EOSC-Nordic Workshop: "FAIRification of Nordic and Baltic data repositories"

Wednesday 22 Apr 2020, 09:00 → 14:30 Europe/Stockholm

FAIR
&
FAIR Implementation Strategies

Erik Schultes
International Science Coordinator
GO FAIR International Support and Coordination Office
http://orcid.org/0000-0001-8888-635X
erik.schultes@go-fair.org
go-fair.org

Automating F, A, I and R

Box 2 | The FAIR Guiding Principles https://www.nature.com/articles/sdata201618

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

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

FAIR Principles say nothing about FAIR implementation priorities, strategies, costs or choices.

FAIR Implementations

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

GO FAIR

Coordinating decision making to optimize FAIR Implementations.

FAIR Implementations

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

GO FAIR

Maximize reuse of existing solutions
Minimize costs and needless reinvention
Maximize interoperation.
Speed discovery
Shorten time to market

FAIR Implementations

There is no FAIR data without machine-actionable metadata

Box 2 | The FAIR Guiding Principles

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

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

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

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: Repository
Data/metadata: biomedical, geological, historical…

Box 2 | The FAIR Guiding Principles https://www.nature.com/articles/sdata201618

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:

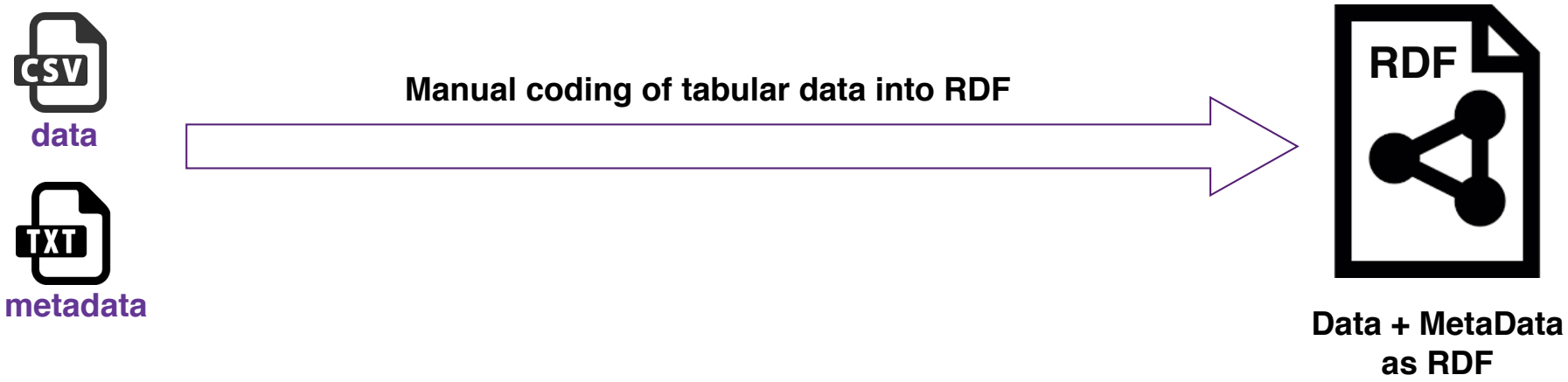
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

For today:

- FAIR Data Point
- FAIR Implementation Profile

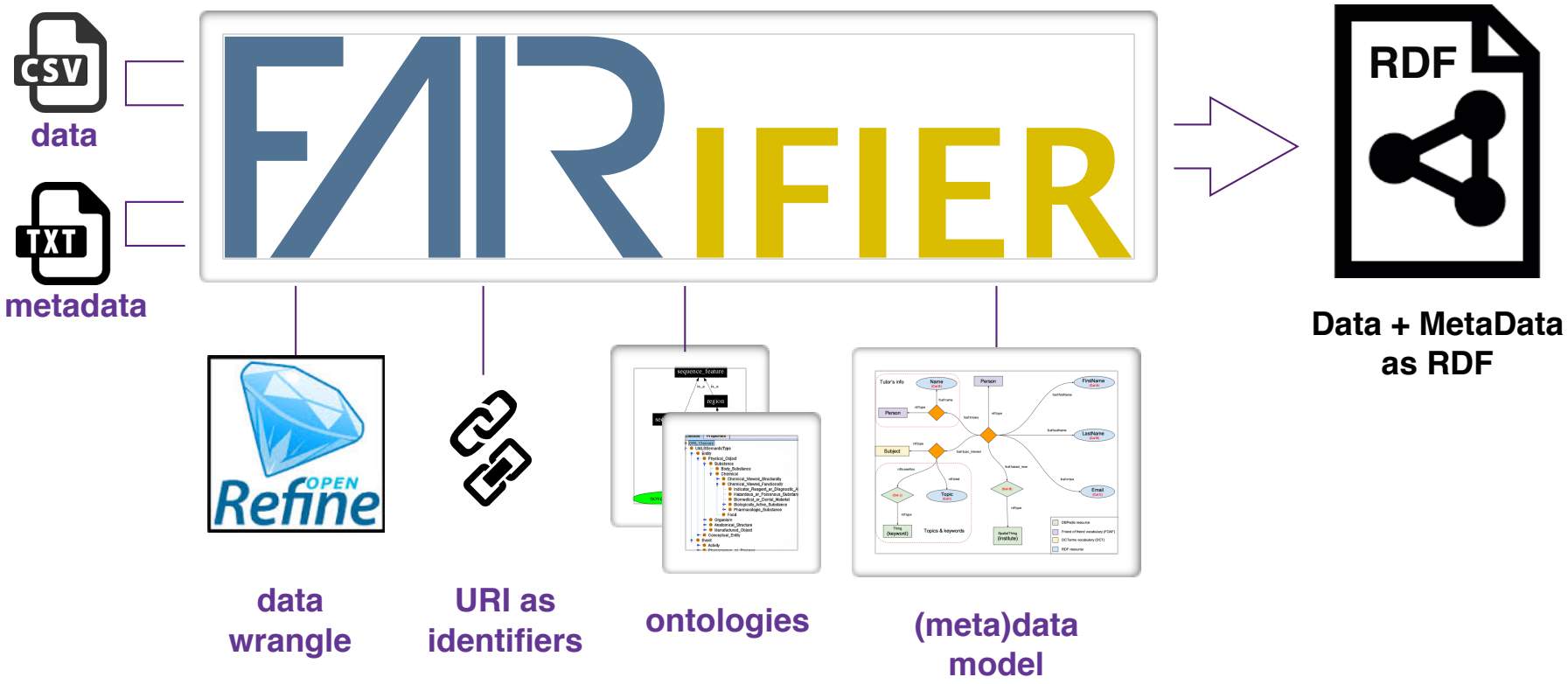
The FAIR Data Point:
A Reference Implementation of the 15 Principles to show how Data can be made FAIR for machines

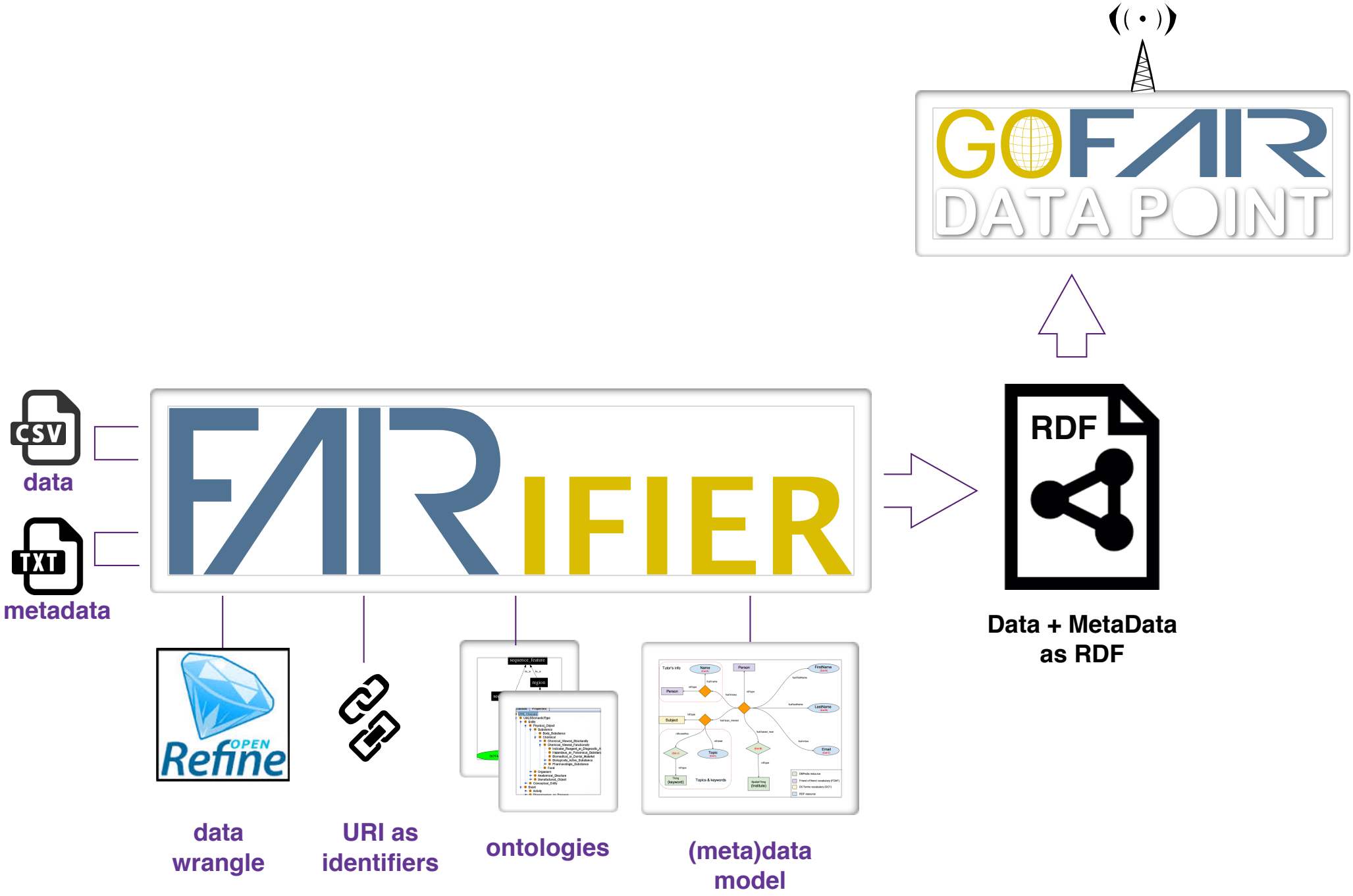
- FAIR-dICT project,DTL: https://www.dtls.nl/fair-data/fair-dict/ **(2016-2017)**
- Interoperability and FAIRness through a novel combination of Web technologies, https://peerj.com/articles/cs-110/
- https://github.com/DTL-FAIRData/FAIRDataPoint/wiki/FAIR-Data-Point-Specification

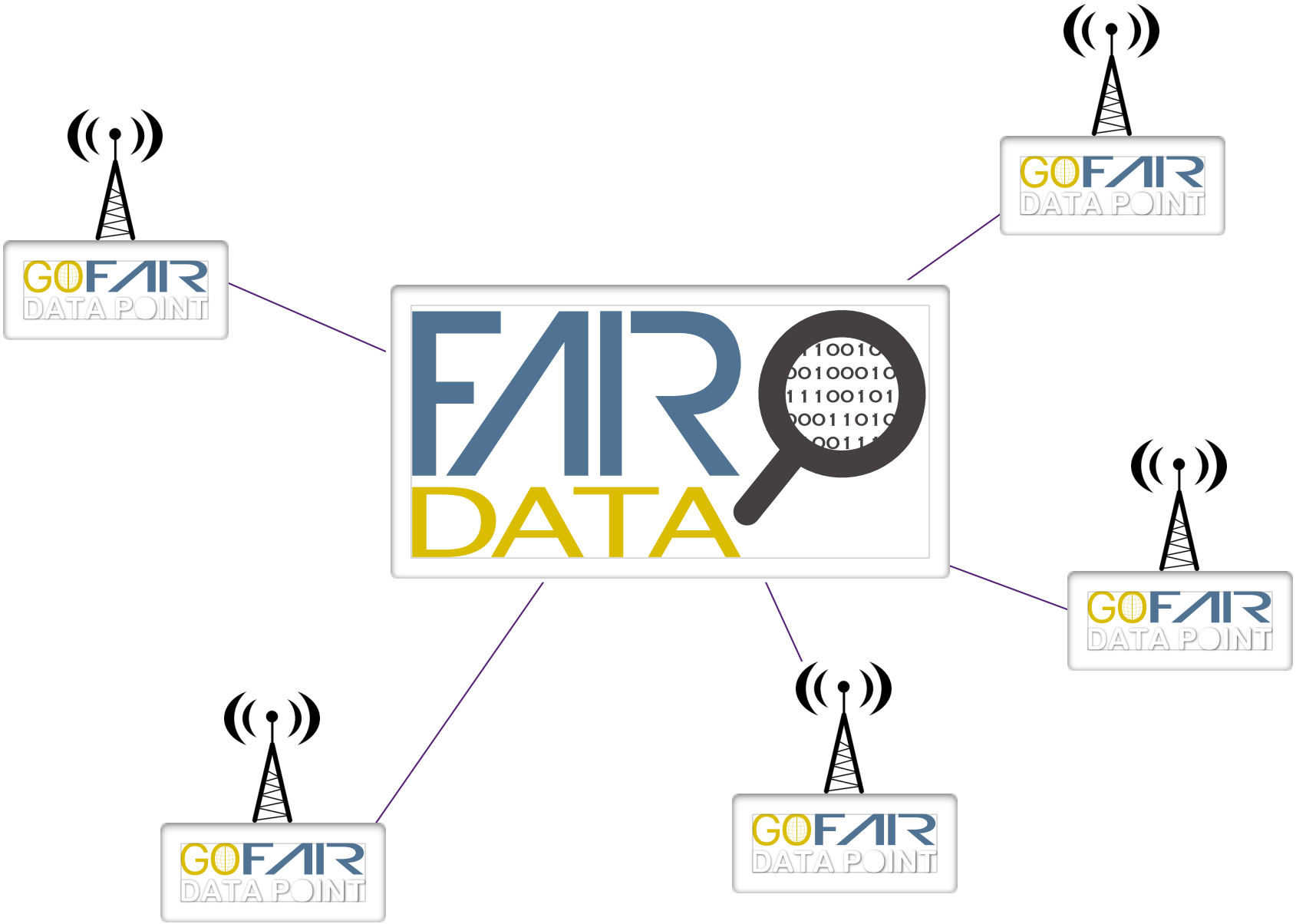


The FAIR Data Point:
A Reference Implementation of the 15 Principles to show how Data can be made FAIR for machines

- FAIR-dICT project,DTL: https://www.dtls.nl/fair-data/fair-dict/ **(2016-2017)**
- Interoperability and FAIRness through a novel combination of Web technologies, https://peerj.com/articles/cs-110/
- https://github.com/DTL-FAIRData/FAIRDataPoint/wiki/FAIR-Data-Point-Specification

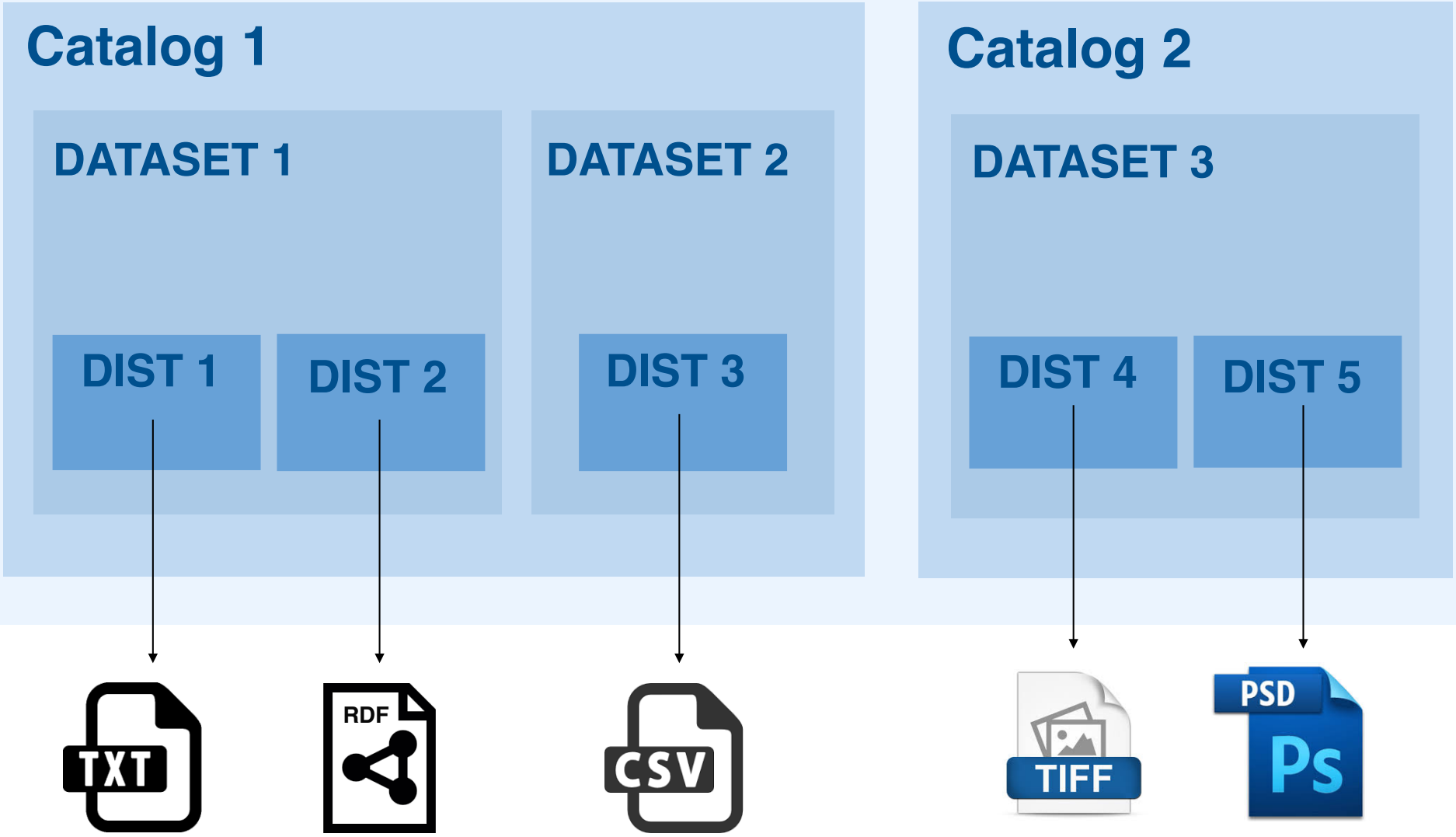



GOFAIR
DATA POINT
CSV
data
TXT
metadata
FAIRIFIER
RDF
Data + MetaData
as RDF
Refine
data
wrangle
URI as
identifiers
ontologies
(meta)data
model


GOFAIR
DATA POINT
GOFAIR
DATA POINT
FAIR
DATA
GOFAIR
DATA POINT
GOFAIR
DATA POINT
GOFAIR
DATA POINT

FAIR Data Point

```
GET <URL>
http://www.w3.org/TR/vocab-dcat/
```



https://github.com/DTL-FAIRData/FAIRDataPoint/wiki/FAIR-Data-Point-Specification

FAIR metadata

http://136.243.4.200:8087/fdp

Title	FDP of biosemantics group	
Metadata ID	fdp	
Description	This is a prototype FDP for hosting research and student projects datasets	
Issued	2017-05-23T09:43:15.57Z	
Modified	2018-08-20T13:09:55	
Version	1.0E0	
License	http://rdflicense.appspot.com/rdflicense/cc-by-nc-nd3.0	
Access Rights	This resource has no access restriction	
Specification	http://rdf.biosemantics.org/fdp/shex/fdpMetadata	
Language	http://id.loc.gov/vocabulary/iso639-1/en	
Publisher	Biosemantic group	
Metrics	Type	https://purl.org/fair-metrics/FM_F1A
	Value	https://www.ietf.org/rfc/rfc3986.txt
	Type	https://purl.org/fair-metrics/FM_A1.1
	Value	https://www.wikidata.org/wiki/Q8777
Catalogs	http://136.243.4.200:8087/fdp/catalog/Transcriptomics	
	http://136.243.4.200:8087/fdp/catalog/multiomics	
	http://136.243.4.200:8087/fdp/catalog/textmining	
	http://136.243.4.200:8087/fdp/catalog/Biosamples	
	http://136.243.4.200:8087/fdp/catalog/Patient_Registries_1.0_998ccbcf-8714-426a-a28e-9335a86adb19	
	http://136.243.4.200:8087/fdp/catalog/SCA3_HD_multi-omics_blood_data	
Institution Country	http://lexvo.org/id/iso3166/NL	
Download RDF	ttl rdf+xml jsonld	

http://136.243.4.200:9080/fdp

orphadata **FAIR Data Point**

Search FAIR Data Point...

Orphadata FAIR Data Point

FAIR Data Point to host free orphanet's dataset(s). These data sets are a partial extraction of the Orphanet data base, freely accessible at www.orpha.net (for consultation purposes only). The data sets are available in nine languages: English, French, German, Italian, Portuguese, Spanish, Dutch, Polish and Czech.

Catalogs

Cross-referencing of rare diseases

Rare Diseases and Cross-referencing

Q929833 Q2330408

Datasets: 9 **Issued:** 27-11-2019 **Modified:** 27-11-2019

Genes associated with rare diseases

Table with ORPHA number of the disease linked to the associated genes, with a characterisation of the relationship between gene and disease (causative, modifier, susceptibility, or...

Q929833 Q7187

Datasets: 1 **Issued:** 27-11-2019 **Modified:** 27-11-2019

Phenotypes associated with rare diseases

Table with diseases listed in Orphanet annotated with HPO phenotypes. The alignment is characterised by frequency (obligatory, very frequent, frequent, occasional, very rare or exc...

Issued
27-11-2019

Modified
27-11-2019

Version
1.0

License
legalcode

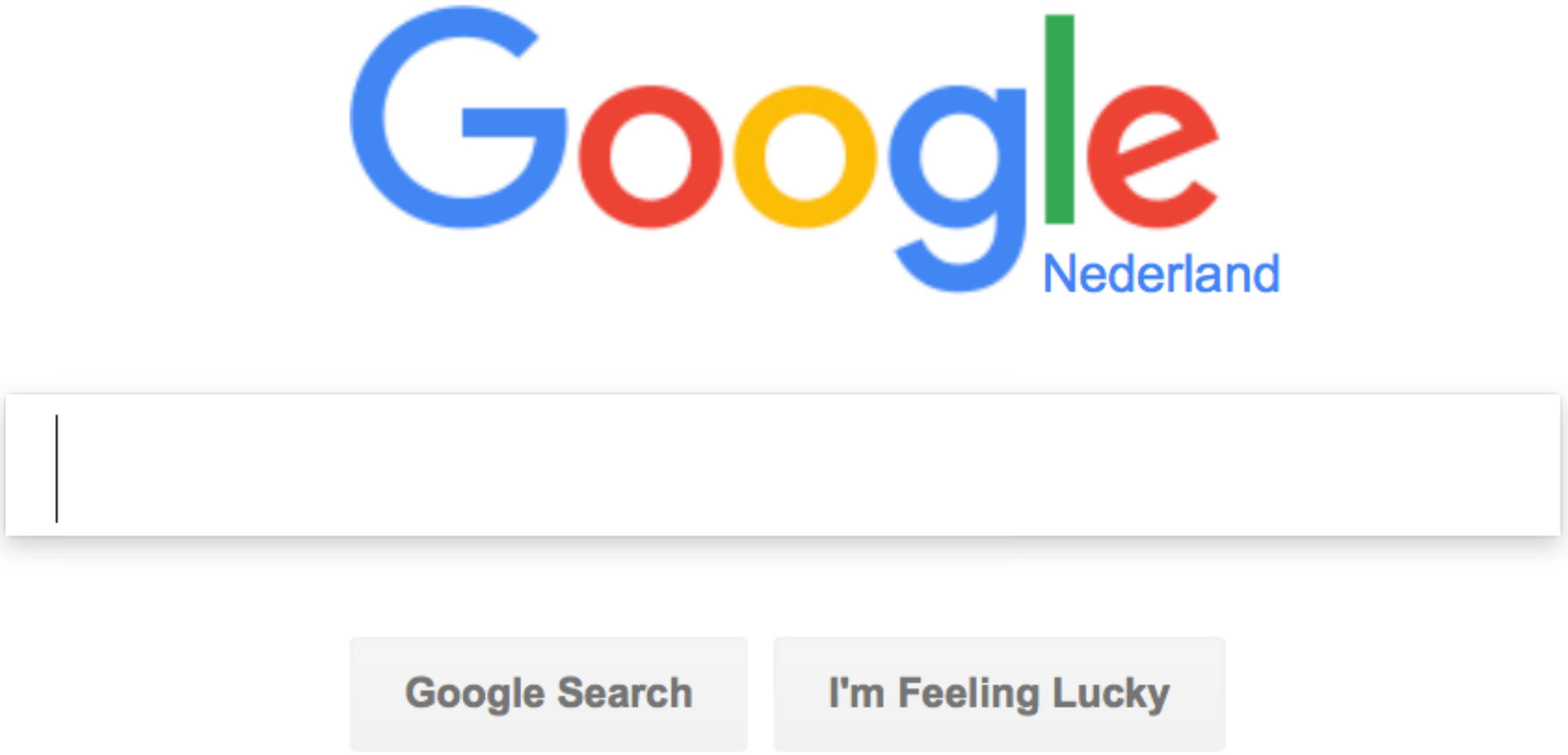
Access Rights
The resource may have some access restrictions

Specification
fdpMetadata

Language
en

Publisher
Orphadata

Download RDF
ttl rdf+xml jsonld



Mark D. Wilkinson
CBGP-UPM/INIA, Madrid

Mark D. Wilkinson
CBGP-UPM/INIA, Madrid

inhibitors p65

All Images Shopping Videos News More Settings Tools

About 544.000 results (0,51 seconds)

RelA/NFkB p65 Inhibitors: Novus Biologicals

https://www.novusbio.com/**inhibitors**/rela-nfkb-**p65**

RelA/NFkB **p65 Inhibitors** available through Novus Biologicals. Browse our RelA/ NFkB **p65 Inhibitor** catalog backed by our Guarantee+.

Inhibiting NF-κB Activation by Small Molecules As a Therapeutic ...

https://www.ncbi.nlm.nih.gov › NCBI › Literature › PubMed Central (PMC)

by SC Gupta - 2010 - Cited by 345 - Related articles

May 21, 2010 - ... pathway (Fig 3). Table 1. A list of small molecules as **inhibitors** of NF-kB pathway .... Blocking NF-κB activation by **inhibitors** of **p65** acetylation.

Suppression of p65 phosphorylation coincides with inhibition of ... - NCBI

https://www.ncbi.nlm.nih.gov/pubmed/16163708

by J Hu - 2005 - Cited by 26 - Related articles

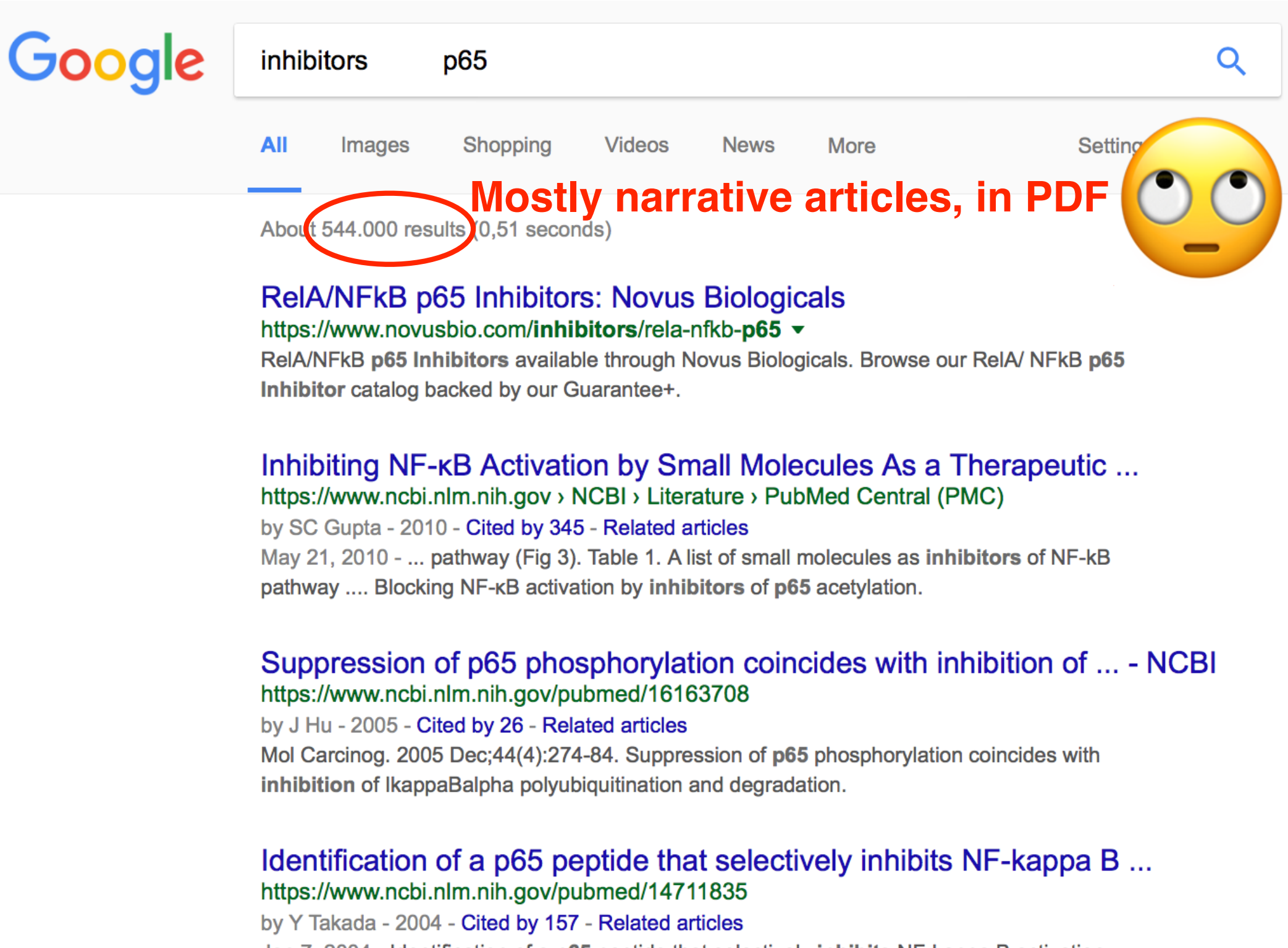
Mol Carcinog. 2005 Dec;44(4):274-84. Suppression of **p65** phosphorylation coincides with **inhibition** of IkappaBalpha polyubiquitination and degradation.

Identification of a p65 peptide that selectively inhibits NF-kappa B ...

https://www.ncbi.nlm.nih.gov/pubmed/14711835

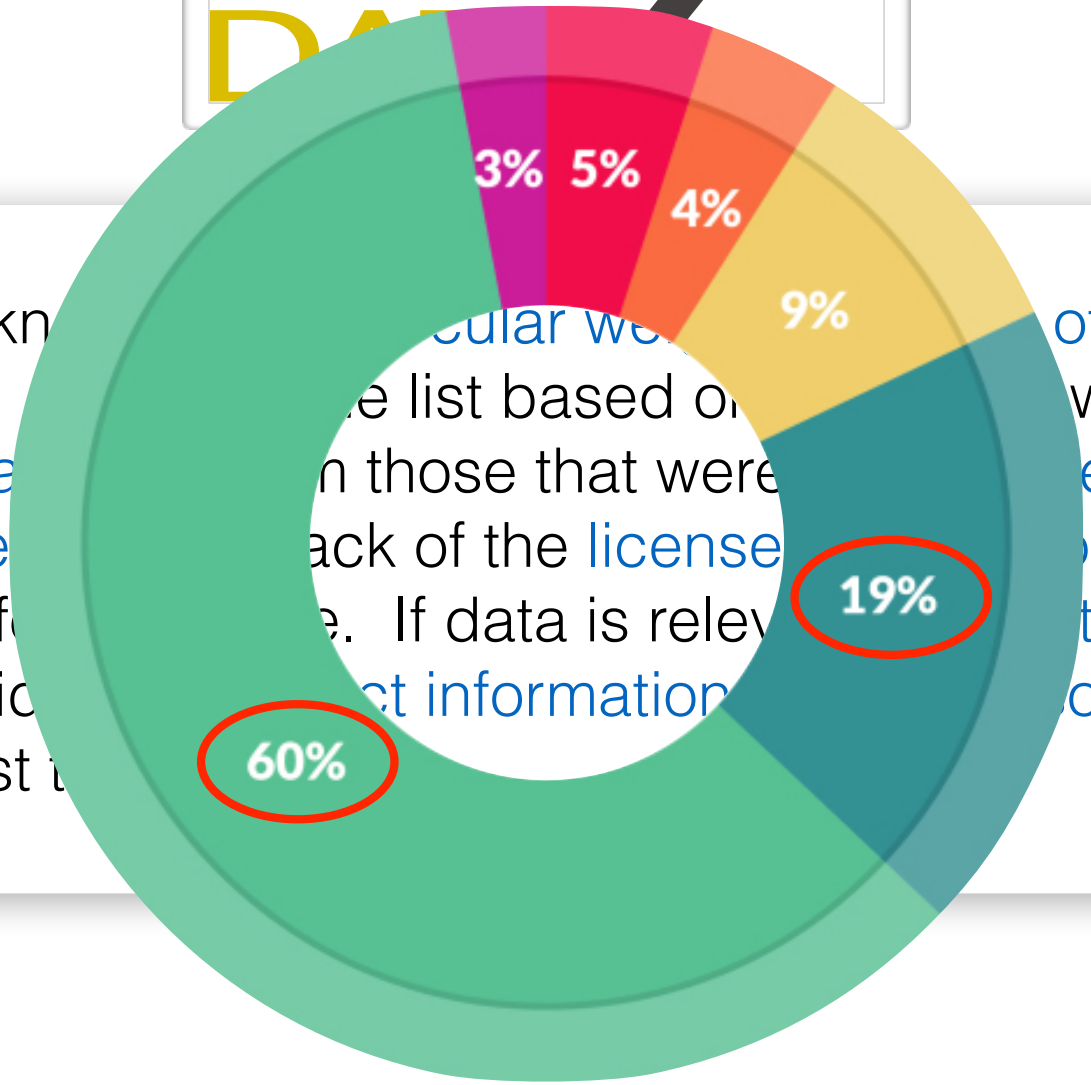
by Y Takada - 2004 - Cited by 157 - Related articles

Jan 7, 2004 - Identification of a **p65** peptide that selectively **inhibits** NF-kappa B activation


Google
inhibitors p65
All Images Shopping Videos News More Setting
Mostly narrative articles, in PDF
About 544.000 results (0,51 seconds)
RelA/NFkB p65 Inhibitors: Novus Biologicals
https://www.novusbio.com/inhibitors/rela-nfkb-p65
RelA/NFkB p65 Inhibitors available through Novus Biologicals. Browse our RelA/ NFkB p65 Inhibitor catalog backed by our Guarantee+.
Inhibiting NF-κB Activation by Small Molecules As a Therapeutic ...
https://www.ncbi.nlm.nih.gov › NCBI › Literature › PubMed Central (PMC)
by SC Gupta - 2010 - Cited by 345 - Related articles
May 21, 2010 - ... pathway (Fig 3). Table 1. A list of small molecules as inhibitors of NF-kB pathway .... Blocking NF-κB activation by inhibitors of p65 acetylation.
Suppression of p65 phosphorylation coincides with inhibition of ... - NCBI
https://www.ncbi.nlm.nih.gov/pubmed/16163708
by J Hu - 2005 - Cited by 26 - Related articles
Mol Carcinog. 2005 Dec;44(4):274-84. Suppression of p65 phosphorylation coincides with inhibition of IkappaBalpha polyubiquitination and degradation.
Identification of a p65 peptide that selectively inhibits NF-kappa B ...
https://www.ncbi.nlm.nih.gov/pubmed/14711835
by Y Takada - 2004 - Cited by 157 - Related articles
Jan 7, 2004 - Identification of a p65 peptide that selectively inhibits NF-kappa B activation

Find me all known low molecular weight inhibitors of the Human p65 Protein. Separate the list based on those that were found in curated databases, from those that were found in self-deposited data archives. Keep track of the license and citation information for each one. If data is relevant, but not public, please provide the contact information for the person I need so I can request the data.

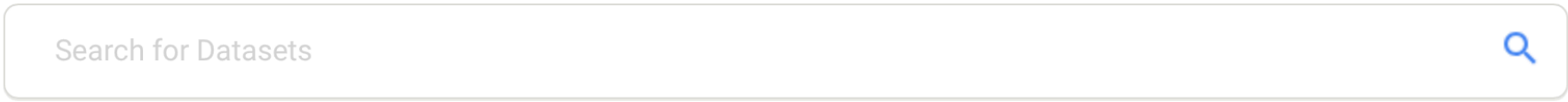
Mark D. Wilkinson
CBGP-UPM/INIA, Madrid



Find me all kn … of the Human p65 Protein. … list based o… were found in curated data… those that were… elf-deposited data archive… ack of the license… on information f… e. If data is relev… t public, please provid… ct information… on I need so I can request t…

Mark D. Wilkinson
CBGP-UPM/INIA, Madrid

September 5, 2018



Try boston education data or weather site:noaa.gov

September 5, 2018

Biosemantics Group LUMC

Try boston education data or weather site:noaa.gov

Gene disease association (LUMC)

semlab1.liacs.nl

Dataset provided by

Biosemantics group leiden

License

Attribution-NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

Available download formats from providers

HTML

Description

Gene disease association dataset from LUMC

L

Gene disease association (LUMC)

semlab1.liacs.nl

Google Structured Data Testing Tool

http://136.243.4.200:8087/fdp/dataset/gene_disease_association

NEW TEST

```
<!DOCTYPE html>
<html>
    <head>
        <!-- Latest compiled and minified CSS -->
        <link rel="stylesheet" href="https://stackpath.bootstrapcdn.com/bootstrap/4.1.0/css/bootstrap.min.css"
integrity="sha384-9gVQ4dYFwwWSjIDZnLEWnxCjeSWFphJiwGPXr1jddIhOegiu1FwO5qRGvFXOdJZ4" crossorigin="anonymous">

        <title>

            Gene disease association (LUMC)

        </title>

        <script type="application/ld+json">
            {"@graph":
[{"@id":"http://136.243.4.200:8087/fdp/dataset/gene_disease_association","@type":"http://schema.org/Dataset","http
://schema.org/creator":{"@id":"http://biosemantics.org"},"http://schema.org/description":"High-throughput
experimental methods such as medical sequencing and genome-wide association studies (GWAS) identify increasingly
large numbers of potential relations between genetic variants and diseases. Both biological complexity (millions
of potential gene-disease associations) and the accelerating rate of data production necessitate computational
approaches to prioritize and rationalize potential gene-disease relations. Here, we use concept profile technology
to expose from the biomedical literature both explicitly stated gene-disease relations (the explicitome) and a
much larger set of implied gene-disease associations (the implicitome). Implicit relations are largely unknown to,
or are even unintended by the original authors, but they vastly extend the reach of existing biomedical knowledge
for identification and interpretation of gene-disease associations. The implicitome can be used in conjunction
with experimental data resources to rationalize both known and novel associations. We demonstrate the usefulness
of the implicitome by rationalizing known and novel gene-disease associations, including those from GWAS. To
facilitate the re-use of implicit gene-disease associations, we publish our data in compliance with FAIR Data
Publishing recommendations [https://www.force11.org/group/fairgroup] using nanopublications. An online tool
(http://knowledge.bio) is available to explore established and potential gene-disease associations in the context
of other biomedical relations.","http://schema.org/distribution":
[{"@id":"http://136.243.4.200:8087/fdp/distribution/gene_disease_association_html"},
{"@id":"http://136.243.4.200:8087/fdp/distribution/gene_disease_association_nquads_gzip"},
{"@id":"http://136.243.4.200:8087/fdp/distribution/gene_disease_association_csv_gzip"}],"http://schema.org/keyword
s":["The Explicitome","The Implicitome","Text mining","Gene disease association
(LUMC)","GDA","LWAS"],"http://schema.org/name":"Gene disease association (LUMC)"},
{"@id":"http://biosemantics.org","@type":"http://schema.org/Thing","http://schema.org/name":"Biosemantic
group"}],"@context":{"rdf":"http://www.w3.org/1999/02/22-rdf-syntax-ns#","rdfs":"http://www.w3.org/2000/01/rdf-
schema#","dcat":"http://www.w3.org/ns/dcat#","xsd":"http://www.w3.org/2001/XMLSchema#","owl":"http://www.w3.org/20
02/07/owl#","dcterms":"http://purl.org/dc/terms/","fdp":"http://rdf.biosemantics.org/ontologies/fdp-
o#","r3d":"http://www.re3data.org/schema/3-0#","lang":"http://id.loc.gov/vocabulary/iso639-1/"}}
        </script>

        <style>
            /* Sticky footer styles
            -------------------------------------------------- */
            html {
              position: relative;
              min-height: 100%;
            }
            body {
              /* Margin bottom by footer height */
              margin-bottom: 60px;
            }
```

Dataset

0 ERRORS 0 WARNINGS

ID: http://136.243.4.200:8087/fdp/dataset/gene_disease_association

@type	Dataset
@id	http://136.243.4.200:8087/fdp/dataset/gene_disease_association
description	High-throughput experimental methods such as medical sequencing and genome-wide association studies (GWAS) identify increasingly large numbers of potential relations between genetic variants and diseases. Both biological complexity (millions of potential gene-disease associations) and the accelerating rate of data production necessitate computational approaches to prioritize and rationalize potential gene-disease relations. Here, we use concept profile technology to expose from the biomedical literature both explicitly stated gene-disease relations (the explicitome) and a much larger set of implied gene-disease associations (the implicitome). Implicit relations are largely unknown to, or are even unintended by the original authors, but they vastly extend the reach of existing biomedical knowledge for identification and interpretation of gene-disease associations. The implicitome can be used in conjunction with experimental data resources to rationalize both known and novel associations. We demonstrate the usefulness of the implicitome by rationalizing known and novel gene-disease associations, including those from GWAS. To facilitate the re-use of implicit gene-disease associations, we publish our data in compliance with FAIR Data Publishing recommendations [https://www.force11.org/group/fairgroup] using nanopublications. An online tool (http://knowledge.bio) is available to explore established and potential gene-disease associations in the context of other biomedical relations.
keywords	The Explicitome
keywords	The Implicitome
keywords	Text mining
keywords	Gene disease association (LUMC)
keywords	GDA
keywords	LWAS
name	Gene disease association (LUMC)
creator	
@type	Thing
@id	http://biosemantics.org/
name	Biosemantic group
distribution	
@type	DataDownload
@id	http://136.243.4.200:8087/fdp/distribution/gene_disease_association_html
distribution	
@type	DataDownload
@id	http://136.243.4.200:8087/fdp/distribution/gene_disease_association_nquads_gzip
distribution	
@type	DataDownload
@id	http://136.243.4.200:8087/fdp/distribution/gene_disease_association_csv_gzip

Virus Outbreak Data Network (VODAN)

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*. In the case of Ebola this is even more harrowing and ironic as the data are *least available* to the population that were *most affected* by the disaster. 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).

Purpose of the Implementation Network

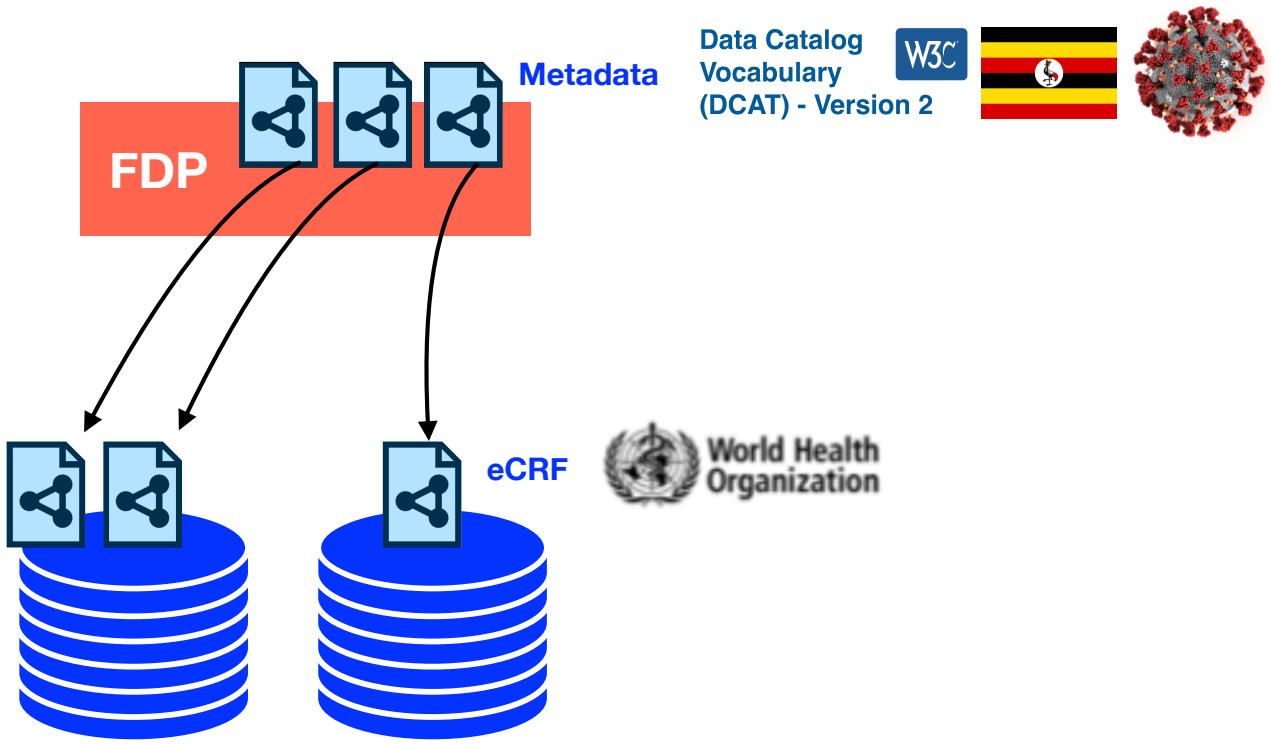
This time, we can do better. We now have the technical ability, as well as the commitment from experts in a series of affected countries, to make the SARS CoV-2 virus data FAIR, meaning that they are Findable, Accessible, Interoperable and thus Reusable by both humans and machines, during this epidemic of COVID-19. The technical components that make this possible can remain in place, waiting in the ready state *for potential future infectious disease outbreaks*.

With a sense of urgency driven by the rapid developments on COVID-19 we come together to launch a GO FAIR Implementation Network to address the current and immediate challenges. For this epidemic, unfortunately, we have to 'FAIRify' COVID-19 data *'after the fact'* and use Chinese, Dutch, Swedish, etc. and English electronic (or even hand-written) health records to create proper FAIR data. The FAIRification will initially focus on the Clinical Research Form (CRF) model following the WHO standards. Multiple IN partners will create input forms that make it easy for local caregivers to create FAIR-CRF data in real time as a first step. As a second step, we will jointly develop (via online work sessions) localized FAIR Data Points (FDP). FDP is a FAIR data repository with 'docking' capabilities as a 'station' for 'trains' (virtual machines (VMs)) that come to 'visit' the data locally, with a specific question to ask. The local data custodian (frequently a hospital or centre for disease control and prevention type of institution) grants permission to VMs to ask the question / run analyses. As the personal data of patients never leaves the underlying database of the local institution, GDPR issues are largely accommodated and in this way data can be 'shared' or rather 'visited' without violating any patient rights and, in the case of a disease outbreak, also governed by the laws and policies of the individual jurisdictions in which the outbreak manifests.

The VODAN FAIR Data Point

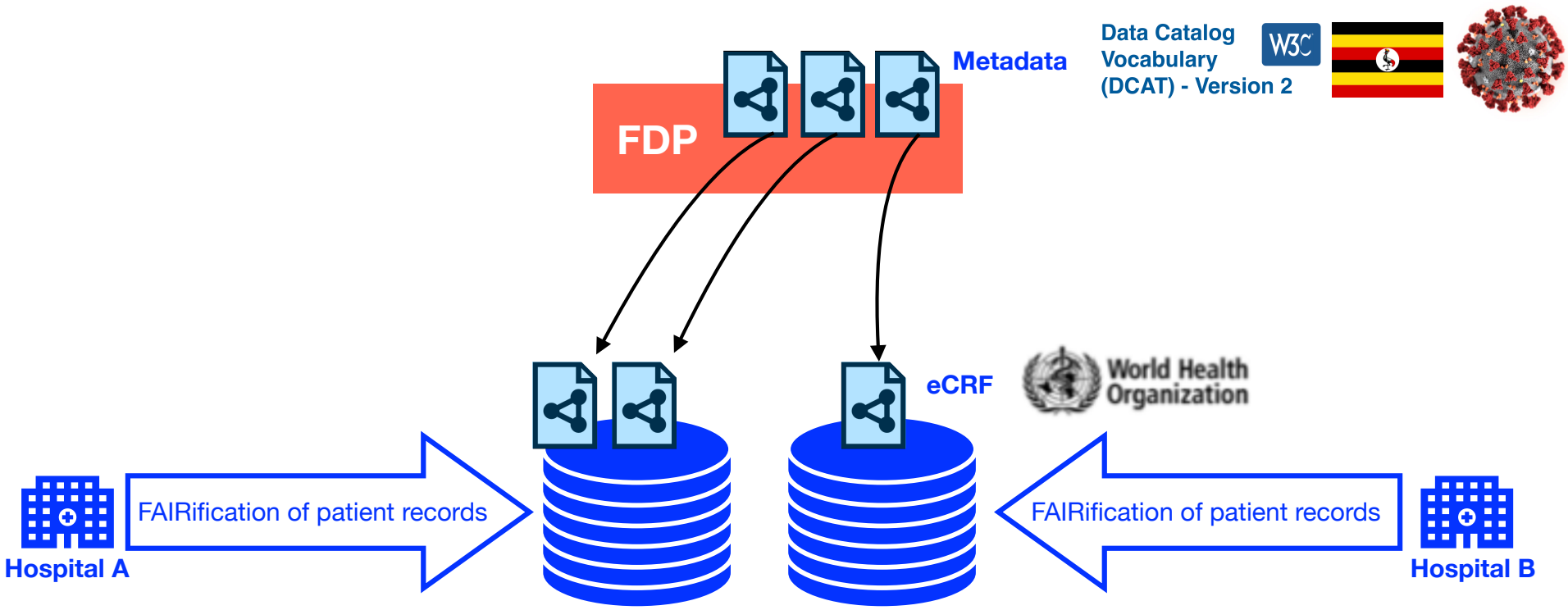
(Enables data visiting)

- Ethiopia: Addis Ababa University
- Ethiopia: Mekelle University
- Uganda: KIU
- Zimbabwe: GZU
- Nigeria: Data Science group
- Kenya: Westlands



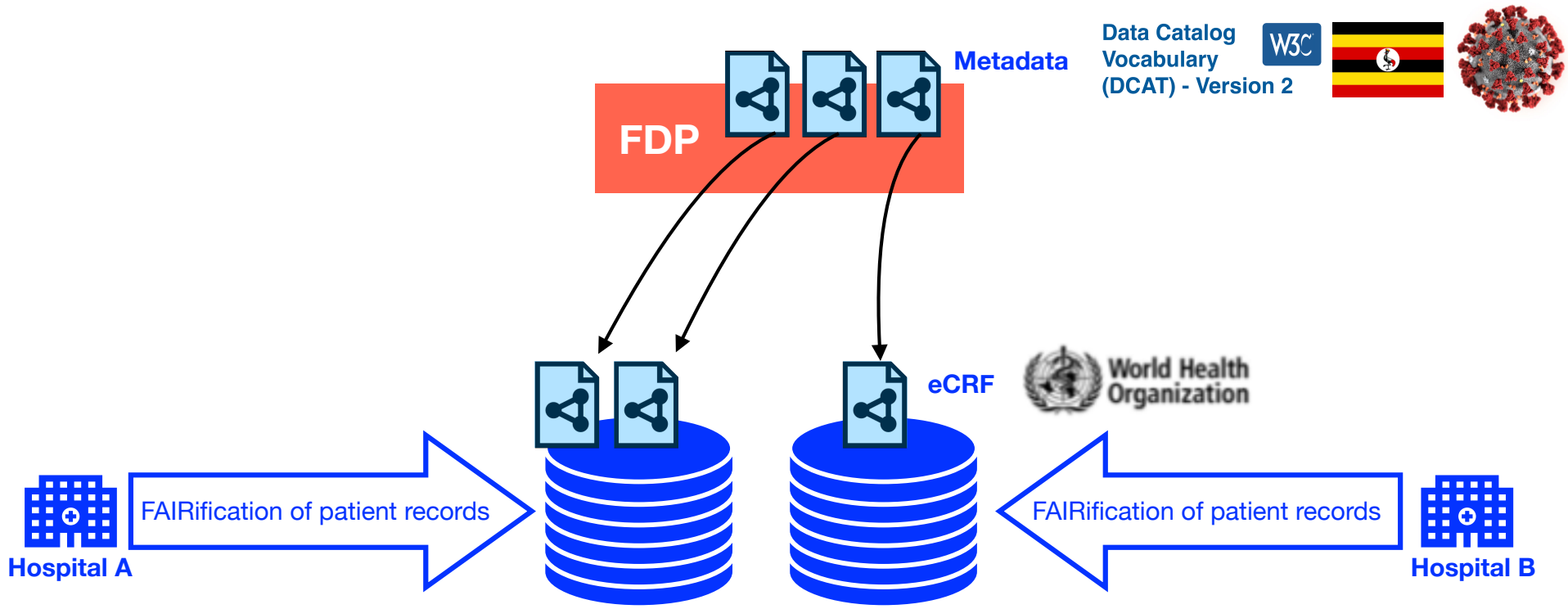
The VODAN FAIR Data Point

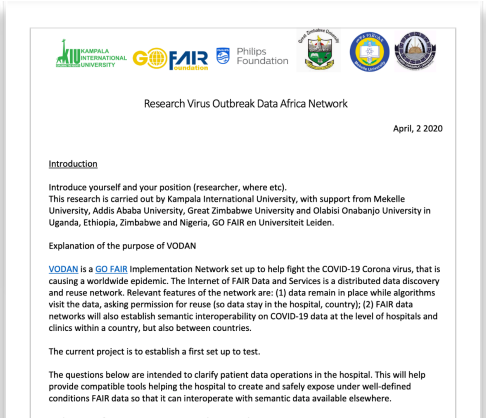
(Enables data visiting)



The VODAN FAIR Data Point

(Enables data visiting)





Research Virus Outbreak Data Africa Network

April, 2 2020

Introduction

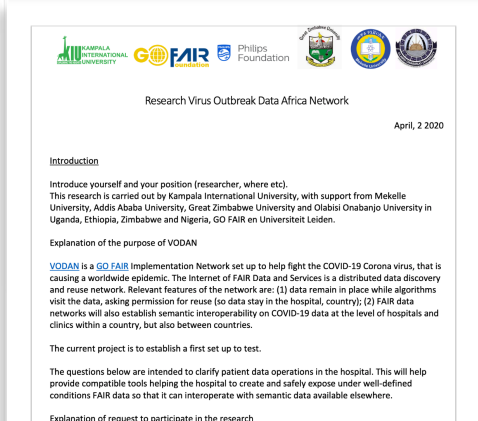
Introduce yourself and your position (researcher, where etc).
This research is carried out by Kampala International University, with support from Mekelle University, Addis Ababa University, Great Zimbabwe University and Olabisi Onabanjo University in Uganda, Ethiopia, Zimbabwe and Nigeria, GO FAIR en Universiteit Leiden.

Explanation of the purpose of VODAN

VODAN is a GO FAIR Implementation Network set up to help fight the COVID-19 Corona virus, that is causing a worldwide epidemic. The Internet of FAIR Data and Services is a distributed data discovery and reuse network. Relevant features of the network are: (1) data remain in place while algorithms visit the data, asking permission for reuse (so data stay in the hospital, country); (2) FAIR data networks will also establish semantic interoperability on COVID-19 data at the level of hospitals and clinics within a country, but also between countries.

The current project is to establish a first set up to test.

The questions below are intended to clarify patient data operations in the hospital. This will help provide compatible tools helping the hospital to create and safely expose under well-defined conditions FAIR data so that it can interoperate with semantic data available elsewhere.



Research Virus Outbreak Data Africa Network

April, 2 2020

Introduction

Introduce yourself and your position (researcher, where etc).
This research is carried out by Kampala International University, with support from Mekelle University, Addis Ababa University, Great Zimbabwe University and Olabisi Onabanjo University in Uganda, Ethiopia, Zimbabwe and Nigeria, GO FAIR en Universiteit Leiden.

Explanation of the purpose of VODAN

VODAN is a GO FAIR Implementation Network set up to help fight the COVID-19 Corona virus, that is causing a worldwide epidemic. The Internet of FAIR Data and Services is a distributed data discovery and reuse network. Relevant features of the network are: (1) data remain in place while algorithms visit the data, asking permission for reuse (so data stay in the hospital, country); (2) FAIR data networks will also establish semantic interoperability on COVID-19 data at the level of hospitals and clinics within a country, but also between countries.

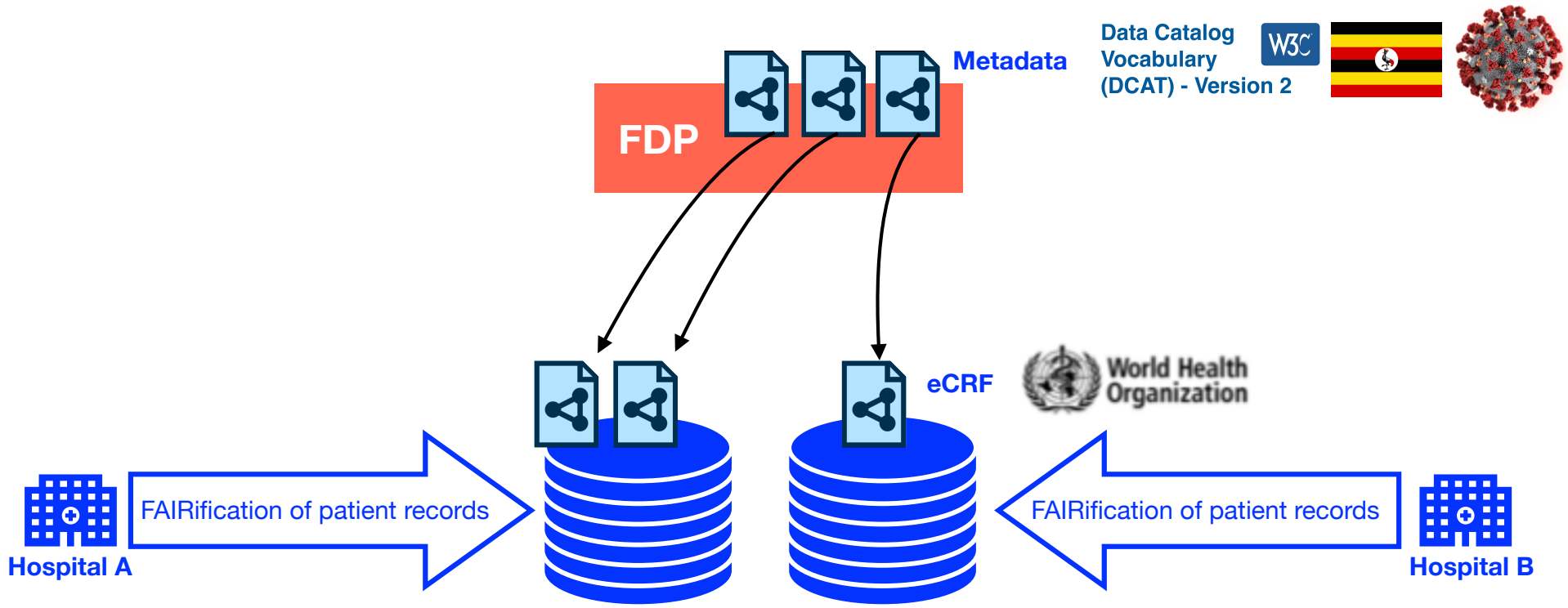
The current project is to establish a first set up to test.

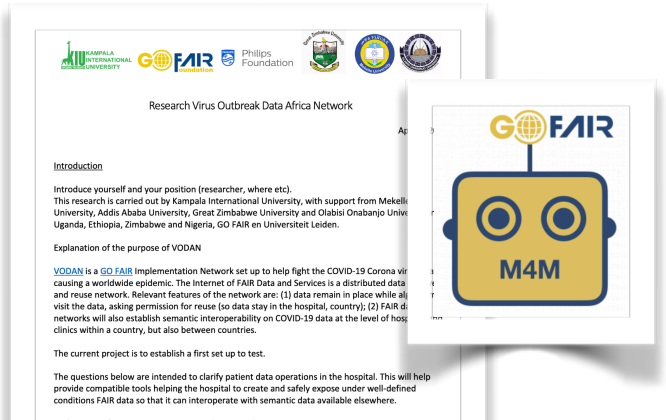
The questions below are intended to clarify patient data operations in the hospital. This will help provide compatible tools helping the hospital to create and safely expose under well-defined conditions FAIR data so that it can interoperate with semantic data available elsewhere.

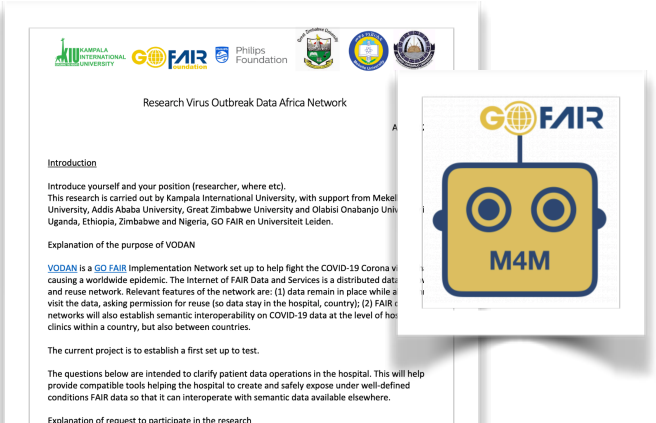
Explanation of request to participate in the research

The VODAN FAIR Data Point

(Enables data visiting)

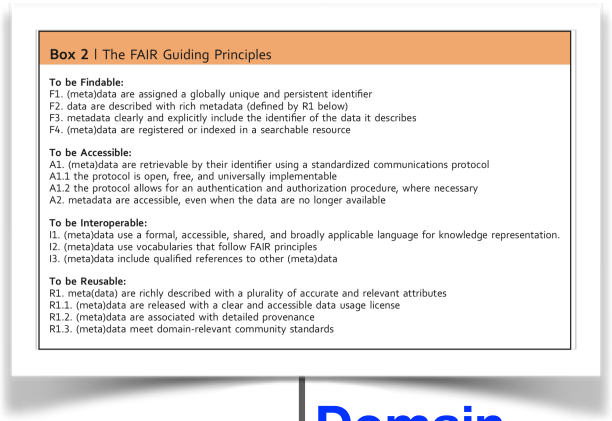






The FAIR Implementation Profile

FAIR Implementation Profiles



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

Domain Communities

Implementation **Choice**

Community chooses to re-use existing resources as needed to implement FAIR

A matrix of Communities x Resources

Implementation **Challenge**

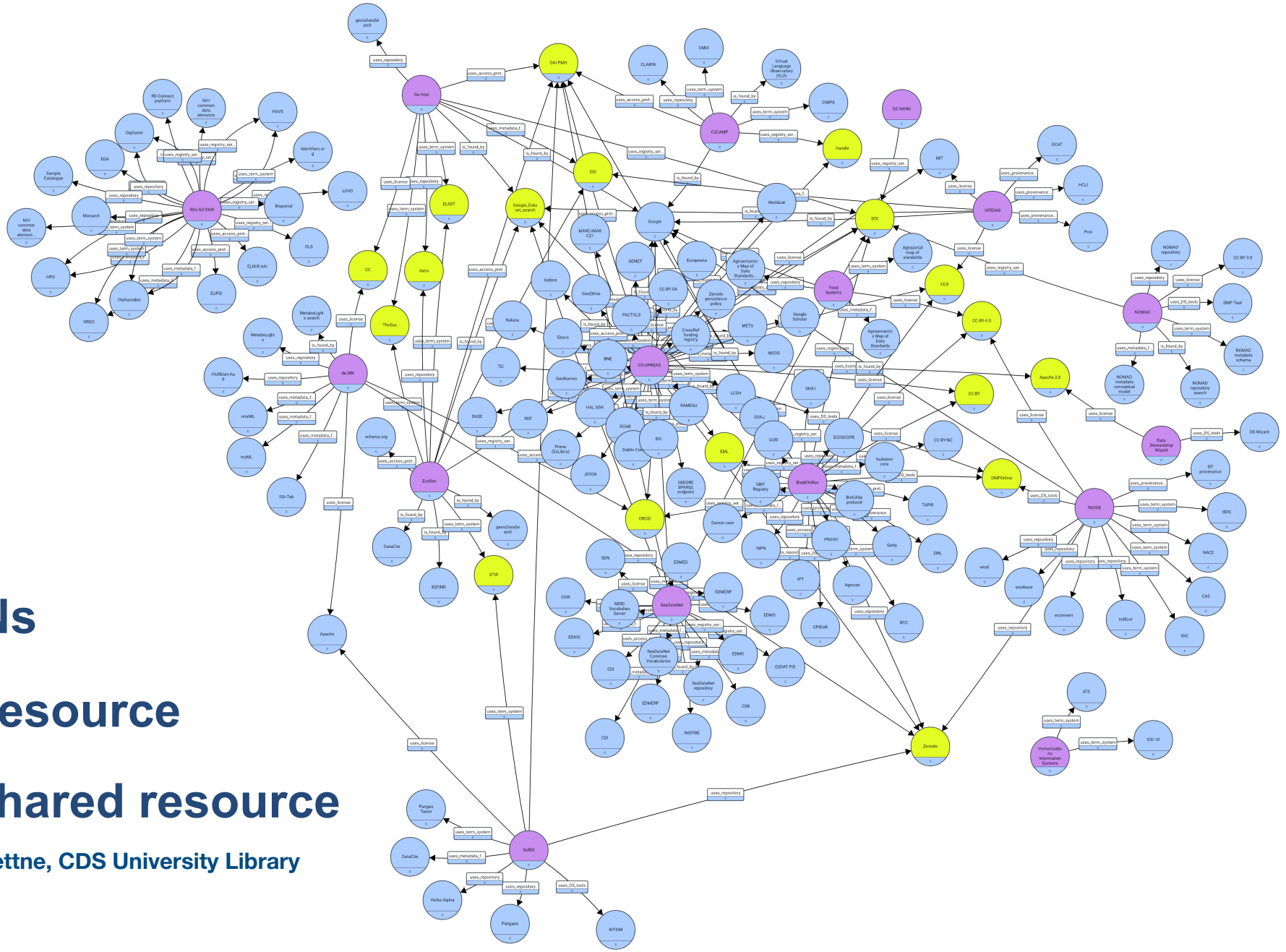
Community accepts challenge to create new resources as needed to implement FAIR *

Declaration / Registration

Optimal Reuse

** every "choice" was once a "challenge"*

FAIR Implementation Profiles



INs

Resource

Shared resource

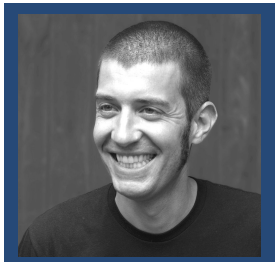
Kristina Hettne, CDS University Library

FAIR Convergence Matrix

Kristina Hettne

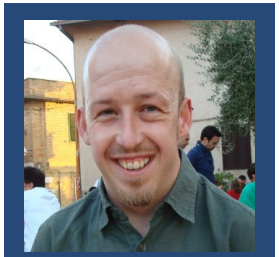
Barbara Magagna

Peter Wittenburg



Tobias Kuhn

Robert Pergl



Pete McQuilton

Universiteit Leiden

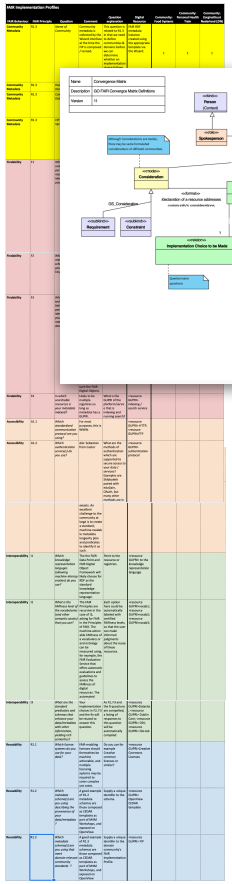
umweltbundesamt
ENVIRONMENT AGENCY AUSTRIA

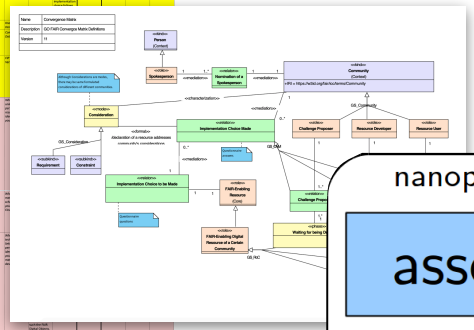
RDA

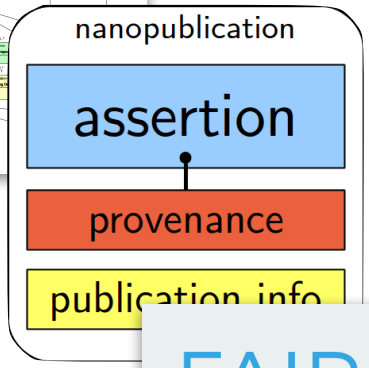
MAX-PLANCK-GESELLSCHAFT

VU

CZECH TECHNICAL UNIVERSITY IN PRAGUE



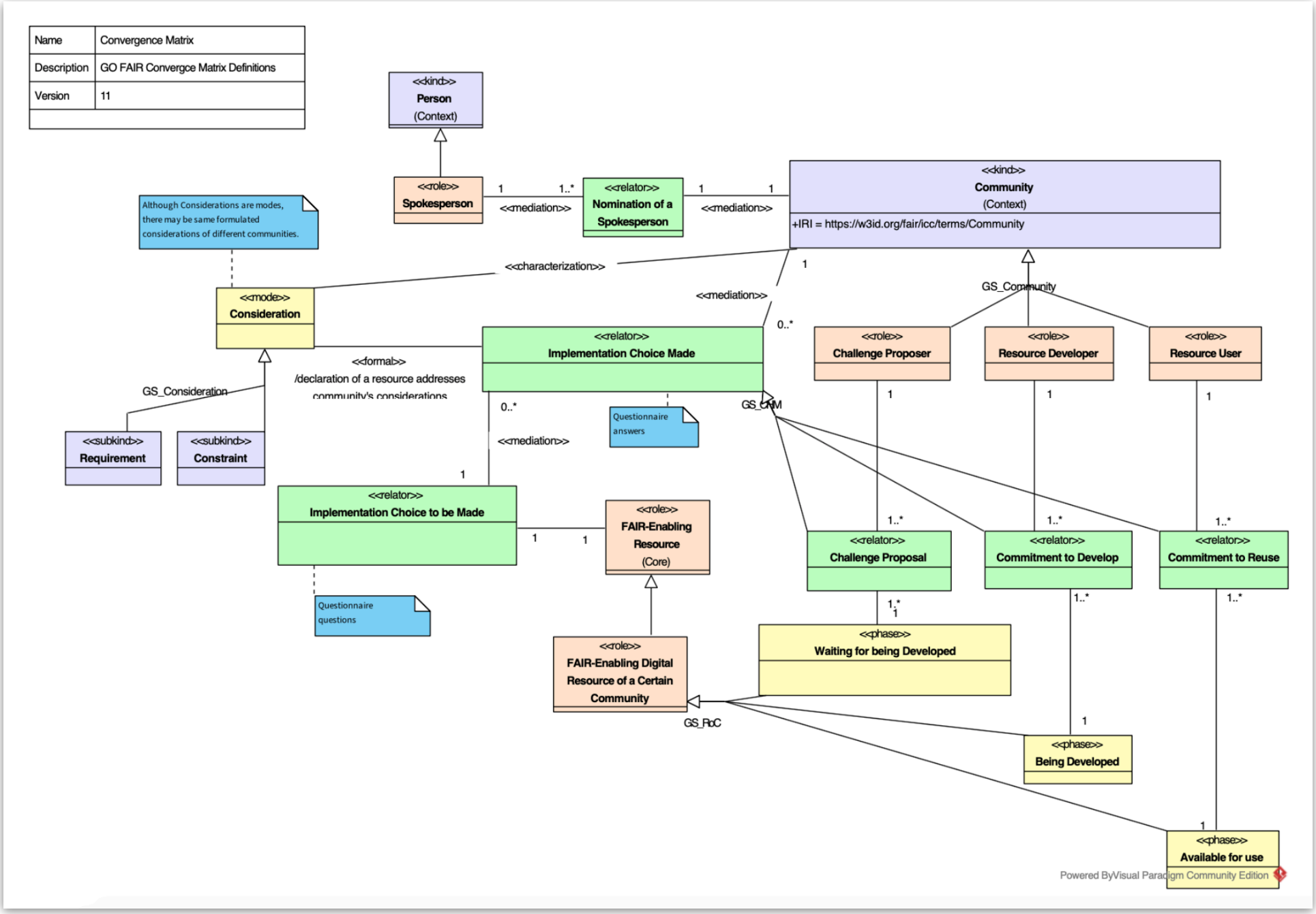




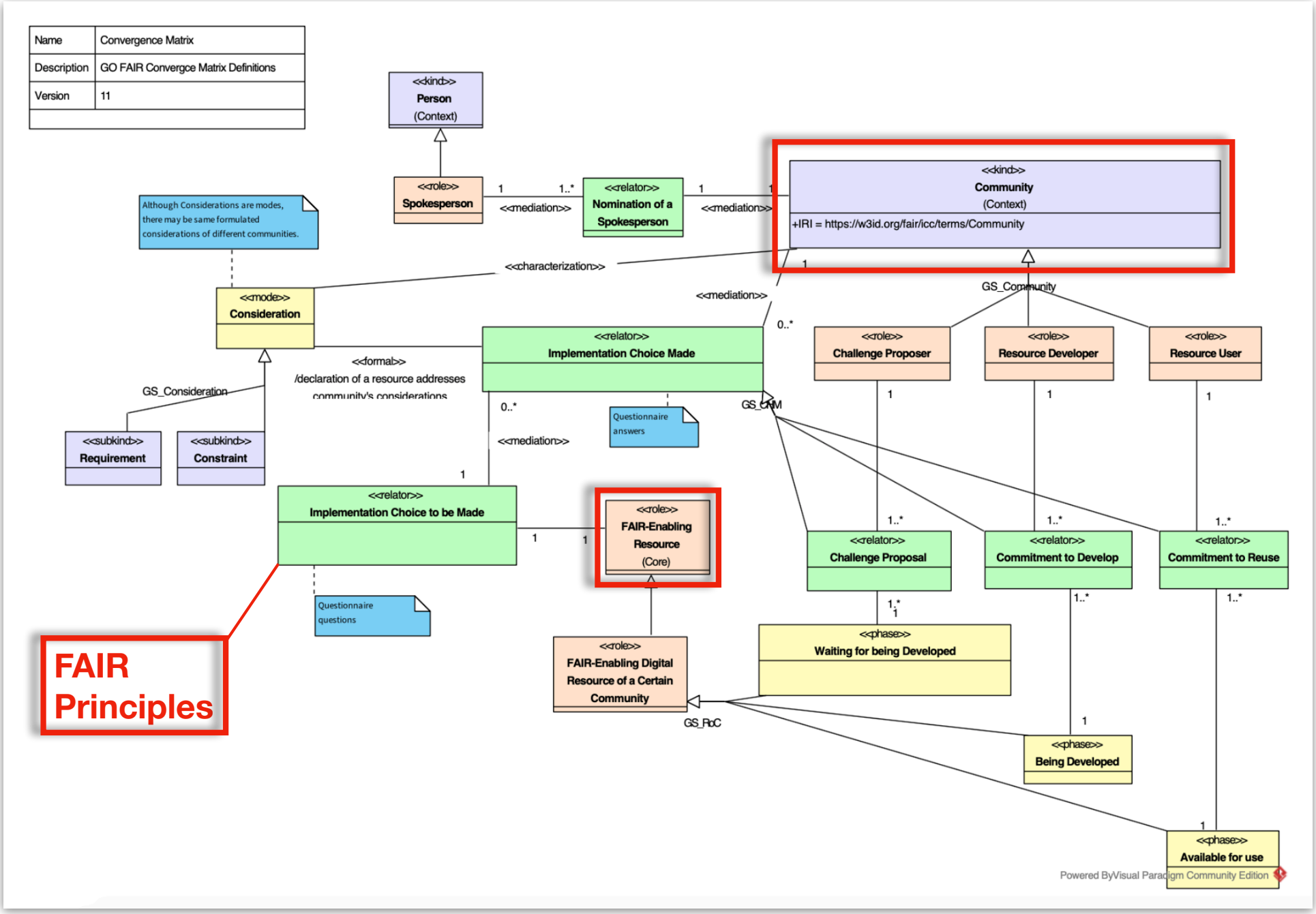
FAIRsharing.org
standards, databases, polic

DSW
DATA STEWARDSHIP WIZARD

Core ontology



Core ontology



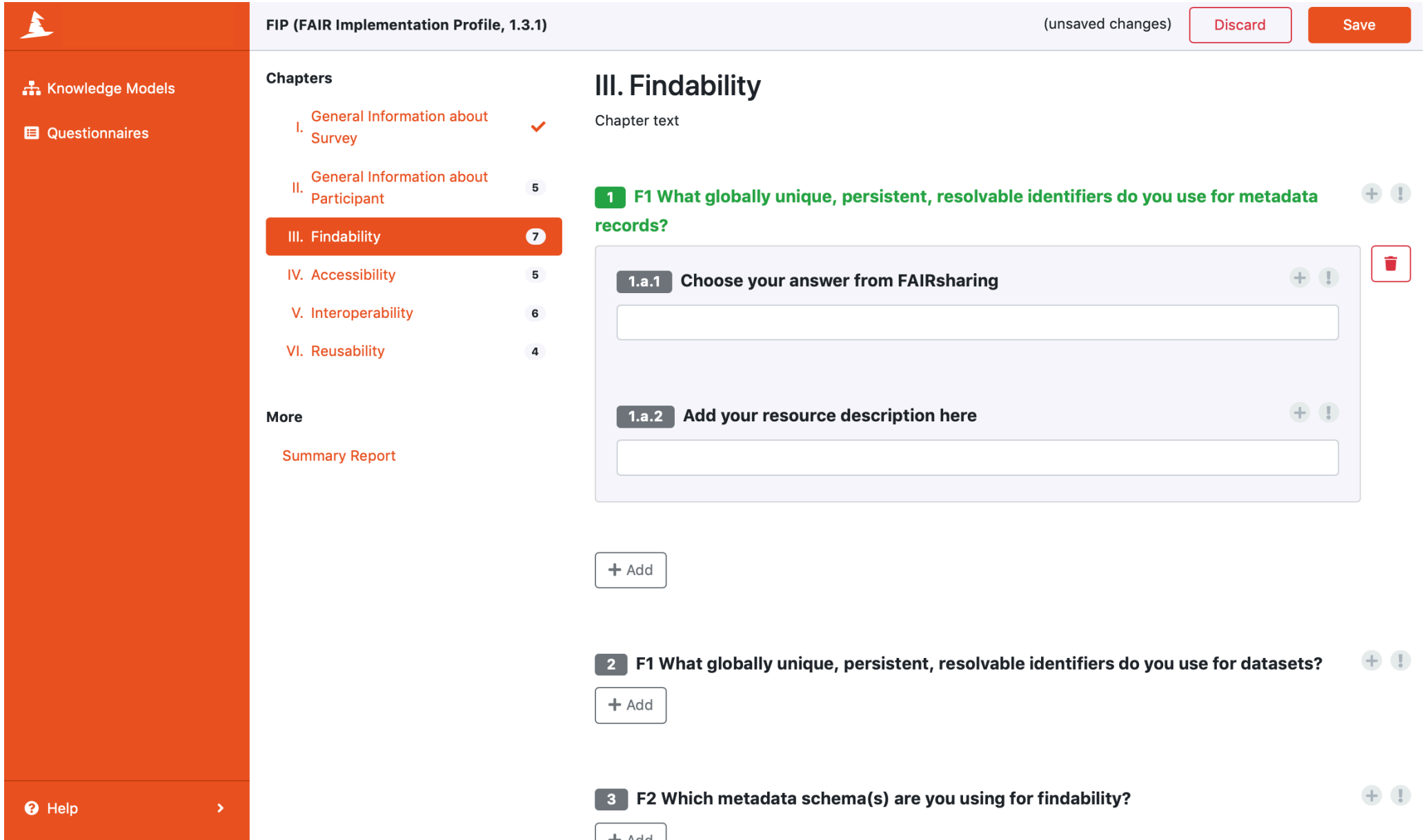
v3 Questionnaire

FAIR Implementation Profiles								
FAIR Behaviour	**FAIR Principle**	**Question**	**Comment**	**Question explanation**	**Digital Resource**	**Community: Food Systems**	**Community: Personal Health Train**	**Community: Longitudinal Nederland (LN)**
Community Metadata	R1.3	*Name of Community*	Community metadata is collected by the Wizard interface at the time the FIP is composed / revised.	This question is related to R1.3 in that we need to define communities & domains before we can determine whether an implementation choice follows "domain-relevant community standard".	FAIR FIP metadata instance created using the appropriate template via the Wizard.	1	1	1
Community Metadata	R1.3	*Knowledge Domain*	same as above	same as above	same as above	1	1	1
Community Metadata	R1.3	*Community Data Steward(s)*	same as above	The person / team who answers the questionnaire on behalf of the community.	same as above	1	1	1
Community Metadata	R1.3	*FIP Date / Version*	same as above	The FIP is a publishable entity and therefore has a publication date/time and is versionable. The Wizard interface tracks FIP provenance.	same as above	1	1	1
Findability	F1	*What globally unique, persistent, resolvable identifiers do you use?*	Different identifier service providers will offer different kinds of resolution services impacting F2 & I, and have different persistence policies impacting F1 and A2.	There is a long list of general purpose identifiers from Bioregistry. The Matrix will also track and list in this "row", any other identifiers that are used by other technology platforms that might be mentioned elsewhere in the Matrix (e.g., ORCID for people, GBIF for species, InChi for chemicals).	resource GUPRI= DOI (offered by Datacite and Crossref); resource GUPRI= DOI (offered by other organizations); resource GUPRI= general Handles; resource GUPRI= ePIC services; resource GUPRI=Local; resource GUPRI=Fuzzy URIs; resource GUPRI=Fuzzy Urlaps;			
Findability	F2	*Which data or metadata schema(s) are you using for Findability?*	A good example of F2 metadata schemas are those composed as CEDAR templates as part of MAM Workshops, and reposed on OpenView.	Supply a unique identifier to the schema.	resource GUPRI= OpenView CEDAR template			
Findability	F3	*What is the technology that ties the persistent identifiers of your data to the metadata description?*	This question is asking for what technology permanently links your data and metadata. The answers could range from metadata schema having predicates that qualify the relations and locations to technology platforms like repositories and FAIR Data Points that guarantee the permalink to infrastructures such the FAIR Digital Objects.	In general if PIDs are assigned they should be findable in the metadata as a separate field, if possible, provide an example.	resource GUPRI= metadata schema; resource GUPRI= a FAIR Data Point; resource GUPRI= FDP			
Findability	F4	*In which searchable resources is your metadata indexed?*	Likely to be multiple registries as long as metadata has a GUPRI.	What is the GUPRI of the platform/service that is indexing and running search?	resource GUPRI= Indexing / search service			
Accessibility	A1.1	*Which standardized communication protocol are you using?*	For most purposes, this is WWW.		resource GUPRI= HTTP; resource GUPRI=FTP			
Accessibility	A1.2	*Which authentication service(s) do you use?*	Ask Sebastian from Castor	What are the methods of authentication which are supported to secure access to your data / services? Examples are Bridgeteeth paired with Bridgeteeth, Oliveth, but many other methods are in	resource GUPRI= authentication protocol			
			exists. An excellent challenge to the community at large is to create a standard, machine-readable metadata property (ideally) and predicates to identify it as such.					
Interoperability	I1	*Which knowledge representation language (allowing machine-interpretation) do you use?*	The FAIR Data Point and FAIR Digital Object Framework will likely choose for RDF as the standard knowledge representation language.	Point to the resource or registries.	resource GUPRI= to the knowledge representation language			
Interoperability	I2	*What is the FAIRness level of the vocabularies (and other semantic assets) that you use?*	The FAIR Principles are recursive in the case of I2, asking for FAIR in the Principles of FAIR. The machine-action of the FAIRness of a vocabulary or and ontology can be measured using, for example, the FAIR Evaluation Service that offers automatic evaluations and guidelines to assess the FAIRness of digital resources. The automated	Each option here could be automatically labeled with certified FAIRness levels, so that the user can make informed judgments about the reuse of these resources.	resource GUPRI=vocab1; resource GUPRI=vocab2; resource GUPRI=vocab3;			
Interoperability	I3	*What are the standard predicates and schemas that enhance your data/metadata with other information, yielding rich semantics?*	Your implementation choices in F2, F3 and the Rs will be tested to answer this question.	As F2, F3 and the R questions are completed, a listing of responses to this question will be automatically compiled.	resource GUPRI=Datacite; resource GUPRI= DOI; resource GUPRI= DOI; resource GUPRI= RA-tab			
Reusability	R1.1	*Which license systems do you use for your data?*	Skill-enabling licenses should themselves be machine-actionable, and multiple licensing systems may be required to cover complex use cases.	Do you use for example Creative common licenses or similar?	resource GUPRI= Creative Commons Licenses			
Reusability	R1.2	*Which metadata schema(s) are you using describing the provenance of your data/metadata?*	A good example of R1.2 metadata schemas are those composed as CEDAR templates as part of MAM Workshops, and reposed on OpenView.	Supply a unique identifier to the schema.	resource GUPRI= OpenView CEDAR template			
Reusability	R1.3	*Which metadata schema(s) are you using that meet domain-relevant community standards ?*	A good example of R1.3 metadata schemas are those composed as CEDAR templates as part of MAM Workshops, and reposed on OpenView.	Supply a unique identifier to the domain community's Skill Implementation Profile.	resource GUPRI= FIP			

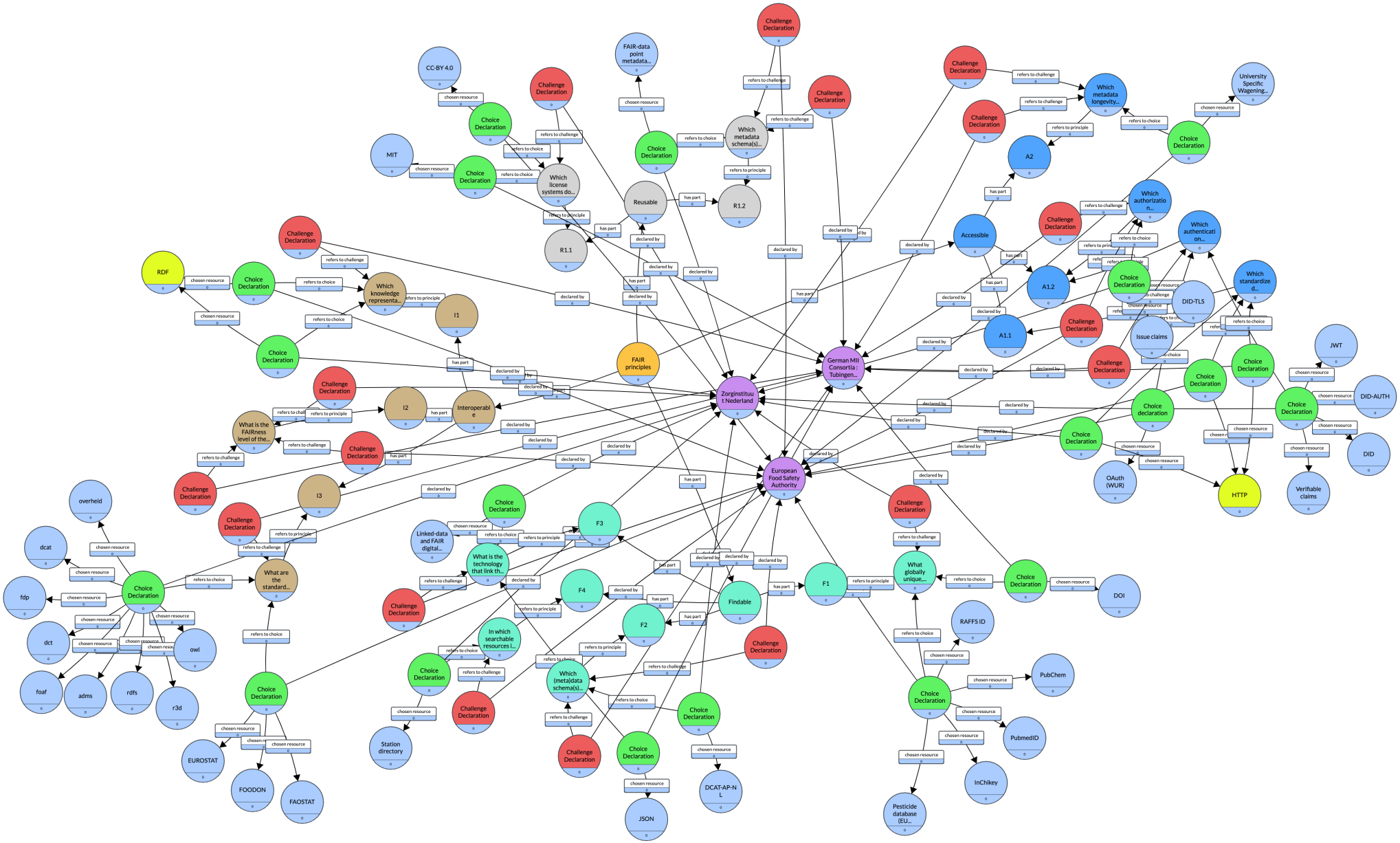
FAIR Convergence Matrix

https://fair-matrix.ds-wizard.org/

Email: “guest@example.com”
Password: “guest"

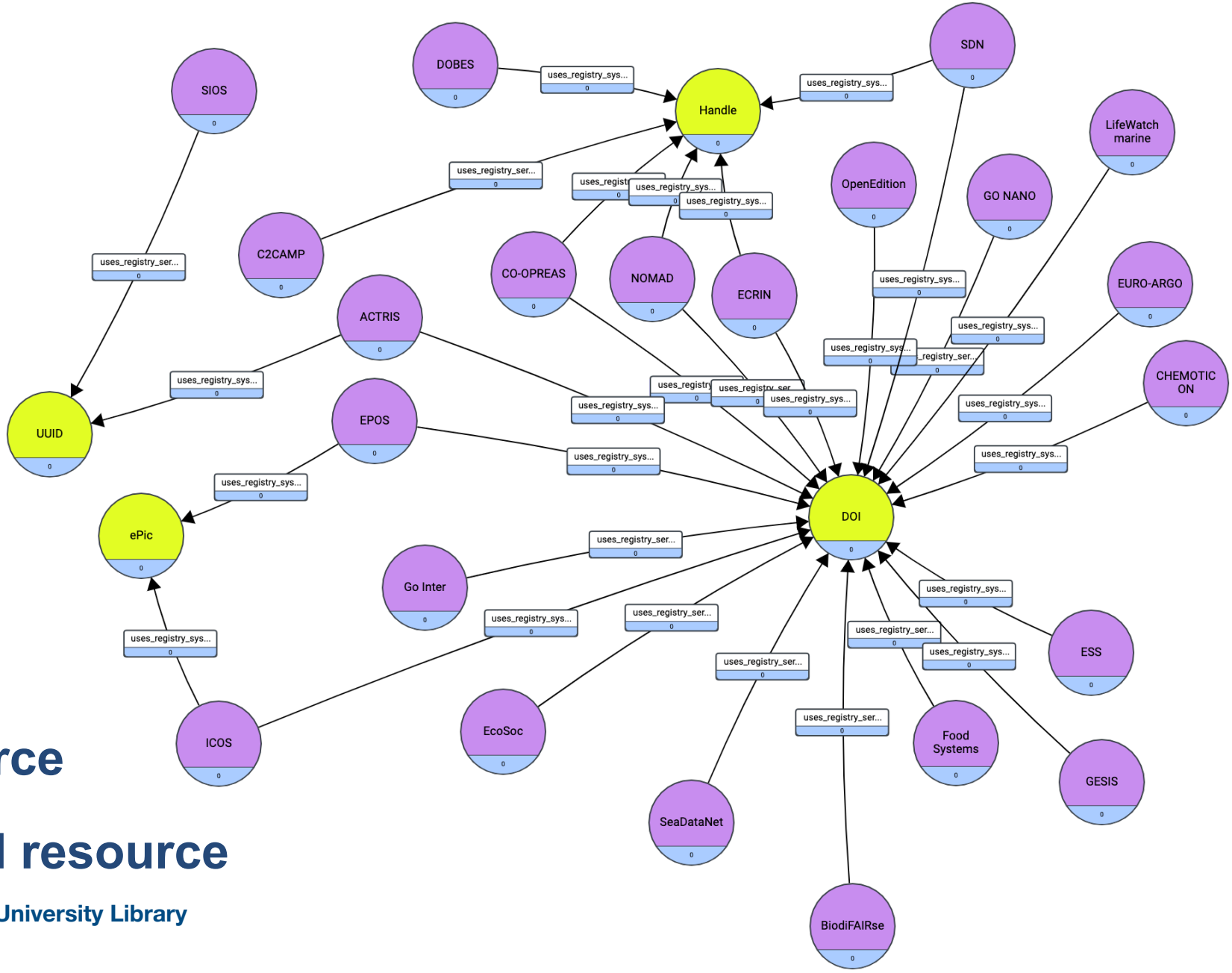


3 Communities devoted to distributed learning



Barbra Magagna, Umweltbundesamt GmbH
Kristina Hettne, CDS University Library

FIPs for 32 European Research Infrastructures



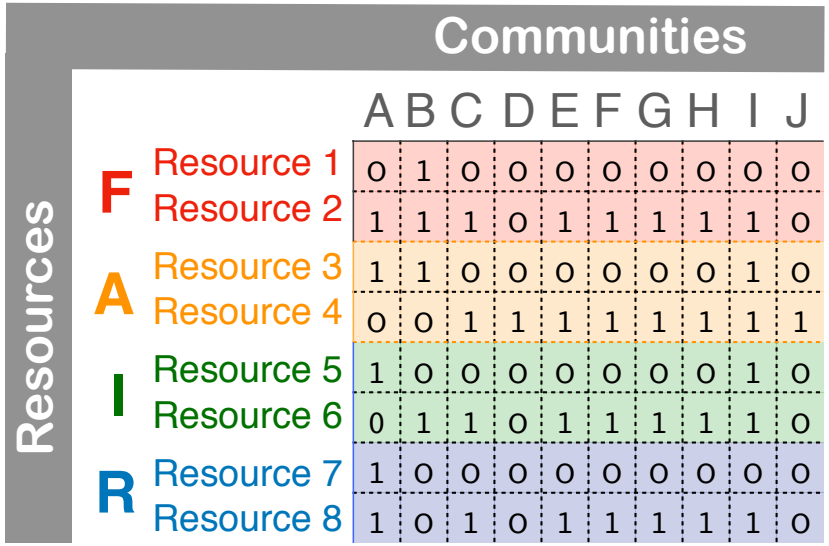
INs

Resource

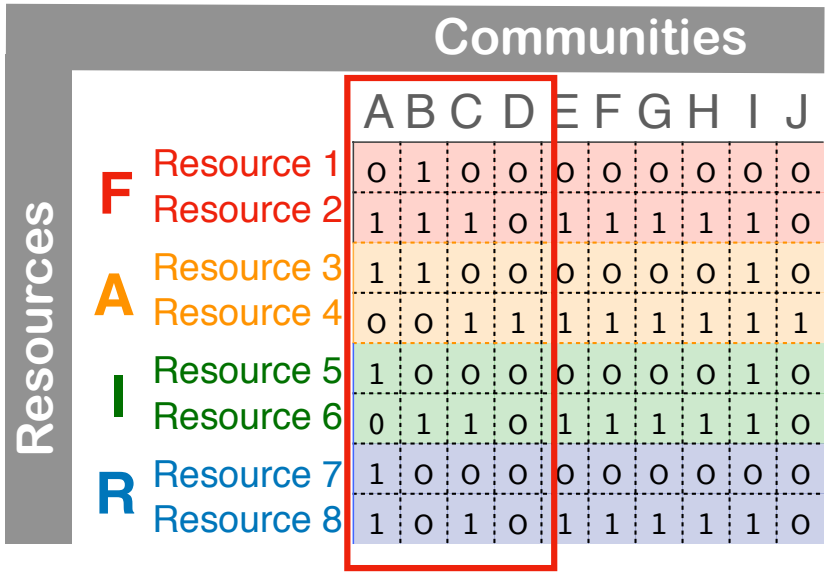
Shared resource

Kristina Hettne, CDS University Library

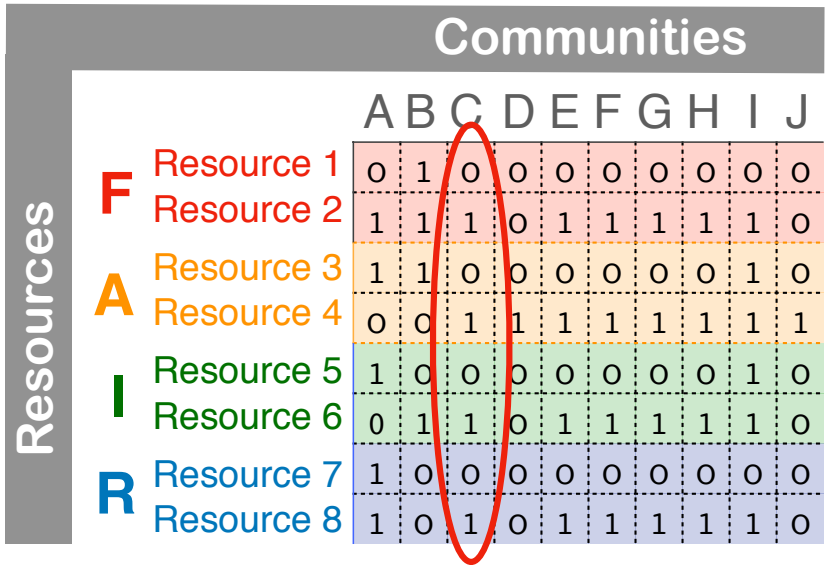
FAIR Implementation Profiles



FAIR Implementation Profiles

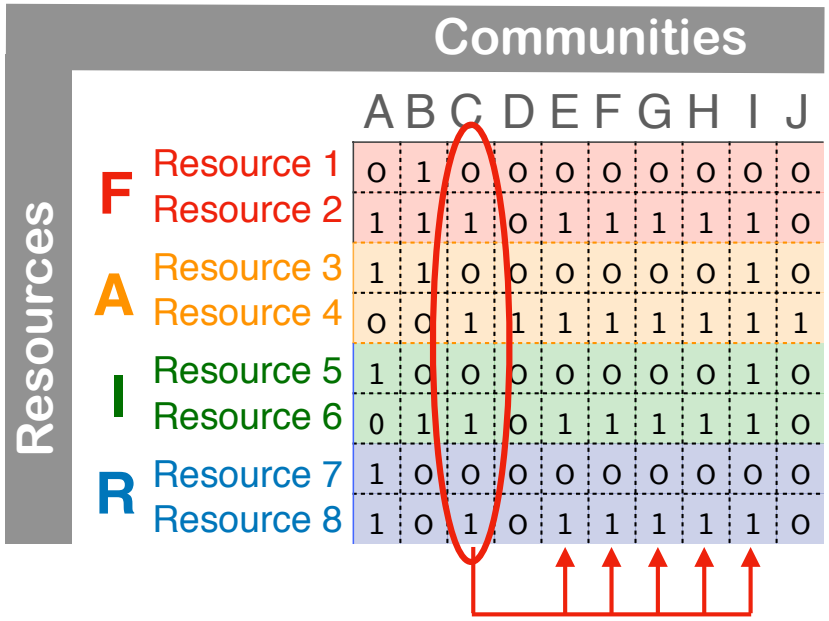


FAIR Implementation Profiles



FAIR Implementation Profile for Community C = {0,1,0,1,0,1,0,1}

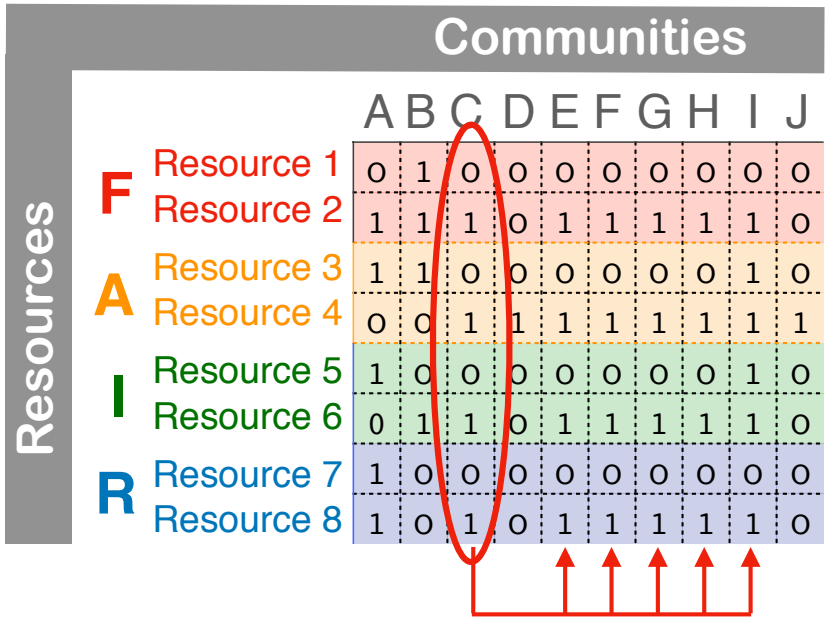
FAIR Implementation Profiles



FAIR Implementation Profile for Community C = {0,1,0,1,0,1,0,1}

Community E-I have reused the FIP of (trusted) Community C

FAIR Implementation Profiles



FAIR Implementation Profile for Community C = {0,1,0,1,0,1,0,1}

Community E-I have reused the FIP of (trusted) Community C

- **Save time on FAIR implementation**
- **Maximize reuse**
- **Maximize interoperation**

DATA Intelligence | CN: 10-1626/G2
ISSN: 2096-7004

Published • Versions 3 | Vol 2 (1) : 158–170 2019

Download

FAIR Convergence Matrix: Optimizing the Reuse of Existing FAIR-Related Resources

Hana Pergl Sustkova, Kristina Maria Hettne, Peter Wittenburg, Annika Jacobsen, Tobias Kuhn, Robert Pergl, Jan Slifka, Peter McQuilton, Barbara Magagna, Susanna-Assunta Sansone, Markus Stocker, Melanie Imming, Larry Lannom, Mark Musen, Erik Schultes hidden

DOI: 10.1162/dint_a_00038

439 | 43 | 0

Abstract & Keywords

Abstract: The FAIR Principles articulate the behaviors expected from digital artifacts that are Findable, Accessible, Interoperable and Reusable by machines and by people. Although by now widely accepted, the FAIR Principles by design do not explicitly consider actual implementation choices enabling FAIR behaviors. As different communities have their own, often well-established implementation preferences and priorities for data reuse, coordinating a broadly accepted, widely used FAIR implementation approach remains a global challenge. In an effort to accelerate broad community convergence on FAIR implementation options, the GO FAIR community has launched the development of the FAIR Convergence Matrix. The Matrix is a platform that compiles for any community of practice, an inventory of their self-declared FAIR implementation choices and challenges. The Convergence Matrix is itself a FAIR resource, openly available, and encourages voluntary participation by any self-identified community of practice (not only the GO FAIR Implementation Networks). Based on patterns of use and reuse of existing resources, the Convergence Matrix supports the transparent derivation of strategies that optimally coordinate convergence on standards and technologies in the emerging Internet of FAIR Data and Services.

Keywords: FAIR Implementation Choices and Challenges; Convergence; FAIR Communities

Acknowledgments

References

DATA Intelligence | CN: 10-1626/G2 ISSN: 2096-7004

Published Issues Submit Paper Subscribe

Published • Versions 3 | Vol 2 (1) : 158–170 2019

Download

FAIR Convergence Matrix: Optimizing the Reuse of Existing FAIR-Related Resources

Hana Pergl Sustkova, Kristina Maria Hettne, Peter Wittenburg, Annika Jacobsen, Tobias Kuhn, Robert Pergl, Jan Slifka, Peter McQuilton, Barbara Magagna, Susanna-Assunta Sansone, Markus Stocker, Melanie Imming, Larry Lannom, Mark Musen, Erik Schultes hidden

DOI: 10.1162/dint_a_00038

439 | 43 | 0

Abstract & Keyword

Abstract: The FAIR Principl and Reusable by machines consider actual implementat implementation preferences approach remains a global the GO FAIR community has any community of practice, Matrix is itself a FAIR reso practice (not only the GO Convergence Matrix suppor technologies in the emerging

Keywords: FAIR Implementa

Acknowledgments

References

Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence

Erik Schultes
GO FAIR International Support and Coordination Office, Leiden, The Netherlands
ORCID 0000-0001-8888-635X

Version 2.0 DRAFT: DO NOT CIRCULATE
December 11, 2019
On OSF: https://osf.io/8sv5f/

Abstract

There are now powerful incentives driving the adoption of FAIR practices among a broad cross section of public and private sector organizations[1] . This adoption process requires considerations on both domain-specific and infrastructural resources. Together, these considerations compose a potentially very large number of decisions on the use and reuse of standards and technologies. Many of the relevant factors in these considerations are likely to exceed the knowledge and experience of individual domain specialists and data stewards. For communities adopting FAIR practices, the collection of their implementation choices compose their *FAIR Implementation Profile* (FIP). The FIPs of numerous communities can be systematically acquired from the FAIR Convergence Matrix[2] and used as the basis to optimize the reuse of existing resources and interoperation within and between domains. Ready-made and well-tested FIPs created by trusted

What is the FIP for your repository ?

Data Service Portal Aila, SND, DataDOI, SEAD, UiT Open Research Data Dataverse, CESSDA DC, CLARIN-DK, CLARINO Bergen Center repository, SprÃ¥kbanken, TROLLing, tekstlab, NIRD Archive, ICOS, DDA, Kansanperinteen arkisto (Folklive archives), QsarDB, JYX, Kielipankki, Bolin Centre Database, GERDA, GEUS, HPA, DNBC, UNITE, PlutoF, FinBIF, CELR META-SHARE, MIDAS, SSRI, DTU data

FAIR Convergence Matrix

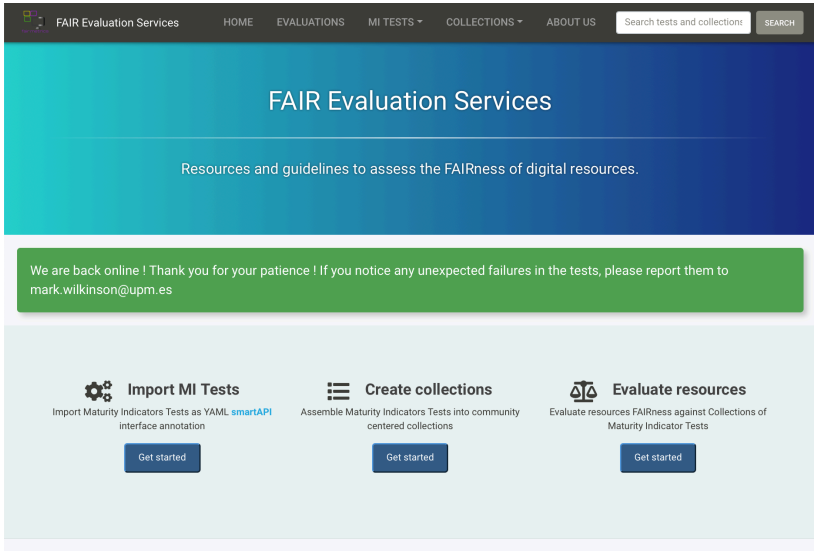
https://fair-matrix.ds-wizard.org/

Email: “guest@example.com”
Password: “guest"

What is the FIP for your repository ?

Data Service Portal Aila, SND, DataDOI, SEAD, UiT Open Research Data Dataverse, CESSDA DC, CLARIN-DK, CLARINO Bergen Center repository, SprÃ¥kbanken, TROLLing, tekstlab, NIRD Archive, ICOS, DDA, Kansanperinteen arkisto (Folklive archives), QsarDB, JYX, Kielipankki, Bolin Centre Database, GERDA, GEUS, HPA, DNBC, UNITE, PlutoF, FinBIF, CELR META-SHARE, MIDAS, SSRI, DTU data

FAIR Convergence Matrix

https://fair-matrix.ds-wizard.org/

Email: “guest@example.com”
Password: “guest"