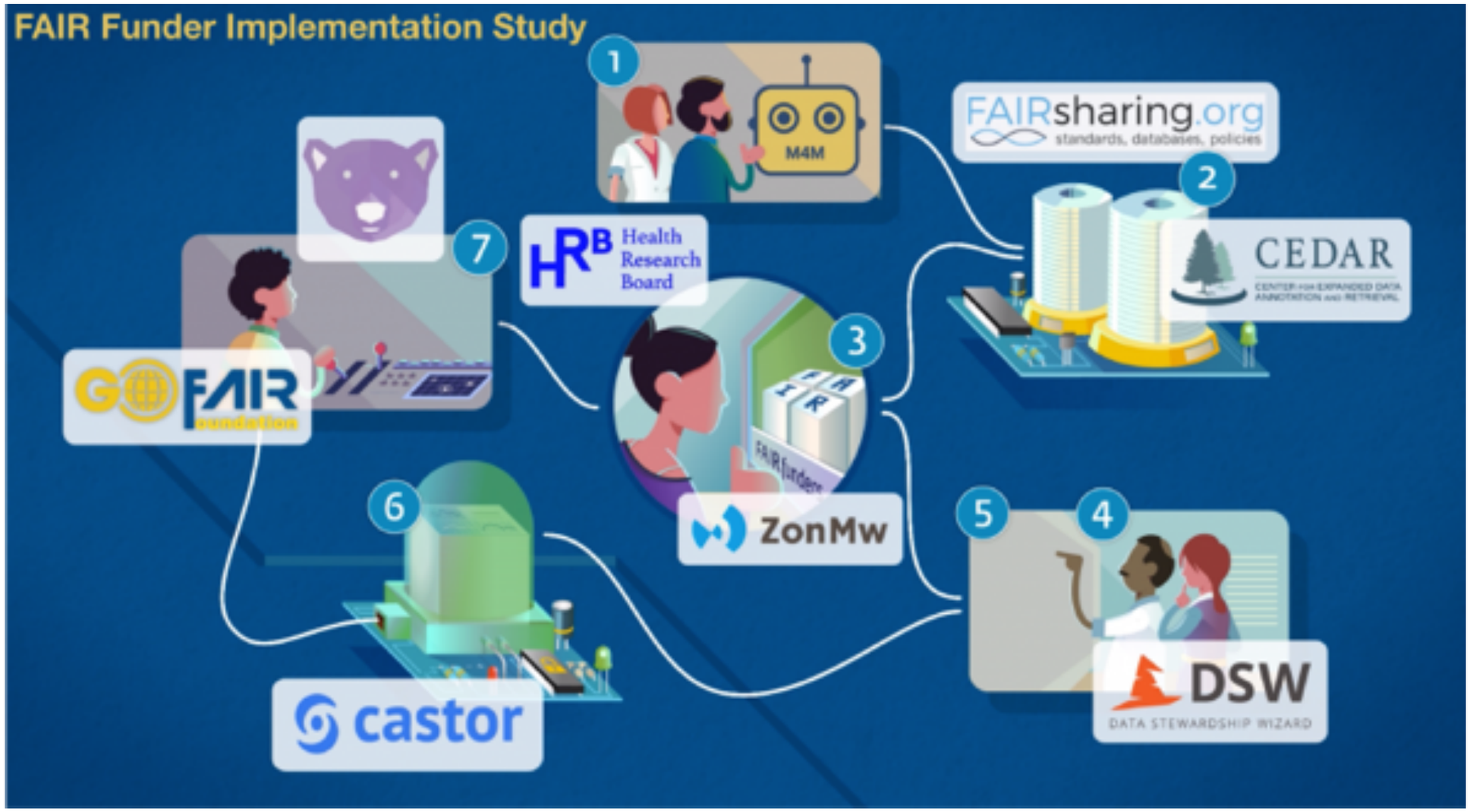
The Future belongs to Data Stewards !!





CODATA

GO FAIR


FAIR Funder Implementation Study
1
M4M
FAIRsharing.org
standards, databases, policies
2
CEDAR
CENTER FOR EXPANDED DATA ANNOTATION AND RETRIEVAL
7
HRB Health Research Board
3
ZonMw
GO FAIR foundation
6
castor
5
4
DSW
DATA STEWARDSHIP WIZARD

FAIR made easy