Leiden University
Medical Center

Empowering the use of registries by applying FAIR principles: lessons learned and future possibilities

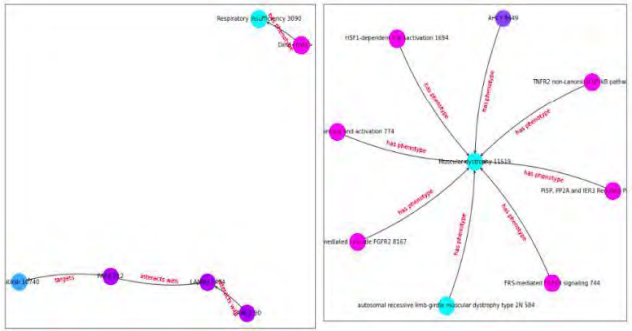
Optimierung der Nutzbarkeit von Daten für Menschen, die mit einer (seltenen) Krankheit leben

Marco Roos
Human Genetics Department, LUMC

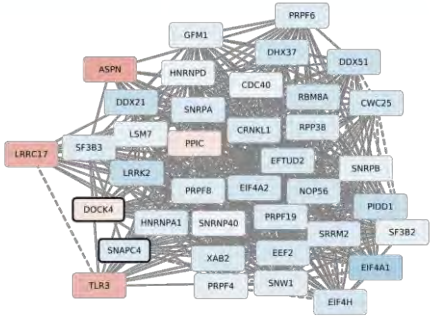
BERLIN, MAI 9, 2023

Why FAIR? – AI/ML/Data integration for rare diseases

Preliminary results kindly made accessible by Pablo Perdomo, Carmen Reep, Daphne Wijnbergen, Núria Queralt-Rosinach, Eleni Mina, Annika Jacobsen, Katy Wolstencroft



Ontological hypotheses generated from DMD knowledge graph of multiple sources by ML algorithm

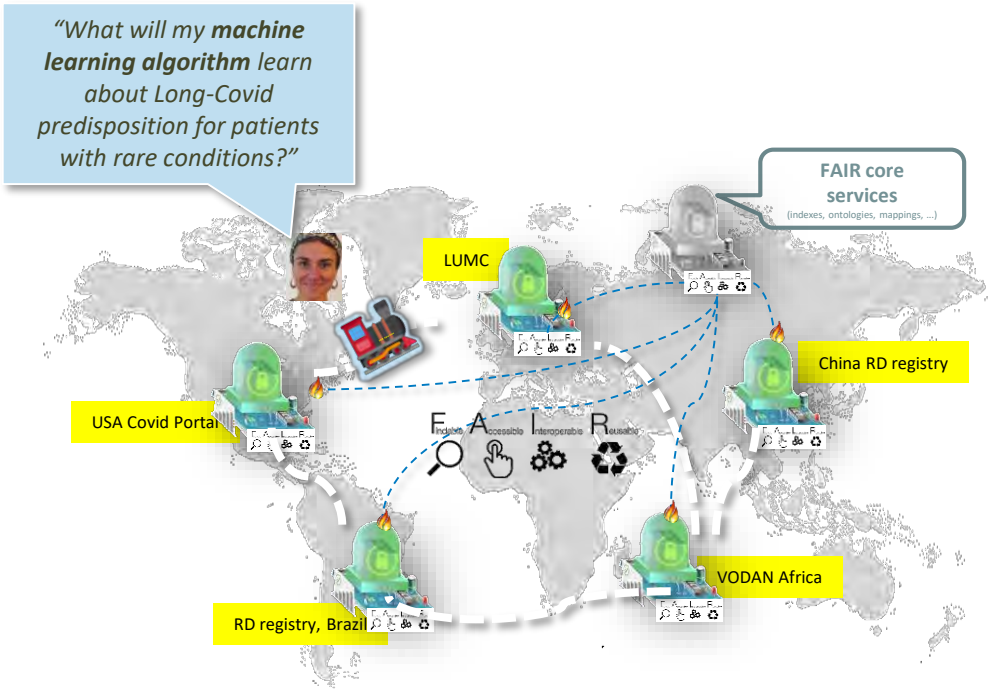


Subnetwork extracted from IBM transcriptomics networks & additional knowledge characterized by ‘RNA splicing’

drug URI	drug label
https://identifiers.org/chembl:CHEMBL29097	CHEMBL29097
https://identifiers.org/chembl:CHEMBL8260	BAICALEIN
https://identifiers.org/chembl:CHEMBL221137	EMBELIN
https://identifiers.org/chembl:CHEMBL267345	AMPHOTERICIN B
https://identifiers.org/chembl:CHEMBL308688	5,7-DIMETHOXYISOFLAVONE
https://identifiers.org/chembl:CHEMBL2110660	IGMESINE
https://identifiers.org/chembl:CHEMBL275809	FR-122047
https://identifiers.org/chembl:CHEMBL161343	ARACHIDONOYL GLYCINE
https://identifiers.org/chembl:CHEMBL585	TRIAMTERENE
https://identifiers.org/chembl:CHEMBL1269845	CHEMBL1269845

Prioritised drug targets for HD learned from knowledge graph

Globally Find, Access, Interoperate, & legitimately Reuse data for data science for patient benefit



Outline

- Motivation to making registry data more findable, accessible, interoperable, and reusable for computational use and short introduction machine understandable Linked Data
- FAIR implementation effort Virtual Platform European Joint Programme Rare Diseases
- Next challenge

Where is your balance?

Effort on Data
“Having data”

Effort on Usability
“Reuse by others”

A question

- How many types of treatment are given to patients who are experiencing “difficulty in swallowing” (Dysphagia) as a symptom around the world?
  - …in the TMF auditorium?

How long would it take to get the answer?

- ❑ Seconds
- ❑ Weeks or months
- ❑ Years
- ❑ Forever

Is this our balance?

- [x] Weeks or months
- [x] Years

Effort on Usability
“Reuse by others”

Effort on Data
“Having data”

Data linking for machines in action

If your (quality of) life depended on other people's data

Any resemblance with real people is purely coincidental

Disclaimer: mock data!!!!

Patient data in four registries

Each contains Common Data Elements
Disease, Symptom, Treatment (if any)

Given: one patient's treatment is a candidate treatment for Monica

What is the candidate treatment for Monica?

EJP RD

Leiden University Medical Center

This Training has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 825575 - European Joint Programme on Rare Diseases

Game: if your life depended on other people's data

Any resemblance with real people is purely coincidental

Disclaimer: mock data!!!!

Patient data in four registries

Each contains Common Data Elements

Disease, Symptom, Treatment (if any)

Given: one patient's treatment is a candidate treatment for Monica

What is the candidate treatment for Monica?

EJP RD

Leiden University Medical Center

This Training has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 825575 - European Joint Programme on Rare Diseases

Find the treatment experiment

Monika 	Annika 	Rajaram 	Pietro 
Krankheit Ringbildung Chromosom 14, Salaam-Anfälle, (Keine Behandlung)	Ring-14-sjúkumynd, sankta Vitusar dansur, eingin viðgerð	பெர்ரி நோய்க்குறி, வலிப்பு தாக்குதல்கள், லாமோட்ரைஜின்	sindrome Perry, sbalzi d'umore estremi, ossalato

Leiden University Medical Center

This Training has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 825575 - European Joint Programme on Rare Diseases

Find the treatment experiment

Monika 	Annika 	Rajaram 	Pietro 
Krankheit Ringbildung Chromosom 14, Salaam-Anfälle, (Keine Behandlung)	Ring-14-sjúkumynd, sankta Vitusar dansur, eingin viðgerð	பெர்ரி நோய்க்குறி, வலிப்பு தாக்குதல்கள், லாமோட்ரைஜின்	sindrome Perry, sbalzi d'umore estremi, ossalato
Ring-14 disease, Salaam seizures, (no treatment)	Ring-14 syndrome, Chorea, (no treatment)	Perry syndrome, Epileptic attacks, lamotrigine	Perry syndrome, extreme mood swings, oxalate

EJP RD

Leiden University Medical Center

This Training has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 825575 - European Joint Programme on Rare Diseases

Find the treatment experiment

Machine understandability starts with human agreement on the semantics
standards follow from there (Ontologies, OMOP, FHIR, etc.)

Monika 	Annika 	Rajaram 	Pietro 
Krankheit Ringbildung Chromosom 14, Salaam-Anfälle, (Keine Behandlung)	Ring-14-sjúkumynd, sankta Vitusar dansur, eingin viðgerð	பெர்ரி நோய்க்குறி, வலிப்பு தாக்குதல்கள், லாமோட்ரைஜின்	sindrome Perry, sbalzi d'umore estremi, ossalato
Monika ***has disease*** Ring-14 disease, and ***has phenotype*** Salaam seizures	Annika ***has disease*** Ring-14 disease, and ***has phenotype*** Chorea	Rajaram ***has disease*** Perry syndrome, and ***has phenotype*** Epileptic seizures. Epileptic seizures ***are treated by*** lamotrigine	Pietro ***has disease*** Ring-14 disease, and ***has phenotype*** Extreme mood swings. Extreme mood swings ***are treated by*** the drug Oxalate

EJP RD

Leiden University Medical Center

This Training has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 825575 - European Joint Programme on Rare Diseases

Find the treatment experiment

Translate to ontological model with globally unique IDs (URIs)

Monika	Annika	Rajaram	Pietro
Monika ***has disease*** Ring-14 disease, and ***has phenotype*** Salaam seizures	Annika ***has disease*** Ring-14 disease, and ***has phenotype*** Chorea	Rajaram ***has disease*** Perry syndrome, and ***has phenotype*** Epileptic seizures. Epileptic seizures ***are treated by*** lamotrigine	Pietro ***has disease*** Ring-14 disease, and ***has phenotype*** Extreme mood swings. Extreme mood swings ***are treated by*** the drug Oxalate
<_> ***obo:OGMS_0000031*** ordo:Orphanet_1440 ***obo:RO_0002452*** obo:HP_0011097.	<_> ***obo:OGMS_0000031*** ordo:Orphanet_1440, ***obo:RO_0002452*** obo:HP_0002072.	<_> ***obo:OGMS_0000031*** ordo:Orphanet_178509, ***obo:RO_0002452*** obo:HP_0011097 ***obo:RO_0002302*** obo:CHEBI_33237	<_> ***obo:OGMS_0000031*** ordo:Orphanet_178509, ***obo:RO_0002452*** obo:HP_0000720 ***obo:RO_0002302*** obo:CHEBI_132952

obo: http://purl.obolibrary.org/obo/
ordo: http://www.orpha.net/ORDO/

Unambiguous machine understandable semantics for:
Person ***has disease*** Disease,
Person ***has phenotype*** Phenotype,
Phenotype ***is treated by*** Treatment

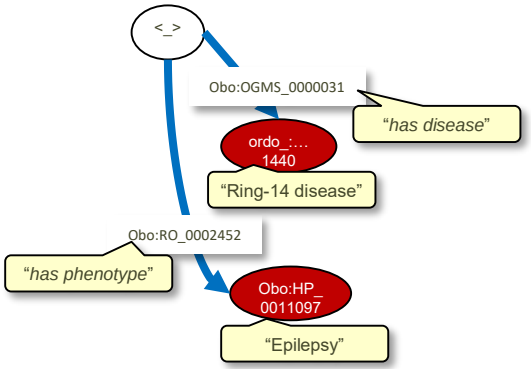
EJP RD

Leiden University Medical Center

This Training has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 825575 - European Joint Programme on Rare Diseases

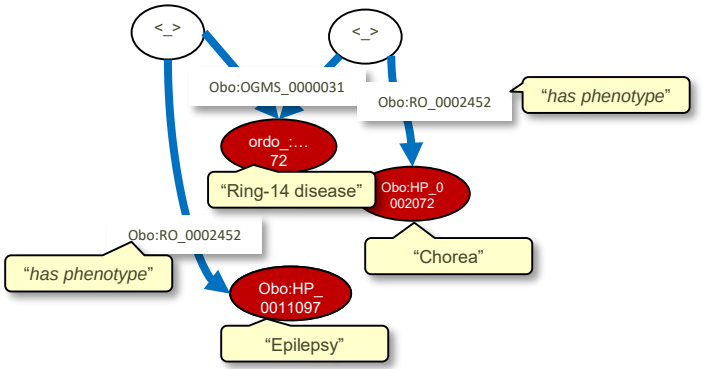
Explore combinations!

Monika	Annika	Rajaram	Pietro
<_> *obo:OGMS_0000031* ordo:Orphanet_1440 *obo:RO_0002452* obo:HP_0011097.	<_> *obo:OGMS_0000031* ordo:Orphanet_1440, *obo:RO_0002452* obo:HP_0002072.	<_> *obo:OGMS_0000031* ordo:Orphanet_178509, *obo:RO_0002452* obo:HP_0011097 *obo:RO_0002302* obo:CHEBI_33237	<_> *obo:OGMS_0000031* ordo:Orphanet_178509, *obo:RO_0002452* obo:HP_0000720 *obo:RO_0002302* obo:CHEBI_132952



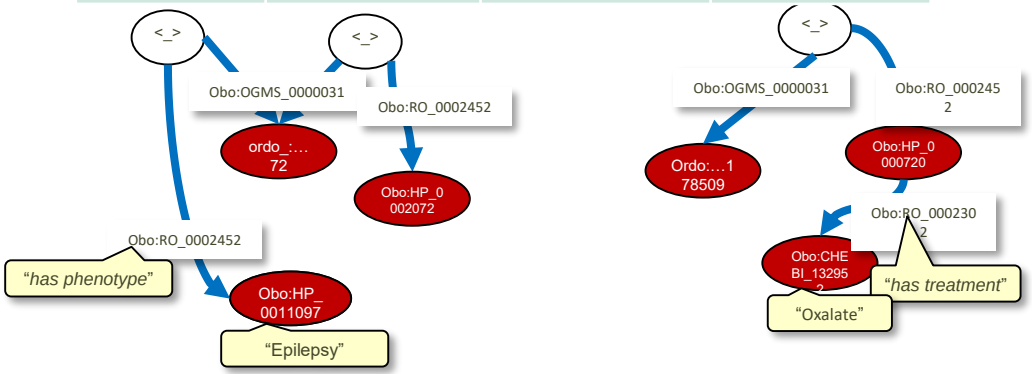
Explore combinations!

Monika	Annika	Rajaram	Pietro
<_> *obo:OGMS_0000031* ordo:Orphanet_1440 *obo:RO_0002452* obo:HP_0011097.	<_> *obo:OGMS_0000031* ordo:Orphanet_1440, *obo:RO_0002452* obo:HP_0002072.	<_> *obo:OGMS_0000031* ordo:Orphanet_178509, *obo:RO_0002452* obo:HP_0011097 *obo:RO_0002302* obo:CHEBI_33237	<_> *obo:OGMS_0000031* ordo:Orphanet_178509, *obo:RO_0002452* obo:HP_0000720 *obo:RO_0002302* obo:CHEBI_132952



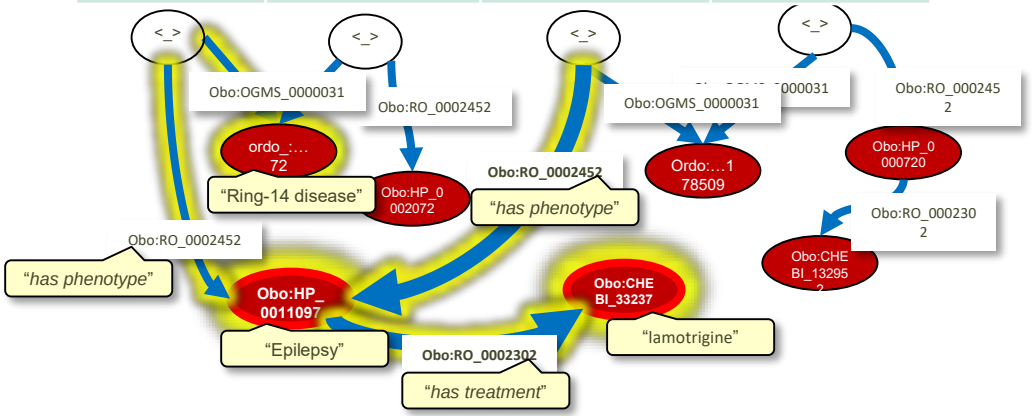
Explore combinations!

Monika	Annika	Rajaram	Pietro
<_> ***obo:OGMS_0000031*** ordo:Orphanet_1440 ***obo:RO_0002452*** obo:HP_0011097.	<_> ***obo:OGMS_0000031*** ordo:Orphanet_1440, ***obo:RO_0002452*** obo:HP_0002072.	<_> ***obo:OGMS_0000031*** ordo:Orphanet_178509, ***obo:RO_0002452*** obo:HP_0011097 ***obo:RO_0002302*** obo:CHEBI_33237	<_> ***obo:OGMS_0000031*** ordo:Orphanet_178509, ***obo:RO_0002452*** obo:HP_0000720 ***obo:RO_0002302*** obo:CHEBI_132952



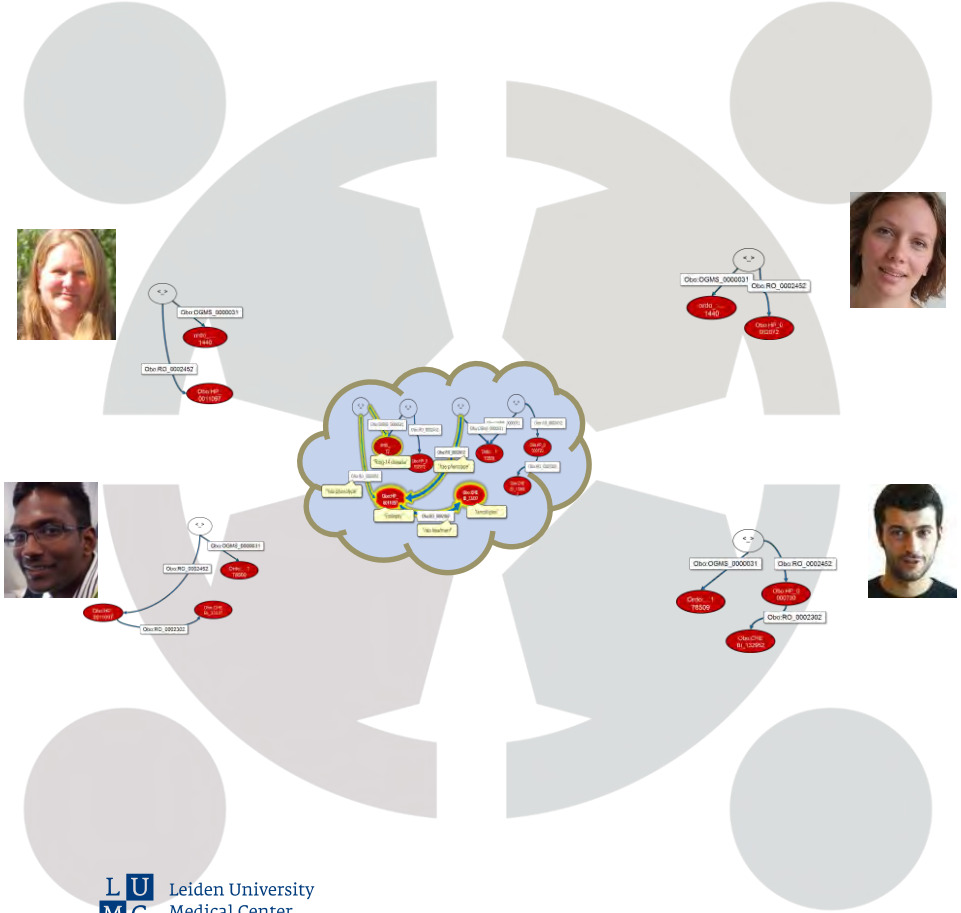
The solution!

Monika	Annika	Rajaram	Pietro
<_> *obo:OGMS_0000031* ordo:Orphanet_1440 *obo:RO_0002452* obo:HP_0011097.	<_> *obo:OGMS_0000031* ordo:Orphanet_1440, *obo:RO_0002452* obo:HP_0002072.	<_> *obo:OGMS_0000031* ordo:Orphanet_178509, *obo:RO_0002452* obo:HP_0011097 *obo:RO_0002302* obo:CHEBI_33237	<_> *obo:OGMS_0000031* ordo:Orphanet_178509, *obo:RO_0002452* obo:HP_0000720 *obo:RO_0002302* obo:CHEBI_132952



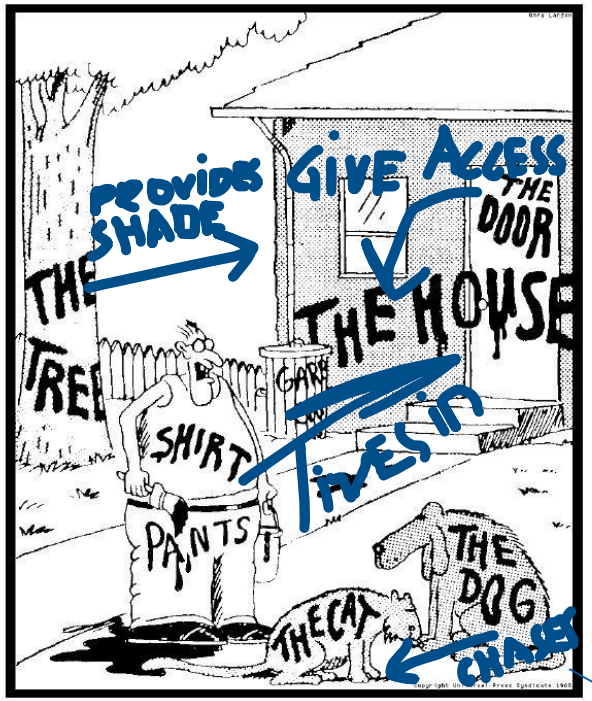
FAIR registries

They form a *virtual* knowledge graph (i.e. not necessarily in one location)



This Training has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 825575 - European Joint Programme on Rare Diseases

Observation of FAIR game



"Now! *That* should clear up a few things around here!"

- Common data elements, nor data quality resolved the lack of interoperability
- Data elements and their relations were converted in ontology-qualified identifiers '*at source*'
- That creates a virtual knowledge graph containing solutions across resources
- *A focus on FAIR may contribute more to (re)usability of data than a focus on data as such*

…in a language that computers understand

Silos, silos, silos

Funded by the European Union
GA n°825575

Silos, silos, silos

Funded by the European Union
GA n°825575

Absolute need to combine

Funded by the European Union
GA n°825575

Achieving the global vision

Usability of data for rare disease care and research must be brought to higher, global levels

Funded by the European Union
GA n°825575

Lorentz workshop 2014
Birth of the FAIR principles

Barend Mons (LUMC/UL)
Initiator FAIR principles
President CODATA

Lorentz Workshop, Leiden, 2014

Meeting of global leaders in data science

Motivation: data use & reuse *unacceptably* inefficient and poorly reproducible

Concluding that data should be

Findable, **A**ccessible, **I**nteroperable, **R**eusable

*for humans **and machines****

*In 2016 refined in terms of 15 FAIR guiding principles for scientific data management and stewardship

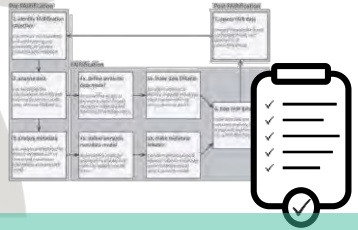
[Wilkinson *et al.*, 2016 https://www.nature.com/articles/sdata201618]

Research & Development of FAIRification

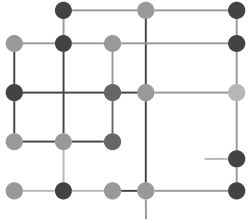
Experience from FAIRification projects and workshops in RD domain since 2014

Training

Organising expertise

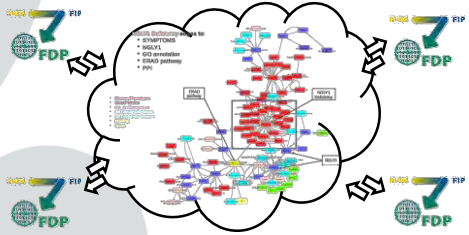


Guidelines & metrics



Align with infrastructure

Tools



FAIR exploitation

EJP RD

Funded by the European Union
GA n°825575

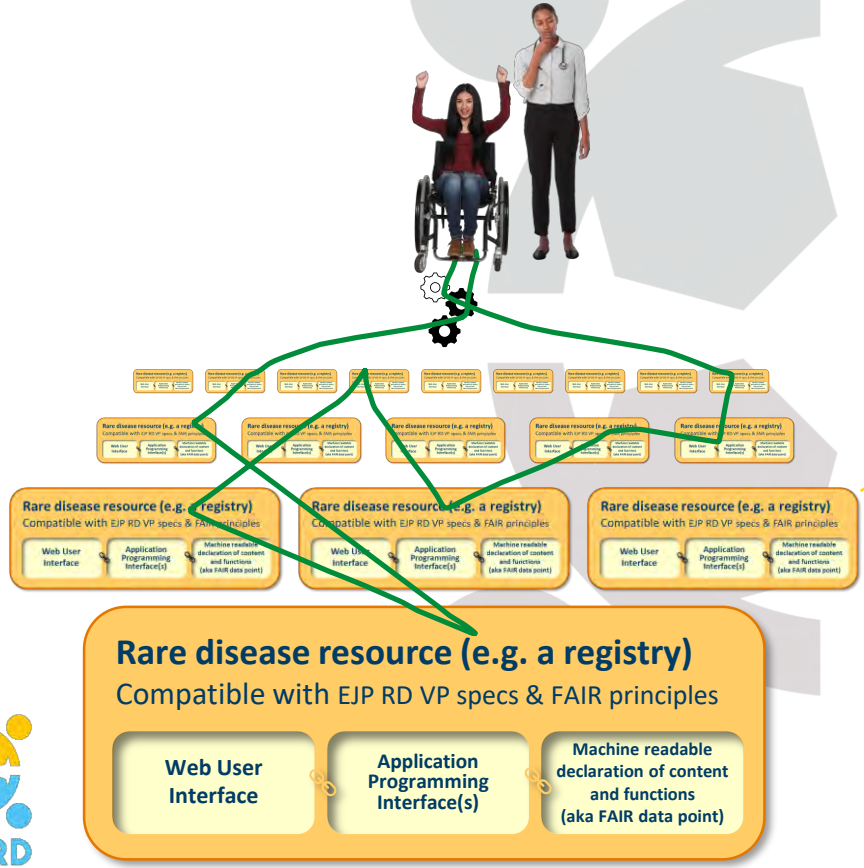
EUROPEAN JOINT PROGRAMME RARE DISEASES

Development of the Virtual Platform network of federated FAIR resources

EJP RD

Funded by the European Union
GA n°825575

EJP RD FAIR Virtual Platform to contribute to other **people's research for benefit of Rare Disease patients**



- Rare disease patients would like you to contribute to the VP network
- Resources create the Virtual Platform network
- Resources declare about their content and functions *in a language that computers understand*
- Analysis algorithms make use of FAIRness, researchers make use of these algorithms

Funded by the European Union
GA n°825575

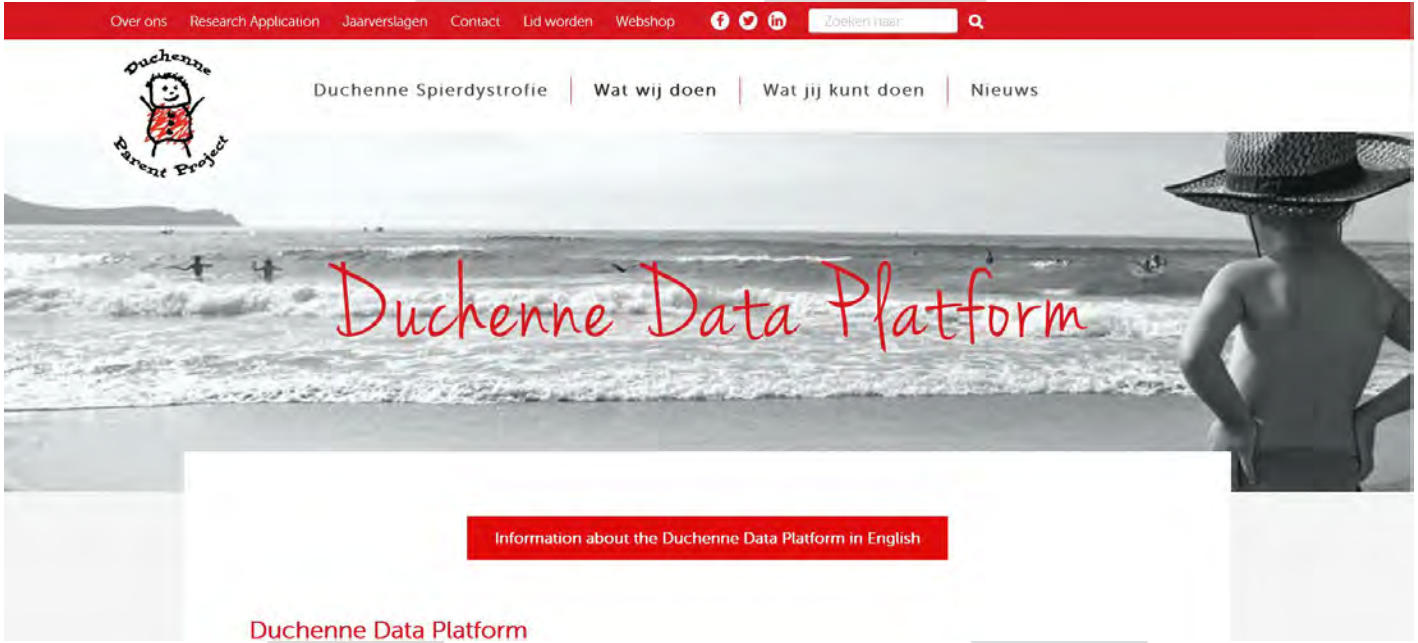
FAIR for machines

example: web interface Duchenne Data Platform

Rare disease resource (e.g. a registry)

Compatible with EJP RD VP specs & FAIR principles

- **Web User Interface**
- **Application Programming Interface(s)**
- **Machine understandable declaration of content**



Funded by the European Union
GA n°825575

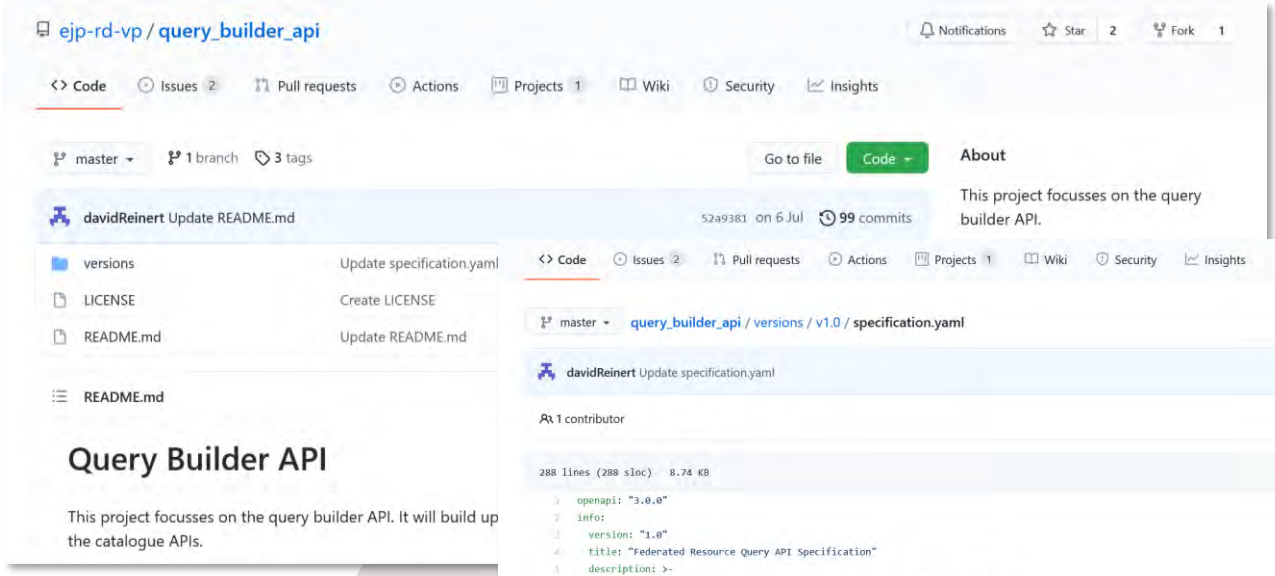
API for engineers…(a special type of human)

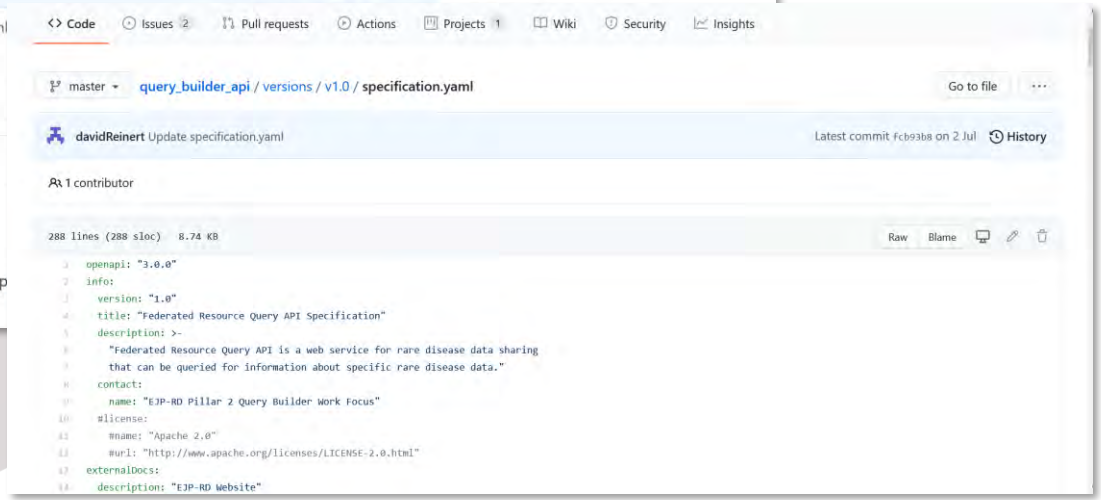
A resource discovery API (programmatic access)

Rare disease resource (e.g. a registry)

Compatible with EJP RD VP specs & FAIR principles

- Web User Interface
- Application Programming Interface(s)
- Machine understandable declaration of content





Funded by the European Union
GA n°825575

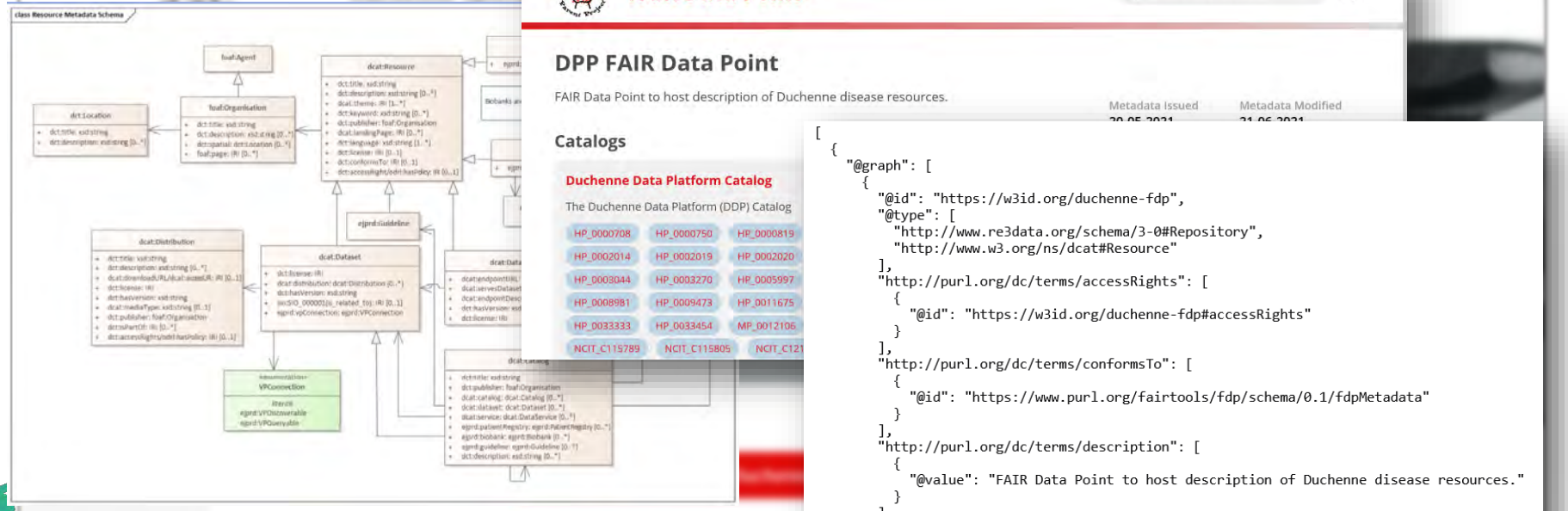
Resource description for machines

example: DCAT metadata & metadata provisioning service ('FAIR Data Point')

Rare disease resource (e.g. a registry)

Compatible with EJP RD VP specs & FAIR principles

- Web User Interface
- Application Programming Interface(s)
- Machine understandable declaration of content



W3C-recommended Data Catalog Vocabulary

EJP RD

...ded by the European Union
GA n°825575

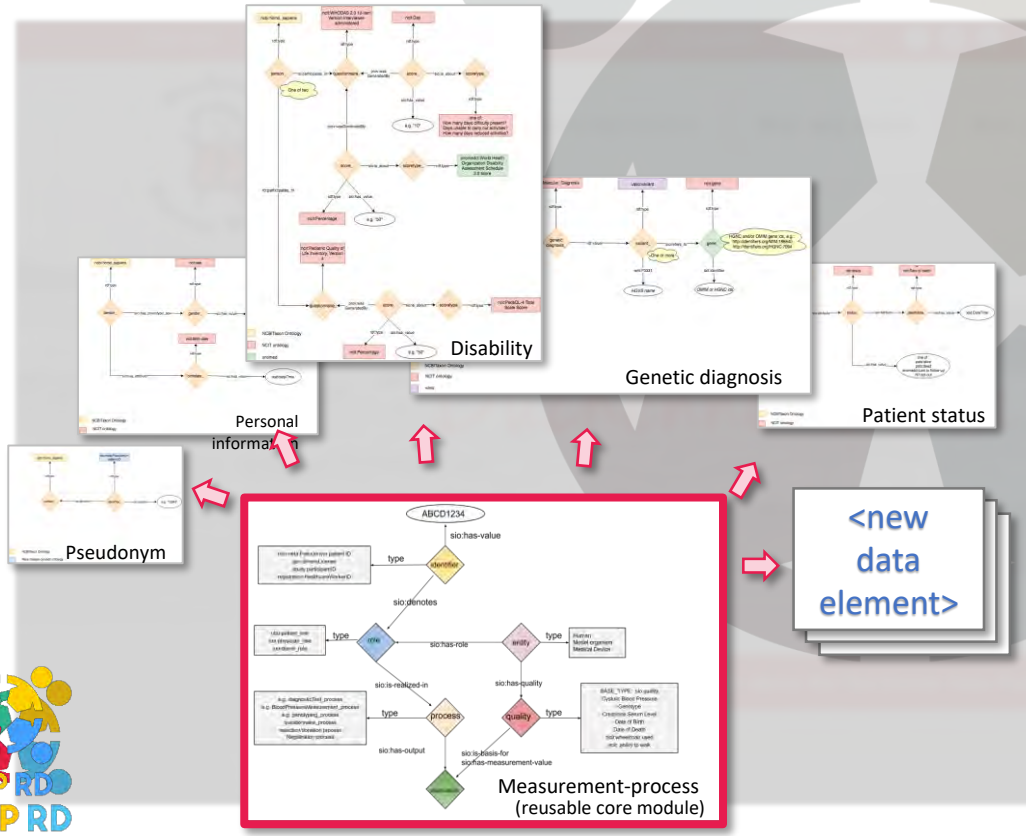
Data elements for machines

Modelling pattern for data elements, aka "DCAT" for data elements

Rare disease resource (e.g. a registry)

Compatible with EJP RD VP specs & FAIR principles

- Web User Interface
- Application Programming Interface(s)
- Machine understandable declaration of content



Semantic modelling pattern for machine understandable observations

- Extendable model '*for machines*' using standard ontologies, applied *at source*
- Modules available for 16 *common* data elements for RD patient registries
- Mappings/conversions, e.g. to FHIR, OMOP, C-DISC, OBO Foundry, GA4GH

EJP RD

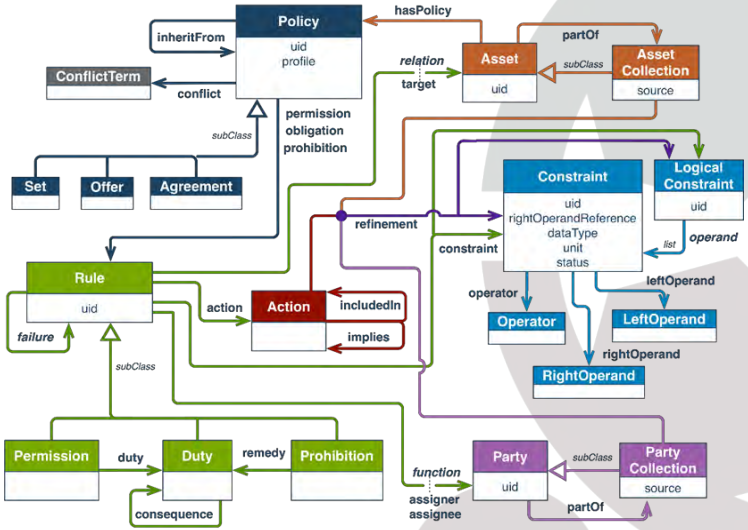
Funded by the European Union
GA n°825575

Data access conditions for machines

Modelling common consent elements for machines

Rare disease resource (e.g. a registry)
Compatible with EJP RD VP specs & FAIR principles

Web User Interface | Application Programming Interface(s) | Machine understandable declaration of content



- ‘Common consent elements (IRDiRC Task Force)
- Developing machine understandable version for triaging access requests (W3C Open Digital Rights Language – ODRL)
- *(Contact me if you are interested in a cross-project working group)*

Funded by the European Union
GA n°825575

Light weight index generated from FAIR metadata

EJP METADATA
DCAT « Template »

EJP METADATA
DCAT « Template »

EJP METADATA
DCAT « Template »

FDP

FDP

FDP

Index generated
from metadata

Ask for
Available
sources

Fetch metadata
to be displayed

Obtain
endpoint URLs
to perform
query API

« Virtual Platform Portal »

FAIR metadata provisioning services
« FAIR Data Point » *at source*

Virtual Platform Network

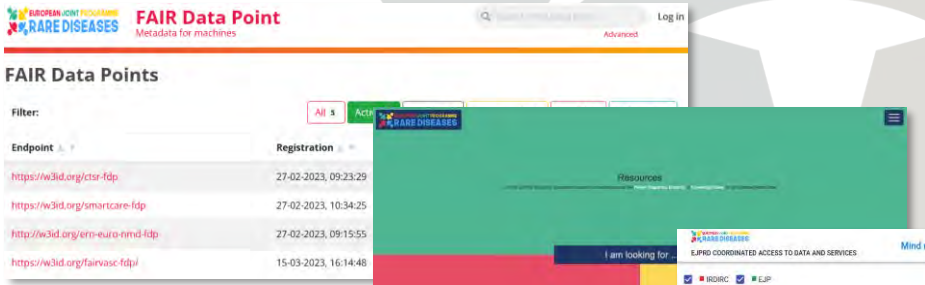
Virtual Platform Portal

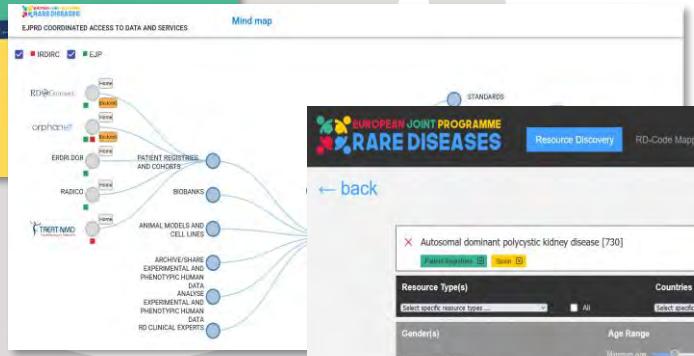
EJP RD

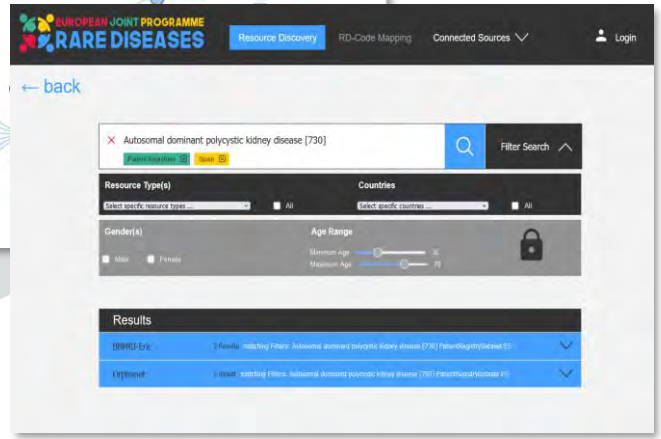
Funded by the European Union
GA n°825575

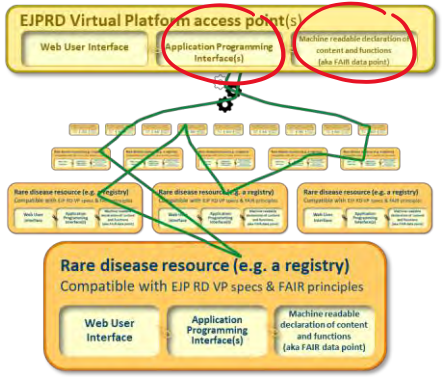
Applications for humans…

example: resource finders on vp.ejprarediseases.org









Or, … create your own computational workflow or user application

Funded by the European Union
GA n°825575

EJP RD effort on getting there…

Your data in original form

FAIRification

FAIR VP compliant resource with FAIR data

Web User Interface

Application Programming Interface(s)

Machine readable declaration of content and functions (aka FAIR data point)

Your data in original form

Ontological Information about resource & access

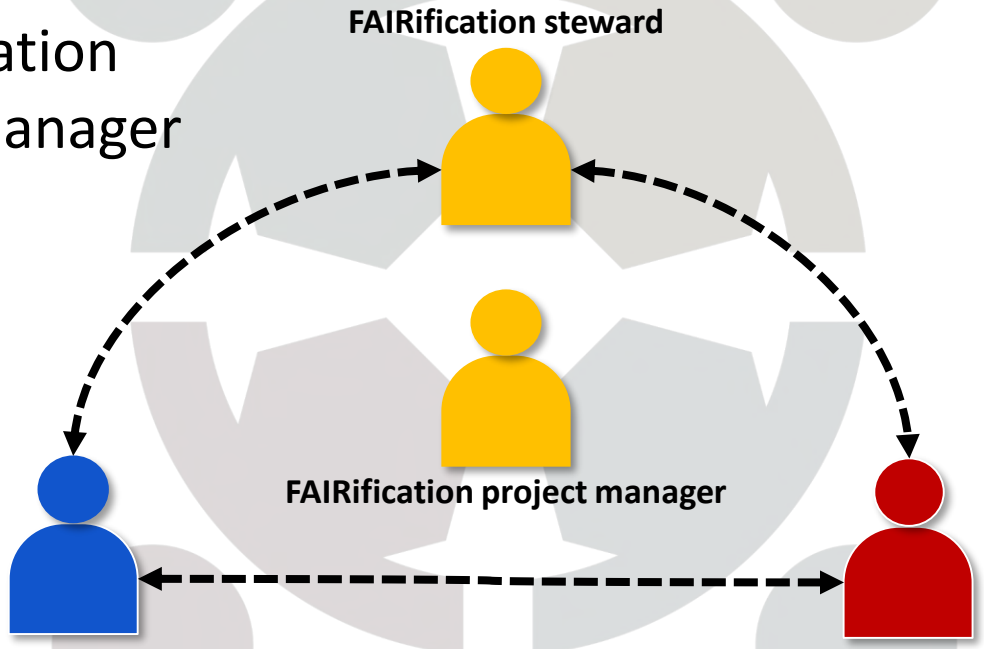
Ontological model describing data elements

EJP RD

Funded by the European Union GA n°825575

R&D of FAIRification
Organising expertise

3 party collaboration
+ FAIR project manager



Tasks

	Task	Period (months)	Associated Milestone/ Deliverable	Comments
Start of preparation phase	Define driving research questions and target data for FAIRification (use case)	1	M1	
	Technical assessment of requirements for the BYOD	1-2	M2	
	Select BYOD experts	3	M2	<Disease D> experts, <Disease D> database experts, FAIR data experts
	Prepare minimal reference ontology for BYOD	2-4	M2	
End of preparation phase	BYOD <Disease D> data owners and FAIR data experts	4	M3	
Start of implementation phase	Technical assessment of requirements for implementing a FAIR Data Point for the <Disease D> resource	4-8	M4, D1	
	Develop and test <Disease D> FAIR Data Point	4-11	M4, D1	Agile development involving pre-releases of FDPs for testing
	Design and execute data analysis for testing FAIR Data Points as substrate for knowledge discovery on <Disease D> (proof-of-principle)	9-11	M5	Executed by collaborating research questos
End of	Release documented	11-12	D1	By <Disease D>

Roadmap & cost estimate

Funded by the European Union
GA n°825575

Network of FAIRification stewards

3-party teams
Software engineer,
FAIRification steward,
Local data steward

Collaborating
FAIRification
stewards

liaison

FAIR VP
architects

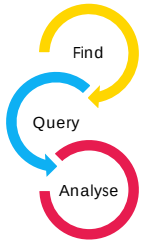
Data & technical
experts

Specific expert

EJP RD

by the
Union
GA n°825575

Incrementally contributing to the Virtual Platform



Onboarding guidance document (ongoing)

Onboarding F2F and follow-up workshops

Level 1

- Metadata discovery
  - basic description of resource
  - overview of content

Level 2

- Discovery of limited parts of datasets
  - output: yes/no, counts, etc..

Level 3

- Querying of datasets
  - record-level querying

Level 4

- Federated Analysis
  - Answer advanced research questions

What is it?

What does it contain?

Answer my complex question

EJP RD

Funded by the European Union
GA n°825575

Tooling by the EJP RD

Your data in original form

FAIRification

- DIY by **VP & FDP specifications**
- CSV to FAIR by **FAIR-in-a-box**
- Create API by **Beacon-in-a-box**
- Out of the box FAIR generation by **VP specs-compliant registry software** such as MOLGENIS, CastorEDC, Duchenne Data Platform
- Question-based guidance by **smart guidance tool** for data stewards
- Project management guidelines

FAIR VP compliant resource with FAIR data

- Web User Interface
- Application Programming Interface(s)
- Machine readable declaration of content and functions (aka FAIR data point)

Your data in original form

- Ontological Information about resource & access
- Ontological model describing data elements

EJP RD

Funded by the European Union GA n°825575

WORLD
DUCHENNE
ORGANIZATION
UPPMD

FAIR

2019

Duchenne
Data
Platform

FAIR
TEAM

2020

FAIR
STATUS
ACHIEVE
D

2021

2022

INTEROPERABILITY
ACHIEVED BETWEEN
TWO FAIR REGISTRIES

2023

INTEROPERABILITY
ACHIEVED BETWEEN
FIVE FAIR REGISTRIES

Duchenne
Data
Platform

SMARTCARE

ERN
EURO-NMD

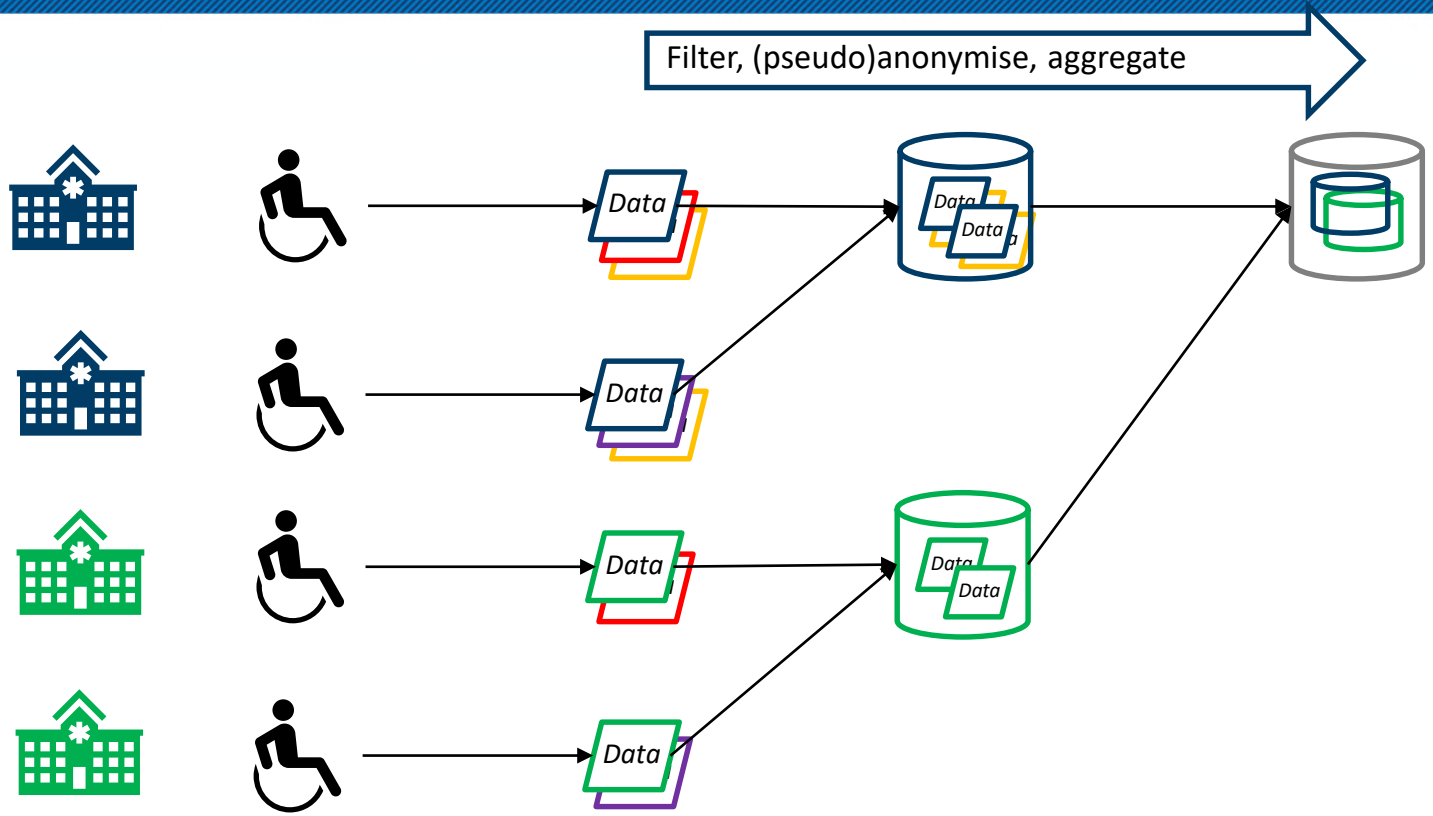
CRAMP

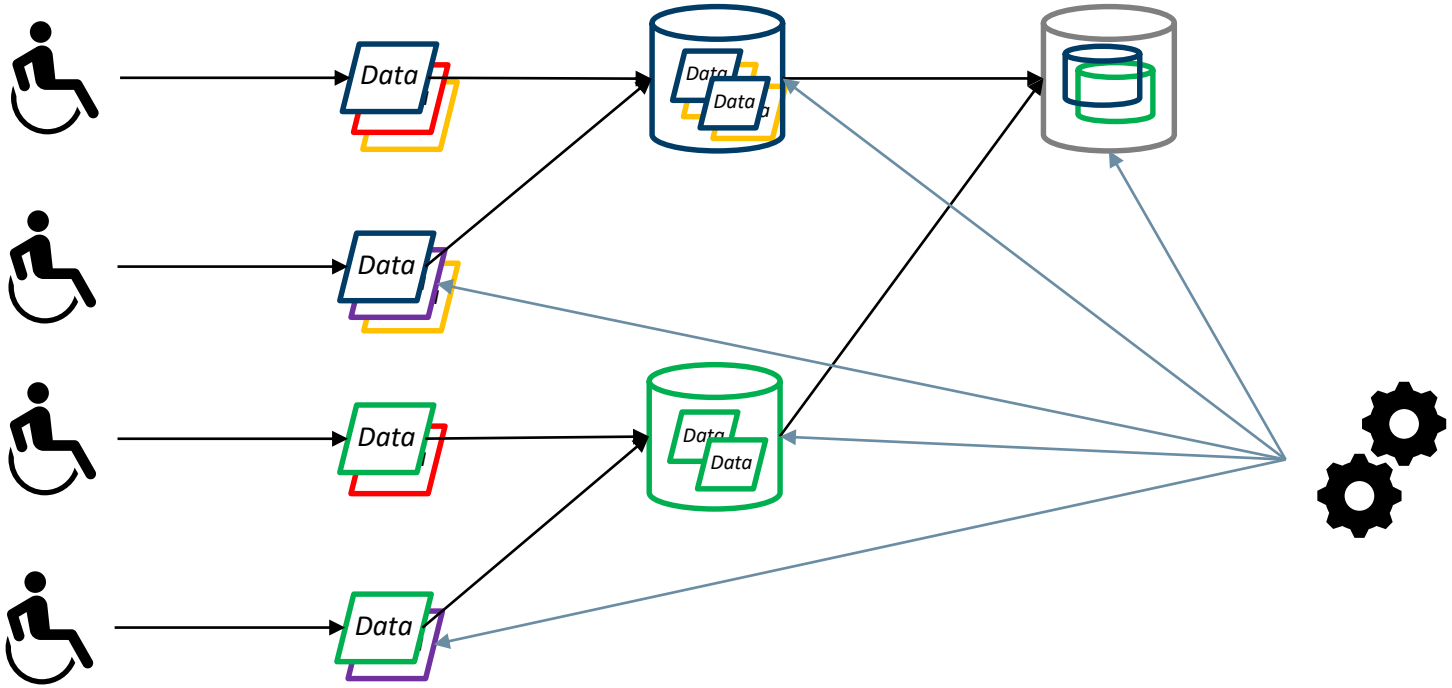
DM Scope

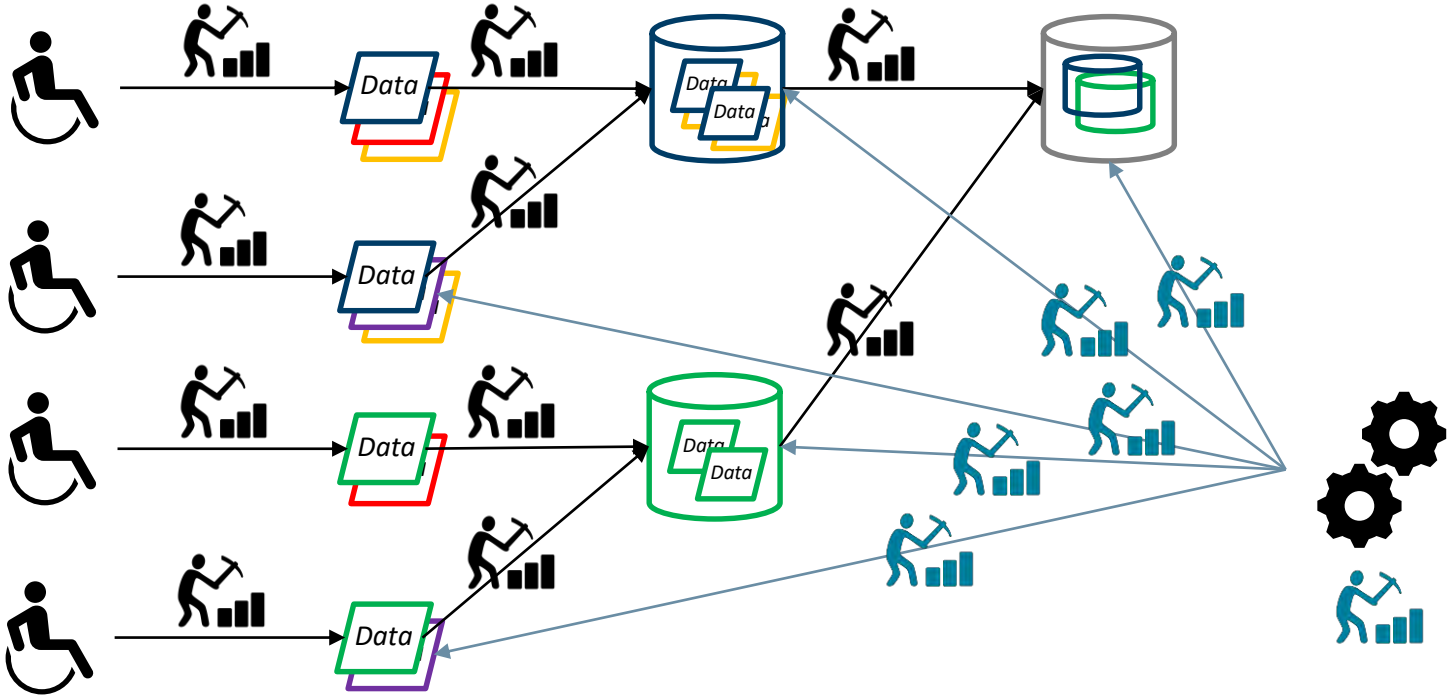
My conclusion

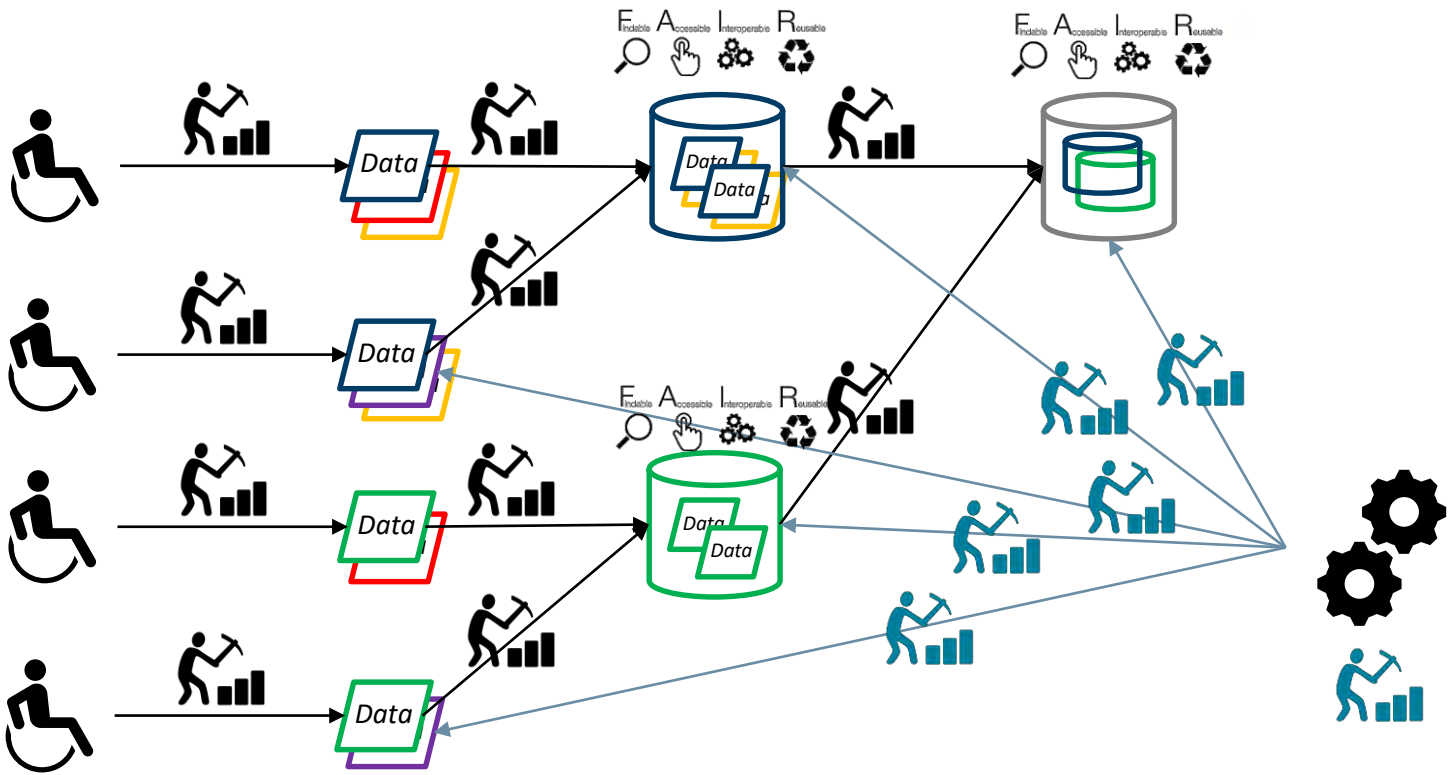
Lizanne, Duchenne Muscular Dystrophy patient, endorses FAIR principles

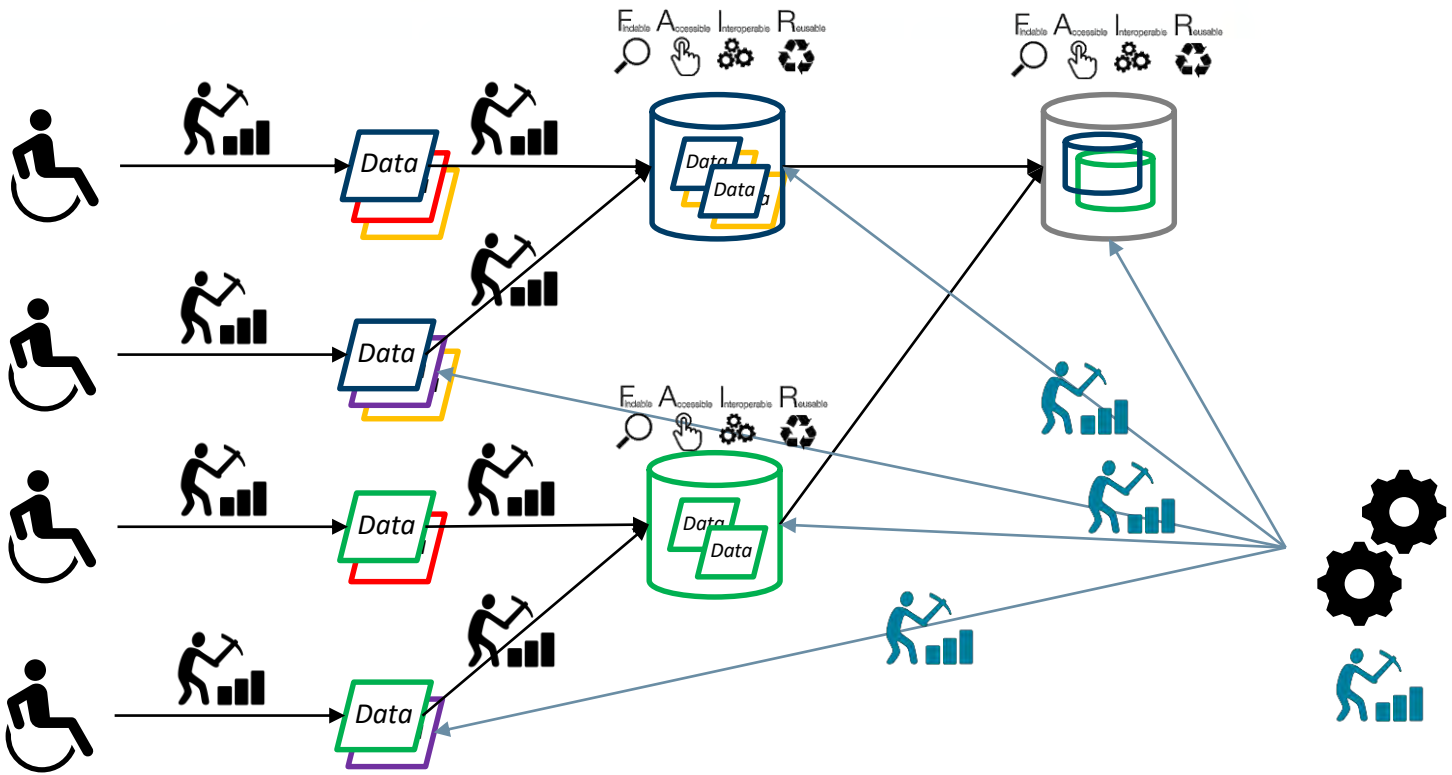
- Having data is not the same as having optimally usable data
- FAIR principles exist entirely to make data as usable as possible, for trusted algorithms, under defined – but efficiently processable – conditions
- Patients will benefit if data generators and data collectors shift balance to effort on making data FAIR as much as they do on generating/collecting data
- Be persistent: it may take years before we reach a critical mass of FAIR data (NB: patient organisations will not wait forever…)
- Future challenge: move FAIR closer to the source → automate creation of registries

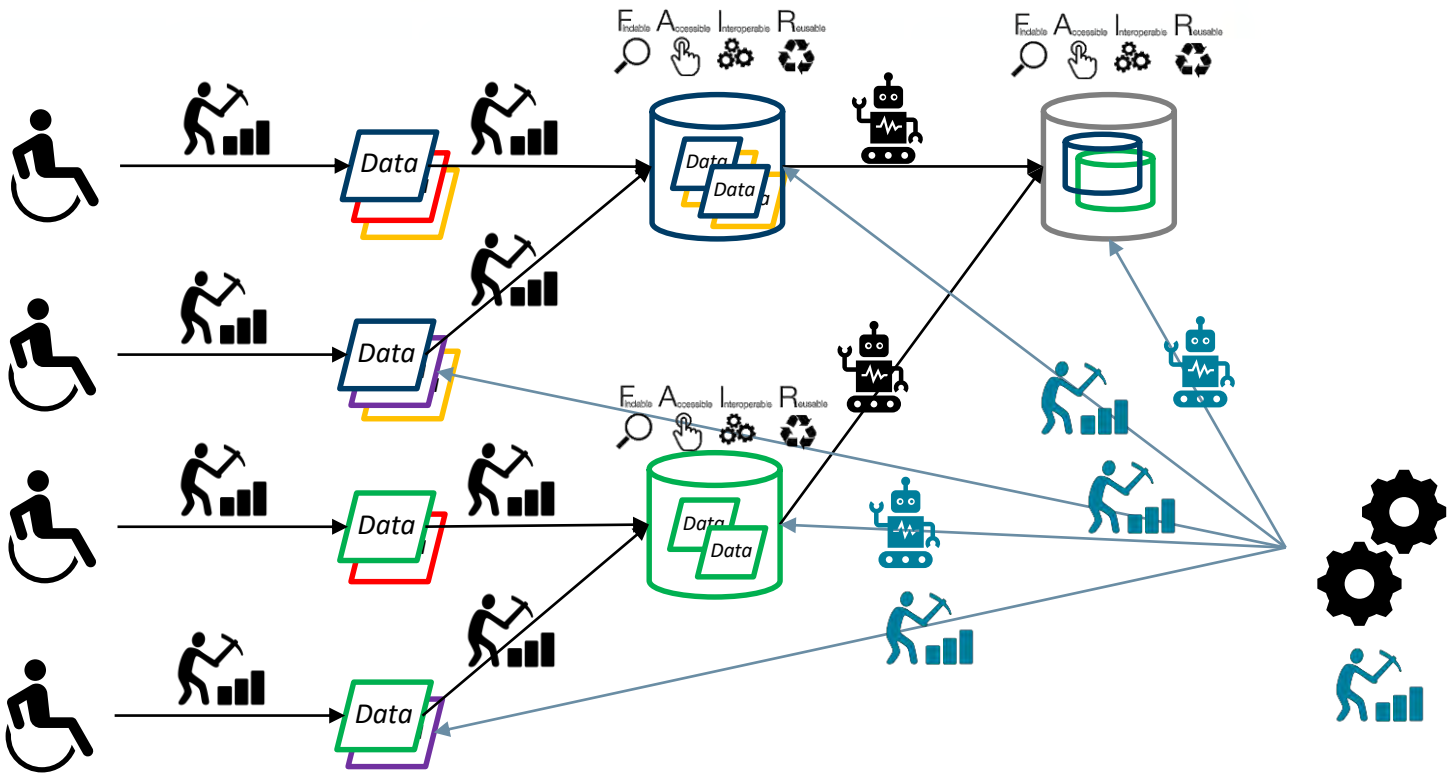

Filter, (pseudo)anonymise, aggregate
Data
Data
Data
Data
Data
Data
Data
Data

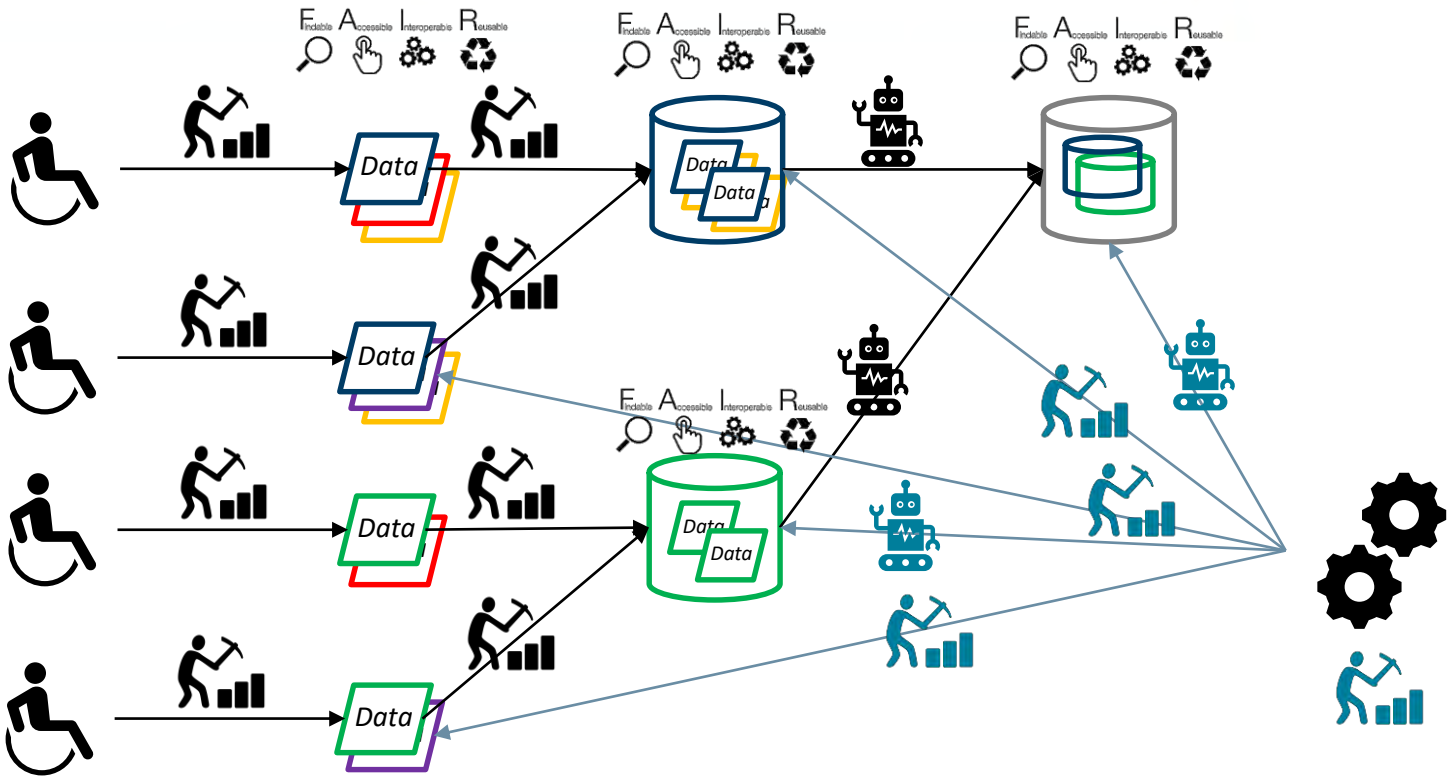

Data
Data
Data
Data
Data
Data
Data
Data

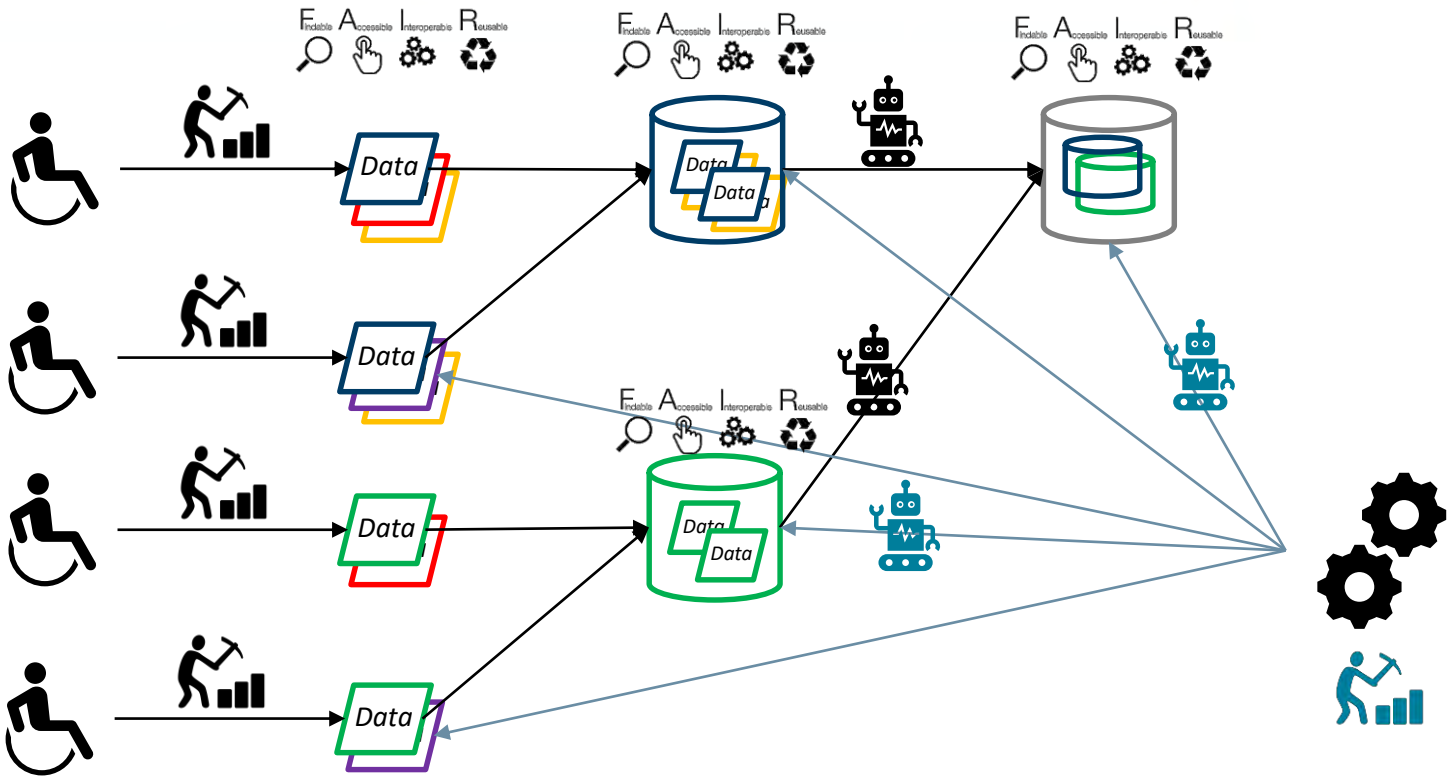

Data
Data
Data
Data
Data
Data
Data
Data

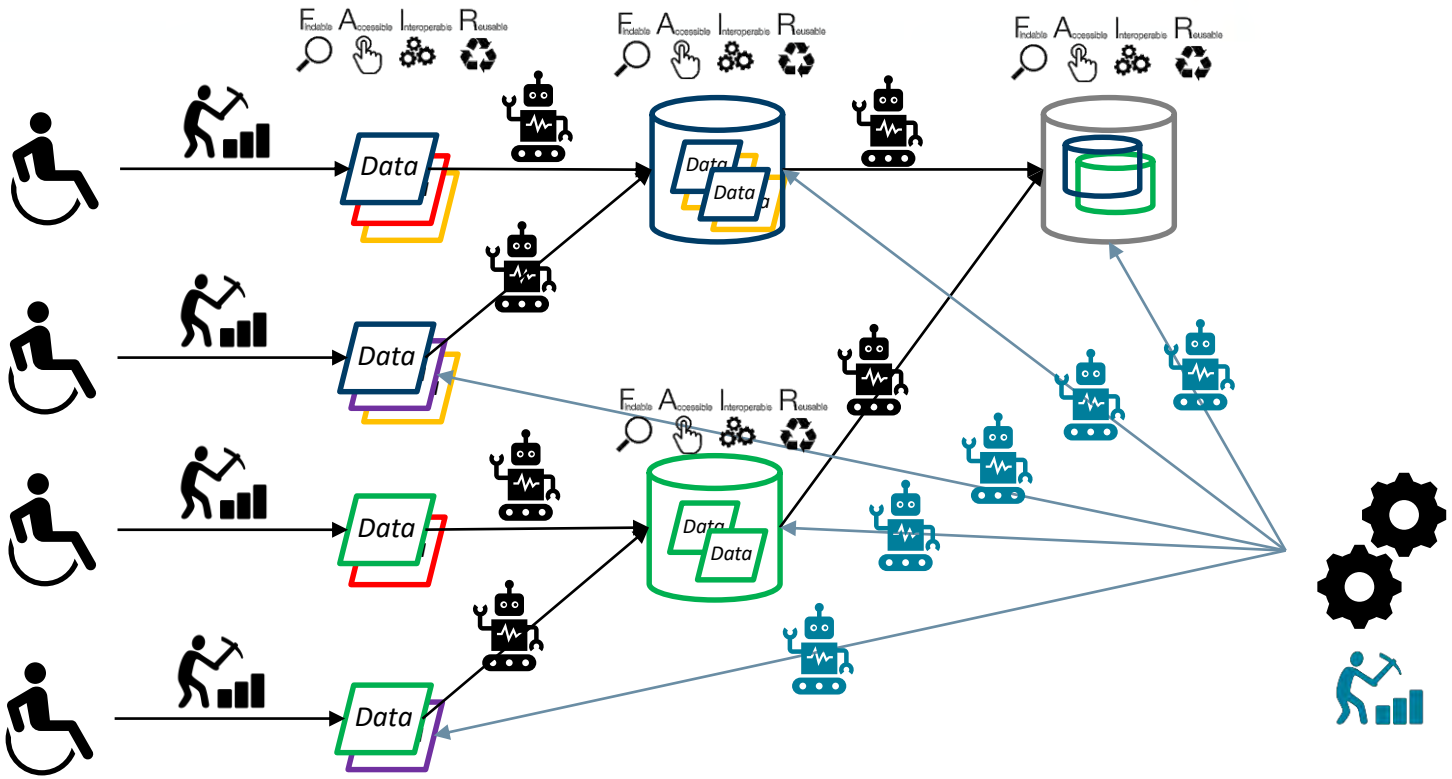

Findable Accessible Interoperable Reusable
Findable Accessible Interoperable Reusable
Findable Accessible Interoperable Reusable
Data
Data
Data
Data
Data
Data
Data
Data


Findable Accessible Interoperable Reusable
Findable Accessible Interoperable Reusable
Findable Accessible Interoperable Reusable
Data
Data
Data
Data
Data
Data
Data
Data
Data

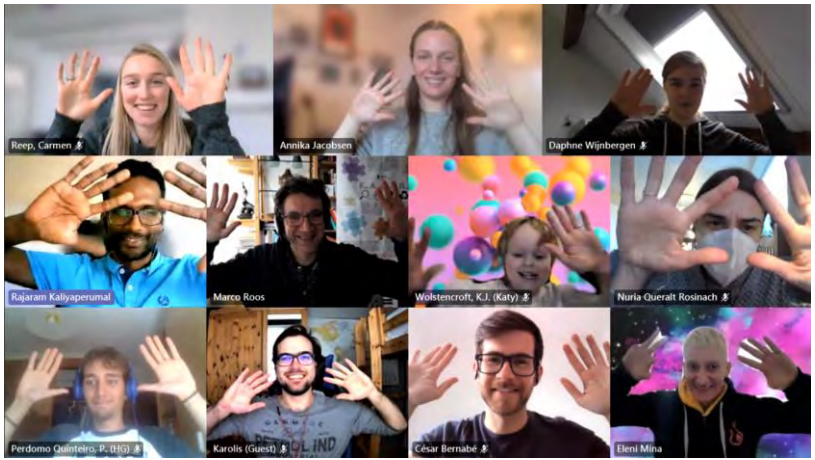

Findable Accessible Interoperable Reusable
Findable Accessible Interoperable Reusable
Findable Accessible Interoperable Reusable
Data
Data
Data
Data
Data
Data
Data
Data
Data


Findable Accessible Interoperable Reusable
Data
Data
Data
Data
Data
Data
Data
Data
Data


Findable Accessible Interoperable Reusable
Data
Data
Data
Data
Data
Data
Data
Data
Findable Accessible Interoperable Reusable
Findable Accessible Interoperable Reusable
Findable Accessible Interoperable Reusable


Findable Accessible Interoperable Reusable
Data
Data
Data
Data
Data
Data
Data
Data
Findable Accessible Interoperable Reusable
Findable Accessible Interoperable Reusable
Findable Accessible Interoperable Reusable

Thank you



Special thanks to the Biosemantics group, rare disease patient representatives, the Duchenne community, The European Joint programme rare diseases, ELIXIR, RD-Connect, the Semantic Web community, many, many more…

Vielen Dank für Ihre Aufmerksamkeit