Update ILIAD registry

an International Library of Intellectual disability and Anomalies of Development

ITHACA online meeting 11 dec 2020

Tech team: Morris Swertz, Fernanda de Andrade, Dieuwke Prins- Joeri van der Velde, supported by MOLGENIS tech team. Legal/contracts: Gert-Jan van de Geijn. Collaborators from GenIDA, OPBG, RETT.
Steering: Alessandra Renieri, Alain Verloes, Klea Vyshka

umcg

Genomics coordination center

MOLGENIS

OUTLINE

- Project overview (reminder)
- State of play
  - Progress
  - Current work
  - Future steps
- Demo (Fernanda)

Project synopsis

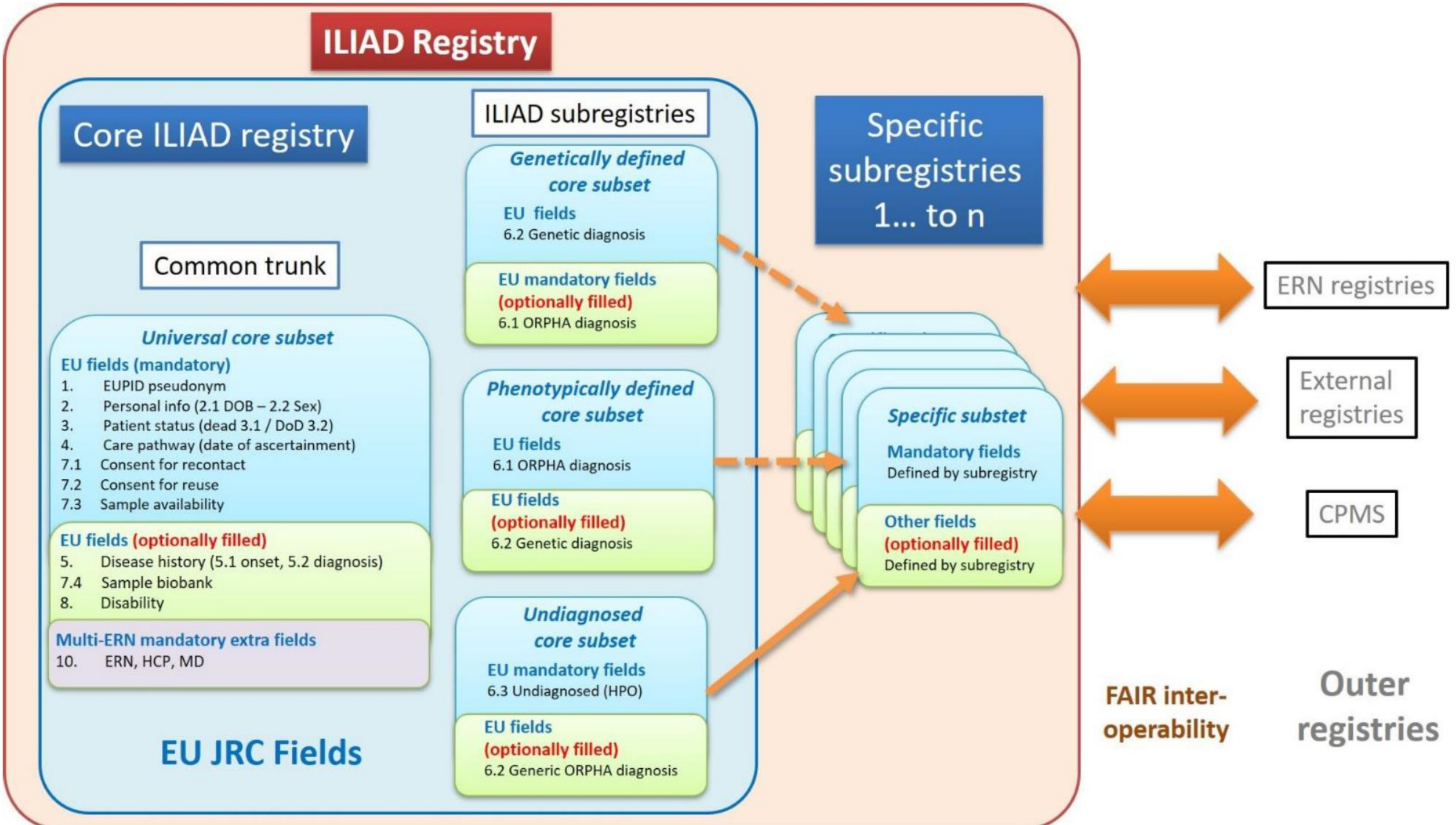
GOAL

ILIAD ‘meta’ registry:
an International Library of
Intellectual disability and Anomalies
of Development

Linking and making visible patient cohorts at the European level in order to follow the natural course of diseases with sufficient patient data

By creating a **central web-based registry** and a network of satellite/client registries linked to the central registry to form the ITHACA **registry federation** (e.g. linking Genida registry)

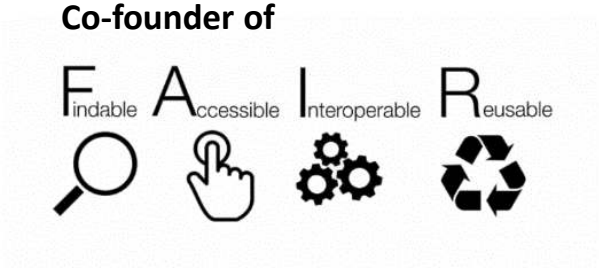
And adding relevant F.A.I.R. metadata and interfaces linking to ERDRI (the EU RD platform) and BBMRI (the EU biobank platform), in collaboration with EJP-RD, to **enable cross ERN collaboration**


ILIAD Registry
Core ILIAD registry
Common trunk
Universal core subset
EU fields (mandatory)
1. EUPID pseudonym
2. Personal info (2.1 DOB – 2.2 Sex)
3. Patient status (dead 3.1 / DoD 3.2)
4. Care pathway (date of ascertainment)
7.1 Consent for recontact
7.2 Consent for reuse
7.3 Sample availability
EU fields (optionally filled)
5. Disease history (5.1 onset, 5.2 diagnosis)
7.4 Sample biobank
8. Disability
Multi-ERN mandatory extra fields
10. ERN, HCP, MD
EU JRC Fields
ILIAD subregistries
Genetically defined core subset
EU fields
6.2 Genetic diagnosis
EU mandatory fields (optionally filled)
6.1 ORPHA diagnosis
Phenotypically defined core subset
EU fields
6.1 ORPHA diagnosis
EU fields (optionally filled)
6.2 Genetic diagnosis
Undiagnosed core subset
EU mandatory fields
6.3 Undiagnosed (HPO)
EU fields (optionally filled)
6.2 Generic ORPHA diagnosis
Specific subregistries 1... to n
Specific substet
Mandatory fields
Defined by subregistry
Other fields (optionally filled)
Defined by subregistry
ERN registries
External registries
CPMS
FAIR inter-operability
Outer registries

F.A.I.R. data principles

Good data management is the key conduit leading to knowledge discovery and innovation. Applies also to ‘non-data’

Co-founder of



To be Findable:

F1. (meta)data are assigned a globally unique and persistent identifier

F2. data are described with rich metadata (defined by R1 below)

F3. metadata clearly and explicitly include the identifier of the data it describes

F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

A1. (meta)data are retrievable by their identifier using a standardized communications protocol

A1.1 the protocol is open, free, and universally implementable

A1.2 the protocol allows for an authentication and authorization procedure, where necessary

A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

I1. (meta)data use a formal, accessible, and broadly applicable language for knowledge representation.

I2. (meta)data use vocabularies that follow FAIR principles

I3. (meta)data include qualified references to other (meta)data

To be Reusable:

R1. meta(data) are richly described with a plurality of accurate and relevant attributes

R1.1. (meta)data are released with a clear and accessible data usage license

R1.2. (meta)data are associated with detailed provenance

R1.3. (meta)data meet domain-relevant community standards

Wilkinson et al, 2016, *Scientific Data: 160018*

MOLGENIS PLATFORM

build using MOLGENIS open source software platform for 'scientific' data, a web-based service.

https://molgenis.org/

Bulk data
Import your own data and metadata model using the EMX format, VCF, OBO

Data request
Find and request (biobank) data sets and items

Data explorer
Filter and download for further analysis

Genome browser
Data sharing and integration DAS protocol

Model registry
Metadata registry of models for biobanks and molecular data

MOLGENIS

Annotators
Data integration for diagnostics and personalized medicine

FAIR integration
Using ontologies to derive harmonization rule for data pooling

R statistics
Use R data api to up/dov data and integrate graphics

Pipelines
Large scale computation on computational clusters, grids and clouds

Reports
Tailormade report functionality

App store
Add your own user interfaces and advanced tools.

Data capture
Questionnaires, forms, importer to easily add data to a running instance

LGPL Free Software

github SOCIAL CODING

elixir Recommended Interoperability Resources

Collaborations

Increase impact by enabling exchange within and between ERNs

By networking ‘all’ registry developers on FAIR principles

ERN CRANIO

ITHACA

SKIN

www.genturis.eu

EUROPEAN JOINT PROGRAMME RARE DISEASES

European Commission Joint Research Centre

Global Alliance for Genomics & Health Collaborate. Innovate. Accelerate.

SolveRD Solving the Unsolved Rare Diseases

RD Connect

BBMRI-ERIC®

elixir

Steering

Morris Swertz (Chair)	Head of Genomics Coordination Center, UMCG
Alain Verloes	ERN ITHACA Coordinator, APHP
Sylvia Huisman	Department of Paediatrics, Amsterdam UMC
Alessandra Renieri	Director of Medical Genetics Unit, AOUS
Jill Clayton-Smith	Professor in Medical Genetics, University of Manchester
David Koolen	Clinical Geneticist Radboud UMC
Carole Herman	ePAG and Patient support

Progress

CONFIGUREd DATA ELEMENTS

Fully Customizable data structure

Model need to be defined can be done in excel(EMX format)

Afterwards changing model with metadata editor

Your Data

sex	birthyear	disease	consent	since
M	1976	NA	Y	1-1-201
F	1977	CV	N	1-1
M	1985	COPD	N	1-1
NA	1985	EB	Y	1-1
F	1940	NA	Y	1-1
F	2001	NA	Y	1-1

Your Meta Data

entity	attribute	dataType	refEntity	required
patient	id	autoid		true
patient	sex	xref	sexValues	
patient	birthyear	int		
patient	disease	xref	diseases	
patient	consent	boolean		true
patient	since	date		

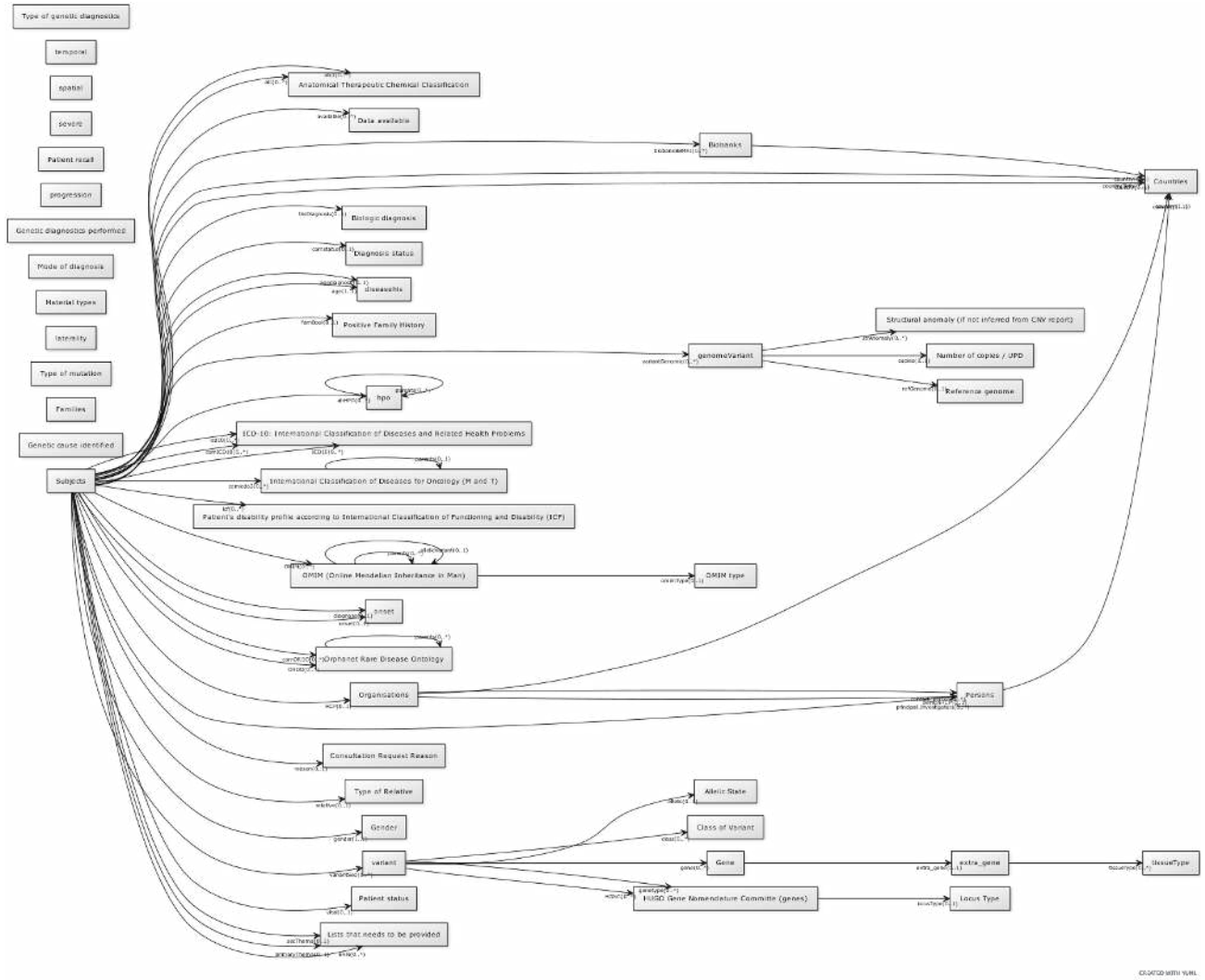
INTER-OPERABILITY

rely on the JRC common data elements, enriched by disease specific elements

5. Disease history	5.1.	Age at onset	Age at which symptoms/signs first appeared	• Antenatal • At birth • Date (dd/mm/yyyy) • Undetermined	
	5.2.	Age at diagnosis	Age at which diagnosis was made	• Antenatal • At birth • Date (dd/mm/yyyy) • Undetermined	
6 Diagnosis	6.1.	Diagnosis of the rare disease	Diagnosis retained by the specialised centre	Orpha code (strongly recommended – see link) / Alpha code/ ICD-9 code/ ICD-9-CM code / ICD-10 code	http://www.orphadata.org/cgi-bin/inc/product1.inc.php
	6.2.	Genetic diagnosis	Genetic diagnosis retained by the specialised centre	International classification of mutations (HGVS) (strongly recommended – see link) / HGNC / OMIM code	http://www.hgvs.org
	6.3	Undiagnosed case	How the undiagnosed case is defined	• Phenotype (HPO) • Genotype (HGVS)	
7. Research	7.1.	Agreement to be contacted for research purposes	Patient's permission exists for being contacted for research purposes	• YES • NO	
	7.2.	Consent to the reuse of data	Patient's consent exists for his/her data to be reused for other research purposes	• YES • NO	
	7.3.	Biological sample	Patient's biological sample available for research	• YES • NO	If YES answer question 7.4
	7.4.	Link to a biobank	Biological sample stored in a biobank	• YES (if appropriate use link) • NO	https://directory.bbmri-eric.eu
8.Disability	8.1.	Classification of functioning/disability	Patient's disability profile according to International Classification of Functioning and Disability (ICF)	• Disability profile / Score	http://www.who.int/classifications/icf/whodasii/en/

State of play with model

277 data elements defined



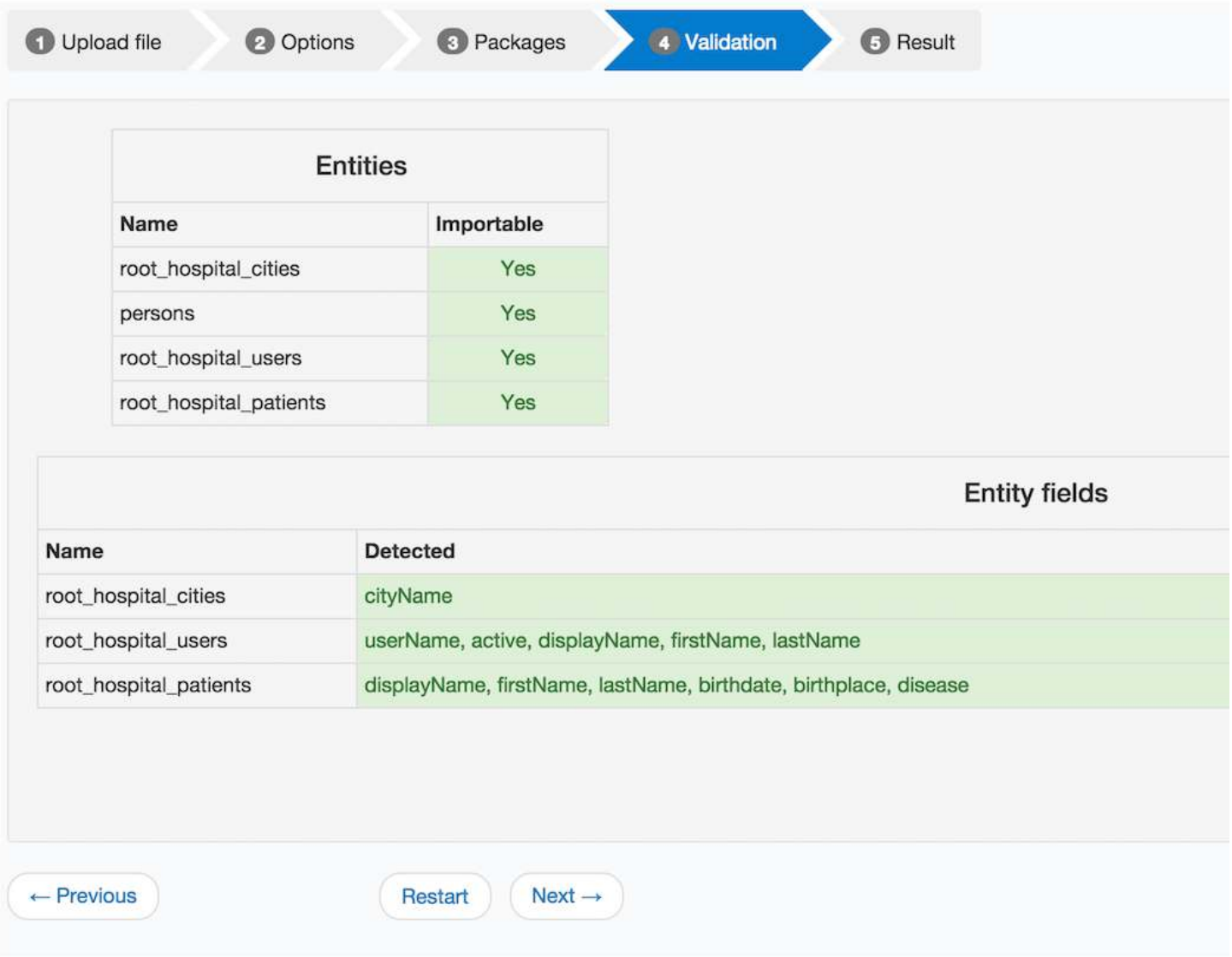
Subjects

description: Collection of all pseudonymized patients

- Attributes:**Period of first symptom onset**: categor
- **Other Remarks**: text
- **Clinical diagnosis**: compound
- **Diagnosis Coding System**: mref(ICD-10: Internati
- **OMIM Code**: mref(OMIM (Online Mendelian Inherit
- **Orphanet nomenclature**: mref(Orphanet Rare Dis
- **Phenotypic Abnormality (HPO)**: mref(hpo)
- **Biologic diagnosis**: categorical(Biologic diagnosis
- **Date of first symptoms onset**: date
- **Age at first symptoms onset**: int
- **Period of current diagnosis**: categorical(onset)
- **Date of current diagnosis**: date
- **Age at diagnosis**: int
- **All codes**: string - *Map all codes to this field (field:*
- **Biologic diagnosis**: compound
- **Link to variant**: mref(variant) - *(same as point 6; G*
- **Link to Genomic variant**: mref(genomeVariant)
- **Family History**: compound
- **Other cases in the family?**: categorical(Positive F
- **If Yes – kinship of Carrier**: xref(Type of Relative)
- **Comorbidities**: compound
- **Other status of diagnosis**: text
- **Diagnosis Coding System**: mref(ICD-10: Internati
- **International Classification of Diseases for Onc**
- **Orphanet nomenclature**: mref(Orphanet Rare Dis
- **Other comorbidities**: text
- **Date of diagnosis**: date
- **Status of diagnosis**: categorical(Diagnosis status)
- **Is the patient following any other medication?**: b
- **Medication**: compound
- **Is the patient receiving medication for his/her ra**
- **Medication for Rare Disease Description**: text
- **Active Ingredient (ATC)**: mref(Anatomical Therape
- **If not ATC**: text
- **Start date medication**: date
- **End date medication**: date
- **Active Ingredient (ATC)**: mref(Anatomical Therape
- **If not ATC**: text
- **General comments**: text
- **Consent**: compound
- **Pseudonym** : compound
- **EUPID/ Person ID***: string
- **Patient enrolment**: compound
- **Patient ID**: string
- **Date enrolled**: date
- **Consent for Care**: bool
- **Date of consent**: date
- **Consent for ERN databases/Registries/research**
- **Date of consent**: date
- **Consent to be contacted for research**: bool
- **Date of consent**: date
- **Retractation to be contacted for research**: bool
- **Consent for usage of biobank material**: bool
- **Date of consent**: date
- **Retractation for biobank use**: bool
- **Date of retractation**: date
- **Retractation for reuse**: bool
- **Date of retractation**: date
- **Last name**: string
- **Date of retractation**: date
- **Research** : compound
- **Patient's biological sample available for researc**
- **Other; what patient Data is available**: categorical
- **Biobank**: mref(Biobanks)
- **Identifying data**: compound
- **First name**: string
- **Record Label**: string - *This Record Label is displa*
- **Patient Status** : compound
- **Patient's Status**: categorical(Patient status)
- **Date of death**: date
- **Care pathway** : compound
- **First Contact date***: date - *Date first contact with s*

BULK Data ingest

Enable upload of datasets.

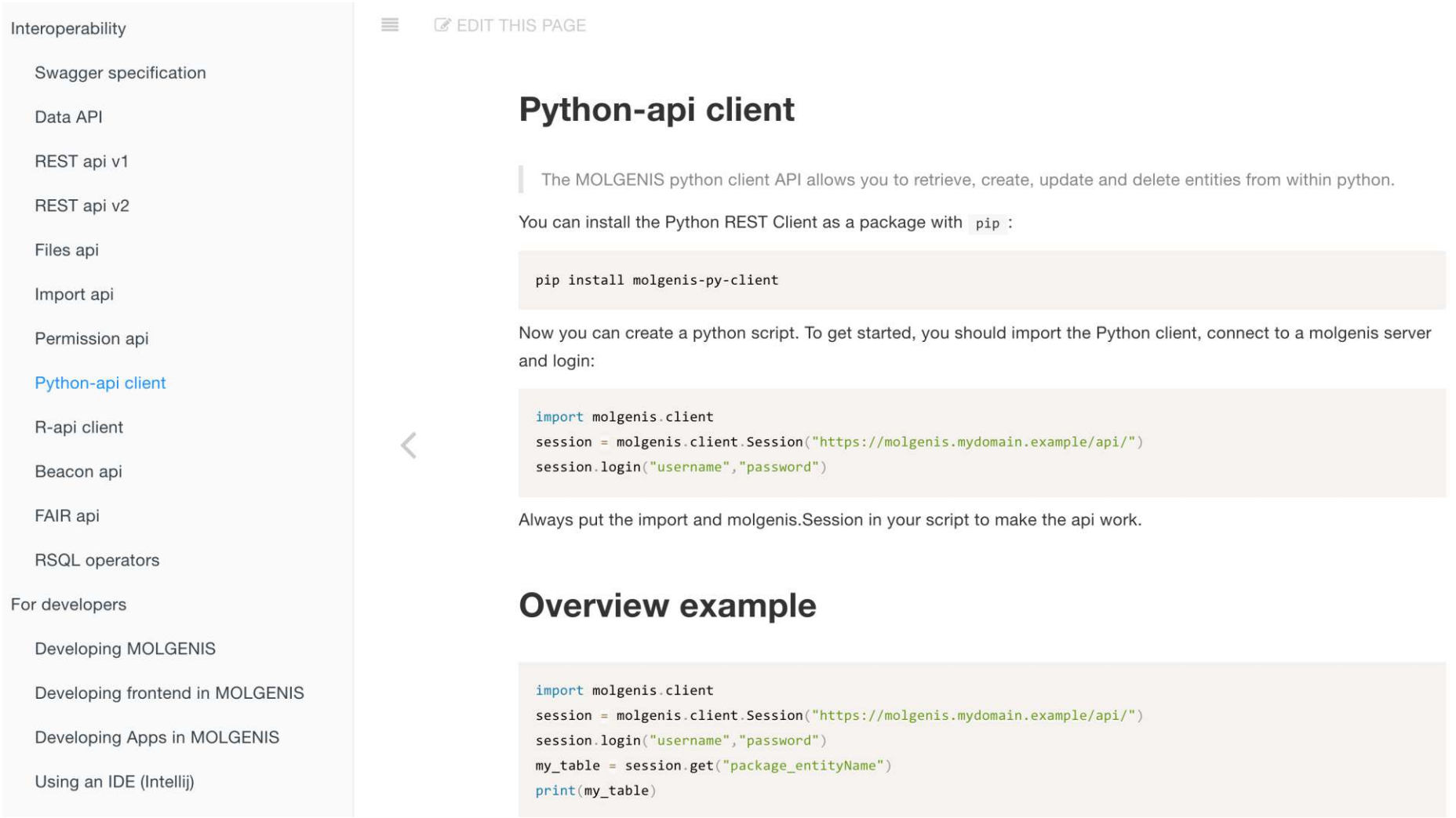


PROGRAM ACCESS

https://molgenis.gitbooks.io/molgenis/

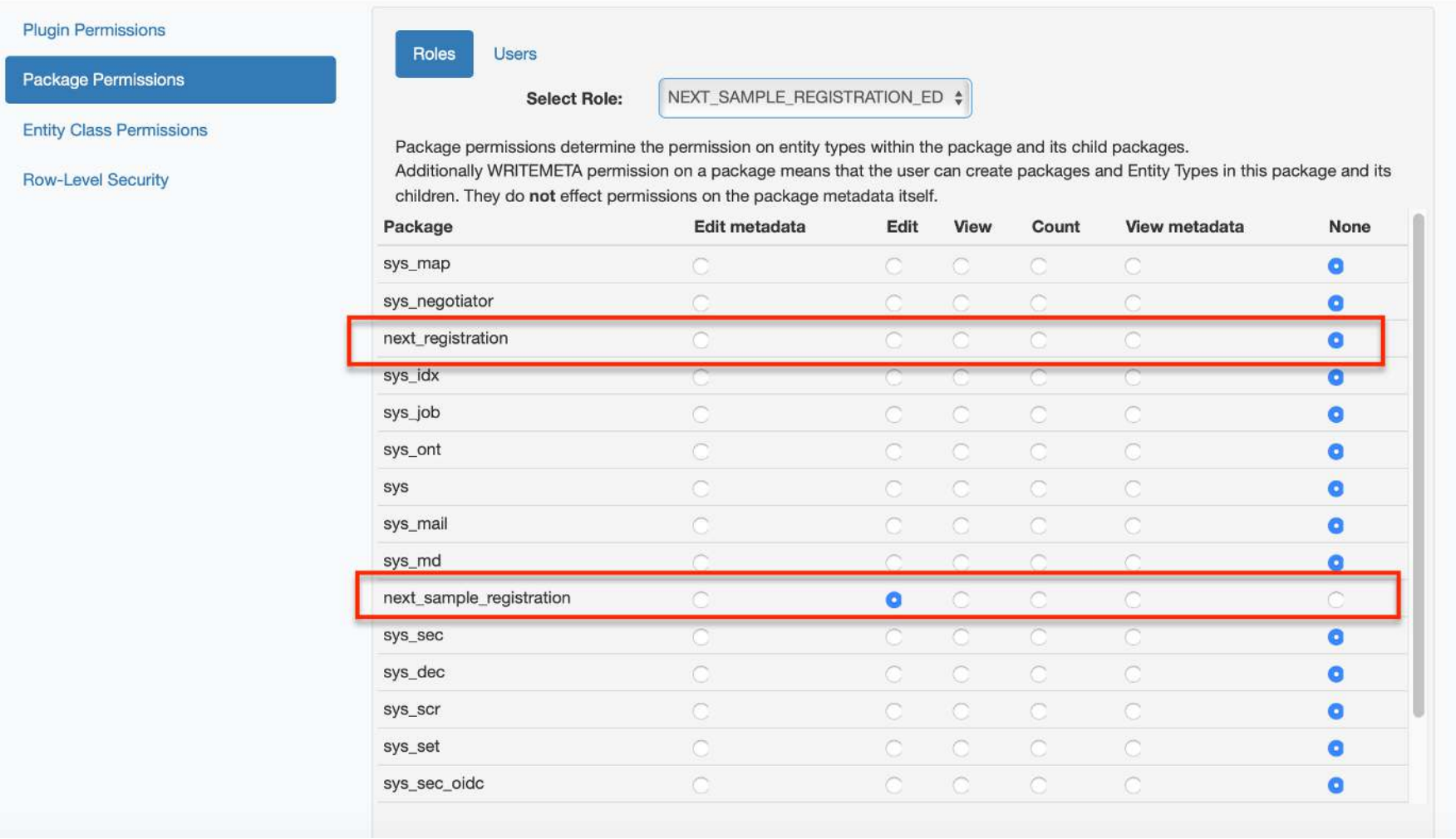
To connect HCPs

To automate data federation



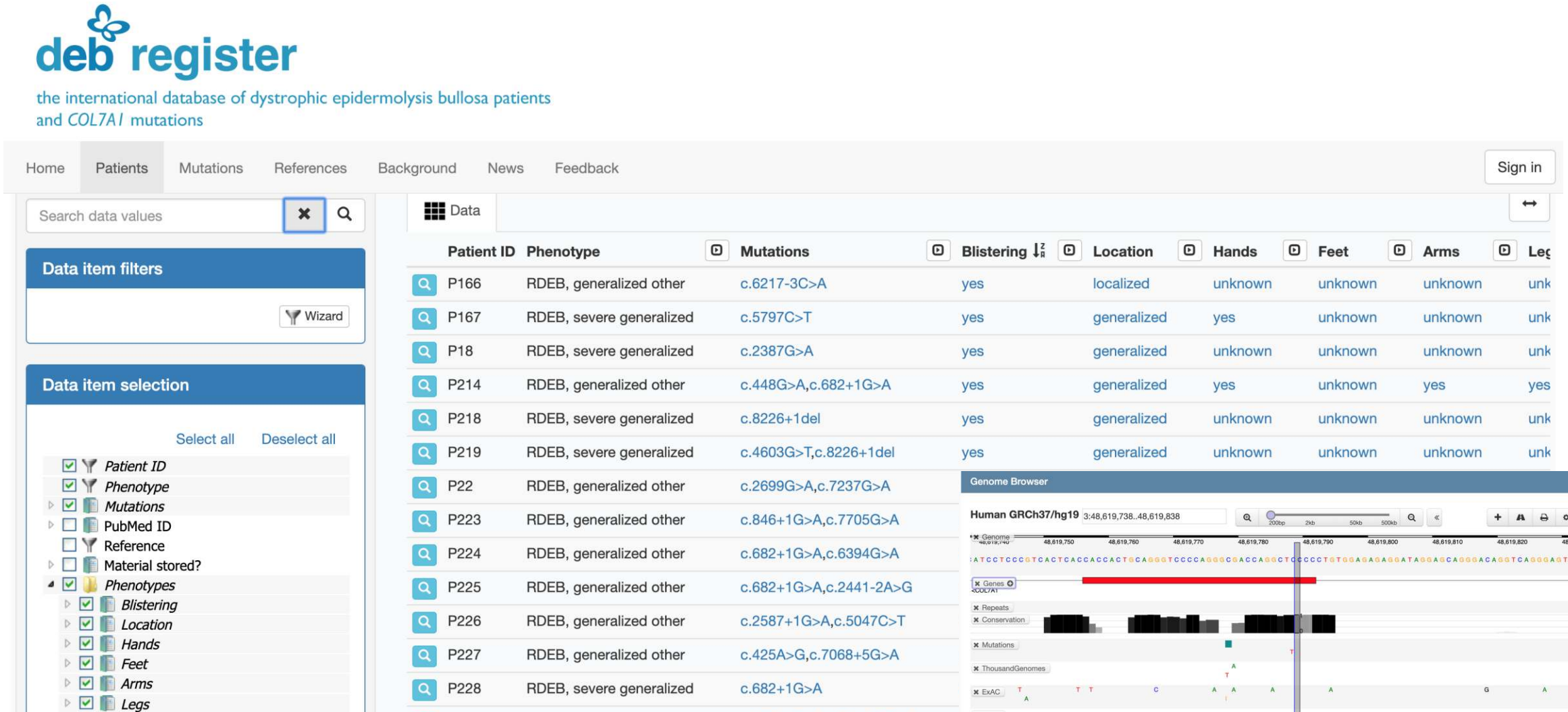
CONTROLLED ACCESS

In permissions system there are different roles (viewer, editor and data manager).



Current activities

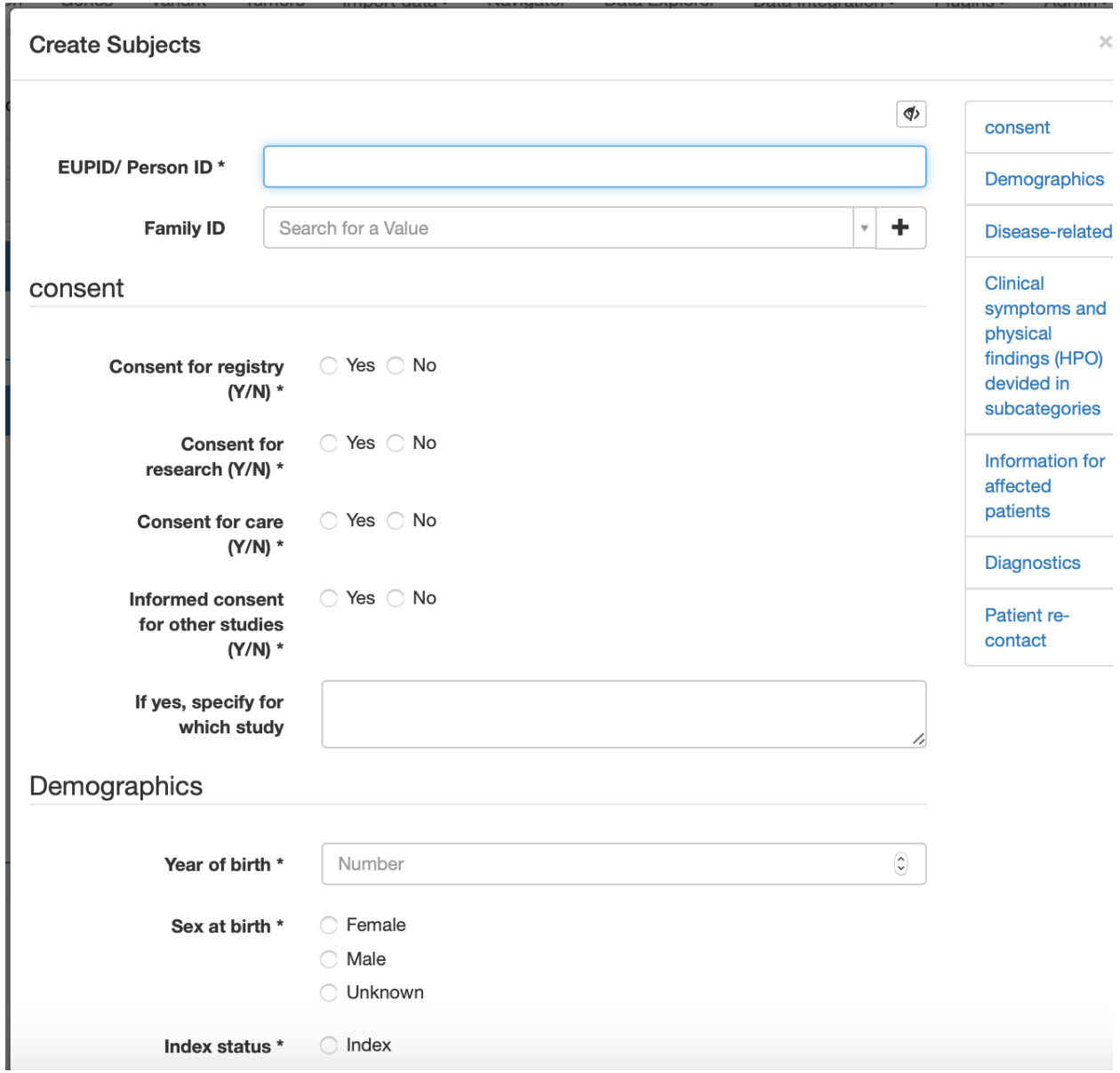
SUITABLE USER INTERFACES TO SEARCH



SUITABLE USER INTERFACES FOR DATA ENTRY

Example from Genturis registry:

Easily create new entries (subjects)



Prototype dashboard

Input needed

All registered patients

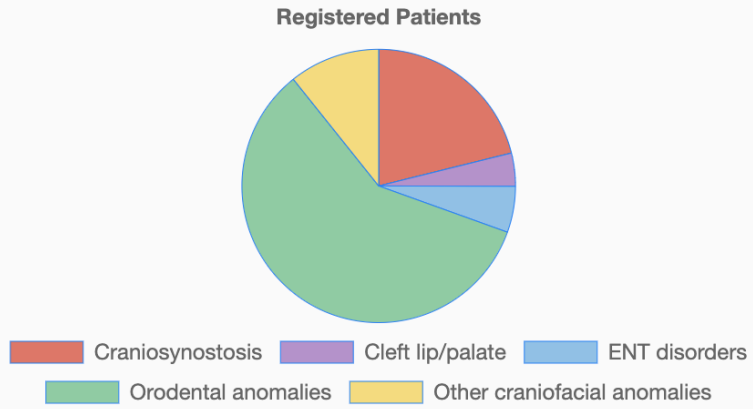
Craniosynostosis

Cleft lip/palate

ENT disorders

Orodental anomalies

Other craniofacial anomalies

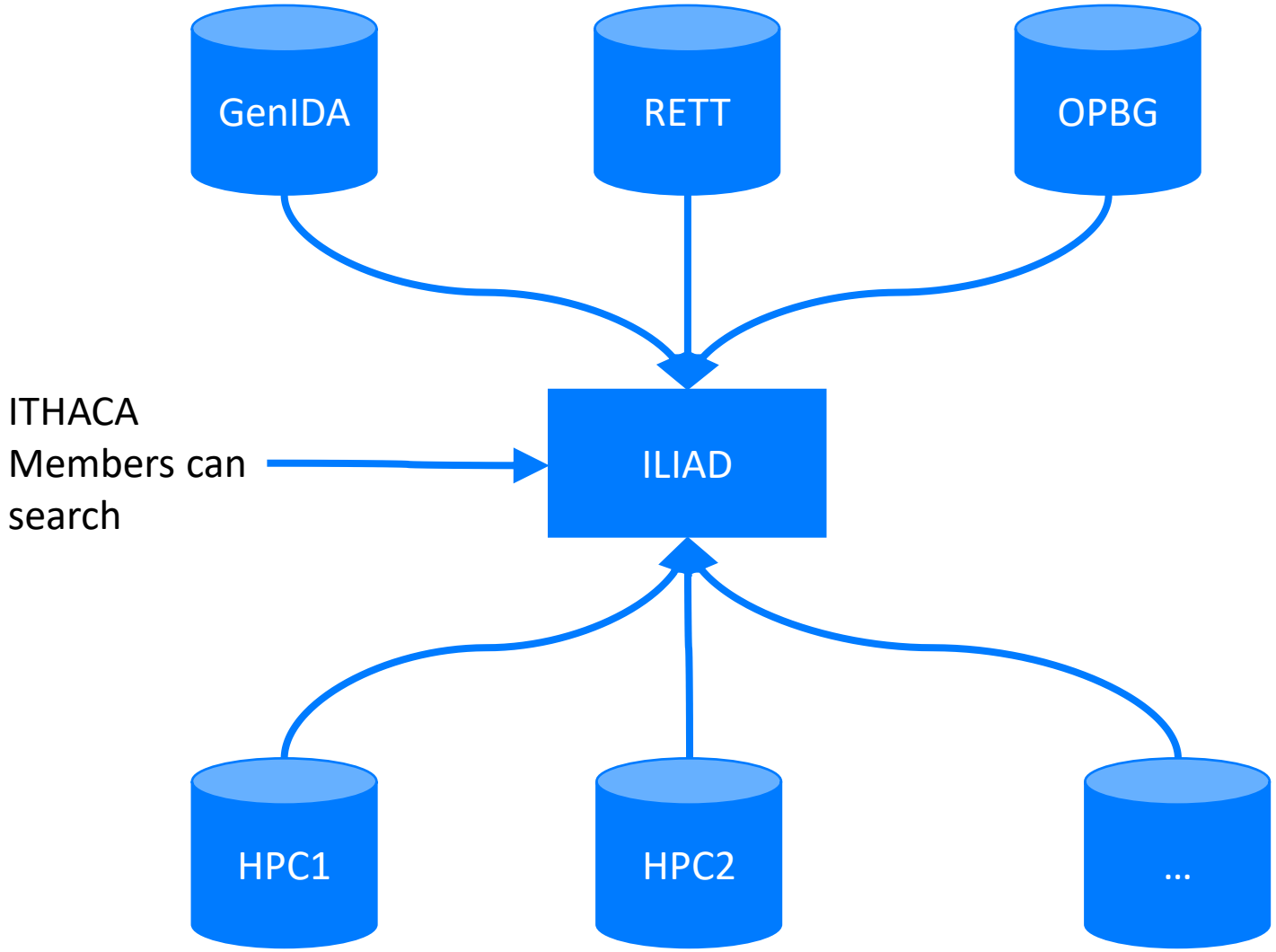


Federating registries and HCPs

Technical connections

Legal connections

Desease specific WGs



EJP FAIR data stewards

Joeri vd Velde from UMCG with 8 others to synergize between ERN registry developers, and with EJP RD developments.

Tool/ Standard name	*Implemented*
ORDO	*10*
Common Data Elements JRC	*9*
ERDRI.dor	*8*
ICD-10	*8*
HPO	*8*
CSV	*7*
HGVS	*6*
HGNC	*6*
ERDRI.mdr	*5*
Excel Reports	*5*
OMIM	*4*
Molgenis	*4*
MOLGENIS metadata structure	*4*
EUPID	*3*
RDF/XML	*3*
RedCap	*3*
Own Registry System	*3*
Epic EHR	*3*
CDE Semantic Model	*2*
ICD-11	*2*

Tool/ Standard name	*Need expert*
CDE Semantic Model	*8*
EUPID	*7*
ERDRI.mdr	*6*
Common Data Elements JRC	*3*
FAIR Data Point	*3*
RML	*3*
EJPRD Metadata Model	*2*
ORDO	*1*
ERDRI.dor	*1*
RDF/XML	*1*
JSON LD	*1*
SPARQL	*1*
ATC	*1*
iCRF Generator	*1*
Turtle	*1*
ICD-10	*0*
HPO	*0*
CSV	*0*
HGVS	*0*
HGNC	*0*

Training

+ Add task

need guidance | need expert input

- GENTURIS, SKIN, CRANIO, ITHACA - needs more information on ERDRI.mdr utility / added value

MN

task is a bit vague | need expert input

- GENTURIS, SKIN, CRANIO, ITHACA - needs more information on EJPRD metadata model

MN

need guidance | need expert input

- GENTURIS, SKIN, CRANIO, ITHACA - what is the use of phenopackets for the registry?

  need to know from other ERNs experience, and what is the utility?

MN

need guidance | task is a bit vague

- GUARD-HEART - Query Language

  They need to know more about "querying".

Community

+ Add task

waiting for answer

- ERN GUARD-HEART: how other ERNs deal with Diagnosis and Genetirc Diagnosis Elements

  They wonder how other ERNs collect information from the two data elements (6.1 Diagnosis and 6.2 Genetic Diagnosis).

need guidance | waiting for answer

- Skin: share ERN tools/standards overview

  request for sharing the EJP-RD FAIRification stewards overview of standards and tools chosen to build registries across all ERNs

need better resource

- Skin: share the database templates other ERNs

  request to share any database templates across ERNs, in addition to used standards/tools etc.

need guidance | task is a bit vague

- Skin: learn from others to overcome datasharing issues

need guidance

- EpiCARE - learn from others

Modelling

+ Add task

need expert input

- EURACAN - CDE for Rare Cancer

  This ERN is focused on rare types of cancer. Since cancer is not a RD, but the types are, they are worried that the CDE (or even other cases) might not fit well for them.

need guidance | waiting for answer | need expert input

- EURACAN - Difference between CDE Model and OMOP-CMD

  The ERN is planning to use the OMO-CMD model as common data elements. They want to know what would be the interoperability impact when using EJPR-RD CDE Model or OMOP-CMD.

need guidance | need better resource | need expert input

- VASCERN - RADICO - plan to collect treatment info, ATC ontology is not free, need alternative. would like a table with free ontologies options

need guidance | need expert input

- ERN GUARD-HEART - Ontology Recommendation for Diagnosis

  For 6.2, which ontology (HVGS, HGNC, OMIN) should ERN choose?

Implementation

+ Add task

need guidance | waiting for answer

- MERITA - Data integration

  MERITA registry considers offering an Excel import function. Yet they are awaiting for knowing whether a common template would be made available.

! 20/11

need expert input

- VASCERN - Radico - EUPID match: how can we see the "similar name" patient that is already registered. how do i know which patient it is similar to?

need expert input

- VASCERN - RADICO - need EUPID specifications. have 250 patients, need bash implementation or automated.

waiting for answer

- VASCERN - RADICO - would like erdri.mdr import template.

Bruna DosSantosVieira

waiting for answer

- VASCERN - RADICO - iCDF generation

Legal

+ Add task

need guidance | waiting for answer

- EURACAN - Data transfer agreement

  This ERN works with a set of federated databases, and would like to know how others have been dealing with data transfer policies among databases.

need guidance | waiting for answer

- GENTURIS - Legal issues data sharing

  GENTURIS wants to know about legal aspects of data sharing, specifically those of software providers (data consumers) and hospitals/institutes (data producers). ERNs are not legal entities, so difficult. Perhaps

need expert input

- ENROL - EUPID

  Legal questions about policy and licenses issues. Need information from experts, if they can clarify how local registries use EUPIDS, so that they can generate pseudonyms and then provide them to the

MN Michael Nitzlnader

need expert input

- RND - Patient consent

  They use their own hospital-patient agreement. Is this allowed?

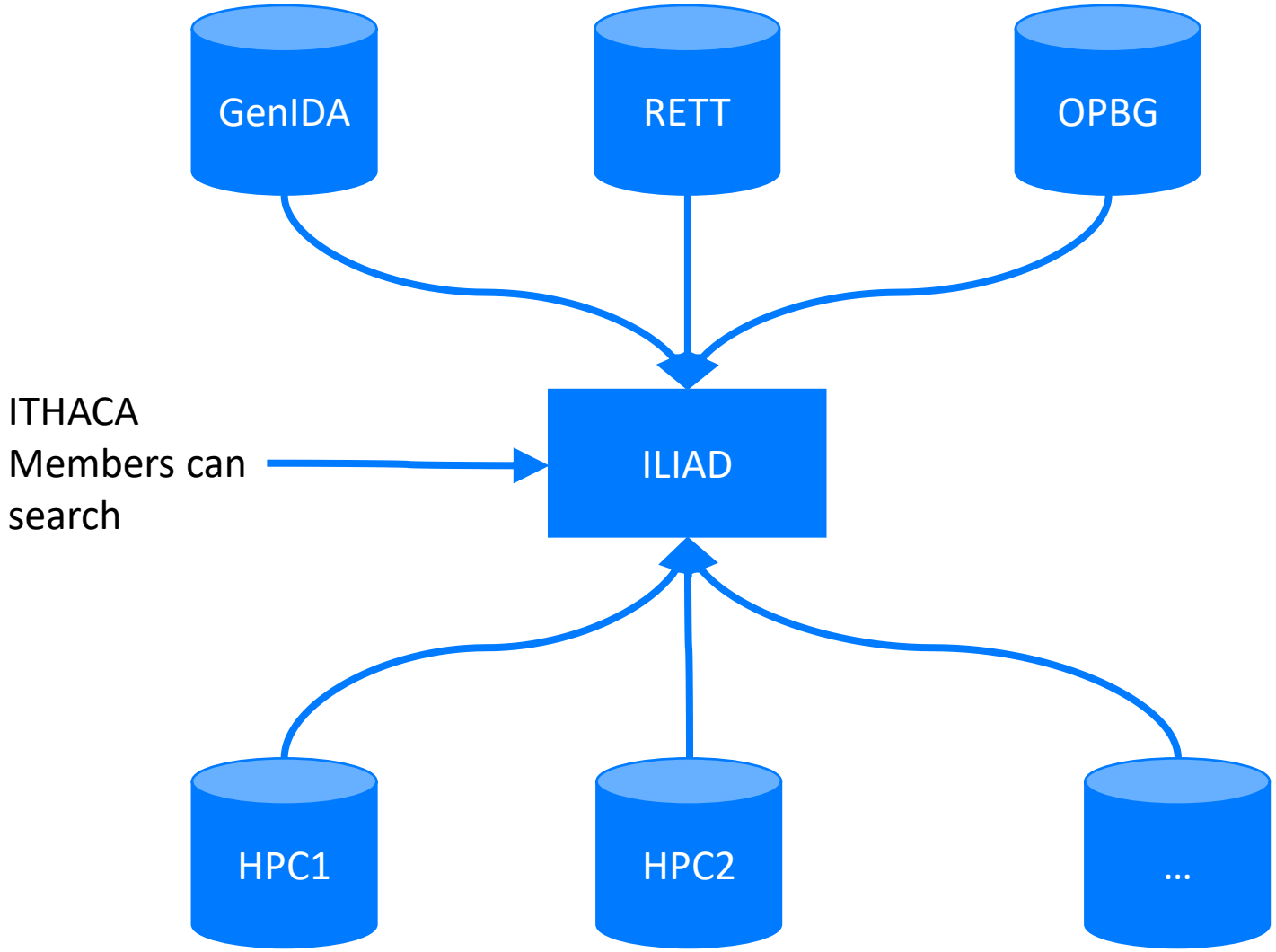
Future steps

Federating registries and HCPs

Technical connections

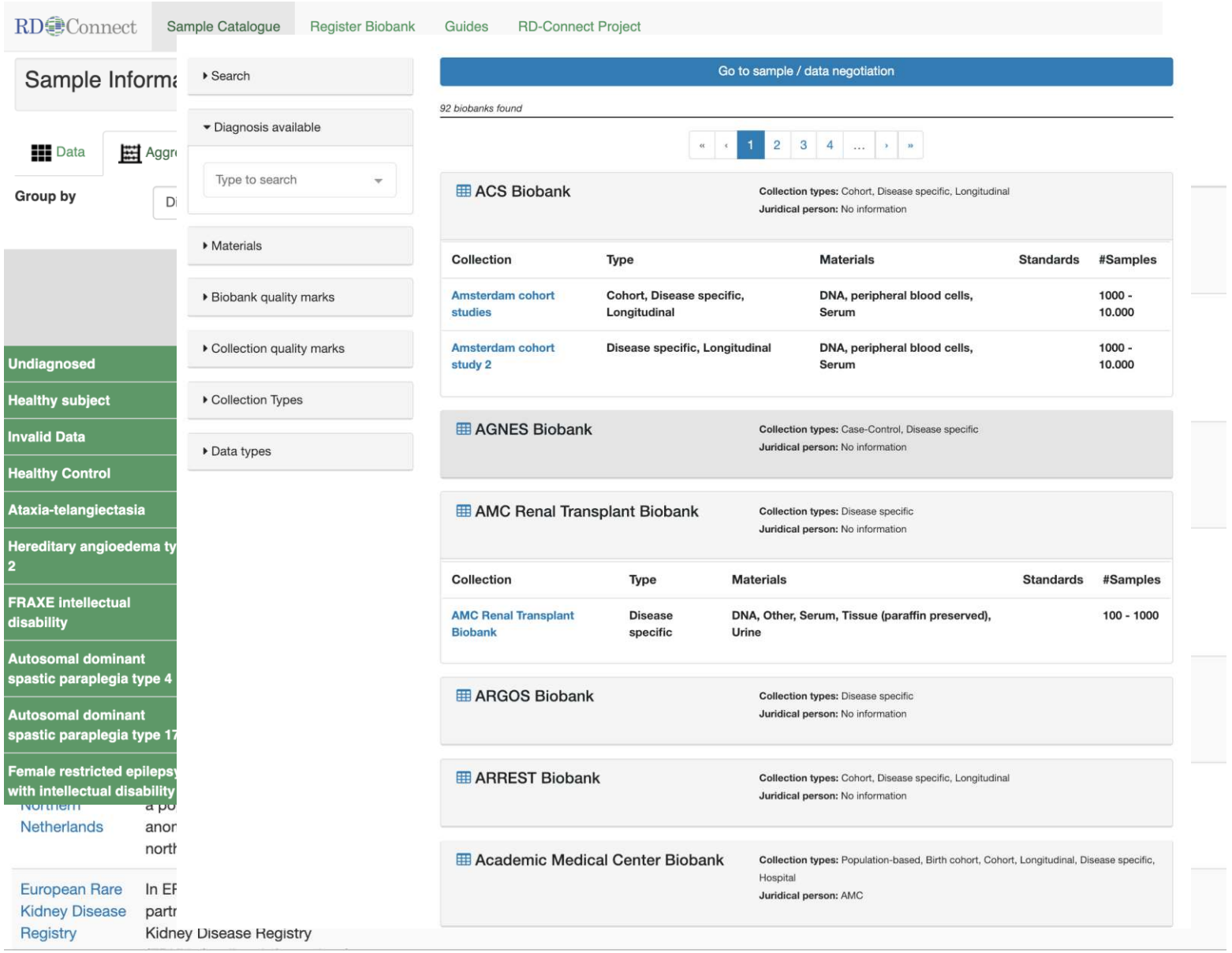
Legal connections

Desease specific WGs



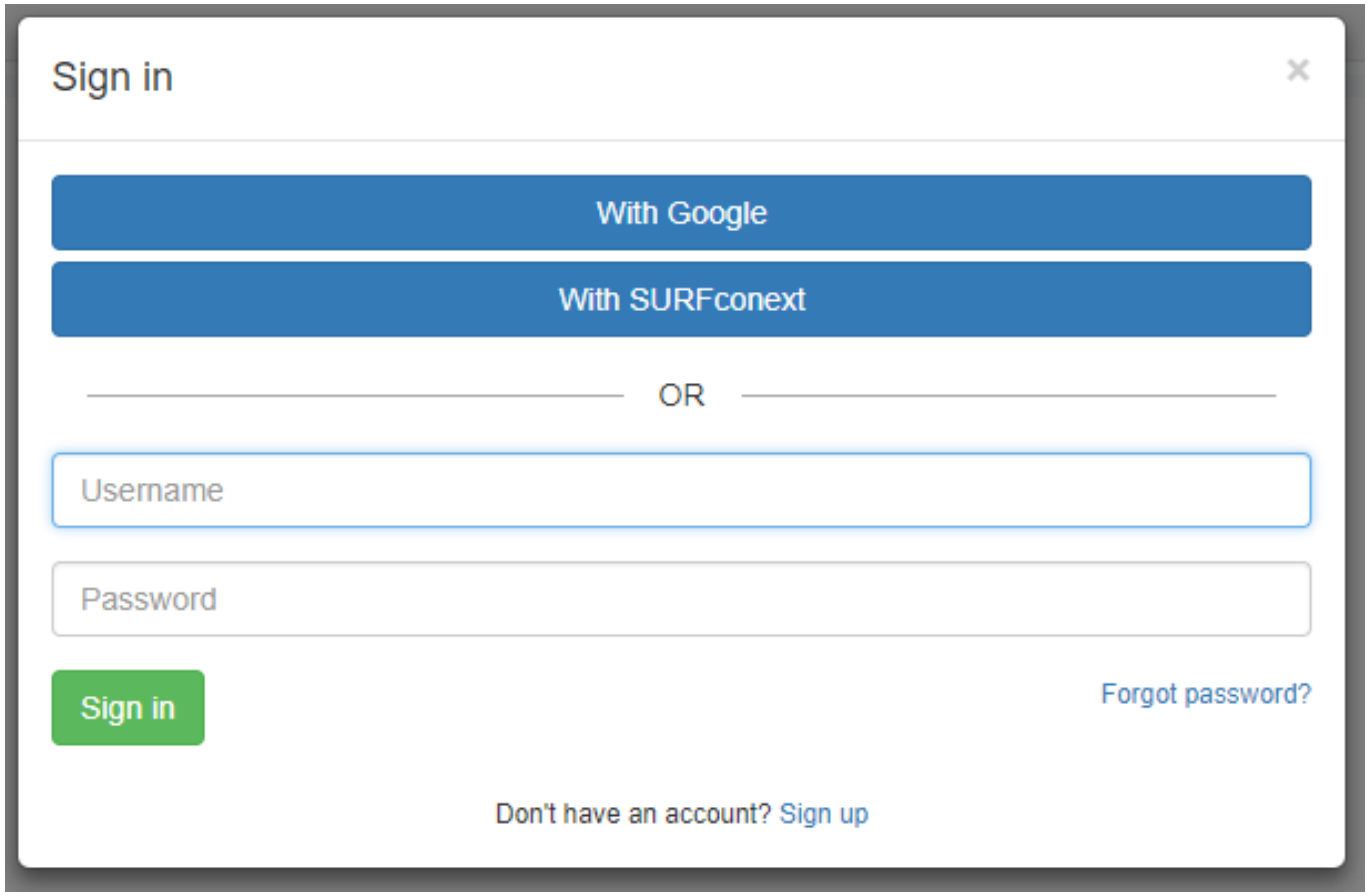
FINDABILITY

To support findability, the registry will be connected catalogues such as the European Directory of Registries (ERDRI.dor), BBMRI catalogue, RD-connect registry & biobank catalogues.



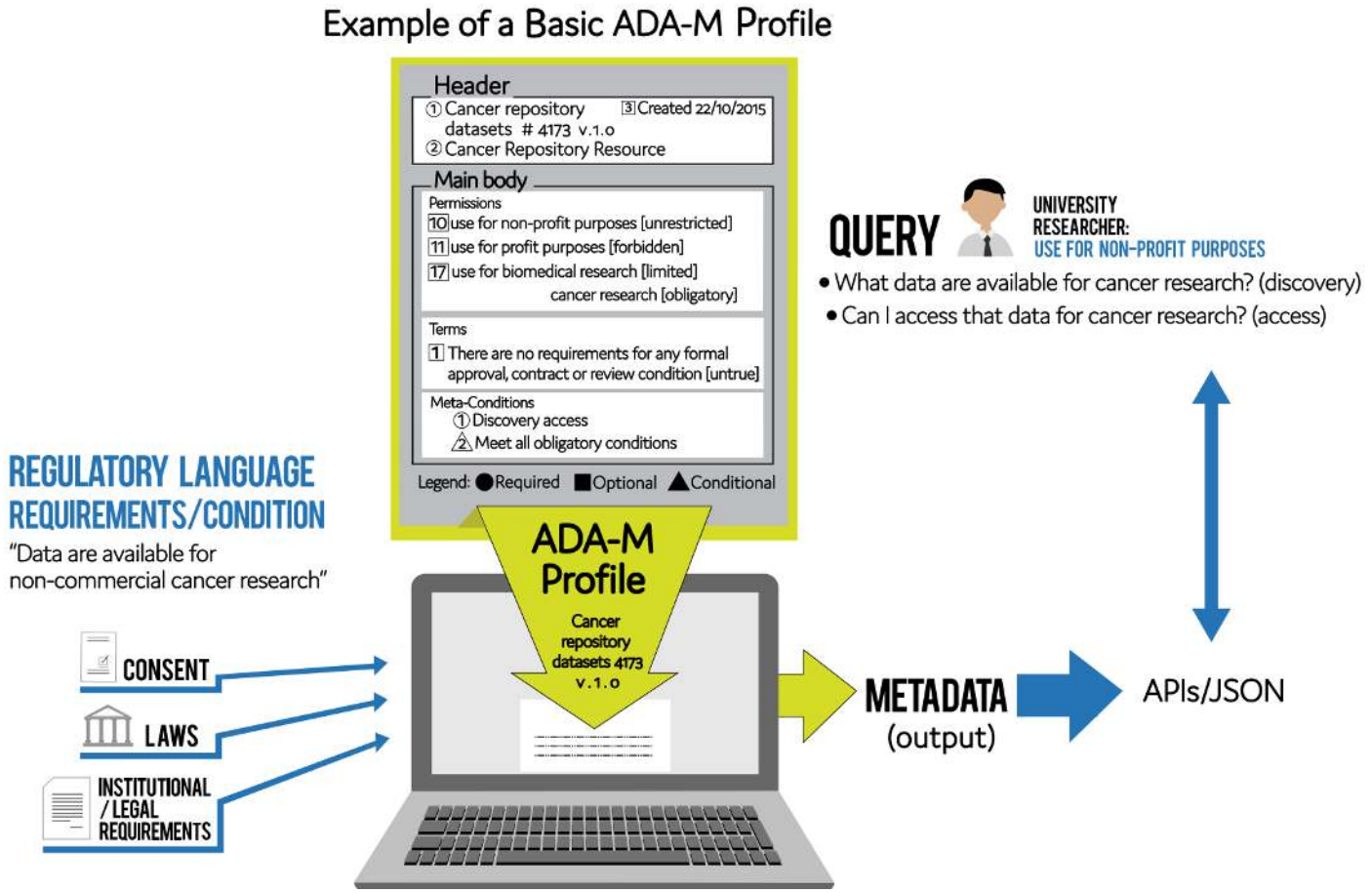
ACCESS

Integrate into existing authentication systems to ensure those accessing have the proper credentials to view sensitive data (e.g. ELIXIR AAI)



ACCESS (2)

Per project different rules, different information on access conditions.

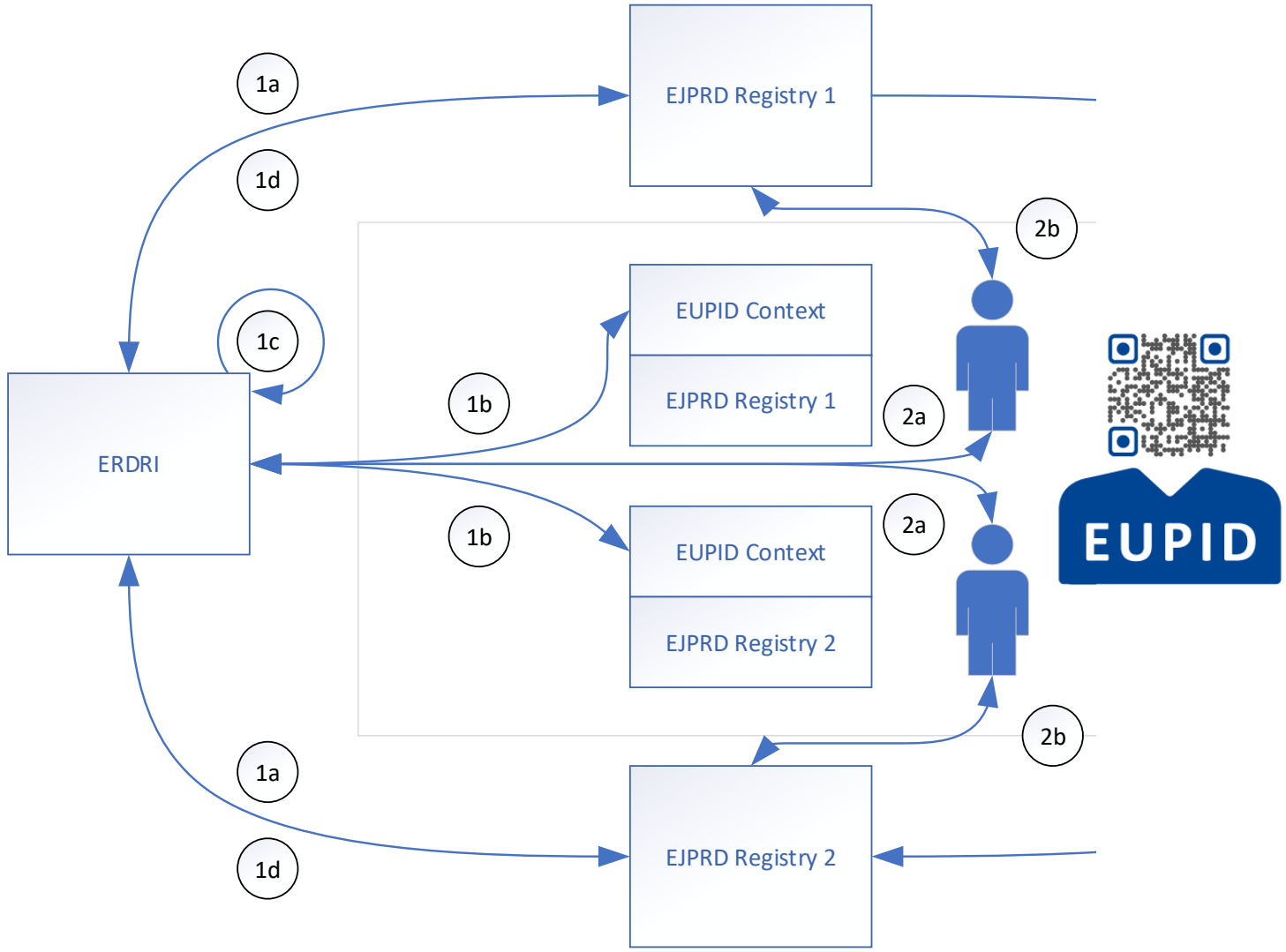


Woolley, J. P., Kirby, E., Leslie, J., Jeanson, F., Cabili, M. N., Rushton, G., … Brookes, A. J. (2018). Responsible sharing of biomedical data and biospecimens via the "Automatable Discovery and Access Matrix" (ADA-M). *Npj Genomic Medicine*, *3*(1), 1–6. https://doi.org/10.1038/s41525-018-0057-4

INTER-OPERABILITY

To enable linkage of patients across ERNs we will use the ERDRI Pseudonymisation Tool (EUPID) to allow linking RD patients cohorts

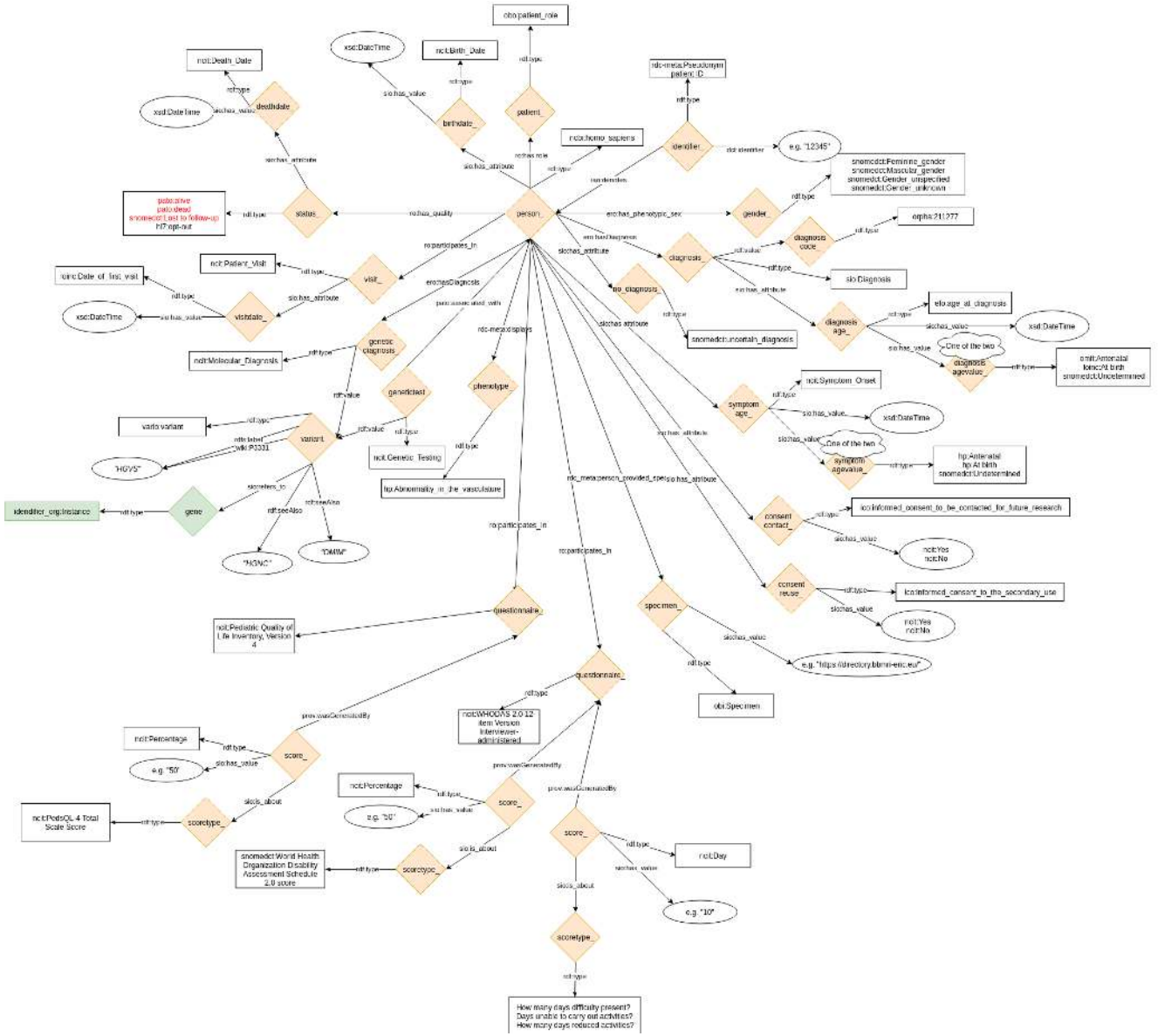
https://eupid.eu/#/home



Courtesy of Michael Nitzlnader, AIT, et al

INTER-OPERABILITY

To ease cross search, we will develop extensions together with other registries using EJP RD, GA4GH, BBMRI, ELIXIR, etc, and make the meta-data available in the Central Metadata Repository(ERDRI.mdr)



Courtesy of Annika Jacobsen, Rajaram Kaliyaperumal, Marco Roos, Peter-Bram ‘
https://github.com/LUMC-BioSemantics/ERN-common-data-elements

DEMO

Acknowledgements:

- ERN teams
- MOLGENIS partners past 15y
- GEN2PHEN
- RD-connect
- Solve-RD
- EJP RD
- ELIXIR
- BBMRI

MOLGENIS

umcg

http://molgenis.org

Morris Swertz m.a.swertz@gmail.com

MOLGENIS support molgenis-support@umcg.nl

Genomics coordination center