Common Data Model

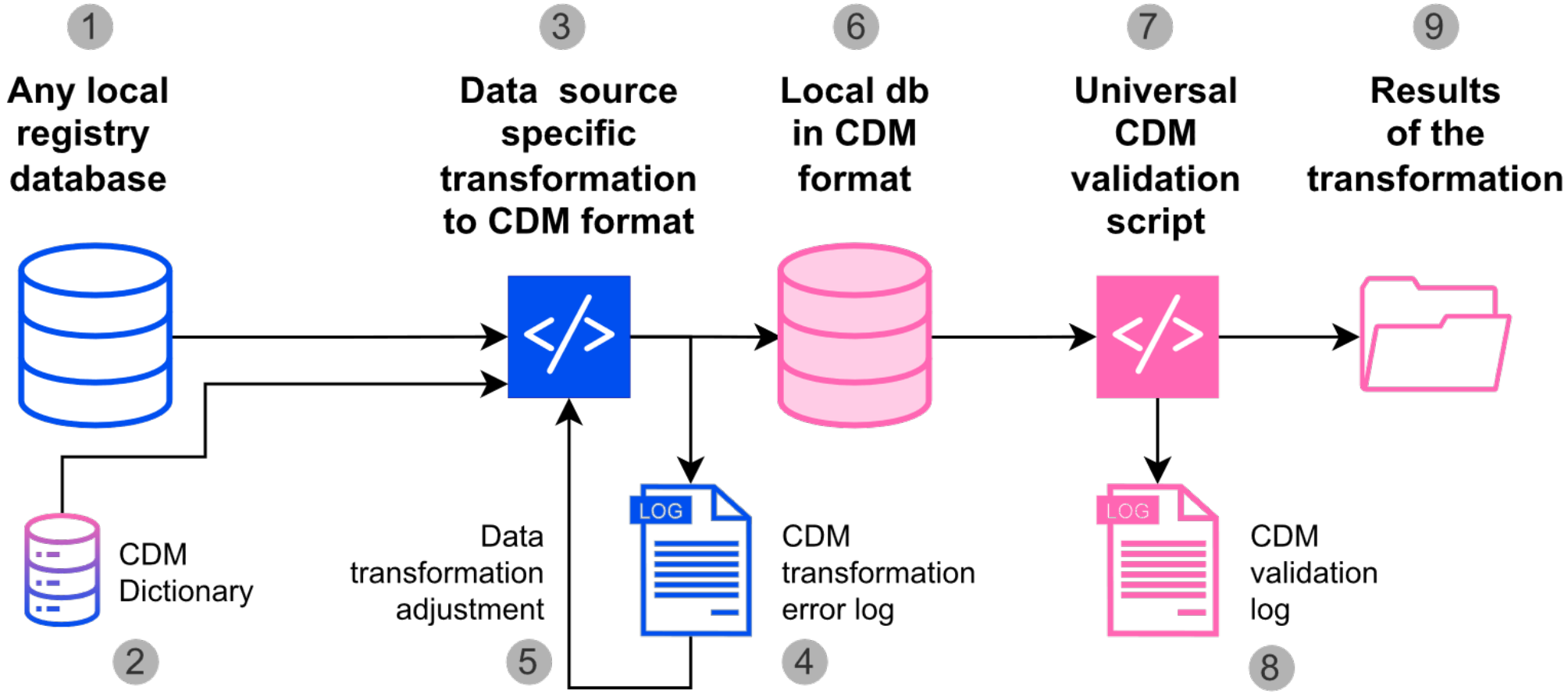
BIG MS
The big Multiple Sclerosis data network

Aims of the BMSD CDM

1	Uniform understanding of analyzed data
2	Precise data description
3	Ability to run federated / merged analysis on data converted to CDM
4	The basis for project-enhanced data models

Common Data Model

BIG MS
The big Multiple Sclerosis data network



INFORMATION	
Registry_ID	Character
Export_date	Date
CDM_transformation_timestamp	Date, time
CDM_version	Character
Responsible_person	Character

PATIENTS_EXTRAS	
Patient_ID	Character
Report_date	Date
Weight	Numeric
Height	Numeric
Smoking	Character

MRI	
Patient_ID	Character
MRI_date	Date
Brain	Logical
Spinal_cord	Logical
Brain_new_T2	Character
Brain_GD	Character
Brain_total_T2	Character
Brain_T2_worsening	Logical
SC_lesions	Character
SC_GD_lesions	Character
SC_worsening	Logical

RELAPSES	
Patient_ID	Character
Relapse_date	Date
Severity	Character
Corticosteroids_treatment	Logical

PATIENTS	
Patient_ID	Character
Birth_date	Date
Sex	Character
Date_of_onset	Date
Date_of_CIS_diagnosis	Date
Date_of_MS_diagnosis	Date
MS_diagnosis_criteria_used	Character
Phenotype_at_MS_diagnosis	Character
Date_of_SP_onset	Date
Date_of_SP_assignment	Date
Latest_phenotype	Character
MS_site_ID	Character
Region_of_MS_site	Character
Region_of_patient_residence	Character
Registry_inclusion_date	Date
Monitoring_discontinuation_date	Date
Monitoring_discontinuation_reason	Character
Cause_of_death	Character
Last_update	Date

CSF	
Patient_ID	Character
CSF_date	Date
CSF_result	Character

VISITS	
Patient_ID	Character
Visit_date	Date
Visit_type	Character
EDSS	Numeric
FSS_system	Character
Score_ambulation	Numeric
Score_bowel_bladder	Numeric
Score_brainstem	Numeric
Score_cerebellar	Numeric
Score_cerebellar_WI	Logical
Score_mental	Numeric
Score_pyramidal	Numeric
Score_sensory	Numeric
Score_visual	Numeric
Score_visual_TP	Logical

MEDICAL_CONDITIONS	
Patient_ID	Character
MedDRA_code	Character
MedDRA_name	Character
Date_of_event_start	Date
Date_of_event_end	Date
Outcome	Character
SAE	Logical
Severity	Character
Treatment_related_ATC	Character
MedDRA_code_version_used	Character
Original_dictionary_used	Character

TREATMENTS	
Patient_ID	Character
ATC_code	Character
Treatment	Character
Start_date	Date
Start_date_type	Character
End_date	Date
Reason_for_discontinuation	Character
Route_of_administration	Character

DIAGNOSTIC_RESULTS	
Patient_ID	Character
Test_date	Date
Test_name	Character
Positive	Logical
Test_value	Numeric
Test_units	Character
Test_remarks	Character

PREGNANCY	
Patient_ID	Character
Pregnancy_ID	Numeric
Outcome_ID	Numeric
Pregnancy_start	Date
Outcome_date	Date
Outcome	Character
Outcome_anomaly	Character

BIG MS

The big Multiple Sclerosis data network

BMSD Common Data Model structure

BMSD CDM v0.10,
PRELIMINARY as of 2023-07

BIG MS
The big Multiple Sclerosis data network

DIAGNOSTIC_RESULTS *Harmonized information on laboratory or clinical examination.*

Variable	Data type	Format / options	Harmonized variable principles
Patient_ID	Character	RID_xxx	*Patient_ID* to relate record in PATIENTS table.
Test_date	Date	YYYY-MM-DD	Exact or imputed date of test.*
Test_name	Character	i.e., "25FWT", "JCV", "absolute lymphocyte count", "CRP","Covid-19 PCR", "Covid-19 Ag",…	Name of the test performed. Based on project or study, this list shall be extended to needs of such study.
Positive	Logical	TRUE, FALSE, NA	Indication, wheter test had positive result or not, where applicable.
Test_value	Numeric		Numeric value used in the respective test. If any units are used, they have to be indicated in *Test_units* variable.
Test_units	Character	i.e., "s", "elements/mm3",…	If numeric value is provided, used units have to be presented here.
Test_remarks	Character		If any special test remarks are made, this variable can accomodate them.

BMSD
Common Data Model
structure

PREGNANCY *Harmonized information on pregnancy.*

Variable	Data type	Format / options	Harmonized variable principles
Patient_ID	Character	RID_xxx	*Patient_ID* to relate record in PATIENTS table.
Pregnancy_ID	Numeric	1, 2,…	ID of pregnancy in patient.
Outcome_ID	Numeric	1,2,…	If more than one fetus or child is present in pregnancy, this variable allows to identify each of them.
Pregnancy_start	Date	YYYY-MM-DD	Date of pregnancy start. If only last menstruation period is known, then this date is used.*
Outcome_date	Date	YYYY-MM-DD	Date of outcome - giving birth, abortion or miscarriage.
Outcome	Character	"Live Birth", "Stillbirth/Fetal Death", "Termination of Pregnancy for Fetal Anomaly (TOPFA)", "Miscarriage/Spontaneous Abortion", "Late Fetal Death", NA	Outcome of pregnancy, resp. specific fetus of the pregnancy. **Live Births**: This is when a baby is born alive, regardless of gestational age. **Stillbirths/Fetal Deaths**: These are pregnancies that result in the birth of a child with no signs of life at or after 20 weeks' gestation. **Terminations of Pregnancy for Fetal Anomaly (TOPFA)**: These are medical procedures performed to end a pregnancy due to detected fetal anomalies. **Miscarriages/Spontaneous Abortions**: These are pregnancies that end spontaneously before 20 weeks' gestation. **Late Fetal Deaths**: This refers to pregnancies where the fetus dies in the womb after 20 weeks' gestation, but before birth, which is not induced.
Outcome_anomaly	Character	i.e., "Q00.1", "Q90",… If more than one anomaly is present in the child, all shall be filled and separated with pipe character "|".	If any anomaly is present, then EUROCAT classification is used and present here. Entered is only ICD10 code (as of now Q00-Q99)

BIG MS
The big Multiple Sclerosis data network

BMSD Common Data Model structure

ICD10 -> MedDRA mapping

BIG MS
The big Multiple Sclerosis data network

9 630 ICD10 codes is mapped to MedDRA (2024 Jan)
provided by MedDRA MSSO for the MedDRA subscribers
https://www.meddra.org/news-and-events/news/icd-10-meddra-mapping-now-available

In PASS ongoing in ReMuS 2 795 exact ICD-10 diagnosis (4-digit) are of interest

71 % is mapped as Equivalent in MedDRA

18.5 % is having Loss of information when transforming to MedDRA

(loss is mostly in extent of disease or it's specific localization)

10.5 % is Unmapped, however, **all of them is possible to map from a higher branch of ICD:**

64 % as equivalent and

35 % with loss of information.

these diagnosis were somehow more exactly specified these groups of diseases: *Ankylosing spondylitis, Atherosclerosis, Chronic viral hepatitis, Mental retardation, Mild mental retardation, Moderate mental retardation, Profound mental retardation, Reactive arthropathy, Rheumatoid arthritis, Seropositive rheumatoid arthritis, Severe mental retardation, Type 1 diabetes mellitus, Unspecified mental retardation, Alcohol use disorder, Haematuria, Substance use disorder, Tobacco use disorder*

ICD10 -> MedDRA mapping

The big Multiple Sclerosis data network

Real analysis of the example in ReMuS

In **c.a. 1000 patients**, **10,720 ICD10 records**

Out of these **4,134** is in the list of PASS interested codes (38.6 %)

Results of mapping:

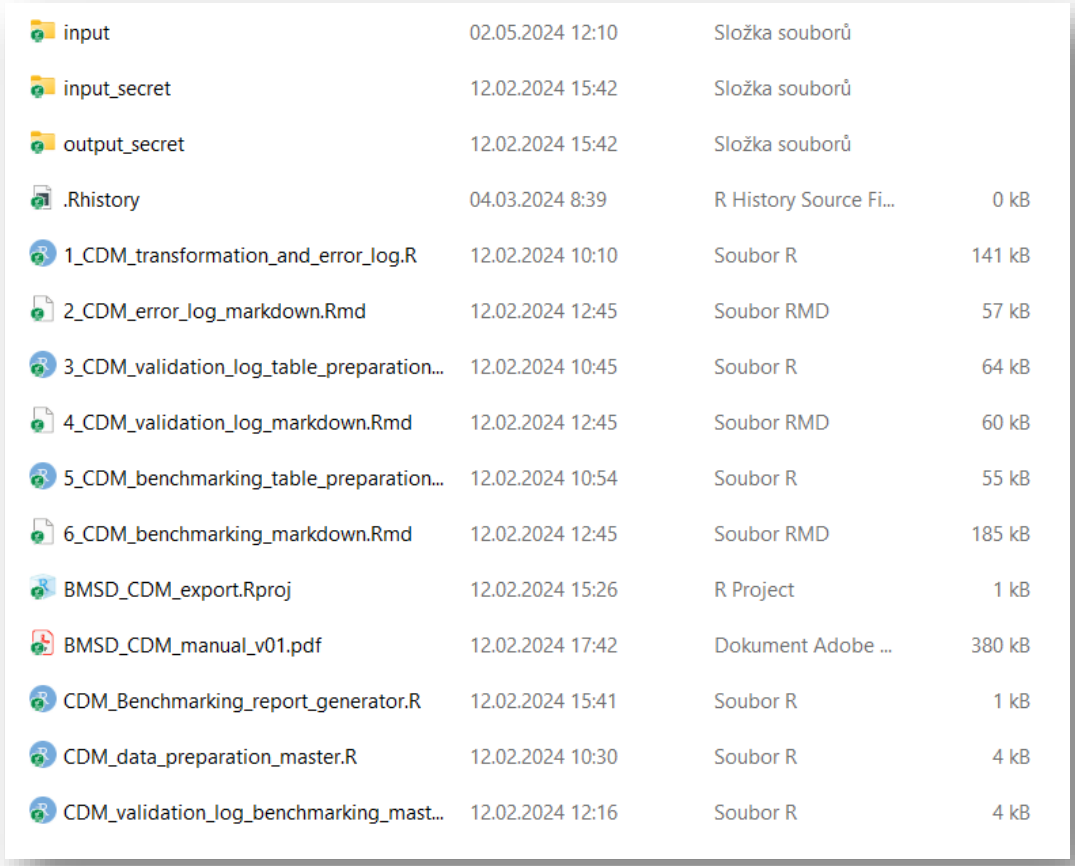
Transformation result	Count of ICD10 records	%
Equivalent	3,985	96.40%
Loss of Information	131	3.17%
Unmapped	18	0.44%
Total	**4134**	**100.00%**

Implementation

BIG MS
The big Multiple Sclerosis data network

14 751

3 750



input	02.05.2024 12:10	Složka souborů	
input_secret	12.02.2024 15:42	Složka souborů	
output_secret	12.02.2024 15:42	Složka souborů	
.Rhistory	04.03.2024 8:39	R History Source Fi...	0 kB
1_CDM_transformation_and_error_log.R	12.02.2024 10:10	Soubor R	141 kB
2_CDM_error_log_markdown.Rmd	12.02.2024 12:45	Soubor RMD	57 kB
3_CDM_validation_log_table_preparation...	12.02.2024 10:45	Soubor R	64 kB
4_CDM_validation_log_markdown.Rmd	12.02.2024 12:45	Soubor RMD	60 kB
5_CDM_benchmarking_table_preparation...	12.02.2024 10:54	Soubor R	55 kB
6_CDM_benchmarking_markdown.Rmd	12.02.2024 12:45	Soubor RMD	185 kB
BMSD_CDM_export.Rproj	12.02.2024 15:26	R Project	1 kB
BMSD_CDM_manual_v01.pdf	12.02.2024 17:42	Dokument Adobe ...	380 kB
CDM_Benchmarking_report_generator.R	12.02.2024 15:41	Soubor R	1 kB
CDM_data_preparation_master.R	12.02.2024 10:30	Soubor R	4 kB
CDM_validation_log_benchmarking_mast...	12.02.2024 12:16	Soubor R	4 kB

Implementation

BIG MS

The big Multiple Sclerosis data network

1_CDM_transformation_and_error_log.R | 2_CDM_error_log_markdown.Rmd | 3_CDM_validation_log_table_preparation.R | 4_CDM_validation_log_markdown.Rmd | 5_CDM_benchmarking_table_preparation.R | 6_CDM_benchmarking_markdown.Rmd

```
###  ____  __  __ ____  ____        _
###  | __ )(_)  __ _  |  \/  / ___|  |  _ \    __ _| |_ __ _
###  |  _ \| |/ _` | | |\/| \___ \  | | | |/ _` | __/ _` |
###  | |_) | | (_| | | |  | |___) | | |_| | (_| | || (_| |
###  |____/|_|\__, | |_|  |_|____/  |____/ \__,_|\__\__,_|
###           |___/
### ******************************************************************************
### PROJECT            : BMSD CDM data transformation and benchmarking
### FILE               : CDM_transformation_and_error_log.R
### AUTHOR             : Jiri Drahota, j.drahota@multiplesclerosis.cz
###                      Gregor Fistravec, g.fistravec@multiplesclerosis.cz
### ENCODING           : UTF-8
### BUILT ON           : R 4.3.1, RStudio 2023.06.0 Build 421
### AIM                : Transform data to CDM format and prepare error log
### DESCRIPTION / NOTES: all critically important information
### RELEVANT LINKS     :
### DEPENDENCIES       :
### ..............................................................................
### QUALITY CONTROL PROCESS
###    LEVEL           : Non-critical / Highly important / Critical
###    TECHNIQUE DONE  : NA / Code review / Double programming / etc.
###    TEAM            : NA / QC responsible team member and team member names
###    DATE            : NA / QC team names
###    RESULT          : QC PASSED / QC NOT PASSED / + remarks
### ..............................................................................
### VERSION LOG        : 2024-02-07 v1.0
###                                 - initial version
###                      : YYYY-MM-DD vXX
###                                 - modifications done
### ..............................................................................
### DISCLAIMER        : Script for data transformation and report generation
###                     was created within ReMuS | The Czech Republic Multiple
###                     Sclerosis Patient Registry by Gregor Fistravec and
###                     Jiri Drahota © 2024
###
### ******************************************************************************

### / / / / / / / / / / / / / / / /
### Script init ----
### / / / / / / / / / / / / / / / /
cat('\f')       # Clear console
set.seed(8524)  # Set seed for reproducibility

### / / / / / / / / / / / / / / / /
### Libraries ----
### / / / / / / / / / / / / / / / /
library(glue)
library(dplyr)
library(lubridate)
library(openxlsx)
```

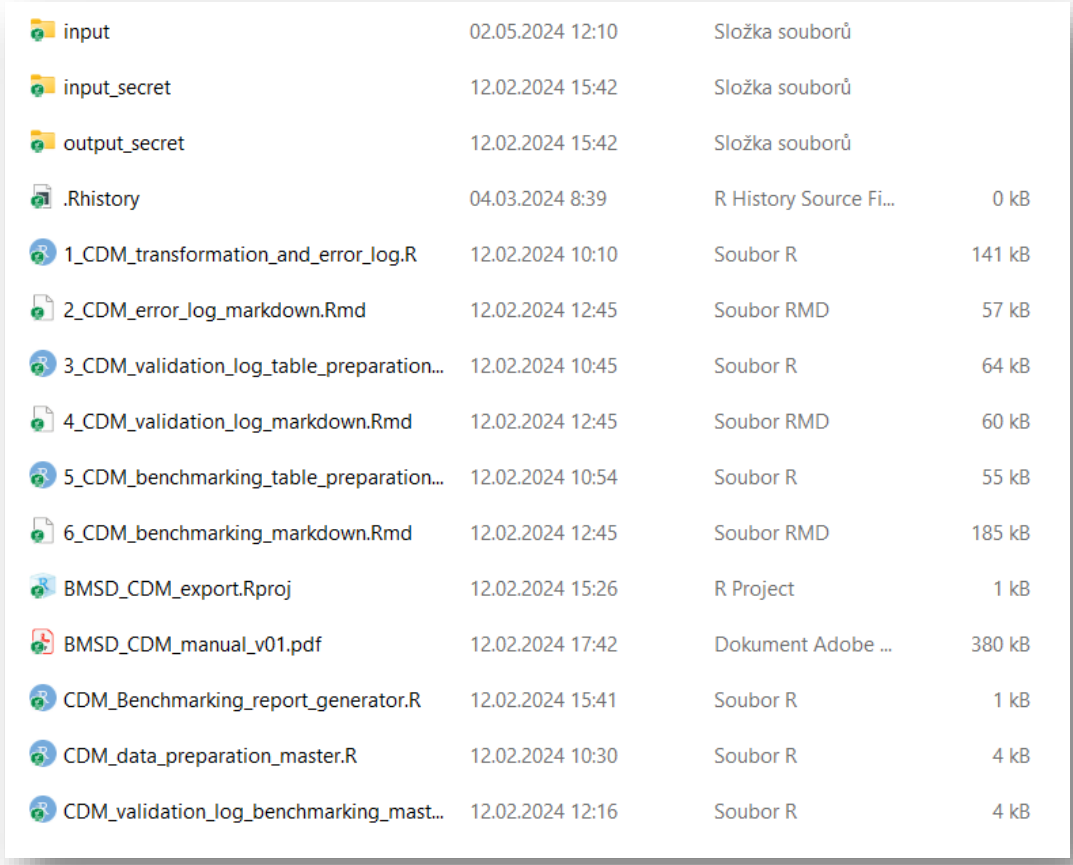
Implementation

BIG MS
The big Multiple Sclerosis data network

ReMuS®
THE CZECH REPUBLIC
MULTIPLE SCLEROSIS
PATIENT REGISTRY

MSBase
Neuro-Immunology Registry

Svenska
neuroregister



input	02.05.2024 12:10	Složka souborů	
input_secret	12.02.2024 15:42	Složka souborů	
output_secret	12.02.2024 15:42	Složka souborů	
.Rhistory	04.03.2024 8:39	R History Source Fi...	0 kB
1_CDM_transformation_and_error_log.R	12.02.2024 10:10	Soubor R	141 kB
2_CDM_error_log_markdown.Rmd	12.02.2024 12:45	Soubor RMD	57 kB
3_CDM_validation_log_table_preparation...	12.02.2024 10:45	Soubor R	64 kB
4_CDM_validation_log_markdown.Rmd	12.02.2024 12:45	Soubor RMD	60 kB
5_CDM_benchmarking_table_preparation...	12.02.2024 10:54	Soubor R	55 kB
6_CDM_benchmarking_markdown.Rmd	12.02.2024 12:45	Soubor RMD	185 kB
BMSD_CDM_export.Rproj	12.02.2024 15:26	R Project	1 kB
BMSD_CDM_manual_v01.pdf	12.02.2024 17:42	Dokument Adobe ...	380 kB
CDM_Benchmarking_report_generator.R	12.02.2024 15:41	Soubor R	1 kB
CDM_data_preparation_master.R	12.02.2024 10:30	Soubor R	4 kB
CDM_validation_log_benchmarking_mast...	12.02.2024 12:16	Soubor R	4 kB

Next steps

- Kick-off meeting with data managers of the onboarding registries
  (in charge: Gregor Fištravec from ReMuS)
- Customisation (text, explanatory notes)
- Test conversions to the CDM format
- Testing pooling / federated analysis