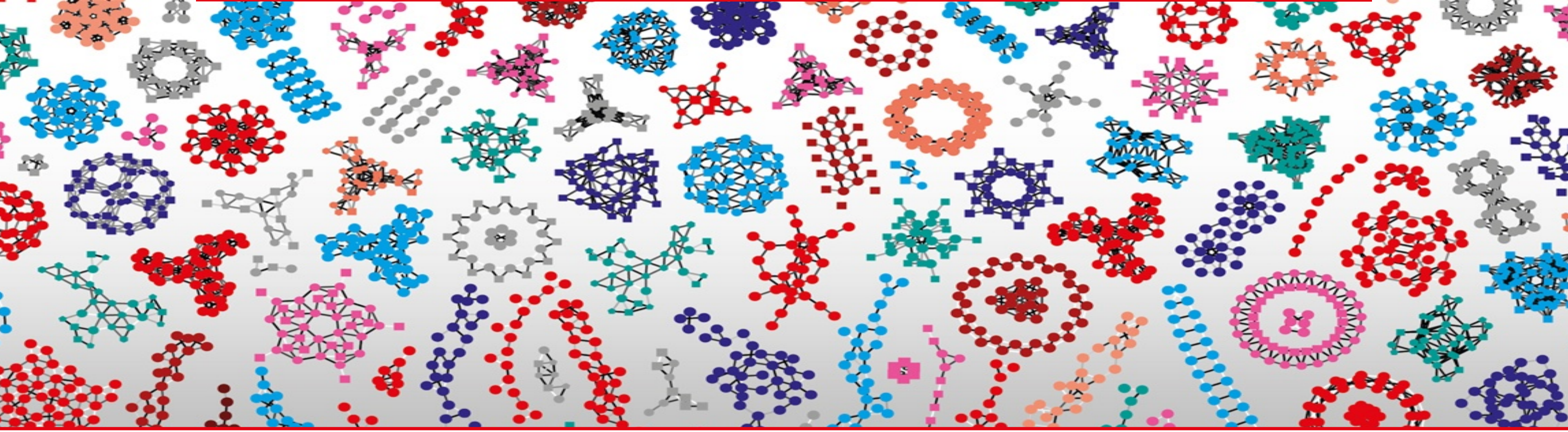


Conversion de données cliniques provenant d’uni et d’hôpitaux pour permettre l’interopérabilité et l’analyse à grande échelle

SIB
Swiss Institute of Bioinformatics

Frédéric Burdet & team, Vital-IT group, SIB Swiss Institute of Bioinformatics

24.06.2019, GUF CDISC, Paris

elixir SWITZERLAND

Contents

- **Présentation du SIB et de Vital-IT**
- Pourquoi harmoniser et convertir les données? Présentation du système d'analyse fédéré
- Processus de conversion, notre utilisation de SDTM
- Exemple d'utilisation du système fédéré

SIB in brief

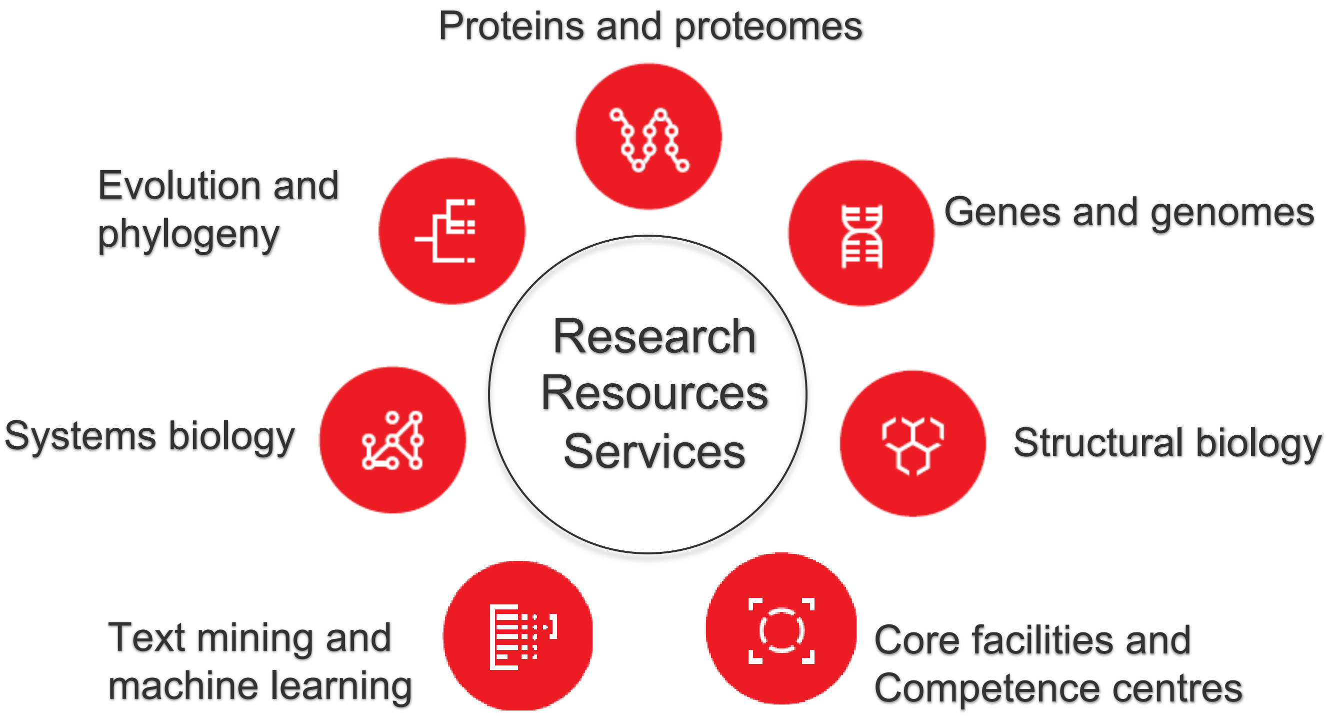
70 groups

800 scientists

20 partner institutions

95 bioinformaticians per million inhabitants

A complete and diverse activity scope



Research Resources Services

Proteins and proteomes

Genes and genomes

Structural biology

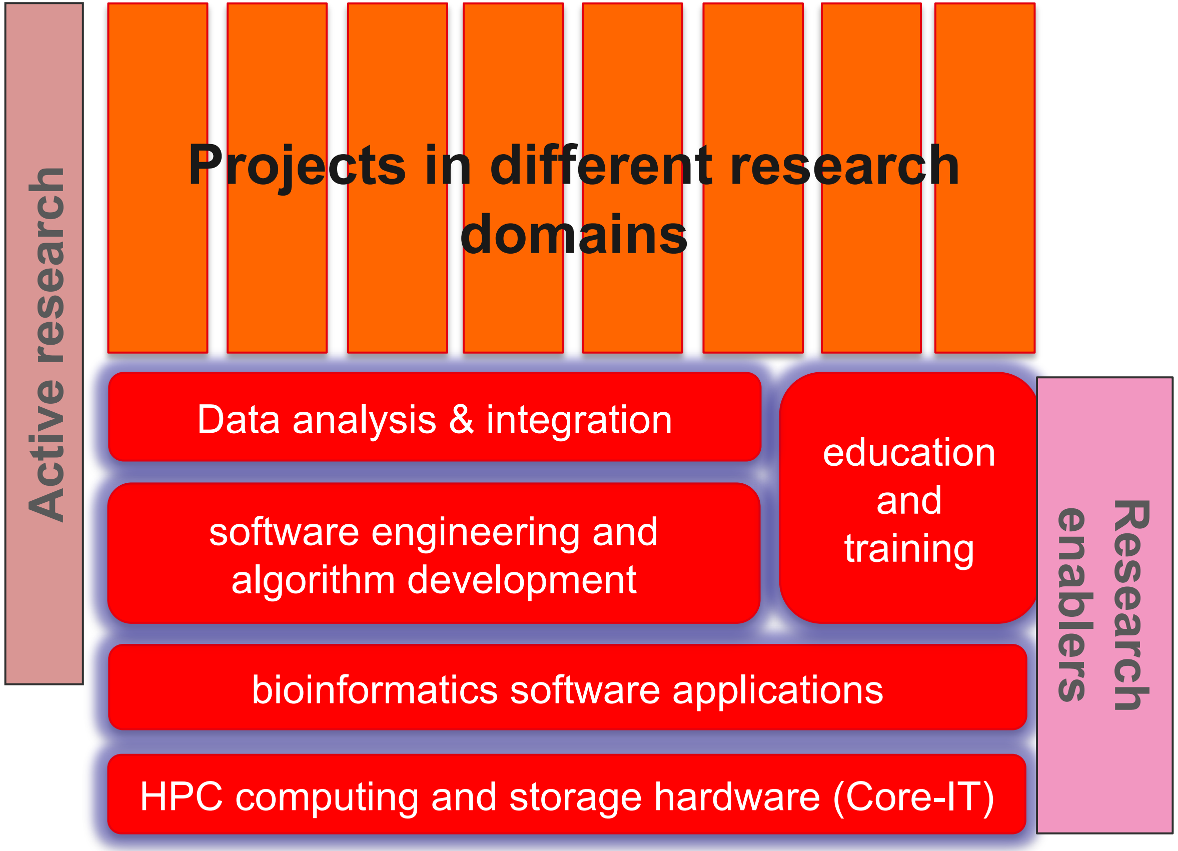
Core facilities and Competence centres

Text mining and machine learning

Systems biology

Evolution and phylogeny

Vital-IT is an enabler and driver of life science research



Active research

Projects in different research domains

Data analysis & integration

software engineering and algorithm development

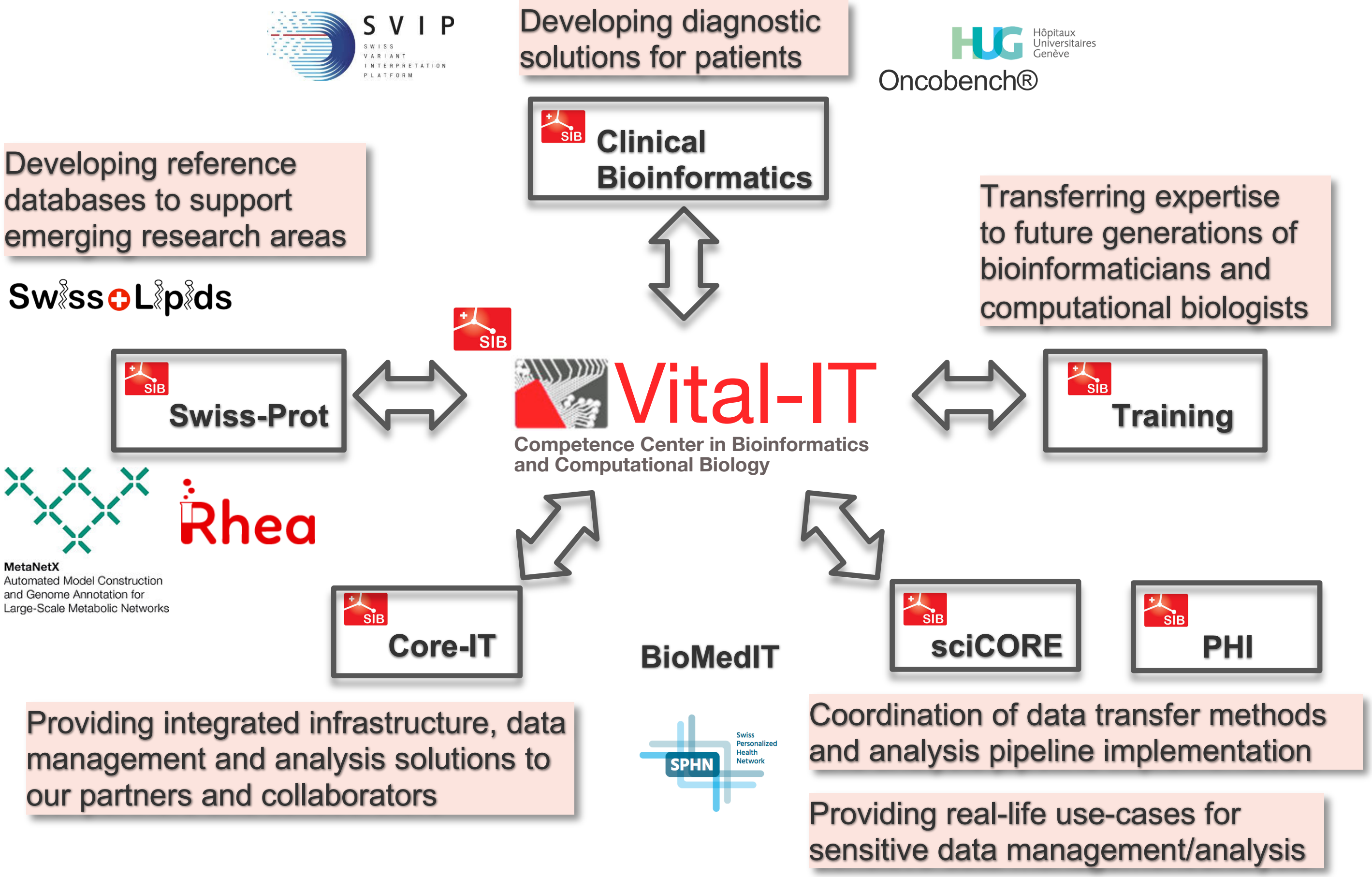
education and training

bioinformatics software applications

HPC computing and storage hardware (Core-IT)

Research enablers

Continued support of SIB-wide activities



SVIP Swiss Variant Interpretation Platform

Developing diagnostic solutions for patients

HUG Hôpitaux Universitaires Genève

Oncobench®

SIB Clinical Bioinformatics

Developing reference databases to support emerging research areas

SwissLipids

Transferring expertise to future generations of bioinformaticians and computational biologists

SIB Swiss-Prot

SIB Vital-IT

Competence Center in Bioinformatics and Computational Biology

SIB Training

MetaNetX

Automated Model Construction and Genome Annotation for Large-Scale Metabolic Networks

Rhea

SIB Core-IT

BioMedIT

SIB sciCORE

SIB PHI

Providing integrated infrastructure, data management and analysis solutions to our partners and collaborators

SPHN Swiss Personalized Health Network

Coordination of data transfer methods and analysis pipeline implementation

Providing real-life use-cases for sensitive data management/analysis

Projets IMI pour lesquels nous avons converti des données en CDISC

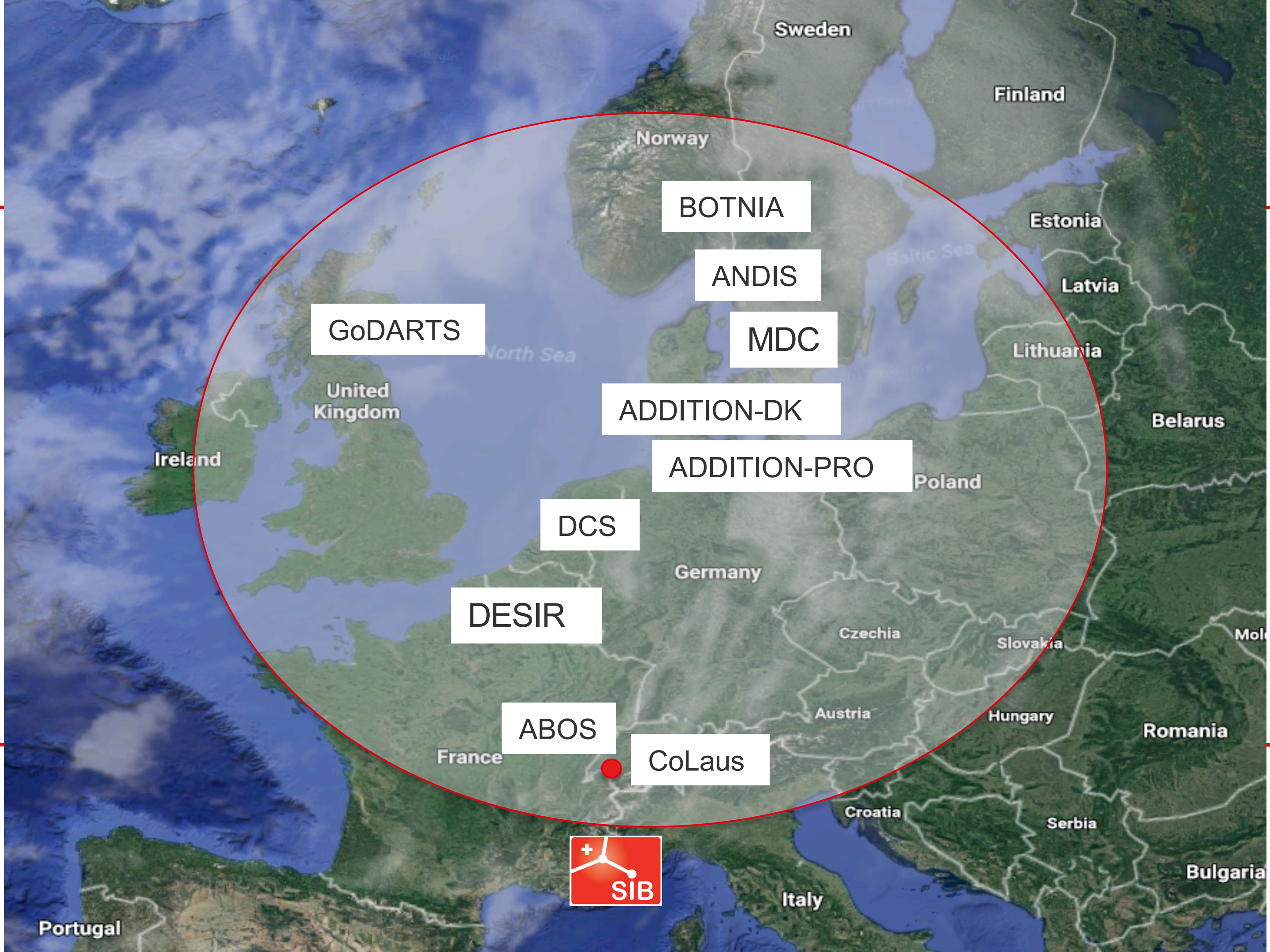
- **IMI :** Innovative Medicines Initiative, **EU public-private** partnership funding health research and innovation

- **RHAPSODY**: Assessing risk and progression of pre-diabetes and type 2 diabetes to enable disease modification
  - **10 cohortes**

- **BEAt-DKD**: Biomarker Enterprise to Attack Diabetic Kidney Disease
  - **5 cohortes**

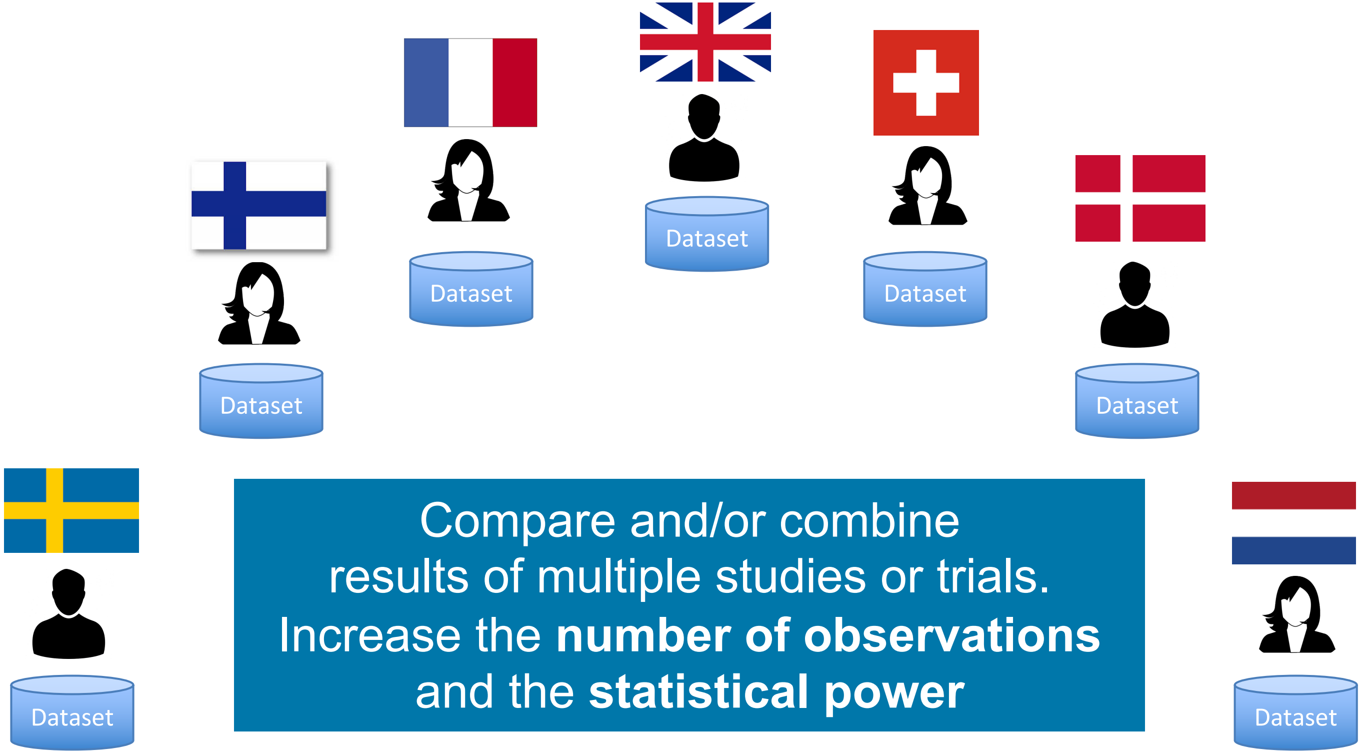
Contenu

- Présentation du SIB et de Vital-IT
- **Pourquoi harmoniser et convertir les données? Présentation du système d’analyse fédéré**
- Processus de conversion, notre utilisation de SDTM
- Exemple d’utilisation du système fédéré



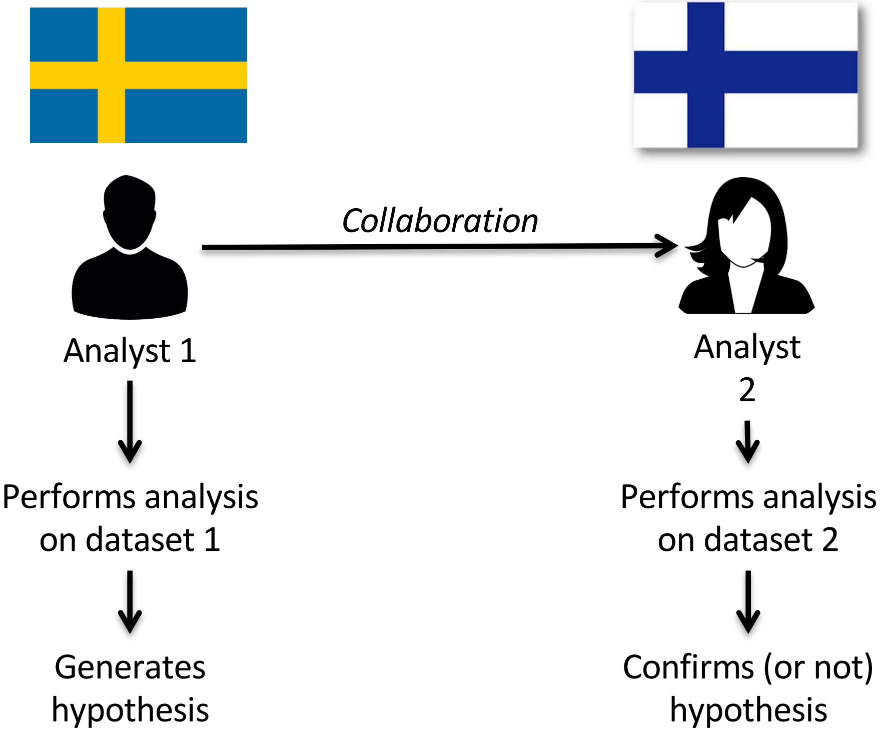
Sweden
Finland
Norway
Estonia
Latvia
Lithuania
Belarus
Poland
United Kingdom
Ireland
North Sea
Baltic Sea
Germany
Czechia
Slovakia
Austria
Hungary
Romania
Mol
France
Croatia
Serbia
Bulgaria
Italy
Portugal
BOTNIA
ANDIS
MDC
GoDARTS
ADDITION-DK
ADDITION-PRO
DCS
DESIR
ABOS
CoLaus
SIB

Meta-analysis is necessary to gain analysis power



Compare and/or combine
results of multiple studies or trials.
Increase the **number of observations**
and the **statistical power**

Ethical and/or **legal/governance** constraints on clinical cohort data mean that often sensitive individual-level (patient) data **cannot be shared** or **copied and cannot be analysed together in a centralized way**



Analyst 1

Performs analysis on dataset 1

Generates hypothesis

Collaboration

Analyst 2

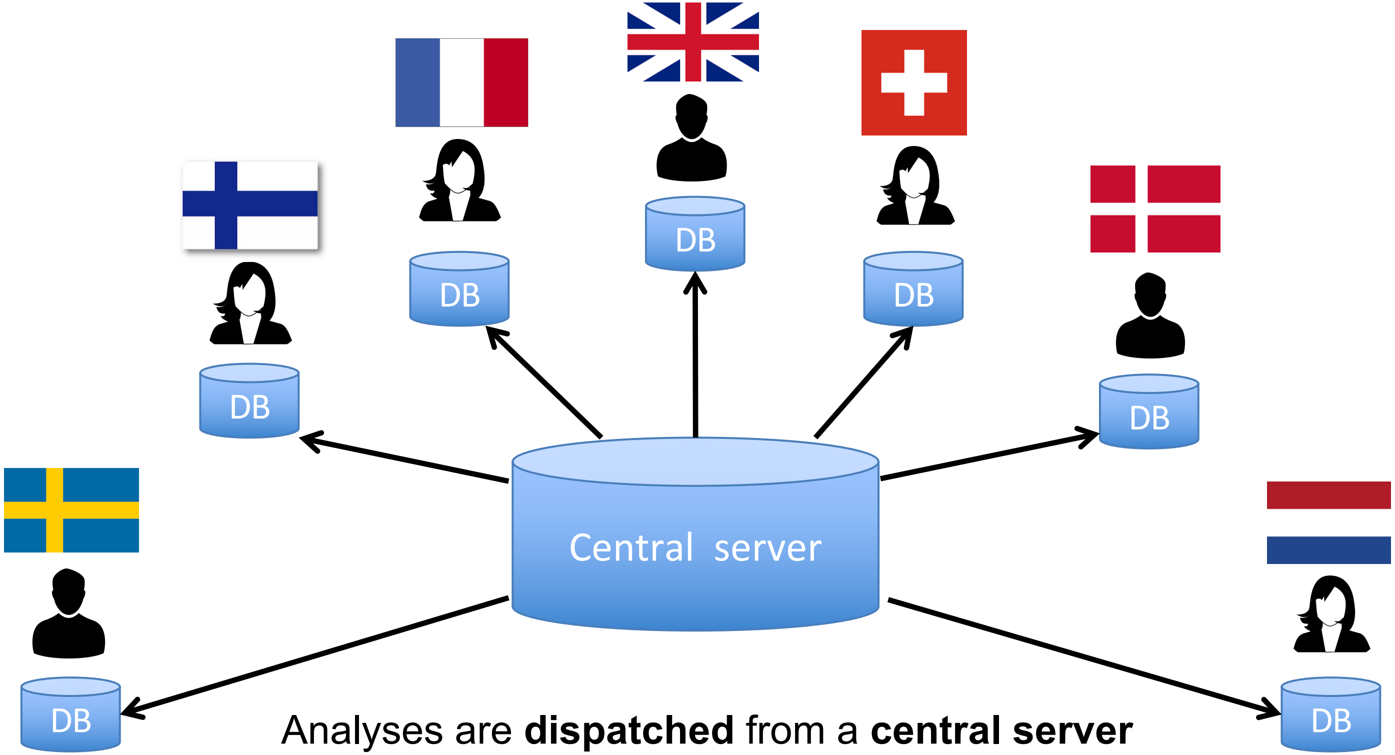
Performs analysis on dataset 2

Confirms (or not) hypothesis

Problèmes de cette méthode ...

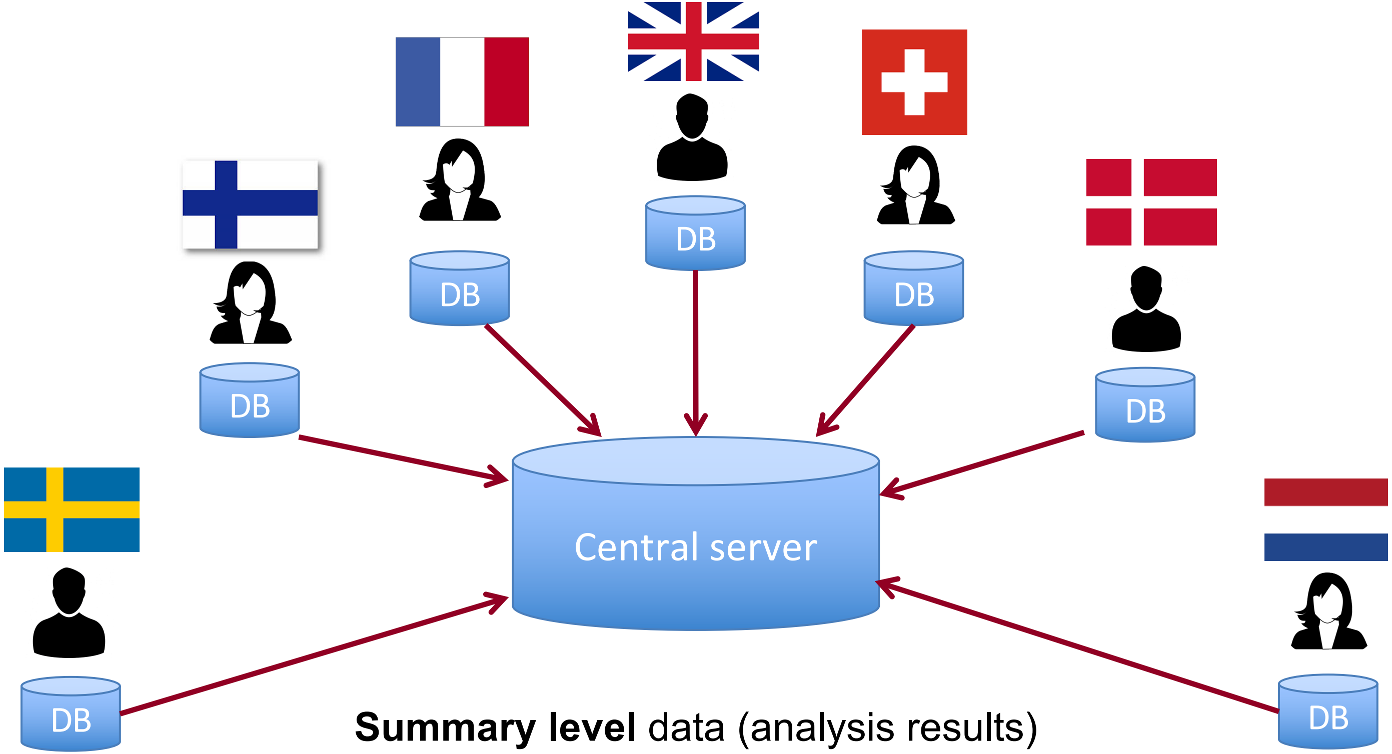
- Les noms de variables et la structure de donnée sont différents, on doit donc faire des analyses séparées
- Les mesures dans chaque set de donnée ne sont pas forcément similaires ou équivalentes
- Il est donc difficile de comparer les analyses, donc d'évaluer si les résultats sont comparables

Federated analysis is a possible solution



Analyses are **dispatched** from a **central server**
and run on each **local server**

Federated analysis is a possible solution



Summary level data (analysis results)
are **sent back** to the **central server**

Avantages du système d’analyse fédéré

- L’analyse peut être effectuée **sans copier les données** sur un autre serveur, évitant d’éventuels problèmes éthiques ou de régulation

- Les data managers et administrateurs système locaux **gardent le contrôle** sur l’utilisation de leurs données

- Les analyses statistiques peuvent être **standardisées à travers les études / cohortes** (p.ex. les méthodes d’analyse, la gestion des variables continues, des time points, …)

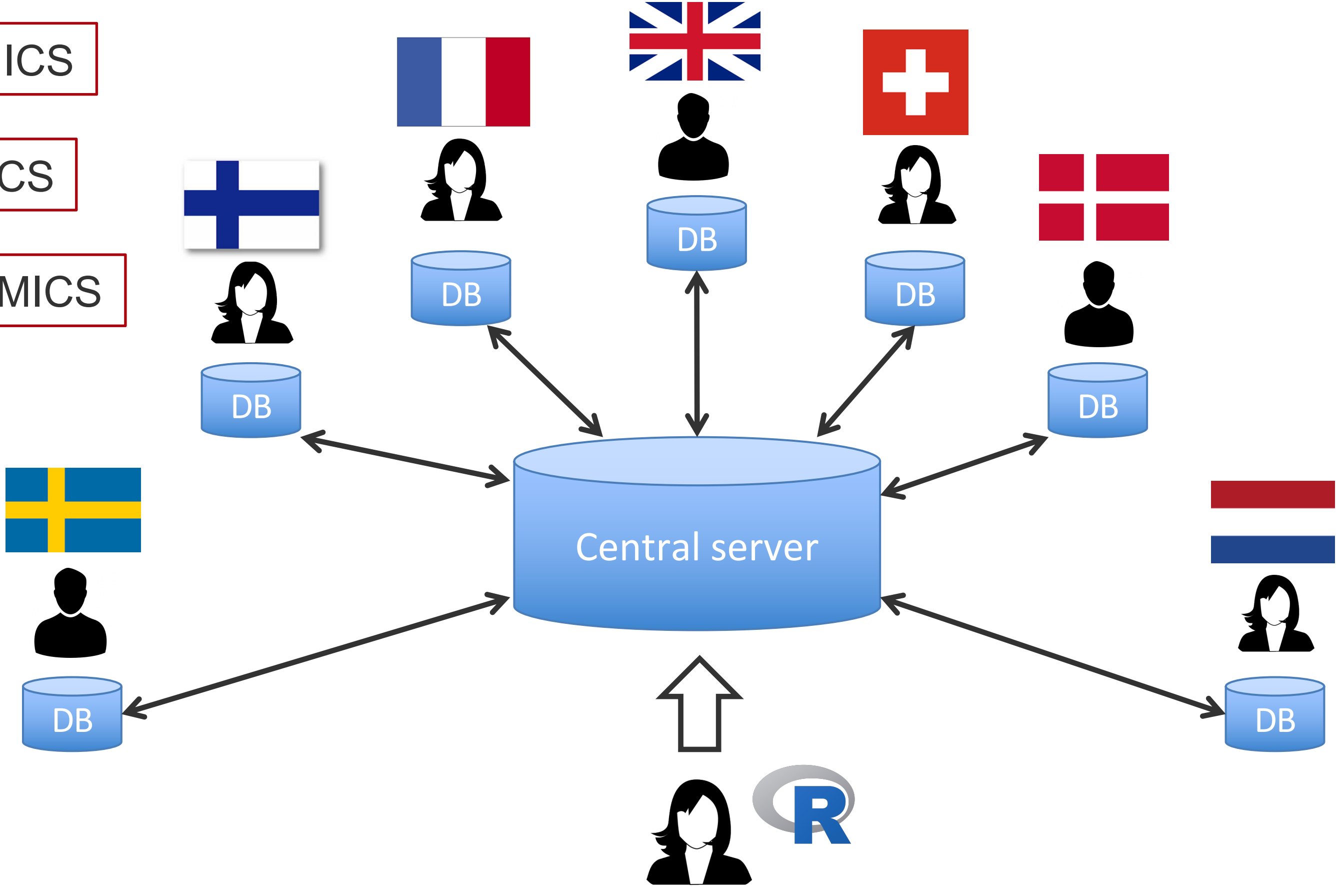
- **Accès à l’ensemble des variables**, contrairement à un sous-ensemble lors d’un transfers => plus flexible

New data can be added and accessed

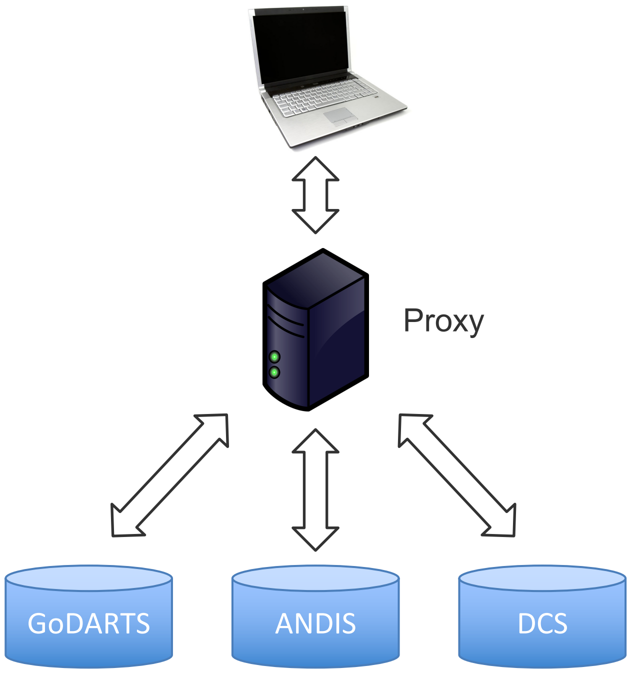
PROTEOMICS

LIPIDOMICS

METABOLOMICS



Two modes of federated analysis are possible



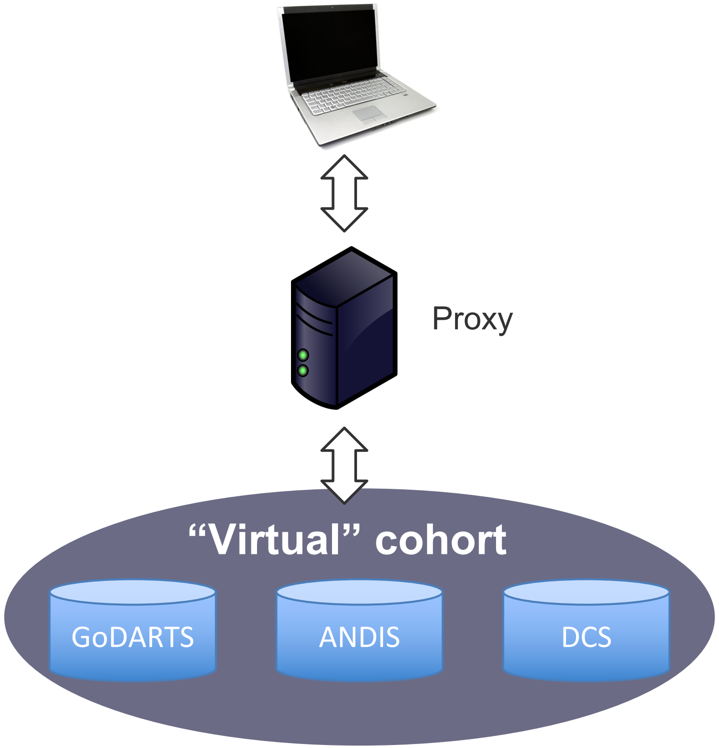
Proxy

GoDARTS

ANDIS

DCS

REPLICATION MODE:
The same analysis is performed on each cohort



Proxy

“Virtual” cohort

GoDARTS

ANDIS

DCS

VIRTUAL COHORT MODE:
The analysis is performed as if there is a single cohort

Contenu

- Présentation du SIB et de Vital-IT
- Pourquoi harmoniser et convertir les données? Présentation du système d'analyse fédéré
- **Processus de conversion, notre utilisation de SDTM**
- Exemple d'utilisation du système fédéré

Pourquoi harmoniser les données?

- Pour s’assurer que les mesures entre cohortes peuvent être **comparées** et **analysées,** avec un format commun

- CDISC – SDTM semblait la meilleure option
  - Obligatoire pour les nouvelles études
  - Recommandé par l’IMI
  - Possibilité d’intégrer facilement des cohortes originellement capturées en SDTM par la suite

Three main steps for setting up the federated database

1. Cohort data harmonization in SDTM
    1. Mapping
    2. Conversion
    3. Vérification

2. Set up of remote IT infrastructure and loading of harmonised data on each node

3. Development of software for accessing the data and performing remote analysis

Challenges techniques

- Pour des raisons de privacité des données, nous ne recevons généralement des données que pour 10 ou 20 patients

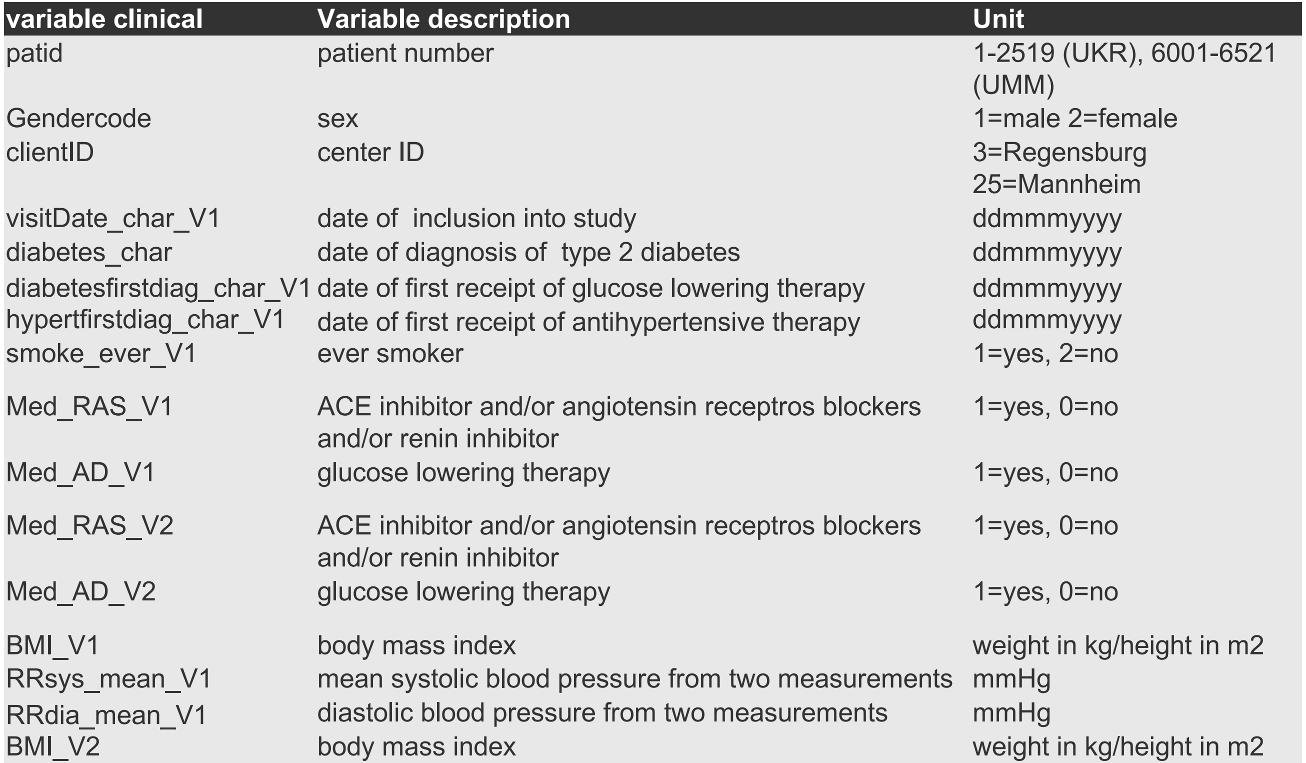
- Les personnes sur les sites universitaires / hopitaux
  - Ne sont pas data manager
  - N'ont souvent pas beaucoup de temps à consacrer à la conversion

- Les données sont souvent dans une seule table, à assigner dans les domaines SDTM
- Les métadonnées sont à "collecter"

Harmonisation “Jamborees”

Données reçues

- Liste des variables



variable clinical	Variable description	Unit
patid	patient number	1-2519 (UKR), 6001-6521 (UMM)
Gendercode	sex	1=male 2=female
clientID	center ID	3=Regensburg 25=Mannheim
visitDate_char_V1	date of inclusion into study	ddmmmyyyy
diabetes_char	date of diagnosis of type 2 diabetes	ddmmmyyyy
diabetesfirstdiag_char_V1	date of first receipt of glucose lowering therapy	ddmmmyyyy
hypertfirstdiag_char_V1	date of first receipt of antihypertensive therapy	ddmmmyyyy
smoke_ever_V1	ever smoker	1=yes, 2=no
Med_RAS_V1	ACE inhibitor and/or angiotensin receptros blockers and/or renin inhibitor	1=yes, 0=no
Med_AD_V1	glucose lowering therapy	1=yes, 0=no
Med_RAS_V2	ACE inhibitor and/or angiotensin receptros blockers and/or renin inhibitor	1=yes, 0=no
Med_AD_V2	glucose lowering therapy	1=yes, 0=no
BMI_V1	body mass index	weight in kg/height in m2
RRsys_mean_V1	mean systolic blood pressure from two measurements	mmHg
RRdia_mean_V1	diastolic blood pressure from two measurements	mmHg
BMI_V2	body mass index	weight in kg/height in m2

- ”Dummy” data: souvent en format excel, format large

Pseudo-code pour le mapping

- Peut-être facilement utilisé par les curateurs
- Peut-être importé en R ou autre langage de programmation pour la conversion
- Intègre le code de mapping et les méta-données en plus de la CDISC Variable Name:
    - **CDISC Variable Mapping**: p. ex "1=Y,2=N,NA=NA"
    - **Associated CDISC Variables**: p.ex "LBTESTCD=GLUC;LBSPEC=PLASMA;VISIT=BASELINE"

Mapping

variable clinical	Variable description	Unit	CDISC Variable Name	CDISC Variable Mapping	Associated CDISC Variables	CDISC table
patid	patient number	1-2519 (UKR), 6001-6521 (UMM)	USUBJID	NA	NA	ALL
Gendercode	sex	1=male 2=female	SEX	1=M,2=F,NA=NA	NA	DM
clientID	center ID	3=Regensburg 25=Mannheim	SITEID	38=Regensburg,39=Mannheim	NA	DM
visitDate_char_V1	date of inclusion into study	ddmmmyyyy	DMDTC	NA	VISIT=BASELINE	DM
diabetes_char	date of diagnosis of type 2 diabetes	ddmmmyyyy	MHDTC	NA	MHTERM=TYPE 2 DIABETES	MH
diabetesfirstdiag_char_V1	date of first receipt of glucose lowering therapy	ddmmmyyyy	MHDTC	NA		
hypertfirstdiag_char_V1	date of first receipt of antihypertensive therapy	ddmmmyyyy	MHDTC	NA	MHTERM=HYPERTENSION	MH
smoke_ever_V1	ever smoker	1=yes, 2=no	SUCAT::SUOCCUR	1&2&NA=TOBACCO_FORMER::1=Y,2=N,NA=NA	VISIT=BASELINE	SU
Med_RAS_V1	ACE inhibitor and/or angiotensin receptros blockers and/or renin inhibitor	1=yes, 0=no	CMCAT::CMOCCUR	0&1&NA=AGENTS ACTING ON THE RENIN-ANGIOTENSIN SYSTEM::0=N,1=Y,NA=NA	VISIT=BASELINE	CM
Med_AD_V1	glucose lowering therapy	1=yes, 0=no	CMCAT::CMOCCUR	0&1&NA=BLOOD GLUCOSE LOWERING DRUGS::0=N,1=Y,NA=NA	VISIT=BASELINE	CM
Med_RAS_V2	ACE inhibitor and/or angiotensin receptros blockers and/or renin inhibitor	1=yes, 0=no	CMCAT::CMOCCUR	0&1&NA=AGENTS ACTING ON THE RENIN-ANGIOTENSIN SYSTEM::0=N,1=Y,NA=NA	VISIT=1	CM
Med_AD_V2	glucose lowering therapy	1=yes, 0=no	CMCAT::CMOCCUR	0&1&NA=BLOOD GLUCOSE LOWERING DRUGS::0=N,1=Y,NA=NA	VISIT=1	CM
BMI_V1	body mass index	weight in kg/height in m2	VSORRES	NA	VSTESTCD=BMI;VISIT=BASELINE	VS
RRsys_mean_V1	mean systolic blood pressure from two measurements	mmHg	VSORRES	NA	VSTESTCD=SYSBP;VISIT=BASELINE	VS
RRdia_mean_V1	diastolic blood pressure from two measurements	mmHg	VSORRES	NA	VSTESTCD=DIABP;VISIT=BASELINE	VS
BMI_V2	body mass index	weight in kg/height in m2	VSORRES	NA	VSTESTCD=BMI;VISIT=VISIT1	VS

Explications du mapping + pseudocode

Original variable	Description	Values	SDTM variable	SDTM mapping	Associated SDTM	Domain
smoke_ever_V1	ever smoker	1=yes, 2=no	SUCAT::SUOCCUR	1&2&NA=TOBACCO_FORMER::1=Y,2=N,NA=NA	VISIT=BASELINE	SU
Med_AD_V1	glucose lowering therapy	1=yes, 0=no	CMCAT::CMOCCUR	0&1&NA=BLOOD GLUCOSE LOWERING DRUGS::0=N,1=Y,NA=NA	VISIT=BASELINE	CM
BMI_V1	body mass index	weight in kg/height in m2	VSORRES	NA	VSTESTCD=BMI;VISIT=BASELINE	VS
BMI_V2	body mass index	weight in kg/height in m2	VSORRES	NA	VSTESTCD=BMI;VISIT=VISIT1	VS

Tables SDTM utilisées

SDTM Tables

DM: demographics

LB: laboratory test results

CM: concomitant medication (i.e. treatments)

MH: medical history (i.e. conditions & diseases)

VS: vital signs (e.g. weight, height, BMI)

SU: substance use (e.g. tobacco)

APMH: associated person medical history

Conversion en R, exemples

- Utilisation de listes pour ajouter les métadonnées

```
"genderCode": {
    "SEX": {
      "1": ["M"],
      "2": ["F"],
      "NA": ["NA"]
    }
  },
  "clientId": {
    "SITEID": {
      "38": ["Regensburg"],
      "39": ["Mannheim"]
    }
  }
```

- Utilisation de “merge” (par exemple pour les dates / codes de visites), de “melt” pour passer de données en large à en long

Melt

patNr	glukosekorr_V1	glukosekorr_V2	glukosekorr_V3	HbA1c_percent_V1	HbA1c_percent_V2	HbA1c_percent_V3	CRP_V1	CHOL_V1
1	156	159.07	250	8.00598	8.00598	9.103976	0.76	228.46
2	165	214	138	7.91448	11.025469	6.907983	5.77	270.87
3	142	150.26	NA	7.639981	8.00598	NA	2.81	192.21
…	100	103.42	NA	5.992987	5.901487	NA	3.28	217.74
…	118	105.49	NA	6.999483	6.907983	NA	0.74	201.24
…	146	NA	NA	7.182482	NA	NA	2.34	284.32

patNr	variable	value
1	glukosekorr_V1	156
1	glukosekorr_V2	159.07
1	glukosekorr_V3	250
1	HbA1c_percent_V1	8.00598
1	HbA1c_percent_V2	8.00598
1	HbA1c_percent_V3	9.103976
1	CRP_V1	0.76
1	CHOL_V1	228.46

Ajouter les méta-données en utilisant la liste de mapping

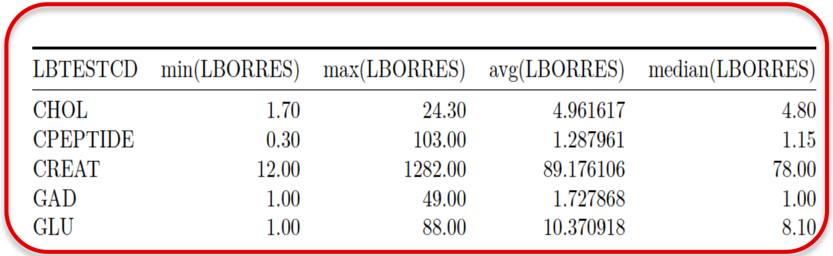
patNr	variable	value	LBORRES	LBTESTCD	LBSPEC	VISIT
1	glukosekorr_V1	156	156	GLUC	PLASMA	BASELINE
1	glukosekorr_V2	159.07	159.07	GLUC	PLASMA	VISIT1
1	glukosekorr_V3	250	250	GLUC	PLASMA	VISIT2
1	HbA1c_percent_V1	8.00598	8.00598	HBA1C	BLOOD	BASELINE
1	HbA1c_percent_V2	8.00598	8.00598	HBA1C	BLOOD	VISIT1
1	HbA1c_percent_V3	9.103976	9.103976	HBA1C	BLOOD	VISIT2
1	CRP_V1	0.76	0.76	CRP	SERUM	BASELINE
1	CHOL_V1	228.46	228.46	CHOL	SERUM	BASELINE

Puis merger avec la table des dates :

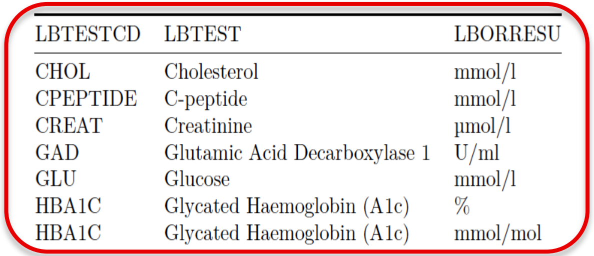
visitOrig	patNr	date	LBDTC	VISIT
visitDate_char_V1	1	03-Feb-10	2010-02-03	BASELINE
visitDate_char_V1	2	04-Feb-10	2010-02-04	BASELINE
visitDate_char_V1	3	18-Feb-10	2010-02-18	BASELINE
visitDate_char_V1	4	18-Feb-10	2010-02-18	BASELINE
visitDate_char_V1	5	19-Feb-10	2010-02-19	BASELINE
visitDate_char_V1	7	22-Feb-10	2010-02-22	BASELINE

Vérification de la conversion

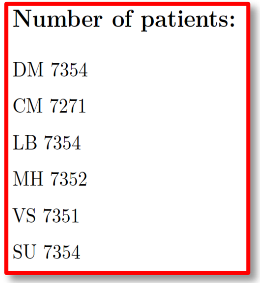
Diagnostic script



LBTESTCD	min(LBORRES)	max(LBORRES)	avg(LBORRES)	median(LBORRES)
CHOL	1.70	24.30	4.961617	4.80
CPEPTIDE	0.30	103.00	1.287961	1.15
CREAT	12.00	1282.00	89.176106	78.00
GAD	1.00	49.00	1.727868	1.00
GLU	1.00	88.00	10.370918	8.10



LBTESTCD	LBTEST	LBORRESU
CHOL	Cholesterol	mmol/l
CPEPTIDE	C-peptide	mmol/l
CREAT	Creatinine	µmol/l
GAD	Glutamic Acid Decarboxylase 1	U/ml
GLU	Glucose	mmol/l
HBA1C	Glycated Haemoglobin (A1c)	%
HBA1C	Glycated Haemoglobin (A1c)	mmol/mol



Number of patients:

DM 7354

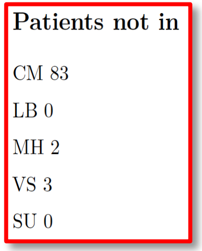
CM 7271

LB 7354

MH 7352

VS 7351

SU 7354



Patients not in

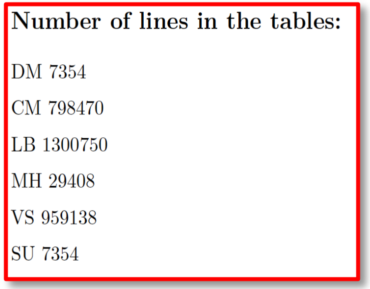
CM 83

LB 0

MH 2

VS 3

SU 0



Number of lines in the tables:

DM 7354

CM 798470

LB 1300750

MH 29408

VS 959138

SU 7354

20 random patients double coded

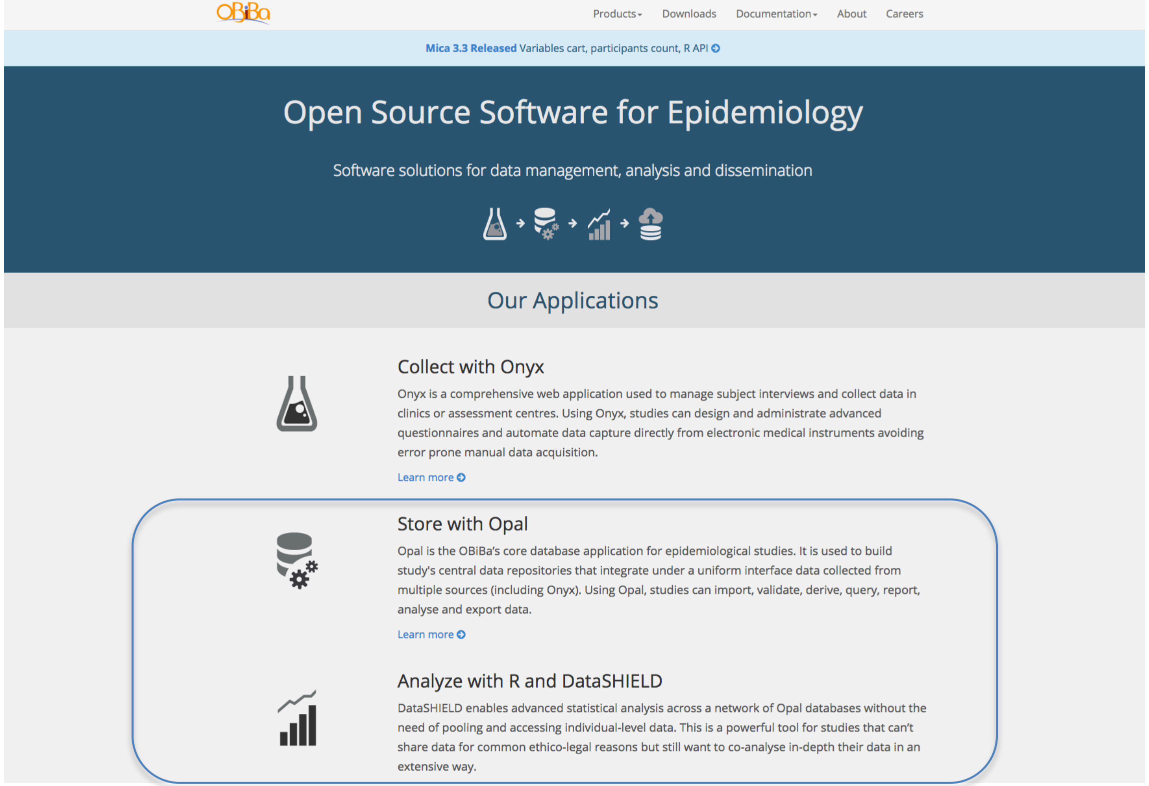
Manually checked by Anne

Quelques remarques

- Nous avons parfois pris quelques libertés avec le format SDTM, le but était plus d’harmoniser que de coller parfaitement au standard (pas de soumission de donnée prévue)

- Plus généralement, ces cohortes sont très différentes de patients recrutés pour une étude clinique (ici pas treatment arm, exposure, adverse events, …)

Infrastructure: Introducing OBiBa software



Mica 3.3 Released Variables cart, participants count, R API

Open Source Software for Epidemiology

Software solutions for data management, analysis and dissemination

Our Applications

Collect with Onyx

Onyx is a comprehensive web application used to manage subject interviews and collect data in clinics or assessment centres. Using Onyx, studies can design and administrate advanced questionnaires and automate data capture directly from electronic medical instruments avoiding error prone manual data acquisition.

Learn more

Store with Opal

Opal is the OBiBa's core database application for epidemiological studies. It is used to build study's central data repositories that integrate under a uniform interface data collected from multiple sources (including Onyx). Using Opal, studies can import, validate, derive, query, report, analyse and export data.

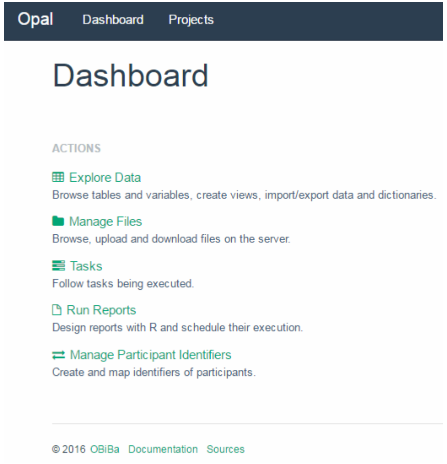
Learn more

Analyze with R and DataSHIELD

DataSHIELD enables advanced statistical analysis across a network of Opal databases without the need of pooling and accessing individual-level data. This is a powerful tool for studies that can't share data for common ethico-legal reasons but still want to co-analyse in-depth their data in an extensive way.

www.obiba.org

Infrastructure: Virtual machine with complete set of software to run a RHAPSODY node



Dashboard

ACTIONS

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

Manage Files
Browse, upload and download files on the server.

Tasks
Follow tasks being executed.

Run Reports
Design reports with R and schedule their execution.

Manage Participant Identifiers
Create and map identifiers of participants.

© 2016 OBiBa Documentation Sources

- Oracle VirtualBox platform 5.0.20
  - Runs on Linux, Solaris, Windows, Mac OS
- Oracle Enterprise Linux 7 as guest OS
- BioShare Opal 2.5.1 (the latest available)
  - MySQL 5.7.12
  - R 3.2.3
  - Opal-rserver
  - R studio
  - DataShield
  - Python API utilities
- Ready to be distributed (4GB)
- Repository and workbench for harmonized RHAPSODY data
- Only 15 GB disk space – needs additional volume(s) to be mounted, direct or NFS
- Needs node-specific analysis R script(s)

Lists of analyses that are possible using the federated database (programmées en R)

Already Implemented:

- Quantiles, summaries, *glm* (DataSHIELD)
- *PCA*
- *Kmeans* clustering
- Fast linear regression
- Gaussian mixtures
- Random forests (*not fully tested*)

Possible to run in "*Virtual cohort*" mode

- *KNN* imputation (*vim*)
- Cox proportional hazards (*coxph*)
- Conditional logistic regression (*clogit*)
- Linear mixed models (*nlme*)

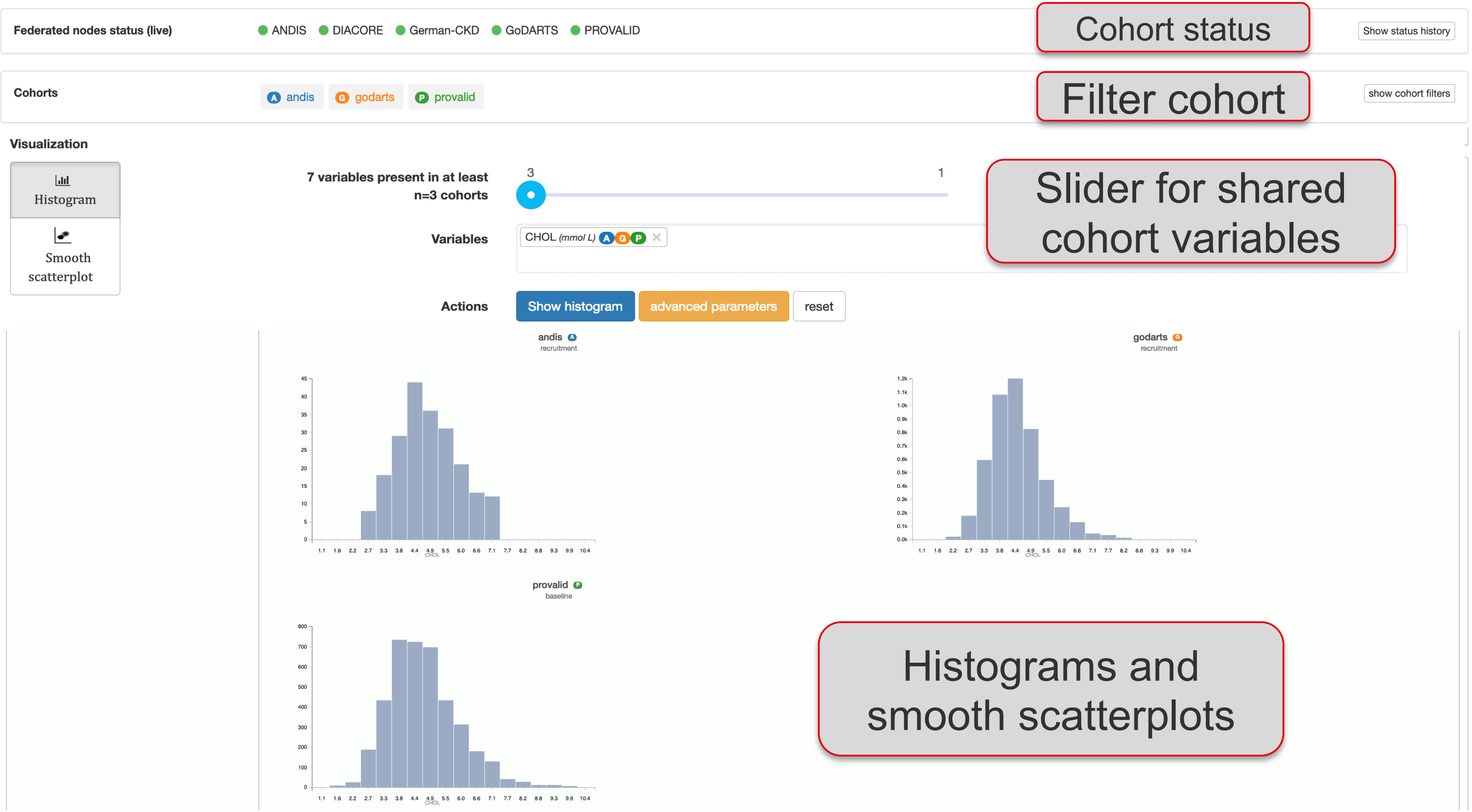
Work in progress:

- Similarity Network Fusion (*SNF*)

Possible to run in *Replication* mode only

In RHAPSODY we have built a **federated database** comprising **10 clinical cohorts**

Available data in federated databases can be browsed on web interface



Federated nodes status (live) ● ANDIS ● DIACORE ● German-CKD ● GoDARTS ● PROVALID

Show status history

Cohorts A andis G godarts P provalid

show cohort filters

Visualization

7 variables present in at least n=3 cohorts

Variables CHOL *(mmol L)* A G P ✕

Actions Show histogram | advanced parameters | reset

Cohort status

Filter cohort

Slider for shared cohort variables

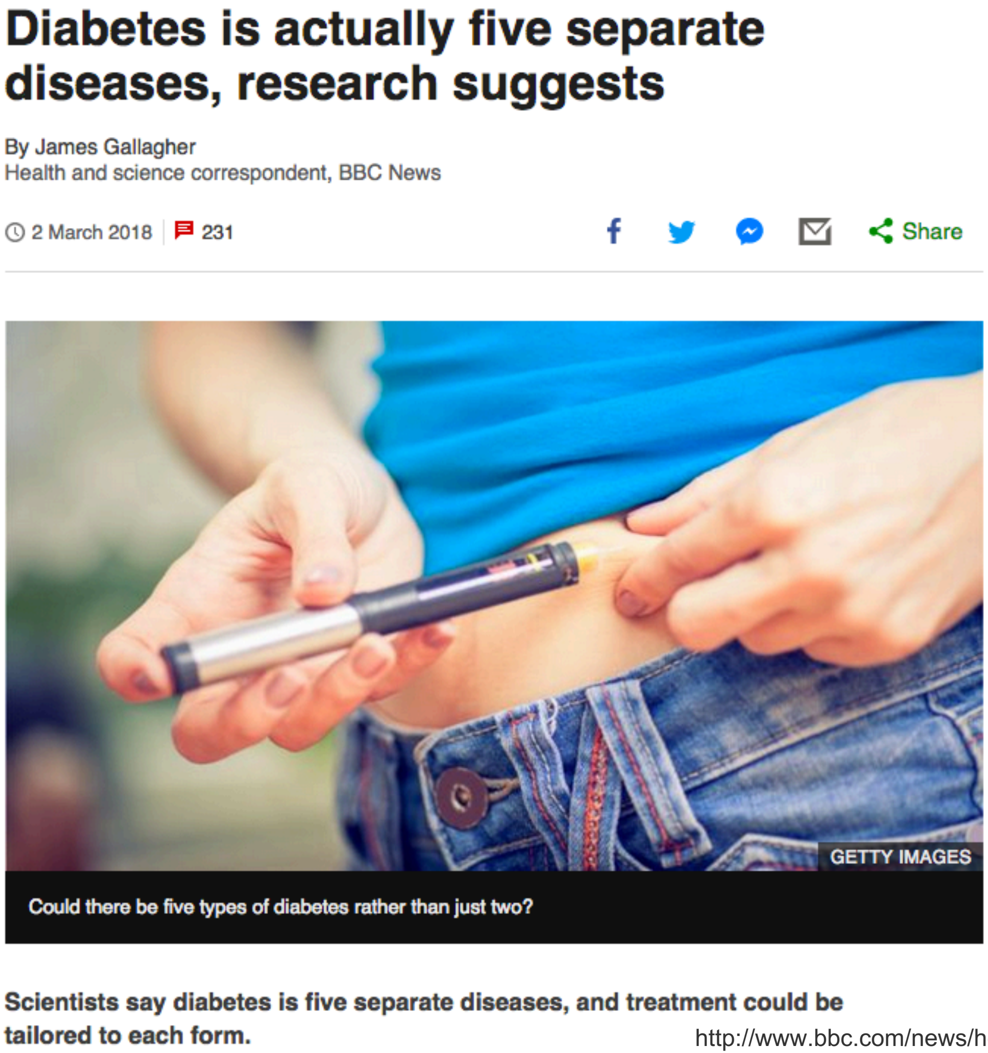
Histograms and smooth scatterplots

Deep clinical phenotypes for ~50K individuals harmonised and federated in RHAPSODY

Cohort	Cohort type	No. Individuals
GoDARTS	Progression	9081
ANDIS	Progression	11549
DCS	Progression	5560
BOTNIA	Pre-diabetes	3354
MDC	Pre-diabetes	3008
DESIR	Pre-diabetes	5212
COLAUS	Pre-diabetes	6187
ABOS	Gastric bypass	249
ADDITION-DK	Progression	1533
ADDITION-PRO	Pre-diabetes	2093
Total		**47826**

Contenu

- Présentation du SIB et de Vital-IT
- Pourquoi harmoniser et convertir les données? Présentation du système d’analyse fédéré
- Processus de conversion, notre utilisation de SDTM
- **Exemple d’utilisation du système fédéré**



Diabetes is actually five separate diseases, research suggests

By James Gallagher
Health and science correspondent, BBC News

2 March 2018 | 231

Could there be five types of diabetes rather than just two?

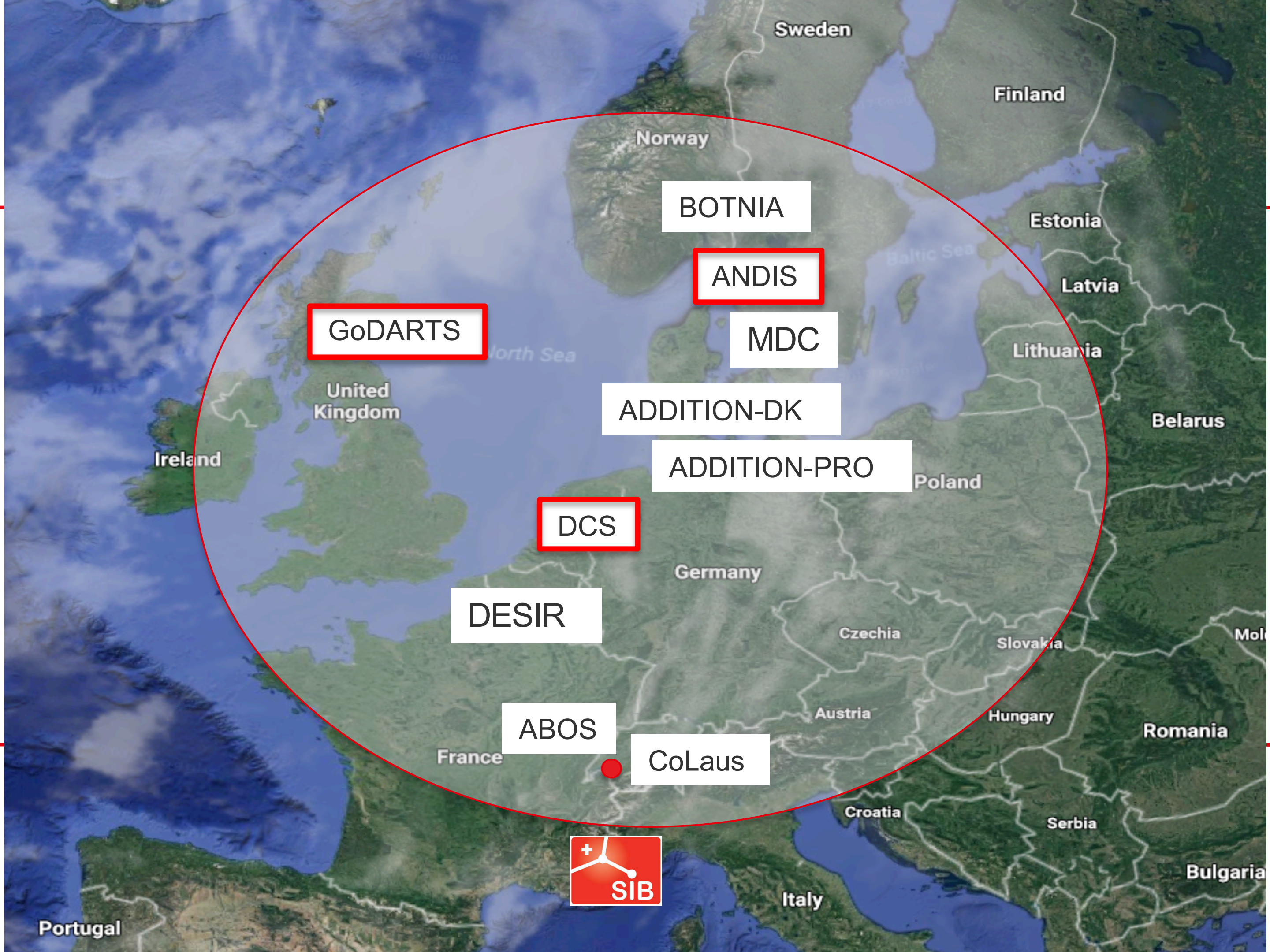
Scientists say diabetes is five separate diseases, and treatment could be tailored to each form.

http://www.bbc.com/news/health-43246261

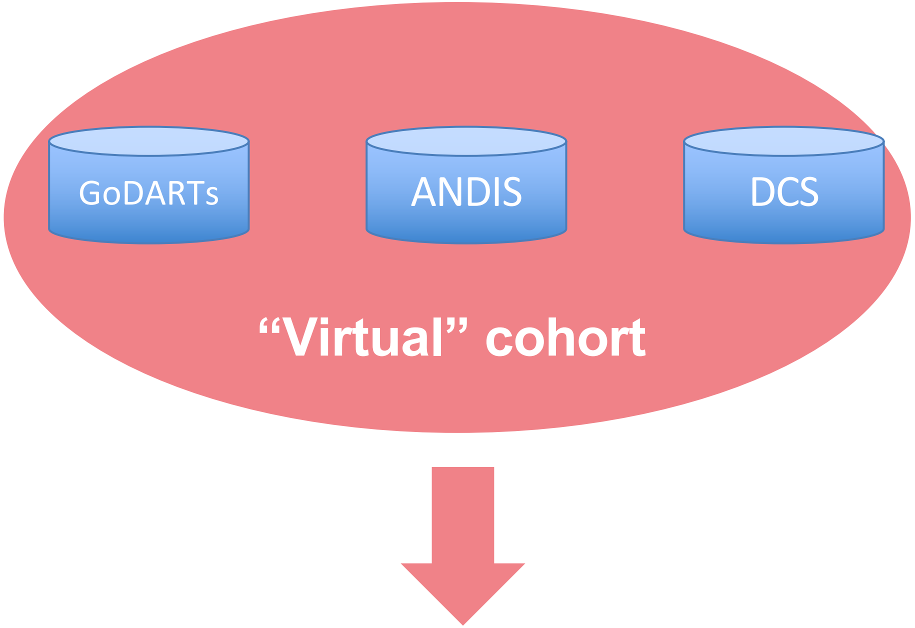
Can we replicate the clusters using the federated database?

Perform *kmeans* clustering using 5 clinical variables (HBA1c, Cpeptide, BMI, Age, HDL) on

ANDIS, **DCS** and **GoDARTS** cohorts through the RHAPSODY federated database


Sweden
Finland
Norway
BOTNIA
Estonia
ANDIS
Latvia
GoDARTS
MDC
North Sea
Lithuania
United Kingdom
ADDITION-DK
Belarus
Ireland
ADDITION-PRO
Poland
DCS
Germany
DESIR
Czechia
Slovakia
Mol
Austria
ABOS
Hungary
Romania
France
CoLaus
Croatia
Serbia
SIB
Bulgaria
Italy
Portugal

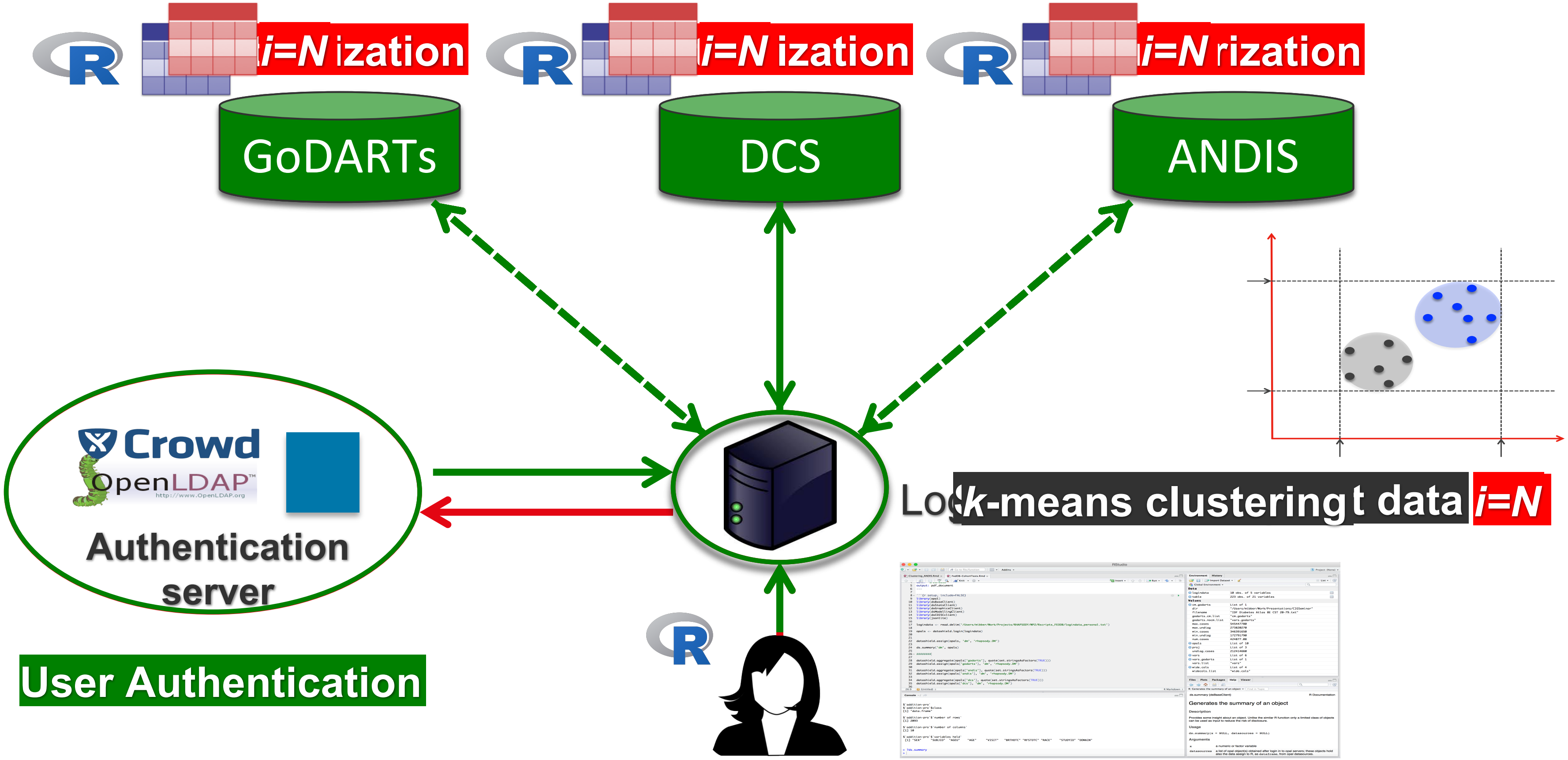
Federated *k*-means clustering using 5 variables



Age
BMI
C-peptide
HbA1c
HDL

Patient clusters

How it works ..



i=N ization

i=N ization

i=N rization

GoDARTs

DCS

ANDIS

Authentication server

User Authentication

Lo k-means clustering t data i=N

RStudio

~~~
output: pdf_document
---

```{r setup, include=FALSE}
library(opal)
library(dsBaseClient)
library(dsStatsClient)
library(dsGraphicsClient)
library(dsModellingClient)
library(dsCDISCclient)
library(jsonlite)

logindata <- read.delim('/Users/mibber/Work/Projects/RHAPSODY/WP2/Rscripts_FEDDB/logindata_personal.txt')

opals <- datashield.login(logindata)


datashield.assign(opals, 'dm', 'rhapsody.DM')

ds.summary('dm', opals)

########

datashield.aggregate(opals['godarts'], quote(set.stringsAsFactors(TRUE)))
datashield.assign(opals['godarts'], 'dm', 'rhapsody.DM')

datashield.aggregate(opals['andis'], quote(set.stringsAsFactors(TRUE)))
datashield.assign(opals['andis'], 'dm', 'rhapsody.DM')

datashield.aggregate(opals['dcs'], quote(set.stringsAsFactors(TRUE)))
datashield.assign(opals['dcs'], 'dm', 'rhapsody.DM')
~~~

```
$`addition-pro`
$`addition-pro`$class
[1] "data.frame"

$`addition-pro`$`number of rows`
[1] 2093

$`addition-pro`$`number of columns`
[1] 10

$`addition-pro`$`variables held`
 [1] "SEX"      "SUBJID"  "AGEU"     "AGE"      "VISIT"    "BRTHDTC" "RFSTDTC" "RACE"     "STUDYID" "DOMAIN"

> ?ds.summary
>
```

```
Data
logindata              10 obs. of 5 variables
table                  223 obs. of 21 variables
Values
cm.godarts             List of 1
dir                    "/Users/mibber/Work/Presentations/CIGSeminar"
filename               "IDF Diabetes Atlas 8E CST 20-79.txt"
godarts.cm.list        "cm.godarts"
godarts.nocm.list      "vars.godarts"
max.cases              545447780
max.undiag             273820270
min.cases              346391650
min.undiag             172791790
num.cases              424877.08
opals                  List of 10
proj                   List of 3
undiag.cases           212414660
vars                   List of 6
vars.godarts           List of 1
vars.list              "vars"
wide.cols              List of 4
widecols.list          "wide.cols"
```

ds.summary {dsBaseClient} R Documentation

Generates the summary of an object

Description

Provides some insight about an object. Unlike the similar R function only a limited class of objects can be used as input to reduce the risk of disclosure.

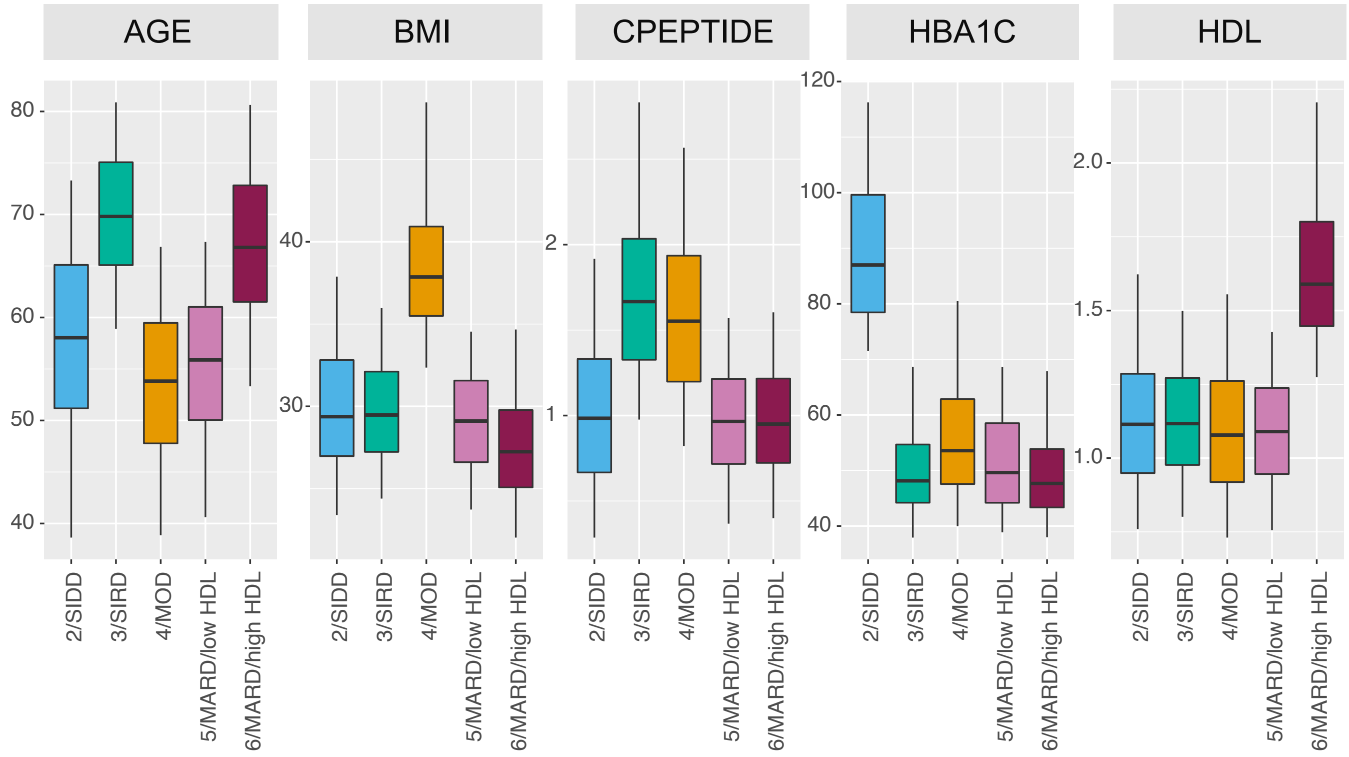
Usage

`ds.summary(x = NULL, datasources = NULL)`

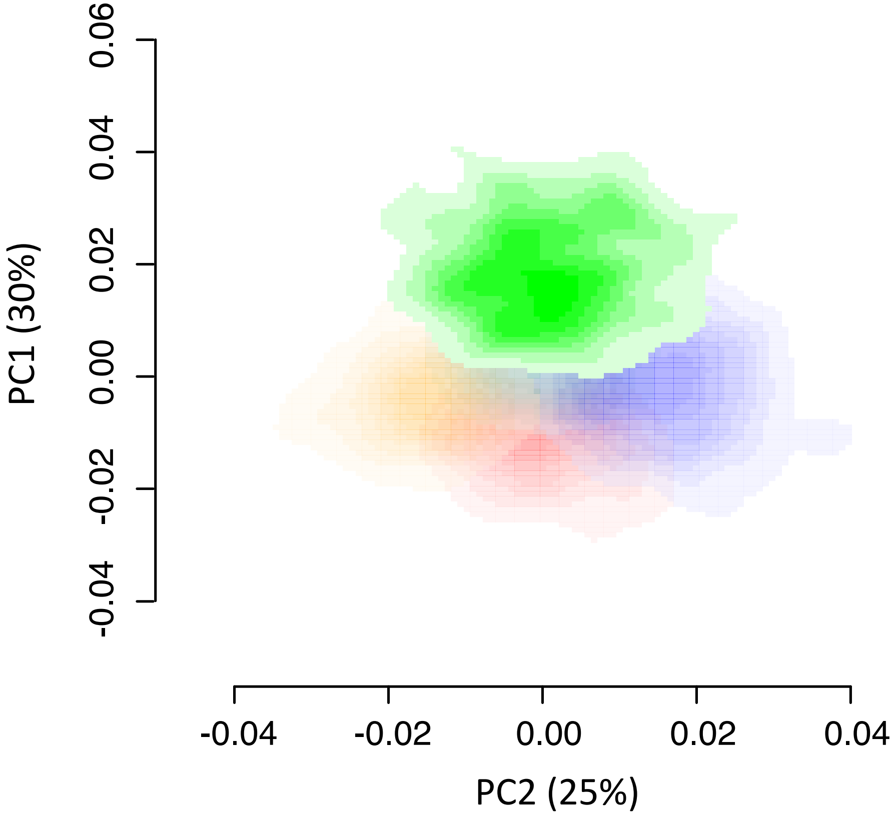
Arguments

x	a numeric or factor variable
datasources	a list of opal object(s) obtained after login in to opal servers; these objects hold also the data assign to R, as dataframe, from opal datasources.

Clustering reproduced on a “virtual” cohort comprising ANDIS, DCS and GoDARTS cohorts



Federated PCA on “virtual” cohort comprised of DCS + ANDIS + GoDARTS (N=5723)



Iulian Dragan

Take home messages

- Nous avons pu montrer que **l’analyse de multiples cohortes** cliniques est possible à travers un **système fédéré** et **l’harmonisation** des données en utilisant CDISC-SDTM
- Une partie des analyses peut être effectuée comme si les données avaient été “poolées” physiquement, sans toutefois **qu’aucune donnée individuelle de patient** ne quitte son environnement local
- L’analyse de donnée en système fédéré peut bénéficier d’une **puissance statistique augmentée** par l’analyse groupée de plusieurs cohorts, tout en respectant les contraintes **légales et éthiques.**

RHAPSODY Core Federated DB Team @ Vital-IT, SIB

Federated Database

Dmitry Kuznetsov

Iulian Dragan

Scientific Portal

Robin Liechti

Lou Götz

Fabio Lehman

Cohort harmonization

Frédéric Burdet

Mark Ibberson – Scientific Lead

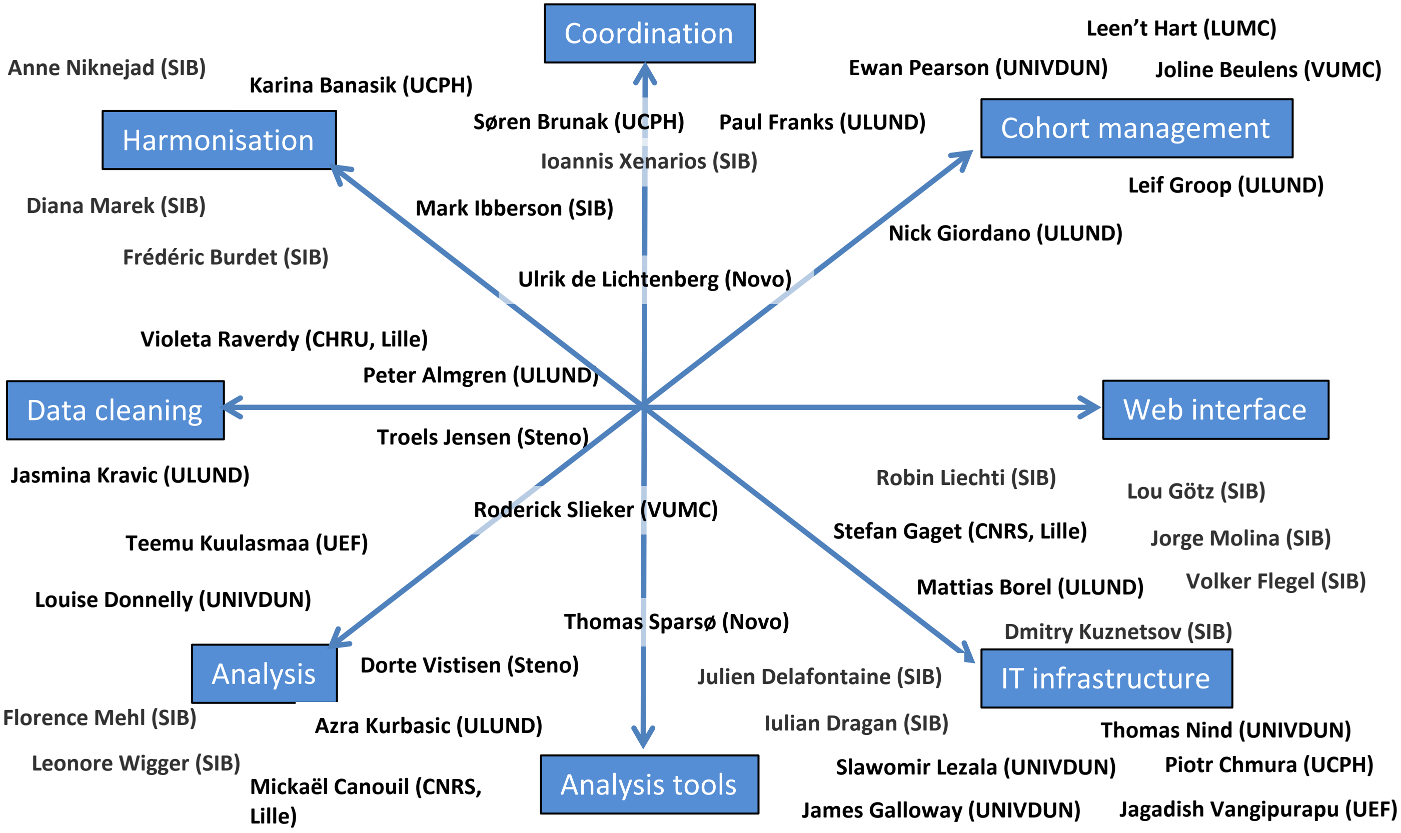
Diana Marek

Anne Niknejad

Vital-IT High Performance Computing Center

RHAPSODY for precision therapy and prevention of diabetes

RHAPSODY
Federated DB Team



Coordination

Leen't Hart (LUMC)

Anne Niknejad (SIB)

Karina Banasik (UCPH)

Ewan Pearson (UNIVDUN)

Joline Beulens (VUMC)

Harmonisation

Søren Brunak (UCPH)

Paul Franks (ULUND)

Cohort management

Ioannis Xenarios (SIB)

Leif Groop (ULUND)

Diana Marek (SIB)

Mark Ibberson (SIB)

Nick Giordano (ULUND)

Frédéric Burdet (SIB)

Ulrik de Lichtenberg (Novo)

Violeta Raverdy (CHRU, Lille)

Peter Almgren (ULUND)

Data cleaning

Web interface

Troels Jensen (Steno)

Jasmina Kravic (ULUND)

Robin Liechti (SIB)

Lou Götz (SIB)

Roderick Slieker (VUMC)

Stefan Gaget (CNRS, Lille)

Jorge Molina (SIB)

Teemu Kuulasmaa (UEF)

Mattias Borel (ULUND)

Volker Flegel (SIB)

Louise Donnelly (UNIVDUN)

Thomas Sparsø (Novo)

Dmitry Kuznetsov (SIB)

Analysis

Dorte Vistisen (Steno)

Julien Delafontaine (SIB)

IT infrastructure

Florence Mehl (SIB)

Azra Kurbasic (ULUND)

Iulian Dragan (SIB)

Thomas Nind (UNIVDUN)

Leonore Wigger (SIB)

Slawomir Lezala (UNIVDUN)

Piotr Chmura (UCPH)

Mickaël Canouil (CNRS, Lille)

Analysis tools

James Galloway (UNIVDUN)

Jagadish Vangipurapu (UEF)

Merci

Slides backup

FMI
Friedrich Miescher Institute
for Biomedical Research

University
of Basel

Swiss TPH

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

University of
Zurich^UZH

Agroscope

Zurich University
of Applied Sciences
zh
aw

SIAF

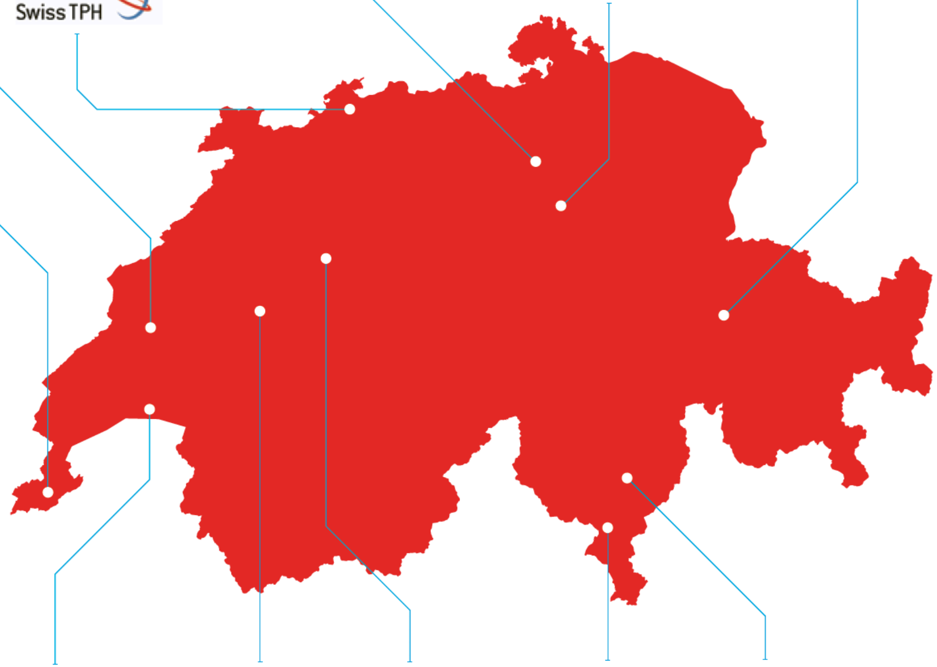
heig-vd
HAUTE ÉCOLE
D'INGÉNIERIE ET DE GESTION
DU CANTON DE VAUD
www.heig-vd.ch

UNIVERSITÉ
DE GENÈVE

HUG
Hôpitaux
Universitaires
Genève

espeRare

heg
Haute école de gestion de Genève
Geneva School of Business Administration



20 institutional partners
all over
Switzerland

EPFL

Unil
UNIL | Université de Lausanne

LUDWIG
CANCER
RESEARCH

UNI
FR
UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

u^b
UNIVERSITÄT
BERN

USI
Università
della
Svizzera
italiana

IOR
Institute of Oncology Research