

# Supporting FAIR data discoverability in clinical research: providing a global metadata repository (MDR) of clinical study object

## Principal Investigators:

- Sergei Gorianin (ECRIN, France, [sergei.gorianin@ecrin.org](mailto:sergei.gorianin@ecrin.org))

## Shepherds:

- Stefano Nicotri (INFN, [nicotri@infn.it](mailto:nicotri@infn.it))
- Giacinto Donvito (INFN, [giacinto.donvito@ba.infn.it](mailto:giacinto.donvito@ba.infn.it))
- Hans van Piggelen (SURF, [hans.vanpiggelen@surf.nl](mailto:hans.vanpiggelen@surf.nl))

Entry in the community requirement database: [ECRIN-ERIC](#)

## About the pilot

### Description of supported work

#### Objective

Supporting FAIR data discoverability in clinical research: providing a global metadata repository (MDR) of clinical study object.

#### General

In order to fully assess and review the evidence generated in clinical research, it is necessary to have access not only to the published results but also to the source individual participant data and related study documents (e.g. study protocol, statistical analysis plan, case report form). As data and document sharing becomes more and more common, however, the researcher is faced with a bewildering mosaic of possible source locations and access modalities. There is an urgent need to develop a central resource that can catalogue all the diverse data and documents associated with a clinical study, and then make that information searchable by using a central web portal.

In the EU H2020-funded project eXtreme DataCloud (XDC; grant agreement 777367) such a service is developed under the coordination of ECRIN-ERIC (European Clinical Research Infrastructure Network) and will be available to open public till the end of 2019.

Currently, the MDR instance in production contains more than 700,000 data objects from one trial registry ([ClinicalTrial.gov](#)) and 6 repositories (PubMed, ZENODO, Data Dryad, WWARN, Edinburgh DataShare, BioLINCC).

The first objective of the proposed project is to extend the MDR demonstrator to run in production in the EOSC environment and be part of the EOSC catalogue and to complete the data base by integrating all major data sources dedicated to clinical research.

The second objective is to include others EOSC services not already included into MDR: we plan to include also EGI Fed Cloud Resources to host the distributed repositories.

## Team

Participant	Role	Name and Surname
Sergei Gorianin	PI	<a href="mailto:sergei.gorianin@ecrin.org">sergei.gorianin@ecrin.org</a>

## Technical Plan

The full technical plan can be found here: [https://docs.google.com/document/d/1wv6sXZt7Cjr-twmjogeY\\_eXpMa5fpj1](https://docs.google.com/document/d/1wv6sXZt7Cjr-twmjogeY_eXpMa5fpj1)

Work planned for Q1	<ul style="list-style-type: none"><li>Investigate a new mechanism of ECRIN metadata 'injection' and upgrading on OneData environment.</li><li>Revision of web-portal.</li><li>Investigate metadata schema and requirements for future harvesting by B2FIND</li></ul>
---------------------	--

Work planned for Q2	<ul style="list-style-type: none"> <li>• Continue the revision of current web-portal, developed within XDC project in collaboration with OneData (web-portal GUI + OneData Environment) and INFN (ElasticSearch + hardware support).</li> <li>• Testing and upgrading the web-portal with respect to updated ECRIN requirements.</li> <li>• Continued investigation on harvesting by B2FIND</li> </ul>
Work planned for Q3	<ul style="list-style-type: none"> <li>• Testing the work produced in Q1 and Q2</li> <li>• Start to develop ElasticSearch-based APIs in collaboration with INFN</li> <li>• <del>Enable harvesting of a single MDR instance by B2FIND test instance</del></li> </ul>
Work planned for Q4	<ul style="list-style-type: none"> <li>• Finalizing the development and integration testing by users.</li> <li>• Support for potential users, including collecting metrics as well as feedback, and feeding back requests for change.</li> <li>• <del>Enable harvesting of one or more MDR instance by B2FIND production instance</del></li> </ul>

## EOSC services and providers

### Provider

- INFN-Bari

### Services

- INFN cloud platform Cloud@ReCaS-Bari
- INFN ElasticSearch
- OneData environment
- ~~B2FIND metadata catalogue~~

## Attachments

MDR - EOSC-Hub EAP.pptx	2020 7 20 by Hans van Piggelen
Interaction of the MDR with B2FIND EUDAT service.docx	2020 7 20 by Hans van Piggelen
EOSC_hub_MDR_earl_adopter_report.docx	2020 7 20 by Hans van Piggelen

