# Applying the FAIR principles to data in a hospital: an interdisciplinary collaboration

Núria Queralt-Rosinach, **Rajaram Kaliyaperumal, César H. Bernabé, Qinqin Long, Simone A. Joosten, Henk Jan van der Wijk, Erik Flikkenschild, Kees Burger, Annika Jacobsen, Barend Mons, and Marco Roos**

Biosemantics Group, LUMC, NL

Open Science FAIR, 22 September 2021

N.QUERALT_ROSINACH@LUMC.NL

Leids Universitair Medisch Centrum
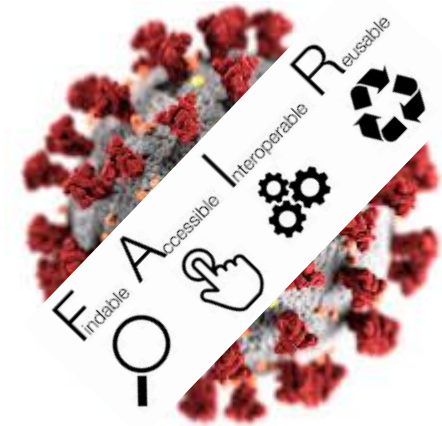
COVID-19 Crisis Triggered Collaboration

Clinical Questions

Questions
- What are the criteria that define the different disease trajectories?
- What are the underlaying mechanistic profiles of the different types of groups?

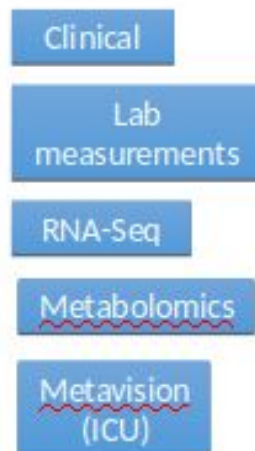**Need to Link Data Across**

Data Management Crisis

Source

- COVID-19 **global** challenge that needs faster interventions
- Clinicians and researchers need to share and collaborate
- But, **observational patient data** is first collected in hospitals

FAIR and Open Science

- Hospital data is **heterogeneous** in nature, description and storage
- **Challenge**: clinical data **interoperability** in and outside hospitals
- FAIR and open science for health data (EJP RD)
- FAIRification is **stakeholders** collaboration

LUMC Data

L U
M C Leids Universitair
          Medisch Centrum

Clinical

Lab measurements

RNA-Seq

Metabolomics

Metavision (ICU)

External Knowledge

FAIR and Open Science in the hospital

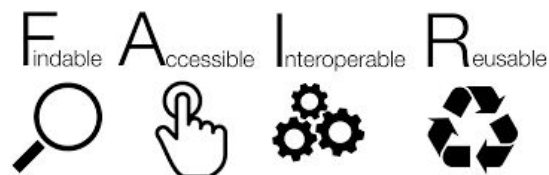**LUMC BEAT-COVID group**

- **Interdisciplinary** stakeholders

Hypothesis

- Use of ontological models for interoperability for machines
- Interdisciplinary collaboration

Goals

- Apply **FAIR** principles *at source* on observational patient data
- Enable **federation** with open knowledge (*visiting data*)

Findable   Accessible   Interoperable   Reusable

Collaborate and Complement

- FAIRification *at source* that **complements** existing hospital systems
- FAIR architecture based on ontological models and Semantic Web
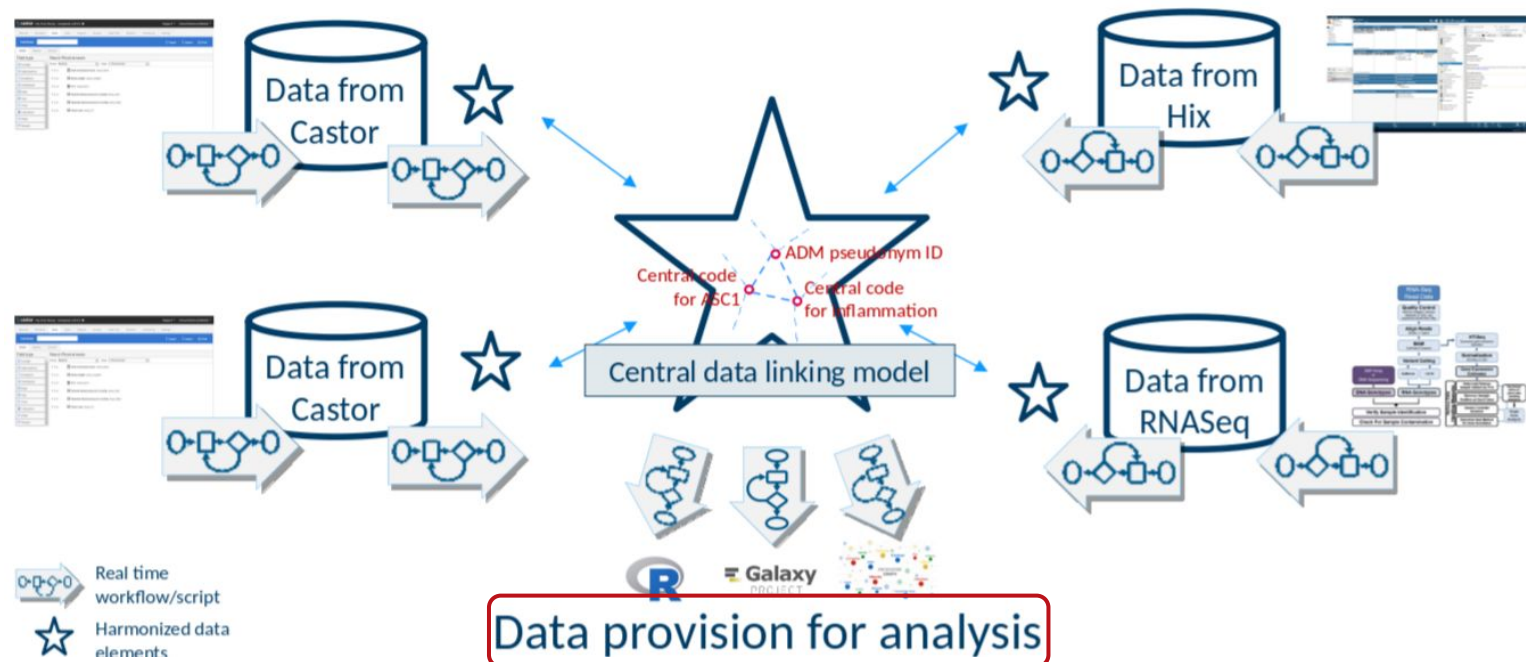
FAIR in parallel and collaborative

- Ontological models driven by **clinical questions**

Questions
- What are the criteria that define the different disease trajectories?
- What are the underlaying mechanistic profiles of the different types of groups?

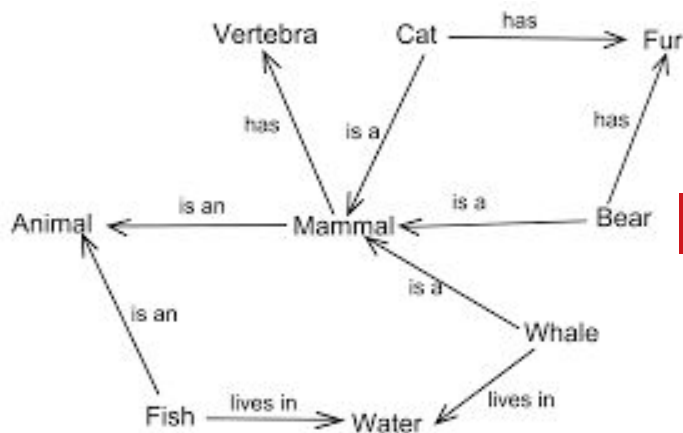# FAIR in parallel and collaborative

- **Ontological models** driven by clinical questions

Data
linking ontological model
(clinical measurements)

Metadata
ontological model
(DCAT2)

FAIR in parallel and collaborative

Improving I in FAIR: *Interoperability*

Improving F,A,R in FAIR: *Visiting data*

Semantic Linking Models

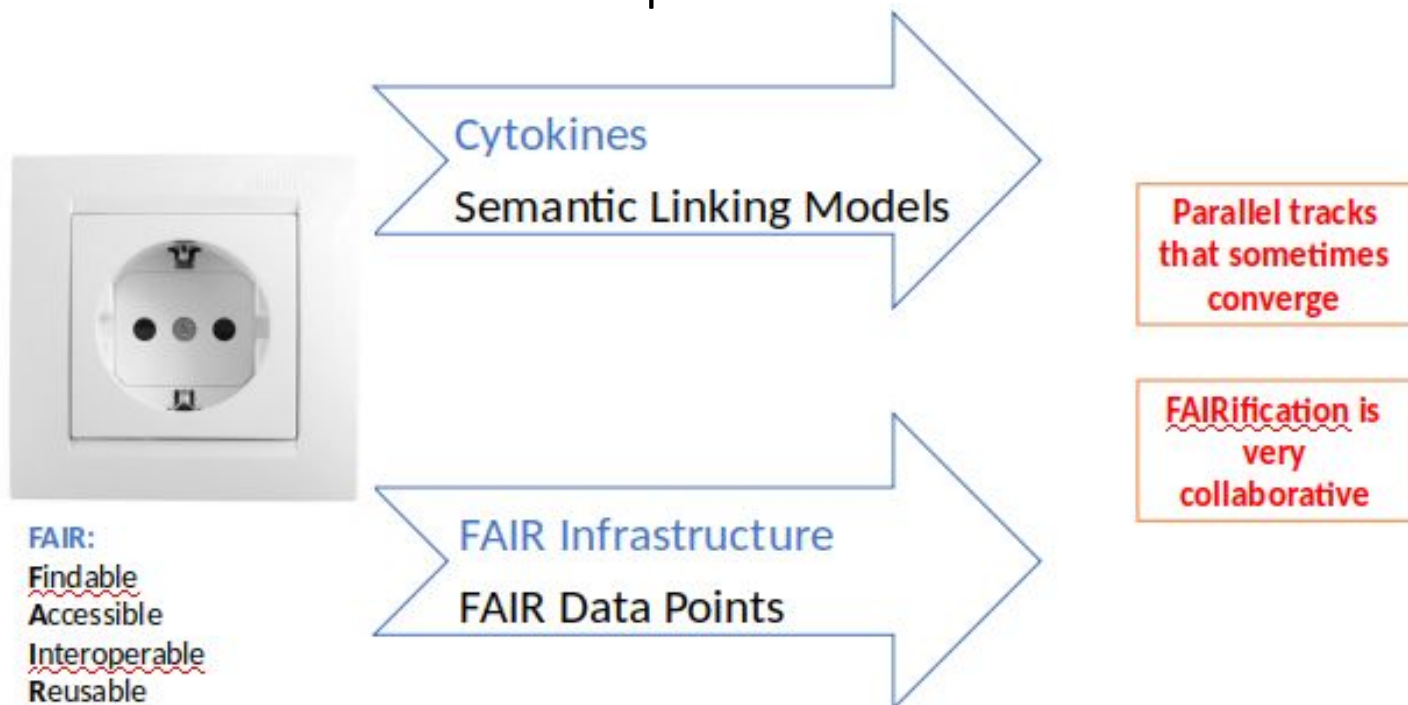(*Linked Data*)

FAIR Data Points (FDPs)

(Publish Metadata)

Community

Clinical data linkable to open knowledge

Machine actionable clinical metadata

FAIR in parallel and collaborative

- **Coordinated** effort that runs in parallel

Difficulties: Social, Funding, Governance (patient data privacy)

FAIR Patient Data

- **FAIRness** evaluation

Metadata **original system**

Metadata original system **complemented** with FDP

Machine Actionable Digital Objects

- **Applications** on top of FAIR patient data
  - Federated queries with Linked Open Data
  - Web APIs
  - Knowledge graph based hypothesis generation tools

```
PREFIX rdf:     <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:    <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd:     <http://www.w3.org/2001/XMLSchema#>
PREFIX owl:     <http://www.w3.org/2002/07/owl#>
PREFIX dct:     <http://purl.org/dc/terms/>
PREFIX obo:     <http://purl.obolibrary.org/obo/>
PREFIX ncit:    <http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#>
PREFIX sio:     <http://semanticscience.org/resource/>
PREFIX prov:    <http://http://http://www.w3.org/ns/prov#>

# Retrieve protein annotation from UniProt
SELECT DISTINCT ?quantitative_trait ?trait_cytokine ?trait_cytokine_label WHERE {

    ?quantitative_trait a obo:IAO_0000109;
            sio:SIO_000628 ?trait_cytokine .

    SERVICE <https://sparql.uniprot.org> {
        ?trait_cytokine rdfs:label ?trait_cytokine_label .
    }
}
```

Discussion and conclusion

- We investigated implementation of FAIR principles in hospital
  - Interoperability for machines (across domains and services)
  - Interdisciplinary collaboration
- We provided a **FAIR Research Data Management** for FAIRification of observational data of hospitalized patients (shared best practices)
- FAIRification adopted in **parallel** for data and metadata and **guided** by users' questions (shared knowledge)

- FAIRification **difficulties**:

  - Interdisciplinary is challenging

  - Data privacy

- We provided COVID-19 **observational patient data** as FAIR digital objects ready to reuse (FAIR evaluation and applications)

A FAIR Research Data Management based on ontological models, interdisciplinary collaboration, open Science, and Semantic Web and FAIR Data Points is providing data infrastructure in the hospital for machine actionable data available for integrative analysis and reusable for applications *in* and *across* open knowledge

Querying FAIR Patient Data Across Hospitals

Distributed Analytics: EU and intercontinental

- FAIR at source
- FDPs: open, secured shared data
- SPARQL queries

- **"count number of patients"**
- **"retrieve LUMC cytokines measurements with protein annotation from UniProt"**

LUMC*
Uganda
Ethiopia
Nigeria
Kenya
Tunisia
Zimbabwe

FAIR Data Points

International Distributed Queries

FDPs publish structured metadata for **machines to interpret how to access**

Acknowledgements

Biosemantics Group, specially:

Marco Roos, Rajaram Kaliyaperumal, César Bernabé, Qinqin Long, Kees Burger, Annika Jacobsen

LUMC BEAT-COVID team, specially:

Sesmu Arbous, Simone Joosten, Hailiang Mei, Jacqueline Janse, Henk Jan van der Wijk, Erik Flikkenschild

Colleagues, mentors, and inspiring scientists, specially:

Mark Wilkinson, Robert Hoehndorf, Carole Goble, Barend Mons

THANK YOU!