Intro

Hello everyone. My name is Sergei Gorianin and I'm a data scientist from ECRIN and today I wish to present "MetaData Repository (MDR) for clinical research.

Slide 1

The MDR is a single portal system which idea is to link clinical studies with related data objects (e.g., study protocol, datasets, statistical analysis plan, publications)

It has been developed within the H2020-funded project XDC in cooperation by ECRIN, OneData and INFN (http://www.extreme-datacloud.eu/)

It is based currently upon the ECRIN metadata schema (version 5) (https://zenodo.org/record/4133889#.X_RoY9hKjcs)

Officially the MDR was launched on 29 April 2020 and put in production for the ECRIN task force on COVID-19, linked as related resource to the European COVID-19 data portal and referenced in the RDA COVID-19 guidelines and recommendations

Currently update and extension of the MDR is in the EOSC-hub (early adopter program) and EOSC-Life (WP1 use case) with expansion of data sources.

Slide 2

On this slide you can see an overview of the current MDR solution.

At the beginning we analyze and process manually each data source. After that, we extract metadata from each of those data sources and put into individual databases and schemas, saving original metadata structures.

Then, we run ETL processes, which include mainly metadata standardization, linkage between themselves and cleaning from duplicates. We use our metadata schemas, which was developed by us as a standardized structure for studies and data objects records. All these schemas are publicly available.

Each record, which was processed by ETL, inserts into our core database.

After that, we convert those records to JSON format, using our metadata schemes as well, and push those JSON files with studies and data objects metadata inside to ONEDATA system. We name this process 'injection', which include, first of all, import of all JSON files to OneProvider, and then, extraction metadata from each of those files and attaching it to the files.

Later, when 'injection' finished, those metadata become available through OneZone for external components, in our case it's ElasticSearch, and for final users on the web-portal.

Slide 3

The web-portal is integrated into OneData system, but it's publicly available.

The web-portal itself is a single-page web-application. On top of it you can see the title and the search panel with different searching options. You can find necessary study by selecting any of several options in the 'Select mode' select box.

Left panel here is represented by studies and data objects filters options.

It's hidden on the screenshot but the bottom part of the web-page contains footer, where you can see 'Disclaimer', 'Data sources and contributing organizations', contact information and collaborators logos.
Also, we've a help page there, where user can find all information how to use our web-portal. And the central part of the web-portal is the content area.

Slide 4
Clinical studies
- 18 trial registries, 2 repositories included
- 600.125 studies included

Data objects related to a study
- 18 registries, 2 repositories, Pubmed included
- 31 data object types (e.g. study protocol, data, publication)
- 913.832 data objects included