ADOPT BBMRI-ERIC
gateway for health

ADOPT BBMRI-ERIC

Grant Agreement no. 676550

DELIVERABLE REPORT

Deliverable no	7.6
Deliverable Title	Report of the Pilot Study
Contractual delivery month	36
Responsible Partner	IOR
Author(s)	Manuela Locatelli, Marina Mordenti, Manila Boarini, Annika Jacobsen, Mark Thompson, David van Enckevort, Marco Roos and Luca Sangiorgi

Report of the Pilot Study

Executive Summary

ADOPT Deliverable D7.6 "Report of the Pilot Study" is a report on the pilot study on Osteogenesis Imperfecta (OI), as a model of rare disease. OI has been selected because it a) is well known; b) is well defined at the molecular and biochemical level; c) presents different grades of severity (from early to late onset, from very severe to very attenuated); d) is clinically complex; e) is possibly treatable; f) shares important features with the other RDs and with many more frequent disorders. Moreover, we can rely on the experience of the OI Registry and Biobank established at the Rizzoli Orthopaedic Institute, where biological samples are linked to the disease registry, providing valuable clinical data.

In order to pursue the objective of reinforcing the infrastructures (tools and procedures developed by the RD community and harmonized with BBMRI-ERIC standards and procedures), a User Case on OI has been conducted in collaboration with RD-CONNECT. We involved biobanks from other infrastructures (in addition to BBMRI-ERIC) and networks like RD-Connect and Telethon Network of Genetic Biobanks (TNGB).

Copyright notice

Document log

Issue	Date (yyyy-mm-dd)	Comment	Author/partner
D7.6	2019-01-11	Initial version of the report	Manuela Locatelli, Marina Mordenti, Manila Boarini, Annika Jacobsen, Mark Thompson, David van Enckevort, Marco Roos and Luca Sangiorgi
D7.6_rev	2019-11-06	Revised version according to reviewer's comments	Manuela Locatelli, Manila Boarini

1. Background

Rare Diseases (RDs) are by definition of low prevalence but represent a large, heterogeneous group of diseases affecting more than 30 million European citizens. Translational efforts are essential to make optimal use of resources (biospecimens and data) and are more critical for research development on rare diseases than for other disease entities for which resources can be made more readily accessible. Many projects have developed shared tools and unified platforms on biobanks and/or registries. They collected biomaterials and data to allow accessibility for research and clinical trials, but the conjunction of biobanks and RD-oriented registries to pool together samples and their clinical information still remained a critical issue.

We are working on establishing community resources and good working procedures by bringing together the RDs community and European infrastructure initiatives, such as ELIXIR, BBMRI-ERIC, and GO FAIR, to align services and procedures at European and international levels. We have selected OI, a rare hereditary disorder, because it is characterized by heterogeneity both on clinical signs (increased bone fragility, blue sclerae, dentinogenesis imperfecta, and hearing loss) and genetic background, sharing important features with other rare and common disorders. The FAIRification process is supported by RD-Connect, ELIXIR, and BBMRI-ERIC.

2. Description of work and efforts

The work for this deliverable has been split into two major tasks:

1. The Rare Disease Working Group (RDWG)
2. FAIRification of the OI registries & biobanks

The Rare Disease Working Group (RDWG)

The RDWG aims to reduce fragmentation and works to optimize the results of various infrastructures and projects which are currently focusing on RDs. Since the end of 2016, this Working Group meets every 2-3 months. RDWG is to date an effective approach to avoid duplication of work on overlapping activities between ADOPT BBMRI-ERIC and other projects/initiatives/networks already in place, such as ERNs, RD-CONNECT and ELIXIR. RDWG, composed of biobank & registry experts, as well as OI specialists, also provides a substantial contribution to the OI Pilot Study.

Tagging biobank collections as Rare Disease collections

We are making the OI Registry (ROI) and Biobank (BIOGEN) data FAIR: Findable, Accessible, Interoperable, and Reusable for humans and machines in order to test and refine the FAIRification procedure. We are following a FAIRification procedure: 1) determine driving user questions, 2) create machine-readable metadata, 3) create machine-readable data by creating a semantic model appropriate to the OI data from specific ontologies such as the Human Phenotype Ontology (HPO) and the Orphanet Rare Disease Ontology (IRDiRC recognized resources), and linking that to the data values in the registry, 4) deposit machine-readable data and metadata on a FAIR data point, a RESTful API based on existing metadata models (re3data, DCAT), and 5) explore added benefits of the FAIR OI data by i.e. answering specific cross-resource questions.

3. Schedule

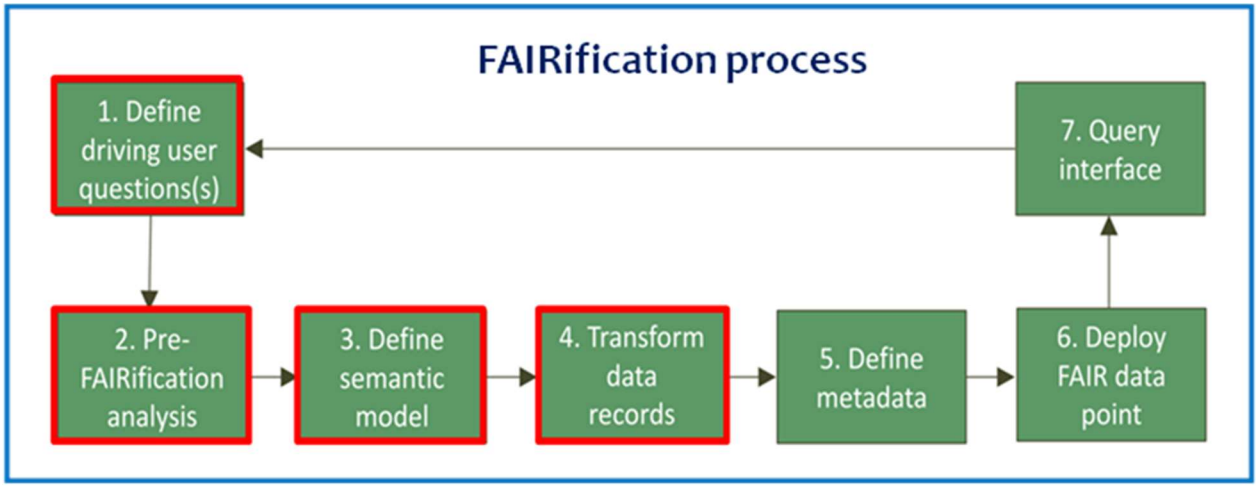
To date the project is still ongoing and the main results of the pilot will be ready in March 2019.

4. Results/Outcome

The activities of the RDWG and FAIRification procedure provided many concrete results for the OI Pilot Study:

- ✓ Definition of a data collection schema for OI information (clinical, genetic, familial, etc.) has been defined, such that a consistent minimum dataset can be produced across OI resources. This has been drafted according to the "European Platform on Rare Disease Registration" (EU RD Platform) by Joint Research Centre of European Commission.
- ✓ Identification of a minimum set of ontologies and/or vocabularies to support data integration: Orphacode/ICD for disease diagnosis, HPO for phenotype and HGVS/OMIM for genetic diagnosis.

- ✓ Definition of six driving questions, based on the data elements in the ROI & BIOGEN (e.g., How many paediatric patients with a missense mutation in COL1A1/A2 genes have dentinogenesis imperfecta? How many samples of these subjects are available?). These questions were created to focus the ROI FAIRification to include the most important data elements related to critical research questions, and to link them to biospecimens and their data.

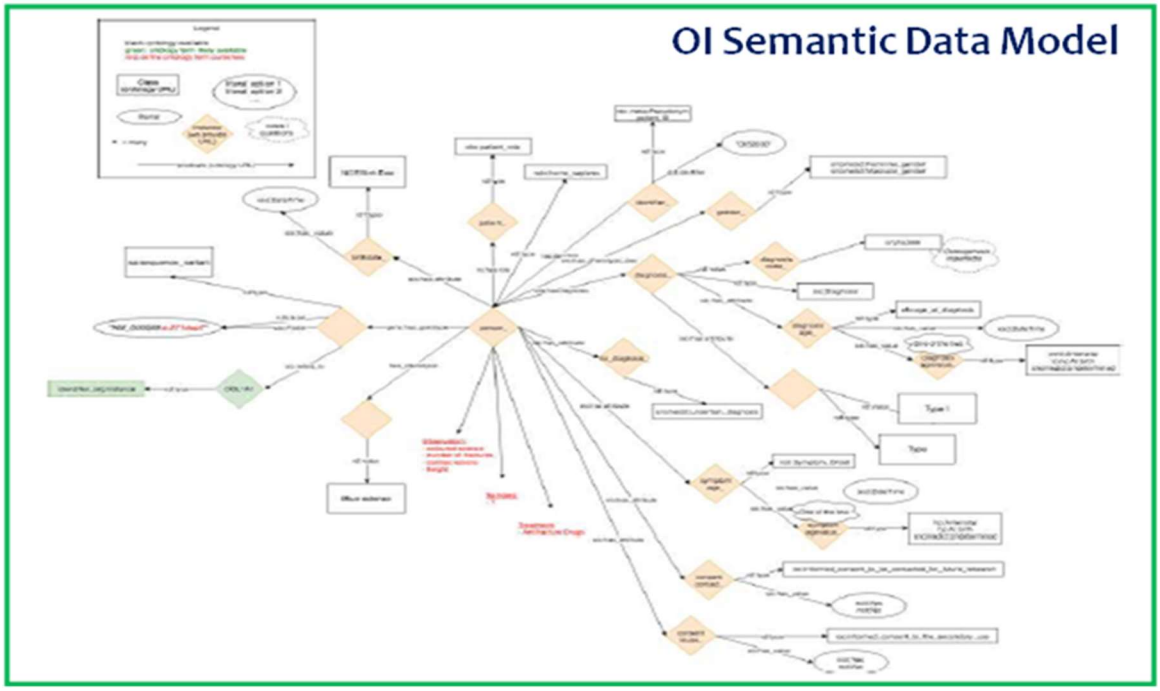


All the data generated is not effectively useful, reproducible, nor reusable, if information is not findable.

To this aim, data and metadata for the registry are being made discoverable for humans and machines. This is a mixture of human-readable and machine-readable sections, capable of highlighting the visibility of the registry.

- ✓ Generation of a semantic data model of the ROI data elements covered by the driving questions. The semantic data model is used to make the human-readable OI Registry data machine-readable by converting its data to a knowledge graph encoded in RDF (Resource Description Framework). The semantic data model used ontologies to describe concepts and created relations between concepts, i.e. 'person' '*has diagnosis*' 'Osteogenesis Imperfecta'. Having the OI Registry data in RDF makes its data interoperable (i.e. ready for computational analysis across resources) with other similarly treated OI Registries, other rare disease registries, and other biological knowledge resources.

Figure 1: O1 Semantic Data Model



The approach proposed for the OI Pilot Study aims to corroborate and integrate data from different sources helping in advancements for OI research. Importantly, the FAIR data and metadata models can be reused and adapted by other rare and common diseases.

5. Further developments

This Use Case has generated much interest, and it will be extended along with:

- Manuela Posada, RD-Connect Spagna
- Biobanks within TNGB Network
- Malcom Pace, BBMRI-ERIC Malta
- BBMRI.be

The main approaches are not only collecting data, but also the method for exchanging data and for making the information interoperable.

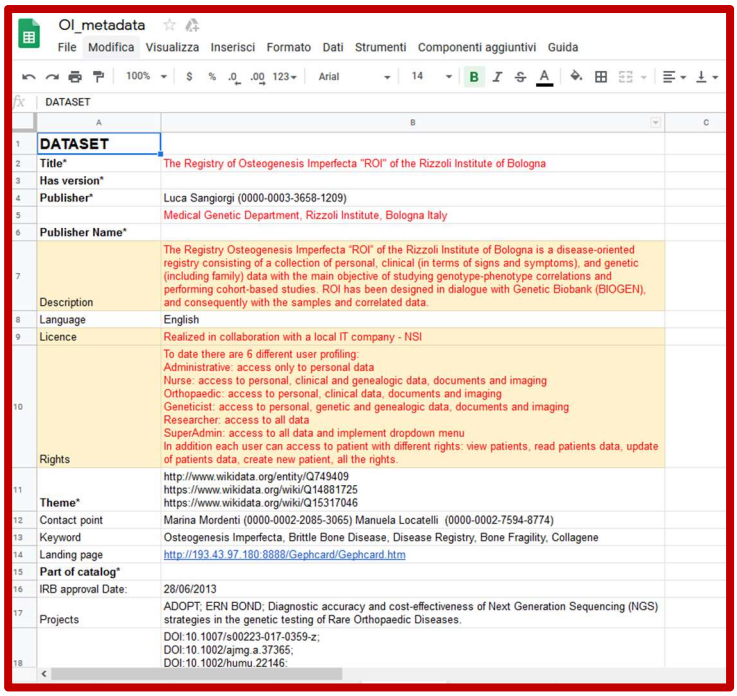
We circulated the form for collecting data:

- 'Osteogenesis_Imperfecta_Form_ADOPT_draft'

and the form for samples' information:

- 'samples minimum data set_draft'.

To make the data findable and accessible we define metadata: information that is given to describe or help you use other information.



OI_metadata

File Modifica Visualizza Inserisci Formato Dati Strumenti Componenti aggiuntivi Guida

DATASET

	A	B	C
1	**DATASET**		
2	**Title***	The Registry of Osteogenesis Imperfecta "ROI" of the Rizzoli Institute of Bologna	
3	**Has version***		
4	**Publisher***	Luca Sangiorgi (0000-0003-3658-1209)	
5		Medical Genetic Department, Rizzoli Institute, Bologna Italy	
6	**Publisher Name***		
7	Description	The Registry Osteogenesis Imperfecta "ROI" of the Rizzoli Institute of Bologna is a disease-oriented registry consisting of a collection of personal, clinical (in terms of signs and symptoms), and genetic (including family) data with the main objective of studying genotype-phenotype correlations and performing cohort-based studies. ROI has been designed in dialogue with Genetic Biobank (BIOGEN), and consequently with the samples and correlated data.	
8	Language	English	
9	Licence	Realized in collaboration with a local IT company - NSI	
10	Rights	To date there are 6 different user profiling: Administrative: access only to personal data Nurse: access to personal, clinical and genealogic data, documents and imaging Orthopaedic: access to personal, clinical data, documents and imaging Geneticist: access to personal, genetic and genealogic data, documents and imaging Researcher: access to all data SuperAdmin: access to all data and implement dropdown menu In addition each user can access to patient with different rights: view patients, read patients data, update of patients data, create new patient, all the rights.	
11	**Theme***	http://www.wikidata.org/entity/Q749409 https://www.wikidata.org/wiki/Q14881725 https://www.wikidata.org/wiki/Q15317046	
12	Contact point	Marina Mordenti (0000-0002-2085-3065) Manuela Locatelli (0000-0002-7594-8774)	
13	Keyword	Osteogenesis Imperfecta, Brittle Bone Disease, Disease Registry, Bone Fragility, Collagene	
14	Landing page	http://193.43.97.180:8888/Gephcard/Gephcard.htm	
15	**Part of catalog***		
16	IRB approval Date:	28/06/2013	
17	Projects	ADOPT; ERN BOND; Diagnostic accuracy and cost-effectiveness of Next Generation Sequencing (NGS) strategies in the genetic testing of Rare Orthopaedic Diseases.	
18		DOI:10.1007/s00223-017-0359-z; DOI:10.1002/ajmg.a.37365; DOI:10.1002/humu.22146;	

Finally, we deploy a FAIR data point, a machine-readable version of the metadata, to search our library.

With this step, the FAIRification process of OI Pilot is completed.

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix dct: <http://purl.org/dc/terms/>.
@prefix dcat: <http://www.w3.org/ns/dcat#>.
@prefix foaf: <http://xmlns.com/foaf/0.1/>.
@prefix fdp: <http://rdf.biosemantics.org/ontologies/fdp-o#>.
@prefix datacite: <http://purl.org/spar/datacite/>.

<> rdf:type dcat:Dataset;
    dct:title "The Registry of Osteogenesis Imperfecta \"ROI\" of the Rizzoli Institute of Bologna";
    dct:hasVersion "1";
    dct:publisher <https://orcid.org/0000-0003-3658-1209>;
    dct:description "The Registry Osteogenesis Imperfecta "ROI" of the Rizzoli Institute of Bologna is a disease-oriented  registry consisting of a
    dct:language <http://id.loc.gov/vocabulary/iso639-1/en>;
    dcat:theme <http://www.wikidata.org/entity/Q749409>, <http://www.wikidata.org/entity/Q14881725>, <http://www.wikidata.org/entity/Q15317046>;
    dcat:contactPoint <https://orcid.org/0000-0002-2085-3065>;
    dcat:keyword "Osteogenesis Imperfecta", "Brittle Bone Disease", "Disease Registry", "Bone Fragility", "Collagene";
    dcat:landingPage <http://193.43.97.180:8888/Gephcard/Gephcard.htm>.
<https://orcid.org/0000-0003-3658-1209> rdf:type foaf:Organization;
    foaf:name "Luca Sangiorgi".
```

6. References

1. Van Dijk FS, Sillence DO. Osteogenesis imperfecta: clinical diagnosis, nomenclature and severity assessment. Am J Med Genet A. 2014 Jun; 164A(6):1470-81. DOI: 10.1002/ajmg.a.36545

2. Wilkinson MD et al. The FAIR Guiding Principles for scientific data management

and stewardship. Sci Data. 2016 Mar 15;3:160018. DOI: 10.1038/sdata.2016.18.

3. Orphanet Report Series. Prevalence of rare diseases: Bibliographic data - Prevalence, incidence or number of published cases listed by diseases (in alphabetical order). INSERM 2016
http://www.orpha.net/orphacom/cahiers/docs/GB/Prevalence_of_rare_diseases_by_alphabetical_list.pdf