NIST/ASM AM Data Management Workshop

FAIR Data Stewardship for FAIR Digital Twins

Erik Schultes, PhD
International Science Coordinator
GO FAIR International Support and Coordination Office
Leiden Center for Data Science

erik.schultes@go-fair.org
https://www.go-fair.org
http://orcid.org/0000-0001-8888-635X

27 October 2020
Slides: https://osf.io/m5x8q/

Automating F, A, I, and R

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

Technical infrastructure (generic operations)
Data/metadata (domain-specific content)

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

Technical infrastructure (generic operations)
Data/metadata (domain-specific content)

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

Technical infrastructure (generic operations)
Data/metadata (domain-specific content)

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 i… …col
A1.1 the protocol is open, free, and univ…
A1.2 the protocol allows for an authenti…
A2. metadata are accessible, even when…

To be Interoperable:
I1. (meta)data use a formal, accessible, … …ge representation.
I2. (meta)data use vocabularies that follo…
I3. (meta)data include qualified referenc…

To be Reusable:
R1. meta(data) are richly described with…
R1.1. (meta)data are released with a cle… …data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

"Collection"-centric Evaluations

- Communities decide which Tests are relevant to them
- These are registered in the Evaluator as a "Collection"
  - Documentation about what Tests are included, and to what communities the Collection would be relevant
- Anyone can execute an evaluation on any Identifier
- Anyone can select any Collection they wish to apply
  - For example, journals may select different evaluation collections than funding agencies, or researchers
- An "Evaluation", therefore, is the application of an identified collection of Tests tests to a given resource of interest.

FAIR Data Systems

Technical infrastructure (generic operations)
Data/metadata (domain-specific content)

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 i… …col
A1.1 the protocol is open, free, and univ…
A1.2 the protocol allows for an authenti…
A2. metadata are accessible, even when…

To be Interoperable:

I1. (meta)data use a formal, accessible, …
I2. (meta)data use vocabularies that follo…
I3. (meta)data include qualified referenc…

To be Reusable:

R1. meta(data) are richly described with …
R1.1. (meta)data are released with a cle… …
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

"Collection"-centric Evaluations

- Communities decide which Tests are relevant to them
- These are registered in the Evaluator as a "Collection"
  - Documentation about what… communities the Collection…
- Anyone can execute an evalua…
- Anyone can select any Collecti…
  - For example, journals may s… than funding agencies, or re…
- An "Evaluation", therefore, is th… of Tests tests to a given resourc…

FAIR Data Systems

Information Technology Laboratory / Software and Systems Division

INFORMATION SYSTEMS GROUP

The Configurable Data Curation System (CDCS)

- About the CDCS
- CDCS Downloads
- CDCS Help and Resources +
- Contact CDCS
- CDCS Development Team

The Configurable Data Curation System (CDCS)

The CDCS represents a platform through which NIST and related communities have begun to mutually engage in discussion, development, and problem-solving. Driven by FAIR data principles, this has given rise to several activities throughout NIST, including:

- Community standards: To increase the availability and quality of community standards, NIST hosted workshops focused on community data model development.
- Interoperability: To increase integration among data platforms (for materials science and beyond), NIST hosted a hackathon focused on such integrations.
- Community development: To support the development of FAIR data communities (in materials science and elsewhere), NIST began an annual convention for growing and nurturing the CDCS community of users.

Technical infrastructure (generic operations)
Data/metadata (domain-specific content)

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:

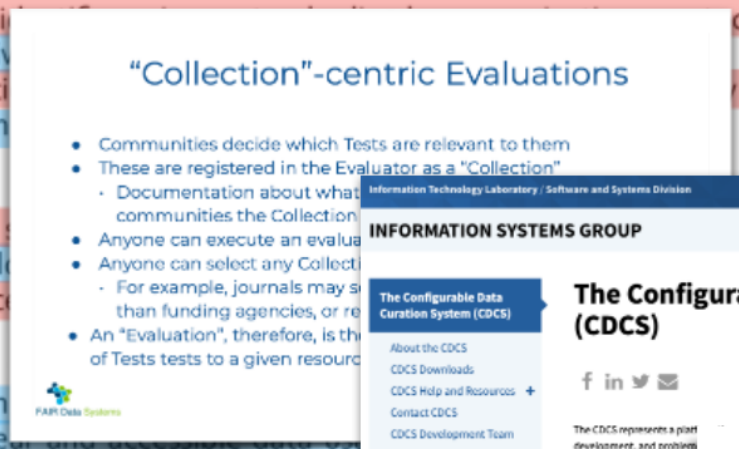
A1. (meta)data are retrievable by their i…col
A1.1 the protocol is open, free, and univ…
A1.2 the protocol allows for an authenti…
A2. metadata are accessible, even when…

To be Interoperable:

I1. (meta)data use a formal, accessible, …
I2. (meta)data use vocabularies that follo…
I3. (meta)data include qualified referenc…

To be Reusable:

R1. meta(data) are richly described with…
R1.1. (meta)data are released with a cle…
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards



AM BENCH 2022

FAIR Principles → FAIR Implementations

How to GO FAIR

https://www.go-fair.org/how-to-go-fair/

Home › How to GO FAIR

How to GO FAIR

Since its beginning in early 2018, the GO FAIR community has been working towards implementations of the **FAIR Guiding Principles**. This collective effort has resulted in a three-point framework that formulates the essential steps towards the end goal, a global Internet of FAIR Data and Services where data are **F**indable, **A**ccessible, **I**nteroperable and **R**eusable (**FAIR**) for machines.

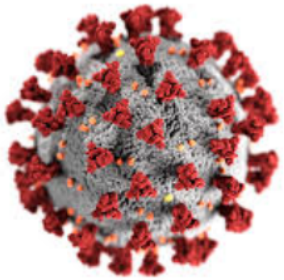
A framework guiding FAIRification

The Three-point FAIRification Framework provides practical "how to" guidance to stakeholders seeking to go FAIR.

Moreover, by following this framework, stakeholders can rest assured that their efforts toward FAIRification will be optimally coordinated with the efforts of other stakeholders in the GO FAIR community. The three-point framework maximizes reuse of existing resources, maximizes interoperability, and accelerates convergence on standards and technologies supporting FAIR data and services.

- Typically, the FAIRification process begins when a community of practice considers its domain-relevant metadata requirements and other policy considerations, and formulates these considerations as machine-actionable metadata components. These considerations can be guided in **Metadata for Machines** (M4M) Workshops.

FAIR Principles → FAIR Implementations



How to GO FAIR

https://www.go-fair.org/how-to-go-fair/

Home › How to GO FAIR

How to GO FAIR

Since its beginning in early 2018, the GO FAIR community has been working towards implementations of the **FAIR Guiding Principles**. This collective effort has resulted in a three-point framework that formulates the essential steps towards the end goal, a global Internet of FAIR Data and Services where data are **F**indable, **A**ccessible, **I**nteroperable and **R**eusable (**FAIR**) for machines.

A framework guiding FAIRification

The Three-point FAIRification Framework provides practical "how to" guidance to stakeholders seeking to go FAIR.

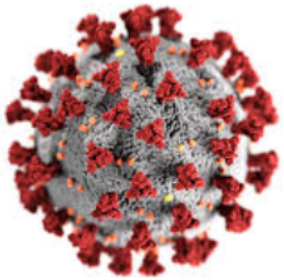
Moreover, by following this framework, stakeholders can rest assured that their efforts toward FAIRification will be optimally coordinated with the efforts of other stakeholders in the GO FAIR community. The three-point framework maximizes reuse of existing resources, maximizes interoperability, and accelerates convergence on standards and technologies supporting FAIR data and services.

- Typically, the FAIRification process begins when a community of practice considers its domain-relevant metadata requirements and other policy considerations, and formulates these considerations as machine-actionable metadata components. These considerations can be guided in **Metadata for Machines** (M4M) Workshops.

Philips Foundation

ZonMw

FAIR Principles → FAIR Implementations



How to GO FAIR

https://www.go-fair.org/how-to-go-fair/

Home › How to GO FAIR

How to GO FAIR

Since its beginning in early 2018, the GO FAIR community has been working towards implementations of the **FAIR Guiding Principles**. This collective effort has resulted in a three-point framework that formulates the essential steps towards the end goal, a global Internet of FAIR Data and Services where data are **F**indable, **A**ccessible, **I**nteroperable and **R**eusable (**FAIR**) for machines.

A framework guiding FAIRification

The Three-point FAIRification Framework provides practical "how to" guidance to stakeholders seeking to go FAIR.

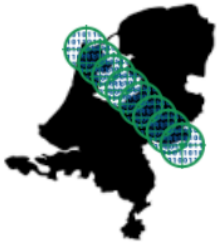
Moreover, by following this framework, stakeholders can rest assured that their efforts toward FAIRification will be optimally coordinated with the efforts of other stakeholders in the GO FAIR community. The three-point framework maximizes reuse of existing resources, maximizes interoperability, and accelerates convergence on standards and technologies supporting FAIR data and services.

- Typically, the FAIRification process begins when a community of practice considers its domain-relevant metadata requirements and other policy considerations, and formulates these considerations as machine-actionable metadata components. These considerations can be guided in **Metadata for Machines** (M4M) Workshops.

Philips Foundation

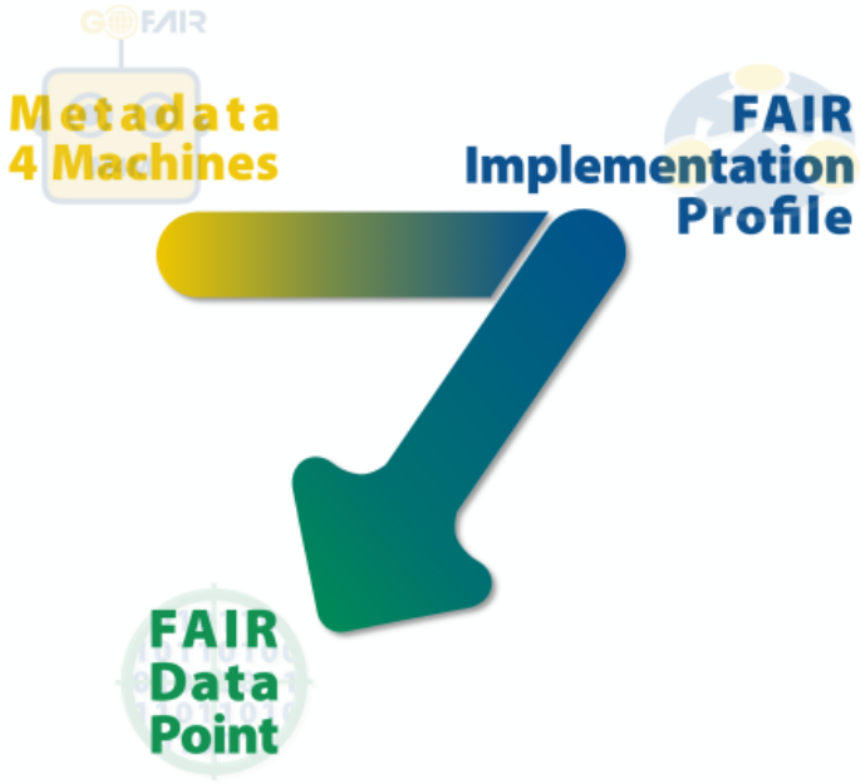
ZonMw

https://www.vodan-totafrica.info/special-news/

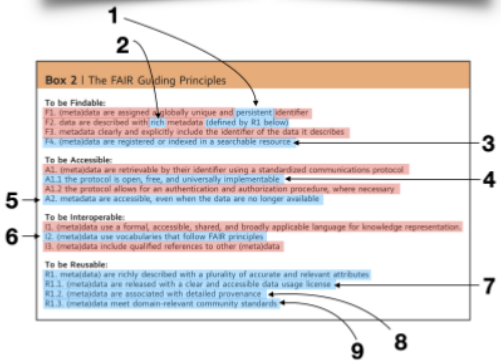


https://www.zonmw.nl/en/research-and-results/fair-data-and-data-management/open-science-in-covid-19-research/

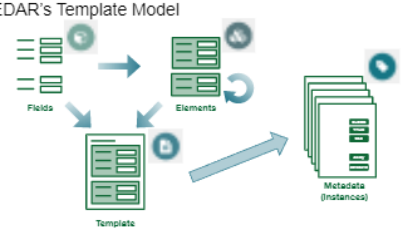
Three-Point FAIRification Framework



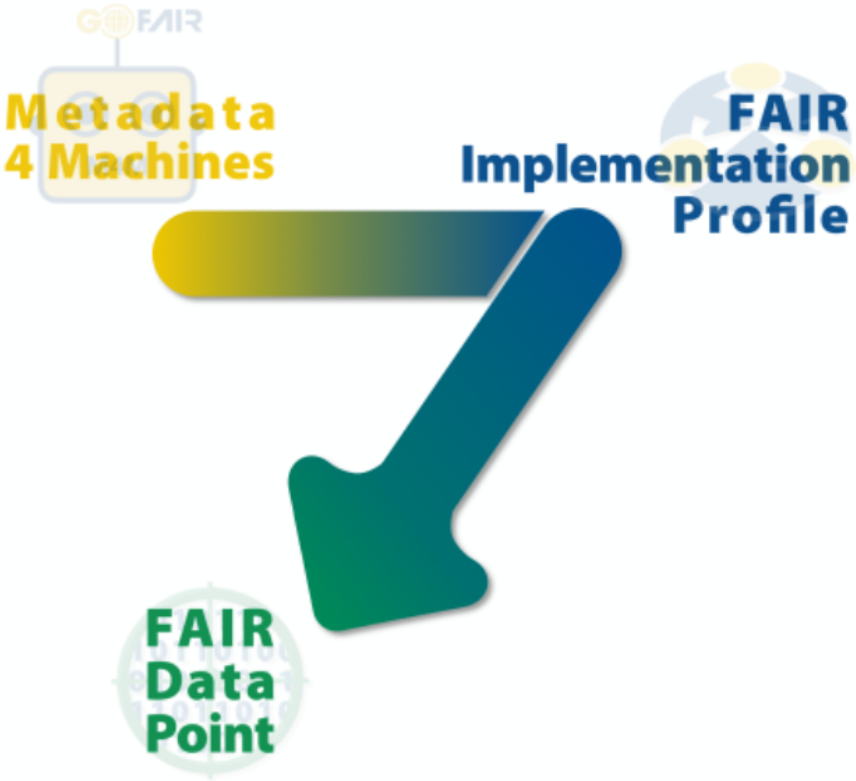
Three-Point FAIRification Framework



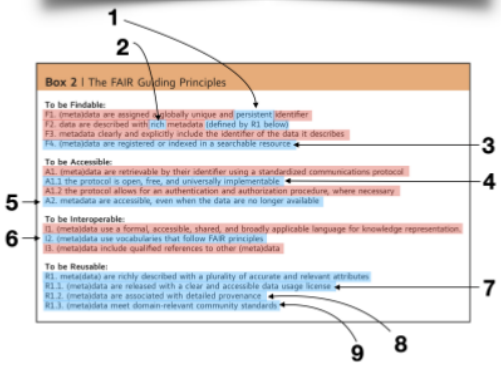
CEDAR's Template Model

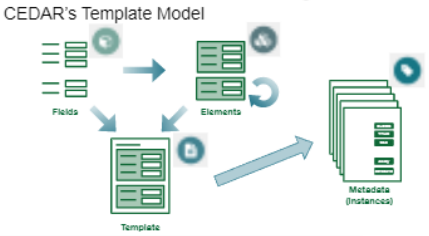


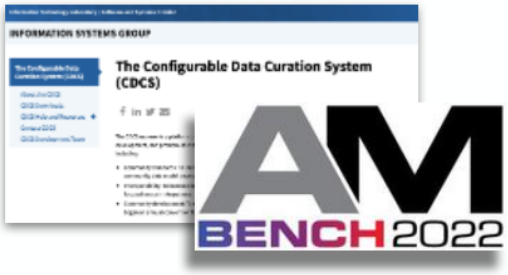
https://www.go-fair.org/today/making-fair-metadata/

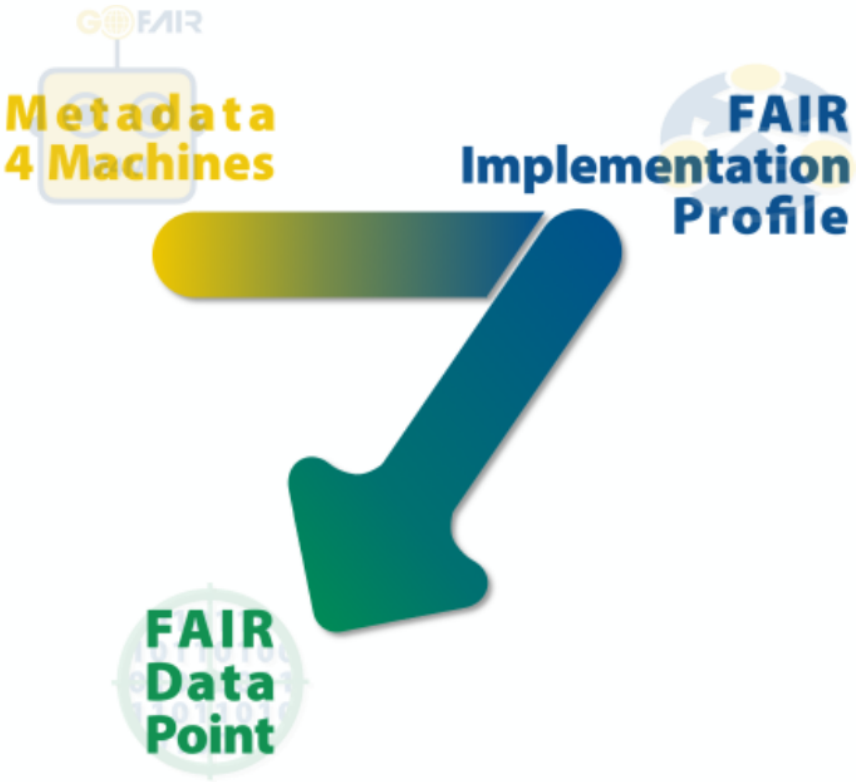


Three-Point FAIRification Framework

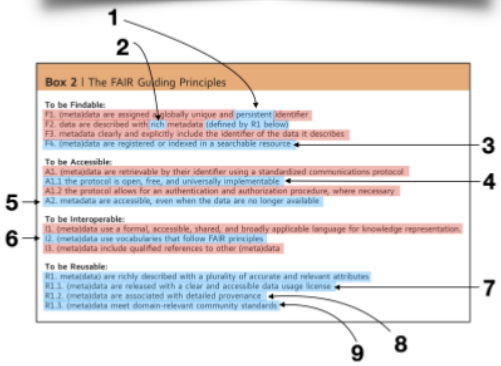




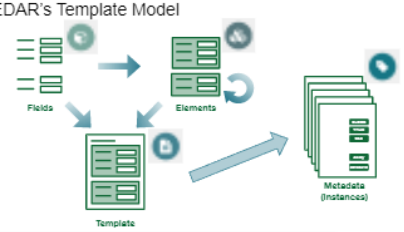


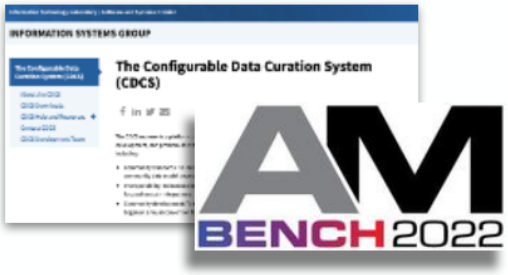


Three-Point FAIRification Framework



CEDAR's Template Model

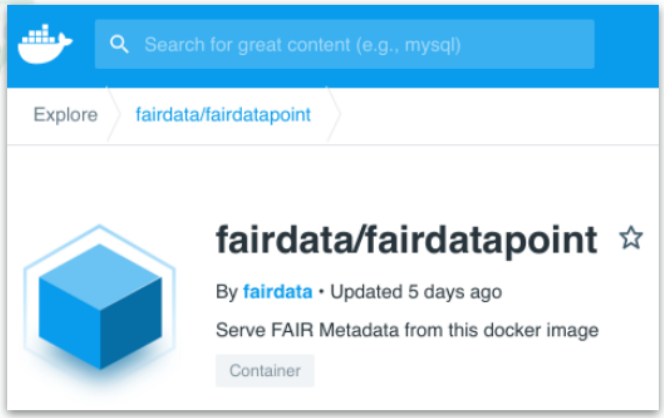


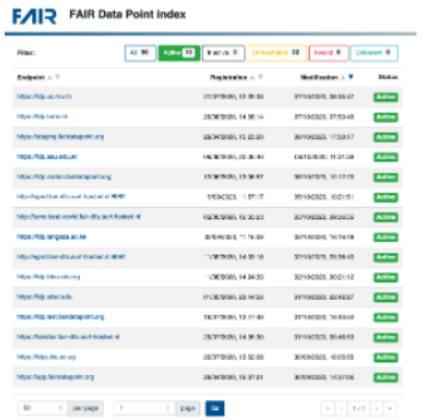


Metadata 4 Machines

FAIR Implementation Profile

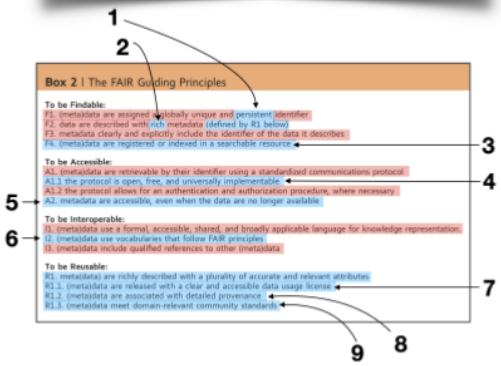
FAIR Data Point



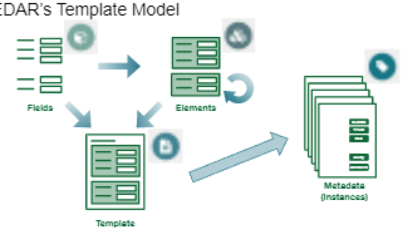


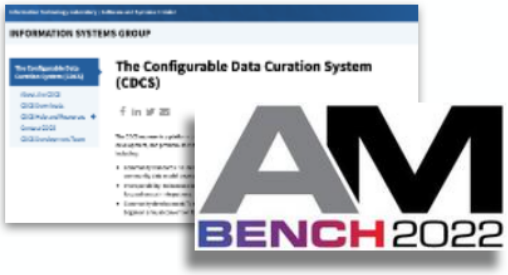
https://home.fairdatapoint.org

Three-Point FAIRification Framework



CEDAR's Template Model

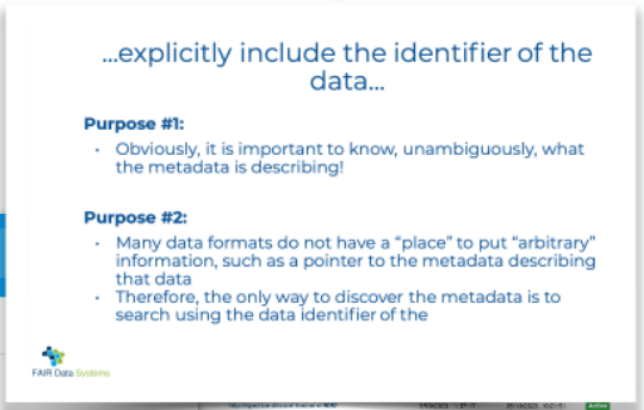


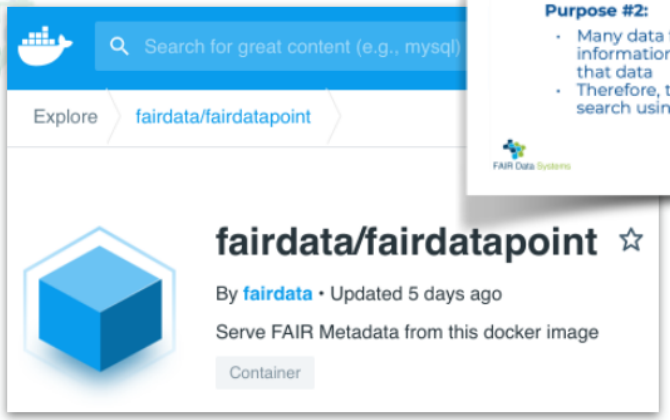


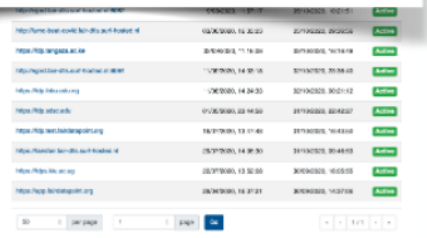
Metadata 4 Machines

FAIR Implementation Profile

FAIR Data Point

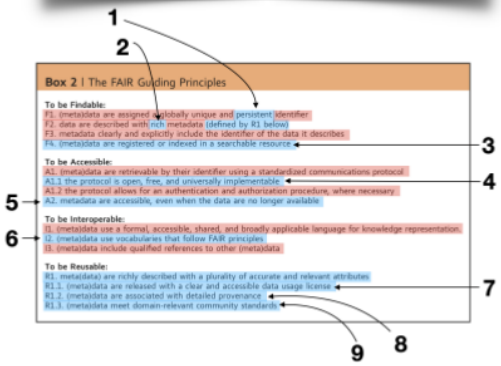




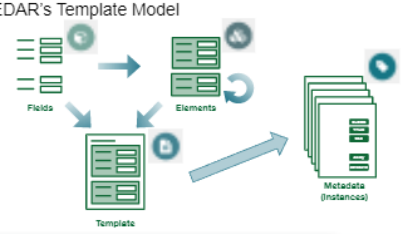


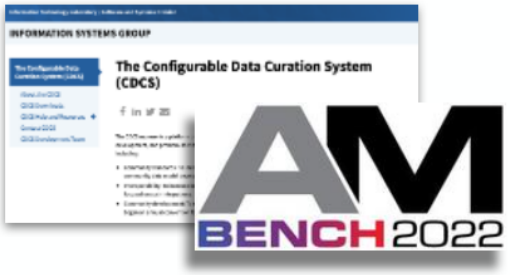
https://home.fairdatapoint.org

Three-Point FAIRification Framework



CEDAR's Template Model

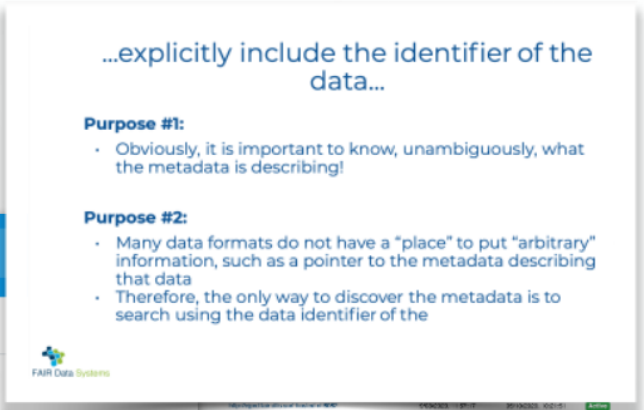


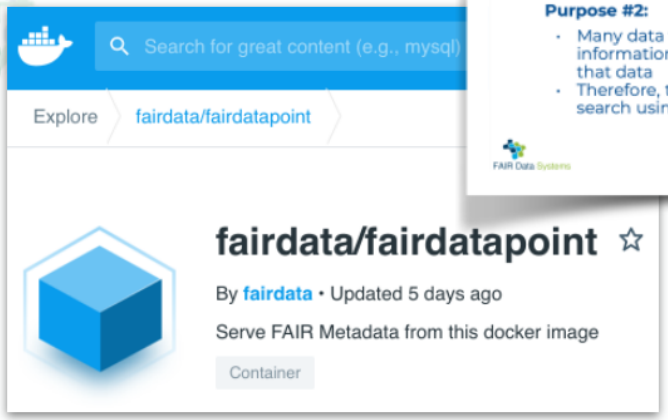


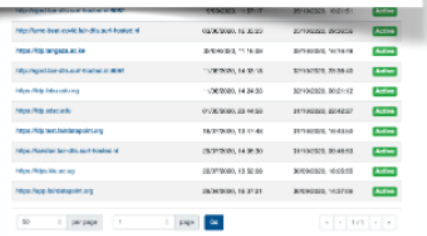
Metadata 4 Machines

FAIR Implementation Profile

FAIR Data Point







https://home.fairdatapoint.org

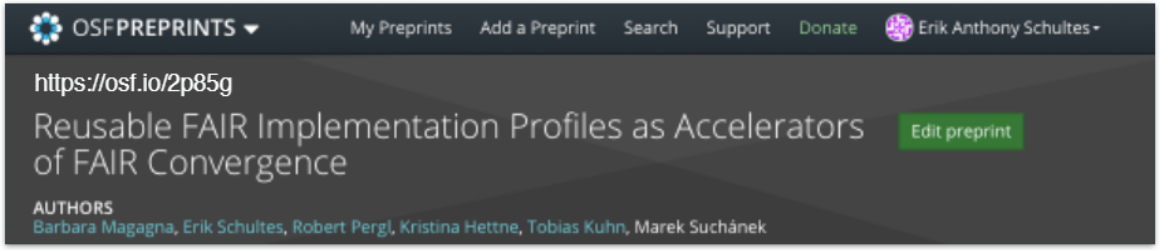
FAIR Implementation Profile

This is a template. Please download or copy to complete your FIP.

Community description	
Name of Community	*e.g. ENVRI*
Description of Community	
Supporting Links	
Research Domain	*e.g. Environmental Sciences*
Data Steward	*e.g. ORCID #*
Date of FIP creation	

FAIR principle	Question	FAIR enabling resource types	Your answers
F1	What globally unique, persistent, resolvable identifiers do you use for metadata records?	Identifier type	*e.g. PURL, DOI*
F1	What globally unique, persistent, resolvable identifiers do you use for datasets?	Identifier type	
F2	Which metadata schemas do you use for findability?	Metadata schema	
F3	What is the technology that links the persistent identifiers of your data to the metadata description?	Metadata-Data linking mechanism	
F4	In which search engines are your metadata records indexed?	Search engines	
F4	In which search engines are your datasets indexed?	Search engines	
A1.1	Which standardized communication protocol do you use for metadata records?	Communication protocol	
A1.1	Which standardized communication protocol do you use for datasets?	Communication protocol	
A1.2	Which authentication & authorisation technique do you use for metadata records?	Authentication & authorisation technique	
A1.2	Which authentication & authorisation technique do you use for datasets?	Authentication & authorisation technique	
A2	Which metadata longevity plan do you use?	Metadata longevity	
I1	Which knowledge representation languages (allowing machine interoperation) do you use for metadata records?	Knowledge representation language	
I1	Which knowledge representation languages (allowing machine interoperation) do you use for datasets?	Knowledge representation language	
I2	Which structured vocabularies do you use to annotate your metadata records?	Structured vocabularies	
I2	Which structured vocabularies do you use to encode your datasets?	Structured vocabularies	
I3	Which models, schema(s) do you use for your metadata records?	Metadata schema	
I3	Which models, schema(s) do you use for your datasets?	Data schema	
R1.1	Which usage license do you use for your metadata records?	Data usage license	
R1.1	Which usage license do you use for your datasets?	Data usage license	
R1.2	Which metadata schemas do you use for describing the provenance of your metadata records?	Provenance model	
R1.2	Which metadata schemas do you use for describing the provenance of your datasets?	Provenance model	

http://bit.ly/FIPminiquestionnaire



FAIR Implementation Profile

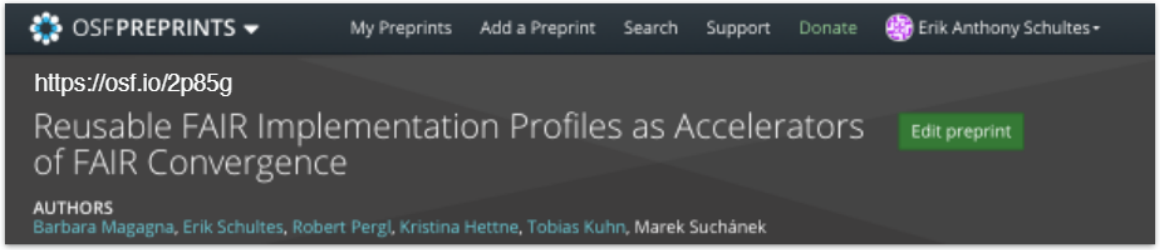
This is a template. Please download or copy to complete your FIP.

Community description	
Name of Community	*e.g. ENVRI*
Description of Community	
Supporting Links	
Research Domain	*e.g. Environmental Sciences*
Data Steward	*e.g. ORCID #*
Date of FIP creation	

FAIR principle	Question	FAIR enabling resource types	Your answers
F1	What globally unique, persistent, resolvable identifiers do you use for metadata records?	Identifier type	*e.g. PURL, DOI*
F1	What globally unique, persistent, resolvable identifiers do you use for datasets?	Identifier type	
F2	Which metadata schemas do you use for findability?	Metadata schema	
F3	What is the technology that links the persistent identifiers of your data to the metadata description?	Metadata-Data linking mechanism	
F4	In which search engines are your metadata records indexed?	Search engines	
F4	In which search engines are your datasets indexed?	Search engines	
A1.1	Which standardized communication protocol do you use for metadata records?	Communication protocol	
A1.1	Which standardized communication protocol do you use for datasets?	Communication protocol	
A1.2	Which authentication & authorisation technique do you use for metadata records?	Authentication & authorisation technique	
A1.2	Which authentication & authorisation technique do you use for datasets?	Authentication & authorisation technique	
A2	Which metadata longevity plan do you use?	Metadata longevity	
I1	Which knowledge representation languages (allowing machine interoperation) do you use for metadata records?	Knowledge representation language	
I1	Which knowledge representation languages (allowing machine interoperation) do you use for datasets?	Knowledge representation language	
I2	Which structured vocabularies do you use to annotate your metadata records?	Structured vocabularies	
I2	Which structured vocabularies do you use to encode your datasets?	Structured vocabularies	
I3	Which models, schema(s) do you use for your metadata records?	Metadata schema	
I3	Which models, schema(s) do you use for your datasets?	Data schema	
R1.1	Which usage license do you use for your metadata records?	Data usage license	
R1.1	Which usage license do you use for your datasets?	Data usage license	
R1.2	Which metadata schemas do you use for describing the provenance of your metadata records?	Provenance model	
R1.2	Which metadata schemas do you use for describing the provenance of your datasets?	Provenance model	

R1.3?

http://bit.ly/FIPminiquestionnaire

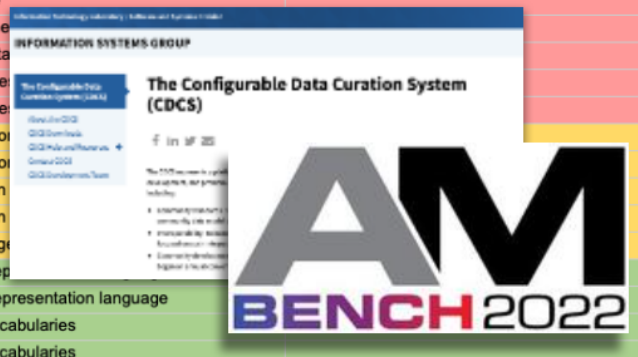


FAIR Implementation Profile

This is a template. Please download or copy to complete your FIP.

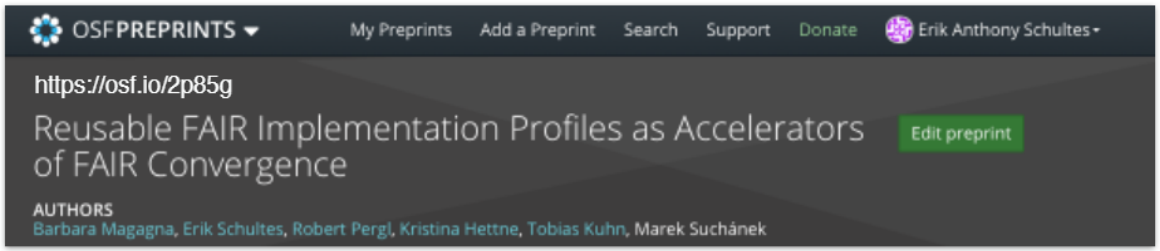
Community description	
Name of Community	*e.g. ENVRI*
Description of Community	
Supporting Links	
Research Domain	*e.g. Environmental Sciences*
Data Steward	*e.g. ORCID #*
Date of FIP creation	

FAIR principle	Question	FAIR enabling resource types	Your answers
F1	What globally unique, persistent, resolvable identifiers do you use for metadata records?	Identifier type	*e.g. PURL, DOI*
F1	What globally unique, persistent, resolvable identifiers do you use for datasets?	Identifier type	
F2	Which metadata schemas do you use for findability?	Metadata sche	
F3	What is the technology that links the persistent identifiers of your data to the metadata description?	Metadata-Data	
F4	In which search engines are your metadata records indexed?	Search engine	
F4	In which search engines are your datasets indexed?	Search engine	
A1.1	Which standardized communication protocol do you use for metadata records?	Communicatio	
A1.1	Which standardized communication protocol do you use for datasets?	Communicatio	
A1.2	Which authentication & authorisation technique do you use for metadata records?	Authentication	
A1.2	Which authentication & authorisation technique do you use for datasets?	Authentication	
A2	Which metadata longevity plan do you use?	Metadata longe	
I1	Which knowledge representation languages (allowing machine interoperation) do you use for metadata records?	Knowledge rep	
I1	Which knowledge representation languages (allowing machine interoperation) do you use for datasets?	Knowledge representation language	
I2	Which structured vocabularies do you use to annotate your metadata records?	Structured vocabularies	
I2	Which structured vocabularies do you use to encode your datasets?	Structured vocabularies	
I3	Which models, schema(s) do you use for your metadata records?	Metadata schema	
I3	Which models, schema(s) do you use for your datasets?	Data schema	
R1.1	Which usage license do you use for your metadata records?	Data usage license	
R1.1	Which usage license do you use for your datasets?	Data usage license	
R1.2	Which metadata schemas do you use for describing the provenance of your metadata records?	Provenance model	
R1.2	Which metadata schemas do you use for describing the provenance of your datasets?	Provenance model	



R1.3?

http://bit.ly/FIPminiquestionnaire

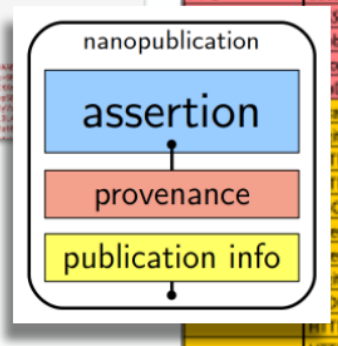


FAIR Implementation Profile

VODAN FIP

https://osf.io/pcx28/

https://osf.io/4zwg5/

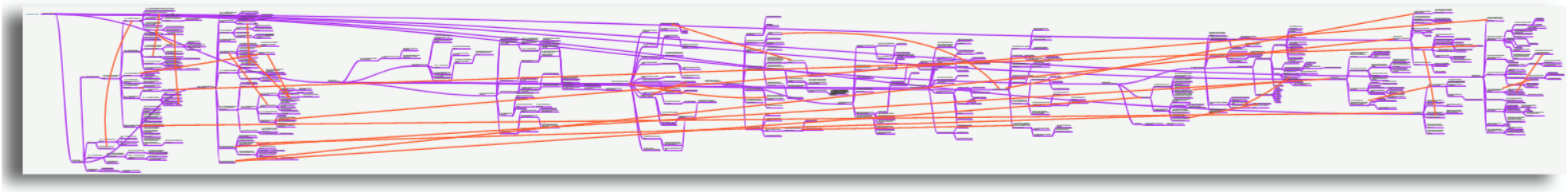


DSW
DATA STEWARDSHIP WIZARD

https://docs.google.com/spreadsheets/d/1jXMwSN0OYJQbtj1iAWGxdWeLCmmCCpLm8Ee5TcyyAfs/edit#gid=340018800

		environmental sciences		oceanography		health sciences	
FAIR Principle	FAIR Enabling Resource	DiSSCo	GBIF	SeaDataNet-CDI	SeaDataNet-Sextant	VODAN	HPA
F1-MD	DOI	1	1	0	1	0	1
	Natural Science Identifier	2	0	0	0	0	0
	Persistent Identifier for eResearch	1	0	0	0	0	2
	Persistent Uniform Resource Locator	0	0	0	0	1	0
	SeaDataNET CDI Global unique identifier	0	0	1	0	0	0
F1-D	Digital Object Identifier	0	1	0	0	0	0
	DOI	0	0	0	1	0	0
	Handle	0	0	1	0	0	2
	Natural Science Identifier	2	0	0	0	0	0
	Persistent Identifier for eResearch	2	0	0	0	0	1
	Persistent Uniform Resource Locator	0	0	0	0	1	2
F2	Darwin Core	1	0	0	0	0	0
	DataCite	0	0	0	0	0	0
	DCAT	0	0	0	0	1	2
	EML GBIF Profile	0	1	0	0	0	0
	ISO 19115	0	0	0	1	0	0
	Open Digital Specimens	2	0	0	0	0	0
	SeaDataNet CDI metadata XML schema	0	0	1	0	0	0
F3	Darwin-core Archive	0	1	0	0	0	0
	Digital Object Identifier	0	0	0	1	0	0
	FAIR Data Point	0	0	0	0	1	2
	Fair Digital Object	2	0	0	0	0	0
	SeaDataNet CDI to PID lookup index	0	0	1	0	0	0
F4-MD	automatic FDP call home registry	0	0	0	0	1	2
	DiSSCo European Collection Objects Index	2	0	0	0	0	0
	Global Biodiversity Information Facility Search Engine	0	1	0	0	0	0
	Global Earth Observation System of Systems	0	0	1	0	0	0
	Google	0	0	0	0	0	0
	Google Dataset Search	0	0	1	0	0	0
	SeaDataNet CDI search user interface	0	0	1	0	0	0
	SeaDataNet Sextant search engines	0	0	0	1	0	0
F4-D	automatic FDP call home registry	0	0	0	0	1	2
	SCo European Collection Objects Index	2	0	0	0	0	0
	bal Biodiversity Information Facility Search Engine	0	1	0	0	0	0
	gle Data Search	0	0	0	0	0	0
	DataNet Sextant search engines	0	0	0	1	0	0
	alog Service for the Web	0	0	0	1	0	0
	ital Object Interface Protocol	1	0	0	0	0	0
	TP REST	1	0	0	0	0	0
	TPS	0	1	0	0	1	1
	C Catalog Services	0	0	1	0	0	0
	en Archives Initiative Protocol for Metadata Harvesting	0	0	1	0	0	0
	enSearch	0	0	1	0	0	0
	ital Object Interface Protocol	1	0	0	0	0	0
	OAP data server	0	0	0	1	0	0
	TP REST	1	0	0	0	0	0
	HTTPS	0	1	0	0	1	1
A1.2-MD	basic access authentication	0	1	0	0	0	0
	DiSSCo Federated Authentication and Authorization Infrastructure	2	0	0	0	0	0
	GBIF.org Authentication technique	0	1	0	0	0	0
	Open access	0	0	0	1	0	0
	Open Data	0	0	0	0	1	1
	SeaDataNet Marine ID AAA service	0	0	1	0	0	0

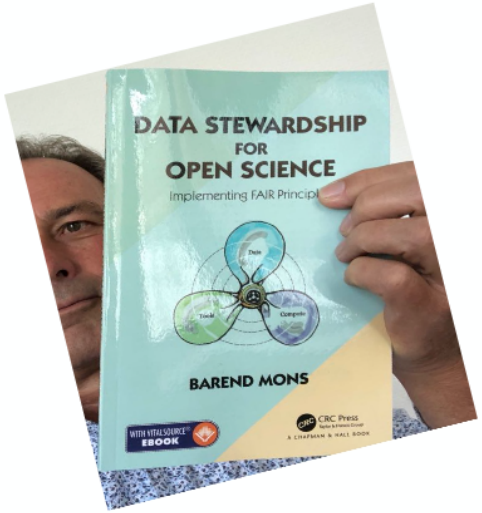
FAIR Data Stewardship



Rob Hooft

DTL

DUTCH TECHCENTRE FOR LIFE SCIENCES



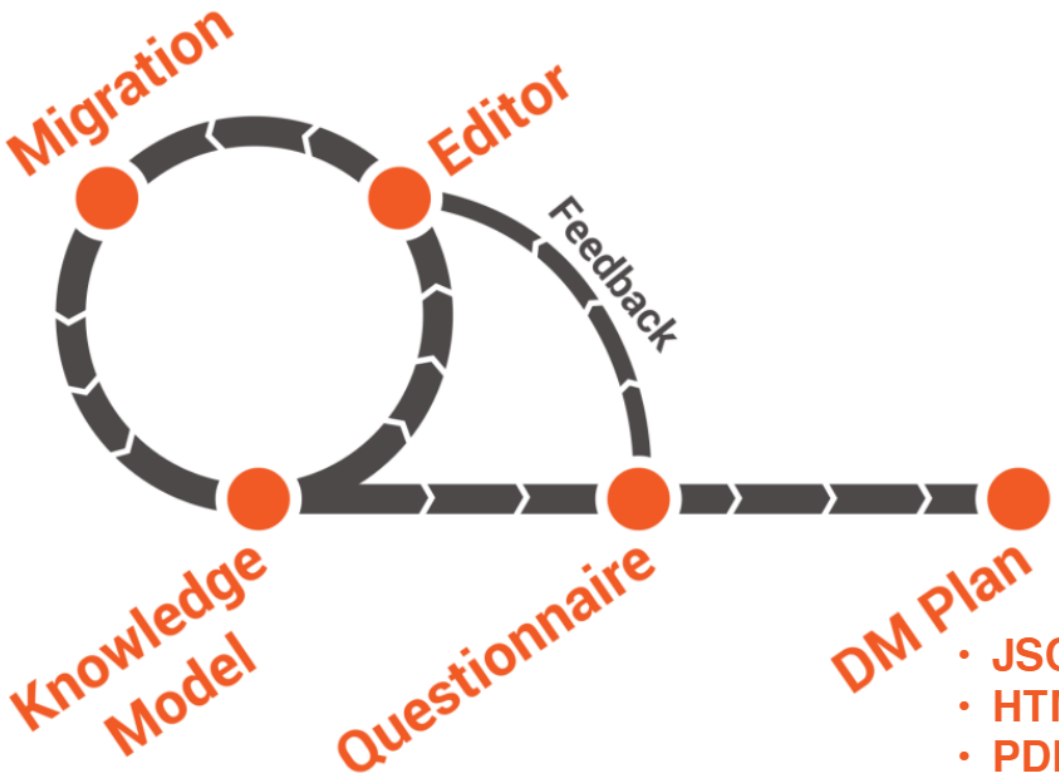
Data Cycle Step 1: Design of Experiment
Data Cycle Step 2: Data Design and Planning
Data Cycle Step 3: Data Capture (Equipment)
Data Cycle Step 4: Data Processing and Curation
Data Cycle Step 5: Data Linking and Integration
Data Cycle Step 6: Data Analysis and Interpretation
Data Cycle Step 7: Publishing

Robert Pergl

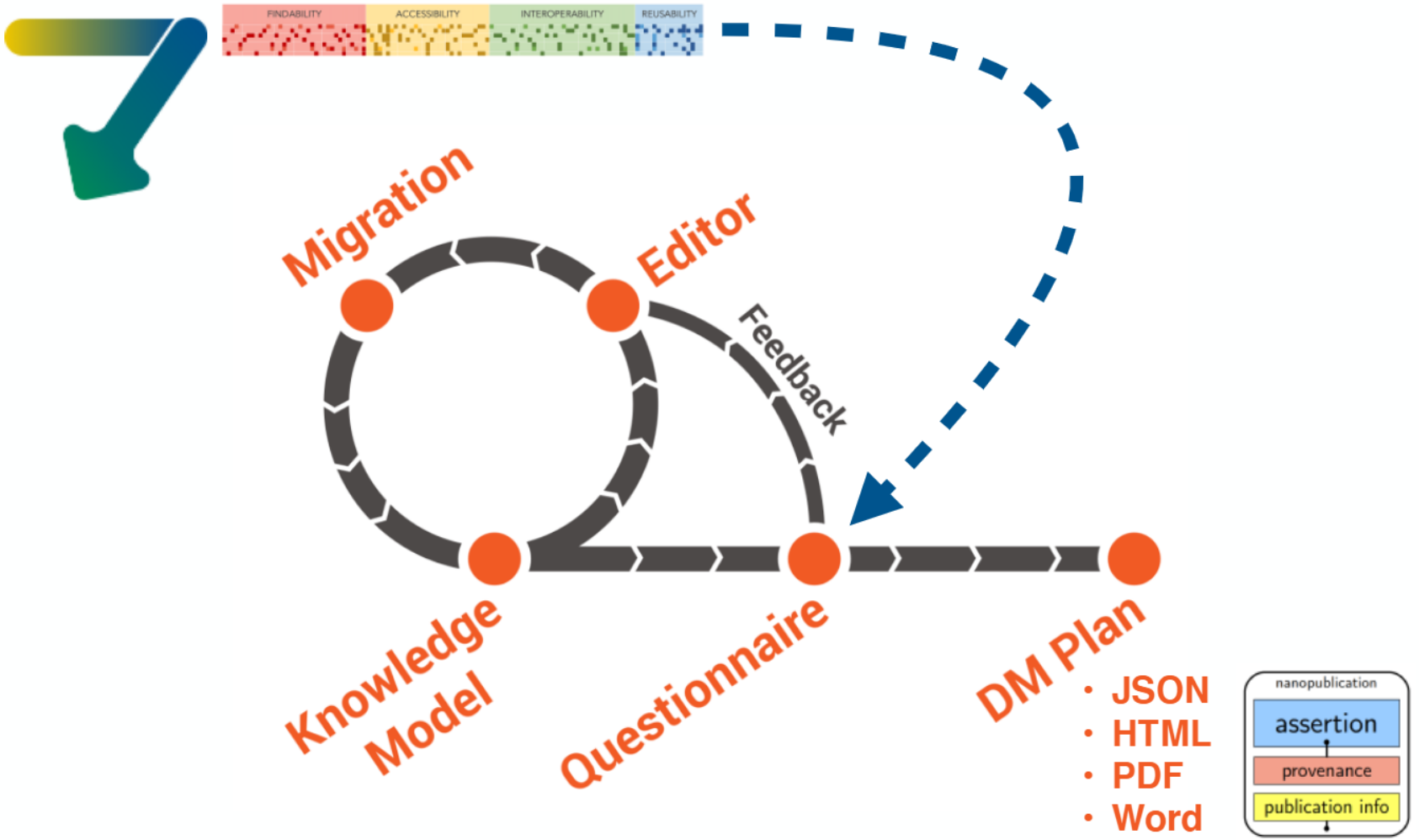
DSW

DATA STEWARDSHIP WIZARD

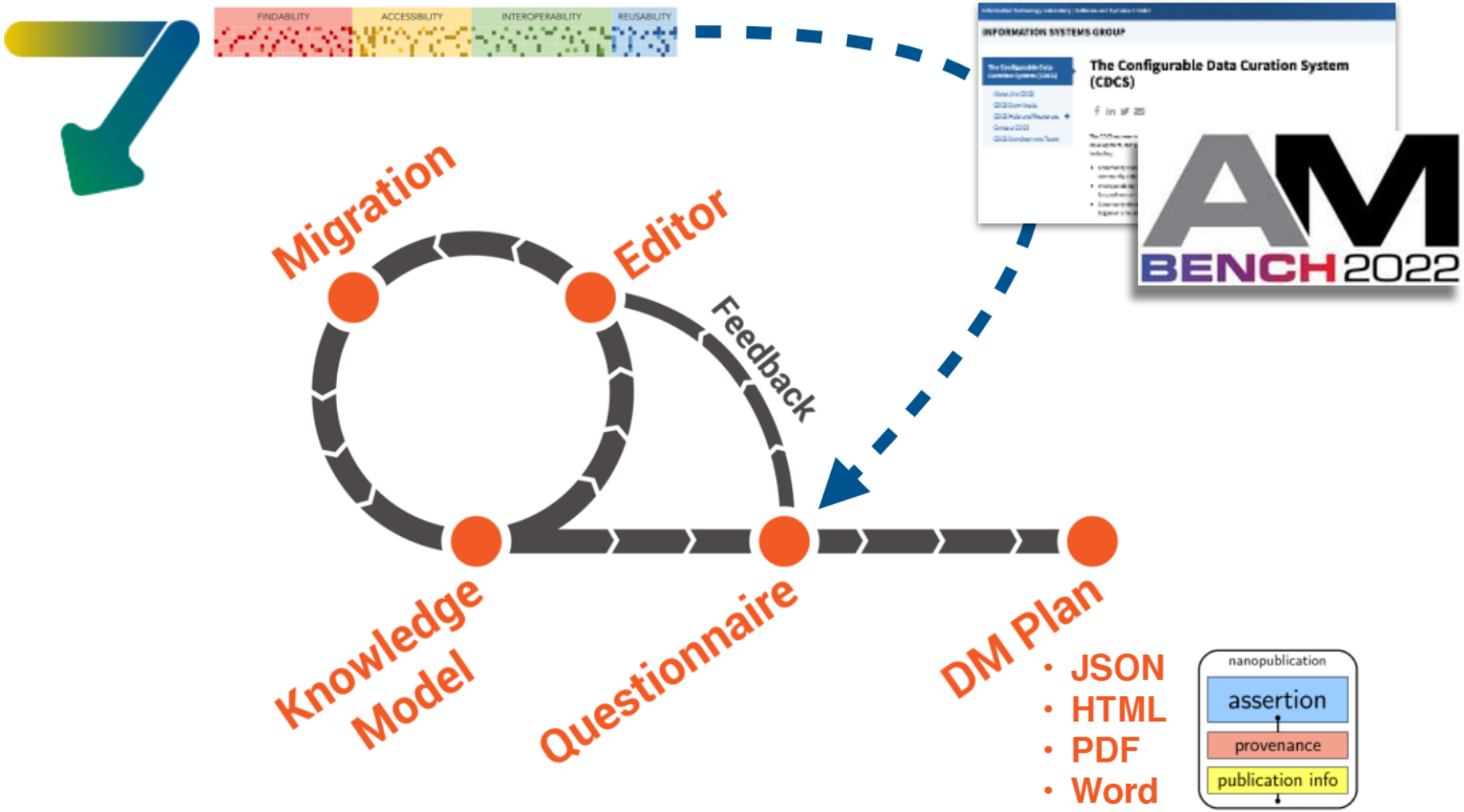
FAIR Data Stewardship



FAIR Data Stewardship



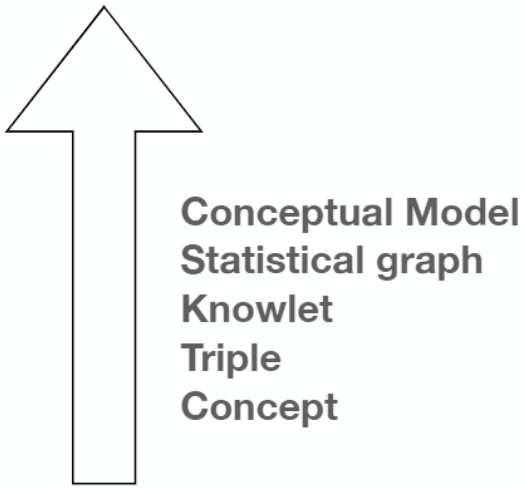
FAIR Data Stewardship



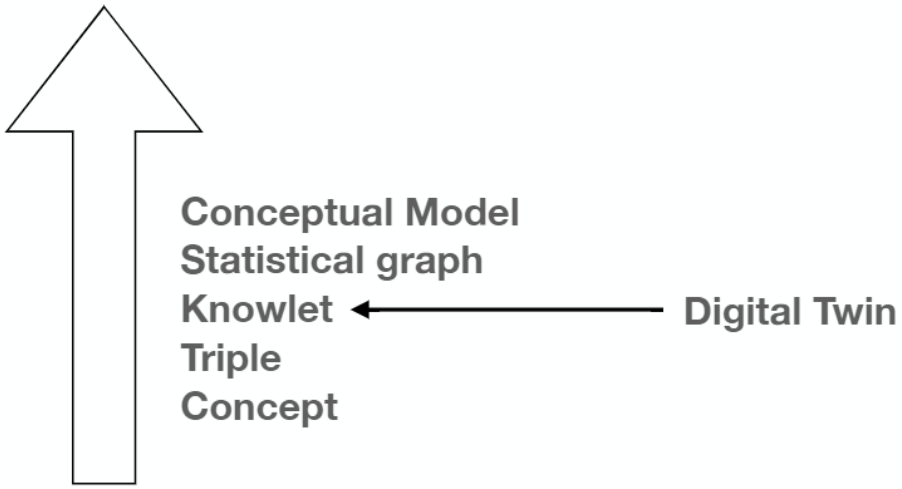
FAIR Digital Twins

Conceptual Modeling Stack

Conceptual Modeling Stack



Conceptual Modeling Stack



Conceptual Modeling Stack

Concept: identifier + definition + synonyms; vocabulary, ontology, PID system

Coronavirus

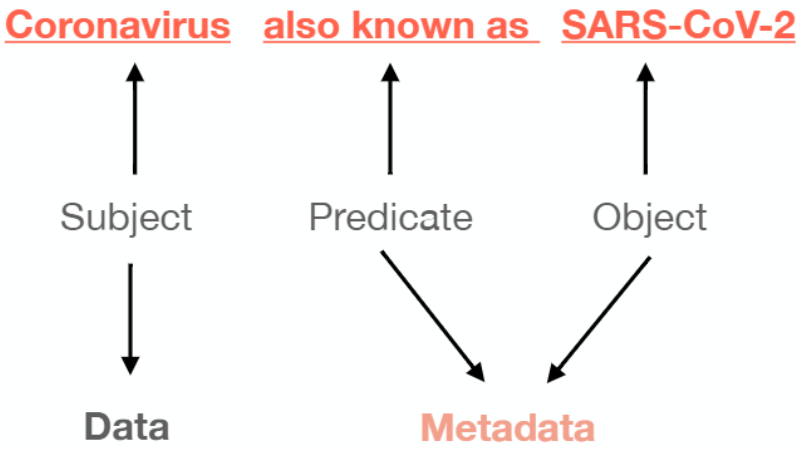
- **https://www.gbif.org/species/10347353**
- **a group of related RNA viruses that cause diseases in mammals and birds**
- **SARS-CoV-2**

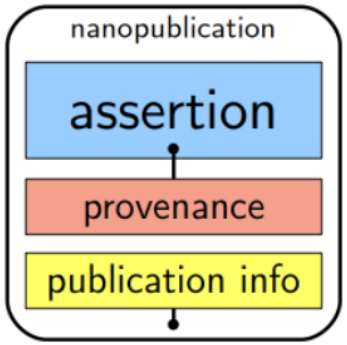
Input form: Protege, VocBench, Concept Wiki, Wikidata

Conceptual Modeling Stack

Triple: Concept (subject) + metadatum (predicate+object)

Triples can be represented as nanopublications - In general, many naopublications have the same cardinal assertion (although each has different provinance and perhaps uses very different ontologies and identfier systems).





Input form: Nanobench, Wizard, CEDAR, Castor

Conceptual Modeling Stack

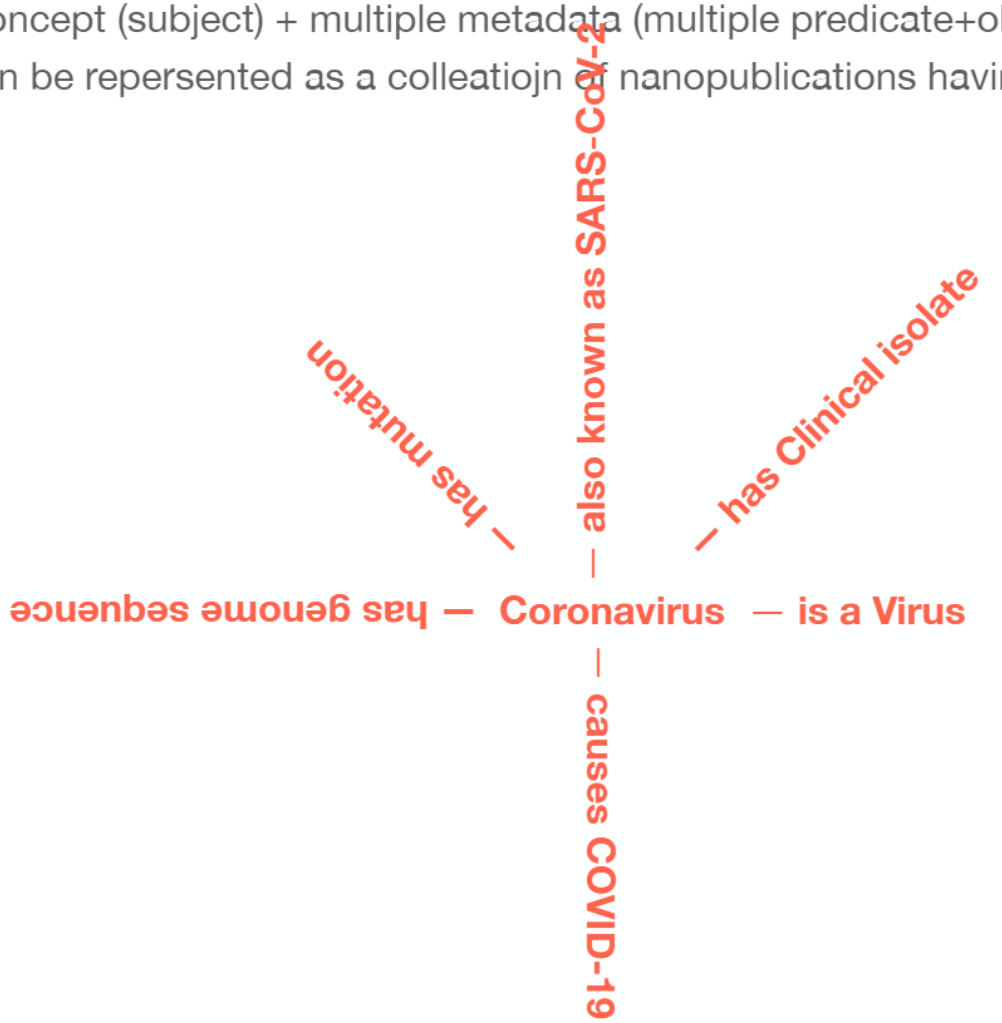
Knowlet: Concept (subject) + multiple metadata (multiple predicate+object pairs)
Knowlets can be repersented as a colleatiojn of nanopublications having the same subject.

Coronavirus	**is a Virus**
Coronavirus	**also known as SARS-CoV-2**
Coronavirus	**causes COVID-19**
Coronavirus	**has Clinical isolate**
Coronavirus	**has genome sequence**
Coronavirus	**has mutation**

Input form: ORKA, Wizard

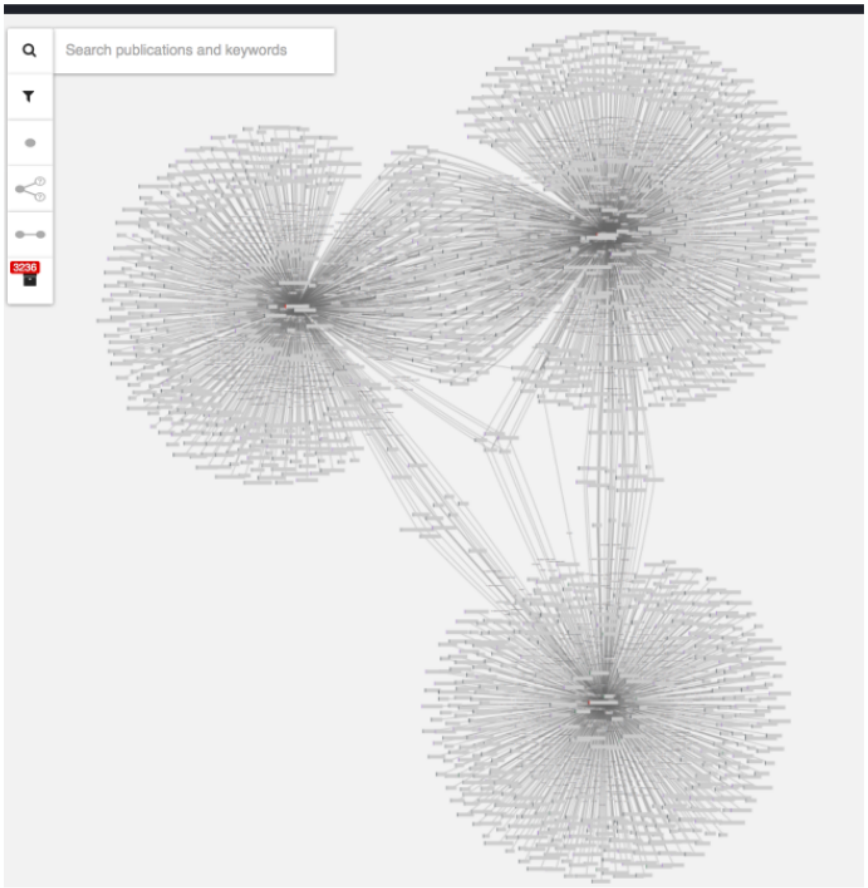
Conceptual Modeling Stack

Knowlet: Concept (subject) + multiple metadata (multiple predicate+object pairs)
Knowlets can be repersented as a colleatiojn of nanopublications having the same subject.

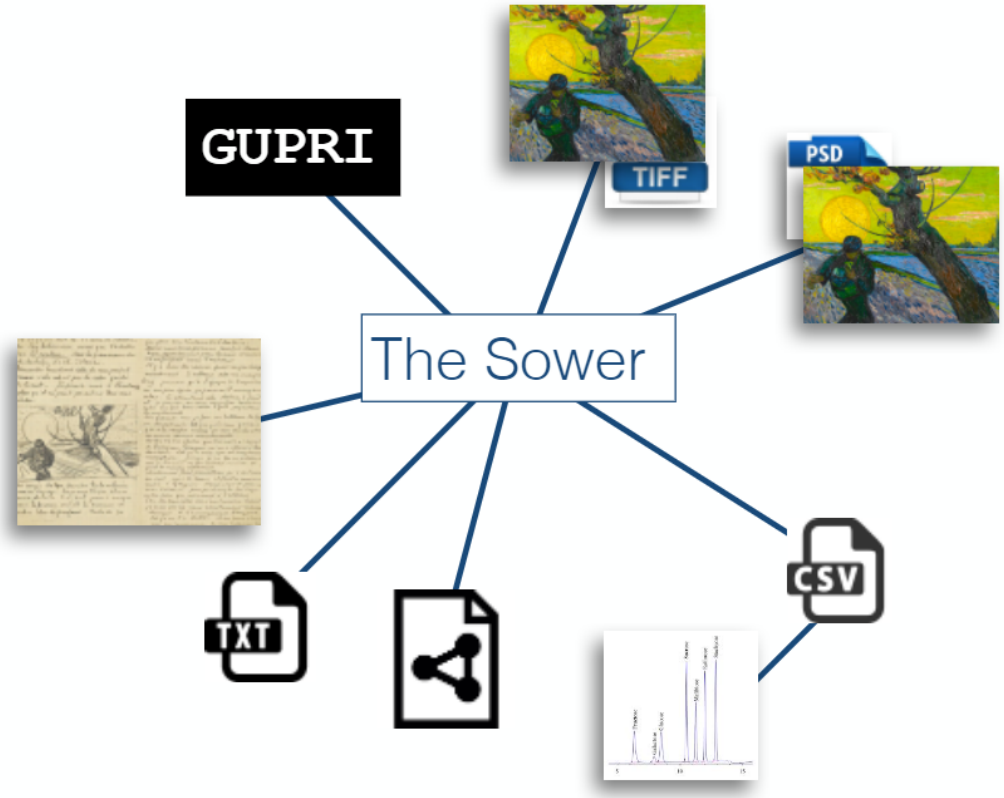


Conceptual Modeling Stack

Knowlet: Concept (subject) + multiple metadata (multiple predicate+object pairs)
Knowlets can be repersented as a colleatiojn of nanopublications having the same subject.



GoghFAIR Digital Twins



https://dissco.tech

Distributed System of Scientific Collections

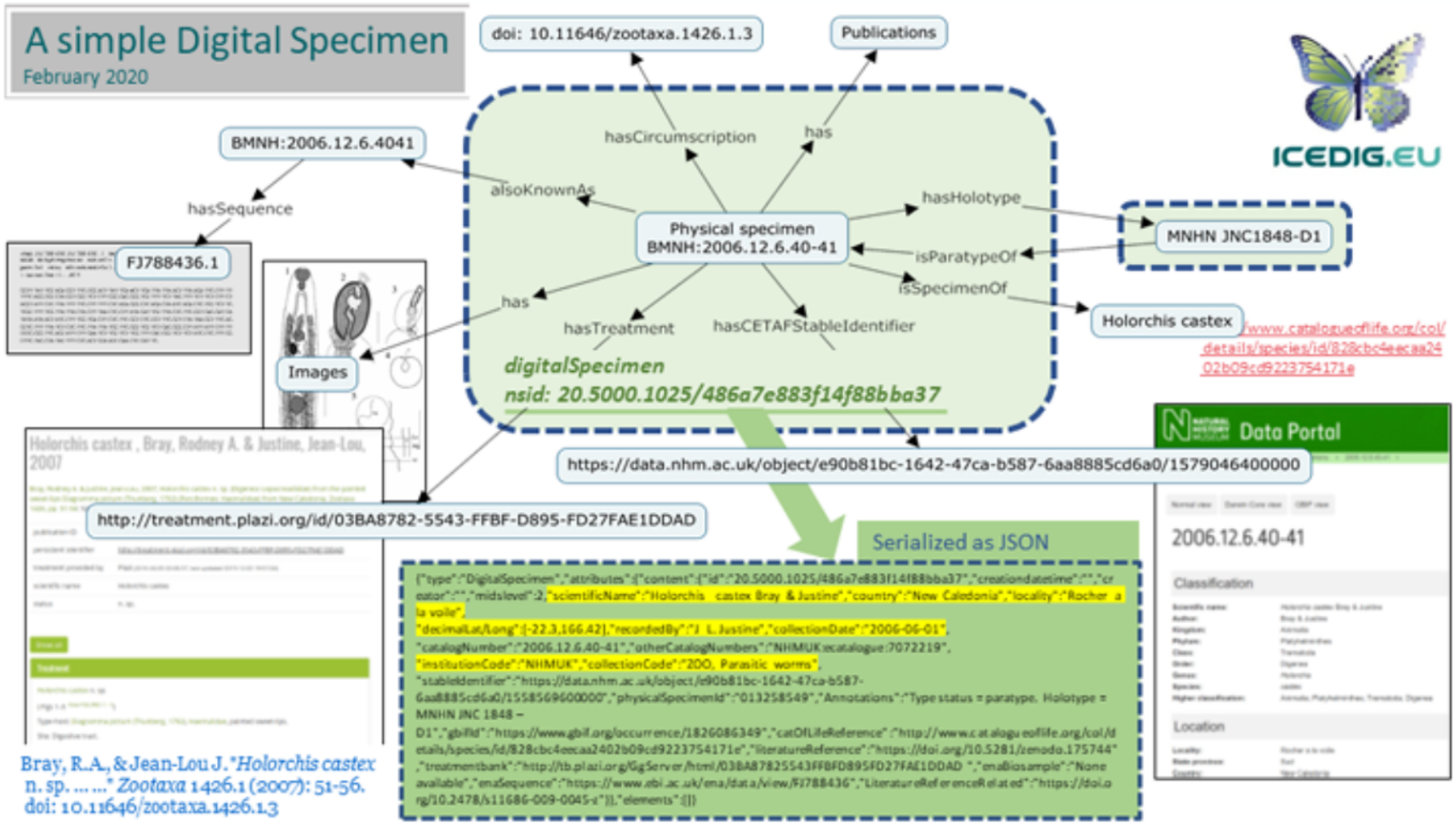
DiSSCoTech

Technical posts about the design of the DiSSCo infrastructure

Home Blog About Contact

Welcome to DiSSCoTech

Data is most reusable where data types are simple and easy to describe, and when the community is organized and collaborative.
— Quote from "FAIR in practice"



https://dissco.tech/2020/03/31/what-is-a-digital-specimen/

https://zenodo.org/record/3685634#.X4RZFi8RpKQ

February 24, 2020

Journal article | Open Access

Humanities Researcher Synergies with Natural Science Collections and Archives

Loo, Tina; Casino, Ana; Gödderz, Karsten; Wijers, Agnes

Task 9.4, Link with Cultural Heritage, falls under WP9, Communication and Dissemination, an effort to identify external actors and ensure their input through effective communication, liaison, networking and dissemination. The task states that a multidisciplinary understanding is needed to effectively plan for the development of a global research infrastructure, thereby requiring that external actors and their potential synergies be identified. It further specifies that the inherent overlap of biodiversity collections with cultural heritage collections, especially in terms of accession books and field notebooks, must be made more explicit, and a roundtable of cultural heritage professionals is identified as a means to discuss these synergies. The scope of this task is limited to identifying synergies that humanities researchers derive from natural science collections, data and archives.

Fulfilling this objective was approached in two ways:

(1) surveying humanities researchers working at the interface of natural science collections and humanities regarding their use of natural science collections, data and archives, and

(2) discussing survey results with other (digital) humanities professionals, and representatives of national and pan-European humanities platforms and research infrastructures in a roundtable format.

Survey results and use cases demonstrated a need by this group of humanities researchers for using a natural science collections data and archive resource and preferably an integrated one, however, the survey was not statistically significant and did not represent the entire demographic.

Discussions at the roundtable covered many subjects. Most notably:

- Planning, organizational and policy concerns were expressed with respect to quantifying the demand for the data resource and having a means to assess whether it warrants investment, roles and responsibilities for driving the initiative and facilitating access and services, and the need to have a clearer understanding of the logistics of this specific landscape.

Publication date:
February 24, 2020

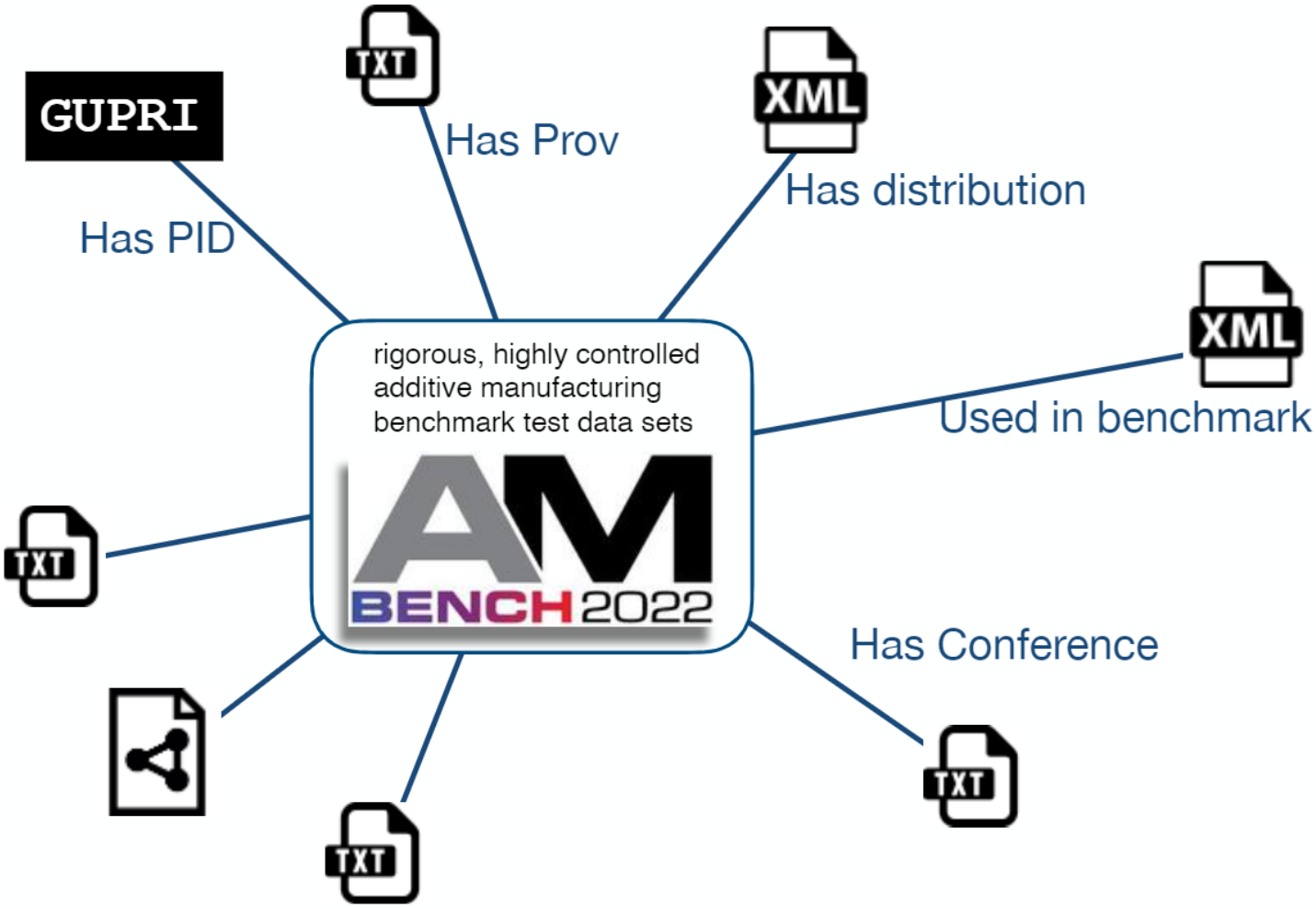
DOI:
DOI 10.5281/zenodo.3685634

Keyword(s):
DiSSCo, Natural Heritage, Cultural Heritage, digital collections,

Grants:
European Commission:
- ICEDIG - Innovation and consolidation for large scale digitisation of natural heritage (777483)

Related identifiers:
References
10.5281/zenodo.3632535 (Project deliverable)
10.1186/1472-6785-13-16 (Journal article)
10.3897/biss.3.37200 (Journal article)

FAIR Digital Twin for AM



Conceptual Modeling Stack

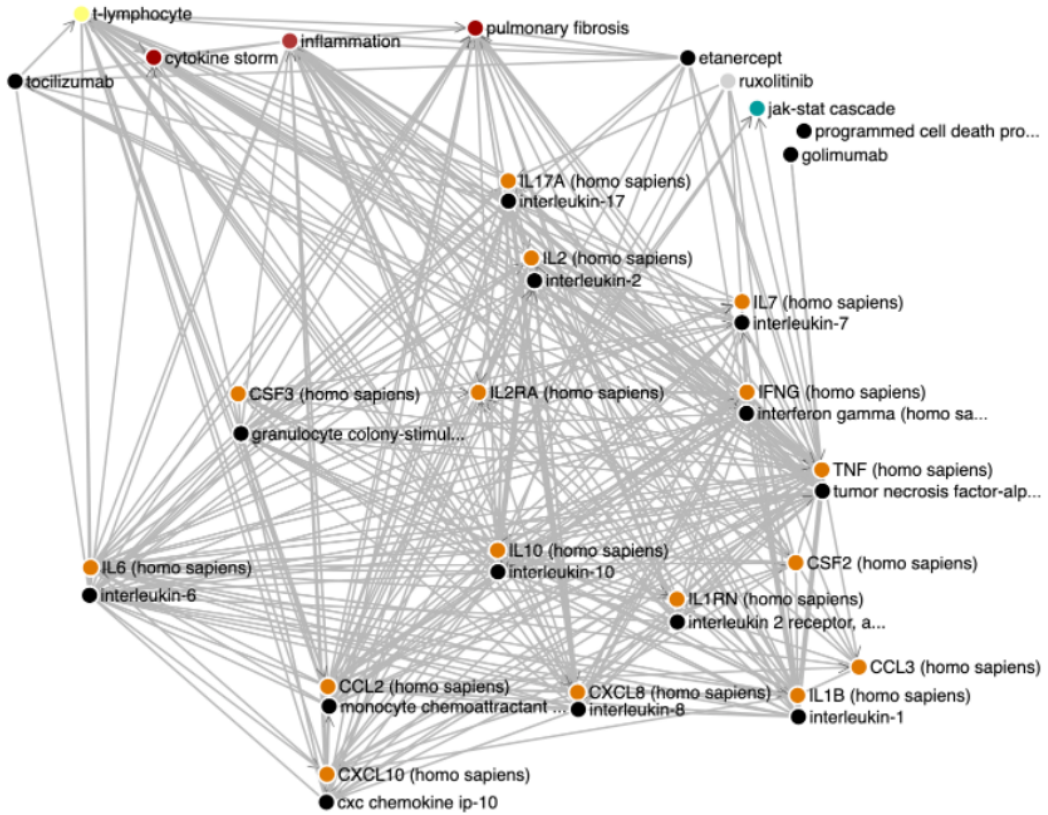
Statistical graph: Associations between Concepts (subjects) determined by fortuitious overlap between metadata components (predicate+object pairs).

Knowlet of Virus, has conceptual overlap with…
Knowlet of Coronavirus, has conceptual overlap with…
Knowlet of SARS-CoV-2, has conceptual overlap with…
Knowlet of COVID-19, has conceptual overlap with…
Knowlet of Clinical isolate, has conceptual overlap with…
Knowlet of genome sequence, has conceptual overlap with…
Knowlet of real world observation of mutation, has conceptual overlap with…

Machine generated: example, the Euretos knowledge graph for Cornonavirus

Conceptual Modeling Stack

Statistical graph: Associations between Concepts (subjects) determined by fortuitious overlap between metadata components (predicate+object pairs).

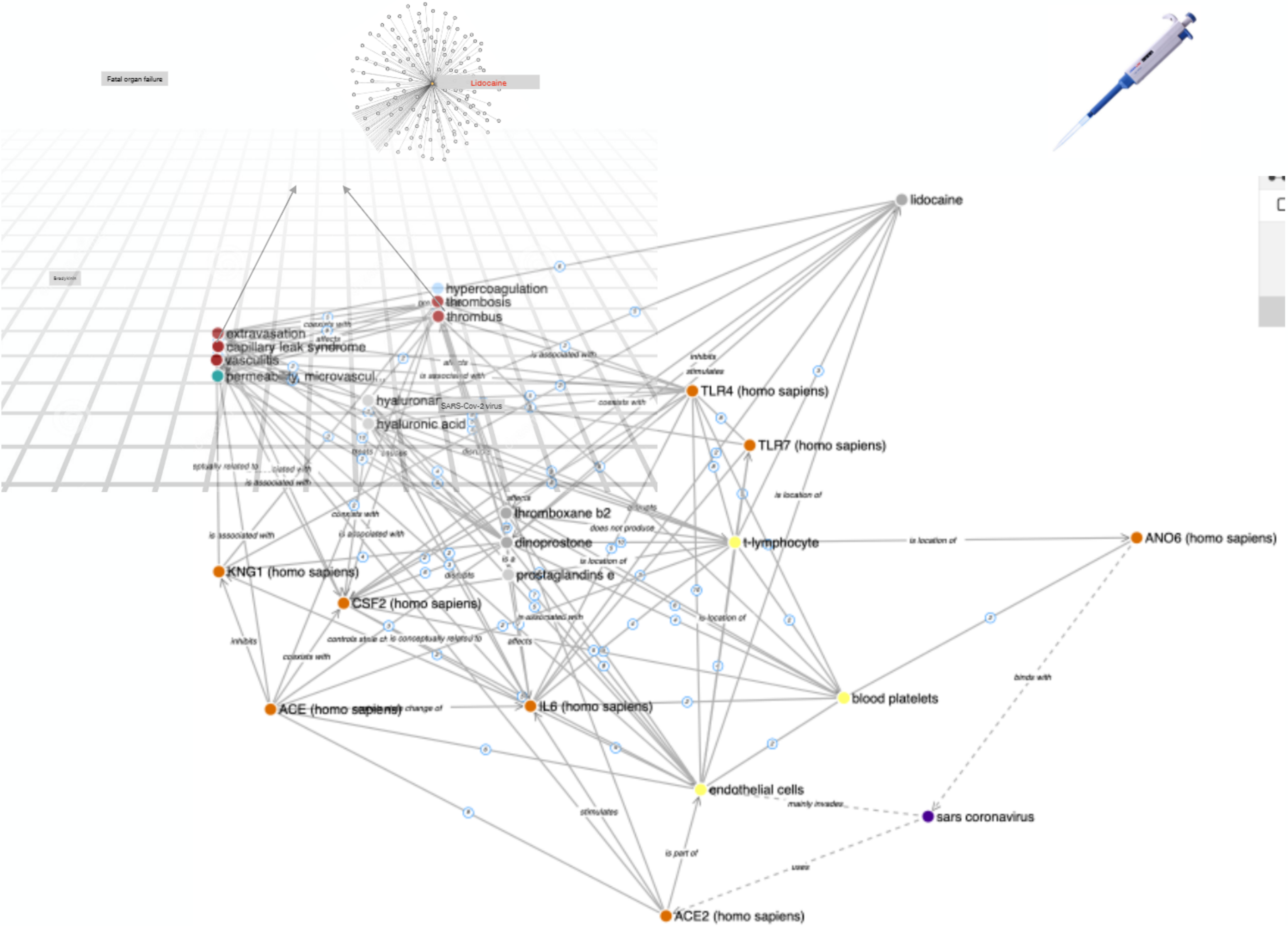


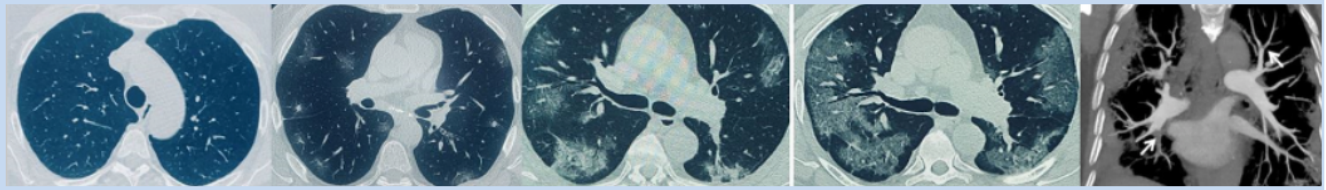
Conceptual Modeling Stack

Conceptual model: Associations between Concepts reflecting the understanding of a human. This is similar to the Statistical graph but for humans most of the associations linking concepts remain inplicit, and the store of knowledge is far less than what the computer can offer. Congruence between the Conceptual model and the Statistical graph (conceptual overlaps) as well as the exposure of novel associations (associations as yet unknown to the human) allow humans to evaluate and improve Conceptual models.

For example, a testable theory about the pathophysiological mechanism of the SARS-CoV-2 virus in patient lung tissues.

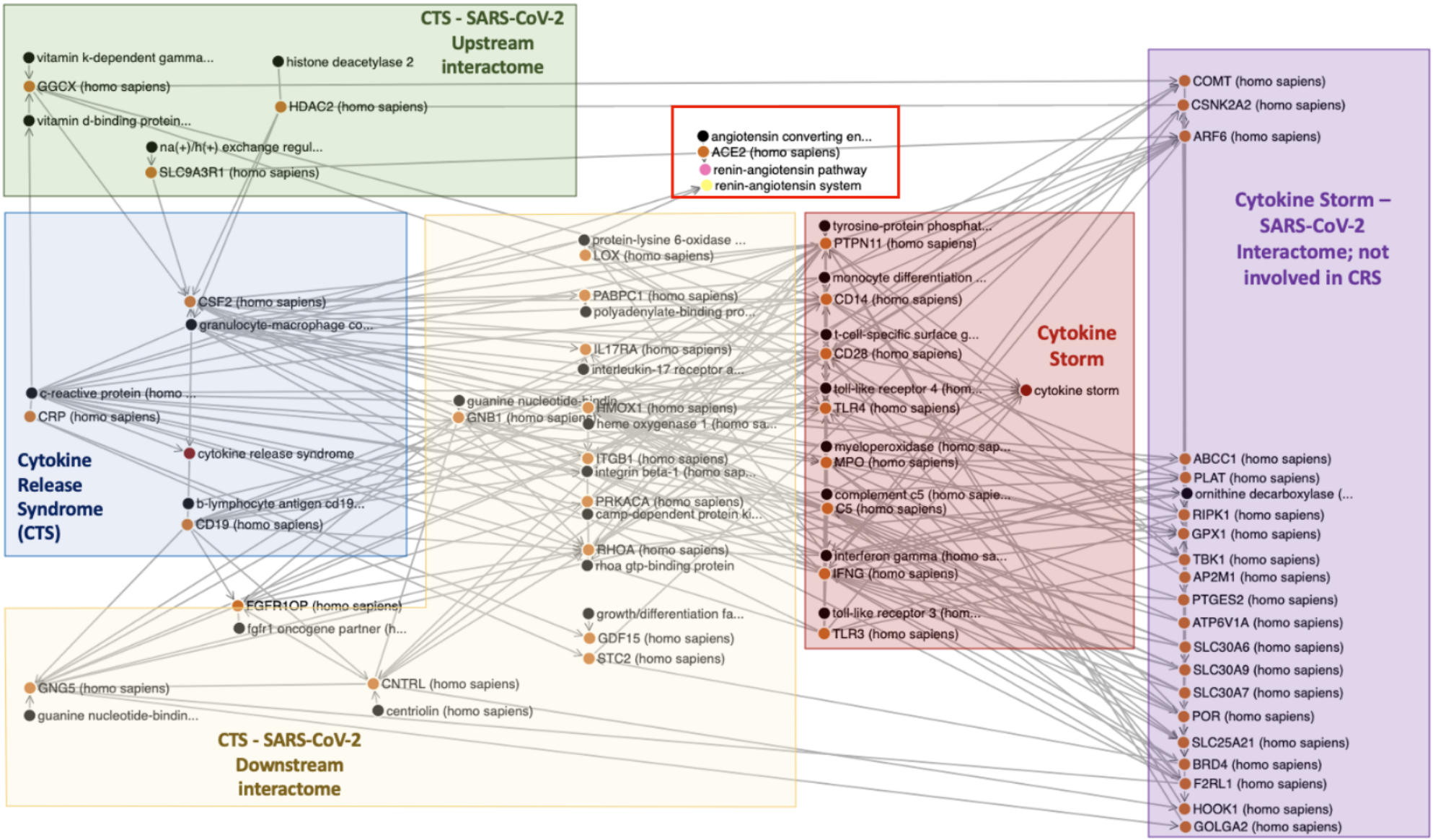
Input form: UML tools, Wizard, CEDAR, Castor


Fatal organ failure
Lidocaine
lidocaine
hypercoagulation
thrombosis
thrombus
extravasation
capillary leak syndrome
vasculitis
permeability, microvascul...
hyaluronan
SARS-Cov-2 virus
hyaluronic acid
TLR4 (homo sapiens)
TLR7 (homo sapiens)
thromboxane b2
dinoprostone
prostaglandins e
t-lymphocyte
ANO6 (homo sapiens)
KNG1 (homo sapiens)
CSF2 (homo sapiens)
ACE (homo sapiens)
IL6 (homo sapiens)
blood platelets
endothelial cells
sars coronavirus
ACE2 (homo sapiens)
is location of
binds with
mainly invades
uses
is part of
stimulates




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
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)
vasculitis
macrophage activation syn...
cell injury
histiocytosis haematophag...
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

Key figure: the gene interactive derived from the 'intersectome' of SARS-Coc-2 predicted proteome with human proteome (aff/MS) and the concepts CRS and CS) Upstream genes/proteins (green) have affinity with Virus proteins interact with CSF2 and CRP, both potentially causing CRS, CRP is also directly associated with the renin-angiotensin pathway that is disturbed in COVID-19 patients As can be seen, there are many downstream connections to the genes directly associated with Cytokine Storm. Note the absence of IL6, IL10 and ACE2 in this pathway (what happens when adding ACE versus ACE2?, how important is this in the very short term? Problems only arise when cytokine s)



Conceptual Modeling Stack