ERNs Frequently Asked Questions and Training Needs

Bruna Vieira, César Bernabé, Franz Schaefer,
Marco Roos and Shuxin Zhang

EUROPEAN JOINT PROGRAMME RARE DISEASES

[ERNs Conference]
[2020-12-15]
[Online]

Walkthrough

- FAIR
- ERNs FAQs and expert answers
  - ERICA Project
- EJP RD next steps
- Prioritize training needs

Funded by the European Union
GA n°825575

Disclaimer

This is a first version of the ERNs FAQs collected by the FAIRification Stewards. The collection is an ongoing process taking place via email, MS Teams calls and local institutional groups.

Please react in the chat and your questions will be addressed after the presentation.

Funded by the European Union GA n°825575

Why FAIRification?

Identified by global leaders in data science, Lorentz Workshop, Leiden, 2014

- **Data reuse over multiple sources *unacceptably inefficient***
- **Quantifiable cost for the European economy > *€10.2 billion per year***
  (unquantifiable elements probably +€16 billion)
  source: EC report 2019-01-16, DG Research and Innovation (EC) & PwC EU Services
- **Ambition: make that NOT TRUE for new ERN registries**

EJP RD

Funded by the European Union
GA n°825575

FAIR principles are straightforward

- Findable, Accessible, Interoperable, Reusable
- for humans and computers

Funded by the European Union
GA n°825575

Implementing FAIR principles *is not*

Funded by the European Union GA n°825575

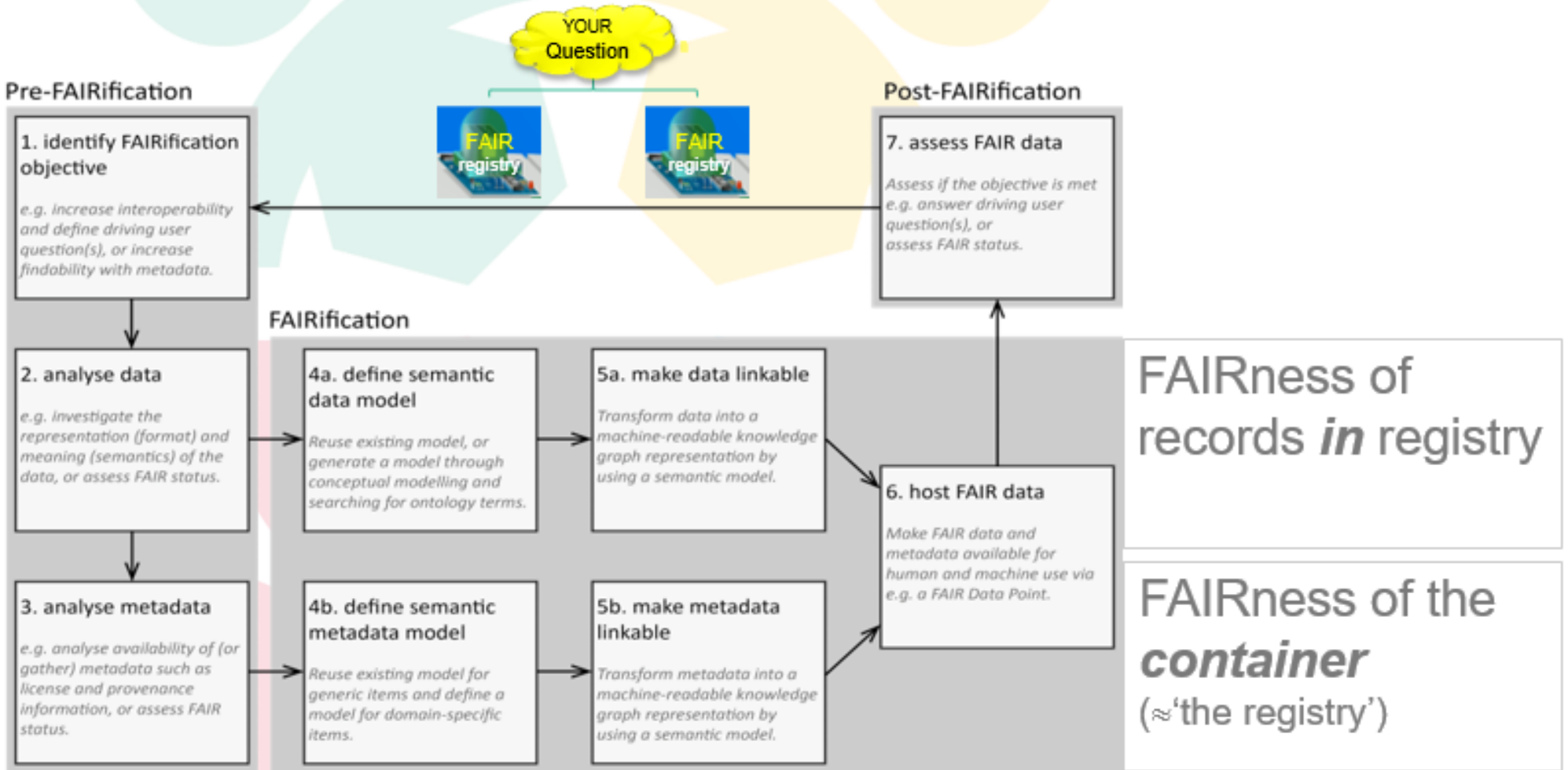
Implementing FAIR **is a** process

EJP RD

Funded by the European Union
GA n°825575

General steps to make a data resource FAIR

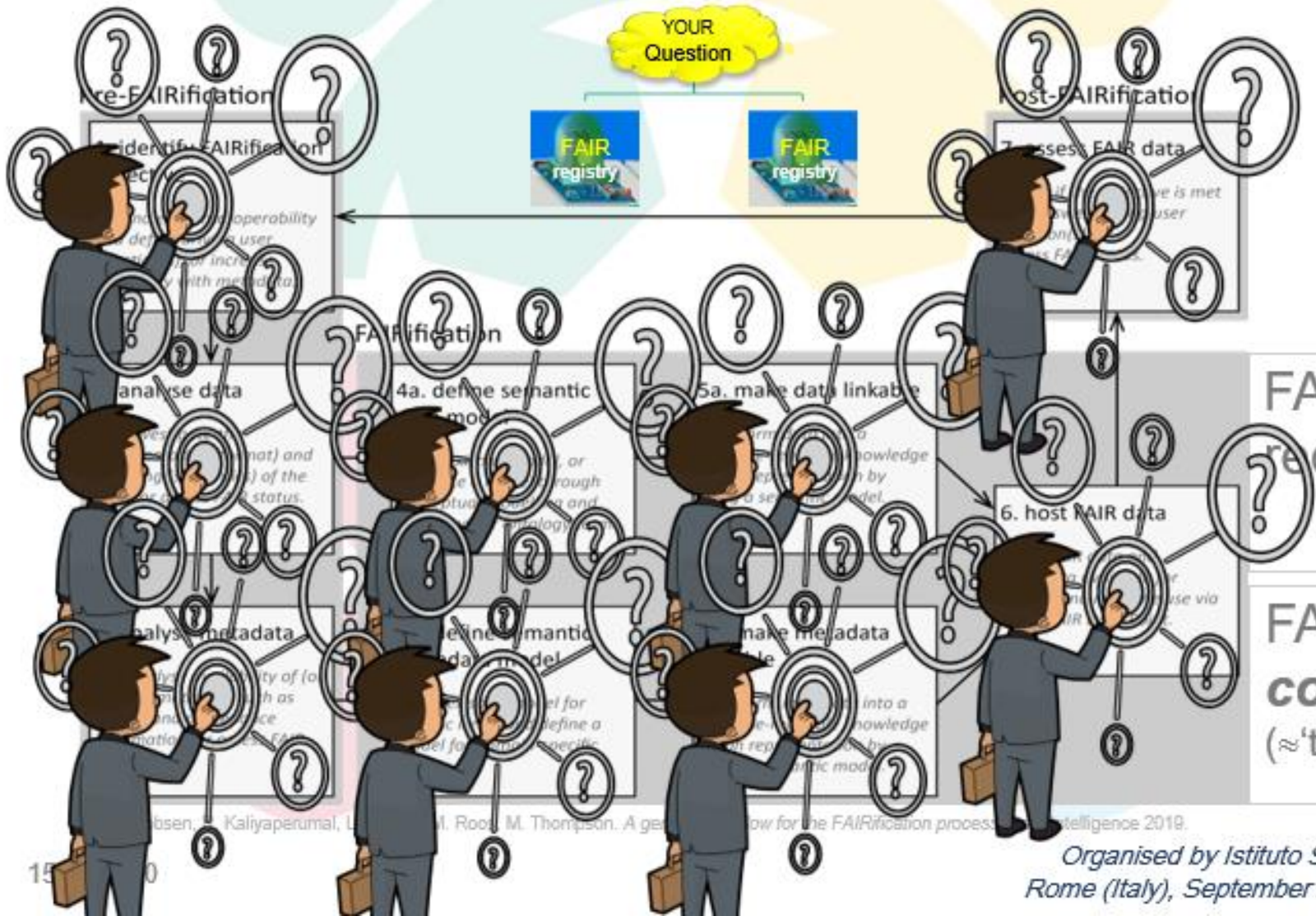
Generic workflow strongly based on RD feedback (e.g. data linkage pilots and Rome summer school)



A. Jacobsen, R. Kaliyaperumal, L. Bonino, M. Roos, M. Thompson. *A generic workflow for the FAIRification process.* Data Intelligence 2019.

Challenge: choices to make a data resource FAIR

Generic workflow strongly based on RD feedback (e.g. data linkage pilots and Rome summer school)



FAIRness of records ***in*** registry

FAIRness of the ***container***
(≈'the registry')

Funded by the European Union GA n°825575

Why is there not ONE easy answer for each step?
FAIRification is a balancing act

Existing (local) systems
Established practices
etc.

"Ideal" implementation of FAIR principles

FAIRification

Funded by the European Union
GA n°101057571

Travel together: core ERN FAIRification team

Help manage FAIRification process of ERN registry

Managed through EJPRD

EJPRD FAIRification steward

Registry software provider

ERN registry steward

Implementing/configuring the software for the registry

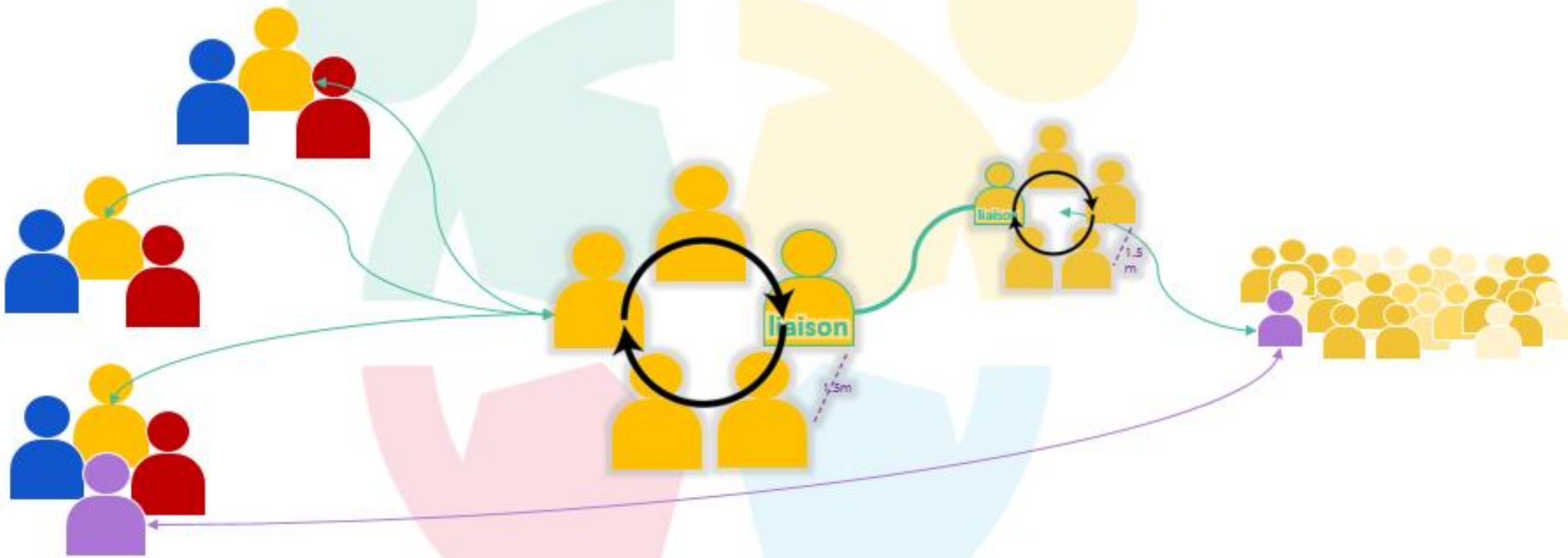
Managed through ERN registry project

Data management within the Registry project

Managed through ERN registry project

Funded by the European Union GA n°825575

Travel together… with the EJPRD



ERN FAIRification teams

EJPRD FAIRification stewards

EJPRD overall architecture teams

EJPRD VP & infrastructure researchers

EJP RD

Funded by the European Union
GA n°825575

Some take-homes before FAQ

- FAIRification is a process to travel through together
- Implementation choices develop between stewards and EJPRD (and other) experts

- **Now: your questions!**

Funded by the European Union
GA n°825575

EUPID

- *Can we add own pseudonymization number to EUPID field (e.g. from mrdm.nl)?*
- *What are the implications of not using EUPID?*
- *Is it mandatory?*

1. EJP RD recommends to work with pseudonyms according to the set of CDEs. The pseudonymisation number of the source can be added as an additional field.
2. Without EUPID we cannot avoid double patient counting between ERN registries and that we cannot link data of the same patient in different registries.
3. EUPID can be used for free via the EU RD Platform. For detailed information on how EUPID works please request specifications from contact@eupid.com. The utilisation of EUPID for patient registration and pseudonymisation is mandatory for the registries.

EJP RD Expert: **contact@eupid.eu**

Funded by the European Union GA n°825575

CDE Minimal Dataset

- *Are all elements mandatory?*
- *Can you better define the CDE elements?*
- *What is the added value of using them?*

1. The inclusion of the CDE in the registries is mandatory as per the EC ERNs grant demands.
2. Further description can be found on the CDE Semantic Model and by request to JRC. EJP and JRC will jointly develop a descriptive document.
3. Implementing the CDEs is the first step towards interoperability and is an essential prerequisite for further research based on registry data.

Experts: **Holger Storf, Franz Schaefer, and JRC**

Funded by the European Union GA n°825575

Ontologies and Vocabularies

- *Genetic Diagnosis: Should all use HGNC, HGVS and OMIM ? What to use it for? Is COSMIC (cancer) ontology a good choice for genetic diagnosis?*
- *Clinical diagnosis: Can we use Orpha code only? Can we use own field classification not in Orphacode?*

1. You may use all vocabularies mentioned in the CDEs pdf depending on which data you collect.
2. HGNC - gene names; HGVS - variants; Orpha codes - “diseases”; ORDO - Orphanet, OMIM and HGNC.
3. Next year EJP will evaluate other vocabularies and ontologies.
4. Additions to Orpha codes can be requested at disease.orphanet@inserm.fr, and mappings to Orpha codes at http://www.rd-code.eu.

EJP RD Experts: **Marc Hanauer, Ana Rath**

Funded by the European Union
GA n°825575

Ontologies

- *What are the recommended free/open ontologies (e.g. ATC)?*

1. The resources mentioned in the CDE pdf are recommended by JRC.
2. EJP RD works on so called 'application ontologies' (e.g. CDE semantic model) which groups recommended ontologies (e.g. from CDEs minimal dataset). The implementation aspects will be communicated in dedicated workshops.

Note: SNOMED CT is not recommended for registries. If used in EHRs based under national licences, mapping to the Orphacode is possible.

EJP RD Experts: **Ana Rath, Marc Hanauer, Rajaram Kaliyaperumal**

Funded by the European Union GA n°825575

CDE Semantic Model

- *What are the added values of implementing the CDE Semantic Model?*

1. It incorporates all recommendations from JRC in the set of CDEs, including ontologies. Data structured within this model is readable and unambiguous by and for computers.
2. The model can be mapped and used in combination with other models such as CDISC/ODM and ODHSI/OMOP. Implementation workshops will be arranged.
3. It increases the potential to link resources (EHR – EDC Systems) easing the data collection process.

EJP RD Experts: **Mark Wilkinson (CDE Modelling group)**

Funded by the European Union GA n°825575

EJP RD Metadata Semantic Model

- *What is it, and what are the added values?*

1. Metadata is "data about data". The benefits of using the metadata model is that it makes your registry findable and accessible for computers.
2. It can be used by the EJP RD Virtual Platform and application programming interface (API) specifications such as the FAIR Data Point and Beacon.

EJP RD Experts: **Rajaram Kaliyaperumal, Marc and Olamidipupo (Metadata WF)**

Funded by the European Union GA n°825575

Data Formats

- *Which data representation format to use? And which formats are ERNs currently using?*

1. The formats will be discussed in the workshops.
2. The cleaner and clearer the data is (one record per row, explicit column names, no empty cells, use same date formats, etc), the easier it is to transform to a FAIR format.
3. ERNs are currently using Excel tables, JSON, CSV, and ontologized data formats for their registries.

EJP Experts: **Mark Wilkinson (RDF), Ronald Cornet (clinical), Sergi Beltran and Anthony Brookes (genomics).**

Funded by the European Union GA n°825575

Data Formats

- *Is there an EJP RD common excel template for registry data exchange?*
- *How can phenopackets be used in registries?*
- *Why do I need phenopackets if my registry is already FAIR?*

1. For now, there is no official excel template for registries or CDEs, but ERNs set up a CDEs codebook also available in table (ask stewards). The aim is to carefully-structure data, that can be transformed into an easily exchangeable format (FAIR).
2. Phenopackets can be used as an exchange format for phenotype data. If your registry is FAIR, it is easier to add extra exchange formats such as Phenopackets. A workshop for Phenopackets will be arranged.

EJP Experts: **Ronald Cornet, Núria Queralt**

Funded by the European Union GA n°825575

FAIR Data Point

- *What is it? And how does it work?*

1. It provides a way for users and computers to read information (i.e., metadata) on how to Find, Access and Reuse your registry.

2. Depending on access mechanisms and license conditions, the actual data can be accessed, queried and/or downloaded.

3. Technical guidance will be given on future workshops, and reference implementations are already available.

EJP Experts: **Rajaram Kaliyaperumal, Marco Roos and Mark Wilkinson**

Funded by the European Union GA n°825575

Data Mapping and Querying

- *What are querying languages (and its options)?*
- *What are mapping languages?*

1. Query languages allow to retrieve information from registries without the need to download and share the whole database. Options depend on your data format. For example, semantic data formats can be queried with SPARQL protocol, and relational databases can be queried by SQL.
2. EJP is defining a tool for specific queries via the Virtual Platform.
3. Mapping languages allow you to map from one data format to another. For example, to transform non-RDF (e.g., CSV) into RDF based formats you could use RML and YARRRML.

EJP RD Experts: **Anthony Brookes, David Reinert, Mark Wilkinson, Núria Queralt and Rajaram Kaliyaperumal**

Funded by the European Union GA n°825575

Rules and Standards

- *What are the **minimal standards** and tools recommended for ERNs to achieve **FAIR** and to be interoperable with the **VP** and other ERNs?*

1. As a starting point, register your registry and data elements at the EU RD Platform (ERDRI), and work with the stewards and EJP RD experts on implementing the **CDE semantic model**, the **Metadata model (FAIR data point)**, and the **Query Builder API**.

2. The **EJP deliverable 12.1** compiles standards and tools. The Virtual Platform specifications will give more clarity on those selected with the main goal to enhance the overall interoperability and FAIRness (planned for March 2021).

EJP RD Experts: **Anthony Brookes, Marco Roos, Marc Hanauer**

Funded by the European Union GA n°825575

Virtual Platform

- *What are the benefits of using the VP?*
- *What are the VP requirements?*

1. The VP enables federated data access, increases the visibility of your resources, and when permitted (access grant, given consent), allows cross-querying of ERNs registries and other databases in a GDPR-compliant manner.
2. The detailed VP specifications are being defined and will be shared at the end of March 2021.

EJP RD Expert: **Karl Kreiner, Dieter Hayn, Günter Schreier**

Funded by the European Union GA n°825575

Feedback from other ERNs

- *Where are other ERNs in the FAIRification process and which approaches were taken?*

1. Very heterogeneous, some ERNs are already collecting FAIR data, whilst others are defining consent forms and dealing with legal issues.
2. You can ask the stewards to share the inventory of implementations to see what other ERNs did so far.
3. EJP RD and ERICA will setup a community sharing platform for ERNs to exchange experience directly.

EJP RD Experts**: FAIRification Stewards, Franz Schaefer**

Funded by the European Union GA n°825575

Legal Issues and Consent

- *Consortium, data sharing and processing agreement; hospital-patient agreement; Country specific legal issues;*
- *EUPID policy and licenses; What is allowed; Needs mention in ICF?*

1. Project ERICA, AREB and exchange process between ERNs. See https://webgate.ec.europa.eu/ern for templates.
2. EUPID Services for patient registration and pseudonymisation, data transition and linkage, and for re-identification, including EJP VP use, are consented. Please contact EUPID for extra EUPID Services documents.

EJP RD Experts: **Annalisa Landi and Viviana Giannuzzi (AREB + ICF wg), contact@eupid.eu (EUPID)**

Funded by the European Union GA n°825575

ERICA: European Rare DIsease Research Coordination & Support Action

2.1 Coordination and Support of ERN registry activities

2.1.1 Harmonization of data protection and access policies for secondary use

2.1.2 Development and promotion of efficient RD data collection and management strategies for ERN registries

2.1.3 Promotion of registry FAIRness

2.1.4 Facilitation of inter-ERN registry research collaborations

Task 2.2 Facilitation of biobanking

Task 2.3 Contribution to development and utilization of EJP RD virtual platform

EJP RD

Funded by the European Union
GA n°825575

WG 'Data Protection' (ERICA 2.1.1)

Members:

Experts from ERNs, EJP RD, ECRIN, EURORDIS, European Commission, National Authorities

Tasks:

To develop GDPR compliant documents for

- Informed consent
- data processing
- data access policies
- data transfer

Funded by the European Union
GA n°825575

WG 'Data Collection and Management' (ERICA 2.1.2)

Members:

- ERN registry coordinators/managers; ePAGs
- ECRIN
- SOLVE-RD diagnostic data management experts

Methods:

- Online surveys
- ‘Best practice’ studies
- Cost efficiency analyses

Tasks:

- Develop tailored strategies and processes for efficient data collection, quality control, processing, accessing
- Promote, monitor and report usage of best working models
- Adopt Solve-RD working model for diagnostic data sharing
- Engage with patient representatives on collection of outcome measures

EJP RD

Funded by the European Union GA n°825575

Training Needs

- **Training material** will be offered by Pillar 3 (e.g. Orpha nomenclature, ORDO)
- **'Coffee room' webinars** for ERN registry developers for theoretical questions
- **Hands-on workshops** between registries and EJP RD experts for all practical questions
  - **Developer Q&As** will be published in StackOverFlow

Funded by the European Union
GA n°825575

Acknowledgments

ERNs, EJP RD Experts, FAIRification Stewards, JRC

EUROPEAN JOINT PROGRAMME
RARE DISEASES