

FGAI4H-N-054

E-meeting, 15-17 February 2022

Source: Editors DEL5.6

Title: DEL5.6: Data sharing practices – Progress Review Presentation

Purpose: Discussion

Contact: Ferath Kherif
CHUV -LREN
Switzerland

E-mail: Ferath.kherif@chuv.ch

Abstract: This PPT discusses the latest updates made on the deliverable Updated DEL5.6: Data sharing practices + Data sourcing.

Data Sharing and Data sourcing



AI for Health

ITU-WHO Focus Group



World Health
Organization

itu.int/go/fgai4h

Data Sharing and Data sourcing

- Initial draft describes the objectives and proposes an initial outline of the planned deliverable “Data Sharing Practices”
- This deliverable aims to provide an overview of the existing **best practices** for data sharing of health-related data, including the requirement to enable **secure** data sharing and issues related to data **governance**.
- The documents described **established solutions** and **novel** approaches based on **distributed and federated environments**.

Data Sharing Principles

Sharing medical data at scale is necessary to improve the development and adoption of artificial intelligence solutions for healthcare and to make these solutions more robust.

Negotiate the balance between potential harm from sharing critical data and the potential benefit of improving care.

Data Sharing Principles

- **Sharing data for appropriate and authorised purposes.** *Why the data is being used*
- **Sharing data only with authorised users.** *Who is using the data*
- **Using data in a safe and secure environment.**
- **Applying appropriate protections to the data.** *Where the data is being used*
- **Ensuring public outputs from data sharing projects do not identify the people or organisations in the data.**

Data Sharing and Data sourcing

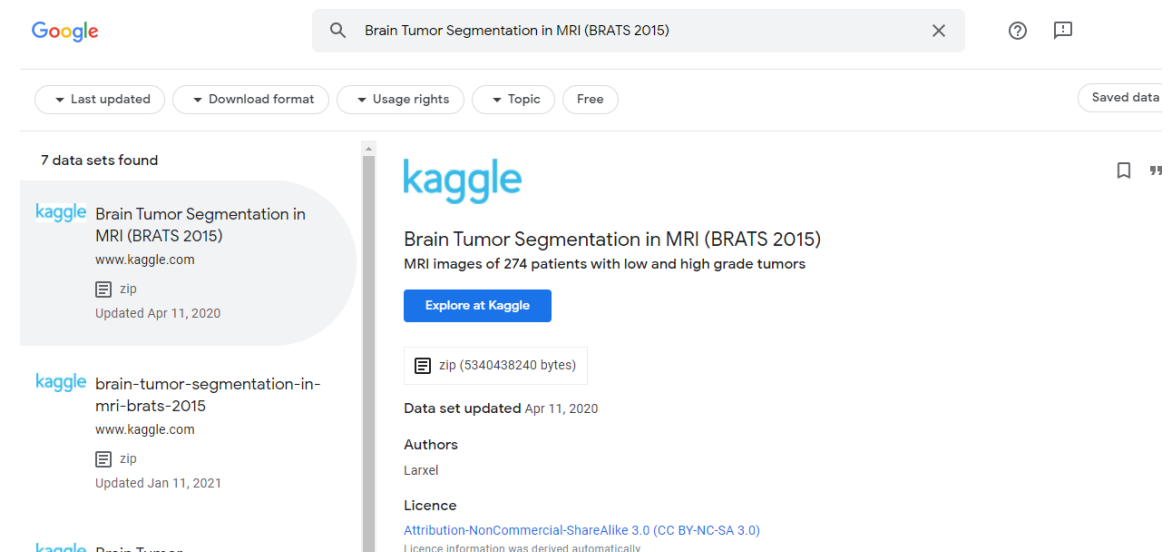
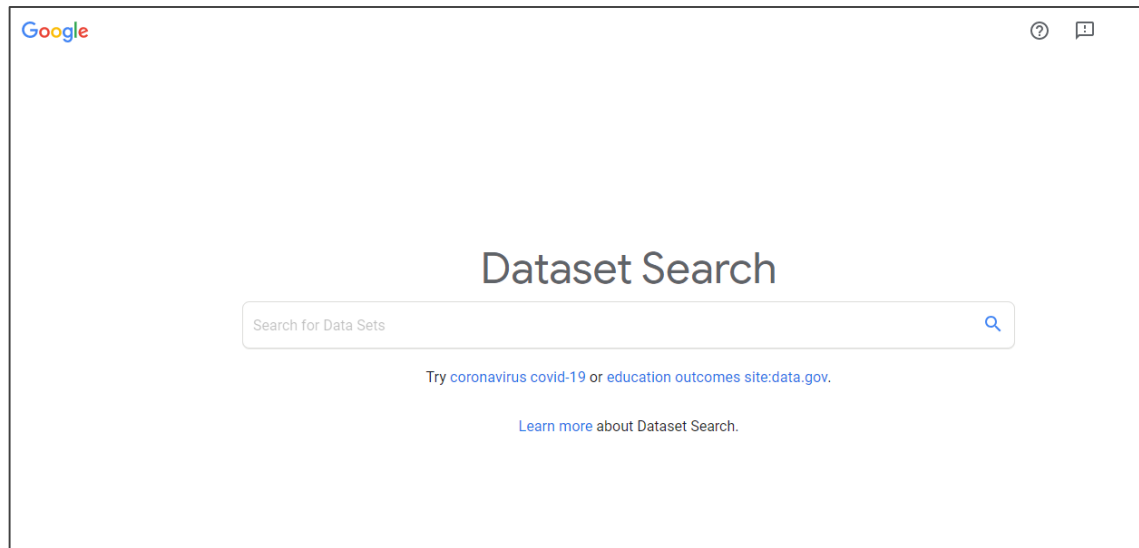
- Necessary steps and requirements to **enable secure data sharing**
- Methods for Data sourcing
 - **Centralized data sources**
 - **Distributed data sources**

Data Sharing and Data sourcing

- Necessary steps and requirements to **enable secure data sharing**
 - Roles and responsibilities of the **data providers, data processors & data receivers**
 - **Data anonymization and de-identification**
 - **Data minimization**
 - **Data confidentiality, Data security and privacy**
 - **Protection and Privacy assessments.**
 - **Responsibilities of data providers and data receivers**
- Necessary steps and requirements to **enable secure data sourcing**
 - **Description of data**
 - **Data ingestion and update**
 - **Results and dissemination of results including IA models**

Best practices for data sharing/Sourcing Platform

- **Data Catalogue & Search**



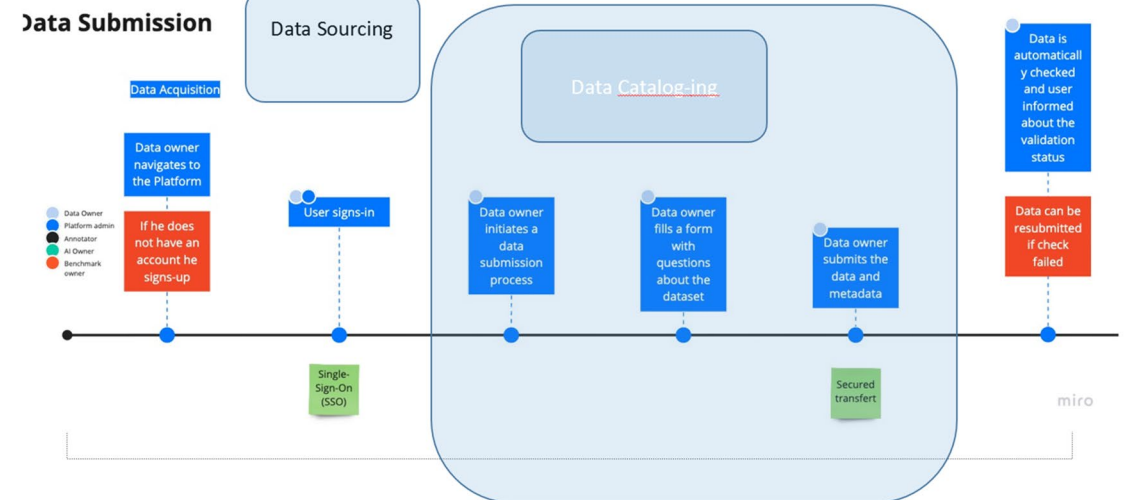


Open-code software

- Develop software tools (e.g., data acquisition, data storage, annotation, prediction, evaluation, and reporting packages)
- Involve developers, regulators, and medical professionals
- Targeted towards a universal tool applicable across borders
- Usable by multiple stakeholders such as notified bodies and doctors

Health is not a local problem and requires a connected structure, universal solutions, and joined principles

Data Submission



Implement tools and define the processes that will support data descriptions and intelligent use in the context of the FG - AI4H objectives, including ethical, annotation, and validation challenges.

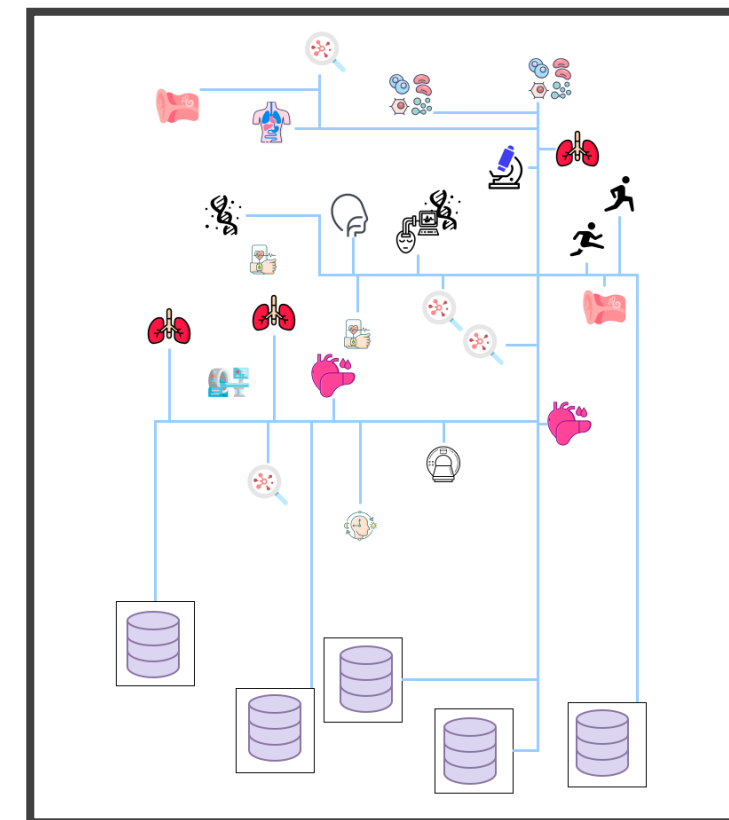
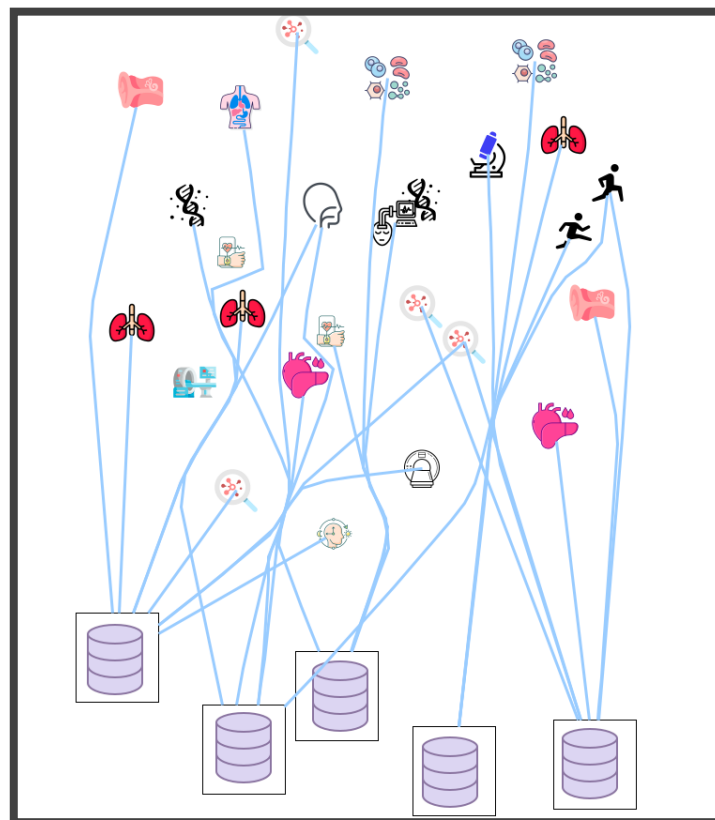
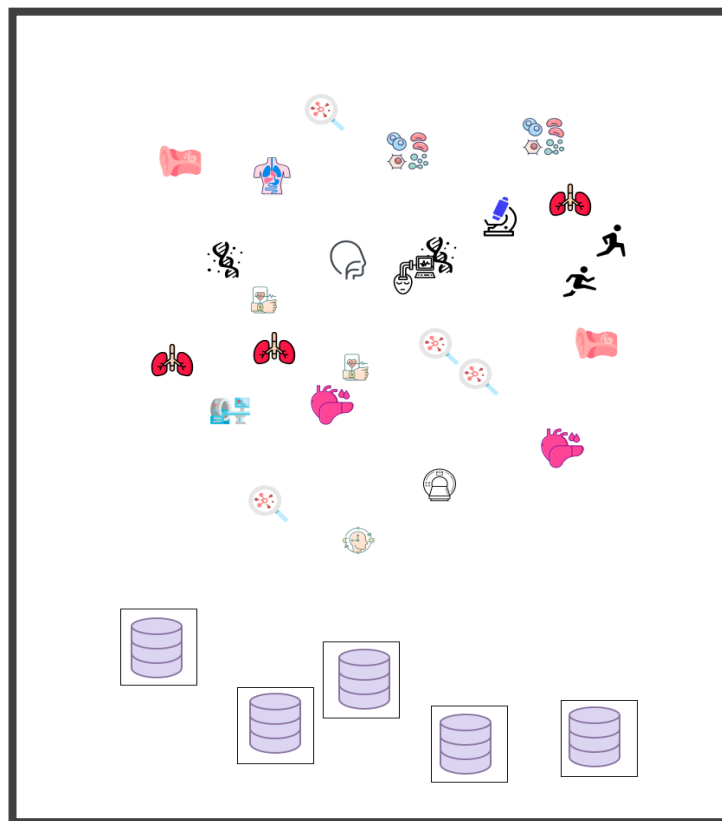
Discover / Capture



Connect Data



Link Data



DATA SPACES (Market) Squares



DATA Catalog Cloud

API ACCESS LAYER



LOCAL DATA

Devices



Metadata Components

Technologies/Platforms tested Data search, data management

- Data lineage support
 - Amundsen
 - DataHub
 - Atlas
 - CKAN
 - Magda - Federation
- Cloud based
 - Google DC
 - AWS
 - Azure

Technologies/Platforms Selected/built

- Data search, data management
- + data lineage support + federation
 - + Annotation metadata + link to ML model
 - + Governance
 - Combined multiple approaches based on **own development and Obiba.org suites** :
 - MetaData capture
 - Data curation according to standards
 - Storage: ontologies/semantic based query
 - Publish data
 - Manage data transfer according to privacy sharing principles (D5.6)
 - Advanced query capabilities according to context information.
 - data transfer to ML models

Data sourcing

Own development and Obiba.org suites

LREN - Web Toolkit

Ferath Khent

Data exploration

Environment
Greenness, noise

Socio-demographics
Demographics, education, profession

Lifestyle
Physical activity, Sleep, Habits

Brain_maps
Volumes, PD, White matter

Behaviour
Addictions, Memory tests

General_Health
Medical history, Blood assay

Brain_maps

Socio-demographics

Lifestyle

Environment

Behaviour

Gender

Select your variables

Collapse All

Search...

Code	Follow up	Description	Info
<input type="checkbox"/> brainstem_pd	FU3	Brainstem volume	
<input type="checkbox"/> brainstem_thrs_mt	FU2	Brainstem volume	
<input type="checkbox"/> brainstem_mt	FU2	Brainstem volume	
<input type="checkbox"/> brainstem_mt	FU3	Brainstem volume	
<input type="checkbox"/> brainstem_thrs_mt	FU3	Brainstem volume	
BRPPU2 (3)			
<input type="checkbox"/> F2_BRPPU2	FU2	PsyColaus: Followup 2 brief psychoti...	
<input type="checkbox"/> F0_BRPPU2	FU0	PsyColaus: Baseline brief psychotic d...	
<input type="checkbox"/> F1_BRPPU2	FU1	PsyColaus: Followup 1 brief psychoti...	
BRPPUOF (3)			
<input type="checkbox"/> F0_BRPPUOF	FU0	PsyColaus: Baseline brief psychotic d...	
<input type="checkbox"/> F1_BRPPUOF	FU1	PsyColaus: Followup 1 brief psychoti...	
<input type="checkbox"/> F2_BRPPUOF	FU2	PsyColaus: Followup 2 brief psychoti...	
BRPPUON (3)			
<input type="checkbox"/> F0_BRPPUON	FU0	PsyColaus: Baseline brief psychotic d...	
<input type="checkbox"/> F1_BRPPUON	FU1	PsyColaus: Followup 1 brief psychoti...	
<input type="checkbox"/> F2_BRPPUON	FU2	PsyColaus: Followup 2 brief psychoti...	
buxexamac (2)			
<input type="checkbox"/> buxexamac	FU2	Does the participant take this medi...	
<input type="checkbox"/> buxexamac	FU1	Does the participant take this medi...	
BUL2 (3)			
<input type="checkbox"/> F1_BUL2	FU1	PsyColaus: Followup 1 bulimia nervosa	
<input type="checkbox"/> F0_BUL2	FU0	PsyColaus: Baseline bulimia nervosa	
Total count: 3010			
See your selection (0)			
<div>Deep phenotype analysis</div> <div>Create project with corresponding dataset</div>			

Mica Home Repository Search Research Data Access

Sign in Sign up

Data portal

Explore the data catalog and make data access requests.

7
Networks

More info

17
Studies

More info

32
Datasets

More info

18,350
Variables

More info

Search networks, studies, datasets and variables to explore the harmonization potential across studies and the variables of interest for your research project.

Search

Get informed about the steps to request data access for a research project.

Data Access Process

Opal Dashboard Projects Search

Search variables...

Administration Help administrator

Dashboard

ACTIONS

Explore Data

Browse tables and variables, create views, import/export data and dictionaries.

Search

Search for variables, a specific entity or entities matching some variable criteria.

Manage Files

Browse, upload and download files on the server.

DataSHIELD

Configure DataSHIELD, grant access to DataSHIELD service.

Tasks

Follow tasks being executed.

Run Reports

Design reports with R and schedule their execution.

My Profile

Change password and create personal access tokens.

BOOKMARKS

Resource

Type

No items.

LATEST NEWS

Opal 4.2 Released

2021-07-05 - R servers clusters, DataSHIELD profiles [more ...](#)

Mica 4.5 Released

2021-06-01 - Improved document listing public pages, Opal Personal Access Token [more ...](#)

Opal 4.1 Released

2021-04-27 - SQL API, Personal Access Token with read-only support [more ...](#)

Opal 4.0 Released

2021-03-08 - Cluster of Rock R servers, RDS import/export [more ...](#)

Opal 3.0 Released

2020-06-11 - Project resources and backup/restore [more ...](#)

Opal 2.16 Released

2019-12-10 - Harmonization using R, new derivation functions [more ...](#)

© 2021 Obiba Documentation Sources

4.2.6

Data Access control

The screenshot displays the Mica web application interface, specifically the 'Data Access' section. The browser address bar shows 'localhost:8445/networks'. The application header includes the Mica logo, navigation links (Home, Repository, Search, Research, Data Access), and user options (Sign in, Sign up). A sidebar on the left lists 'Networks', 'Studies', and 'Datasets'. The main content area is titled 'Data Access Requests' and contains a description: 'Manage the data access requests and the associated users.' Below this, there are two tabs: 'Data Access Requests' (with a count of 0) and 'Registered Users' (with a count of 0). A blue button labeled '+ New Data Access Request' is visible. The footer of the application states 'Powered by OBiBa Mica' and 'Copyright © 2021 Mica. All rights reserved.'


MVP: Data sourcing


LREN - Web Toolkit

Data curation

Data curation

Here we test and measure the reliability of our data. Check it out and add your comments.

 **User-driver curation**
Choose a file you are interested in, explore it and add comments to combine expertises!

 **Automatic data curation**
Is the data as described? Our daily checks...

LREN - Web Toolkit

Data curation / Automatic data curation / Dashboard

413 errors detected

1 comments received

Dashboard : data curation in a glance

Here we assess the data quality of all the variables in your file. The dashboard reports inconsistencies detected in the raw data.

50.3 ACCESSIBILITY

36.4 METADATA INTEGRITY

How to add comments:

1. Choose the variables to comment by selecting the checkbox in the table below

2. Click on the button "Add comment"

Show only variables with issues

Drag a column header here to group by that column

Search...

	File	Code	Necessary TOML fields missing	Variables having only NaN values	Continuous variables likely to be categorical (too few values)	Categorical variables for which the data and the TOML do not agree	Categorical variables likely to be continuous (too many values)	Continuous variables have zero standard deviation	Comments	Overall problem count (%)	Type
<input type="checkbox"/>	20180927_Dra...	F2_INT_JI		✓	✓	✓	✓	✓		0	continuous
<input type="checkbox"/>	20180927_Dra...	F2_INT_AAAA		✓	✗	✓	✓	✓		20	continuous
<input type="checkbox"/>	20180927_Dra...	F1_INT_MM		✓	✗	✓	✓	✓		20	continuous
<input type="checkbox"/>	20180927_Dra...	INT_AAAA		✓	✗	✓	✓	✓		20	continuous
<input type="checkbox"/>	20180927_Dra...	F1_INT_AAAA		✓	✗	✓	✓	✓		20	continuous
<input type="checkbox"/>	20180927_Dra...	INT_JI		✓	✓	✓	✓	✓		0	continuous
<input type="checkbox"/>	20180927_Dra...	INT_MM		✓	✗	✓	✓	✓		20	continuous
<input type="checkbox"/>	20180927_Dra...	F2_INT_MM		✓	✗	✓	✓	✓		20	continuous
<input type="checkbox"/>	20180927_Dra...	F1_INT_JI		✓	✓	✓	✓	✓		0	continuous
<input type="checkbox"/>	20180927_Dra...	F1_ALCDPON		✓	✓	✓	✓	✓		0	continuous
<input type="checkbox"/>	20180927_Dra...	F2_ALCABOF		✓	✓	✓	✓	✓		0	continuous
<input type="checkbox"/>	20180927_Dra...	F1_ALCDP2		✓	✓	✓	✓	✓		0	categorical
<input type="checkbox"/>	20180927_Dra...	F0_ALCABOF		✓	✓	✓	✓	✓		0	continuous
<input type="checkbox"/>	20180927_Dra...	F1_ALCDPOF		✓	✓	✓	✓	✓		0	continuous
<input type="checkbox"/>	20180927_Dra...	F2_ALCDPON		✓	✓	✓	✓	✓		0	continuous
<input type="checkbox"/>	20180927_Dra...	F0_ALCDPOF		✓	✓	✓	✓	✓		0	continuous
<input type="checkbox"/>	20180927_Dra...	F2_ALCDPOF		✓	✓	✓	✓	✓		0	continuous
<input type="checkbox"/>	20180927_Dra...	F2_ALCABON		✓	✓	✓	✓	✓		0	continuous
<input type="checkbox"/>	20180927_Dra...	F2_ALCDP2		✓	✓	✓	✓	✓		0	categorical
<input type="checkbox"/>	20180927_Dra...	F0_ALCDP2		✓	✓	✓	✓	✓		0	categorical

Data Quality

DACQORD Principle

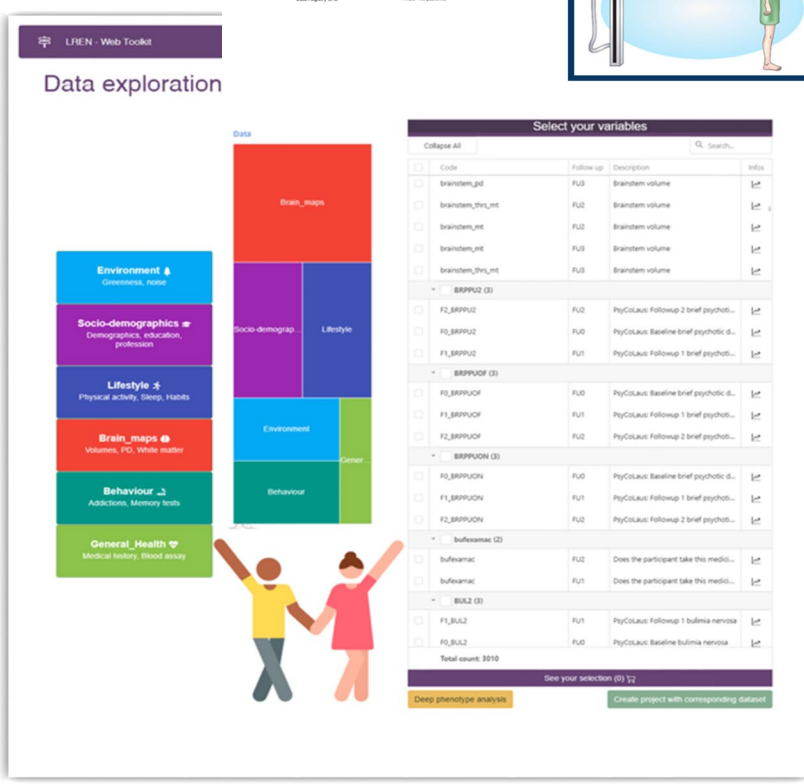
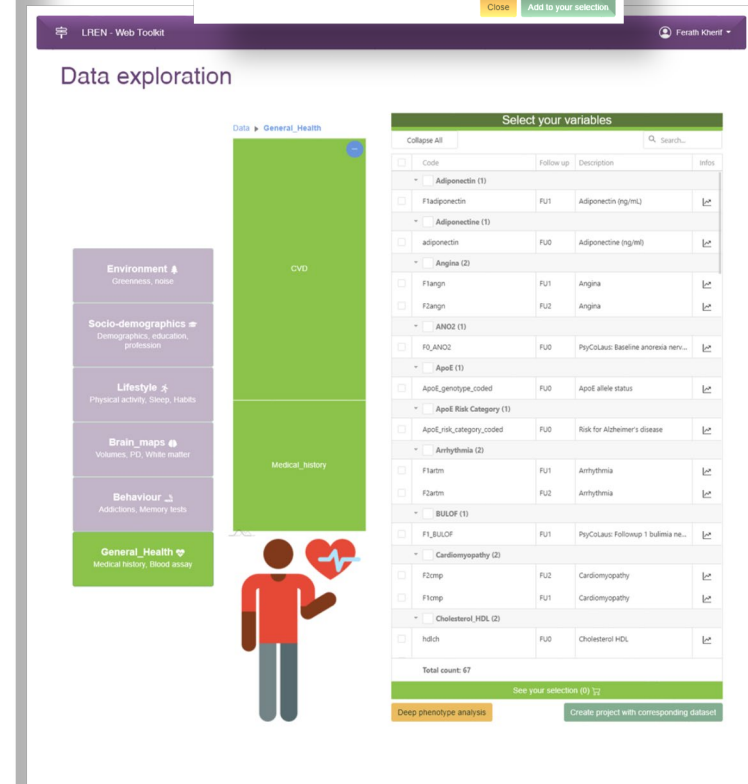
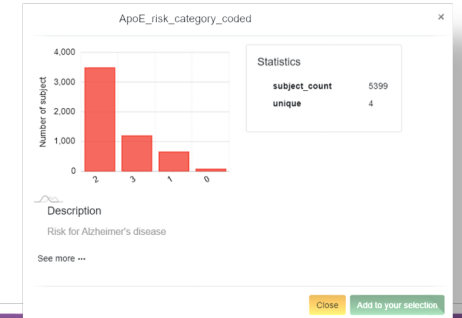
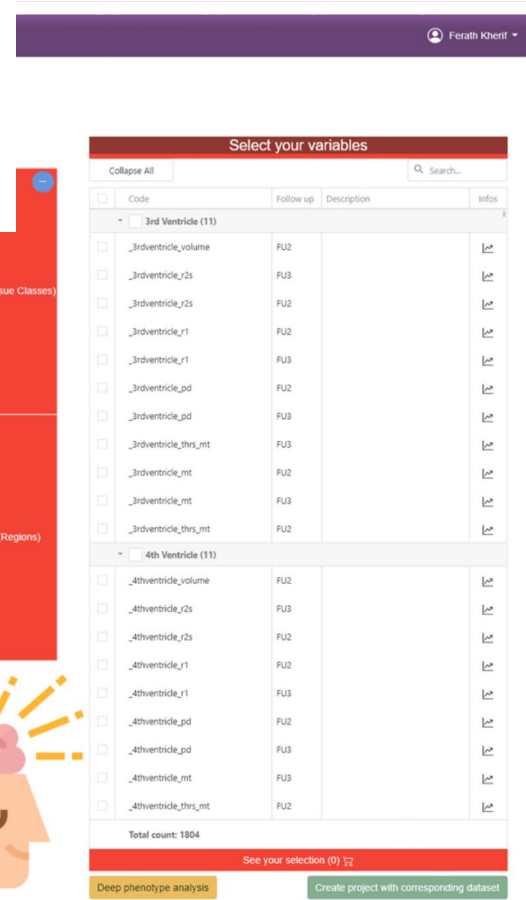
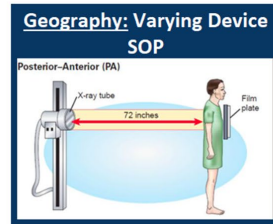
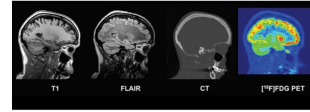
MVP

Metadata Tag

[illegible]

Metadata

- To enrich dataset
- To search dataset
- To evolve Eval data intelligence
- Map metadata to clinical relevance



ITU Tech

Thanks for your attention

Contact

Ferath.kherif@chuv.ch



AI for Health

ITU-WHO Focus Group



World Health
Organization