

MetaData Repository (MDR) for clinical study objects

Sergei Gorianin

Updated: 25.02.2020

FYI, just to remind

Findable

The FAIR Data Principles are a set of guiding principles in order to make data findable, accessible, interoperable and reusable (Wilkinson et al., 2016).

Accessible

These principles provide guidance for scientific data management and stewardship and are relevant to all stakeholders in the current digital ecosystem.

Interoperable

They directly address data producers and data publishers to promote maximum use of research data. Research libraries can use the FAIR Data Principles as a framework for fostering and extending research data services.

Reusable

Problem – clinical trials

Findable



Accessible



▭ *Principles & Recommendations for clinical data sharing (Ohmann et al. BMJ Open 2017)*

Interoperable



▭ *Definition of processes (Ohman et al. F1000Research 2018)*

▭ *Quality criteria for repositories (Banzi et al. Trials 2019)*

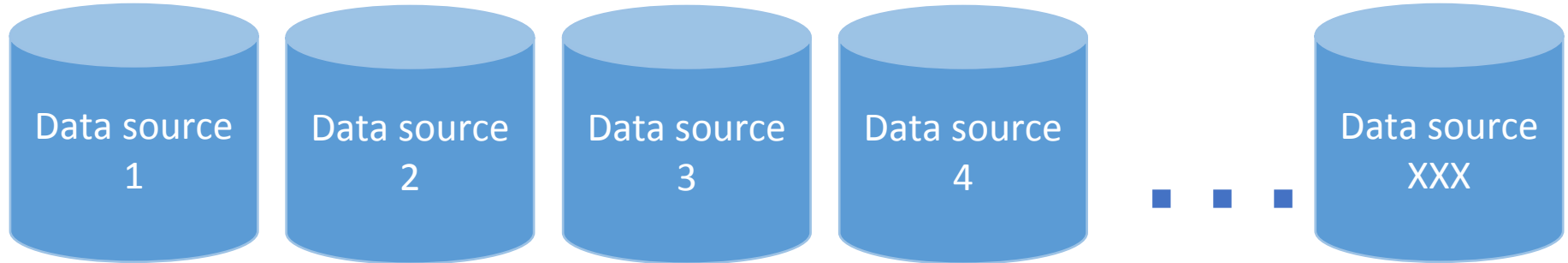
Reusable



▭ *Pilot repository (Tilki et al. Submitted)*

▭ *Recommendations for anonymisation technical solutions (ongoing)*

MDR – idea to improve the **Findability** ?



Let's summarize the problems

- Plenty of data sources
- Unstandardized data
- Unlinked data
- Unidentified data
- Collecting data is ~~very~~ quite problematic
- Possible duplicates
- etc...

**How great would be to have all
standardized and useful
information in a single place...**



MDR... Why?

To maximize the **discoverability** of all these data objects, it is necessary to collect the *metadata* about them, including object provenance, location and access details, into a single system.

To that end ECRIN is currently developing an **MDR** (MetaData Repository) to standardize, assemble and display the metadata about clinical studies and the data objects generated by them, and provide access to them through a single system, accessed via a web portal.

A **study** is any clinical research study with humans as study participants, and which is therefore subject to ethical approval, whether or not the study is interventional (a 'clinical trial'), or observational (including disease registry data), or a case study.

A **data object** is any information available in electronic form (a 'data stream') and may be a document (e.g. a pdf), a dataset in one of a variety of formats (spreadsheet, csv files, database file, XML etc.), a media (audio, video) file, an image (e.g. a conference poster) or simply a web page with useful text.

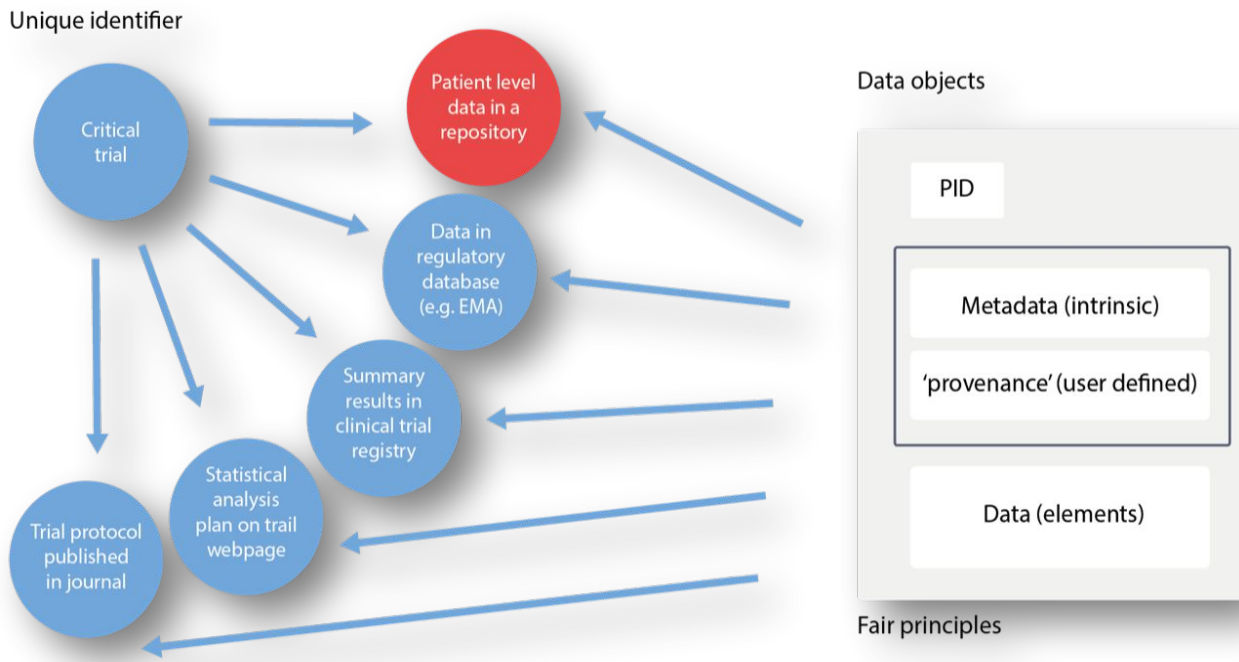


Fig. 1 – Structure of clinical trial (study, data object)

MDR... Implementation

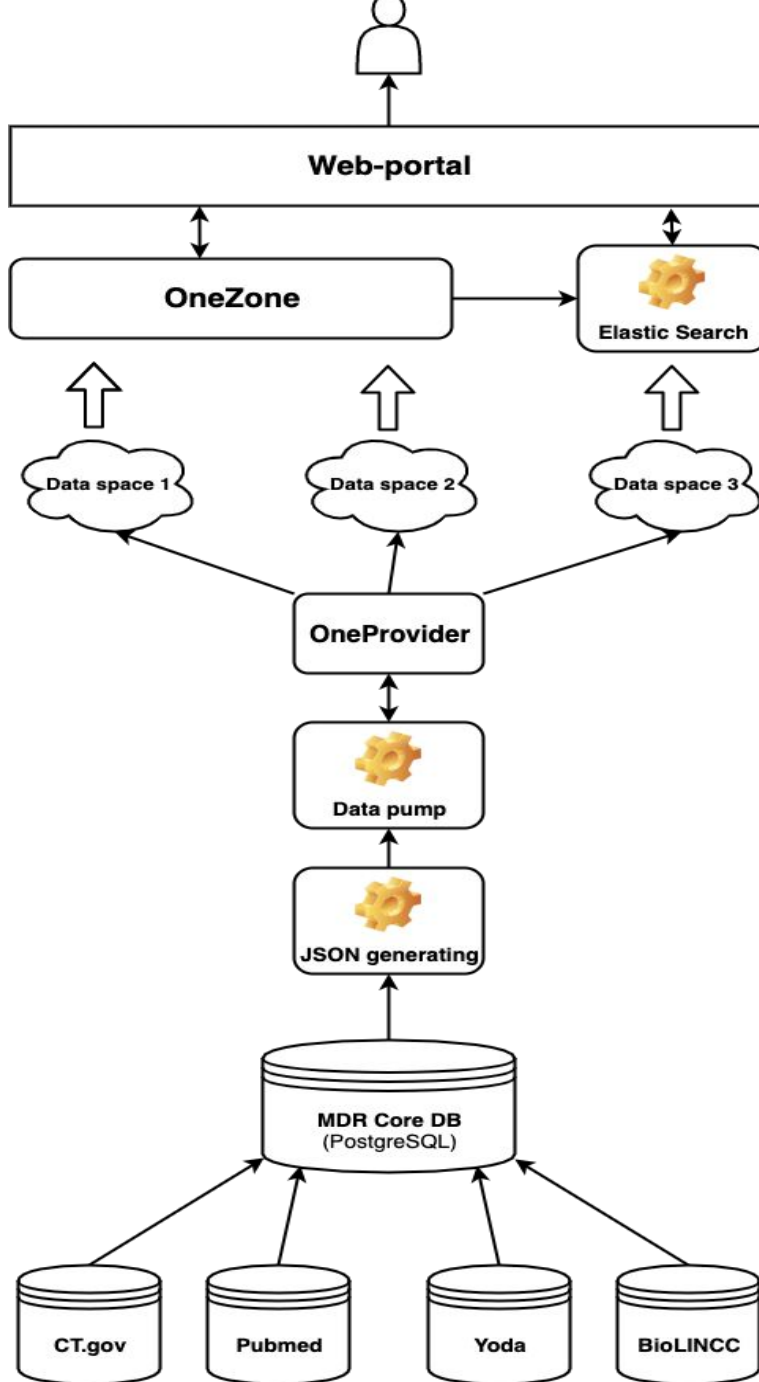
The web portal is developed in collaboration with ONEDATA (Cyfronet, Poland); functionality for discoverability of studies and related data objects is provided by INFN (Italy)*. So far metadata from 4 data sources have been collected using different modalities (e.g. XMLs processing, APIs, scraping of web pages) and stored as JSON objects/relational DB form on the test bed server at INFN. Data will be then ingested, uploaded and made available for the users.

The ONEDATA solution will implement the metadata collection and transport from multiple OneProvider to the central OneZone service; it will also deal with the ACLs management to implement the data protection.

** Within H2020 XDC project*

ONEDATA

ONEDATA



User interface for Web portal (developed by ONEDATA)

MDR Metadata Repository for clinical trial data objects

Select mode: Study characteristics (dropdown) | Study title contains: cancer | AND (dropdown) | Study topics include: (empty) | Find (button)

Filter: Studies (selected) | Data objects | Results: 1000 | Load... | Save... | Clear

Type: Filter options... | Deselect all

- Interventional
- Observational
- Observational Patient Registry
- Expanded access
- Funded programme
- Not yet known

Status: (dropdown)

Gender eligibility: (dropdown)

Interventional studies filters

- Phase: (dropdown)
- Intervention model: (dropdown)
- Allocation type: (dropdown)
- Primary purpose: (dropdown)
- Masking: (dropdown)

Observational studies filters

- Observational model: (dropdown)
- Time perspective: (dropdown)

Filter (button)

Results list:

- Women's Health Study (WHS): A Randomized Trial of Low-dose Aspirin and Vitamin E in the Primary Prevention of Cardiovascular Disease and Cancer [X]
- The Use of Chemotherapy Plus Radiotherapy Plus Azidothymidine in Patients With AIDS-Related Lymph Node Cancer [X]
- A Study of Patients With AIDS Syndrome [X]
- Studies on Tumors of the Thyroid [X]
- Clinical, Laboratory, and Epidemiologic Characterization of Individuals and Families at High Risk of Cancer [X]
- Psychological Benefits of a Normalized Camping Experience for Children With Cancer [X]
- Description (show more)*: Background: Cancer has an enormous impact on the psychological and social well-being of the family unit. The life-threatening connotations of cancer single out the ill child from his peer/family group as one who is different, and often unable to maintain a normal lifestyle. Physical sequelae of cancer and its treatment accentuate the differences between these children and their normal peers/siblings. It is impor...
- Study Overview (show more)*: Psychological Benefits of a Normalized Camping Experience for Children With Cancer :: Study Overview - [X]
Access: https://clinicalstudies.info.nih.gov/cgi/detail.cgi?A_1983-C-0022.html
- Trial Registry entry (show more)*: Psychological Benefits of a Normalized Camping Experience for Children With Cancer :: CTG Registry entry 1999 [X]
Access: <https://clinicaltrials.gov/ct2/show/study/NCT00001186>
- A Multimodality Treatment Approach to Patients With Inflammatory Cancer of the Breast and Locally Advanced Non-Inflammatory Stage III Breast Cancer and Stage IV Breast Cancer [X]
- Interferon and Octreotide to Treat Zollinger-Ellison Syndrome and Advanced Non-B Islet Cell Cancer [X]
- Combination Chemotherapy (FLAC) Combined With Granulocyte-Macrophage Colony Stimulating Factor in Locally Advanced and Metastatic Breast Cancer [X]
- Effect of Preoperative Chemotherapy on Axillary Lymph Node Metastases in Stage II Breast Cancer: A Prospective Randomized Trial [X]
- Phase I Trial of FLAC (5-Fluorouracil, Leucovorin, Adriamycin, Cytosan) Plus GM-CSF (Granulocyte-Macrophage Colony Stimulating Factor) Plus Dose Escalation of IL-3 (Interleukin-3) in Metastatic ... [X]

Data sources and contributing organisations: Please contact us [here](#), to report an error or to arrange for your for your data to be included in the MDR.

ONE DATA | ECRIN (EUROPEAN CLINICAL RESEARCH INFRASTRUCTURE NETWORK) | eXtreme DataCloud | EO SC pilot (The European Clinical Science Cloud for Research Pilot Project)

Current progress

- 4 data sources (PubMed, CT.gov, BioLINCC, Yoda) are connected
 - Single and universal metadata schema was developed*
 - about 800.000 studies and data objects uploaded
- ✓ JSON Schema - http://ecrin-mdr.online/index.php/JSON_Schemas
 - ✓ ECRIN Clinical Research Metadata Schema <https://zenodo.org/record/3562911>

Next steps

- Refine current **metadata model** to increase usefulness to users, and further investigate alignment with other approaches, e. g. data tags.
- Modify **data extraction** to better handle periodic interrogation of the same source (i.e. only handle new or revised data).
- Get metadata from more major **studies registries** (about 6); establish links between new and already extracted studies and data objects.
- Extend extraction to other **data repositories** (about 10).
- Work on **search engine** (ElasticSearch or PostgreSQL alternative); preparation of API supporting data access.
- Develop and deploy **web-portal**.
- Look at developing support tools.
- **Containerization** of developed tools and components.
- Integration with **AAI**, developed and provided by EOSC-Life.
- **Publish as FAIR data resource in the EOSC.**

Acknowledgments



S. Battaglia, S. Canham, C. Ohmann



S. Nicotri, A. Italiano, G. Donvito



J. Kudzia, L. Dutka,
M. Orzechowski, M. Borzecki



More details?

Our wiki:

http://ecrin-mdr.online/index.php/Project_Overview

Thank you!

Any questions?