THE RARE DISEASE DATA LINKAGE PLAN

IMPLEMENTING FAIR DATA PRINCIPLES TO CUT ACROSS BIOBANKS, REGISTRIES, MOLECULAR DATA

Marco Roos, Claudio Carta, David van Enckevort, Rajaram Kaliyaperumal, Annika Jacobsen, Rachel Thompson, Robert Reihs, Heimo Mueller, Pedro Sernadela, Marc Hanauer, Mark Thompson, Mark Wilkinson, Luiz Bonino, Marco Crimi, Peter-Bram 't Hoen, Domenica Taruscio

Special acknowledgement: Barend Mons, rare disease patients and representatives, participants of rare disease BYODs and hackathons

Pointers

- Related posters: Annika Jacobsen, Tamás Szabó, David van Enckevort, Claudio Carta
- Hackers: find Rajaram
- Wednesday: make yourself FAIR!!

  Demonstration of FAIR data tooling: find Rajaram, Kristina, Eleni
- Thursday: e-Rare workshop on data sharing

To promote and support FAIR RD Registries

This edition dedicated to European Reference Networks

Bringing data from across RD data resources to RD specialists across European Health Care Providers

Are you interested?
Send an e-mail to:
rareregistries-school@iss.it

Outline

- Recap rare disease data linkage
- Privacy preserving record linkage
- Deployment of the rare disease data linkage plan
- Experiment: find a treatment for ~~Monica~~ anonymous person
- Q&A

Recap

Where it started

1st RD-Connect annual meeting, session on Linked Data and Ontologies (Peter Robinson, Marco Roos)

→ Linked Data and Ontology task force

→ Now a cross-project task force

Where it started: BYODs and related workshops

- **2nd International Summer School on: "RARE DISEASE AND ORPHAN DRUG REGISTRIES" ,**

Rome, September 15-19 , 2014

- **1st RD-Connect Bring Your Own Data meeting**

Rome 26-27 November 2014

- **3rd International Summer School on: "Rare Disease and Orphan Drug Registries"**

Rome, September 21-23 , 2015

- **Rd-Connect workshop: data linkage and ontologies**

Rome, September 24-25 , 2015

- **Course: Managing and Integrating Life Science Information**

Under the auspices of the Dutch Bioinformatics & Systems Biology Research School
Utrecht, Netherlands, Rome, 30 November - 4 December 2015

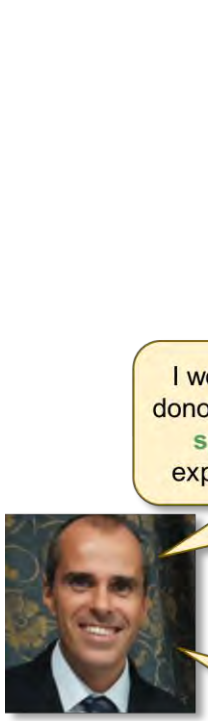
BYODs and related workshops

Workshop/training: "Including biobanks & registries in the RD-Connect platform" (During RD-Connect Annual meeting)
Barcelona, 9-11 March 2016

Workshop 6 – FAIR Data and Data Stewardship
(at 15th European Conference on Computational Biology, ECCB, 2016**)**
The Hague, The Netherlands, September 3, 2016

4th International Summer School on: "Rare Disease and Orphan Drug Registries".
Rome, 26 – 28 September 2016
+RD-Connect BYOD Workshop to Link Rare Disease Registries
Rome, Rome, 26 – 28 September 2016

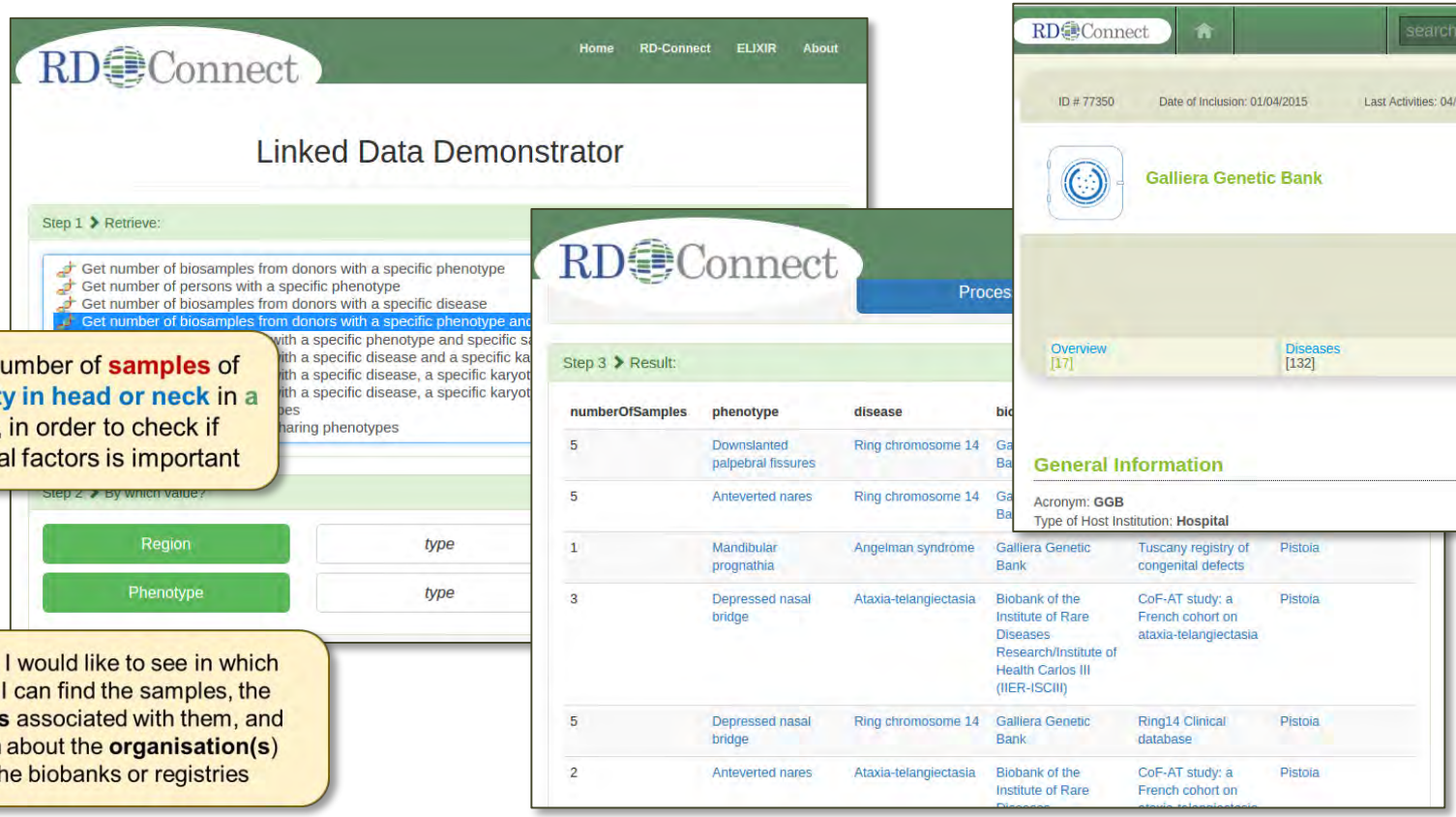
RD-Connect: proof of concept questions across biobanks and registries
ELIXIR: Test interoperability components



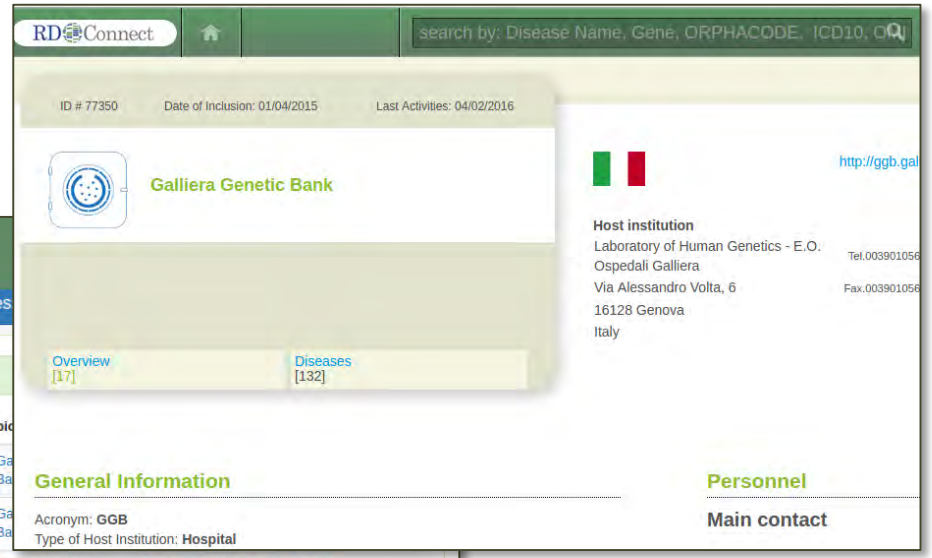
I would like to know the number of **samples** of donors with **an abnormality in head or neck** in **a specific region** of Italy, in order to check if exposure to environmental factors is important

In addition, I would like to see in which **biobanks** I can find the samples, the **phenotypes** associated with them, and information about the **organisation(s)** behind the biobanks or registries

Driving user questions



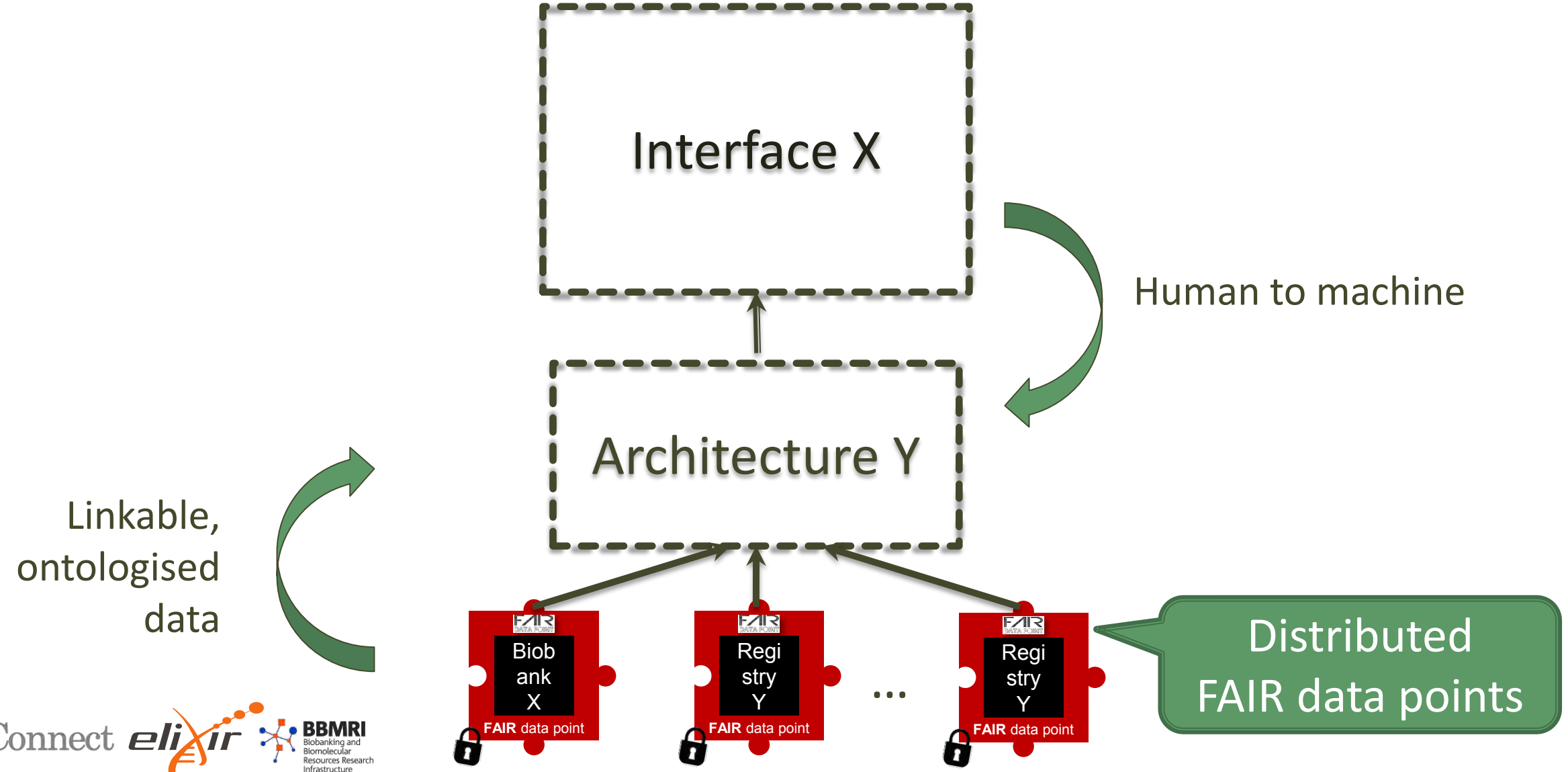
Demonstrator UI



ID Cards

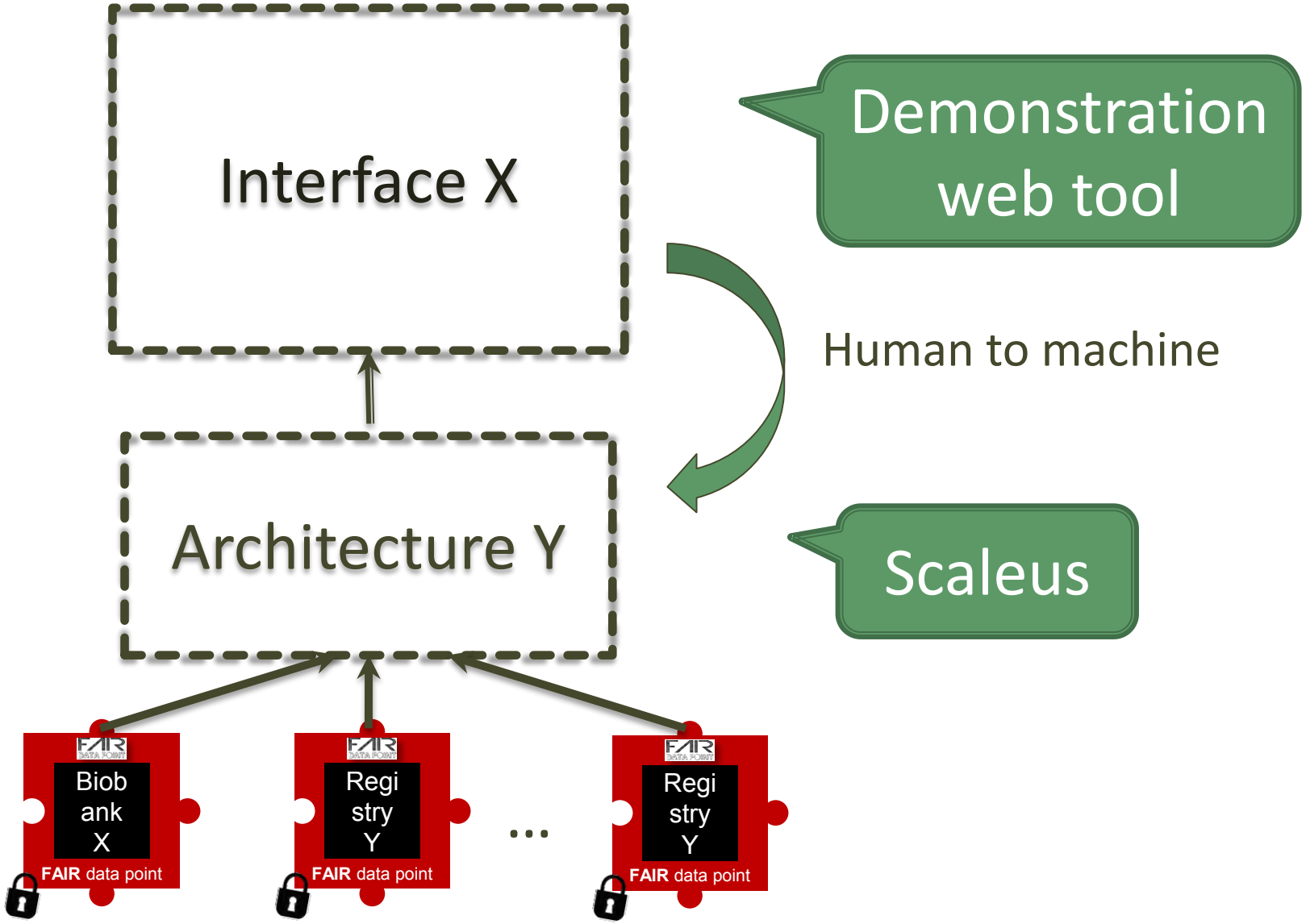
RD-Connect elixir BBMRI Biobanking and Biomolecular Resources Research Infrastructure

Building the demonstrator



RD Connect elixir BBMRI Biobanking and Biomolecular Resources Research Infrastructure

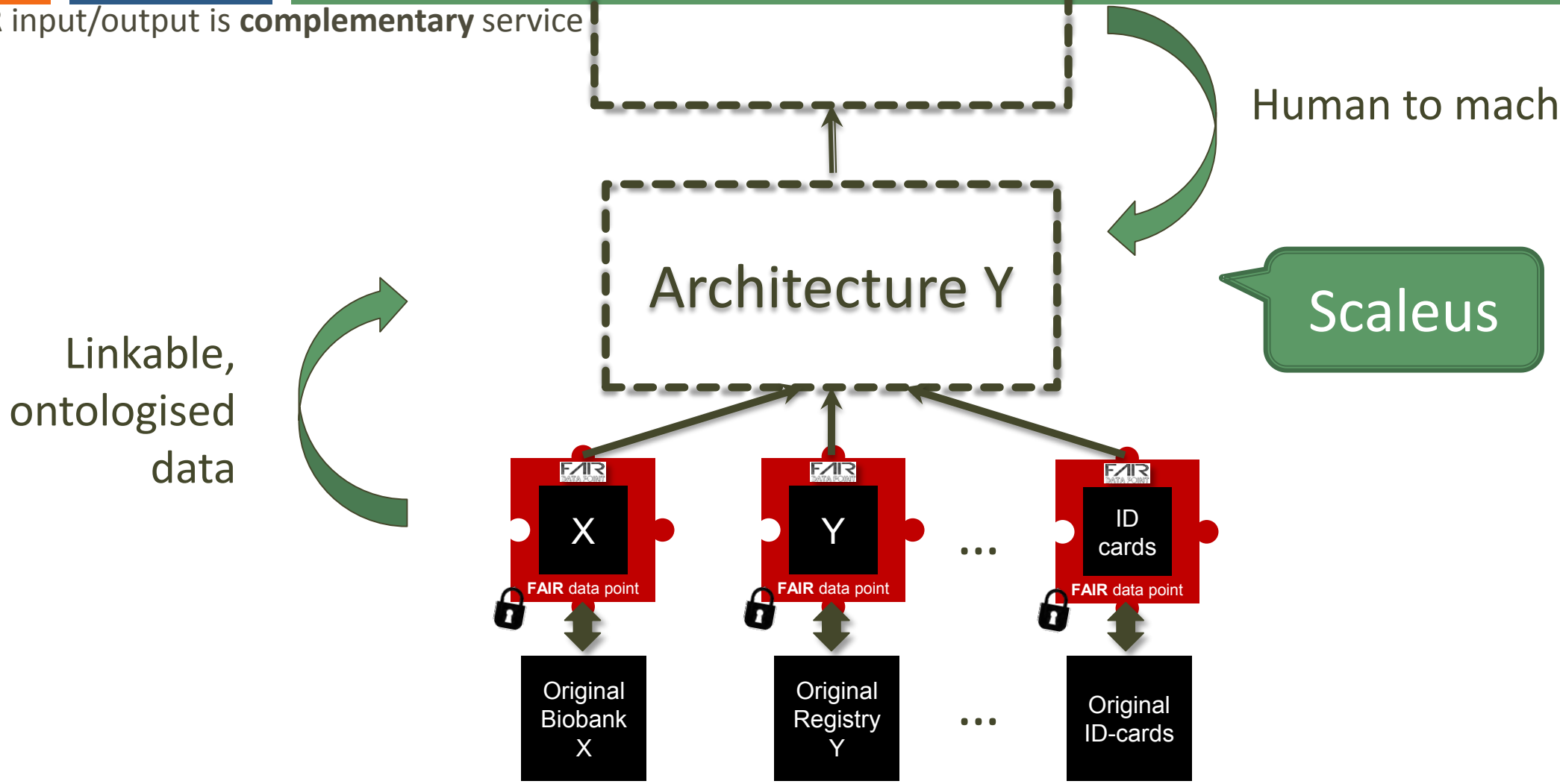
Building the demonstrator



RD Connect elixir BBMRI Biobanking and Biomolecular Resources Research Infrastructure

Building the demonstrator

Note: FAIR input/output is **complementary** service



2017: Towards a FAIR data service for the rare disease community

The rare disease data linkage plan

- Goal: independently FAIR resources
- Help make RD resources linkable '*at the source*'
- Speed up by tooling and SOPs
- Align with ELIXIR interoperability roadmap

+ Address privacy preserving use of personal identifiers

RD-Connect elixir BBMRI Biobanking and Biomolecular Resources Research Infrastructure

Overview

RD community meetings (BYODs, jamborees, …)

Driving user questions

Test interoperability components

Recommen-dations for interoperability backbone

Demo for RD community

Organise rare disease data linkage plan to make RD resources linkable at the source (WP8+WP5 +stakeholders)

Develop SOPs and semantic reference models
Deploy FAIR tools and protocols at RD data sources

Study interoperability components for molecular data

FAIR data software development (Luiz Bonino *et al.*)

GA4GH/IRDiRC PPRL

ELIXIR Interoperability Platform development

2015-2016 | 2016-2017 | 2017

M8.3 D8.3

M8.4 D8.2

RD Connect elixir BBMRI Biobanking and Biomolecular Resources Research Infrastructure

PPRL Working Group update

Global Alliance
for Genomics & Health

elixir

RD Connect

Introduction

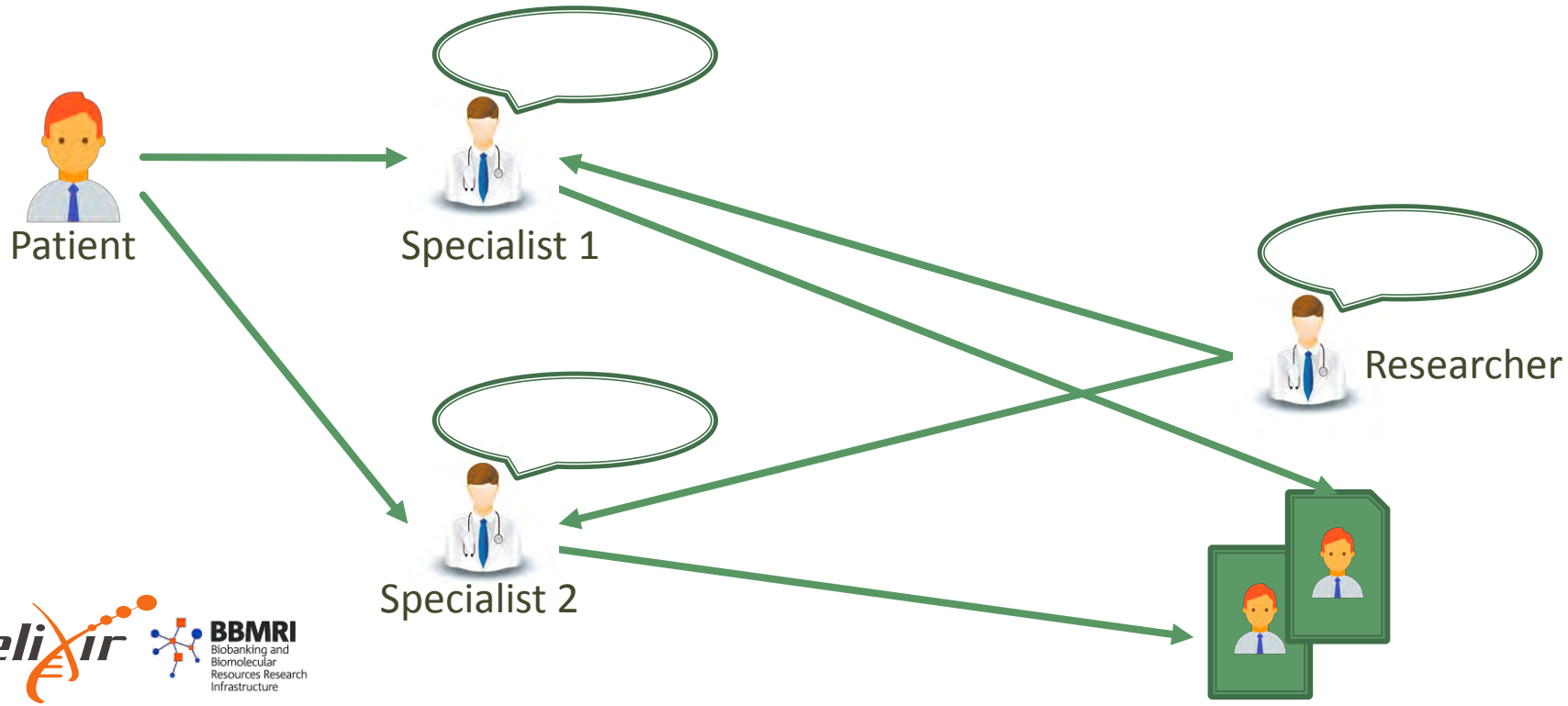
- Joint working group of IRDiRC and GA4GH
- Started in 2016
- Formerly known as PUID working group
- Aims to come up with a recommendation for a solution to the problems of research to link patient records

PPRL Working Group Status

- Evaluated the landscape resulting in two primer documents concerning the technical and ethico legal landscape
- Looking for a solution to serve the research community
- Working on basis of use cases collected at the start of the WG (Originally 7 use cases, which have been reduced to 3 use cases based on scope and duplication)

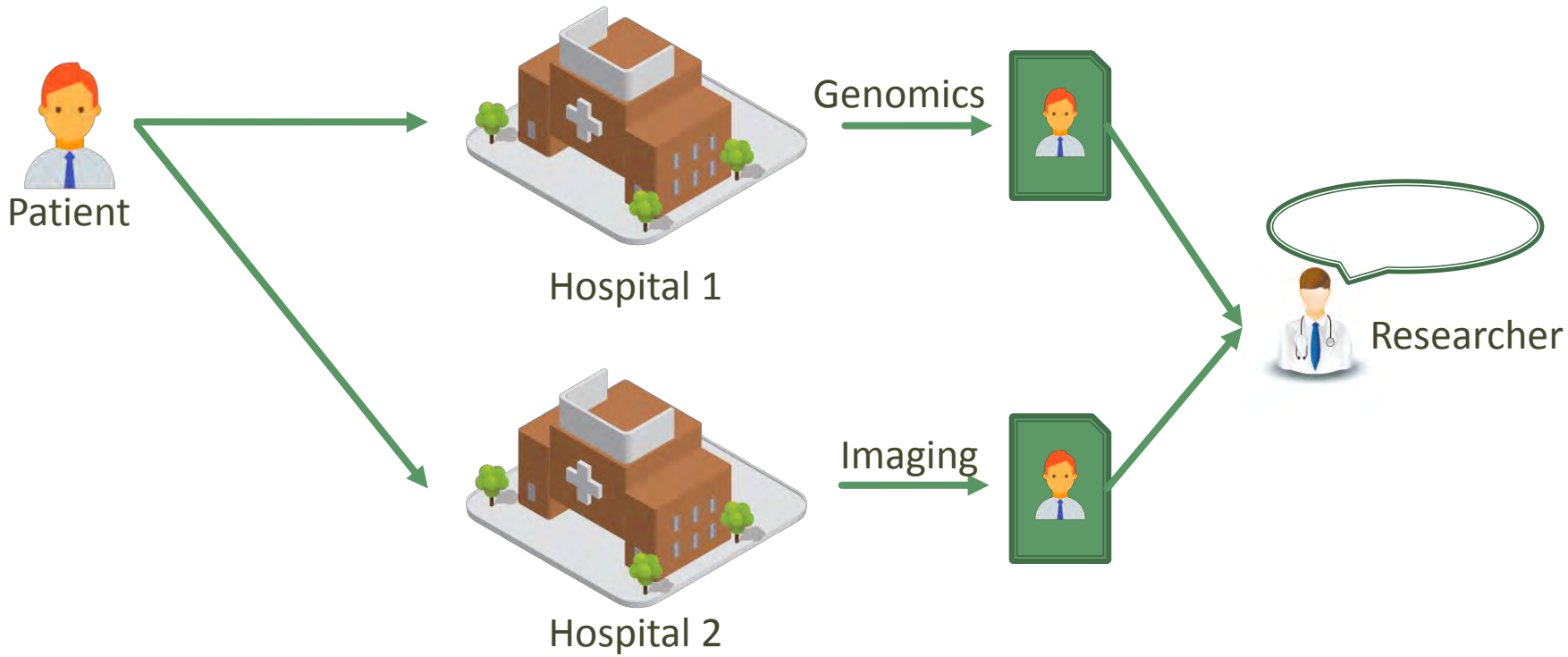
Use case 1

De-duplication of records associated with the same participant within the same database or when databases are aggregated



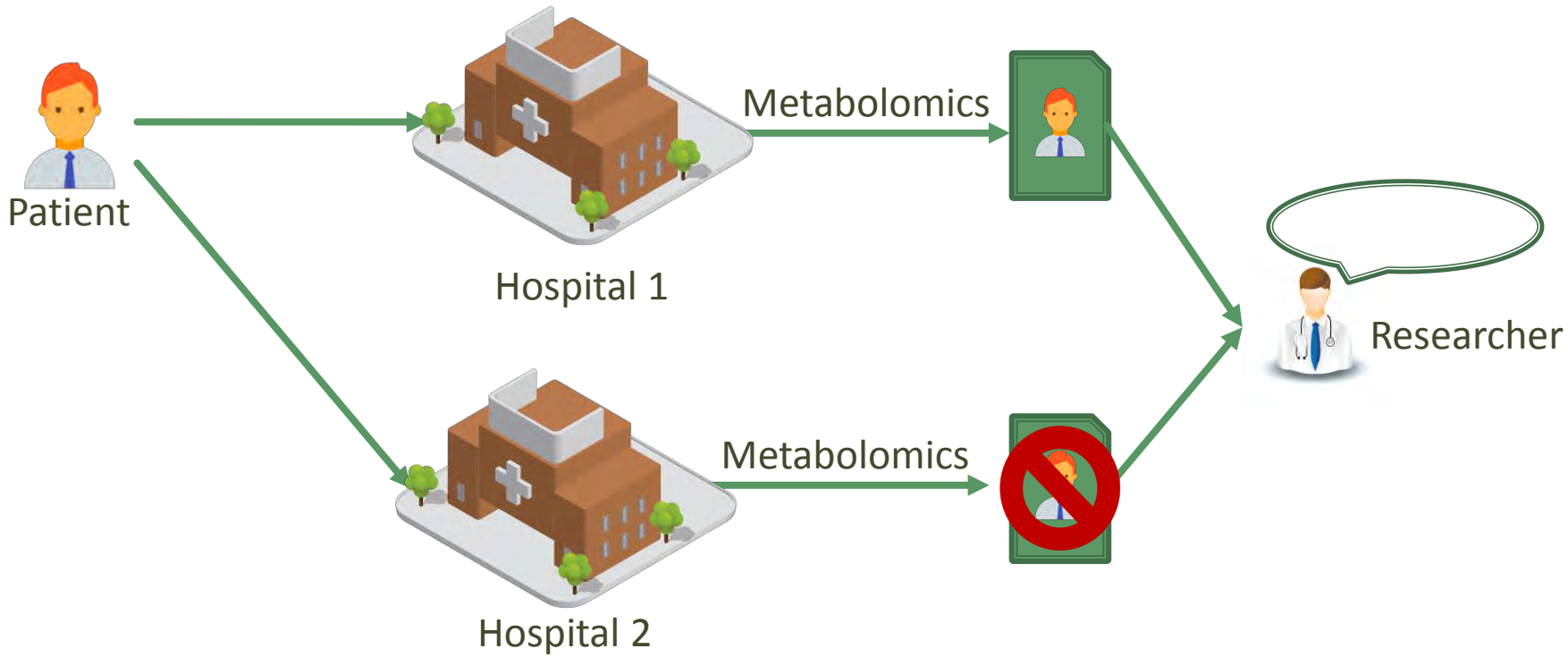
Use case 2

Discovery of additional data associated with the same individual across multiple anonymous datasets



Use case 3

Ensuring that each individual is uniquely represented in a study sample in order to increase research reliability and validity



Ideal world solution

- All these use cases would be trivial to solve if we would all use the same unique patient identifier
- However for numerous legal and ethical reasons this is not possible

Data protection laws (GDPR)

- Data protection laws apply only to data relating to an identifiable individual
- De-identified data is therefore unregulated by data protection
- Pseudonyms used for record linkage are linked to identifying information **therefor Data protection laws do apply**

Source: PPRL WG Ethico-Legal Primer

PPRL principles

Privacy **P**reserving **R**ecord **L**inkage

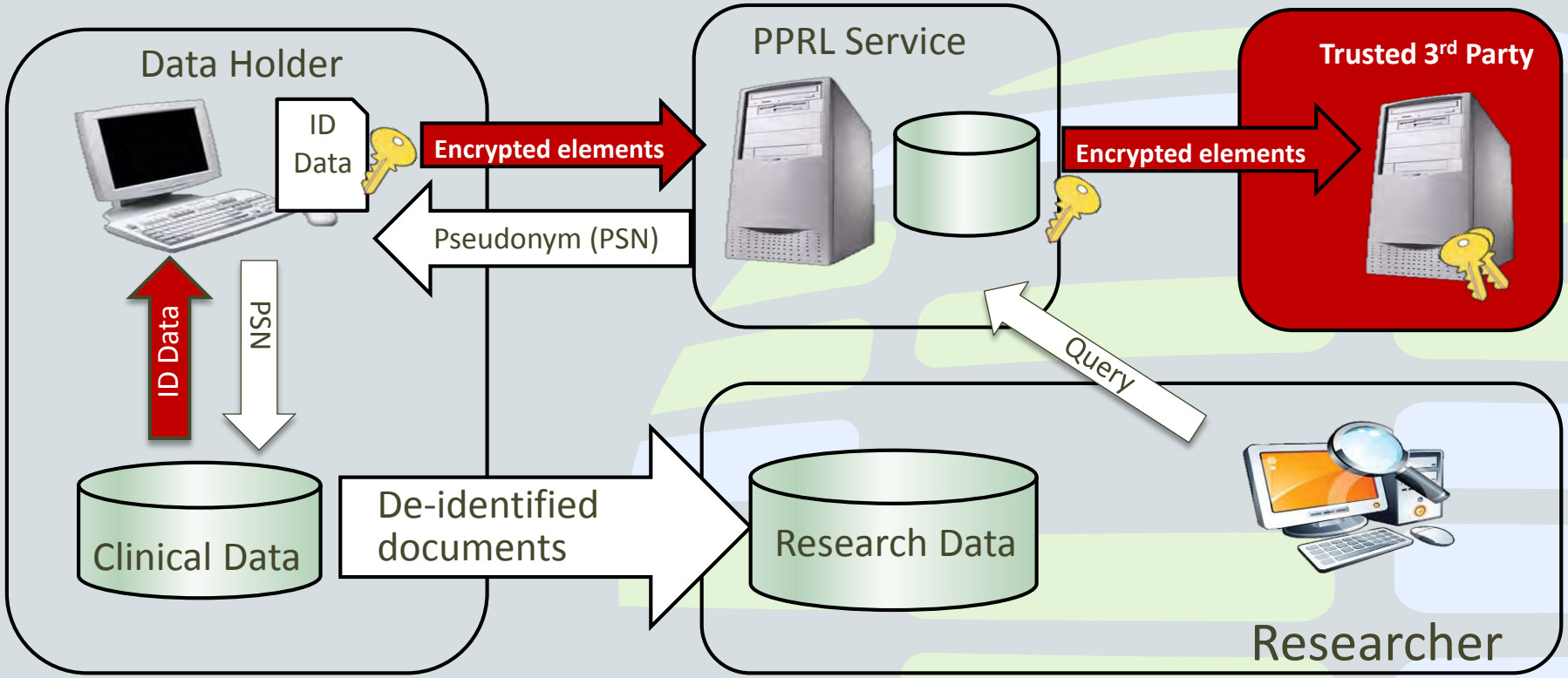
- The principle that you can link records from different sources that belong to the same individual
- Without knowing the identity of the individual
- With a high level of precision and recall
- However at the sources you have access to identifying data

Identifying data

- Direct identifiers
  - If a one-to-one mapping of individual-to-code is possible, the code is considered a direct identifier
  - E.g. national identity numbers
- A set of Quasi identifiers
  - E.g. sex, first name, last name, middle name, date of birth, and city, municipality, and country of birth

Identifiers need to be:

- Stable: will not change over time
- Available: can be asked within the scope of the consent



Simplified PPRL Workflow

PPRL behind the scenes

- Use a computational function that encodes the identifiers from the original dataset such that:
  - The encoding is not reversible (e.g. hashing function)
  - Is robust to minor data entry errors in the original identifiers (e.g. phonetic equivalence)
- Store the results in a database and link it to a independently generated pseudonym (e.g. random number)
- Return the pseudonym to the data holder

How does PPRL link data

- In the database the pseudonyms are linked to the individual (based on the encoded data)
- If authorized the system can retrieve and return other pseudonyms that are associated with the same individual

PPRL Requirements

- To link records from different sources you need to uniquely*) identify an individual in the system
- This requires at least:
  - One common direct identifier
  - Or a sufficient set of common quasi identifiers

*) To a certain level of certainty

Recommendations

- PPRL does not work in separate silos: one (federated) solution across ERNs
- PPRL depends on common identifiers at the source: standardised identifiers increase the ability to link
  - E.g. First name, last name, date of birth, location of birth (all as specified on birth certificate (or equivalent))
- Keep provenance information

Next steps

- Start a pilot project with EUPID (RD-Connect)
- Technical assessment and audit of all potential solutions (lead by GA4GH, BBMRI-ERIC)
- Standardize identifiers (all)
- Explore PPRL in FAIR data point scenarios

RD-Connect Pilot

- First planning right now, first steps:
  - Agreement on terms for pilot with AIT
  - Define use case
  - Define scale
  - Define level of technical integration

Deployment of rare disease data linkage plan

elixir

RD Connect

16 May 2017

Rare disease data linkage plan

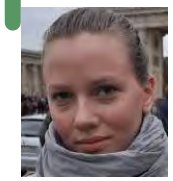
FAIR principles applied to rare disease data and software

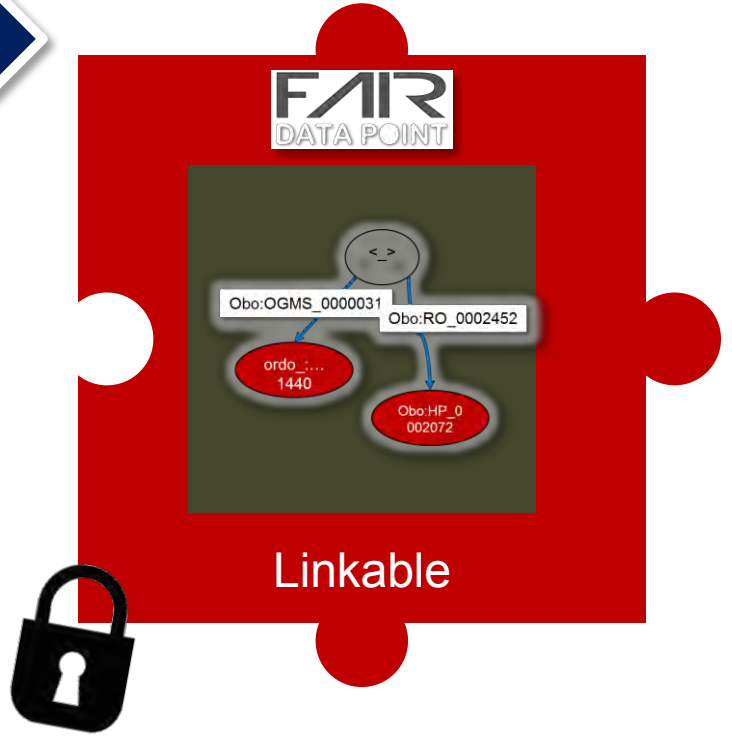
RD DATA

RD data custodian

Knowledge exchange

Domain experts & FAIR data stewards together

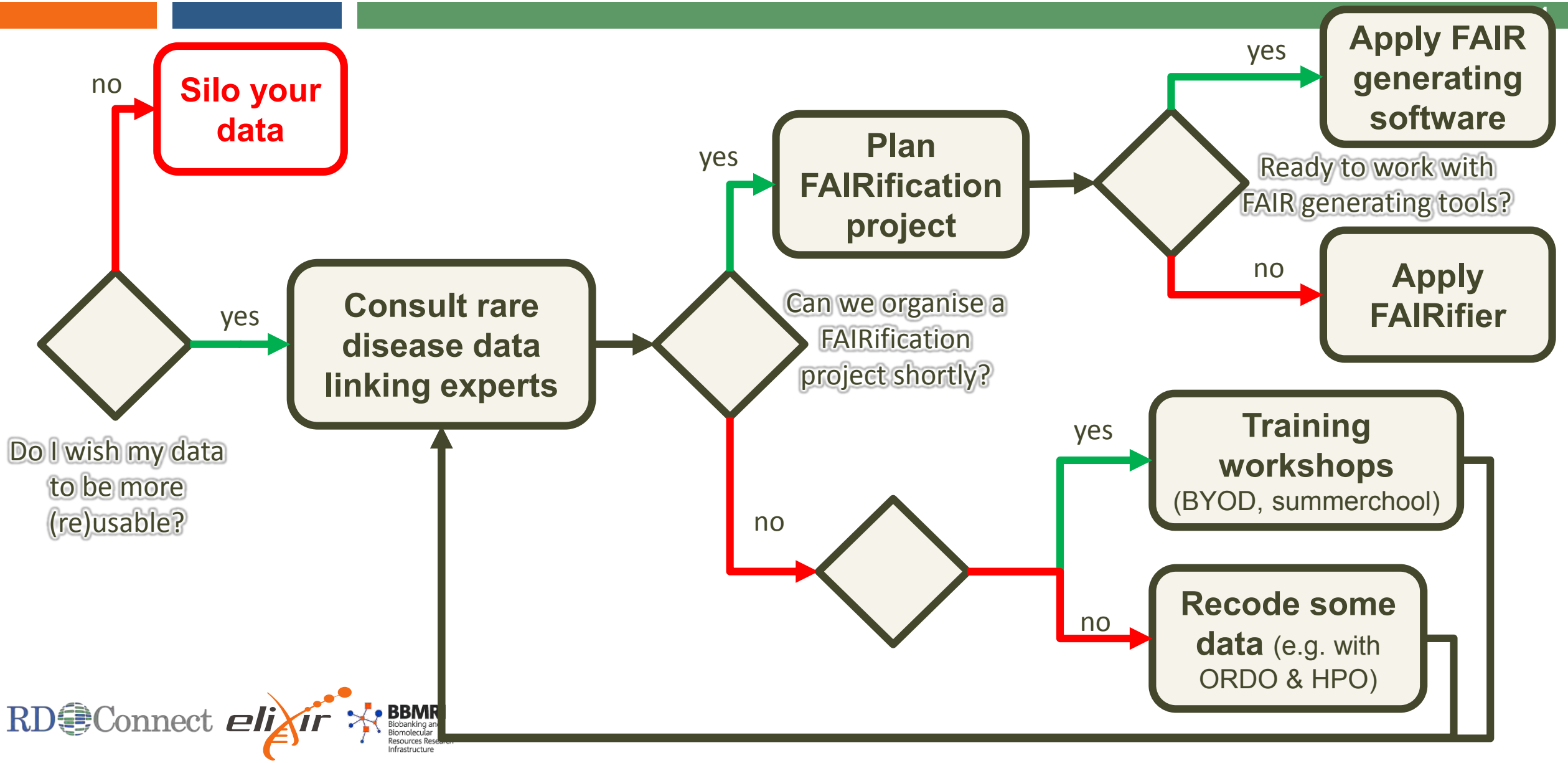




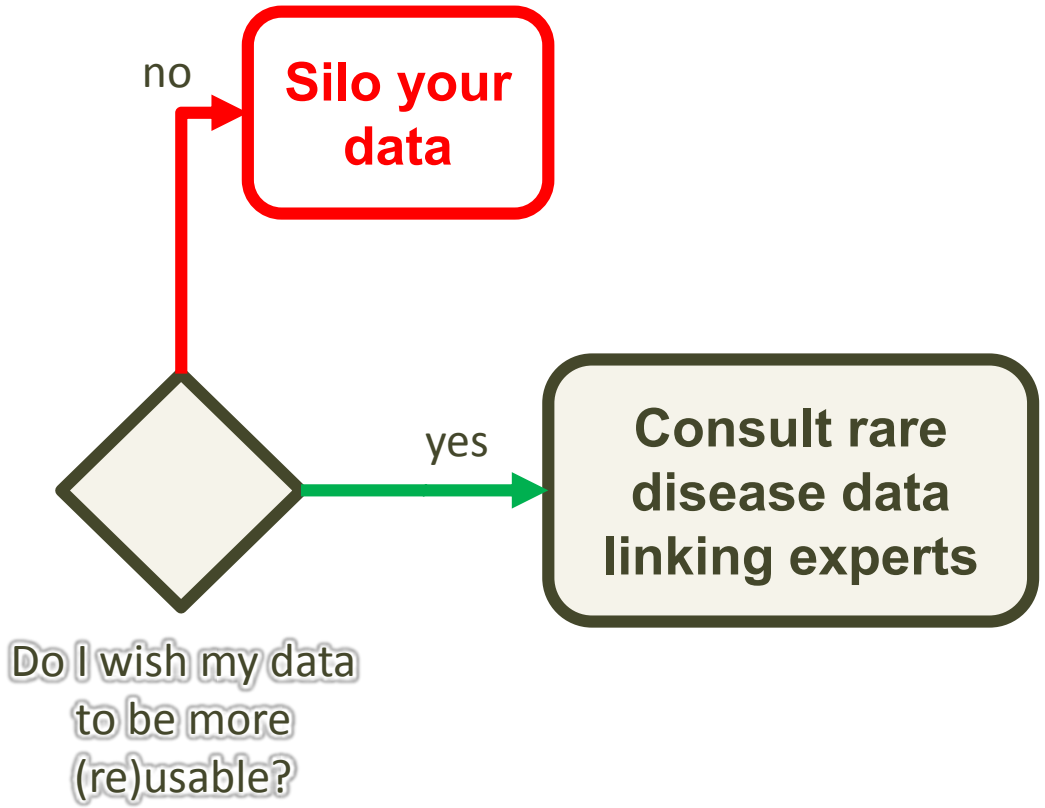
RD data custodian

RD Connect elixir BBMRI Biobanking and Biomolecular Resources Research Infrastructure

Consultancy phase



Tip for stakeholders: contact experts



fair-rd-info@elixir-europe.org

Claudio Carta, Liaison/consultant FAIR training, BYOD workshops, FAIRification in RD domain

Mark Thompson, FAIR software architect, link to FAIR engineering team

David van Enckevort, technical lead RD data linkage plan, link to RD software engineers

Mascha Jansen, Coordinator FAIR data projects & BYODs for DTL/ELIXIR-NL

Annika Jacobsen, 'Roving' FAIR data expert for rare diseases

Marco Roos, ELIXIR infrastructure for rare diseases, project lead RD data linkage plan

RD-Connect elixir BBMRI Biobanking and Biomolecular Resources Research Infrastructure

FAIRification steps

1. Define at least one ‘driving user question’
   Scope tasks and test outcome (cmp. BYOD: end-goal for example data)
2. Disambiguate data with global codes
   Apply ‘Linked Data’ to make machine-readable
3. Make accessible via a FAIR data point
   Add license + standard terms about data set
4. Test: answer driving user question

FAIRification steps: who does what?


Annika and Andra
FAIR data experts
for the rare disease case

1. Define at least one ‘driving user question’
   FAIR data steward + case expert(s)
2. Disambiguate data with global codes
   FAIR data expert (+ local trainee – preferred!)
3. Make dataset accessible via a FAIR data point
   FAIR data expert + case expert
   FAIR software developers + local IT expert(s)
4. Answer driving user question
   Case expert + FAIR data expert

Example: NMD experts in FAIR NMD case

NMD experts contribute to

1. Defining at least one ‘driving user question’
2. Choosing the right codes to represent their data
3. **Testing FAIR features of registry software RDRF**
4. Describing NMD data **set**s for a FAIR data point
5. Testing NMD FAIR data point(s)
6. Optional: deep data analytics with FAIR data

Disease area list rare disease data linkage plan

marco.roos.nl@gmail.com

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Comments Share

Neuromuscular Dystrophy

Disease area	Involved domain expert (consultant)	Stakeholder(s) (Informed)	Extra Funding (cash/ in-kind)	Summary	Social challenges			Technical challenges			Registry system				
											Excel	Castor	Osse	RDRF	PatientCrc
Huntington's Disease/ScalII (LUMC data)	Petra van Overveld	Willeke van Roon-Mom, Eleni Mina (LUMC)	Yes (in-kind)												
Neuromuscular Dystrophy	**Rachel Thompson (UNEW)**	**TREAT-NMD Global Database Advisory Committee (TGDOC). Contact Becca Leary (UNEW)**	**Yes, for domain expertise (in-kind). For SW side it will vary by SW solution.**	**65 national registries collecting a short set of clinical data items on patients with genetically confirmed DMD**	**Not all registries have the capacity (technical or financial) to update their systems**	**Some registries are somewhat protective of their data and only want it to be queryable when they say so**		**65 national registries = 65 different software systems!**	**Data items are theoretically harmonized but in practice there are a lot of differences in the way they are captured.**					**Y**	**Y**
Allan-Herndon-Dudley syndrome	Veronica Popa		TBD	Parent of a child with AHDS keen to set up a registry from scratch. Possibility to make it FAIR from the start.	Veronica has had challenges getting clinicians on board with the registry idea										
Vascular Malformation/ Hemangioma		Mariette (VSOP), Carine (Radboud),	Yes (partly)	Patient signal registry; content provided by patient (representatives). Grant								Y			

FAIR data project blueprint

Tasks

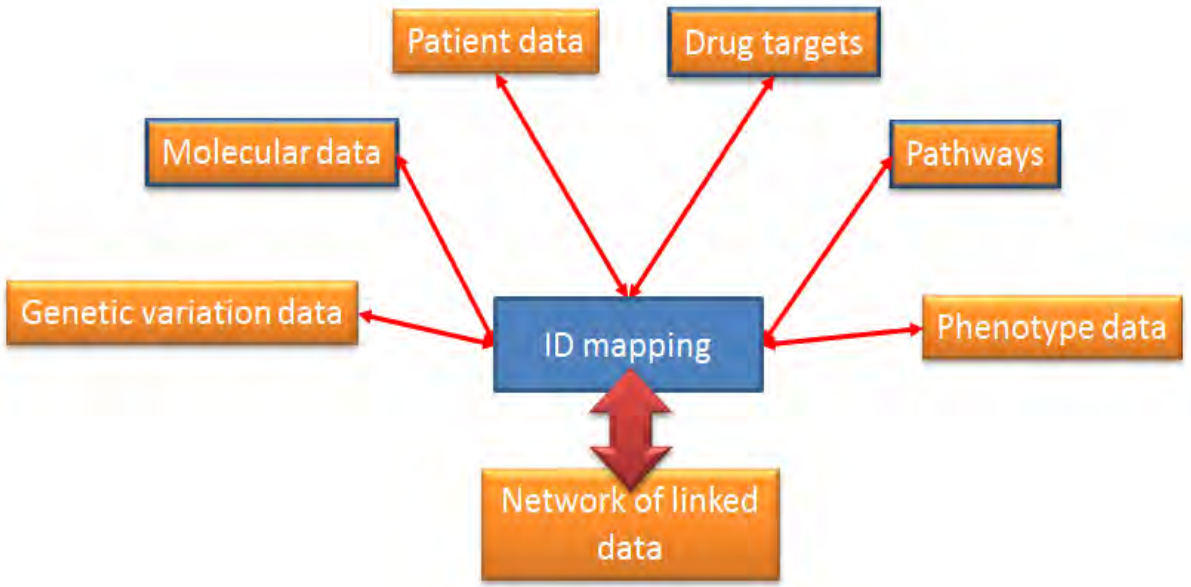
	Task	Period (months)	Associated Milestone Deliverable	Comments
Start of preparation phase	Define driving research questions and target data for FAIRification (use case)	1	M1	
	Technical assessment of requirements for the BYOD	1-2	M2	
	Select BYOD experts	3	M2	<Disease D> experts, <Disease D> database 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	4-5	M4, D1	
	Develop and test <Disease D> FAIR Data Point	4-11	M4, D1	Agile development including 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 scientists
End of	Release documented	11-12	D1	By <Disease D>

- Timeline
- List of stakeholder benefits
- Effort & Budget
  - 6-10PM (case dependent)
  - ~50% covered by infrastructures
  - ~50% in-kind/in-cash/via third party
- Stand-alone or as DMP (contribute to RD FAIR data service)

RD-Connect elixir BBMRI Biobanking and Biomolecular Resources Research Infrastructure

ELIXIR implementation study

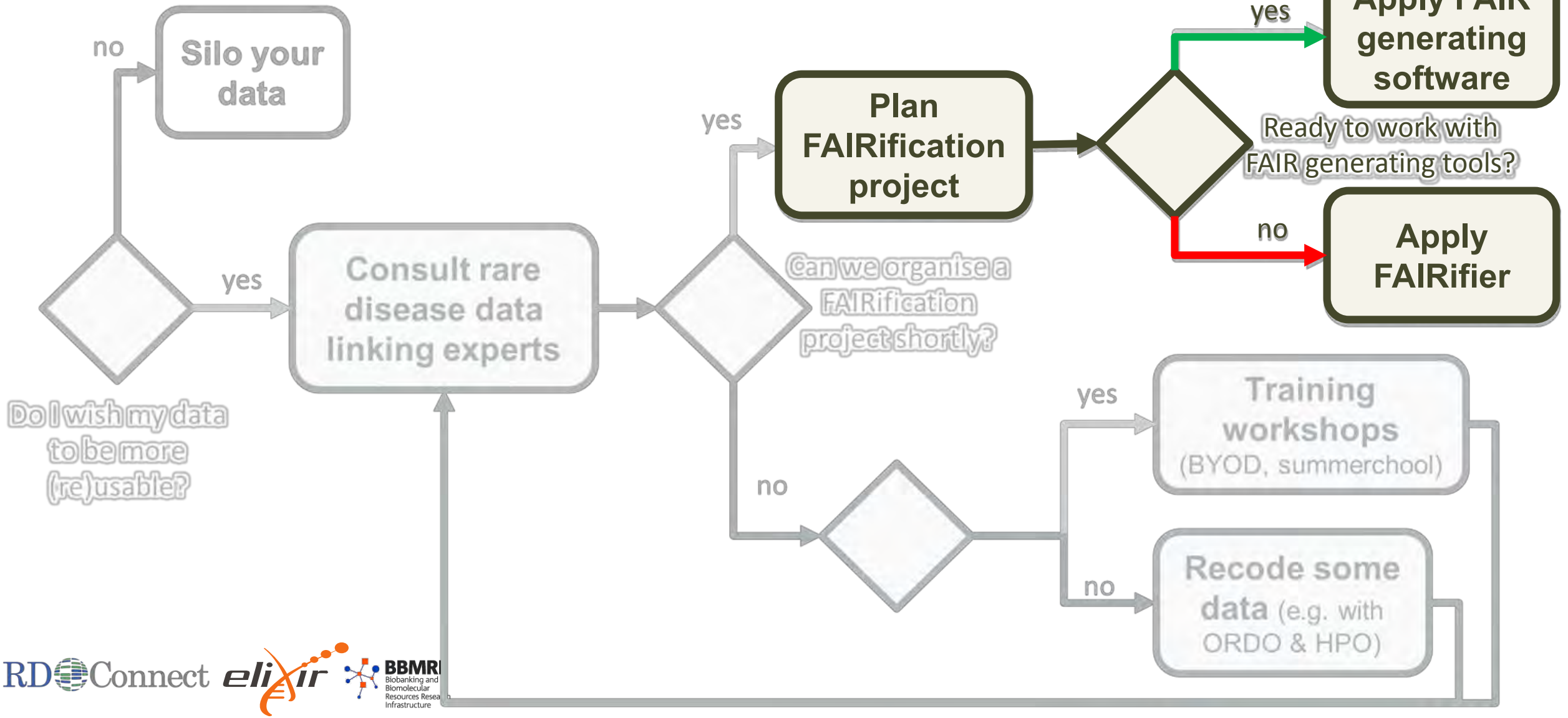
Phenotypes-genotypes-pathways (RETT syndrome)



- Solve driving questions, e.g. *"Which pathways are affected in Rett syndrome, including model systems?"*
- Evaluate usability of data sources (EGA, LOVD, Orphanet)
- Interoperability status and workflow for complex molecular data queries
- Variant to variant mapping
- Update knowledge resources

Study leader: Frederieke Ehrhart (Maastricht University)

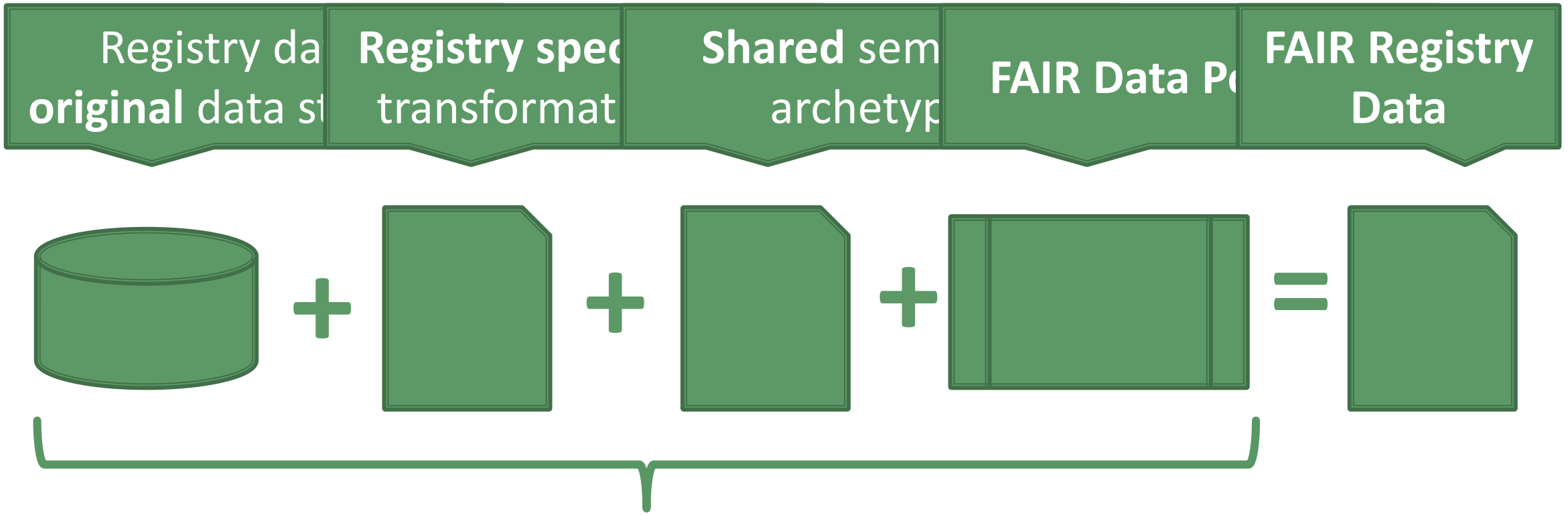
FAIRification phase



FAIR Architecture Brainstorm

- Goal:
  - Common architecture FAIR Data generation in Registry Software
  - Sharing semantic archetypes and transformations

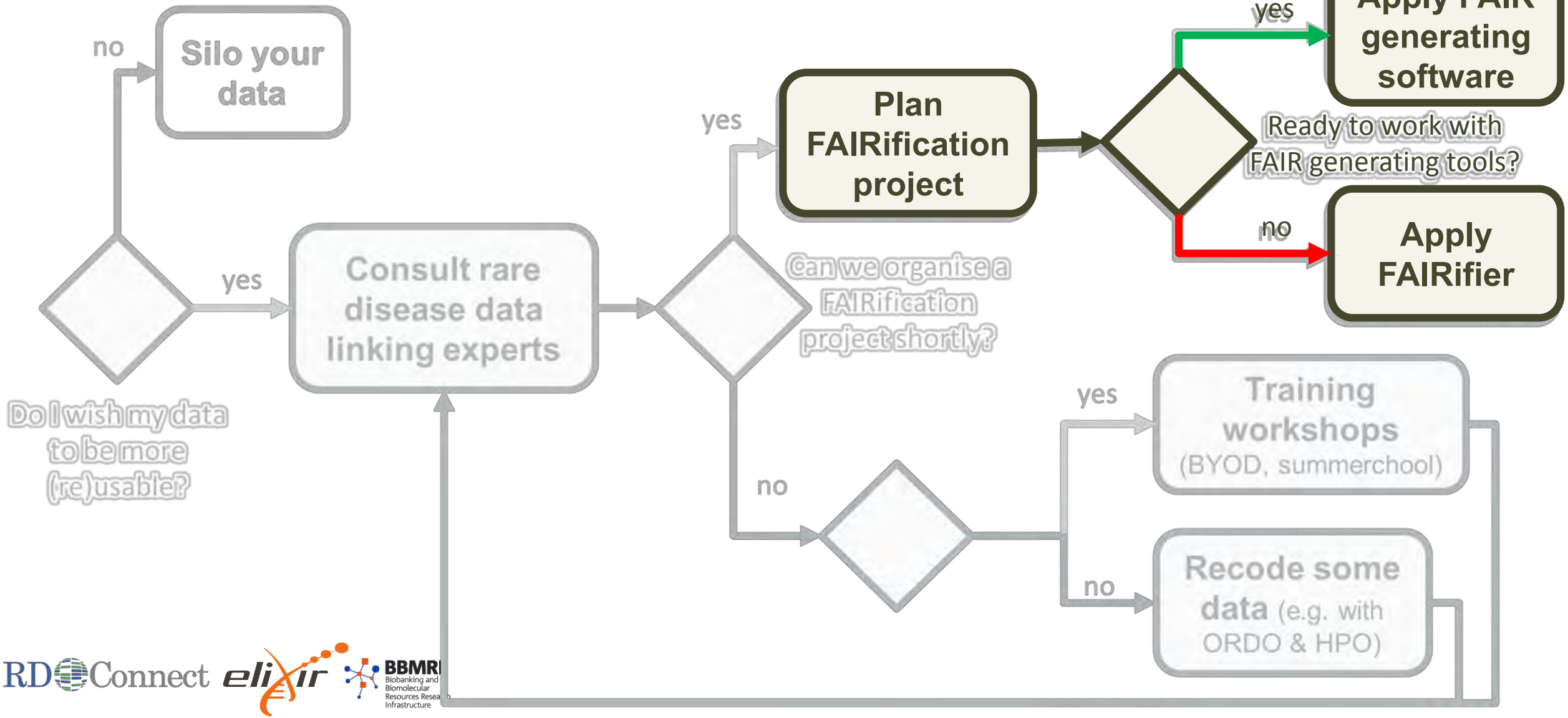
Proposed architecture



FAIR Registry Software: RDRF, OSSE, Castor…

RD Connect elixir BBMRI Biobanking and Biomolecular Resources Research Infrastructure

FAIRification phase



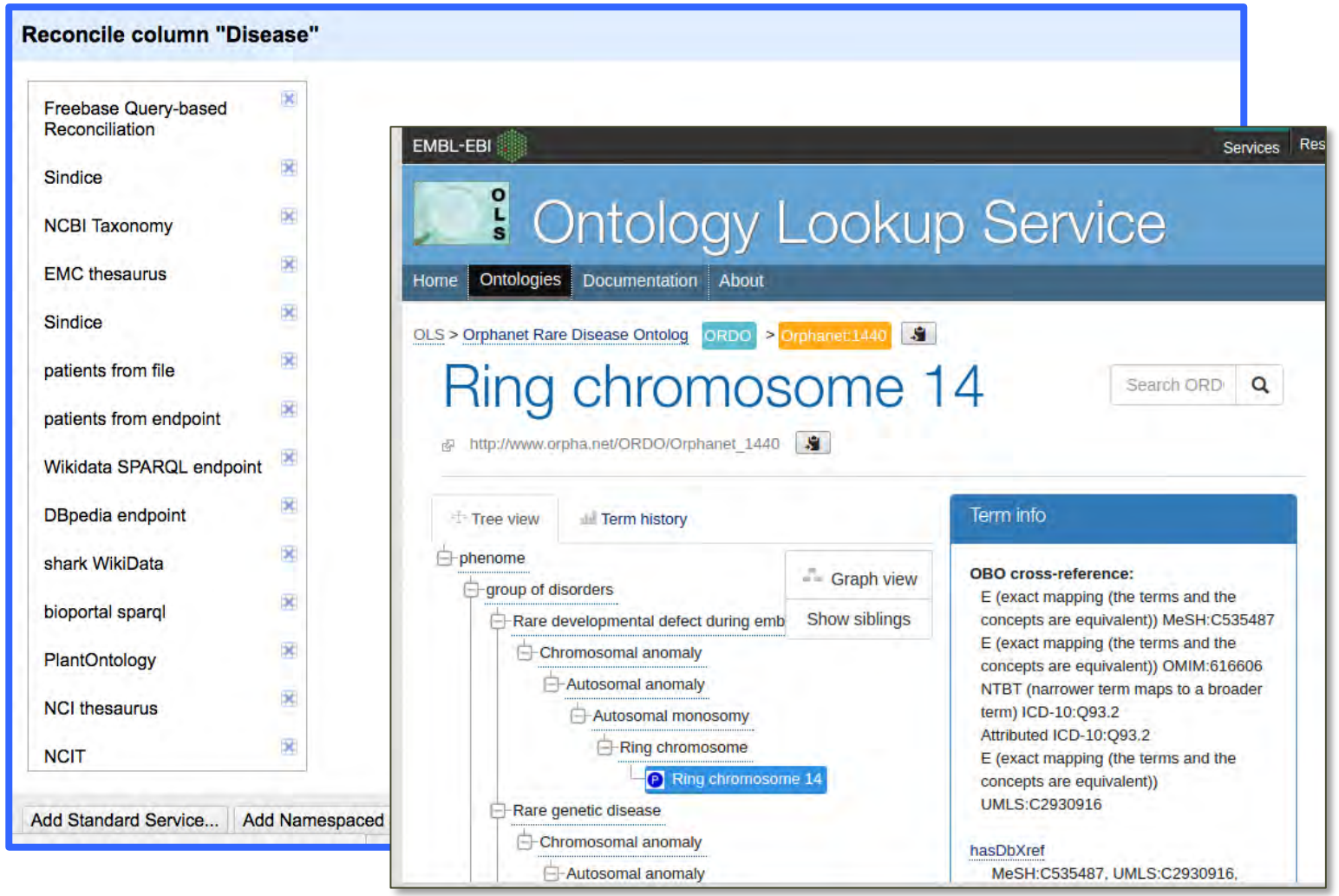
FAIRifier

102 rows Extensions: undefined RDF

Show as: **rows** records Show: 5 10 25 **50** rows « first ‹ previous **1 - 50** next › last »

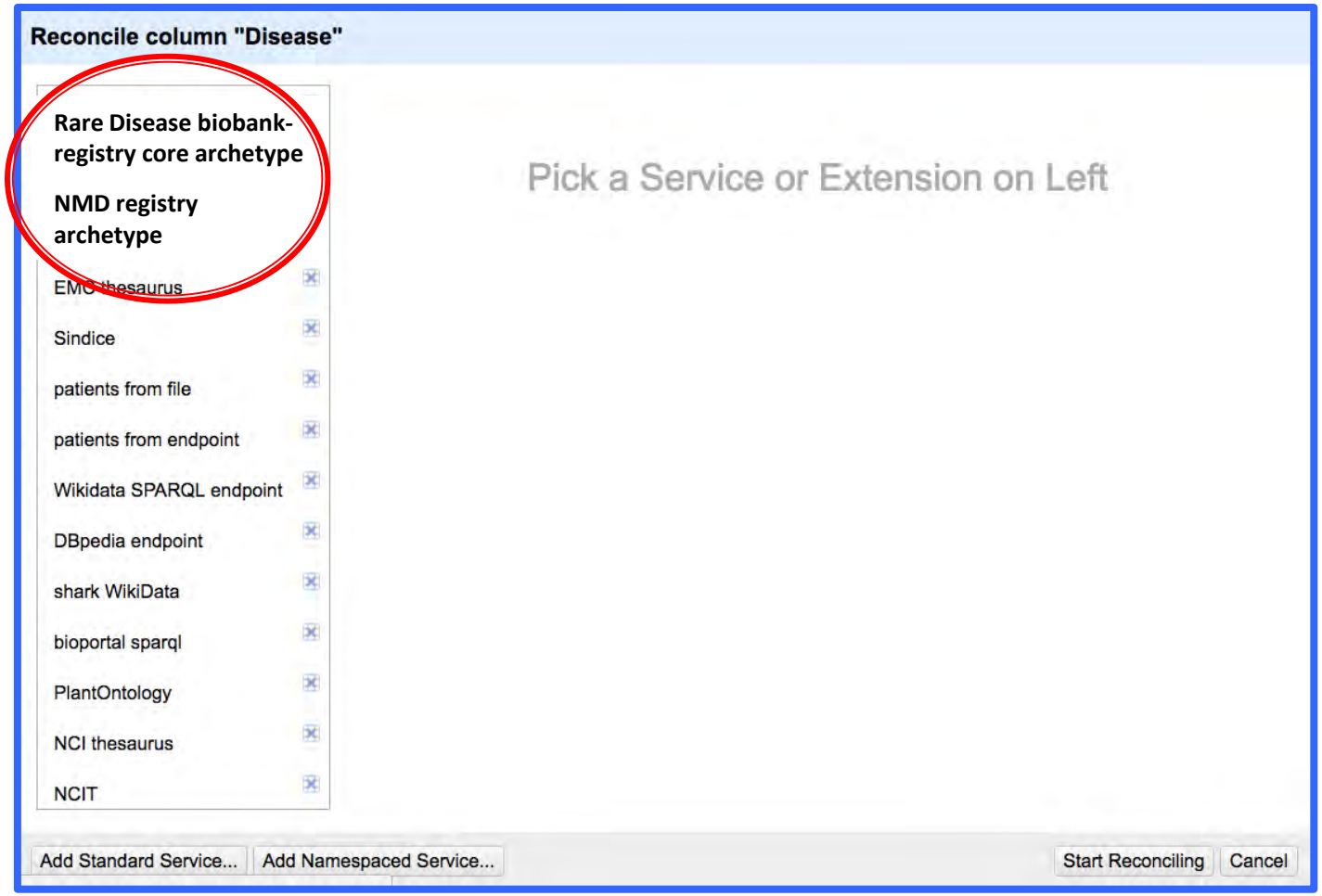
Sample ID	Sample type	Gender	Disease	MOLECULAR KA	BREAKPOINT L	Diagnosis type	Export to cata
5041	DNA	M	CHROMOSOME RING 14	T_DEL	DISTAL	Molecular, Cytogenetics	Yes
5059	Leukocytes	M	CHROMOSOME RING 14	T_DEL	DISTAL	Molecular, Cytogenetics	Yes
5622	Lymphoblast	M	CHROMOSOME RING 14	T_DEL	DISTAL	Molecular, Cytogenetics	Yes
5543	DNA	M edit	CHROMOSOME RING 14	T_DEL	DISTAL	Molecular, Cytogenetics	Yes
5018	Leukocytes	M	CHROMOSOME RING 14	T_DEL	DISTAL	Molecular, Cytogenetics	Yes
5019	Lymphoblast	M	CHROMOSOME RING 14	T_DEL	DISTAL	Molecular, Cytogenetics	Yes
3088	DNA	F	CHROMOSOME RING 14	T_DEL	DISTAL	Molecular, Cytogenetics	Yes
3097	Leukocytes	F	CHROMOSOME RING 14	T_DEL	DISTAL	Molecular, Cytogenetics	Yes
3098	Lymphoblast	F	CHROMOSOME RING 14	T_DEL	DISTAL	Molecular, Cytogenetics	Yes
4729	Leukocytes	F	CHROMOSOME RING 14			Cytogenetics	Yes
3203	Lymphoblast	F	CHROMOSOME RING 14			Cytogenetics	Yes
4342	Leukocytes	M	CHROMOSOME RING 14	DEL/DUP	DISTAL	Molecular, Cytogenetics	Yes
3223	Lymphoblast	M	CHROMOSOME RING 14	DEL/DUP	DISTAL	Molecular, Cytogenetics	Yes
3262	DNA	M	CHROMOSOME RING 14	DEL/DUP	DISTAL	Molecular, Cytogenetics	Yes
3277	Lymphoblast	F	CHROMOSOME RING 14	T_DEL	DISTAL	Molecular, Cytogenetics	Yes
3283	DNA	F	CHROMOSOME RING 14	T_DEL	DISTAL	Molecular, Cytogenetics	Yes
3299	Leukocytes	F	CHROMOSOME	T_DEL	DISTAL	Molecular,	Yes

- Load original data into FAIRifier



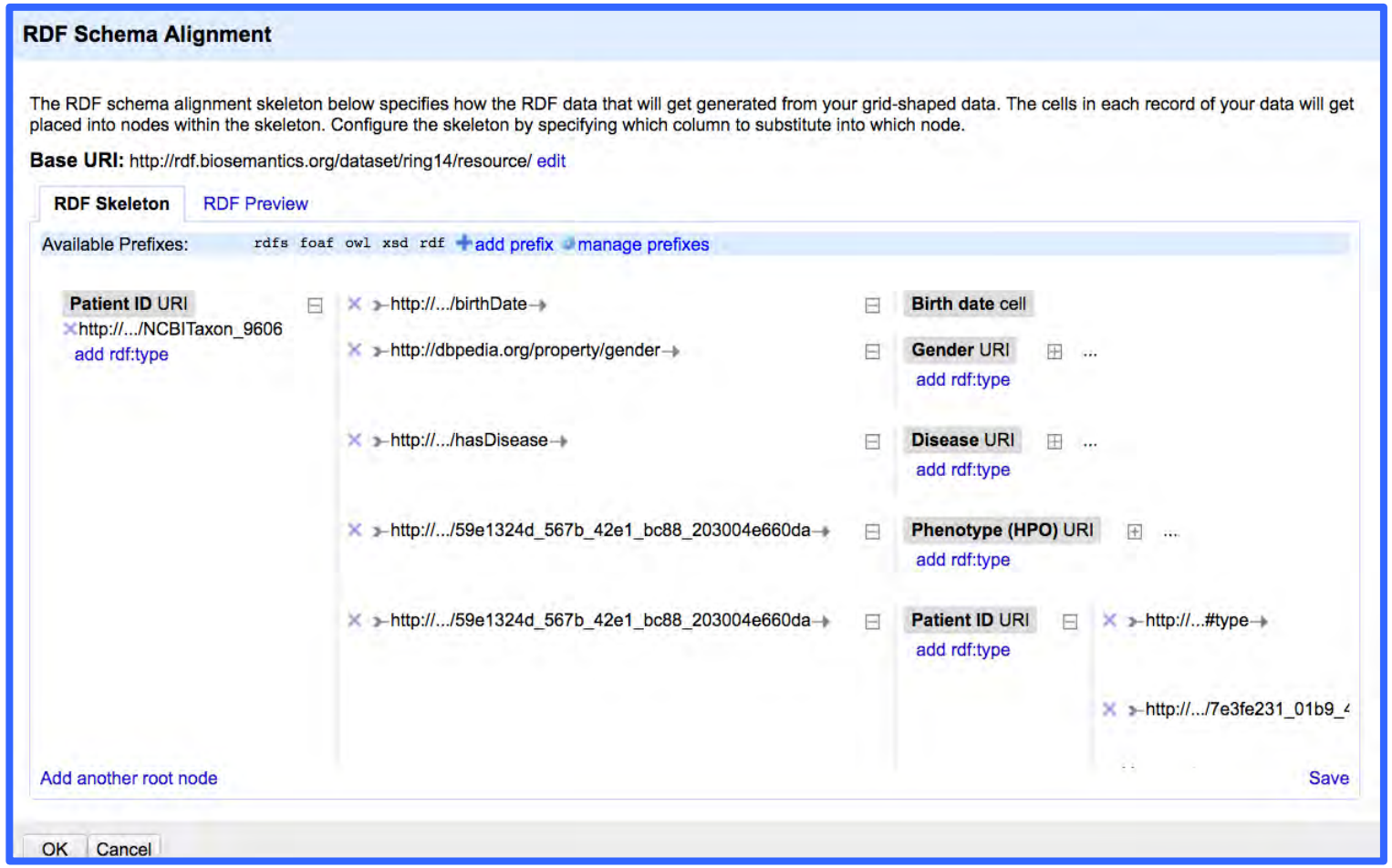
- Recode data with ontology terms
- Use OLS to identify correct terms

FAIRifier

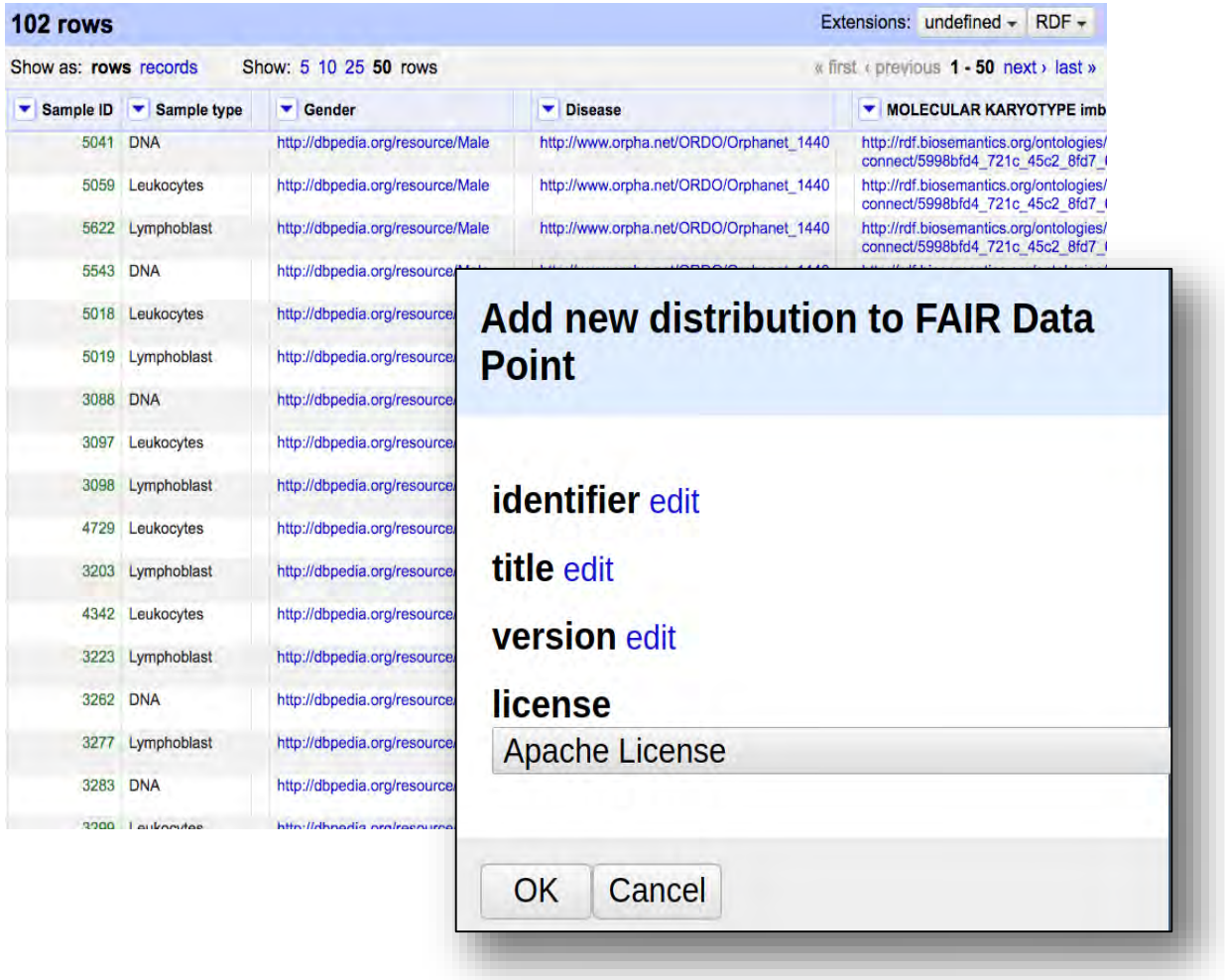


- Soon: choose from semantic archetypes shared by other FAIR data stewards

FAIRifier



- Add and check semantic relations for your data



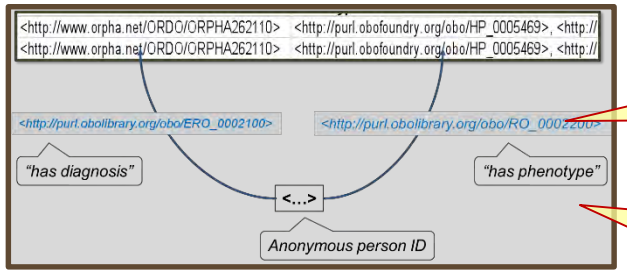
- Make linkable data available via a FAIR data point

Centred around standards

FIND, ACCESS

INTEROPERATE
REUSE

Data access protocol
(REST)



Disorder	Symptoms
Chromosome 14 disease	Epileptic attacks as young child
Deletion of part of the long arm of chromosome 14	Learning deficit and growth problems

Semantic access control model

Metadata
DCAT, bioschema.org

Shared semantic archetype
ORDO+HPO+OBIB+ORE, EDAM+IAO
https://biosharing.org/bsg-s000676

Linked data format

Original data format

Experiment: Record linking

1. Data linking experiment: find a treatment for Monika
2. Privacy-preserving data linking experiment: count Ring-14 cases; stratify patients by phenotype

Experiment: find the treatment!

‘Patients’ with a disease, a phenotype, and some with a treatment

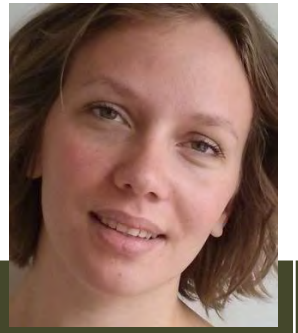
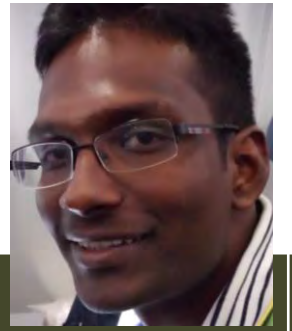



Disclaimer: mock examples!!!!

Experiment: find the treatment!

Which treatment <?T> ***is treating*** phenotype <?P> that is also a phenotype for Monica who ***has disease*** <?D>

Find the treatment experiment

Monika 	Annika 	Rajaram 	Pietro 	FAIR
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	Local registry

Find the treatment experiment

Monika	Annika	Rajaram	Pietro	FAIR
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	Local registry
Ring-14 disease, Salaam seizures, (no treatment)	Ring-14 syndrome, Chorea, (no treatment)	Perry syndrome, Epileptic attacks, lamotrigine	Perry syndrome, extreme mood swings, oxalate	English

Find the treatment experiment

Monika	Annika	Rajaram	Pietro	FAIR
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	Local registry
Ring-14 disease, Salaam seizures, (no treatment)	Ring-14 syndrome, Chorea, (no treatment)	Perry syndrome, Epileptic attacks, lamotrigine	Perry syndrome, extreme mood swings, oxalate	English

Find the treatment experiment

Which treatment <?T> *is treating* phenotype <?P> in a person who *has disease* Ring-14 disease and *has phenotype* <?P>

Monika	Annika	Rajaram	Pietro	FAIR
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	Local registry
Ring-14 disease, Salaam seizures, (no treatment)	Ring-14 syndrome, Chorea, (no treatment)	Perry syndrome, Epileptic attacks, lamotrigine	Perry syndrome, extreme mood swings, oxalate	English

Scenario

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

Monika	Annika	Rajaram	Pietro	FAIR
Ring-14 disease, Salaam seizures, (no treatment)	Ring-14 disease, Chorea, (no treatment)	Perry syndrome, Epileptic attacks, lamotrigine	Perry syndrome, extreme mood swings, oxalate	English
ORPHA1440, HP:0011097	ORPHA1440, HP:0011097	ORPHA178509, HP:0011097, CHEBI_6367	ORPHA178509, HP:0000720 CHEBI_132952	Coded

Putting the pieces together

Which treatment <?T> ***is treating*** phenotype <?P> in any person who ***has disease*** Ring-14 disease and ***has phenotype*** <?P>

ORPHA178509, HP:0000720, CHEBI_132952

ORPHA1440, HP:0011097

ORPHA178509, HP:0011097, CHEBI_6367

ORPHA1440, HP:0002072

Scenario

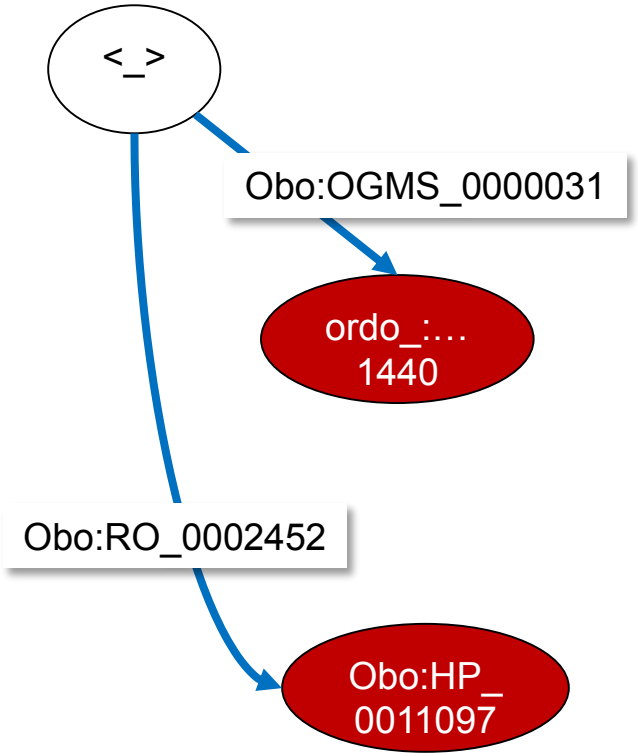
Monika	Annika	Rajaram	Pietro	FAIR
ORPHA1440, HP:0011097	ORPHA72, HP:00027072	ORPHA178509, HP:0011097, CHEBI_6367	ORPHA178509, HP:0000720 CHEBI_132952	Coded
Monika has disease Ring-14 disease, and has phenotype Salaam seizures	Annika has disease Ring-14 disease, and has phenotype Chorea	Rajaram has disease 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	full meaning

Scenario

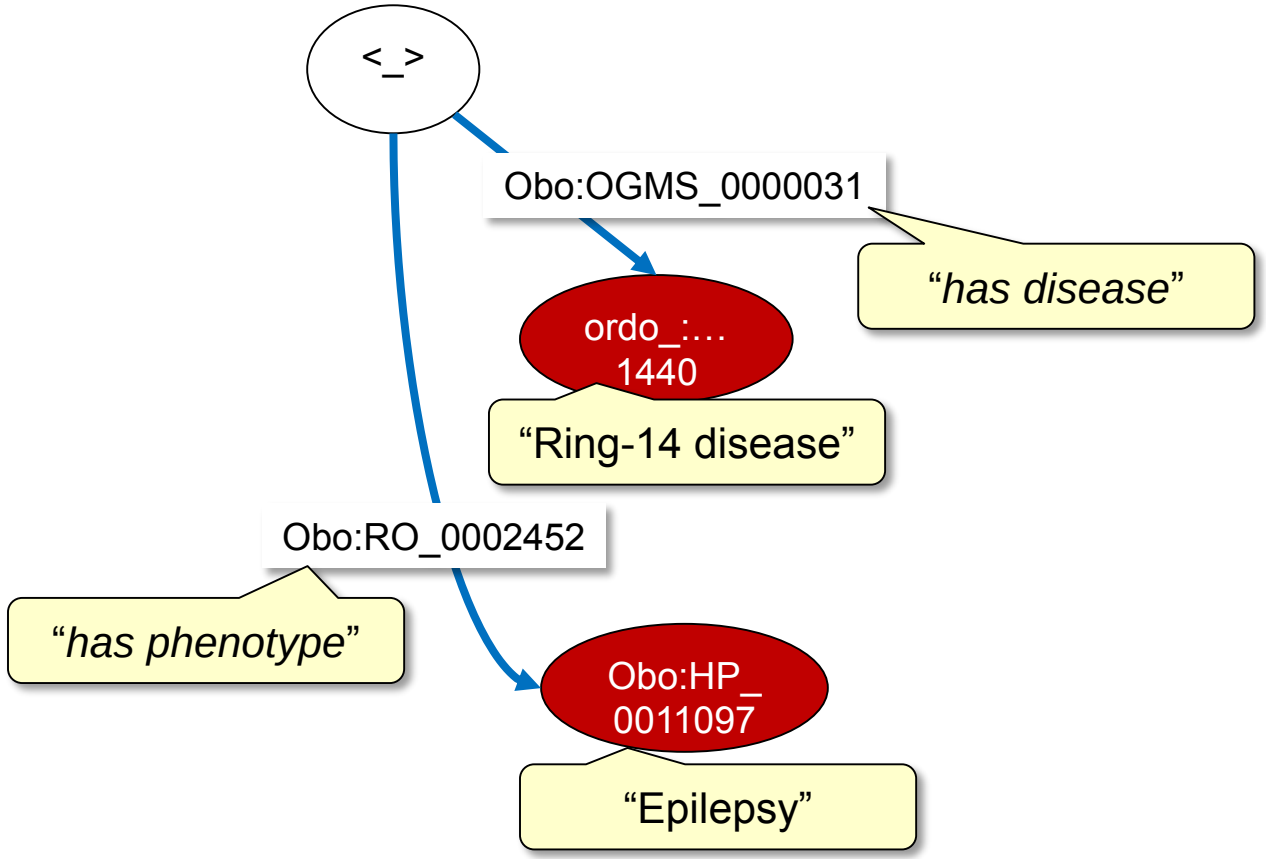
obo: http://purl.obolibrary.org/obo/
ordo: http://www.orpha.net/ORDO/

Monika	Annika	Rajaram	Pietro	FAIR
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 the drug lamotrigine	Pietro has disease Ring-14 disease, and has phenotype Extreme mood swings. Extreme mood swings are treated by the drug Oxalate	Full meaning
<_> 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	Interoperable & Machine readable

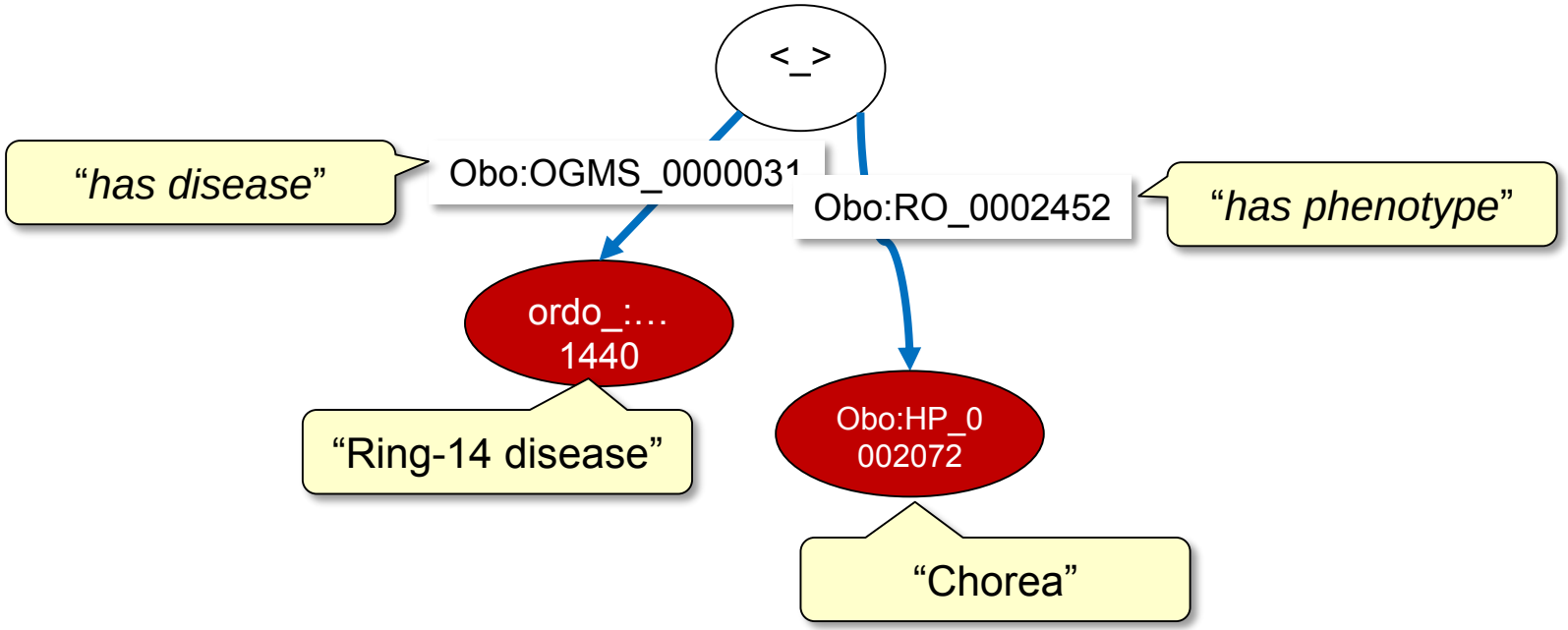
Monika	Annika	Rajaram	Pietro
<_> obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0011097.	<_> obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0002072.	<_> obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0011097 obo:RO_0002302 obo:CHEBI_33237	<_> obo:OGMS_0000031 ordo:Orphanet_178509,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



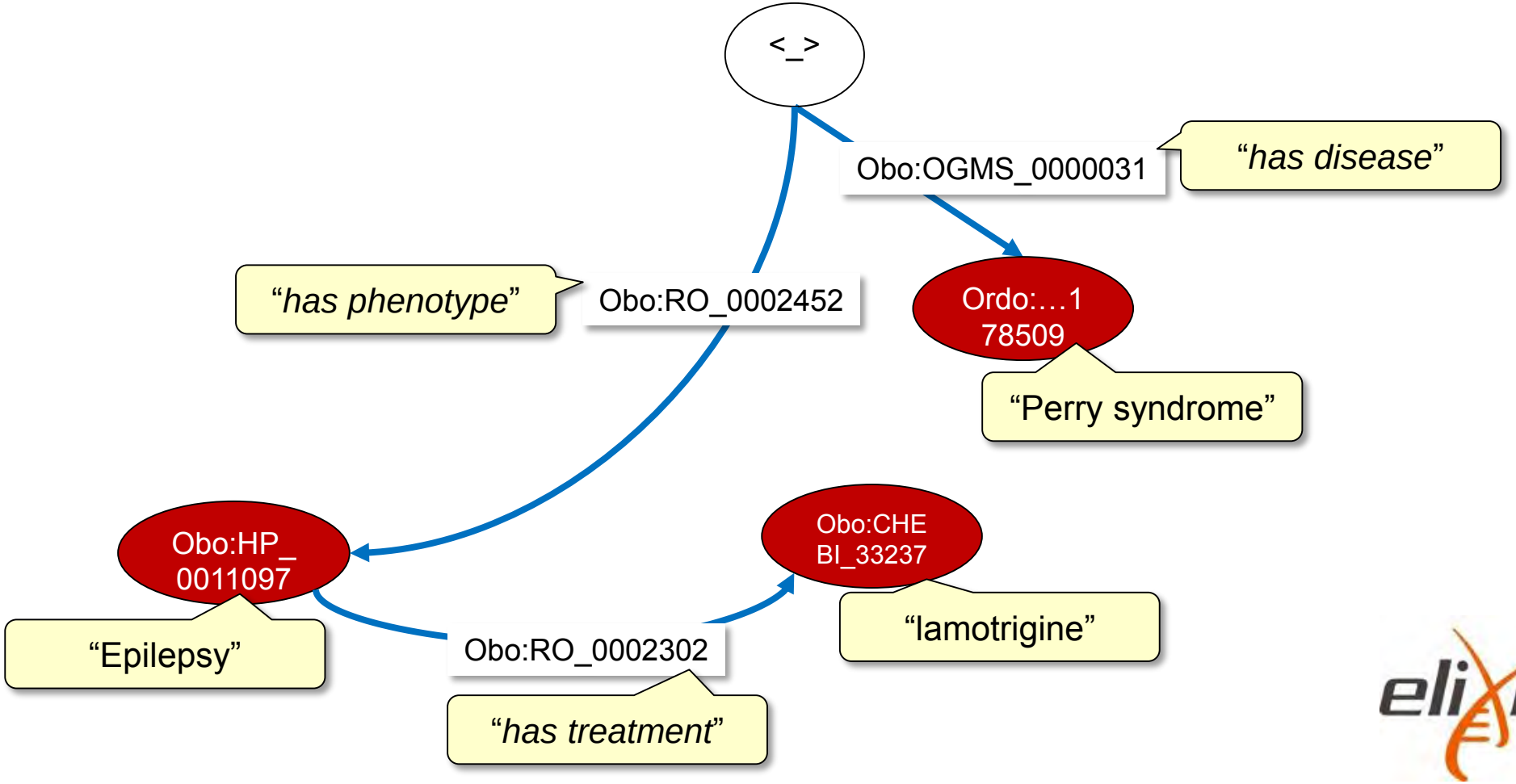
Monika	Annika	Rajaram	Pietro
<_> obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0011097.	<_> obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0002072.	<_> obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0011097 obo:RO_0002302 obo:CHEBI_33237	<_> obo:OGMS_0000031 ordo:Orphanet_178509,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



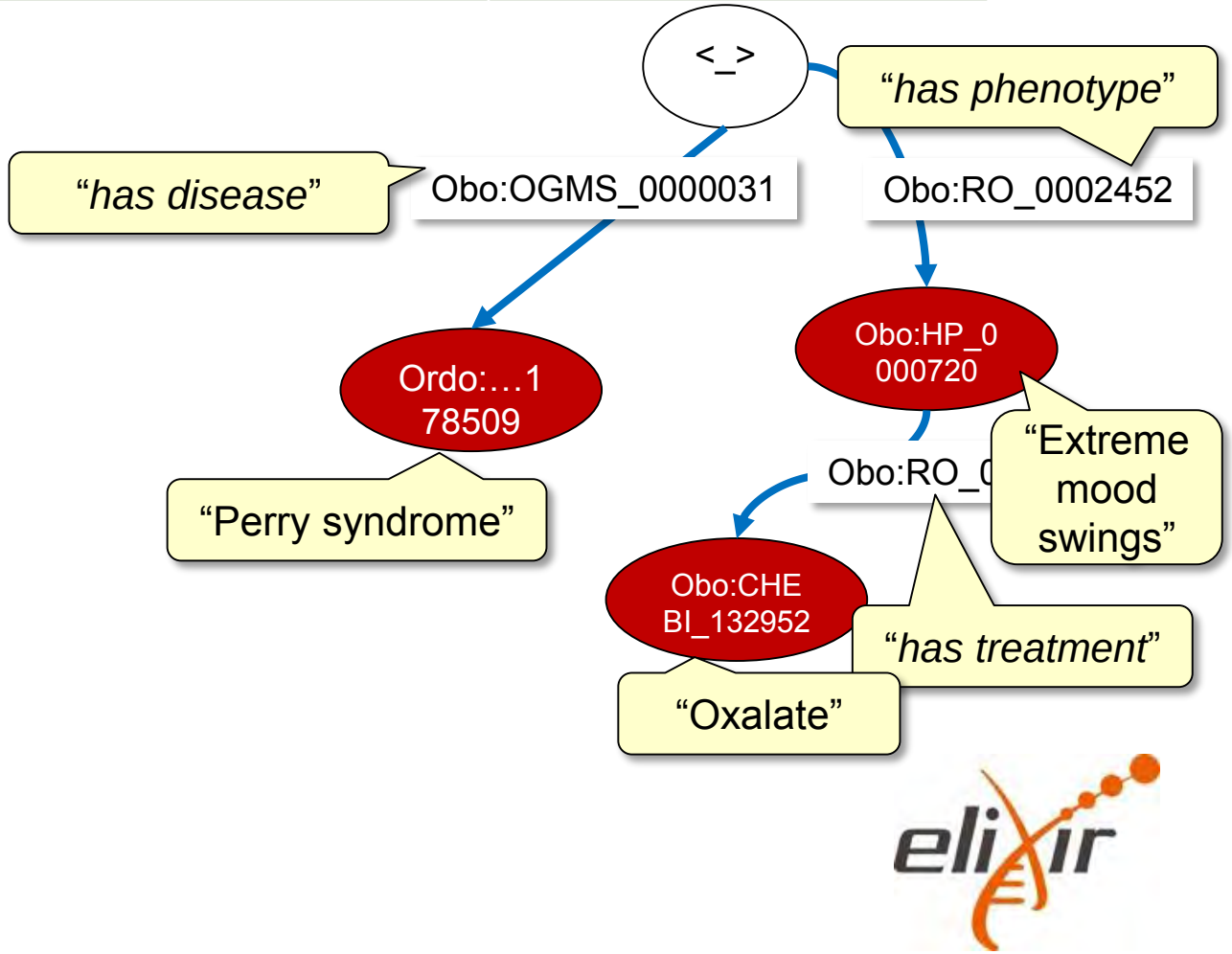
Monika	Annika	Rajaram	Pietro
<_> obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0011097.	<_> obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0002072.	<_> obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0011097 obo:RO_0002302 obo:CHEBI_33237	<_> obo:OGMS_0000031 ordo:Orphanet_178509,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



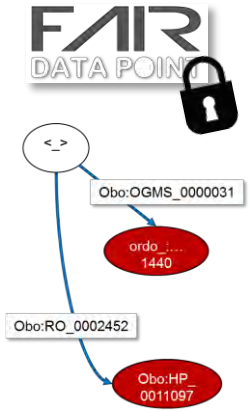
Monika	Annika	Rajaram	Pietro
<_> obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0011097.	<_> obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0002072.	<_> obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0011097 obo:RO_0002302 obo:CHEBI_33237	<_> obo:OGMS_0000031 ordo:Orphanet_178509,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

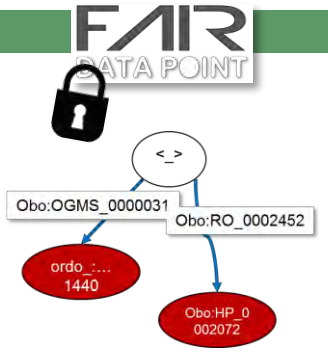


Monika	Annika	Rajaram	Pietro
<_> obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0011097.	<_> obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0002072.	<_> obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0011097 obo:RO_0002302 obo:CHEBI_33237	<_> obo:OGMS_0000031 ordo:Orphanet_178509,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



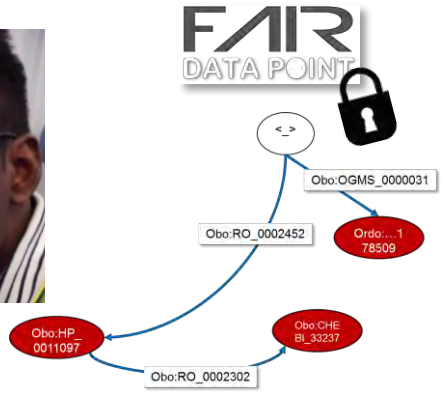
FAIR data landscape

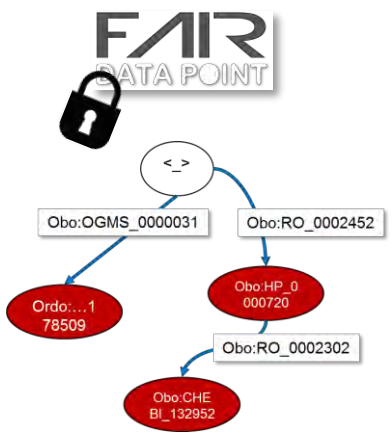




At this stage we have four **independently FAIR*** data sources, under control of the local data manager

(e.g. HCP, patient organisation, patient)

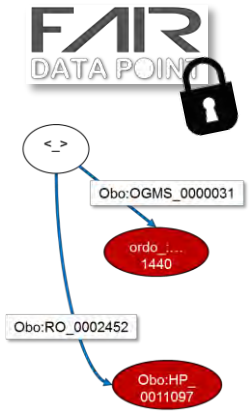




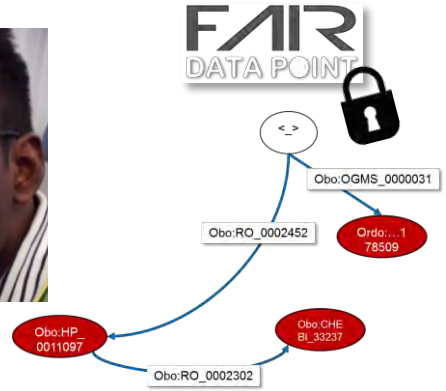
RD Connect elixir BBMRI Biobanking and Biomolecular Resources Research Infrastructure

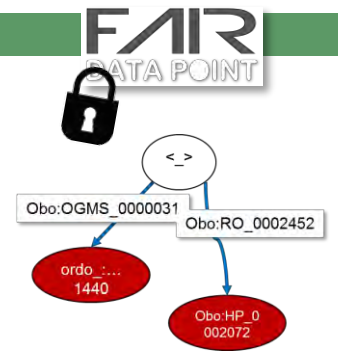
* Strictly speaking, the **interoperability** aspect mostly

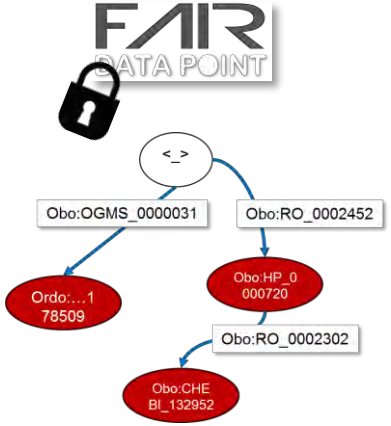
FAIR data landscape



Data at each source is **self-explaining** through global standards that a computer program can understand

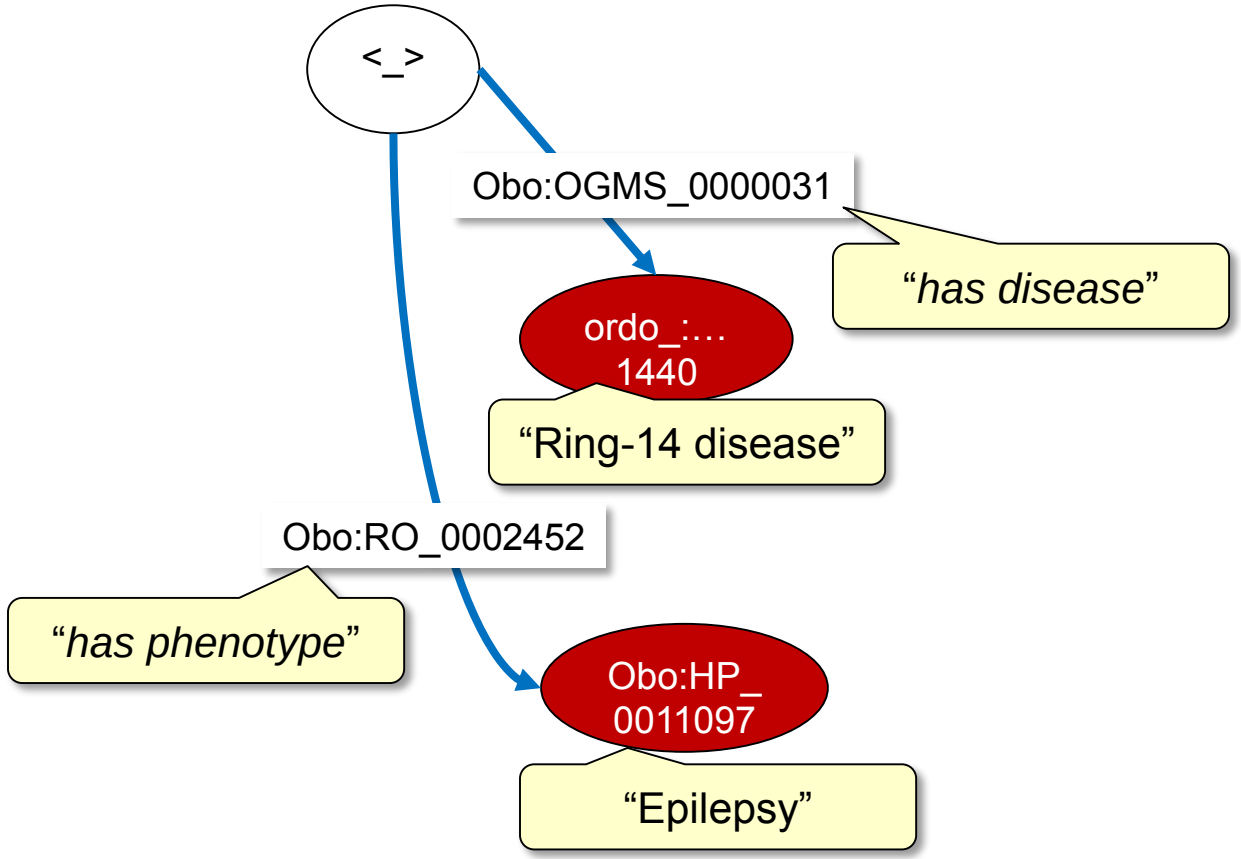




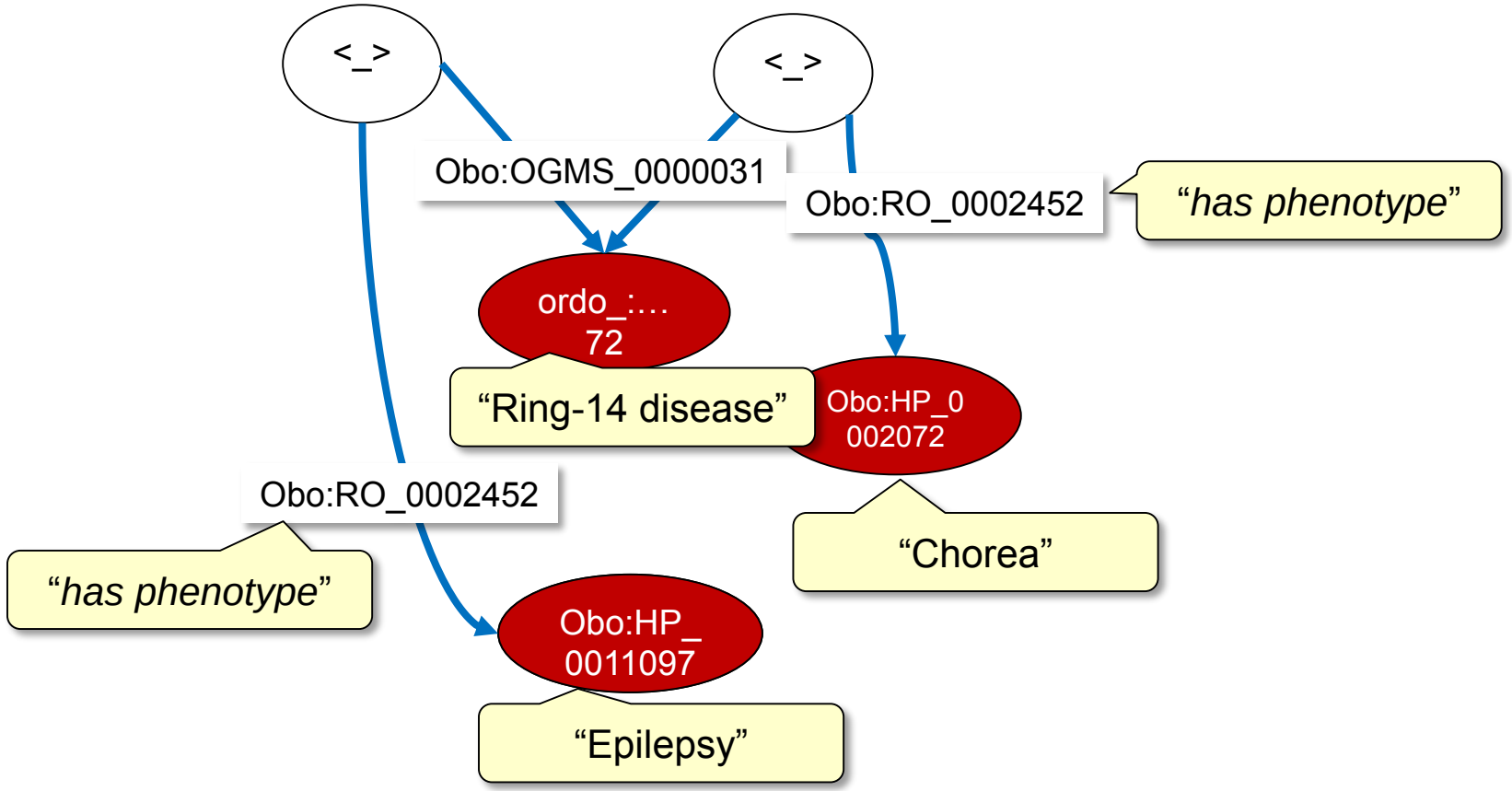


RD Connect elixir BBMRI Biobanking and Biomolecular Resources Research Infrastructure

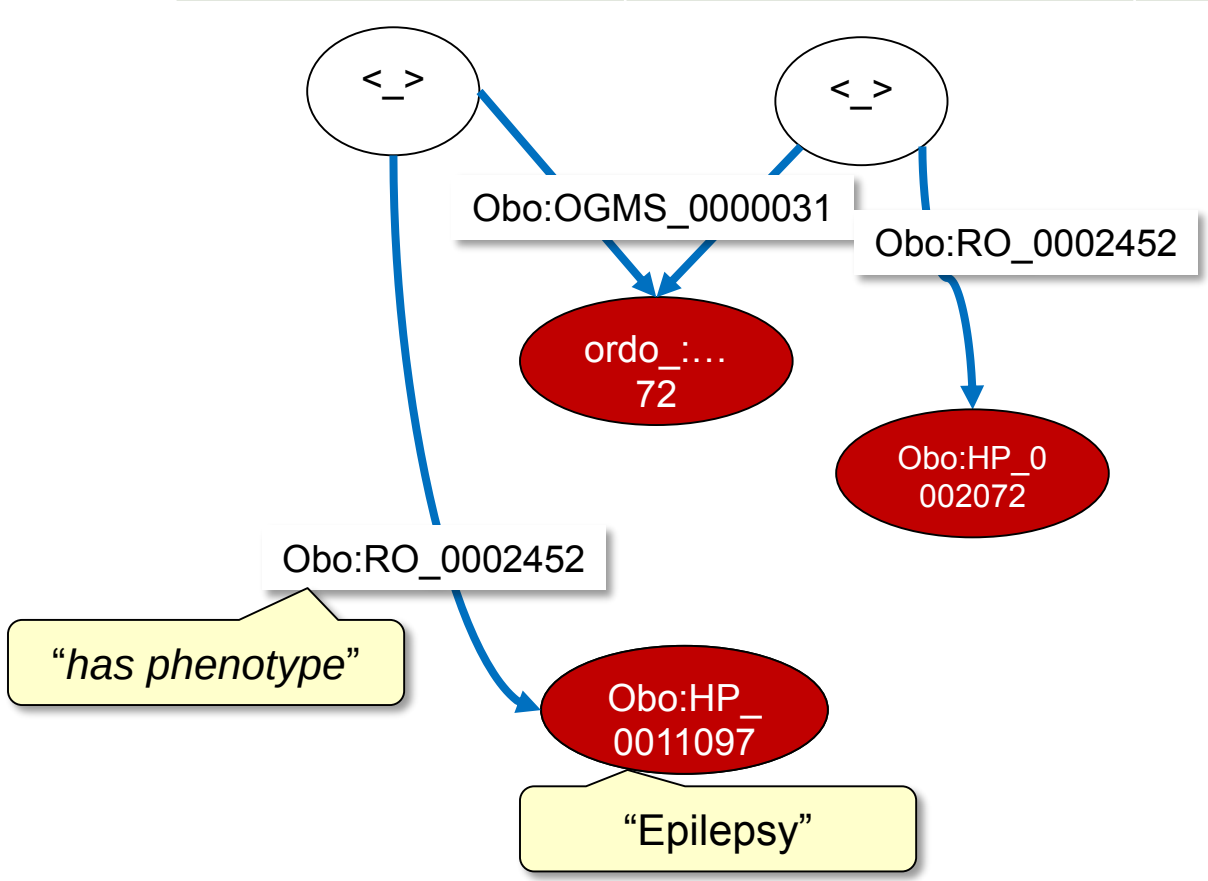
Monika	Annika	Rajaram	Pietro
<_> obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0011097.	<_> obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0002072.	<_> obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0011097 obo:RO_0002302 obo:CHEBI_33237	<_> obo:OGMS_0000031 ordo:Orphanet_178509,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

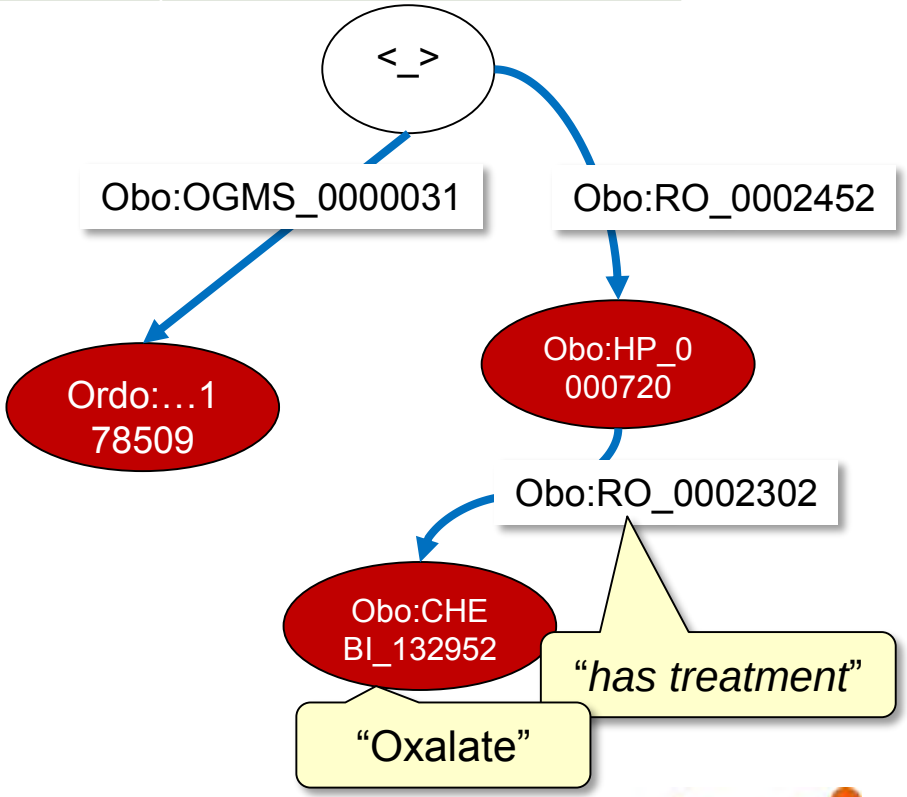


Monika	Annika	Rajaram	Pietro
<_> obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0011097.	<_> obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0002072.	<_> obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0011097 obo:RO_0002302 obo:CHEBI_33237	<_> obo:OGMS_0000031 ordo:Orphanet_178509,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

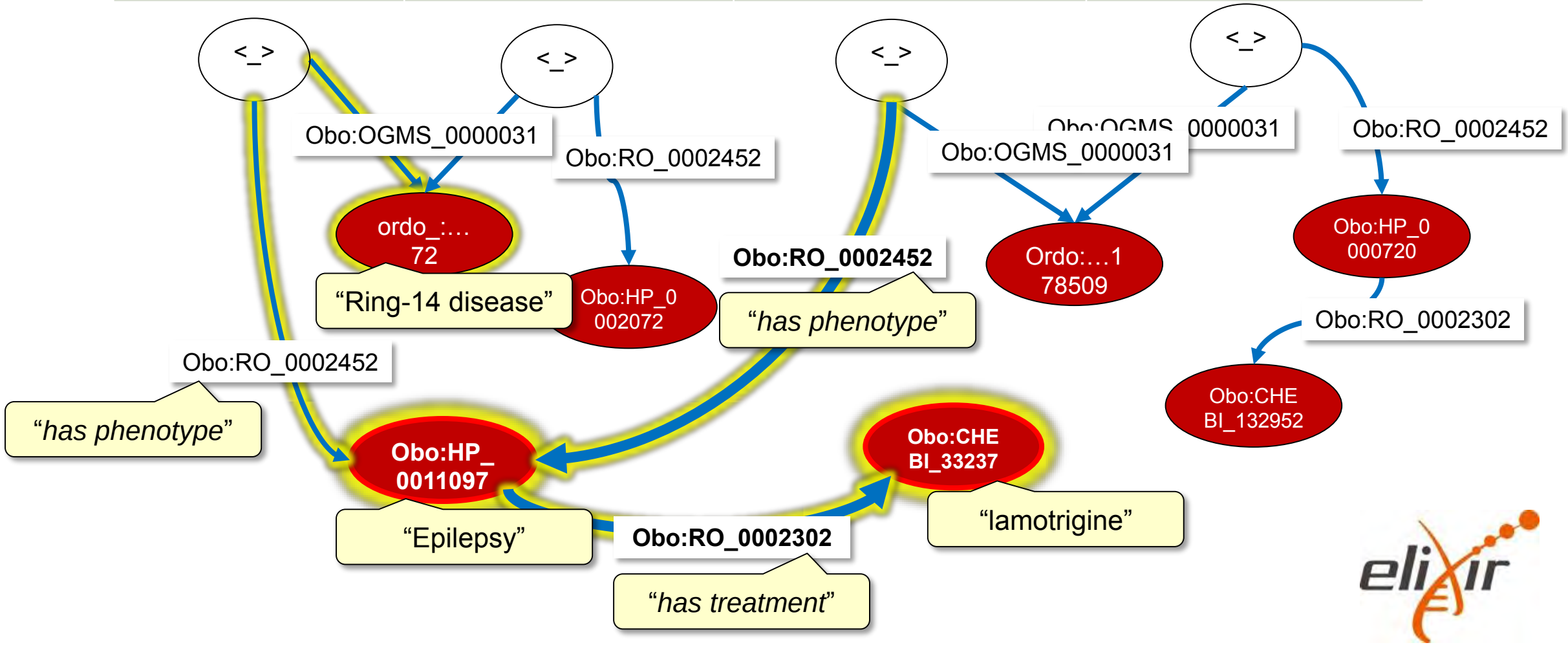


Monika	Annika	Rajaram	Pietro
<_> obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0011097.	<_> obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0002072.	<_> obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0011097 obo:RO_0002302 obo:CHEBI_33237	<_> obo:OGMS_0000031 ordo:Orphanet_178509,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

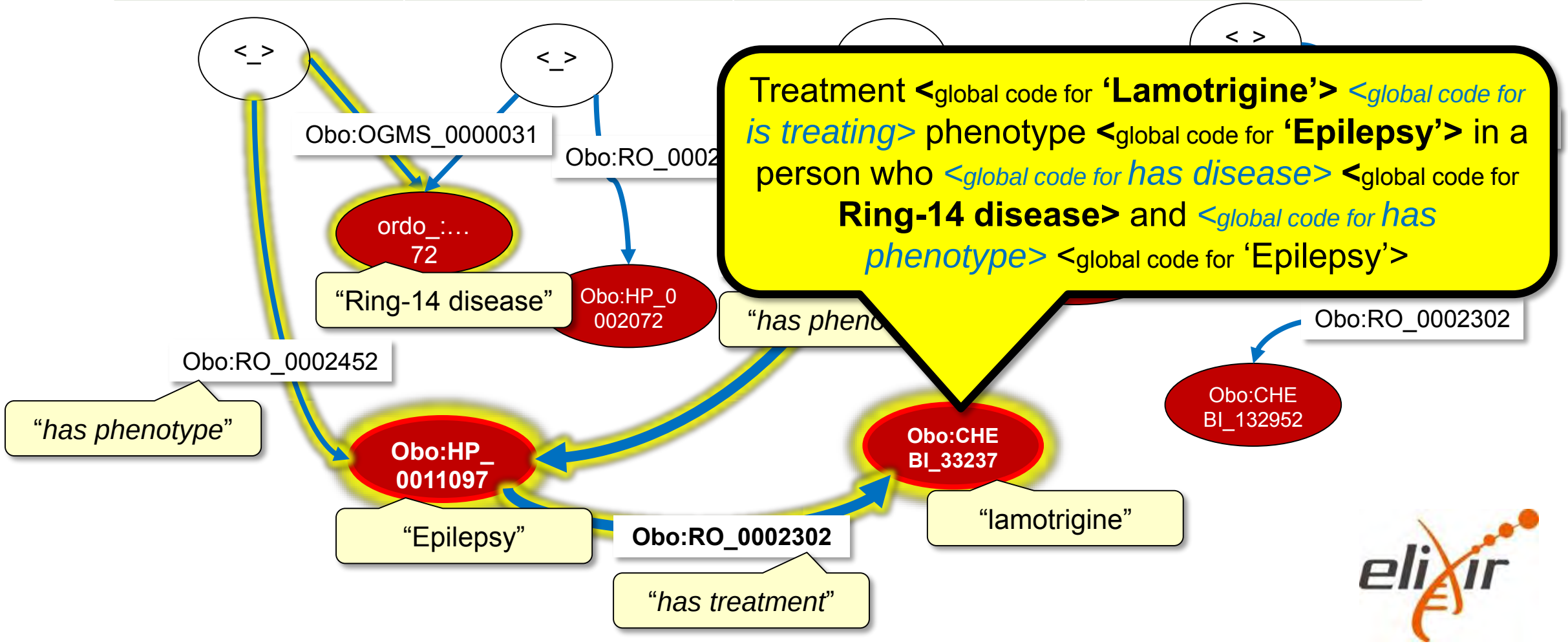




Monika	Annika	Rajaram	Pietro
<_> obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0011097.	<_> obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0002072.	<_> obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0011097 obo:RO_0002302 obo:CHEBI_33237	<_> obo:OGMS_0000031 ordo:Orphanet_178509,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

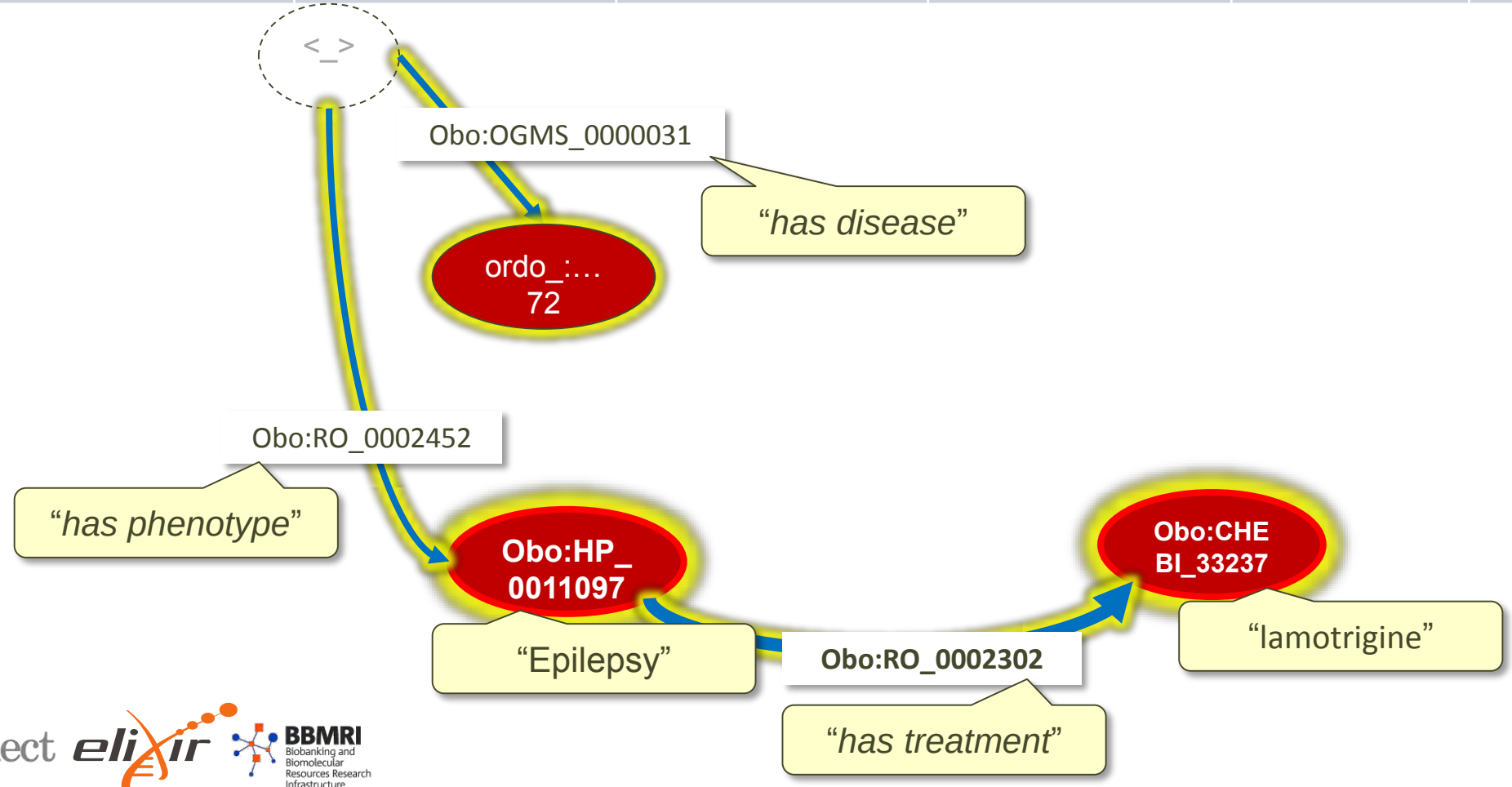


Monika	Annika	Rajaram	Pietro
<_> obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0011097.	<_> obo:OGMS_0000031 ordo:Orphanet_1440, obo:RO_0002452 obo:HP_0002072.	<_> obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0011097 obo:RO_0002302 obo:CHEBI_33237	<_> obo:OGMS_0000031 ordo:Orphanet_178509,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



Result

Disease | global machine readable code (URI)		**Treatment** | URI		**Phenotype** | URI	
Ring-14 disease	ordo:Orphanet_1440	Lamotrigine	Obo:CHEBI_33237	Epilepsy	Obo:HP_0011097



Results: codes are references to more information!

Disease | global machine readable code (URI)		**Treatment** | URI		**Phenotype** | URI	
Ring-14 disease	ordo:Orphanet_1440	Lamotrigine	Obo:CHEBI_33237	Epilepsy	Obo:HP_0011097

1. http://purl.obolibrary.org/obo/CHEBI_6367 **(CHEBI)**:
   - lamotrigine *in Ontobee*: CHEBI, DRON
   - 6-(2,3-dichlorophenyl)-1,2,4-tri
   - 0 *in Ontobee*: CHEBI
   - InChI=1S/C9H7Cl2N5/c10-5-3
   - 255.008 *in Ontobee*: CHEBI
   - 256.09100 *in Ontobee*: CHEBI
   - 3,5-diamino-6-(2,3-dichlorophenyl)-1,2,4-triazine *in Ontobee*: CHEBI, DRON
   - C9H7Cl2N5 *in Ontobee*: CHEBI, DRON
   - Lamictal *in Ontobee*: CHEBI, DRON
   - Nc1nnc(c(N)n1)-c1cccc(Cl)c1Cl *in Ontobee*: CHEBI, DRON
   - PYZRQGJRPPTADH-UHFFFAOYSA-N *in Ontobee*: CHEBI
   - lamotrigina *in Ontobee*: CHEBI, DRON
   - lamotriginum *in Ontobee*: CHEBI, DRON
   - InChIKey=PYZRQGJRPPTADH-UHFFFAOYSA-N *in Ontobee*: DRON

Class:lamotrigine Definition: A member of the class of 1,2,4-triazines in which the triazene skeleton is substituted by amino groups at positions 3 and 5, and by a 2,3-dichlorophenyl group at position 6.

1. http://purl.obolibrary.org/obo/HP_0011097 **(HP)**:
   - Epileptic spasms *in Ontobee*:
   - Salaam convulsions *in Ontobe*
   - Salaam seizures *in Ontobee*:
   - West syndrome *in Ontobee*:

Class:Epileptic spasms Definition: A sudden flexion, extension or mixed extension-flexion of predominantly proximal and truncal muscles which is usually more sustained than a myoclonic movement but not as sustained as a tonic seizure.

Scenario

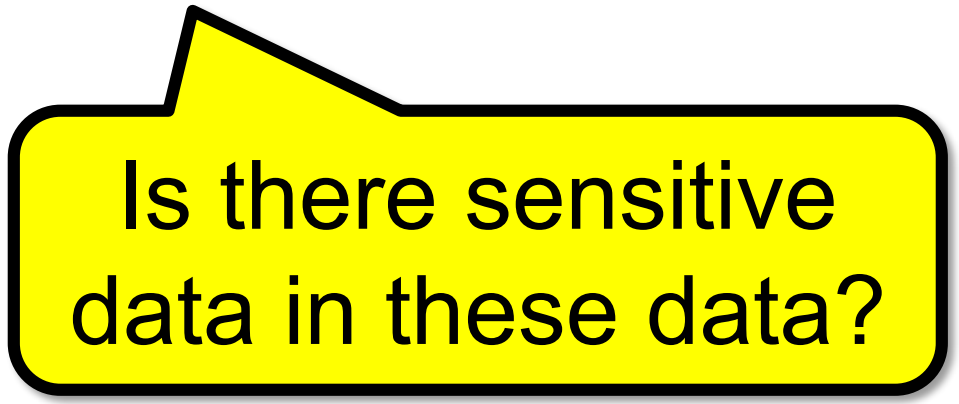
obo: http://purl.obolibrary.org/obo/
ordo: http://www.orpha.net/ORDO/

Monika	Annika	Rajaram	Pietro	FAIR
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 the drug lamotrigine	Pietro has disease Ring-14 disease, and has phenotype Extreme mood swings. Extreme mood swings are treated by the drug Oxalate	Full meaning
<_> 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 :HP_0011097	<_> obo:OGMS_0000031 ordo:Orphanet_178509, obo:RO_0002452 obo:HP_0000720 O_0002302 HEBI_132952	Interoperable & Machine readable

Is there sensitive data in these data?

Results: codes are references to more information!

Disease | global machine readable code (URI)		**Treatment** | URI		**Phenotype** | URI	
Ring-14 disease	ordo:Orphanet_1440	Lamotrigine	Obo:CHEBI_33237	Epilepsy	Obo:HP_0011097



Results: codes are references to more information!

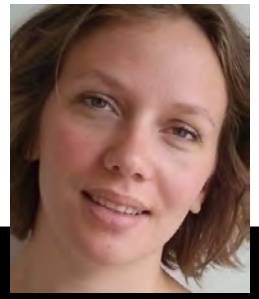
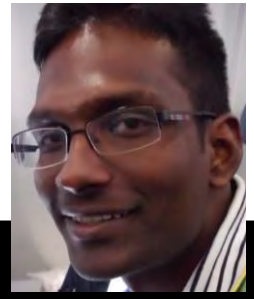
Disease | global machine readable code (URI)		**Treatment** | URI		**Phenotype** | URI	
Ring-14 disease	ordo:Orphanet_1440	Lamotrigine	Obo:CHEBI_33237	Epilepsy	Obo:HP_0011097

The ‘personal health train’ exploits this: algorithms travel to secure (FAIR) local data; no sensitive data is exposed outside of the FAIR data points

Counting individuals by **privacy protecting record linkage (PPRL)**

ELIXIR All Hands 2017, 21-23
March, Rome, Italy

Counting by PPRL

Monika 	Annika 	Rajaram 	Pietro 	Jacobsen 
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	Ziekte gekarakteriseerd door ringvorming in chromosoom 14, aangemeld in verband met epileptische aanval

How many cases of **Ring-14 disease**?
How many patients can I stratify for a trial for **epileptic attacks *and* chorea**?

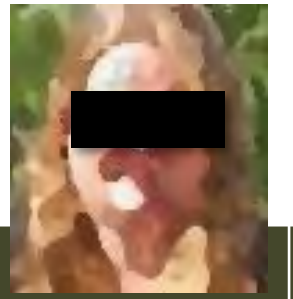
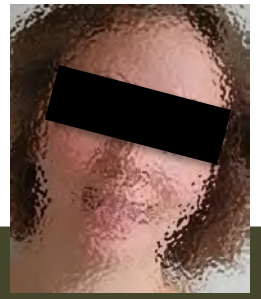
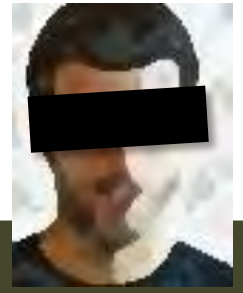
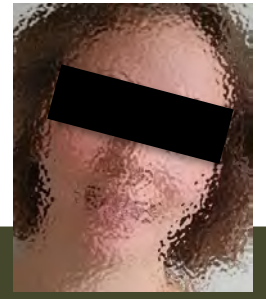
Reminder: FAIR data landscape

Independently FAIR data sources, under control of the local data manager

(e.g. HCP, patient organisation, patient)

RD Connect elixir BBMRI Biobanking and Biomolecular Resources Research Infrastructure

Counting by PPRL

--- 	--- 	--- 	--- 	--- 
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	Ziekte gekarakteriseerd door ringvorming in chromosoom 14, aangemeld in verband met epileptische aanval

How many cases of **Ring-14 disease**?
How many patients can I stratify for a trial for **epileptic attacks *and* chorea**?

RD Connect elixir BBMRI Biobanking and Biomolecular Resources Research Infrastructure

Counting by PPRL

| --- | --- | --- | --- | --- |
| **<5f9f32eb>**
obo:OGMS_0000031
ordo:Orphanet_1440,
obo:RO_0002452,
obo:HP_0011097. | **<5982a734>**
obo:OGMS_0000031
ordo:Orphanet_1440,
obo:RO_0002452
obo:HP_0002072. | **<46ce5dec>**
obo:OGMS_0000031
ordo:Orphanet_178509,
obo:RO_0002452
obo:HP_0011097
obo:RO_0002302
obo:CHEBI_33237 | **<98bbd5d5>**
obo:OGMS_0000031
ordo:Orphanet_178509,
obo:RO_0002452
obo:HP_0000720
obo:RO_0002302
obo:CHEBI_132952 | **<7d3d3c76>**
obo:OGMS_0000031
ordo:Orphanet_72,
obo:RO_0002452
obo:HP_0011097. |

How many cases of **Ring-14 disease**?
How many patients can I stratify for a trial for **epileptic attacks *and* chorea**?

RD Connect elixir BBMRI Biobanking and Biomolecular Resources Research Infrastructure

Counting by PPRL

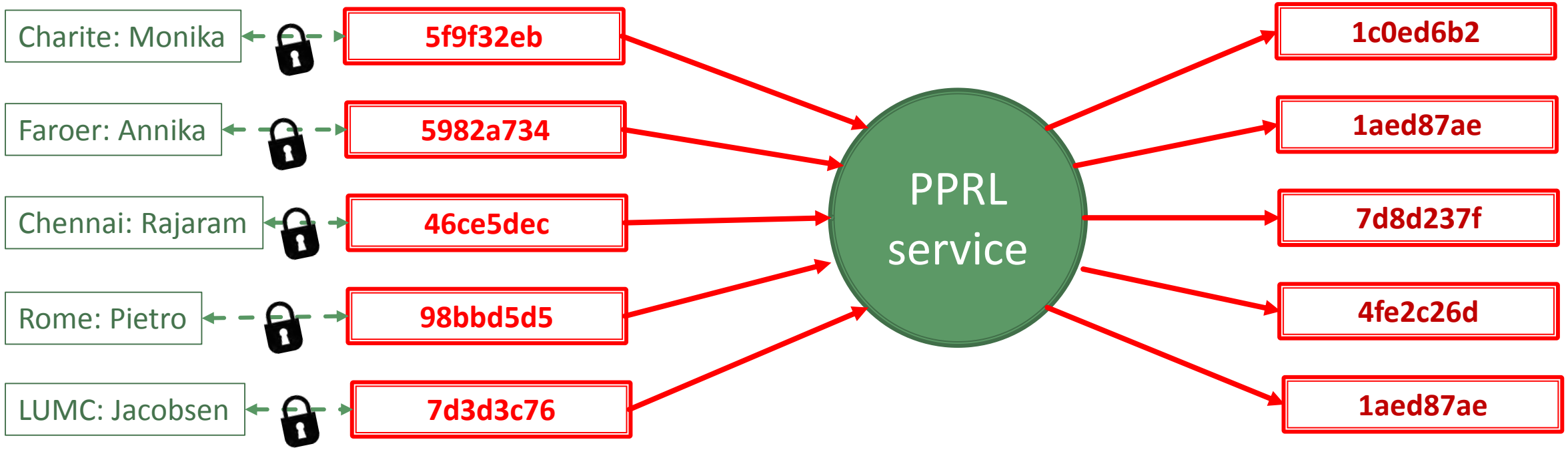
| --- | --- | --- | --- | --- |
| <5f9f32eb>
obo:OGMS_0000031
ordo:Orphanet_1440,
obo:RO_0002452,
obo:HP_0011097. | <5982a734>
obo:OGMS_0000031
ordo:Orphanet_1440,
obo:RO_0002452
obo:HP_0002072. | <46ce5dec>
obo:OGMS_0000031
ordo:Orphanet_178509,
obo:RO_0002452
obo:HP_0011097
obo:RO_0002302
obo:CHEBI_33237 | <98bbd5d5>
obo:OGMS_0000031
ordo:Orphanet_178509,
obo:RO_0002452
obo:HP_0000720
obo:RO_0002302
obo:CHEBI_132952 | <7d3d3c76>
obo:OGMS_0000031
ordo:Orphanet_72,
obo:RO_0002452
obo:HP_0011097. |

Ring-14 count = 3

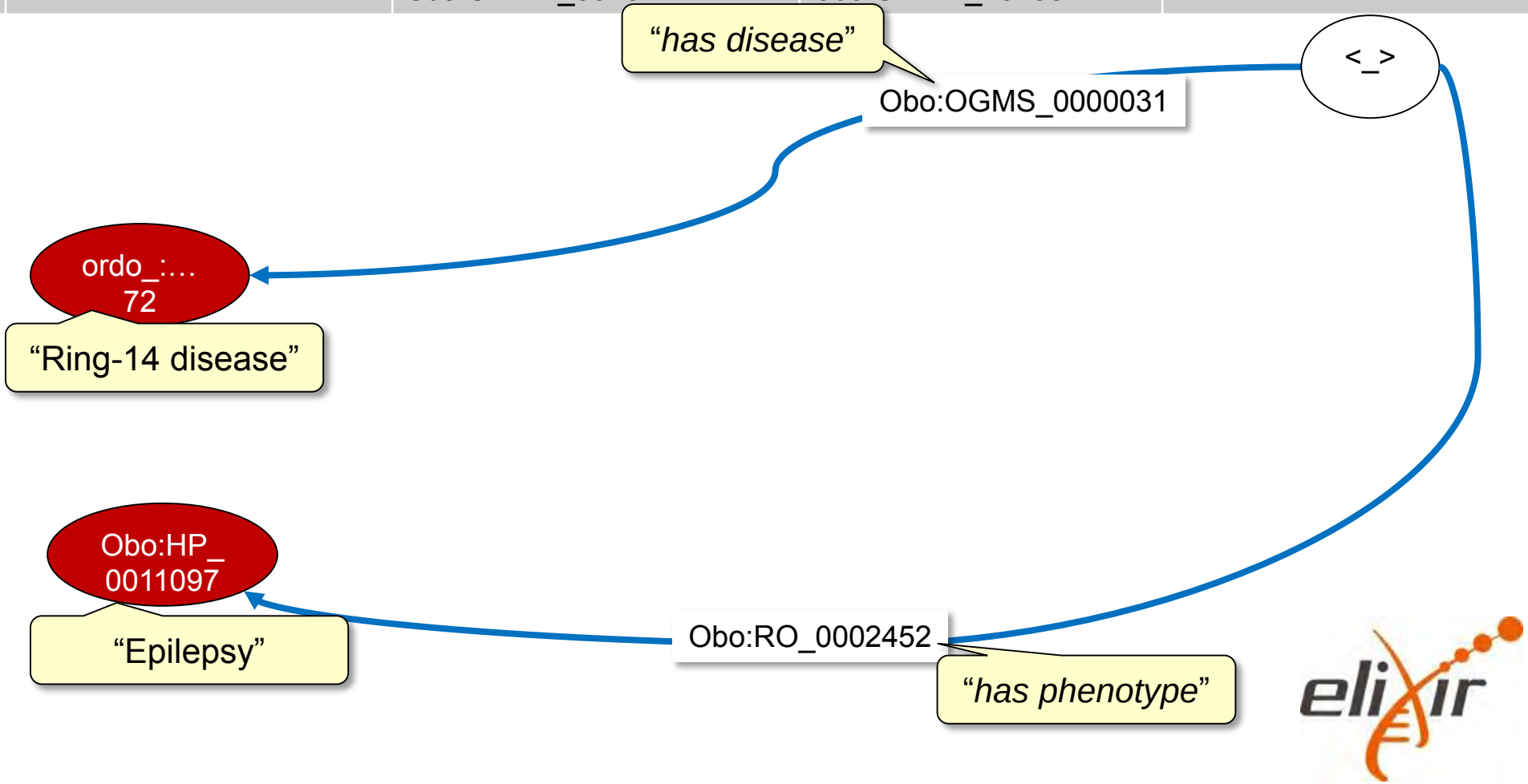
No individuals match stratification criteria

How many cases of **Ring-14 disease**?
How many patients can I stratify for a trial for **epileptic attacks** ***and*** **chorea**?

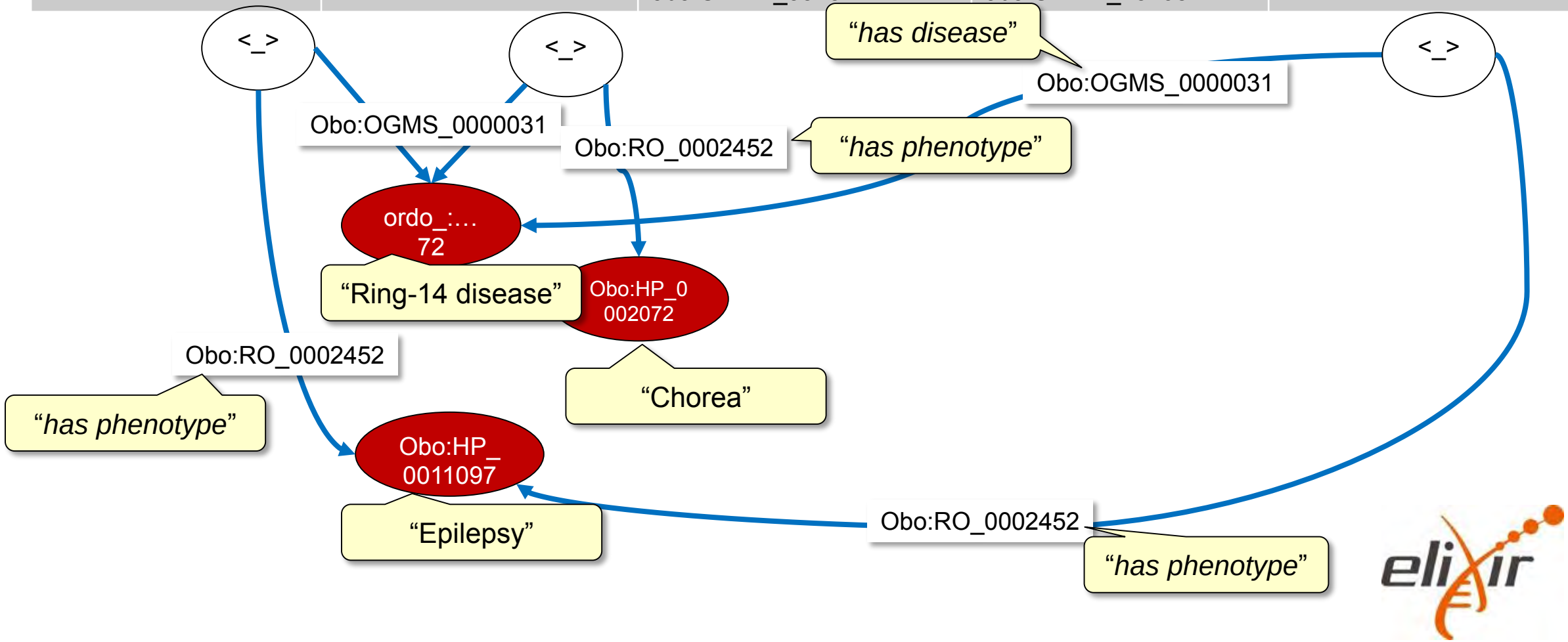
Privacy protecting identification



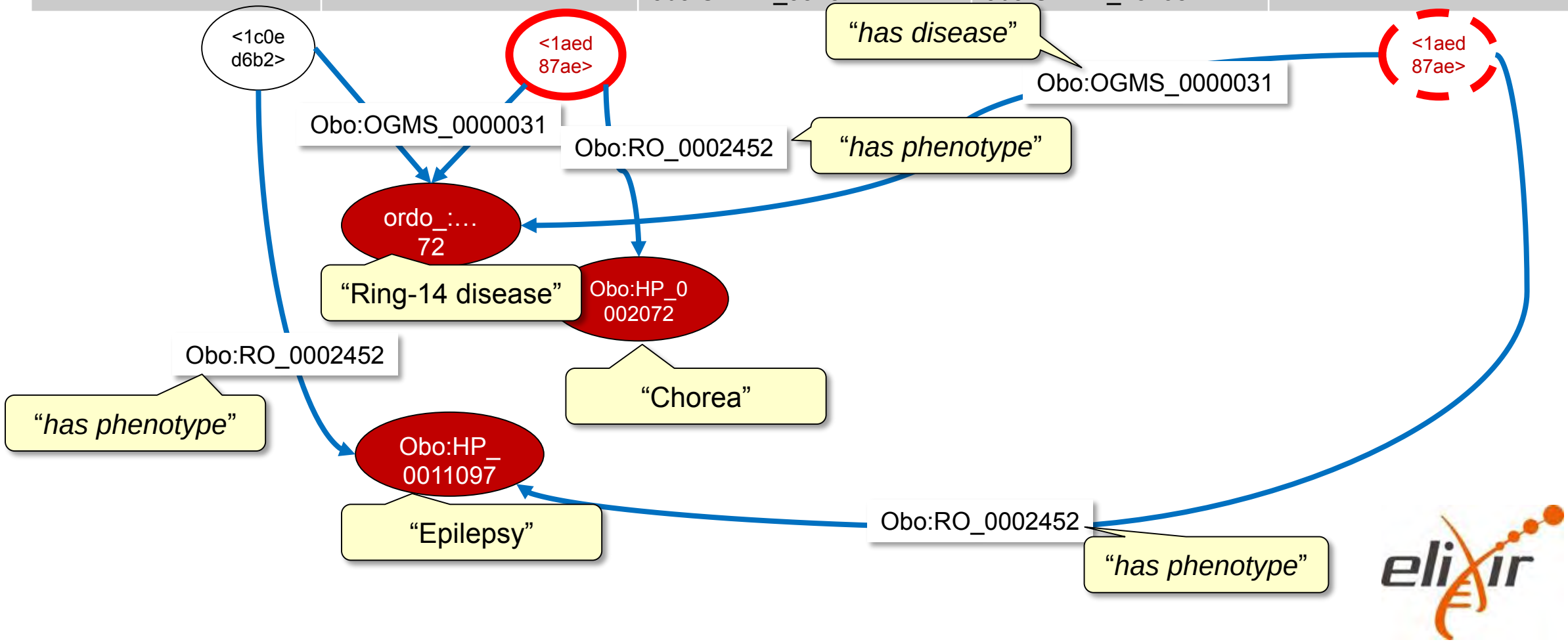
| --- | --- | --- | --- | --- |
| <_>
obo:OGMS_0000031
ordo:Orphanet_1440
obo:RO_0002452
obo:HP_0011097. | <_>
obo:OGMS_0000031
ordo:Orphanet_1440,
obo:RO_0002452
obo:HP_0002072. | <_>
obo:OGMS_0000031
ordo:Orphanet_178509,
obo:RO_0002452
obo:HP_0011097
obo:RO_0002302
obo:CHEBI_33237 | <_>
obo:OGMS_0000031
ordo:Orphanet_178509,
obo:RO_0002452
obo:HP_0000720
obo:RO_0002302
obo:CHEBI_132952 | <_>
obo:OGMS_0000031
ordo:Orphanet_1440,
obo:RO_0002452
obo:HP_0011097. |



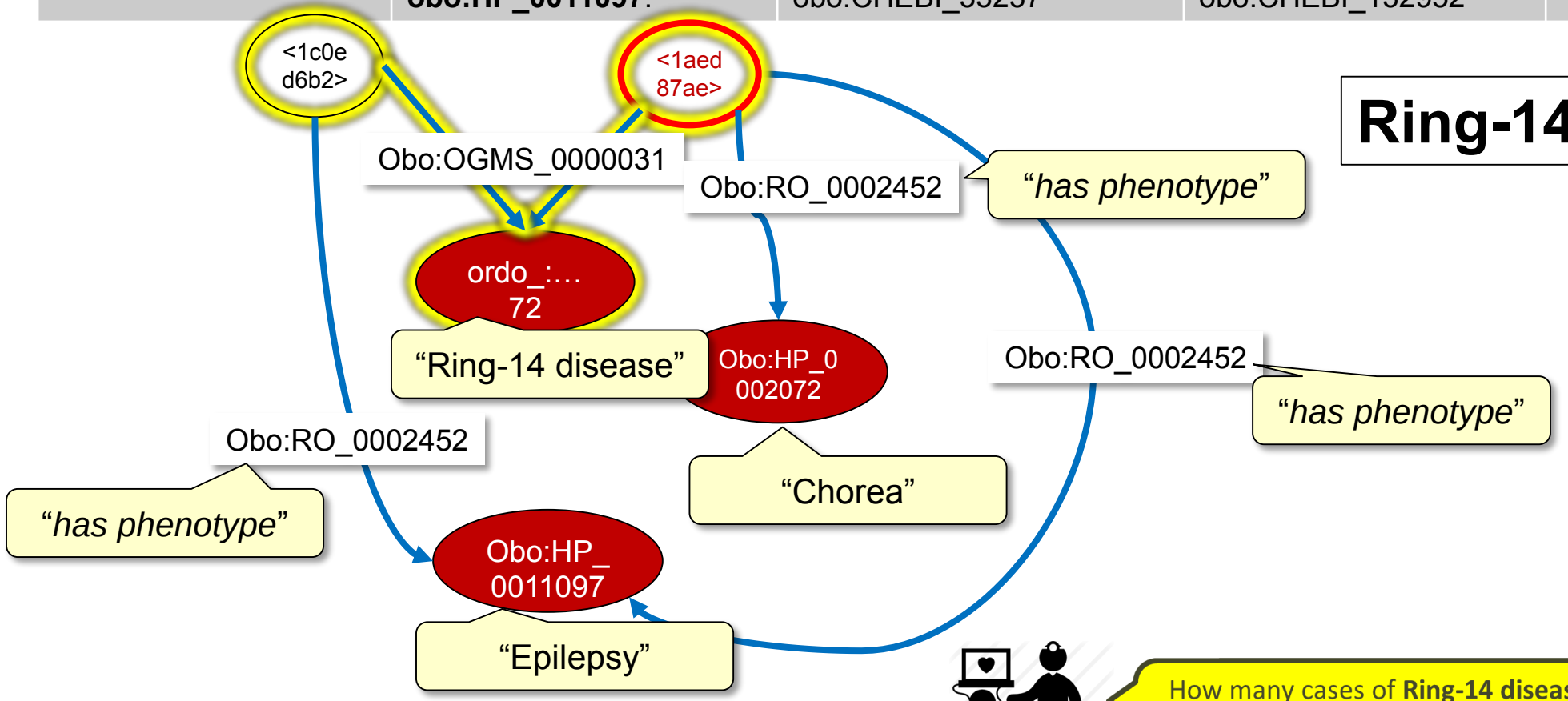
| --- | --- | --- | --- | --- |
| <_>
obo:OGMS_0000031
ordo:Orphanet_1440
obo:RO_0002452
obo:HP_0011097. | <_>
obo:OGMS_0000031
ordo:Orphanet_1440,
obo:RO_0002452
obo:HP_0002072. | <_>
obo:OGMS_0000031
ordo:Orphanet_178509,
obo:RO_0002452
obo:HP_0011097
obo:RO_0002302
obo:CHEBI_33237 | <_>
obo:OGMS_0000031
ordo:Orphanet_178509,
obo:RO_0002452
obo:HP_0000720
obo:RO_0002302
obo:CHEBI_132952 | <_>
obo:OGMS_0000031
ordo:Orphanet_1440,
obo:RO_0002452
obo:HP_0011097. |



| --- | --- | --- | --- | --- |
| **<1c0ed6b2>**
obo:OGMS_0000031
ordo:Orphanet_1440
obo:RO_0002452
obo:HP_0011097. | **<1aed87ae>**
obo:OGMS_0000031
ordo:Orphanet_1440,
obo:RO_0002452
obo:HP_0002072. | **<7d8d237f>**
obo:OGMS_0000031
ordo:Orphanet_178509,
obo:RO_0002452
obo:HP_0011097
obo:RO_0002302
obo:CHEBI_33237 | **<4fe2c26d>**
obo:OGMS_0000031
ordo:Orphanet_178509,
obo:RO_0002452
obo:HP_0000720
obo:RO_0002302
obo:CHEBI_132952 | **<1aed87ae>**
obo:OGMS_0000031
ordo:Orphanet_1440,
obo:RO_0002452
obo:HP_0011097. |



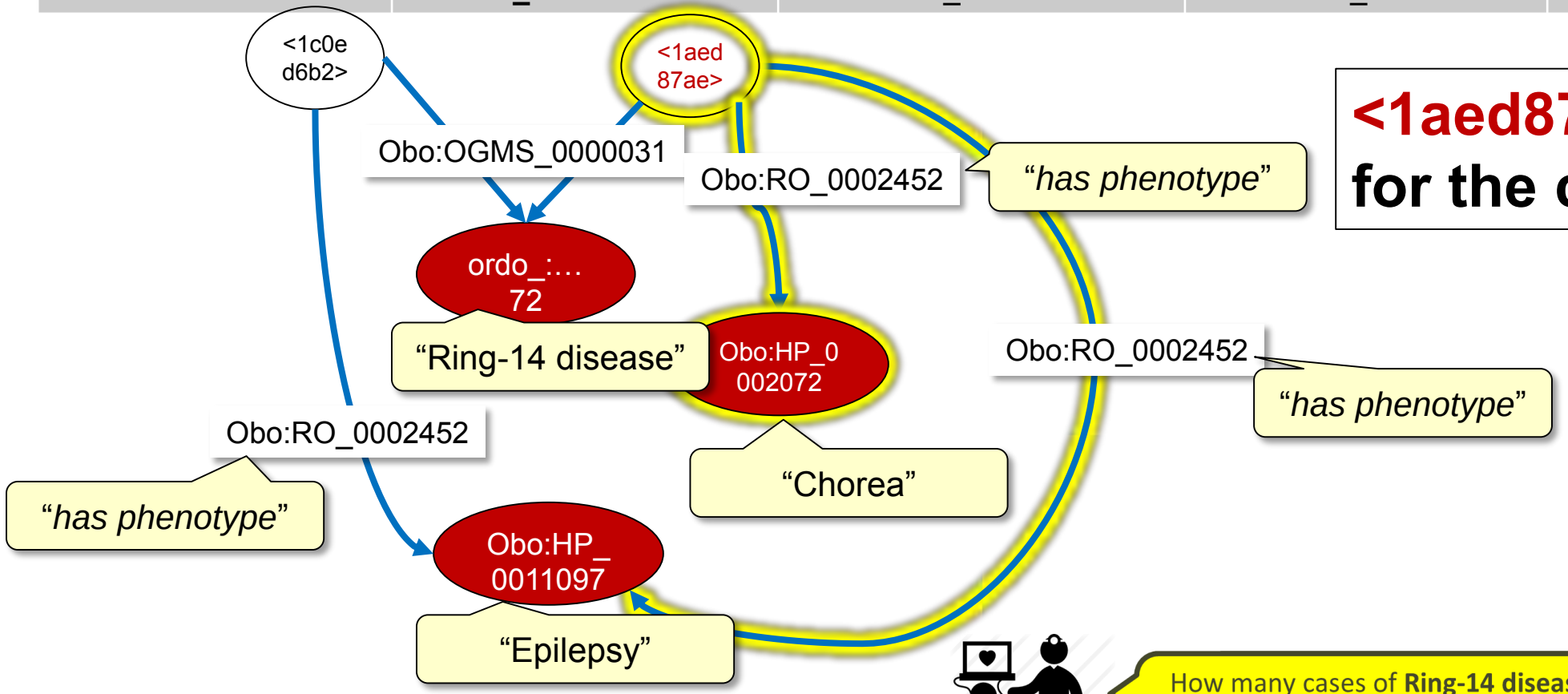
| --- | --- | --- | --- | --- |
| **<1c0ed6b2>**
obo:OGMS_0000031
ordo:Orphanet_1440
obo:RO_0002452
obo:HP_0011097. | **<1aed87ae>**
obo:OGMS_0000031
ordo:Orphanet_1440,
obo:RO_0002452
obo:HP_0002072;
obo:RO_0002452
obo:HP_0011097. | **<7d8d237f>**
obo:OGMS_0000031
ordo:Orphanet_178509,
obo:RO_0002452
obo:HP_0011097
obo:RO_0002302
obo:CHEBI_33237 | **<4fe2c26d>**
obo:OGMS_0000031
ordo:Orphanet_178509,
obo:RO_0002452
obo:HP_0000720
obo:RO_0002302
obo:CHEBI_132952 | ~~**<1aed87ae>**~~
~~obo:OGMS_0000031~~
~~ordo:Orphanet_1440,~~
~~obo:RO_0002452~~
~~obo:HP_0011097.~~ |



Ring-14 count = 2

How many cases of **Ring-14 disease**? How many patients can I stratify for a trial for **epileptic attacks *and* chorea**?

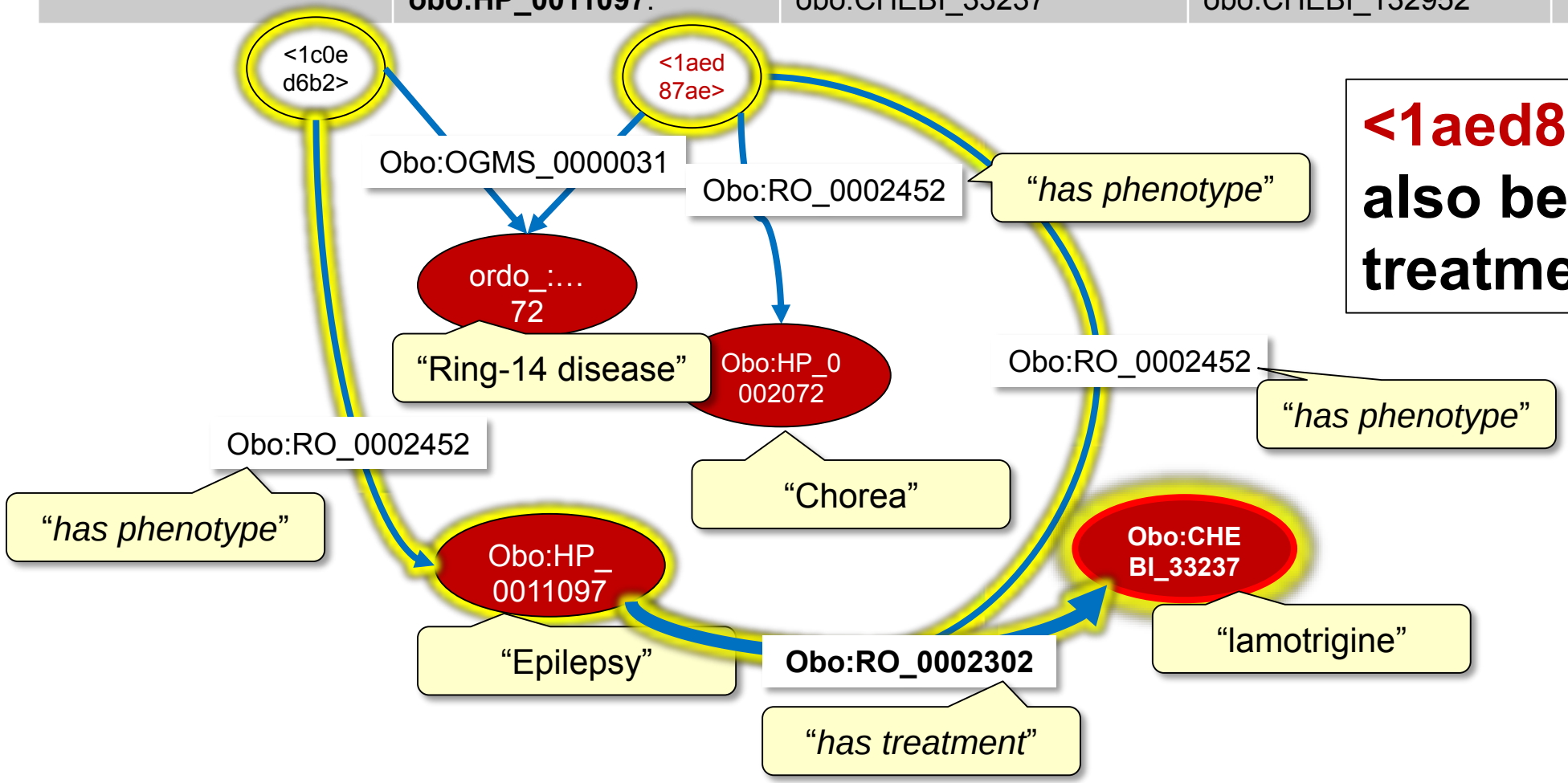
| --- | --- | --- | --- | --- |
| **<1c0ed6b2>**
obo:OGMS_0000031
ordo:Orphanet_1440
obo:RO_0002452
obo:HP_0011097. | **<1aed87ae>**
obo:OGMS_0000031
ordo:Orphanet_1440,
obo:RO_0002452
obo:HP_0002072;
obo:RO_0002452
obo:HP_0011097. | **<7d8d237f>**
obo:OGMS_0000031
ordo:Orphanet_178509,
obo:RO_0002452
obo:HP_0011097
obo:RO_0002302
obo:CHEBI_33237 | **<4fe2c26d>**
obo:OGMS_0000031
ordo:Orphanet_178509,
obo:RO_0002452
obo:HP_0000720
obo:RO_0002302
obo:CHEBI_132952 | ~~**<1aed87ae>**~~
~~obo:OGMS_0000031~~
~~ordo:Orphanet_1440,~~
~~obo:RO_0002452~~
~~obo:HP_0011097.~~ |



<1aed87ae> is fit for the clinical trial

How many cases of **Ring-14 disease**? How many patients can I stratify for a trial for **epileptic attacks** ***and*** **chorea**?

| --- | --- | --- | --- | --- |
| **<1c0ed6b2>**
obo:OGMS_0000031
ordo:Orphanet_1440
obo:RO_0002452
obo:HP_0011097. | **<1aed87ae>**
obo:OGMS_0000031
ordo:Orphanet_1440,
obo:RO_0002452
obo:HP_0002072;
obo:RO_0002452
obo:HP_0011097. | **<7d8d237f>**
obo:OGMS_0000031
ordo:Orphanet_178509,
obo:RO_0002452
obo:HP_0011097
obo:RO_0002302
obo:CHEBI_33237 | **<4fe2c26d>**
obo:OGMS_0000031
ordo:Orphanet_178509,
obo:RO_0002452
obo:HP_0000720
obo:RO_0002302
obo:CHEBI_132952 | ~~**<1aed87ae>**~~
~~obo:OGMS_0000031~~
~~ordo:Orphanet_1440,~~
~~obo:RO_0002452~~
~~obo:HP_0011097.~~ |



<1aed87ae> may also benefit from treatment

Conclusions and outlook

Conclusions

- Linked data + ontologies + PPRL + FAIR data points provide a comprehensive implementation of FAIR principles in RD domain (Findable, Accessible, Interoperable, Reuseable for humans and computers)
- Rare disease data linkage plan is a service for the rare disease community for (i) making data FAIR (ii) knowledge exchange on FAIR data stewardship

Outlook

Financing model matrix
Hooft, Roos, Thompson, *et al.*

European Reference Networks

Deploy FAIR for NMD, RETT

Consult with stakeholders for next cases

Continue BYODs for new cases

Sustainability

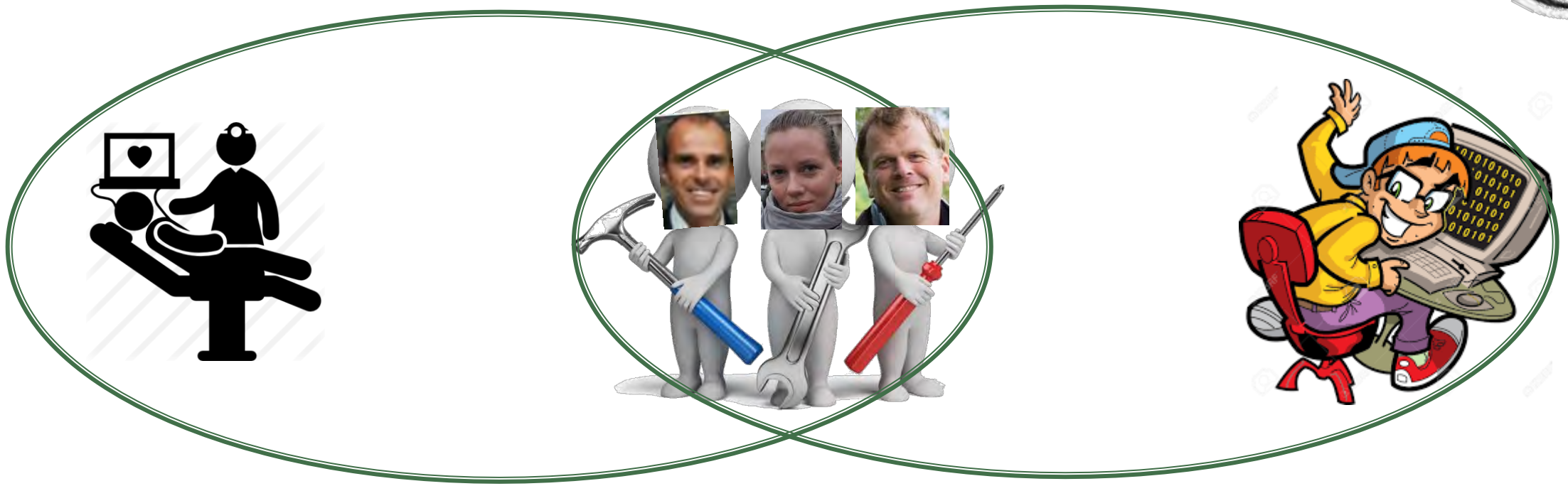
- Develop FAIR consultancy and deployment services with RD stakeholders (patient organisations, ERN HCPs, SMEs)
- Action: FAIR guiding principles as IRDiRC recognized resource

Foster FAIR expert networks (e.g. FAIR experts in Italy)

Engage with European Reference Networks

RD-Connect elixir BBMRI Biobanking and Biomolecular Resources Research Infrastructure

Preliminary model for service for ERNs
based on rare disease data linkage plan



Stakeholders
(Patient organisations, ERNs)

FAIR data stewards
FAIR data consultants

IT infrastructure developers
(Ontologies, mapping services, FAIR software)

Collaboration and financing

FAIR data teams that we link to rare disease case


Luiz Bonino


David van Enckevort

FAIR Data engineering team
Luiz Bonino
Rajaram Kaliyaperumal
Kees Burger
Nuno Nunes
Shamanou van Leeuwen
Mark Thompson

DTL FAIR DATA ENGINEER TEAM

Skunk team / FAIR metrics
Mark Wilkinson
Michel Dumontier

Rare disease software developers
David van Enckevort
Annika Jacobsen, Andra Waagmeester; Heimo Muller, Robert Reihs, Pedro Sernadella, Jose Oliveira, Marc Hanauer, Ana Rath, Roxana Merino, Matthias Brochausen, Developers of Castor, RDRF, OSSE, MOLGENIS

Liaisons/case owners: Rachel Thompson, Libby Wood, Claudio Carta, Domenica Taruscio, Marco Crimi, Estrella Gomes, Marina Mordenti, Freddie Ehrhart

Elixir: Team Evelo, Team Goble, Team Gut/Beltran, Team Parkinson, Team Poch

GoFAIR: Barend Mons; **DTL**: Mascha Jansen, Celia van Gelder, Erik Schultes, Albert Mons

Patients and patient organisations

Mark your agenda's

- Next week: BYOD omics data sources in context rare eye diseases, Strassbourg, France
- BYOD EYE-ERN, plan: October 2017
- Orphanet & ID-Cards FAIR hackathon (here?)
- Annual summer school + BYOD for rare disease data managers in context of ERNs, Rome, September 2017

Thank you…

PLEASE CHECK RD-CONNECT, ELIXIR-EUROPE AND ISS-CNMR WEB SITE FOR UPCOMING RD-CONNECT, E-RARE WORKSHOPS IN MAY, AND INTERNATIONAL SUMMER SCHOOL IN SEPT. !!!

www.rd-connect.eu/events

www.iss.it/cnmr/

www.elixir-europe.org/events

RD Connect elixir BBMRI Biobanking and Biomolecular Resources Research Infrastructure

ELIXIR and the Rare Disease community

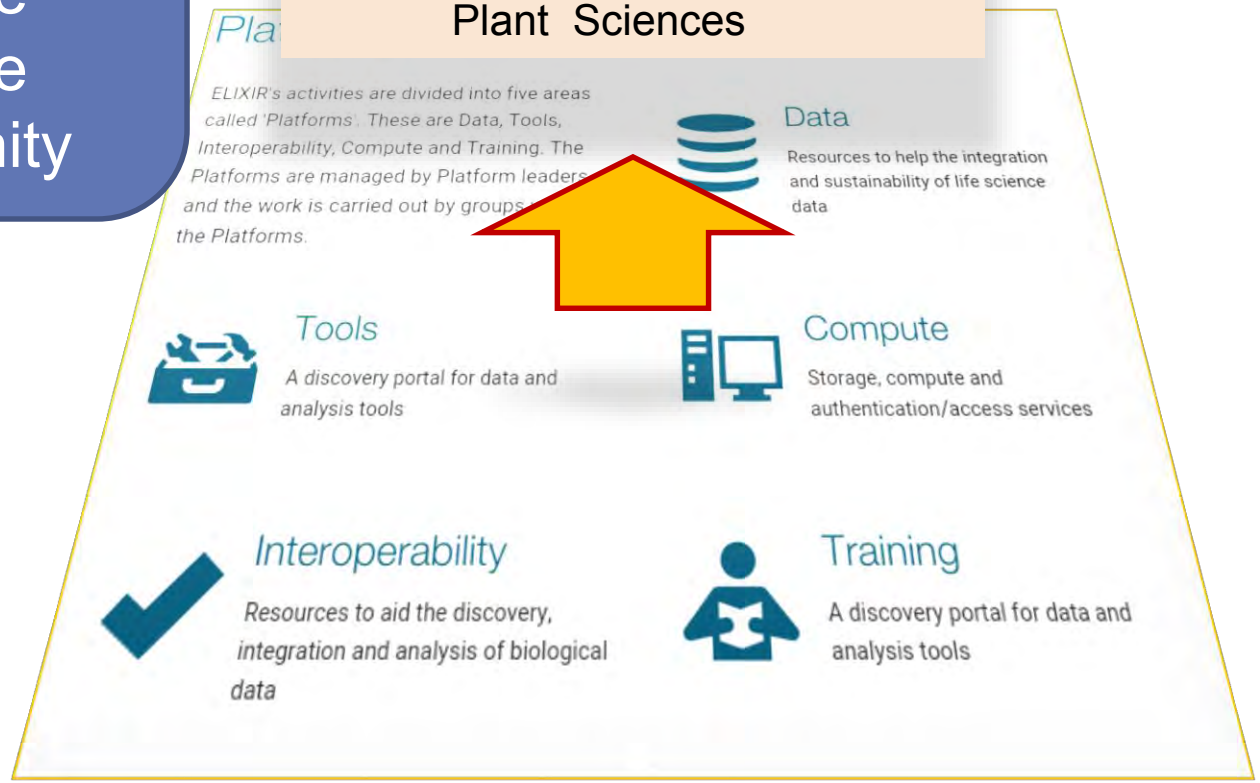
16 May 2017

ELIXIR-EXCELERATE use cases

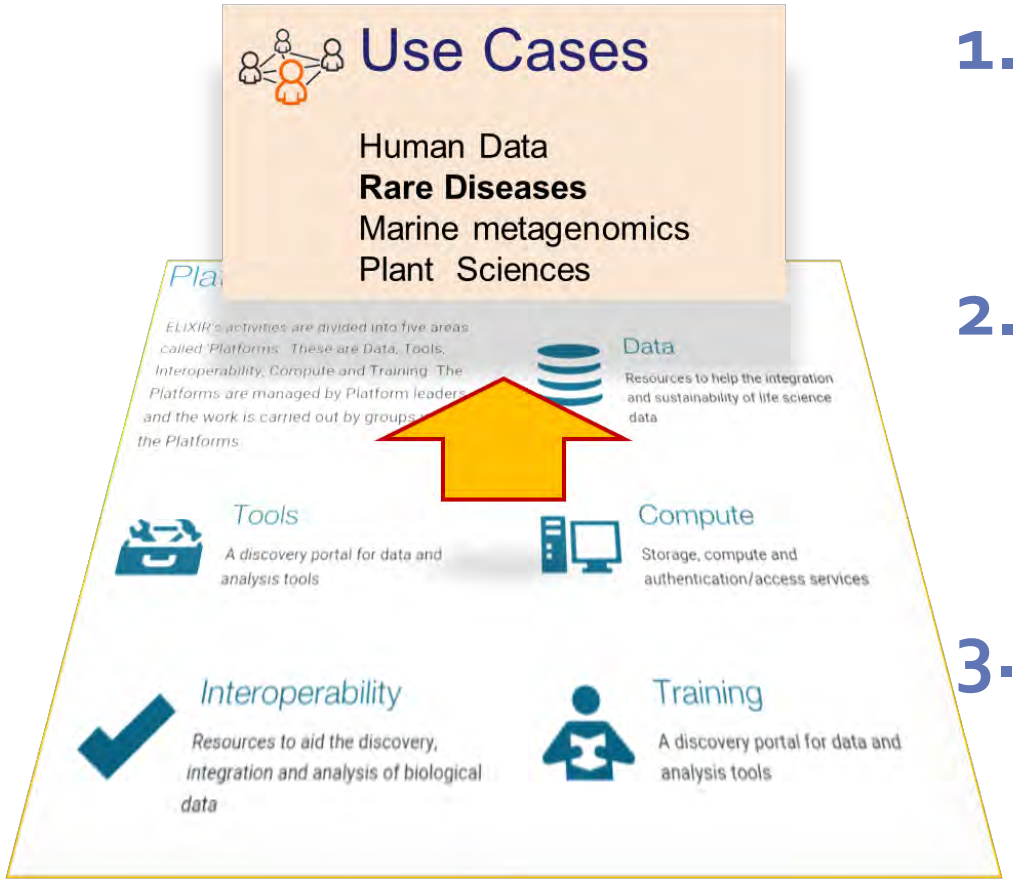
Establish infrastructure **for and with** the rare disease community

Use Cases

Human Data
Rare Diseases
Marine metagenomics
Plant Sciences



EXCELERATE Objectives



Collaborate on

1. **The ELIXIR registry of data resources and analysis tools and benchmarking**
   Tools, Compute, Data
2. **Technical framework for the comparison and standardization of services**
   Interoperability
3. **Training courses, workshops, jamborees**
   Training

T8.1 - ELIXIR portfolio of data resources developed in collaboration with the rare diseases communities

RD data resources and tools survey

Prioritization of tools

elixir Tools and Data Services Registry

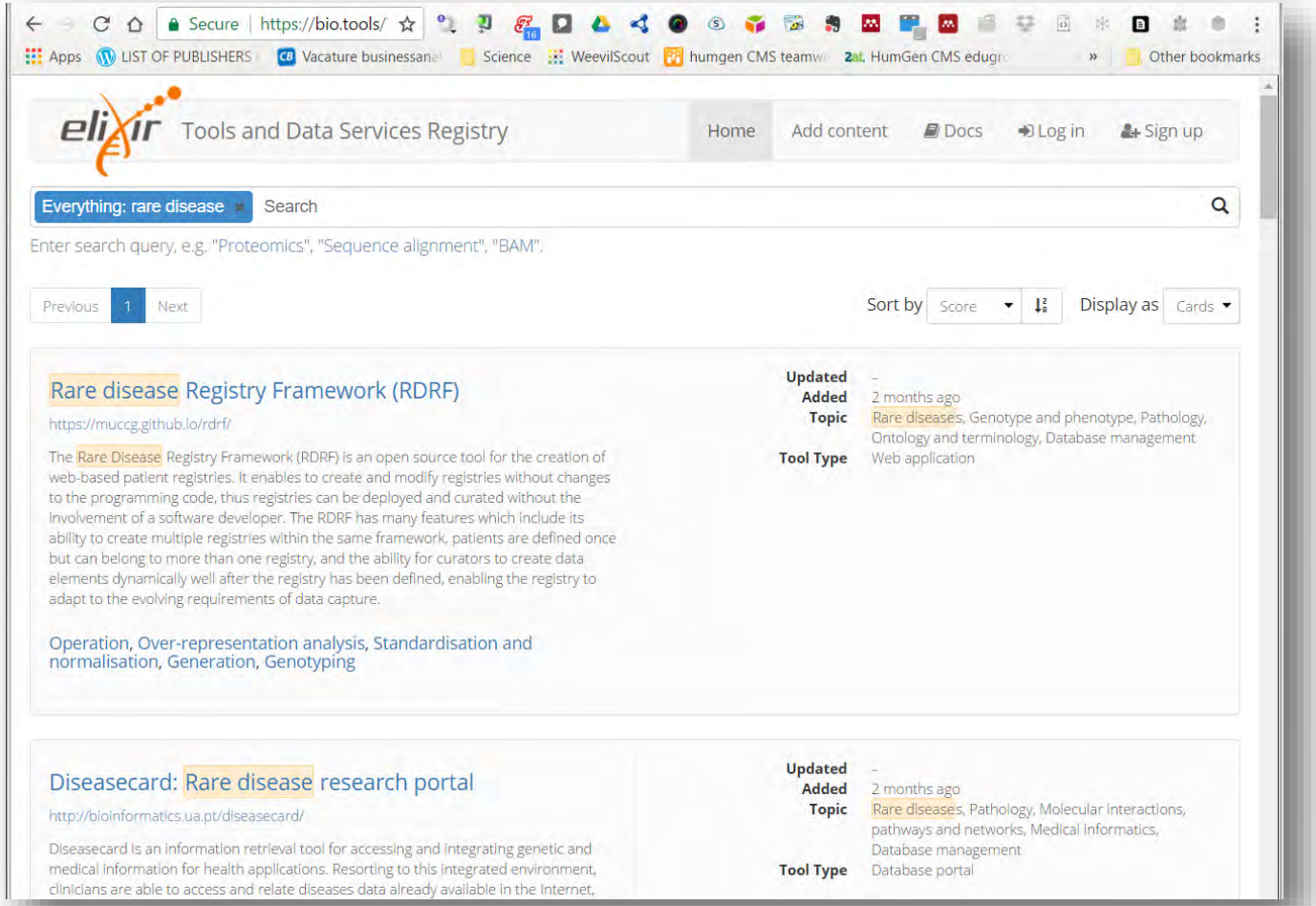
51 tools useful for RD prioritized in the first round and published in the ELIXIR Tools and Data Services Registry.

✓ Milestone 8.1: Publication of the catalogue of resources, data sources and methods assessed in collaboration with the rare-diseases communities.

→ Deliverable 8.1: Portfolio of ELIXIR data resources and tools for the rare diseases communities (M24).

excelerate

ELIXIR portfolio of data resources developed in collaboration with the rare diseases communities



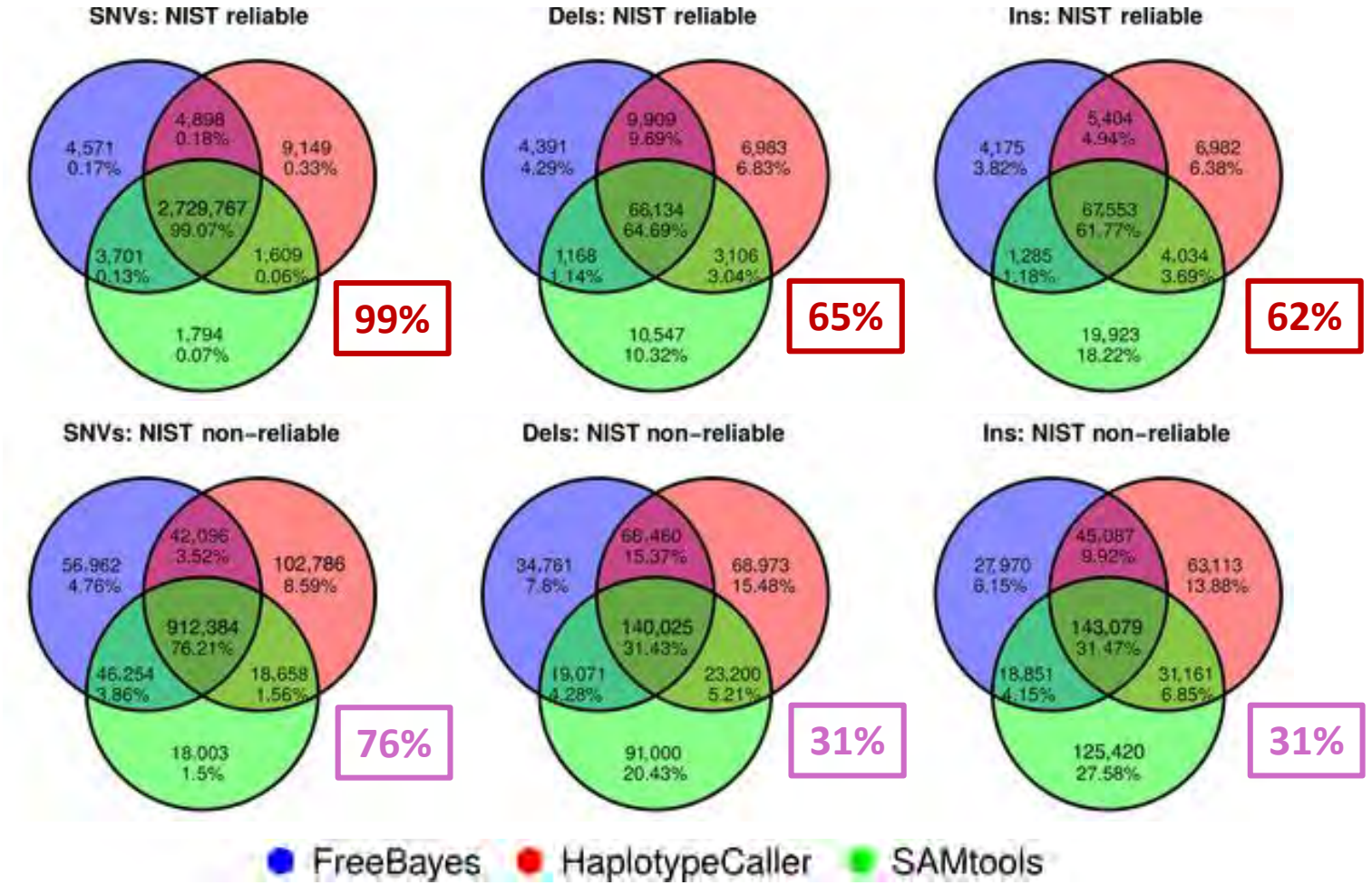
Next steps

- Continuous revision following RD community needs and recommendations (T8.1.1)
- Apply tagging to prioritise tools for rare disease community.

T8.1.2 Benchmarking and standardizing genomics pipeline

Reliably Callable

Not Reliably Callable



Laurie *et al*. Human Mutation, 2016

→ D8.6. Documentation on adequate quality reporting standards for genomics datasets (M48)

→ Standardized data and user-friendly analysis via RD-Connect web platform

excelerate

T8.3 Assessing RD training needs and capacity → training portfolio

Survey (WP8+WP11)

Analysis

elixir Training portfolio

- Survey based on WP11 general survey
- Advertised, and collecting data
- Direct feedback from RD liaisons
  - Reuse training needs assessment in other projects
- Analysis in spring 2017
  - Results will be used to plan further RD training events
  - Compare with WP11 general training survey results

excelerate

Tutorials & workshops targeting different RD user communities

RD-Connect+ELIXIR

- CNAG Data Analysis Workshop, ELIXIR-All Hands, Barcelona, March 2016
- ELIXIR Variant Calling Workshop, Helsinki, June 2016
- Next generation registries, going FAIR, going Gold, Rome, July 2016
- Annual summer school and BYOD, Rome, September 2016
- FAIR data tutorial, ECCB, Den Haag, September 2016
- FAIR data tutorial, SWAT4LS, Amsterdam, December 2016
- RD community workshop with RD-Connect, Edinburgh, ECRD, May 2016
- Invited lecture, Fundación Ramon Areces, Madrid, November 2016
- Invited lecture, Sample & Data Banking Course, Bologna, December 2016
- Panel EURORDIS multi-stakeholder event, Brussels, February 22, 2017
- PPRL and FAIR sessions, RD-Action workshop, Brussels, April 27, 2017
- E-Rare workshop on data sharing, Berlin, May 4, 2017

FAIR service infrastructure development & deployment in rare disease community

Global initiatives

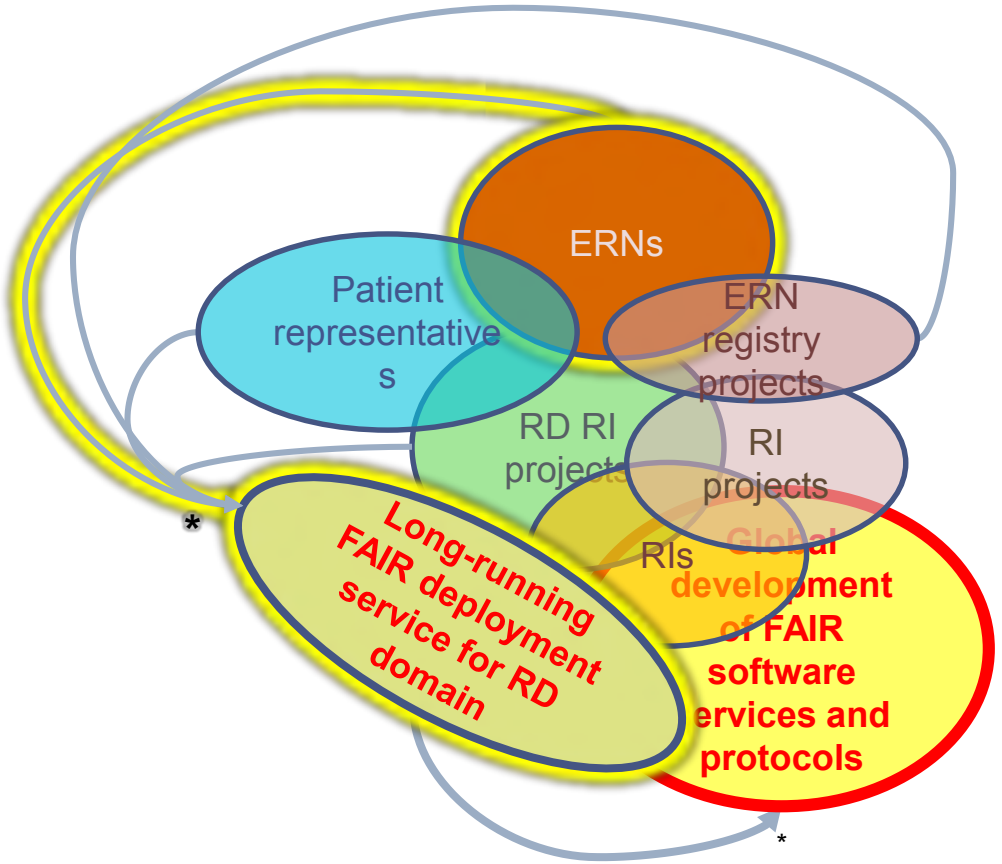
- FAIR tools and protocol developers
- FAIR deployment service

(Research) Infrastructures

- Research infrastructures: ELIXIR, BBMRI
- ERNs
- European Open Science Cloud (and GoFAIR networks)
- NIH Commons

Projects

- ERN registry projects
- RI projects: ELIXIR-EXCELERATE, ELIXIR-CORBEL, BBMRI-ADOPT
- RD RI projects: RD-Connect (→ RD-Bridges)
- Big Data to Knowledge (BD2K) USA

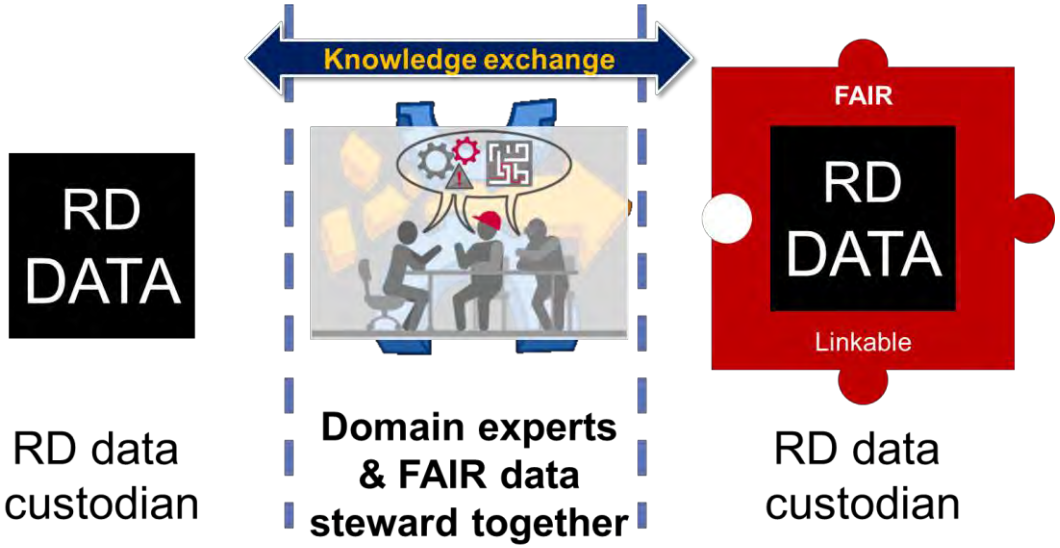


* NB stakeholders/stakeholder projects contribute to the services that they make use of

Starting point: Bring Your Own Data workshop

Preparation

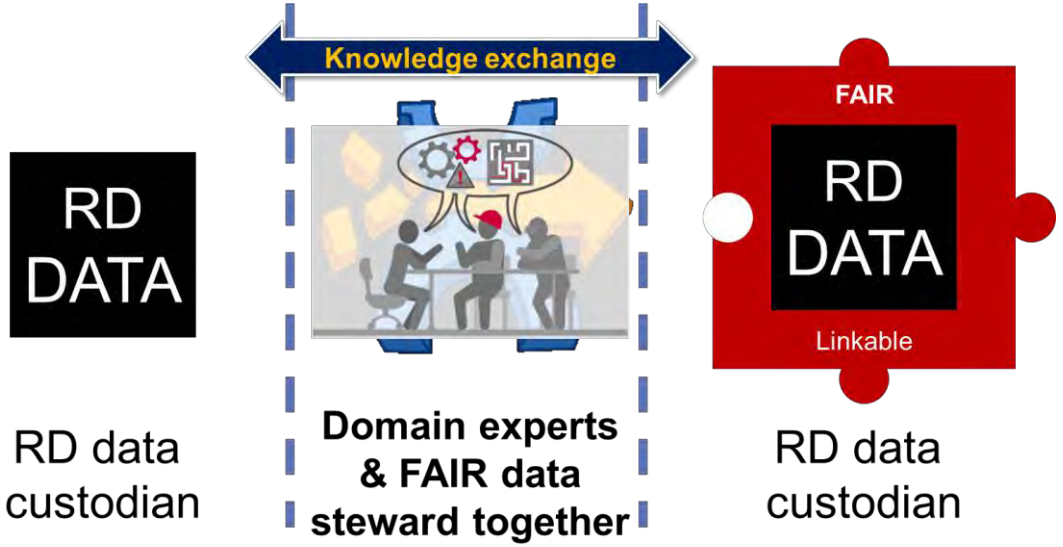
1. Invite data owners and FAIR data experts
2. Webinar(s) to explain the concept
3. Chase data owners to bring
    1. excerpt from real data
    2. driving questions



Starting point: Bring Your Own Data workshop

At BYOD

1. Introduce method
2. Refine driving questions with FAIR data experts
3. Scramble data if needed
4. OpenRefine-FAIRifier to convert data
    1. Ontology lookup services (BioPortal/OLS) to find terms from appropriate ontologies
    2. FAIRifier to apply them to the data
5. Share linkable data in linked data cache
6. Answer driving questions with FAIR data experts



FAIRify rare disease resources

Dataset level

1. Create FAIR data point via metadata editor
2. Incorporate FAIR data point service into local IT infrastructure

Data level

1. Define semantic model for data set & driving integration questions
    1. Ontology lookup services (BioPortal/OLS) to find ontology terms
       Note: terms come from multiple ontologies for a typical data set
    2. Future: apply application ontology tools
2. Add FAIR data point to FAIR data demonstrator

***OR* work with software provider to generate FAIR data**

BYOD for RD data managers

- Apply procedure for data excerpts
- OpenRefine-FAIRifirer to convert
- Cache linkable data
- Answer driving questions
- DIY FAIRification sketch
- Discuss
    - Skills required for FAIRification
    - How to acquire those skills

Outcome

ELIXIR: implementation study to test interoperability components

→ **recommendations interoperability platform & RD community**

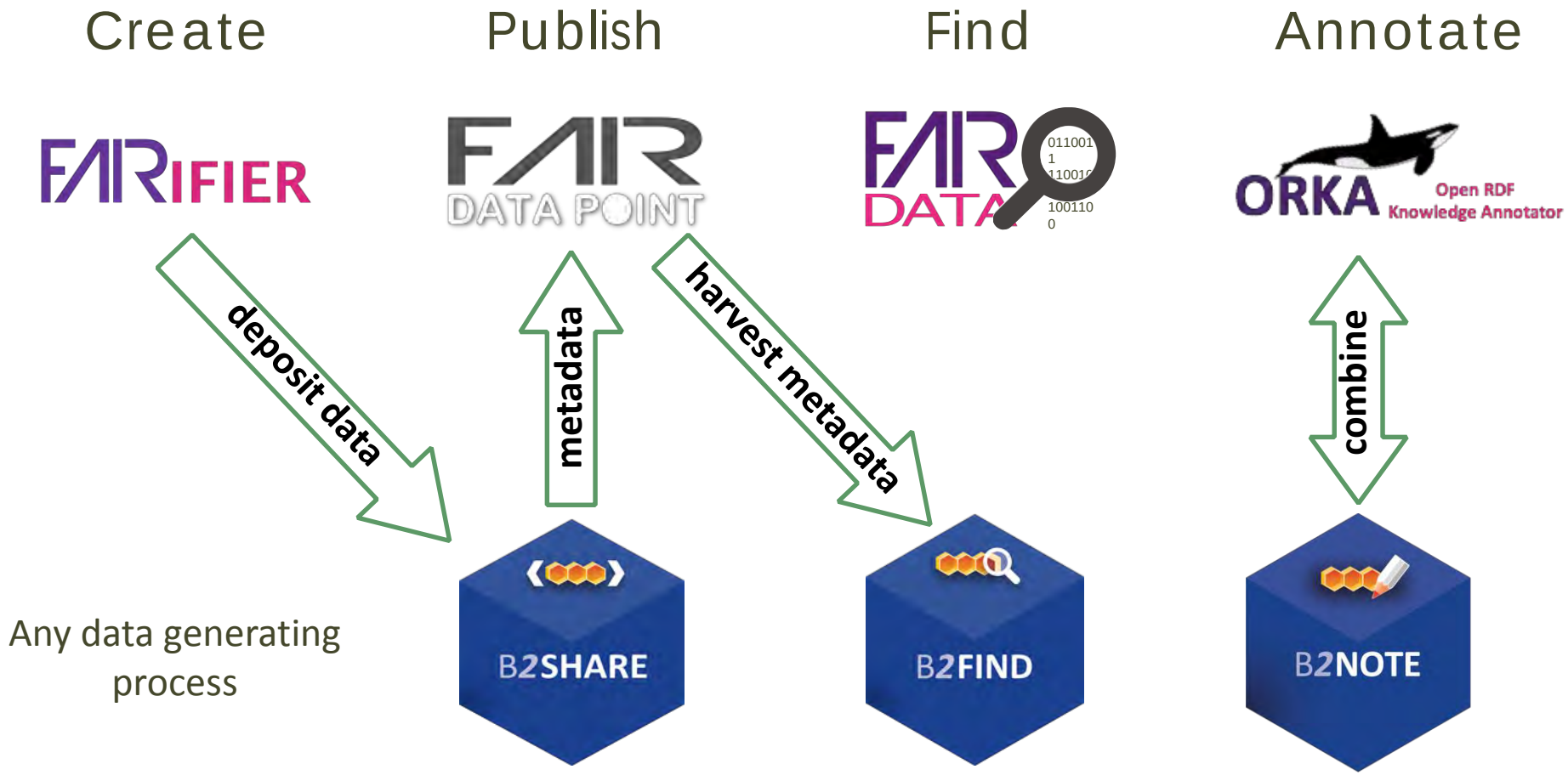
RD-Connect: proof-of-concept demonstration

→ **rare disease data linkage plan**

International FAIR collaborators

→ **adoption, co-development, feedback**

EUDAT Data pilot



Summary of recommendations from proof-of-concept / ELIXIR implementation study

Recommendations

- FAIR data API + standardized metadata: data set level
- Apply rules for persistent identifiers and mapping services for common identifiers
- Ontologies and Linked Data: data level

Recommendations

Ontologies and Linked Data: data level

- ORDO and HPO for diseases and phenotypes
- **Semantic archetypes** (=application ontologies)
  - Apply for data annotation (=prepare for CRQs)
  - built from existing ontologies by modelling experts
  - share on biosharing.org to stimulate convergence
  - derive new recommendations (e.g. OBIB for sample collections)

Recommendations

- Work with software providers
  (patient registry tools, cataloguing tools, HCP systems)
- Emphasize knowledge exchange between experts
  - Train and support registry data managers
  - BYODs and FAIR support service
  - Role of FAIR data stewards

Acknowledgements

GA4GH / IRDiRC PPRL WG

Dixie Baker, Peter Christen, Ken Gersing, Max Hauessler, Mark Phillips, Cenk Sahinalp, Adrian Thorogood, Bartha Maria Knoppers, et al.

RD-Connect

Hanns Lochmuller, Rachel Thompson, et al.

EU-PID

Günter Schreier, Michael Nitzlnader, et al.

BBMRI-ERIC

Petr Holub, Klaus Kuhn, et al.

D4LS SIG Veilige Datakoppeling

Erik Flikkenschild, Jasper Bovenberg, Gerard van Grootheest, et al.

FAIR data software services

Create

Publish

Find

Annotate

Adopt

RDRF
ID-Cards
Orphanet
MOLGENIS
Castor
OSSE
…