

Making data FAIR to boost research on Rare Diseases

Claudio Carta

Istituto Superiore di Sanità, Centro Nazionale Malattie Rare, Roma, Italia

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**

Rare disease data is

- Sparse
- Highly distributed
- Heterogeneous
- Different format
- Poor interoperability
- Sensitive
-





Absolute need to combine

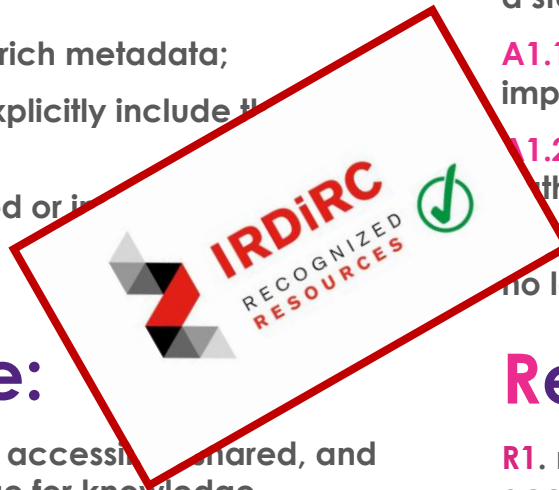
Usability of data for rare disease research and care **must be** brought to higher, global levels

Guiding principles for FAIR data

Human ↔ Computer

Findable:

- F1.** (meta)data are assigned a globally unique and persistent identifier;
- F2.** data are described with rich metadata;
- F3.** metadata clearly and explicitly include the title of the data it describes;
- F4.** (meta)data are registered or indexed in a searchable resource;



Interoperable:

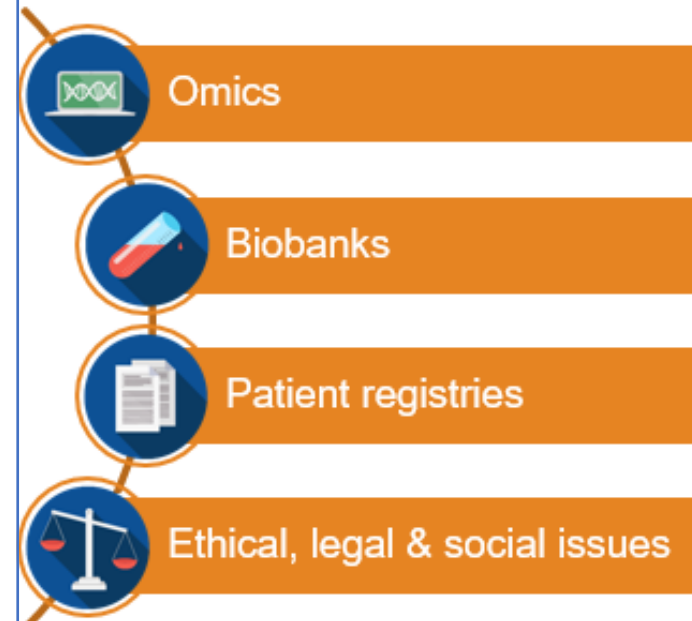
- I1.** (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2.** (meta)data use vocabularies that follow FAIR principles;
- I3.** (meta)data include qualified references to other (meta)data;

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.** (meta)data are accessible, even when the data are no longer available;

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;



What FAIR is not...

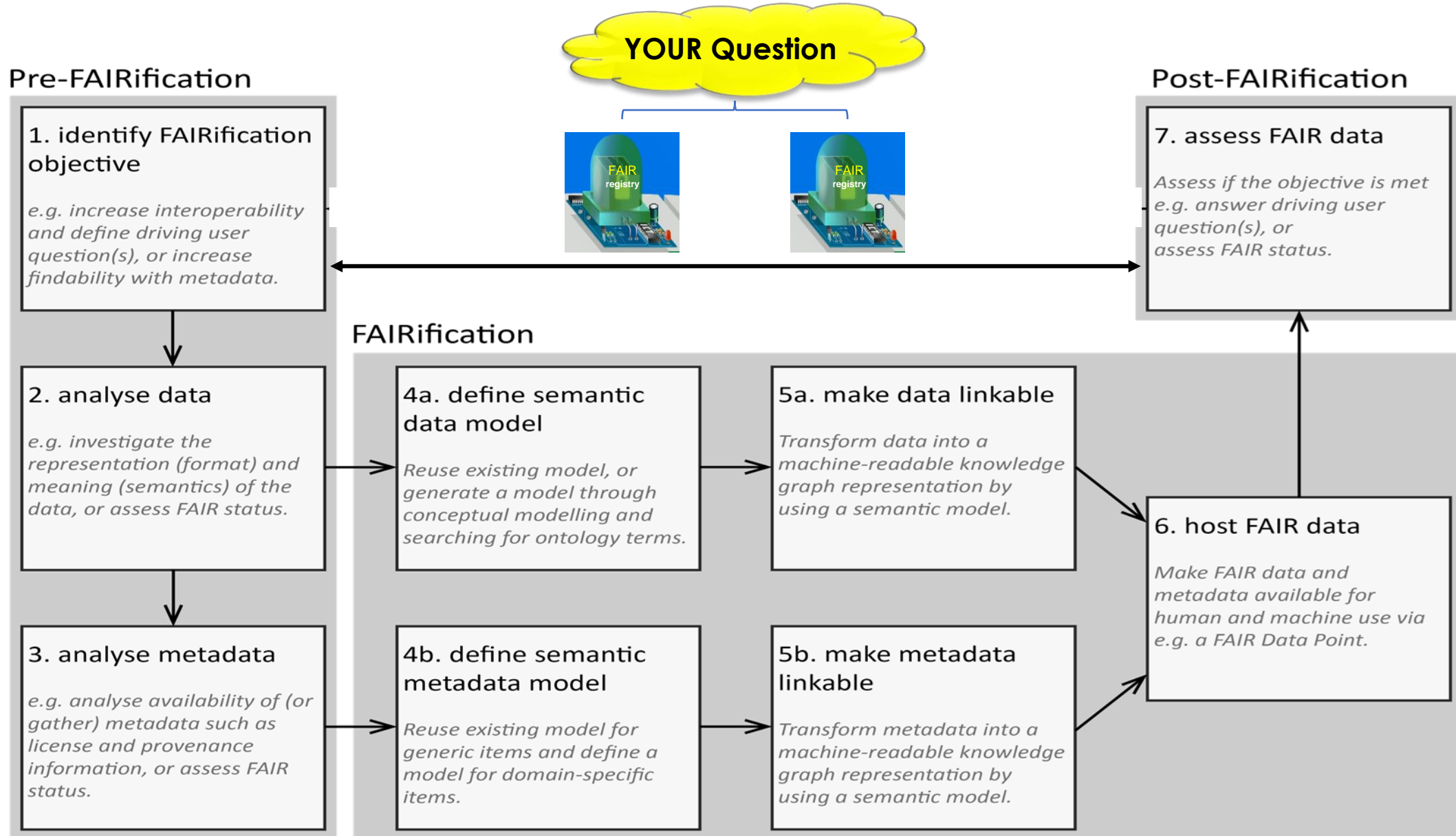
- ✓ FAIR is **not** a standard
- ✓ FAIR is **not** equal to 'Semantic Web'
- ✓ FAIR is **not** equal to 'Open' or 'Free'

Data can be Open but not FAIR

Data can be FAIR but not Open

- ✓ FAIR is **not** for humans only
-

Template process to make rare disease data FAIR

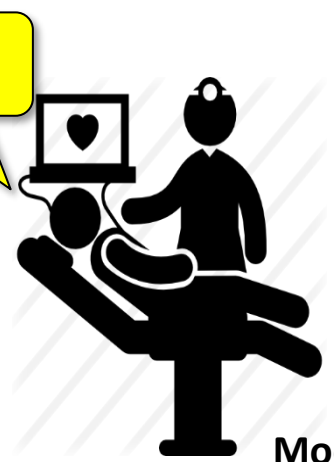
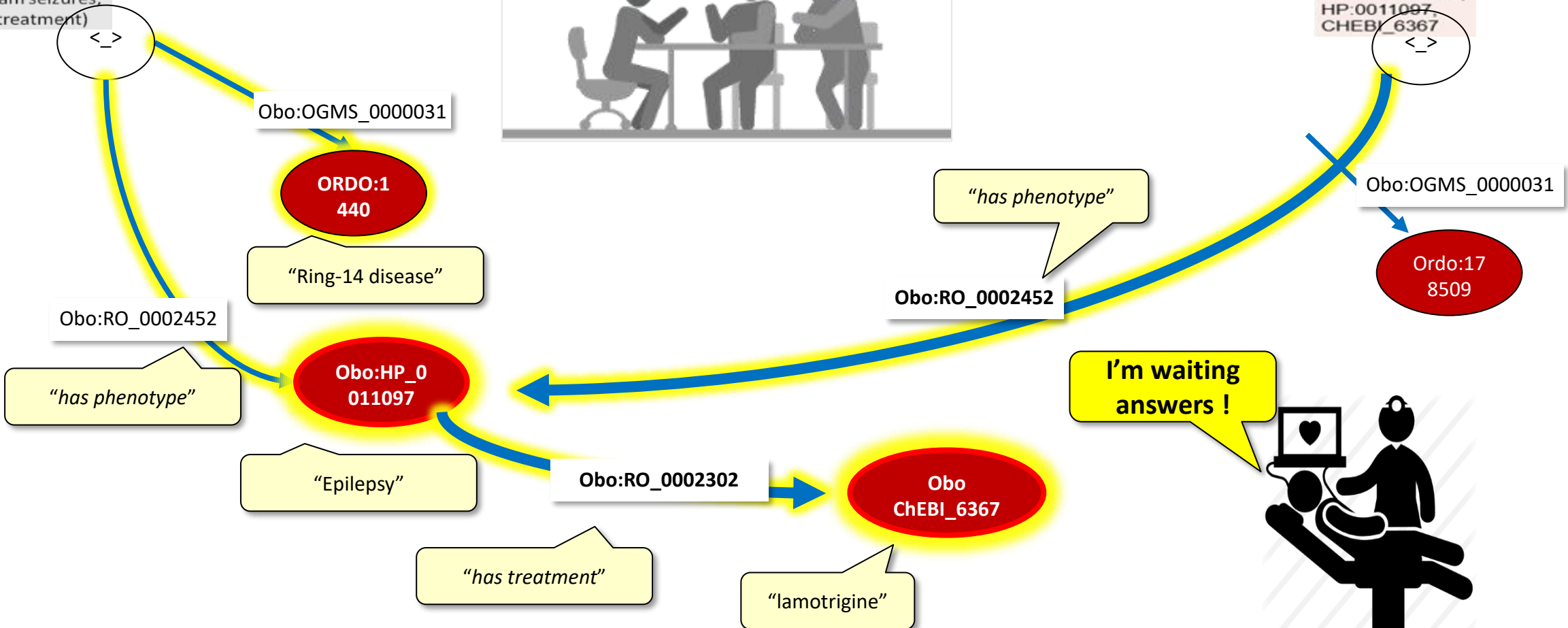


Patient «A»

Krankheit
Ringbildung
Chromosom 14,
Salaam-Anfälle,
(Keine
Behandlung)
Ring-14 disease,
Salaam seizures,
(no treatment)

Patient «B»

பெர்ரி
நோய்க்குறி,
வலிப்பு
தாக்குதல்கள்,
லாமோட்ரைஜி
ன்
Perry syndrome,
Epileptic attacks,
lamotrigine
ORPHA178509,
HP:0011097,
CHEBI_6367



Mock Example

Steps involved in data integration

Experiment/
Survey/
Anamnesis

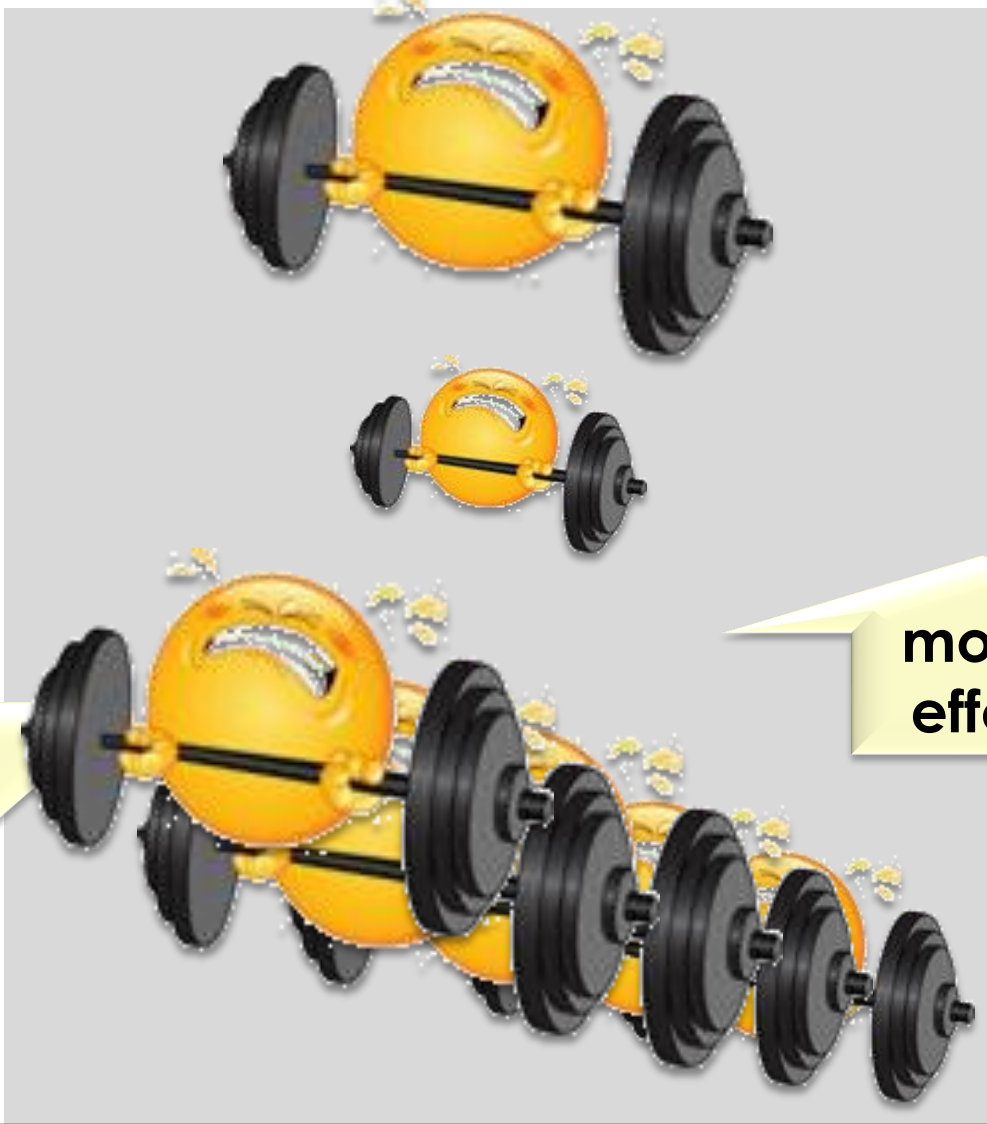
Prepare for analysis

Integrate Data

Analyse

Apply results

Inefficient distribution of effort



Increased efficiency by preparing data for integration

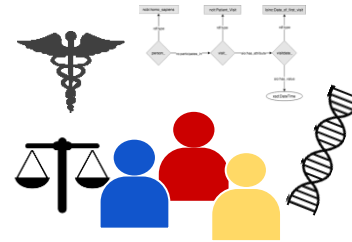


Research & Development of FAIRification

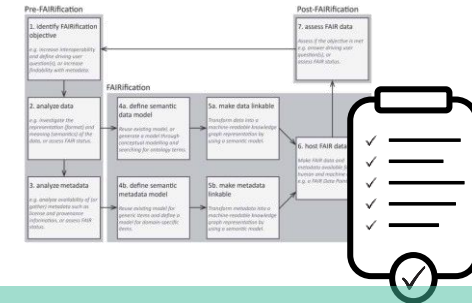
Experience from FAIRification projects and BYODs since 2014



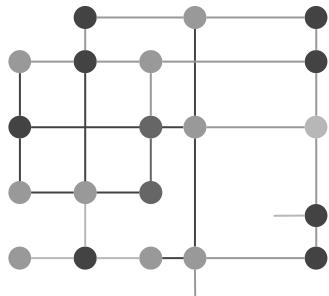
Training



Organising expertise



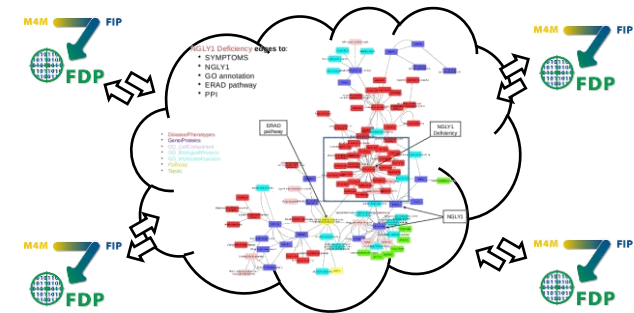
Guidelines & metrics



Align with infrastructure



Tools



FAIR exploitation

THANKS FOR YOUR ATTENTION

Claudio Carta

Istituto Superiore di Sanità, Roma, Italia

Acknowledgments:

Marco Roos, Ronald Cornet, Marc Hanauer, Mark Wilkinson, Rajaram Kaliyaperumal, Annika Jacobsen, Esther van Enckevort, Bruna dos Santos Vieira, Nirupama Benis, Martjin Kersloot, Barend Mons, Erik Schultes, Domenica Taruscio, ELIXIR EU, ELIXIR-IT, ELIXIR-NL, EJP RD partners, LUMC Biosemantics group, EJP RD FAIRification Stewards

This presentation is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nos 857650 (EOSC-Pillar), 824087 (EOSC-Life), and 676559 (ELIXIR-EXCELERATE), and was supported by the Italian Computing and Data Infrastructure (ICDI).

This presentation is released under a CC-BY 4.0 licence. To cite this presentation, please copy and paste:

Pavone, Gina, Lazzeri, Emma, Carta, Claudio, Chiara, Matteo, & Quaglia, Federica. (2022, March 11). Practical course on FAIR Data Stewardship in Life Science. Zenodo. <https://doi.org/10.5281/zenodo.6323375>.

