Why FAIR?

Enabling data intensive research !

SCIENTIFIC DATA

Open Access | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, [...] Barend Mons

Scientific Data **3**, Article number: 160018 (2016) | Cite this article

130k Accesses | **1680** Citations | **1571** Altmetric | Metrics

An Addendum to this article was published on 19 March 2019

The machine knows what I mean

Abstract

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measureable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

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

Virus Outbreak Data Network (VODAN)

The VODAN Implementation Network is one of the joint activities carried out by **CODATA, RDA, WDS, and GO FAIR** (Link to the **Data Together Statement**).
Read the full statement on **Data Together COVID-19 Appeal and Actions**.

Active GO FAIR Implementation Network

The spread of the virus causing the COVID-19 outbreak is far from over. During this epidemic and in earlier occasions, we have seen severely suboptimal data management and data reuse. Moreover, access to the immensely valuable data of past and current epidemics is not always equally accessible for different affected populations and countries. For instance, the data from the past Ebola epidemics are very difficult to find, to access, and if accessible, they are not interoperable, *let alone reusable*. Under the urgent need to harness machine-learning and future AI approaches to discover meaningful patterns in epidemic outbreaks, we need to do better and ensure that data are FAIR (in this sense also meaning **F**ederated, **AI**–**R**eady).

ZonMw

VODAN AFRICA

The VODAN-IN approach: distributed analytics over FAIR data

https://vimeo.com/143246458

VODAN AFRICA

FAIR reference (EK) data stations

FAIR RWO data stations

Trains - FAIR algorithms

The end of data sharing

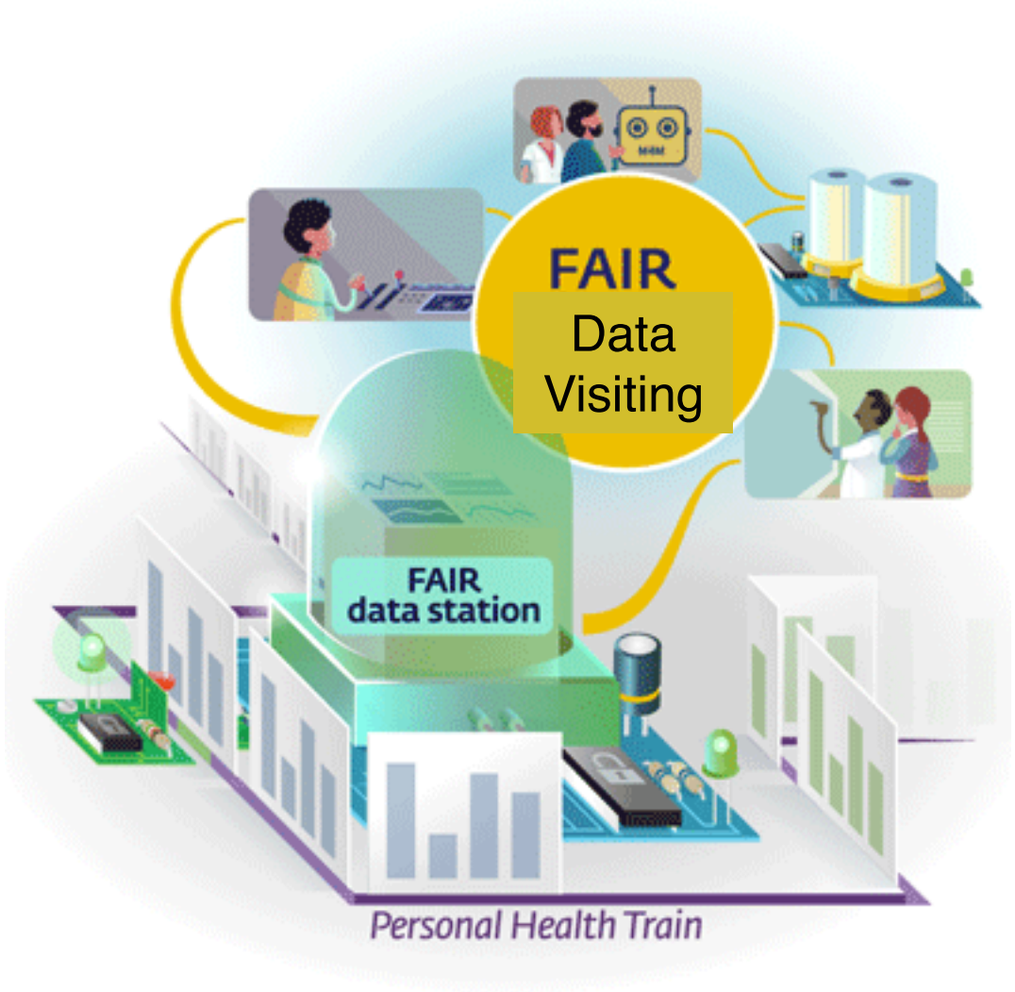


Fig.5

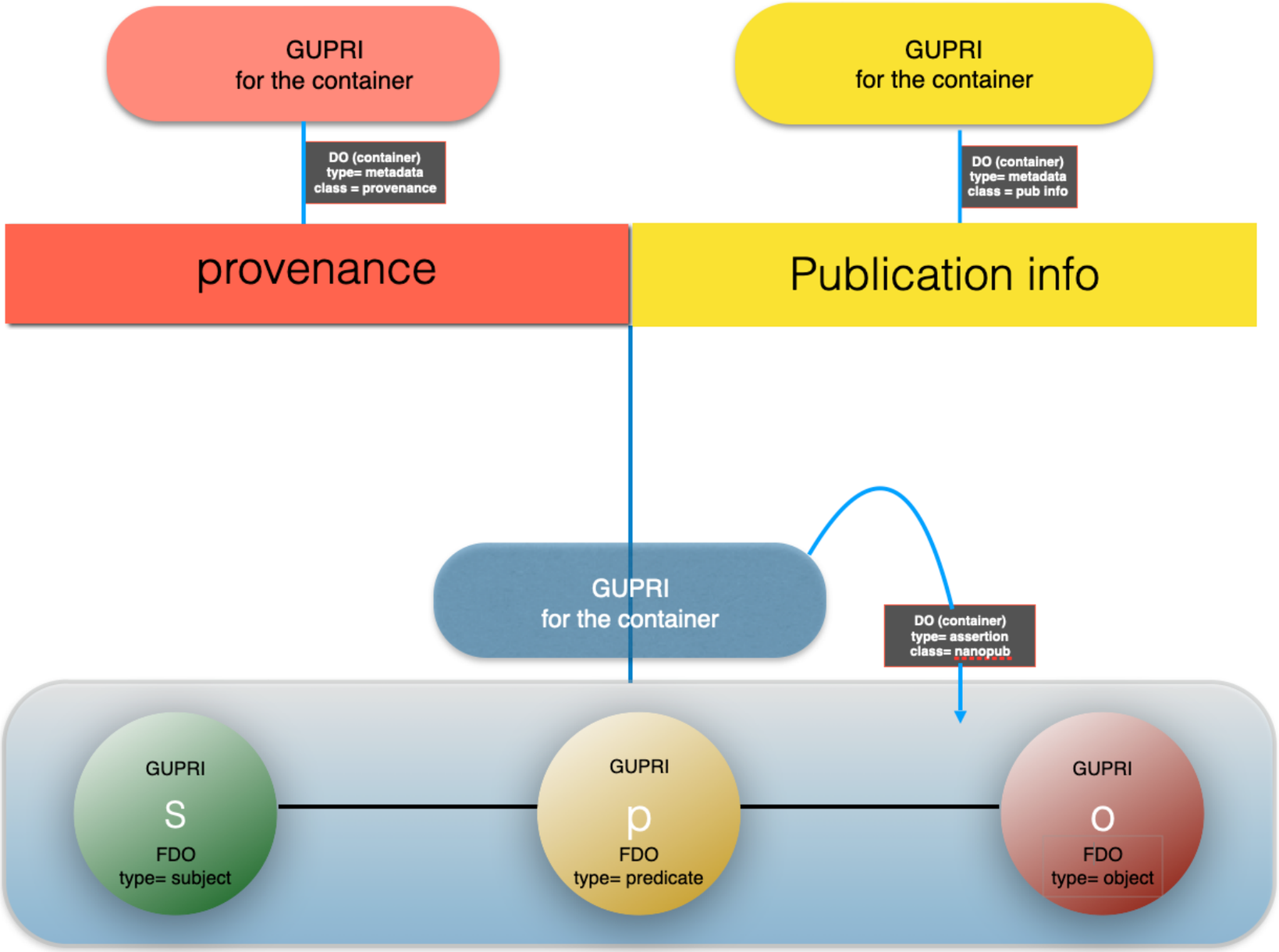

GUPRI
for the container
DO (container)
type= metadata
class = provenance
provenance
GUPRI
for the container
DO (container)
type= metadata
class = pub info
Publication info
GUPRI
for the container
DO (container)
type= assertion
class= nanopub
GUPRI
S
FDO
type= subject
GUPRI
p
FDO
type= predicate
GUPRI
o
FDO
type= object

Fig.8

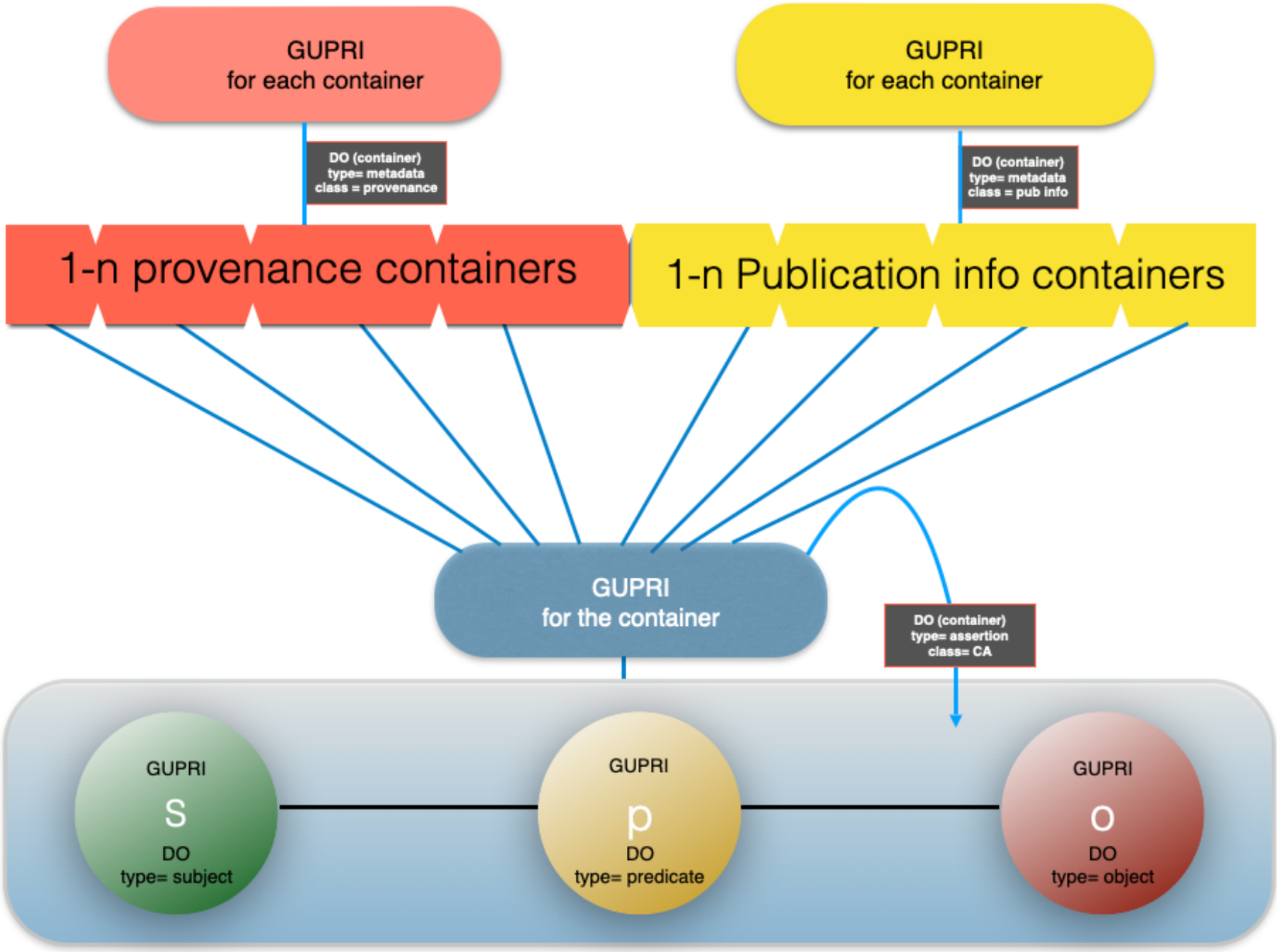
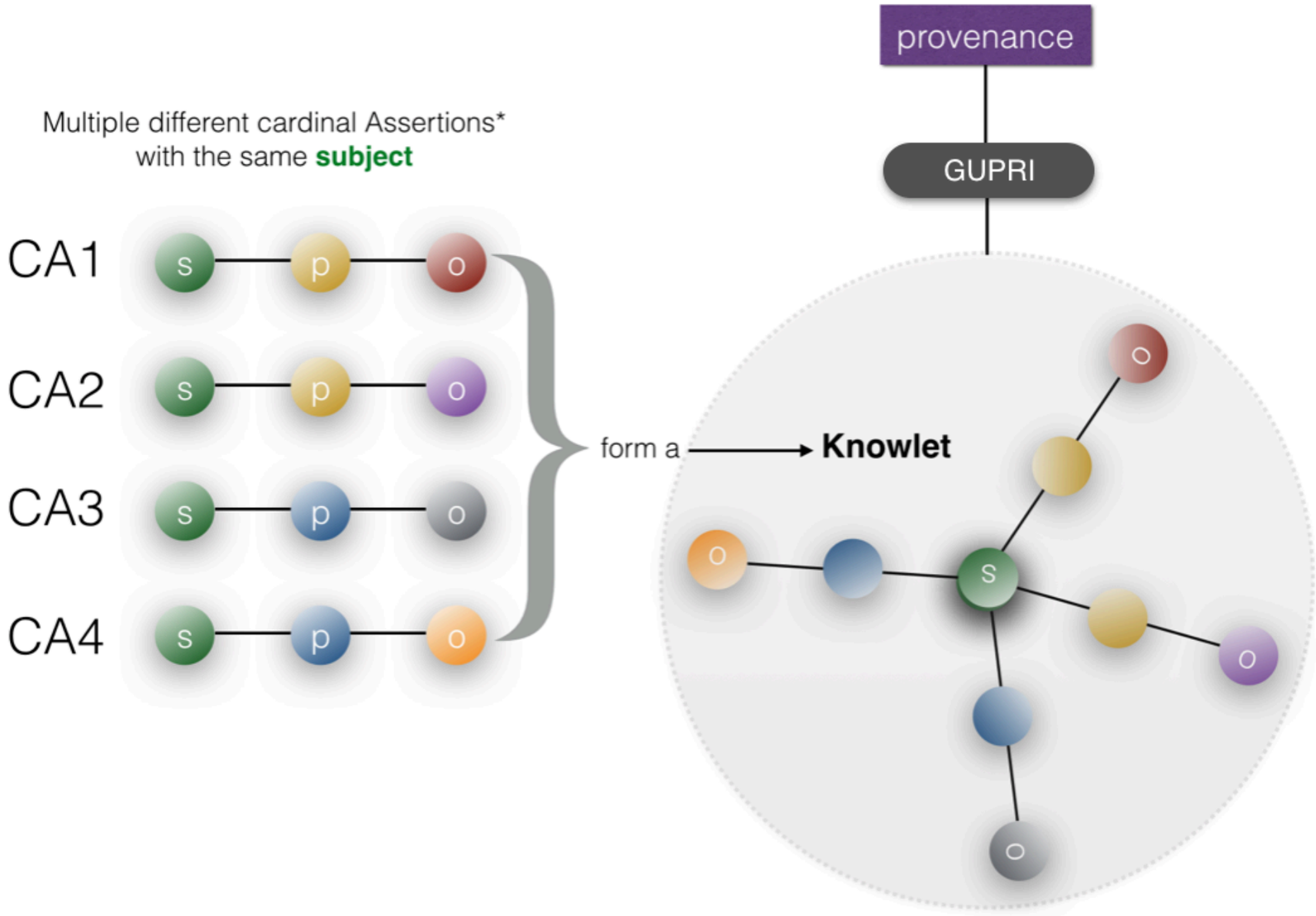
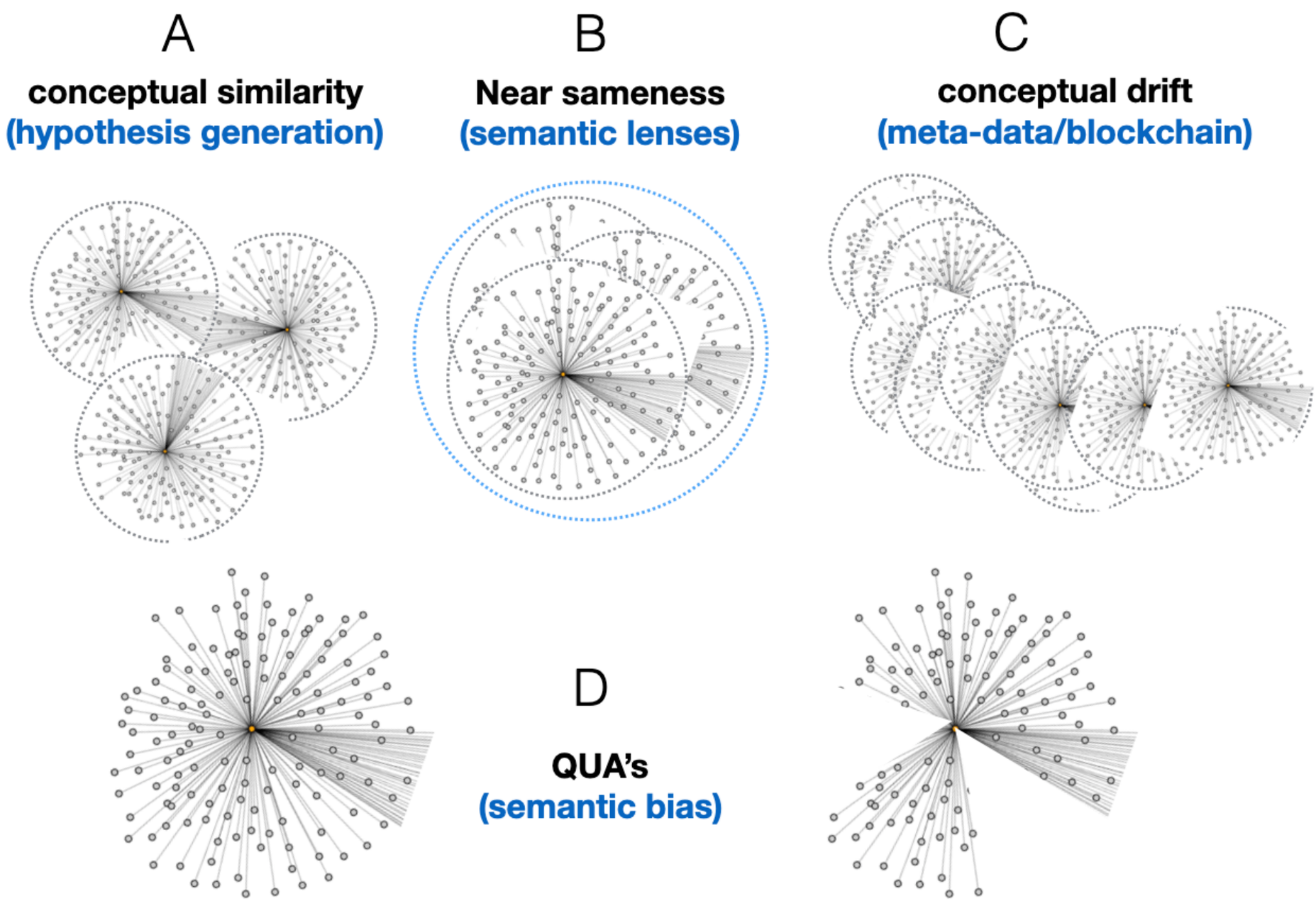

GUPRI
for each container
GUPRI
for each container
DO (container)
type= metadata
class = provenance
DO (container)
type= metadata
class = pub info
1-n provenance containers
1-n Publication info containers
GUPRI
for the container
DO (container)
type= assertion
class= CA
GUPRI
S
DO
type= subject
GUPRI
p
DO
type= predicate
GUPRI
o
DO
type= object

Fig.10

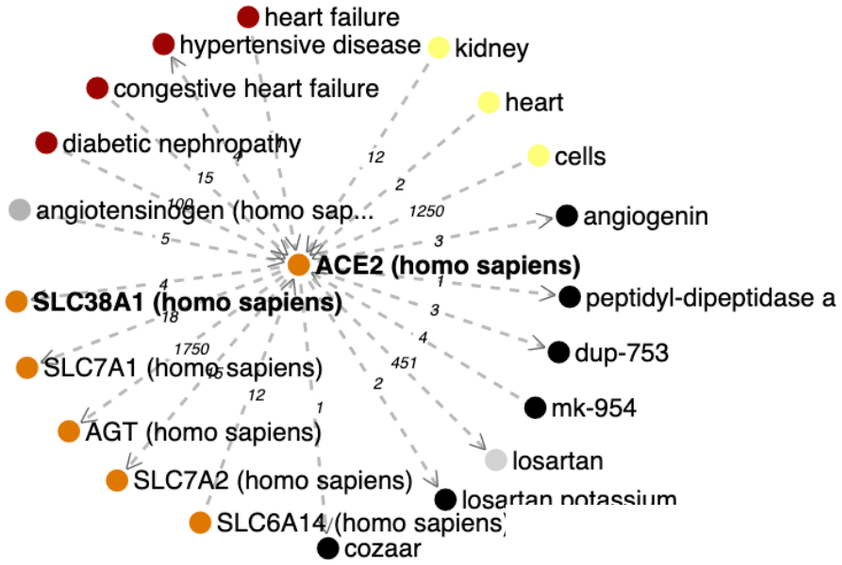


*GUPRI's and Provenance not depicted for simplicity reasons

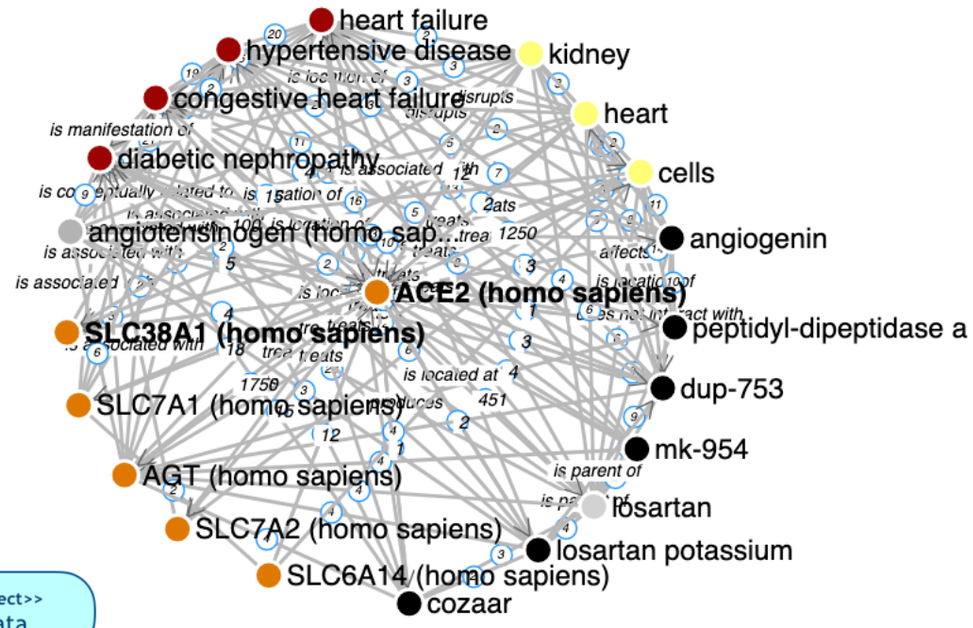
Fig.10

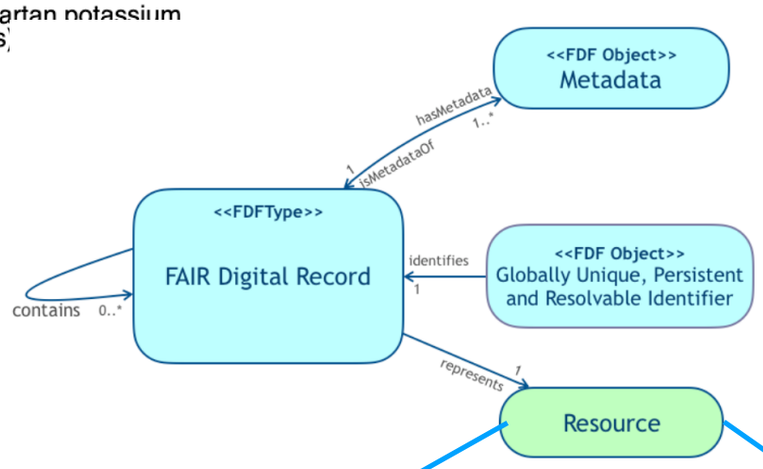

A
conceptual similarity
(hypothesis generation)
B
Near sameness
(semantic lenses)
C
conceptual drift
(meta-data/blockchain)
D
QUA's
(semantic bias)

Knowlet (1-many)



Graph (many to many)





Physical Object

A tissue sample of
A real tumour

Mental construct

‘Pancreatic cancer’

DTL DUTCH TECHCENTRE FOR LIFE SCIENCES

health RI enabling data driven health

GO FAIR foundation

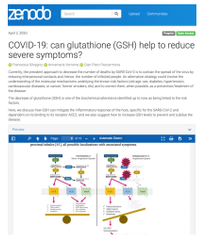
Disease Modelling Workflow (COVID-19)

New Publications
Real World Observations
Clinical and Self reporting

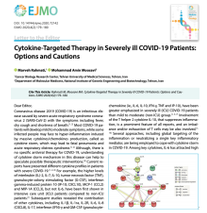
AI-ready Established Knowledge
Plus selected connectors
Subhypotheses

Radiology

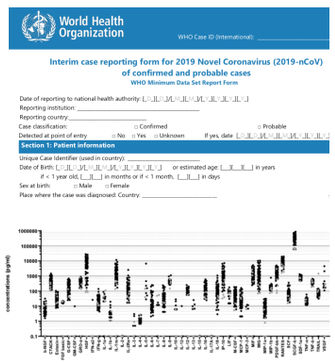
Diagnosis, Prevention, and Treatment of Thromboembolic Complications in COVID-19: Report of the National Institute for Public Health of the Netherlands


zenodo

COVID-19: can glutathione (GSH) help to reduce severe symptoms?


EJMO

Cytokine-Targeted Therapy in Severely Ill COVID-19 Patients: Options and Cautions


World Health Organization

Interim case reporting form for 2019 Novel Coronavirus (2019-nCoV) of confirmed and probable cases

WHO Minimum Data Set Report Form

A1

FAIR abstract Paper1

FAIR abstract Paper 2

FAIR abstract Paper 3

A2

Case Report Forms/ EPD

Clinical measurements

Self Reporting (apps)

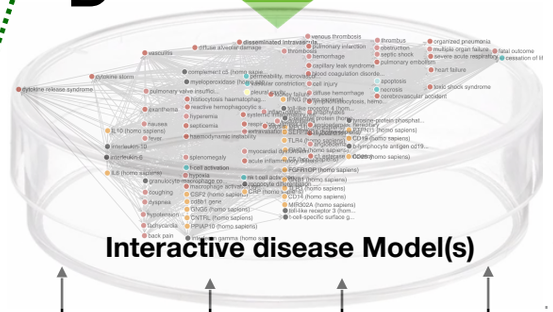
FAIR

EURETOS AI PLATFORM (Other tools if available)

C
Greatest Common Denominator (GDC) 'Connectome'

Multiple Algorithms

D



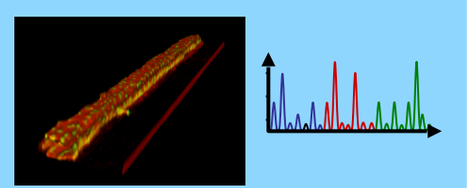
B

- SARS-Cov-2/human proteome
- ACE2/ACE
- Cytokine Storm (CRS)
- Renin-angiotensin system
- Further sub-hypotheses

E

- Expert introduced concept
- Drug, mechanism, cell type etc.

F

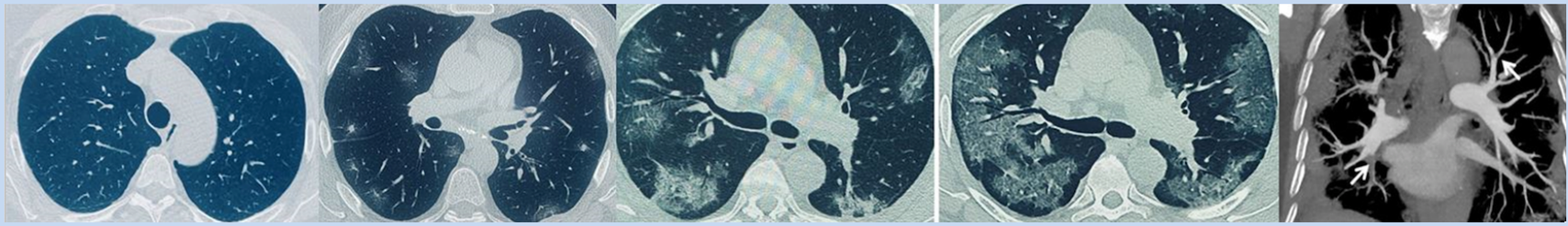


G
Curation, annotation and Hypothesis discussion
Disease phase alignment
Innovative OA publication
Visualisation

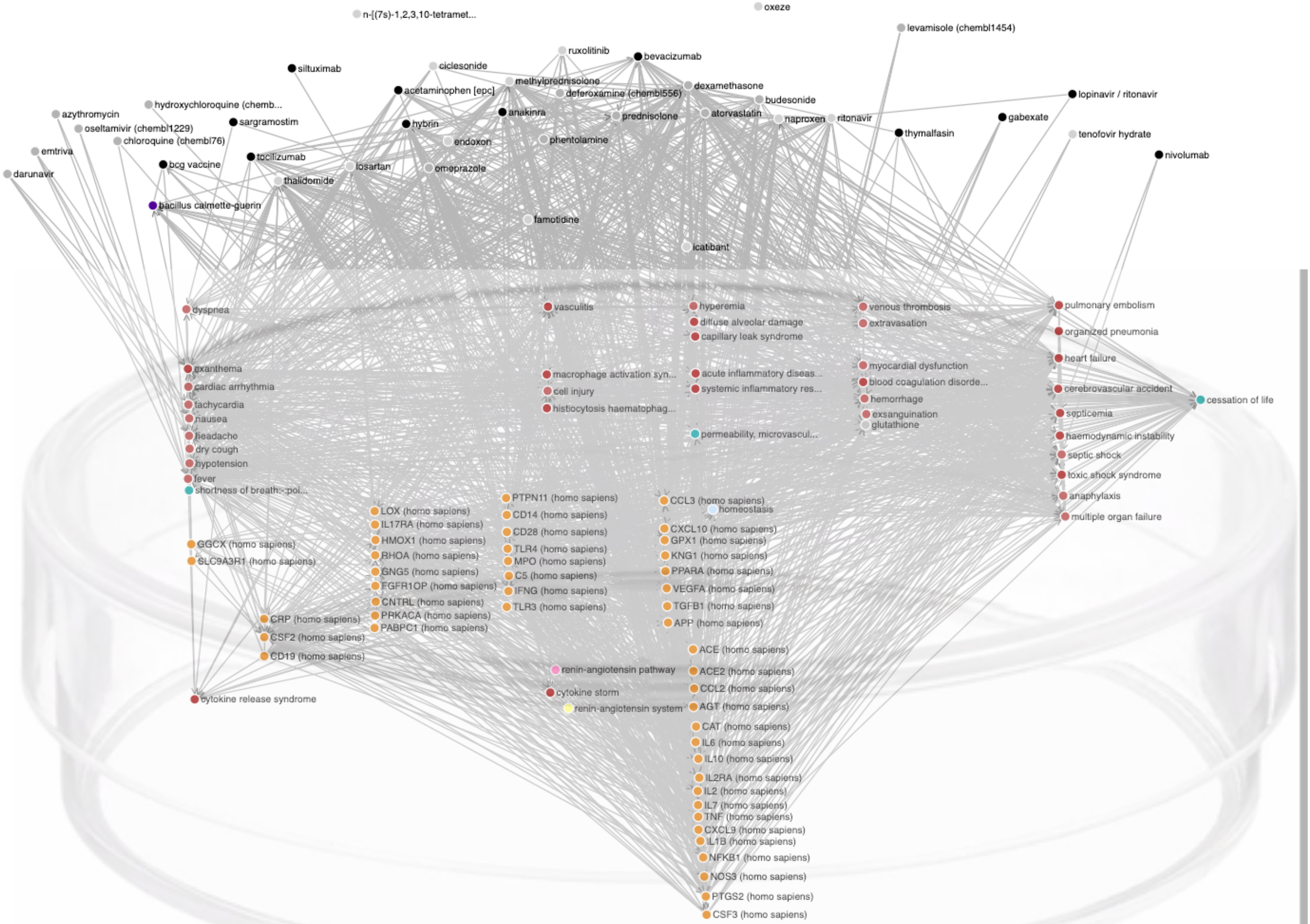
H

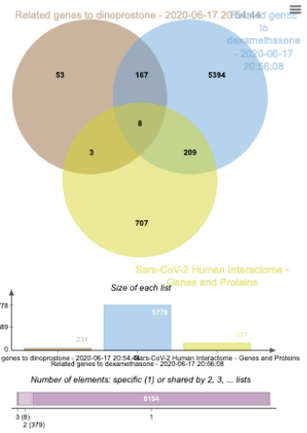
Community annotation options

Systematic *in silico* Rationalisation



I
II
III
IV
V
dyspnea
exanthema
cardiac arrhythmia
tachycardia
nausea
headache
dry cough
hypotension
fever
shortness of breath:-:poi...
GGCX (homo sapiens)
SLC9A3R1 (homo sapiens)
CRP (homo sapiens)
CSF2 (homo sapiens)
CD19 (homo sapiens)
cytokine release syndrome
vasculitis
macrophage activation syn...
cell injury
histiocytosis haematophag...
LOX (homo sapiens)
IL17RA (homo sapiens)
HMOX1 (homo sapiens)
RHOA (homo sapiens)
GNG5 (homo sapiens)
EGFR1OP (homo sapiens)
CNTRL (homo sapiens)
PRKACA (homo sapiens)
PABPC1 (homo sapiens)
PTPN11 (homo sapiens)
CD14 (homo sapiens)
CD28 (homo sapiens)
TLR4 (homo sapiens)
MPO (homo sapiens)
C5 (homo sapiens)
IFNG (homo sapiens)
TLR3 (homo sapiens)
renin-angiotensin pathway
cytokine storm
renin-angiotensin system
hyperemia
diffuse alveolar damage
capillary leak syndrome
acute inflammatory diseas...
systemic inflammatory res...
permeability, microvascul...
CCL3 (homo sapiens)
homeostasis
CXCL10 (homo sapiens)
GPX1 (homo sapiens)
KNG1 (homo sapiens)
PPARA (homo sapiens)
VEGFA (homo sapiens)
TGFB1 (homo sapiens)
APP (homo sapiens)
ACE (homo sapiens)
ACE2 (homo sapiens)
CCL2 (homo sapiens)
AGT (homo sapiens)
CAT (homo sapiens)
IL6 (homo sapiens)
IL10 (homo sapiens)
IL2RA (homo sapiens)
IL2 (homo sapiens)
IL7 (homo sapiens)
TNF (homo sapiens)
CXCL9 (homo sapiens)
IL1B (homo sapiens)
NFKB1 (homo sapiens)
NOS3 (homo sapiens)
PTGS2 (homo sapiens)
CSF3 (homo sapiens)
venous thrombosis
extravasation
myocardial dysfunction
blood coagulation disorde...
hemorrhage
exsanguination
glutathione
pulmonary embolism
organized pneumonia
heart failure
cerebrovascular accident
septicemia
haemodynamic instability
septic shock
toxic shock syndrome
anaphylaxis
multiple organ failure
cessation of life


n-[(7s)-1,2,3,10-tetramet...
oxeze
levamisole (chembl1454)
ruxolitinib
bevacizumab
siltuximab
ciclesonide
methylprednisolone
dexamethasone
acetaminophen [epc]
deferoxamine (chembl556)
budesonide
lopinavir / ritonavir
hydroxychloroquine (chemb...
azythromycin
anakinra
prednisolone
atorvastatin
naproxen
ritonavir
gabexate
sargramostim
oseltamivir (chembl1229)
hybrin
thymalfasin
tenofovir hydrate
chloroquine (chembl76)
endoxon
phentolamine
emtriva
nivolumab
tocilizumab
bcg vaccine
losartan
omeprazole
darunavir
thalidomide
bacillus calmette-guerin
famotidine
icatibant
dyspnea
vasculitis
hyperemia
diffuse alveolar damage
capillary leak syndrome
venous thrombosis
extravasation
pulmonary embolism
organized pneumonia
heart failure
exanthema
cardiac arrhythmia
tachycardia
nausea
headache
dry cough
hypotension
fever
shortness of breath:-poi...
macrophage activation syn...
cell injury
histiocytosis haematophag...
acute inflammatory diseas...
systemic inflammatory res...
permeability, microvascul...
myocardial dysfunction
blood coagulation disorde...
hemorrhage
exsanguination
glutathione
cerebrovascular accident
cessation of life
septicemia
haemodynamic instability
septic shock
toxic shock syndrome
anaphylaxis
multiple organ failure
LOX (homo sapiens)
IL17RA (homo sapiens)
HMOX1 (homo sapiens)
RHOA (homo sapiens)
GNG5 (homo sapiens)
FGFR1OP (homo sapiens)
CNTRL (homo sapiens)
PRKACA (homo sapiens)
PABPC1 (homo sapiens)
PTPN11 (homo sapiens)
CD14 (homo sapiens)
CD28 (homo sapiens)
TLR4 (homo sapiens)
MPO (homo sapiens)
C5 (homo sapiens)
IFNG (homo sapiens)
TLR3 (homo sapiens)
CCL3 (homo sapiens)
homeostasis
CXCL10 (homo sapiens)
GPX1 (homo sapiens)
KNG1 (homo sapiens)
PPARA (homo sapiens)
VEGFA (homo sapiens)
TGFB1 (homo sapiens)
APP (homo sapiens)
GGCX (homo sapiens)
SLC9A3R1 (homo sapiens)
CRP (homo sapiens)
CSF2 (homo sapiens)
CD19 (homo sapiens)
renin-angiotensin pathway
cytokine storm
renin-angiotensin system
ACE (homo sapiens)
ACE2 (homo sapiens)
CCL2 (homo sapiens)
AGT (homo sapiens)
CAT (homo sapiens)
IL6 (homo sapiens)
IL10 (homo sapiens)
IL2RA (homo sapiens)
IL2 (homo sapiens)
IL7 (homo sapiens)
TNF (homo sapiens)
CXCL9 (homo sapiens)
IL1B (homo sapiens)
NFKB1 (homo sapiens)
NOS3 (homo sapiens)
PTGS2 (homo sapiens)
CSF3 (homo sapiens)
cytokine release syndrome

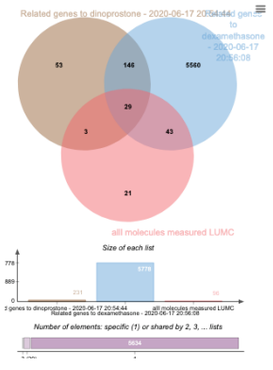


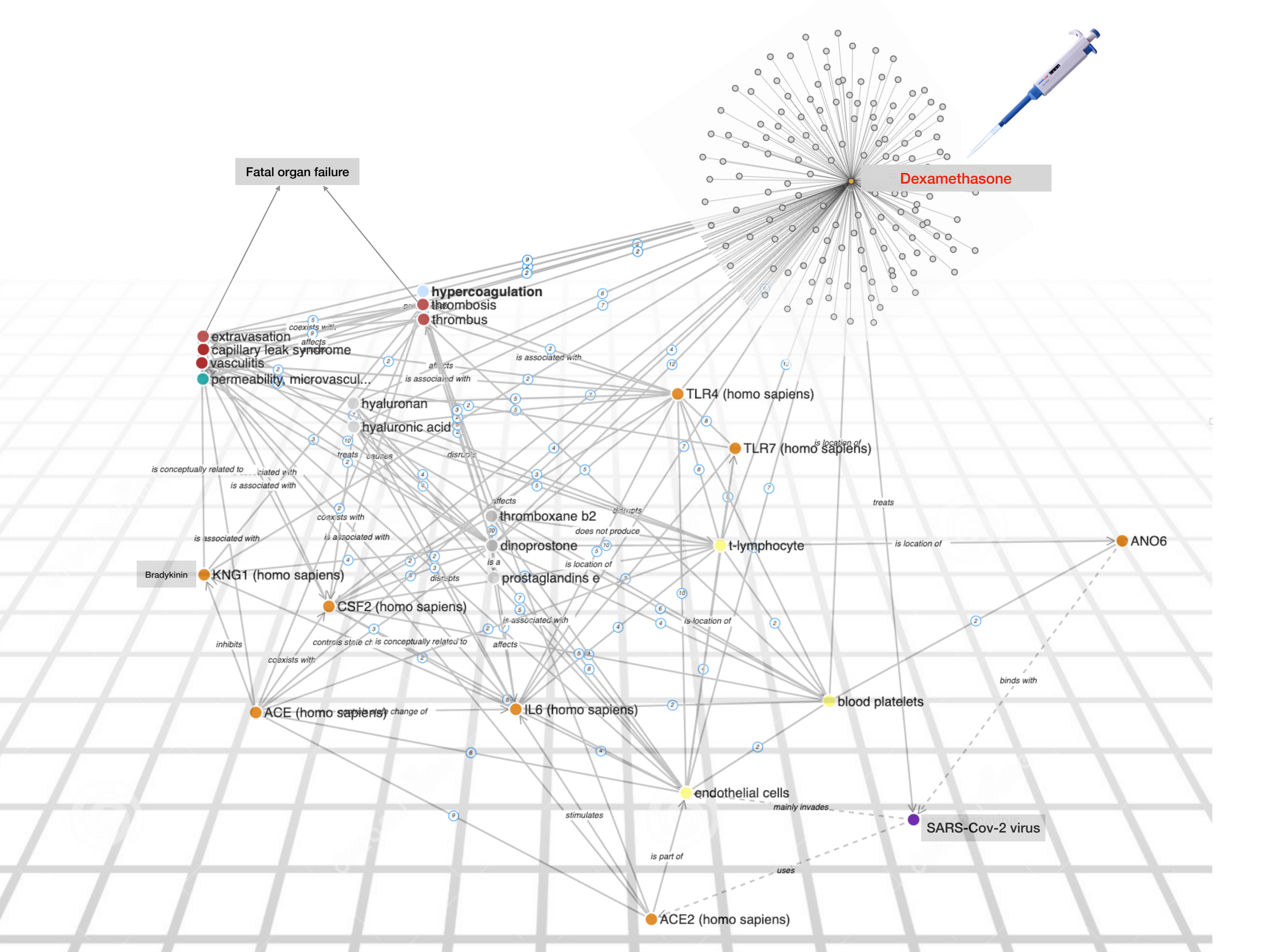
Cardinal assertion

EUROTOS AI PLATFORM

Provenance

Supporting or contesting Evidence





Fatal organ failure
Dexamethasone
hypercoagulation
thrombosis
thrombus
extravasation
capillary leak syndrome
vasculitis
permeability, microvascul...
hyaluronan
hyaluronic acid
TLR4 (homo sapiens)
TLR7 (homo sapiens)
thromboxane b2
dinoprostone
prostaglandins e
t-lymphocyte
ANO6
Bradykinin
KNG1 (homo sapiens)
CSF2 (homo sapiens)
ACE (homo sapiens)
IL6 (homo sapiens)
blood platelets
endothelial cells
SARS-Cov-2 virus
ACE2 (homo sapiens)
coexists with
affects
is associated with
treats
causes
disrupts
is conceptually related to
does not produce
is location of
inhibits
controls state change of
stimulates
mainly invades
binds with
is part of
uses

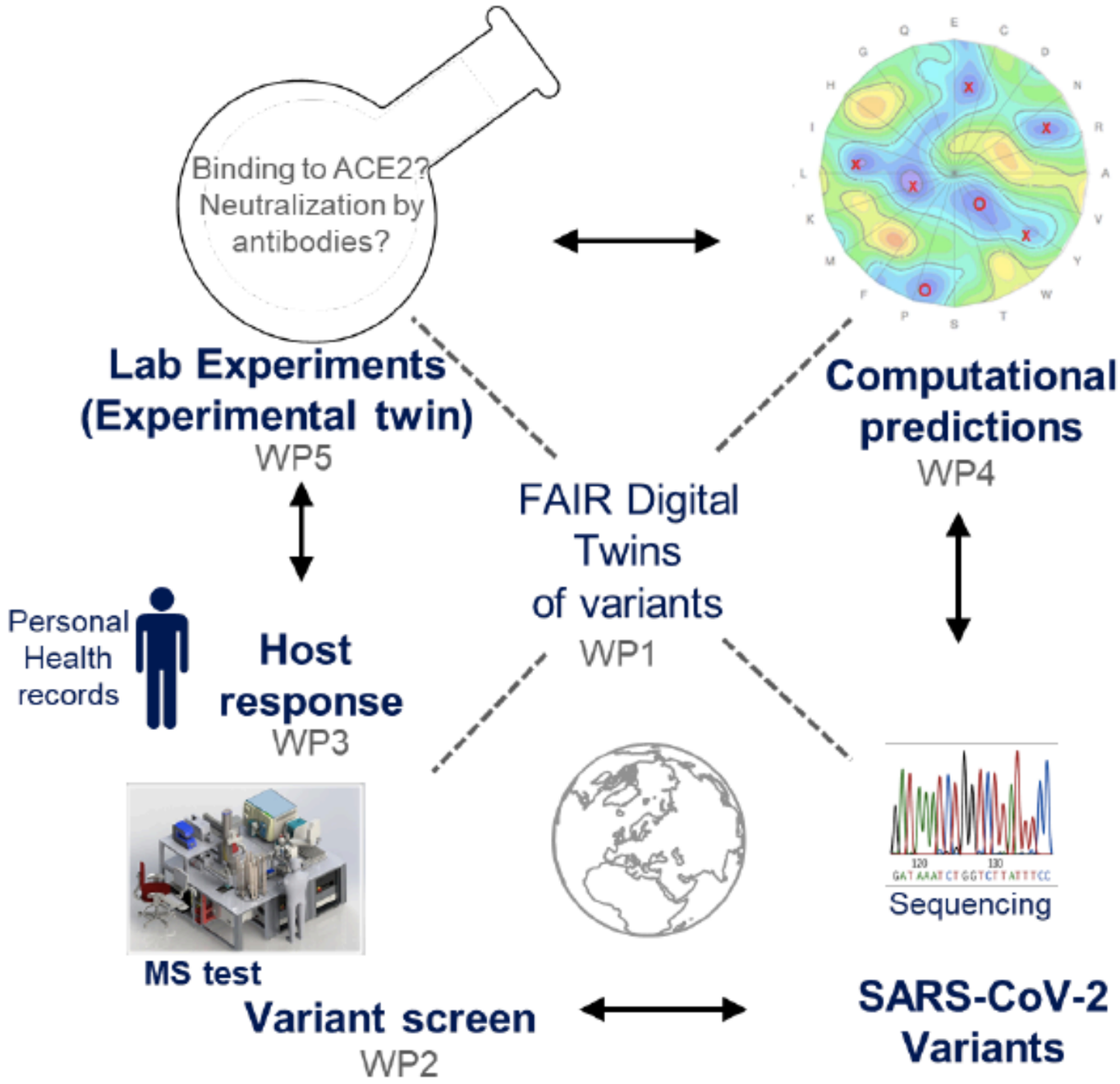
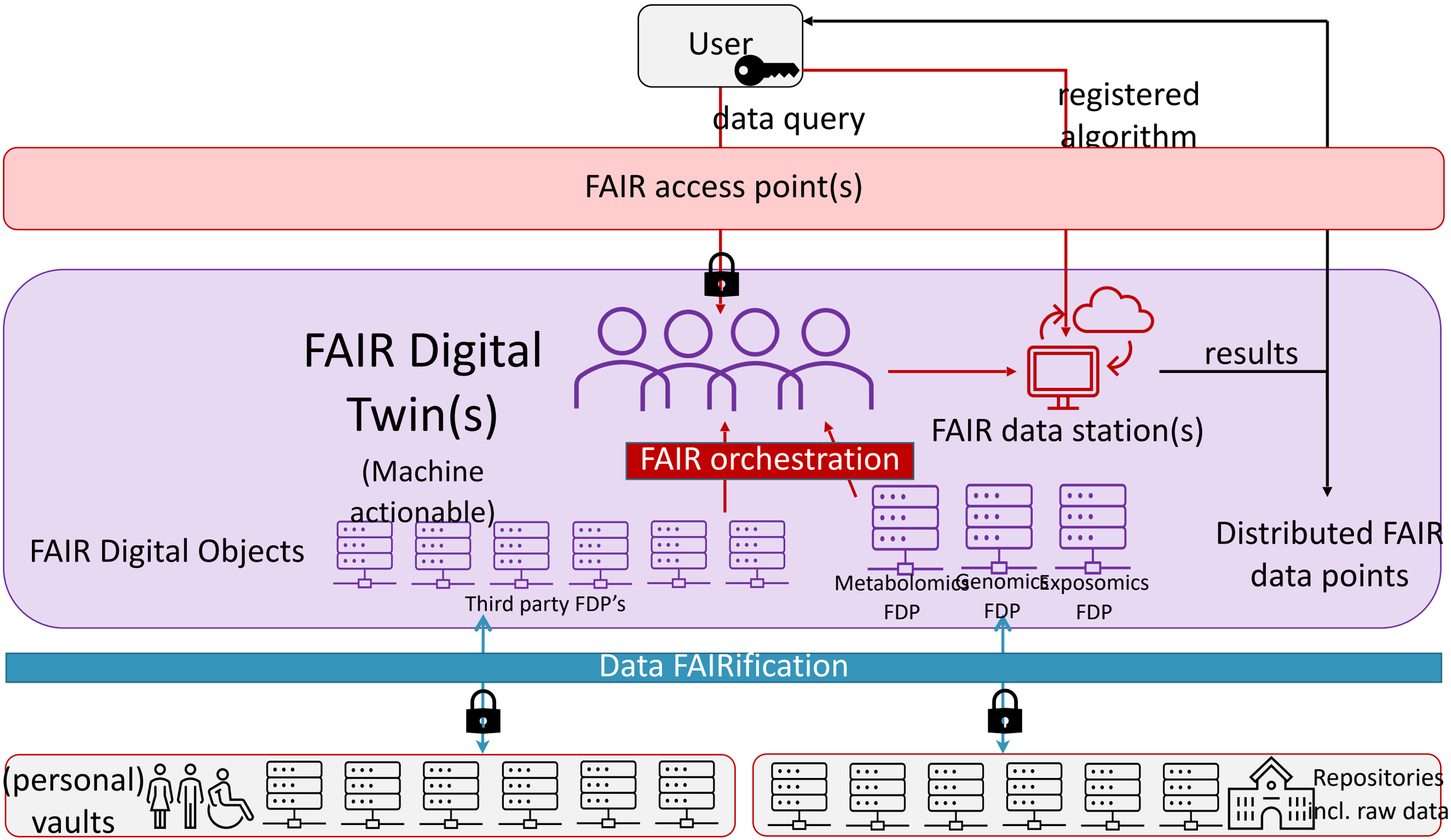


Figure 1. The essence of the STAYAHEAD approach: The combination of a rapid MS test to directly detect viral variants (WP2), as well as broader data on the host (WP3) with computational predictions of likely but still 'virtual' variants of high concern (WP4) would enable to stay ahead of the virus with early detection and pre-emptive vaccine strategies. The validation of predictions in vitro (WP5) traditional sequencing will be used to demonstrate the feasibility of this approach. All data stored as FAIR digital Twins will be made available for a wide range of AI applications (WP1)


User
data query
registered algorithm
FAIR access point(s)
FAIR Digital Twin(s)
(Machine actionable)
FAIR orchestration
FAIR data station(s)
results
FAIR Digital Objects
Third party FDP's
Metabolomics FDP
Genomics FDP
Exposomics FDP
Distributed FAIR data points
Data FAIRification
(personal) vaults
Repositories incl. raw data