http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

- Open
- Artificial Intelligence
- Big Data
- Data Sharing

Policy Briefing:

How to Reconcile GDPR with Artificial Intelligence:

Europe's Competitive Edge in Technological Innovation?

18 April 2018 | Berlaymont, Rue de la Loi, 200

13h00 – 14h00 | **Room 12/040**

European **Political Strategy** Centre

Barend Mons

Director
GO FAIR International Support and Coordination Office

Visiting address: Poortgebouw N-01, Rijnsburgerweg 10, 2333 AA Leiden,The Netherlands, E-mail: barend.mons@go-fair.org, Mobile: +31 6 24879779, Skype: dnerab

www.go-fair.org

- European
- Open
- Science
- Cloud

The FAIR Data Principles set out requirements for data to be processed in an automated way

WORLD ECONOMIC FORUM
COMMITTED TO IMPROVING THE STATE OF THE WORLD

Findable:

"Easy to find by both humans and computer systems and based on mandatory description of the metadata that allow the discovery of interesting datasets"

- e.g. Able to locate data by individual patient, patient segment, intervention, outcome metric

Accessible:

"Stored for long term such that they can be easily accessed and / or downloaded with well-defined license and access conditions (Open Access when possible), whether at the level of metadata, or at the level of the actual data content"

- e.g. Patients should be able to access parts of their own data via a patient controlled record

Interoperable:

"Ready to be combined with other datasets by humans as well as computer systems"

- Semantic interoperability: mapped data taxonomies across diseases and population groups e.g. consistent methodology & scale for measuring pain / quality of life
- Technical interoperability: specifications to allow different systems to communicate with each other

Reusable:

"Ready to be used for future research and to be processed further using computational methods"

- e.g. Outcomes data should be available for the long-term for systematic analysis or clinical research (with permission from data owner)

Important that interoperable datasets can be interpreted by computer systems: to (semi) automatically combine different data sources for richer knowledge discovery

Source: Dutch Techcentre for Life Sciences

Governance recommendations of the HLEG EOSC (IFDS)

G1: Aim at the **lightest possible**, internationally effective governance.

G2: Guidance **only** where guidance is due.

G3: Define **Rules of Engagement** for formal participation in the EOSC.

G4: Federate the **Gems across Member States.**

GO FAIR will obviously also honour the P and I recommendations of the HLEG

EOSC Summit 2018 – Brussels, 11 June 2018

GO FAIR in a nutshell

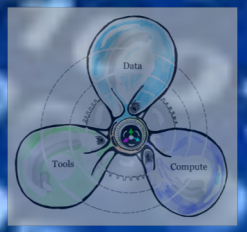
Towards EOSC as the Internet of FAIR data and services

Prof. Barend Mons
GO FAIR ISCO
Leiden, Hamburg, Paris

- GO FAIR supports **bottom up**, **achievable** community practices for establishing the EOSC as part of the **Global Internet of FAIR Data and Services (IFDS)**
- Co-founded & financed by 3 Member States (NL, DE, FR) but **open to all**, GO FAIR aims to **kick-start** the development of the EOSC through **communities of excellence** 'Implementation Networks' committed to collectively engage in the IFDS
- Supported by three pillars - GO CHANGE (culture), GO TRAIN (data stewards) & GO BUILD (technologies or components).
- Any country can join, coordinating national participation in networks and contribute GO FAIR expertise

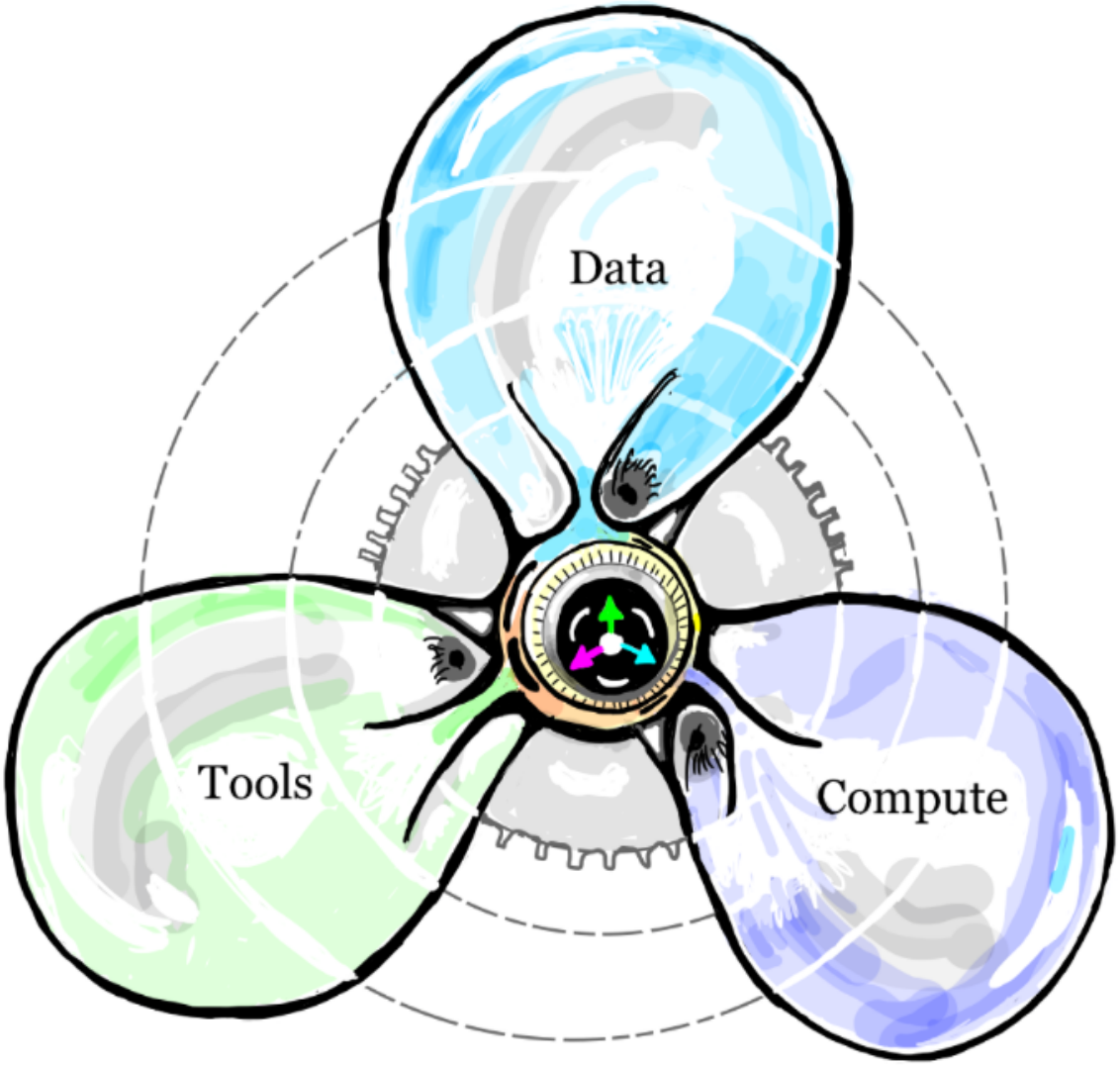
Implement FAIR principles co-create the IFDS

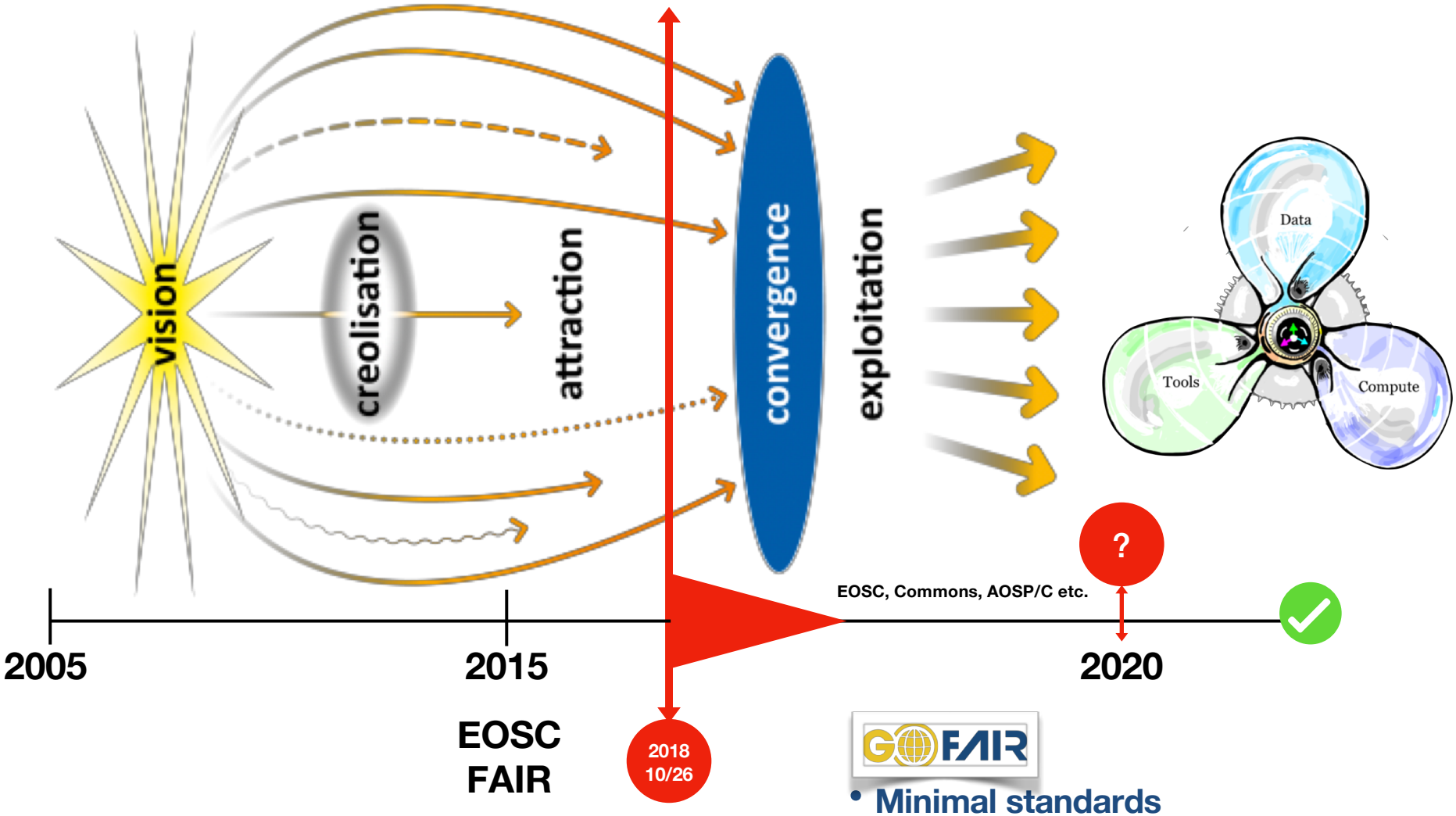
The Internet for Social Machines



=The Internet of FAIR Data and Services

The Internet of FAIR data and Services

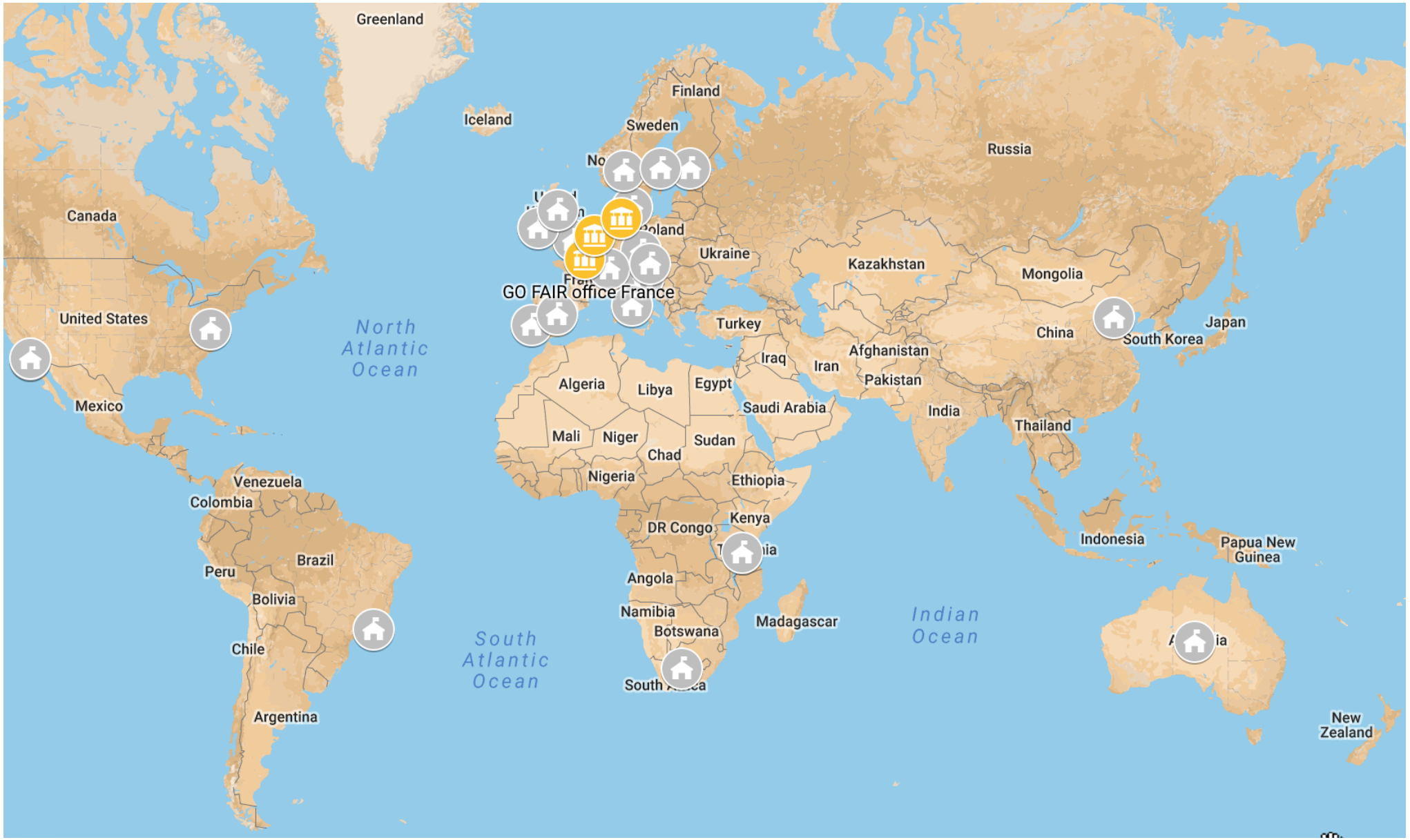




GO FAIR

- Minimal standards
- Voluntary participation
- Critical mass
- Rough consensus and running code

ISCO'S and prospective National offices



1a

Example: Personal Health Train based on distributed data

WORLD ECONOMIC FORUM
COMMITTED TO IMPROVING THE STATE OF THE WORLD

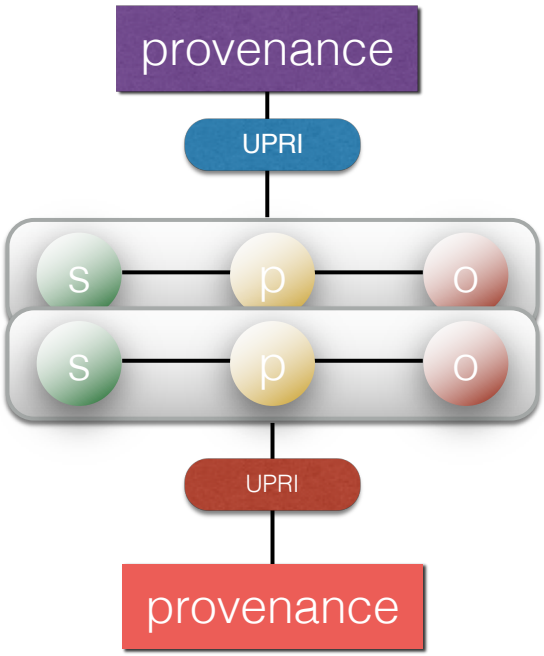
Component	Description
 **FAIR data station**	• Owned by data owner (e.g. hospital) • Contains findable, accessible, interoperable, reusable data, to which access can be granted to others
 **Personal Health Train**	• Algorithms owned by users or service providers (e.g. researchers) • Drives by data stations to perform analyses and collect insights from data without extracting or aggregating data beyond the firewall
 **Train track**	• Owned by public entity (e.g. government organization) • Is a secure environment to which access can be granted by the public entity to service providers

Source: MAASTRO; Dutch Techcentre for Life Sciences

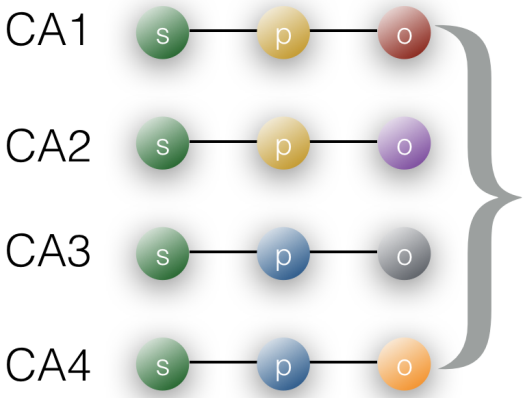
see: GODAN Discussion paper.


Personal Health Train

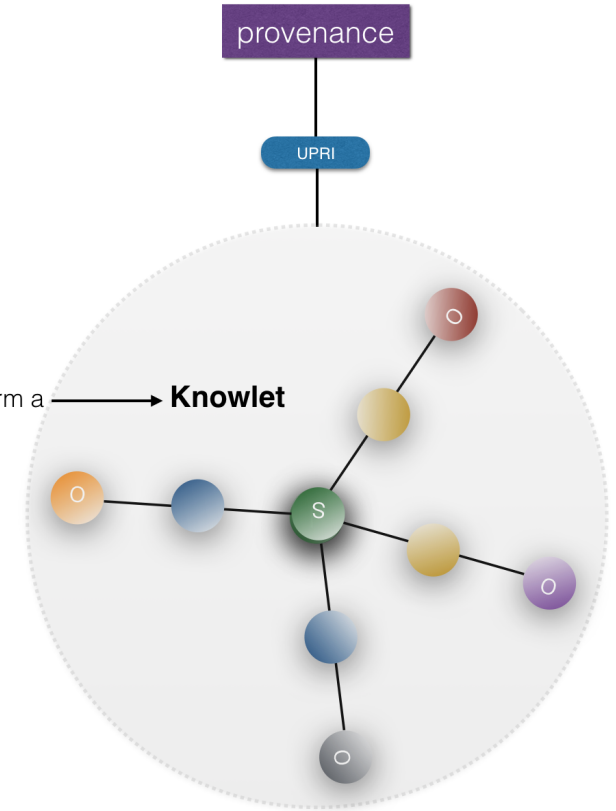

7 WAYS
TO
prevent
change



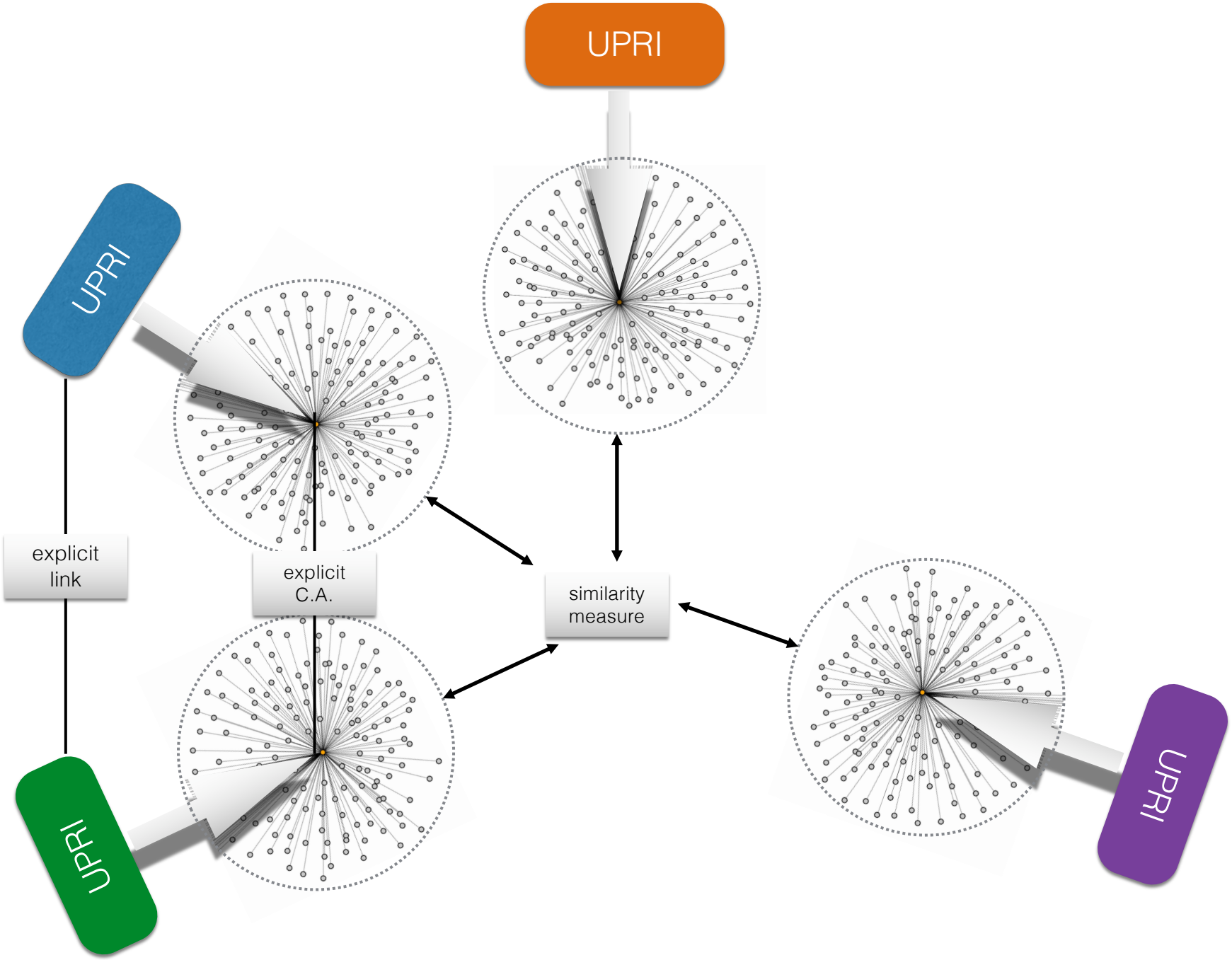
Multiple different cardinal Assertions*
with the same **subject**



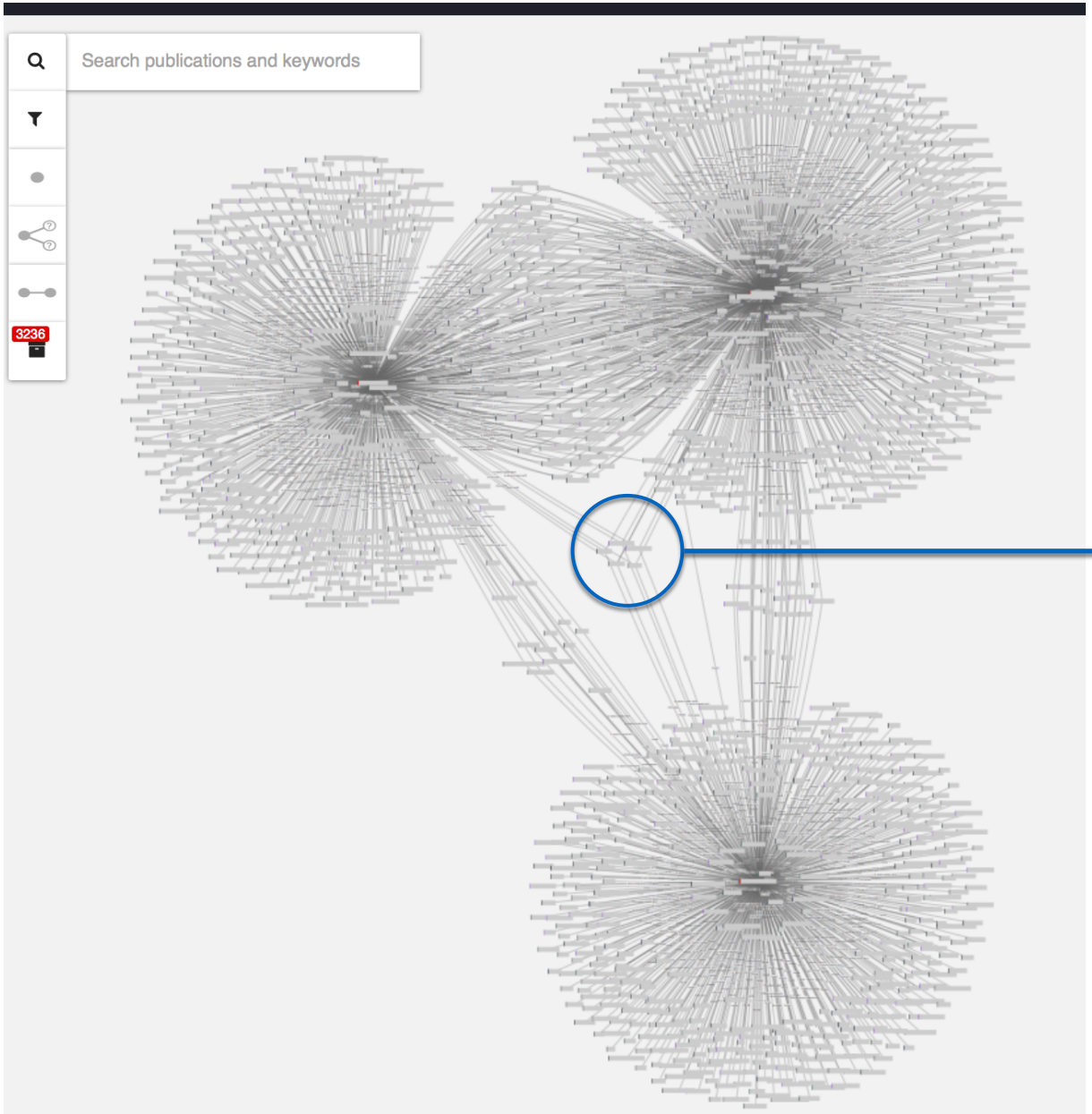
form a → **Knowlet**



* UPRI's and Provenance not depicted for simplicity reasons


UPRI
UPRI
explicit
link
explicit
C.A.
similarity
measure
UPRI
UPRI

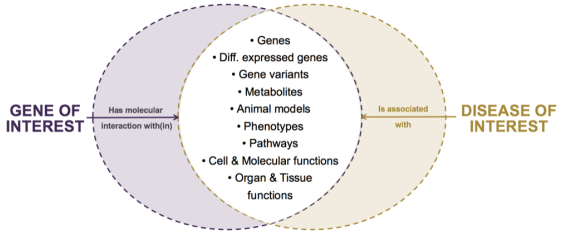
EURETOS



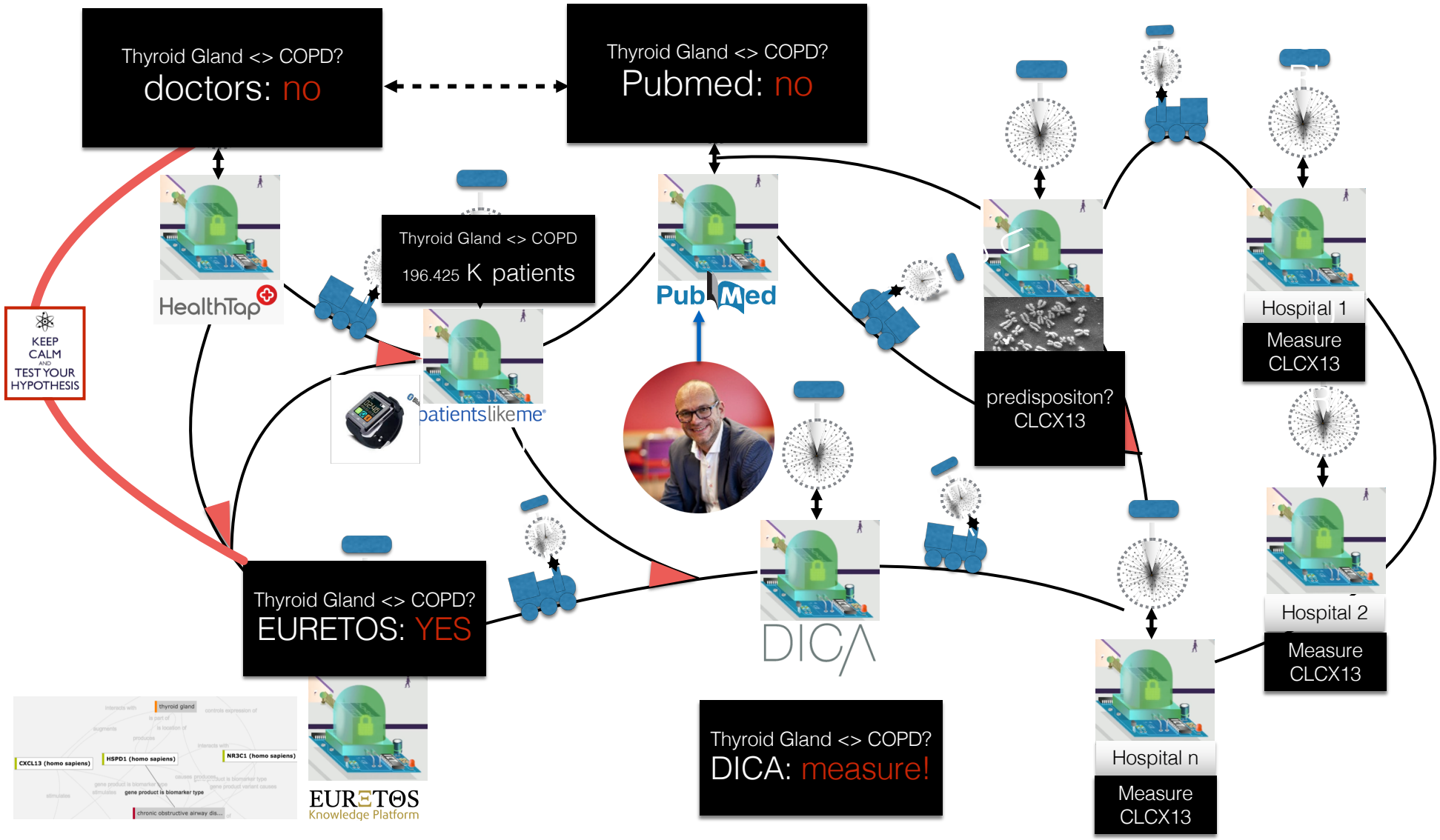
EURETOS

Human readable report
With all references

Evaluate 10 key aspects that are relevant to assess a gene as a viable target or biomarker for a disease including healthy expression, differential expression, gene variants, animal models, gene, protein & metabolite interactions, phenotypes, pathways, cell & molecular function and organ & tissue functions.



share: in this case 5 objects are shared between all three knowlets (in this case: metabolic syndrome


Thyroid Gland <> COPD?
doctors: no
Thyroid Gland <> COPD?
Pubmed: no
HealthTap
Thyroid Gland <> COPD
196.425 K patients
patientslikeme
PubMed
KEEP CALM AND TEST YOUR HYPOTHESIS
predispositon?
CLCX13
Hospital 1
Measure
CLCX13
Thyroid Gland <> COPD?
EURETOS: YES
DICA
Hospital 2
Measure
CLCX13
Thyroid Gland <> COPD?
DICA: measure!
Hospital n
Measure
CLCX13
EURETOS
Knowledge Platform

old (largely failed) situation

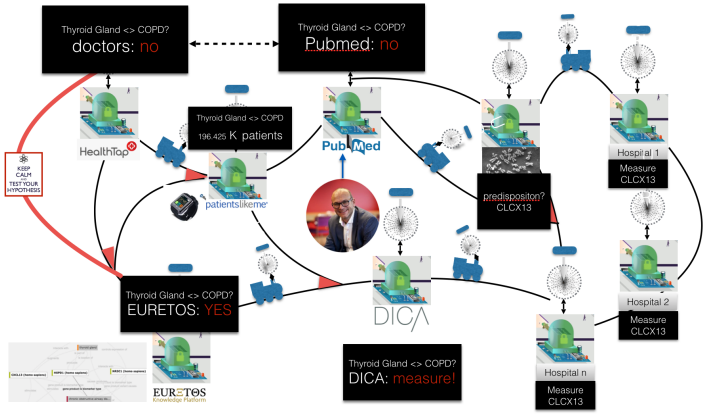
proprietary app

proprietary data warehouse

- Streaming ETL
- Central updating
- GDPR issues
- Enormous costs
- Black Box

New (IFDS) situation

proprietary app



Distributed FAIR data stations

- No cumbersome ETL
- Decentral updating + provenance
- No GDPR issues
- distributed costs
- Transparent

Questions/hypotheses
generated by people

Citizens
Medical Doctors
Researchers
policy makers

?

HealthTap

EURETOS
Knowledge Platform

Questions/hypotheses
generated by Machines

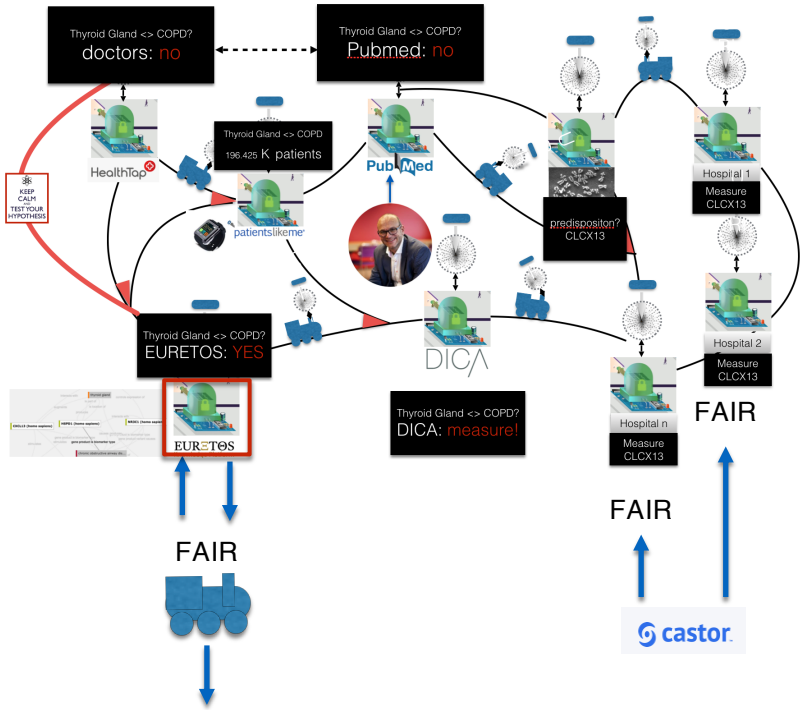
castor
FORMS

SCRIPTS

SPARQL

FAIR

ENTER PHT
any place



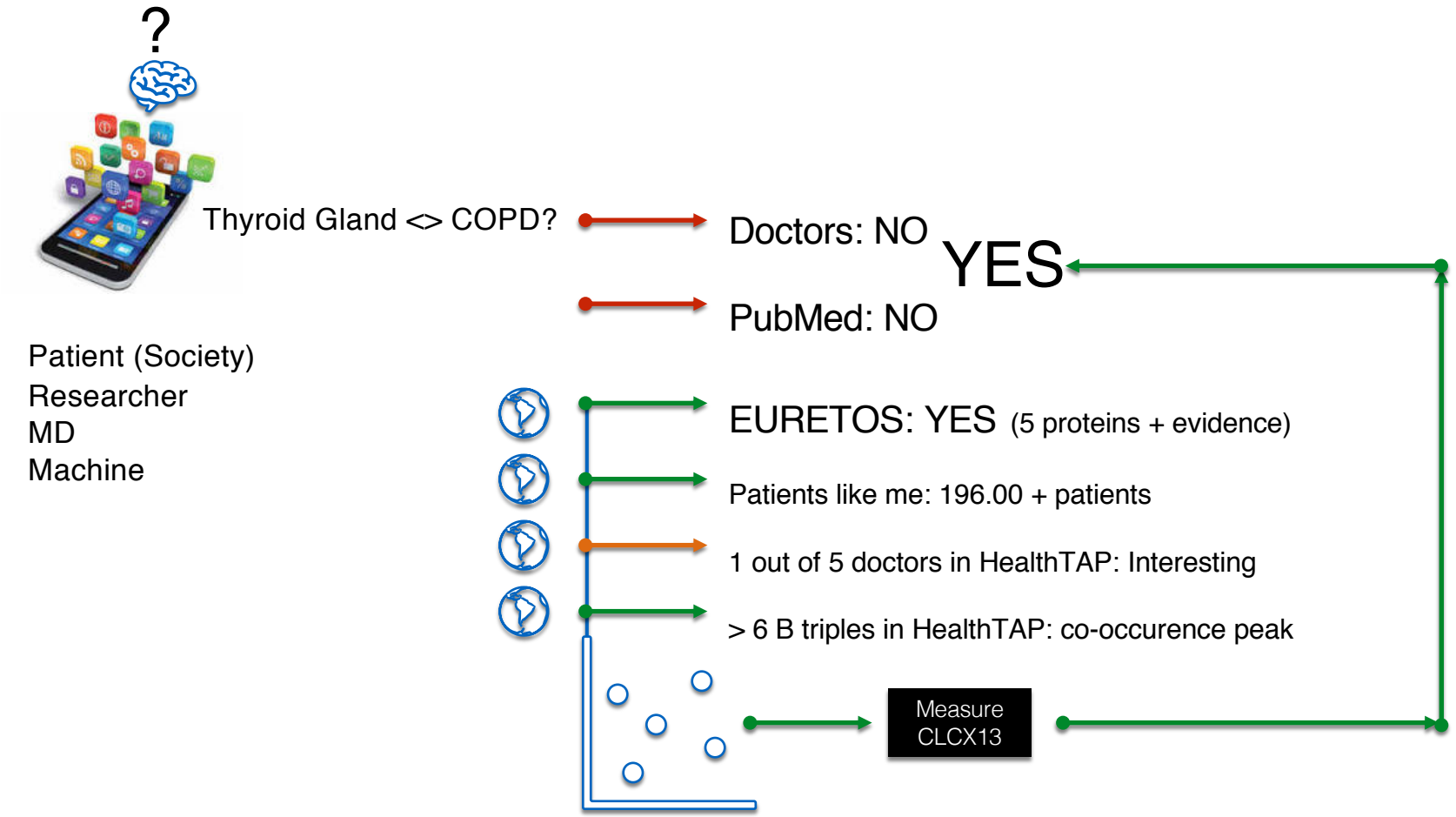
FAIR

FAIR

FAIR

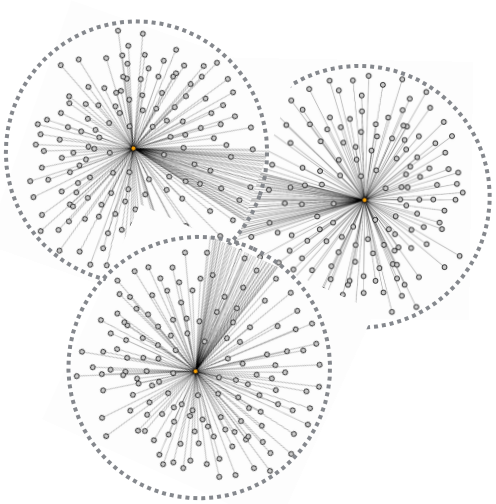
castor

report with full provenance
and references



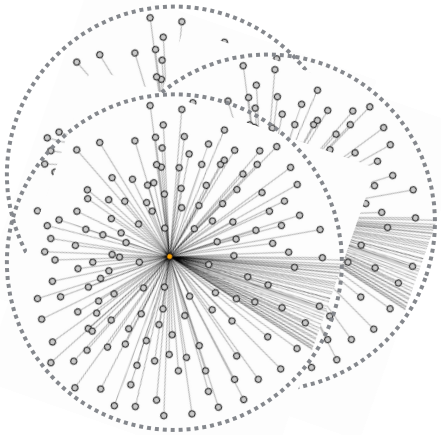
CONNECTING THE DOTS WITH DISTRIBUTED LEARNING

A



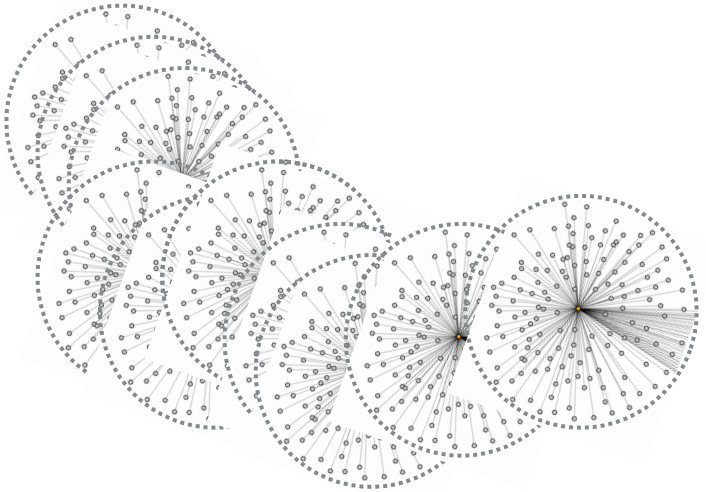
Dynagraph

B



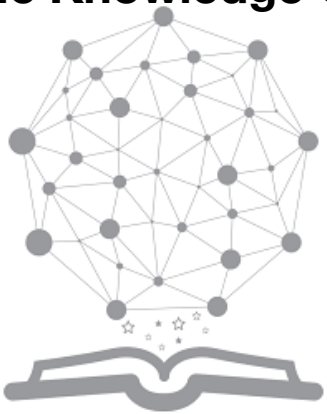
Near Sameness

C



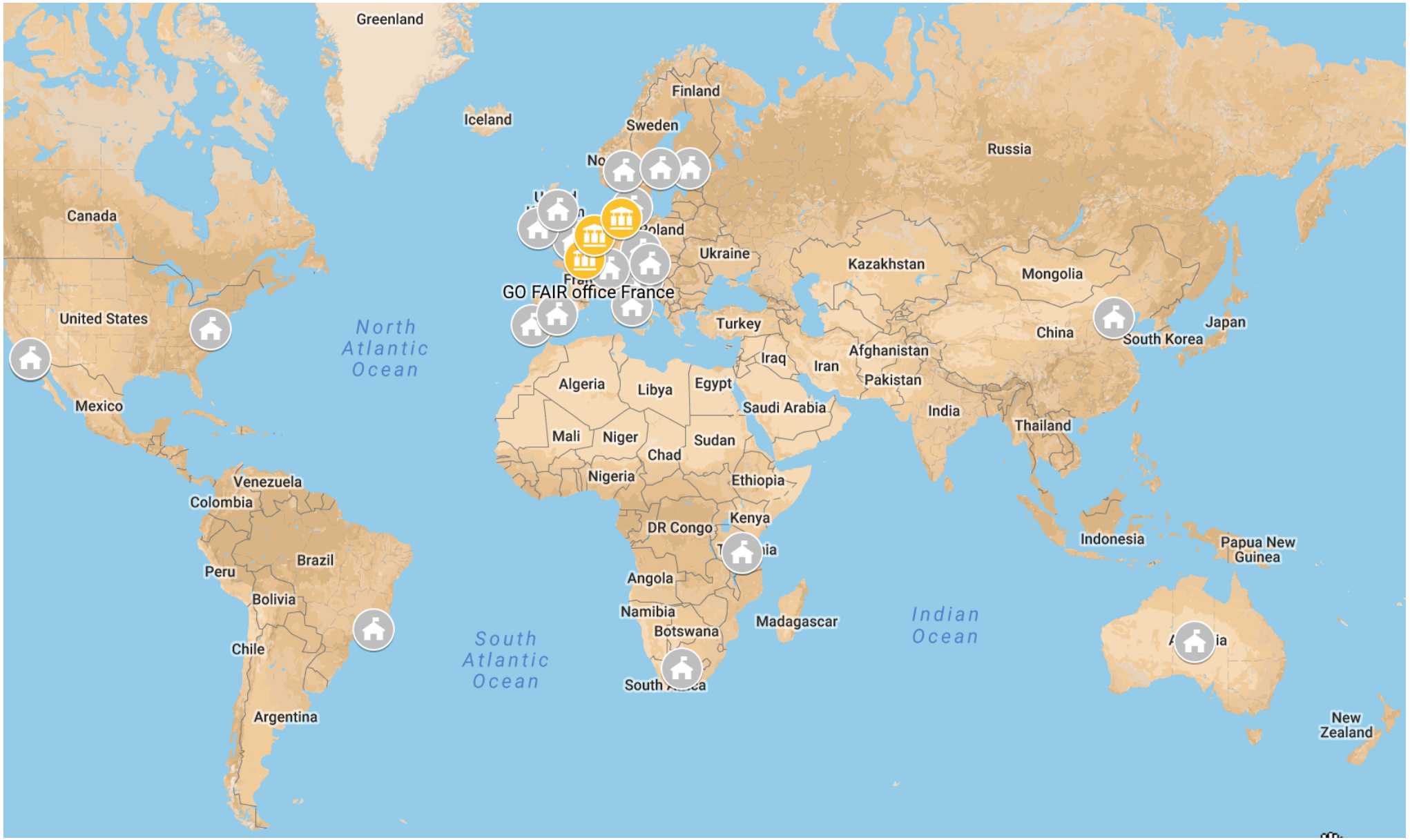
Conceptual Drift

Static Knowledge Graph

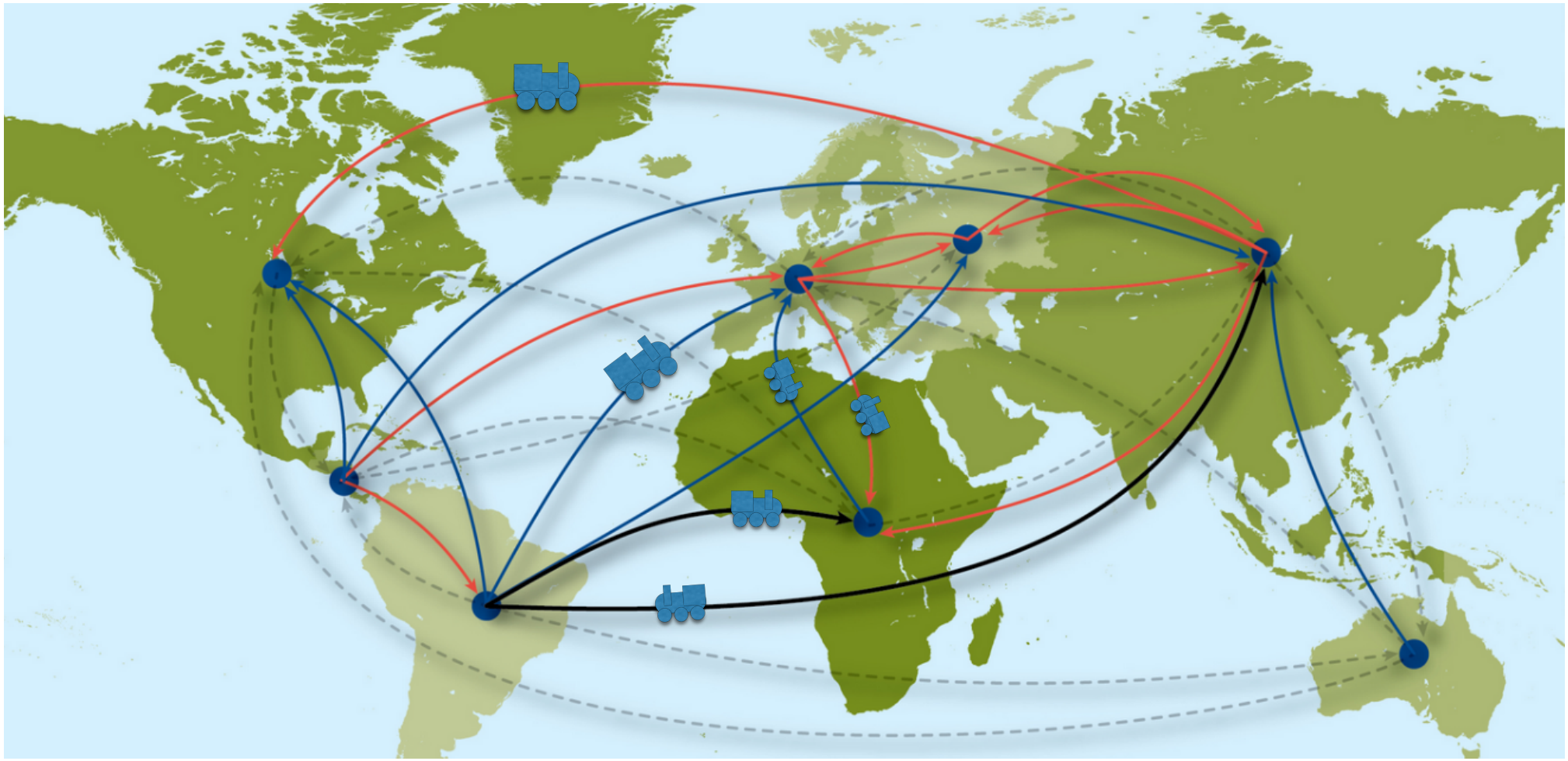


reading……

ISCO'S and prospective National offices



FAIR data points are globally distributed



FAIR Trains can sent by anyone, anywhere

concluding statements

- Avoid Hype Terms (or define them precisely)
- Provenance is (even) more important than quality
- Principle of Reciprocity and Equality > no wider divide!
- Citizens should be partners in research not data sources only
- balance between human readable and machine actionable (don’t mix)
- This all only works when data and services are FAIR