FAQs

These topics were taken from the questions asked by ERN representatives to the FAIRification stewards' team. The answers are from different experts in the EJP RD.

1. **FAIRification training**
    - The EJP RD coffee rounds and technical workshops can be found at https://ejprd.sharepoint.com/sites/EJPRD-ERN-EVENTS *.
    - ELIXIR's Training e-Support System (TeSS): https://tess.elixir-europe.org/
2. **CDE Minimal Dataset**
    - **Are all elements mandatory**?
        - The inclusion of the CDEs in the registries is mandatory as per the EC ERN grant demands.
    - **Can you better define the CDE elements?**
        - Further description of the CDEs can be found on the CDE Semantic Model GitHub page and by request to JRC. EJP RD and JRC will jointly develop a descriptive document with more detailed definitions of the CDEs.
    - **What is the added value of using them?**
        - Implementing the CDEs is the first step towards interoperability and is an essential prerequisite for further research based on registry data.
3. **What to do with disease specific data elements?**
    - Make a note of the disease-specific elements and bring them to your FAIRification steward to be discussed within EJP-RD
    - The EJP RD is building a list of Domain specific CDEs, this list along with semantic models will be released soon in collaboration with the JRC.
4. **Ontologies and Vocabularies**
    - **Genetic Diagnosis: Should everyone use HGNC, HGVS and OMIM? What to use it for**?
        - HGNC - is used for human gene names
        - HGVS - is used for identifying variants
        - Orpha codes - are used to unambiguously denote rare diseases
        - OMIM - codes more common genetic disorders
        - Using these coding systems in your dataset should be a priority.
    - **Clinical diagnosis: Can we only use Orpha codes?**
        - You may use all vocabularies mentioned in the CDEs pdf depending on which data you collect.
        - Workshop on Orpha codes *
        - Coffee session on Orpha codes *
    - **Can we use own field classification not in Orphacode?**
        - Additions to Orpha codes can be requested at disease.orphanet@inserm.fr, and mappings to Orpha codes at http://www.rd-code.eu.
5. **CDE Semantic Model - What are the added values of implementing the CDE Semantic Model?**

- 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.
- The model can be mapped and used in combination with other data models such as CDISC/ODM and ODHSI/OMOP. Implementation workshops will be arranged.
- It increases the potential to link resources (EHR – EDC Systems) easing the data collection process.
- Workshops on the CDE semantic model Part 1 and Part 2 *

6. **EJP RD Metadata Semantic Model - What is it, and what are the added values?**
    - Metadata is "data about data". The benefit of using the metadata model is that it makes your registry findable and accessible for computers.
    - It can be used by the EJP RD Virtual Platform and application programming interface (API) specifications such as the FAIR Data Point and Beacon.
    - Workshop on the EJP RD semantic metadata model *
7. **Data Formats**
    - **Which data representation format to use?**
        - The cleaner and clearer the data is (one record per row, explicit column names, no empty cells, use of the same date formats, etc), the easier it is to transform to a FAIR format.
        - Workshop on data formats *
    - **Which formats are ERNs currently using?**
        - ERNs are currently using Excel tables, JSON, CSV, and ontologized data formats (RDF) for their registries.
    - **How can Phenopackets be used in registries? Why do I need Phenopackets if my registry is already FAIR?**
        - Phenopackets is a standard exchange format for phenotype data. If your registry is FAIR, it is easier to add extra exchange formats such as Phenopackets.
        - Workshop on Phenopackets *
8. **FAIR Data Point - What is it? And how does it work?**
    - It provides a way for users and computers to read information (i.e., metadata) on how to Find, Access and Reuse your registry data.
    - Depending on access mechanisms and license conditions, the actual data can be accessed, queried and/or downloaded.
    - Reference implementations are already available.
    - Workshop on the FAIR Data Point
9. **Data Mapping and Querying**
    - **What are querying languages (and its options)?**
        - Query languages allow one to retrieve information from registries without the need to download and share the whole database. There are several query languages to choose from, but your 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. The EJP RD is defining a tool for specific queries via the Virtual Platform based on JSON and REST API specifications.
    - **What are mapping languages?**

        - 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.
    - Workshop on data formats and mapping languages
10. **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?**
    - 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.
    - The appendix to the EJP deliverable 12.1 compiles standards and tools for FAIRification of data.
    - The Virtual Platform specifications gives more clarity on a selection of standards for the first version with the main goal to enhance the overall interoperability and FAIRness.
    - More information can be found on this page for FAIR guidance for FAIR consortia.
    - Coffee session on Practical requirements for Registry FAIRification
11. **Virtual Platform**
    - **What are the benefits of using the VP?**
        - The VP enables federated data access, increases the visibility of your resources, and when permitted (access granted, consent given), allows cross-querying of ERN registries and other databases in a GDPR-compliant manner.
    - **What are the VP requirements?**
        - The Virtual Platform specifications give more clarity on a selection of standards for the first version with the main goal to enhance the overall interoperability and FAIRness of resources.
    - Workshop on using the Virtual Platform
    - Training sessions on how to use the VP will be conducted in 2023
12. **Feedback from other ERNs - Where are other ERNs in the FAIRification process and which approaches were taken?**
    - Very heterogeneous, some ERNs are already collecting FAIR data, whilst others are defining consent forms and dealing with legal issues.
    - You can ask the stewards to share the inventory of implementations to see what other ERNs did so far.
    - EJP RD and ERICA will setup a community sharing platform for ERNs to exchange experience directly.
13. **Legal Issues and Consent**
    - Related documents can be found on the ERICA website

* Access to these pages will need to be requested via this form