Acknowledging rare disease patients & patient representatives, participants and teachers of all rare disease ‘Bring Your Own Data’ workshops, the members of the rare disease cross-project FAIR data task force, Annika Jacobsen, Claudio Carta, David van Enckevort, Mark Thompson, Rajaram Kaliyaperumal, Kees Burger, Luiz Bonino, Mark Wilkinson, Frederieke Ehrhart, Rachel Thompson, Lieze Thielemans, Claire Shovlin, Leo Schultze Kool, Marina Mordento, Luca Sangiorgi, Domenica Taruscio, Gary Saunders, Ana Rath, Marc Hanauer, RD-Connect and its members, the RDs GO FAIR seed group, RD-Connect community, ELIXIR-NL/EXCELERATE, BBMRI-ERIC/ADOPT/NL, VWData, ODEX4All, FAIR-dICT, European Joint Program Rare Diseases
Address international goals of the rare disease community and accelerate efficiency of global rare disease data analysis

Respectful = use my data, make them as usable as possible, inform me, involve me!

All people living with a rare disease receive an accurate diagnosis, care, and available therapy within one year of coming to medical attention

10 objectives to Foster and strategically oversee transition to FAIR principles by the rare disease community

- towards collecting a critical mass of FAIR data
- Enable support, define sustainable FAIRification service strategy
- Ensure rare disease patients and patient representatives are engaged
- Ensure FAIR sharing is respectful and responsible towards patients
- Stimulate FAIRness assessment of rare disease resources
- Ensure standards are identified, aligned, extended, not reinvented
- Identify overlap/duplication of effort towards implementing FAIR
- Foster examples of answering key questions beyond current capabilities
- Ensure analytics are capable of exploiting FAIR data
- Ensure RD analytical tools are available through FAIR repositories

Position in middle of community and infrastructures

Lightweight organisation liaising with community representatives and training & building programs

Build on & stimulate train and build networks (GO TRAIN, GO BUILD, GA4GH, ELIXIR, BBMRI, EJPRD, etcetera)

All people living with a rare disease receive an accurate diagnosis, care, and available therapy within one year of coming to medical attention

Respectful = use my data, make them as usable as possible, inform me, involve me!

ERNs

Patient

ELIXIR

RD FAIR data TF

EJPRD

ELSI experts

GO FAIR

NIH
Rare disease research

RARE DISEASES

1:17 people
30 million in Europe

7% OF THE POPULATION ARE AFFECTED BY RARE DISEASES

THE EU CLASSES A DISEASE AS ‘RARE’ WHEN LESS THAN 1 IN 2000 SUFFER

OVER 7000 DISEASES BIOSAMPLES, DISEASE & PATIENT INFO, OMICS, GENOTYPE-PHENOTYPE
Patients demand respectful use of data

- Use our data
- Make our data as usable as possible
- Inform us
- Involve us
Enable all people living with a rare disease to receive an accurate diagnosis, care, and available therapy within one year of coming to medical attention

... 1000 new therapies for rare diseases will be approved...

Vision International Rare Diseases Research Consortium (IRDiRC)
European Reference Networks of rare disease expert centres

1. ERN BOND
   European Reference Network on Rare Bone Disorders

2. ERN CRANIO
   European Reference Network on Rare craniofacial anomalies and ENT disorders

3. Endo-ERN
   European Reference Network on Rare Endocrine Conditions

4. ERN EpICARE
   European Reference Network on Rare and Complex Epilepsies

5. ERKNet
   European Rare Kidney Diseases Reference Network

6. ERN RND
   European Reference Network on Rare Neurological Diseases

7. ERNICA
   European Reference Network on Rare inherited and congenital anomalies

8. ERN LUNG
   European Reference Network on Rare Respiratory Diseases

9. ERN Skin
   European Reference Network on Rare and Undiagnosed Skin Disorders

10. ERN EURACAN
    European Reference Network on Rare Adult Cancers (solid tumours)

11. ERN EuroBloodNet
    European Reference Network on Rare Haematological Diseases

12. ERN EURO-NMD
    European Reference Network for Rare Neuromuscular Diseases

13. ERN EYE
    European Reference Network on Rare Eye Diseases

14. ERN GENTURIS
    European Reference Network on Genetic Tumour Risk Syndromes

15. ERN GUARD-HEART
    European Reference Network on Uncommon And Rare Diseases of the HEART

16. ERN ITHACA
    European Reference Network on Rare Congenital Malformations and Rare Intellectual Disability

17. MetabERN
    European Reference Network for Rare Hereditary Metabolic Disorders

18. ERN PaedCan
    European Reference Network for Paediatric Cancer (haemat-o-oncology)

19. ERN RARE-LIVER
    European Reference Network on Rare Hepatological Diseases

20. ERN ReCONNET
    Rare Connective Tissue and Musculoskeletal Diseases Network

21. ERN RITA
    Rare Immunodeficiency, Autoinflammatory and Autoimmune Diseases Network

22. ERN TRANSPLANT-CHILD
    European Reference Network on Transplantation in Children

23. VASCERN
    European Reference Network on Rare Multisystemic Vascular Diseases

24. ERN euroGEN
    European Reference Network on Rare and Complex Urogenital Diseases
Under the networks: unrefined, siloed data

1. ERN BOND
   European Reference Network on Rare Bone Disorders
2. ERN CRANIO
   European Reference Network on Rare craniofacial anomalies and ENT disorders
3. Endo-ERN
   European Reference Network on Rare Endocrine Conditions
4. ERN EpiCARE
   European Reference Network on Rare and Complex Epilepsies
5. ERKNet
   European Rare Kidney Diseases Reference Network
6. ERN RND
   European Reference Network on Rare Neurological Diseases
7. ERNICA
   European Reference Network on Rare inherited and congenital anomalies
8. ERN LUNG
   European Reference Network on Rare Respiratory Diseases
9. ERN Skin
   European Reference Network on Rare and Undiagnosed Skin Disorders
10. ERN EURACAN
    European Reference Network on Rare Adult Cancers (solid tumours)
11. ERN EuroBloodNet
    European Reference Network on Rare Haematological Diseases
12. ERN EURO-NMD
    European Reference Network for Rare Neuromuscular Diseases
13. ERN EYE
    European Reference Network on Rare Eye Diseases
14. ERN GENTURIS
    European Reference Network on Genetic Tumour Risk Syndromes
15. ERN GUARD-HEART
    European Reference Network on Uncommon And Rare Diseases of the HEART
16. ERN ITHACA
    European Reference Network on Rare Congenital Malformations and Rare Intellectual Disability
17. MetabERN
    European Reference Network for Rare Hereditary Metabolic Disorders
18. ERN PaedCan
    European Reference Network for Paediatric Cancer (haemat-oncology)
19. ERN RARE-LIVER
    European Reference Network on Rare Hepatological Diseases
20. ERN ReCONNET
    Rare Connective Tissue and Musculoskeletal Diseases Network
21. ERN RITA
    Rare Immunodeficiency, Autoinflammatory and Autoimmune Diseases Network
22. ERN TRANSPLANT-CHILD
    European Reference Network on Transplantation in Children
23. VASCERN
    European Reference Network on Rare Multisystemic Vascular Diseases
24. ERN eUROGEN
    European Reference Network on Rare and Complex Urogenital Diseases and Conditions
Silos, silos, silos
Silos, silos, silos
Absolute need to enable analysis across resources
Usability of data for efficient rare disease care and research must be brought to higher, global levels
European Open Science Cloud

International activities relevant for rare diseases
(BBMRI-ERIC, GA4GH, IRDiRC, Monarch, OpenPHACTS, Pistoia, GO FAIR, etc.)

European Open Science Cloud

European Reference Networks for rare diseases

National, local, and community activities relevant for rare diseases
(medical institutes, patient organisations, research labs, etc.)

ELIXIR rare disease community activities

Robust infrastructure at EU level for and with the rare disease community

VWData, Data2Person

EOSCLife

EJP RD

RDs GO FAIR

RD Connect Community

European Reference Networks
Disrupting infrastructure for rare diseases
European Joint Program COFUND Rare Diseases

- 5 year
- 100M Euro program (55M EU funding)
- >80 partners, including all ERNs, BBMRI-ERIC, ELIXIR
- 17M Euro for FAIR infrastructure
- Driving project for GA4GH
Target: FAIR-based Virtual Platform
A mixed infrastructure of centralised and federated resources

A powerful substrate for translational research

- **Centralized services for collections** (resource-level)
  - Sample, biobanks, registries, infrastructures and tools catalogue
  - Analysis platform for omics data
  - Curated rare disease information and data

- **Federated services for data elements** (record-level)
  - FAIR ‘at source’
  - Data, patients, and samples - linkable and discoverable
  - Consents and data use conditions

- **For humans and computers**
  - Humans explore
  - Computers find, access and perform reproducible analyses
Centralised services - year 1

• Chart rare disease needs and available functionalities
• Normalise metadata for data resources
• Release first proof-of-principle Virtual Platform: catalogues Findable and Accessible
• Integrate privacy preserving record linkage with other access solutions
• Adapt usability of RD-Connect phenotypic data collections to stakeholder needs
Federated services - year 1

- Compare and align core interoperability standards of relevance
- Chart landscape of existing stakeholders, standards, tools, working groups
- Start multi-team software development towards FAIR ecosystem for RDs
- Organise practical FAIRification support to deliver first EJPRD FAIR stations
- Today: connect to German PHT & MI
European Open Science Cloud

International activities relevant for rare diseases
(BBMRI-ERIC, GA4GH, IRDiRC, Monarch, OpenPHACTS, Pistoia, GO FAIR, etc.)

European Reference Networks for rare diseases

National, local, and community activities relevant for rare diseases
(medical institutes, patient organisations, research labs, etc.)

ELIXIR rare disease community activities

Robust infrastructure at EU level for and with the rare disease community

VWData, Data2Person
Selection of previous/ongoing steps towards rare disease FAIR vision

- Orphanet (ELIXIR core resource)
- RD-Connect portal & genome-phenome analysis platform
- Rare disease sample catalogue and registry/biobank finder
- USA: Monarch initiative
- HPO, ORDO, DUO, Orphacodes basis for metadata (in addition to global health metadata standards)
Selection of previous/ongoing steps towards rare disease FAIR vision

- FAIRification pipeline & software
  - FAIR data access point with consent checking
  - ADA-M (Automatable Discovery & Access Matrix), consent codes
  - Privacy preserving record linkage (EUPID)
- Access automation
- NIH common data elements
- Rome summer school for rare disease data managers with data linkage Bring Your Own Data workshop
Selection of previous/ongoing steps towards rare disease FAIR vision

- FAIRification pipeline & software
- Access automation
  - FAIR data access point with consent checking
  - ADA-M, consent codes
  - Privacy preserving record linkage (EUPID)
- NIH common data elements
- Rome summer school for rare disease data managers with data linkage Bring Your Own Data workshop
**FAIR data stewardship services: tools and people**

(DTL, UPM, ISS, LUMC, ELIXIR-NL/UK/SL, BBMRI, RD-Connect, FAIR-dICT, ...)

<table>
<thead>
<tr>
<th>Plan</th>
<th>Create</th>
<th>Publish</th>
<th>Find</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Stewardship Wizard</td>
<td>FAIRIFIER</td>
<td>FAIR DATA POINT</td>
<td>FAIR DATA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Courses</th>
<th>Bring Your Own Data Workshop</th>
<th>FAIR Hackathons</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Data linkage plan as a service</th>
<th>FAIR expert task forces</th>
<th>Project planning with stakeholders</th>
</tr>
</thead>
</table>

**Courses**

- Project planning with stakeholders
- Data linkage plan as a service
- FAIR expert task forces
- FAIR Hackathons
FAIR data point: metadata model driven REST API

Schema to define how to find and access data sets

Complements existing infrastructure

Ontologised & linkable data at record level

REST interface: use schema for controlled access
Rare disease data linkage plan

Knowledge exchange

Your knowledge of your data

your DATA

Source data steward(s)
FAIR data steward(s)
Domain expert(s)

Your knowledge

your DATA

Cost: ~4-8PM
~40-80K Euro
Generalised FAIRification pipeline

1. Define driving question(s)
2. Assess data source, plan, scope
3. Define meaning of data for machines
4. Apply to make data machine readable
5. Add information for finding and accessing data
6. Identify location to host the data
7. Apply to make data machine readable
8. Add information for finding and accessing data
9. Identify location to host the data
10. Apply interfaces to build into data analysis tools

Create model from ontologies (HPO, ORDO, OBIB, ...)
Apply metadata standards
Test
Analysis
Running FAIRification pilots

- **Osteogenesis imperfecta**
  (EXCELERATE+BBMRI ADOPT, BOND ERN): Marina Mordenti/Luca Sangiorgi (Rizzoli hospital Bologna/BOND ERN)

- **Vascular Anomalies**
  (RadboudUMC funds) Leo Schultze Kool/Karlijn Groenen (RadboudUMC), Claire Shovlin/LiezeThielemans (Imperial College London) (VASCERN), Castor EDC

- **Rare muscular dystrophies**
  (EXCELERATE+RD-Connect) LUMC Neurology department+Library+Advanced Data Management group & Rachel Thompson (Newcastle University)

- **Rett Syndrome variants**
  (ELIXIR) Rett organisations and Frederieke Ehrhart (UM)
Enabling information retrieval and analysis across RD resources
Used for rare diseases and plants!

Driving user questions

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.

Biobank/registry info

(Biobank & Registry Finder)

Demonstrator UI
Current challenge: scaling up
Reuse human & machine readable knowledge

11 May 2019
Can we identify commonalities and complementarities between developing the FAIR Deutche Bahn and the EJPRD mixed centralised/federated infrastructure?

How do we organise possible collaboration?

Do we see bottlenecks, i.e. a role for GO FAIR?

Can we beat NS+DB at developing the FAIR hi-speed train between Amsterdam and Berlin?
Thank you

**FAIR data engineering team**
Rajaram Kaliyaperumal, Kees Burger, Nuno Nunes, Shamanou van Leeuwen, Mark Thompson, **Luiz Bonino**, Mark Wilkinson, Michel Dumontier

**RDs GO FAIR seed group**
Claudio Carta, Ronald Cornet, David van Enckevort, Marc Hanauer, Gulcin Gumus, Ian Harrow, Victoria Hedley, Annika Jacobsen, Dipak Kalra, Deborah Mascalzoni, Veronica Popa, Ana Rath, Marco Roos, Yaffa Rubinstein, Gary Saunders, Rachel Thompson, Mark Wilkinson, **Mark Thompson, Morris Swertz, Domenica Taruscio, Peter-Bram ’t Hoen**

**Training material**
ELIXIR, FAIR-dICT, DTL, Claudio Carta, Mascha Jansen, Celia van Gelder, Erik Schultes, Albert Mons

**Use case collaborators** Rachel Thompson, Libby Wood, Claudio Carta, Marco Crimi, Estrella Gomes, **Marina Mordenti**, Luca Sangiorgi, David van Enckevort, Fleur Kelpin, Claire Showlin, Lieze Thielemans, Hevas, Leo Schulzke Kool, Caroline van den Bosch, Carine van Vleuten, Erik Niks, VSOP, **Annika Jacobsen**, Freddie Ehrhart, Andra Waagmeester; Heimo Muller, Robert Reifs; Pedro Sernadella, Jose Oliveira; Marc Hanauer, Ana Rath; Roxana Merino; Matthias Brochauens; Developers of Castor, RDRF, OSSE, MolGenis; participants Bring Your Own data workshops and patient representatives, rare disease linked data and ontology task force

**ELIXIR**: Chris Evelo and team (Maastricht), Carole Goble and team (Manchester), Ivo Gut and team (CNAG), Helen Parkinson and team (EBI), Olivier Poch and team (Strasbourg), Orphanet, Brane Leskosek and team (Ljubljana)

**The BioSemantics group and the Human Genetics Department LUMC**

https://rd-connect.eu/about-rd-connect/become-member/
Conclusion: FAIR challenges for building FAIR (Haupt)Bahnhöfer for rare diseases

- **Findability**
  *First challenge*
  Orphanet and ‘Finder’ services on RD-Connect platform, bioschemas, GA4GH beacons

- **Accessibility**
  *Unavoidable challenge*
  ‘ELSI’ + Technology experts to automate consent checking

- **Interoperability**
  *Critical challenge*
  Interdisciplinary collaboration, Ontologies & linked data, mapping to domain standards

- **Reusability**
  *Highest challenge*
  Requires F+A+I, collaboration with rare disease data quality guideline developers
Foster the adoption of FAIR data principles by the rare disease community

Ensure that rare disease patients and patient representatives are engaged.

Ensure that FAIR sharing in the rare disease community is respectful and responsible towards rare disease patients.

https://rd-connect.eu/about-rd-connect/become-member/

fair-rd-info@elixir-europe.org
FAIRification: Interdisciplinary learning experience

- Disease experts (MDs, patient representatives)
- “Local” rare disease data stewards
- Rare disease FAIR data stewards
- Registry software providers
- FAIR software engineers
- Cross-project rare disease FAIR data task force
Annual ‘Rome’ summer school & Bring Your Own Data workshop for rare disease data managers

- Data owners bring their data
- FAIRification by
  - data linking experts
  - resource data experts
  - domain experts
    (MDs, patients & patient representatives)
- Discussion on FAIR project management
BYOD Hype cycle

Make all my data FAIR in two days!
BYOD Hype cycle

Make all my data FAIR in two days!

End of day 1: confusion

Make all my data FAIR in two days!
Make all my data FAIR in two days!

End of day 1: confusion

Day 2: discuss FAIRification project planning

BYOD Hype cycle
Test activities - year 1
Ride X-omics train

• Chart stakeholder X-omics data and research questions
• Add rare disease specific pathways to WikiPathways
• Link FAIRified data to knowledge resources for analysis, such as for drug repurposing
Example: drug prioritization for Huntington’s Disease

Eleni Mina, Kristina Hettne, Eelke van der Horst, Katerina Nosikova, Elizaveta Besedina, Katy Wlstencroft, Peter-Bram ’t Hoen, Marco Roos

Select genes by phenotype matching in Monarch

Filter on feasibility for treating HD

Prioritized drug compounds

Select drug compounds in OpenPHACTS
Pain distribution preparing for multi-source analysis

Experiment/survey/...

Data publication

Integrative analysis

Application

1 month

N x 6 months
FAIR Pain distribution preparing for multi-source analysis

- Experiment/survey/
- Data publication
- Integrative analysis
- Application

6 months
N x 1 month

5 months more research
start 5 months earlier
Siloes

- Biobank
- Registry
- Biological study
- Clinical trial
- Natural history
European Joint Program Rare diseases

**Vision**

**NOW**
- ERN research
- Computational research

**END of EJPRD**
- Enhanced ERN research

Funded by the European Union

[GA n°825575]
Strategy

The fuel

Disease research

Enhanced research

VP development

Developers learning from Disease experts

The mechanics

Disease experts learning from computational experts
Timeline for year 1

Preparatory phase (Jan-May)
(Stakeholder surveys and interviews, monthly TCs)

Annual Retreat (May 22-24, Rome)
(Surveys evaluation, fine-grained roadmap, priorities established)

Building phase = Proof-of-principle (Jun-Dec)
(follow-up in TCs, course adjustments, field testing)