

Datashield, Opal & Tools - loading data, connection instance and running 3rd party tools

Niels Hulstaert

Niels.hulstaert@vib-ugent.be
computational omics and systems biology group
VIB / Ghent University, Ghent, Belgium



Introduction

Opal

DataSHIELD

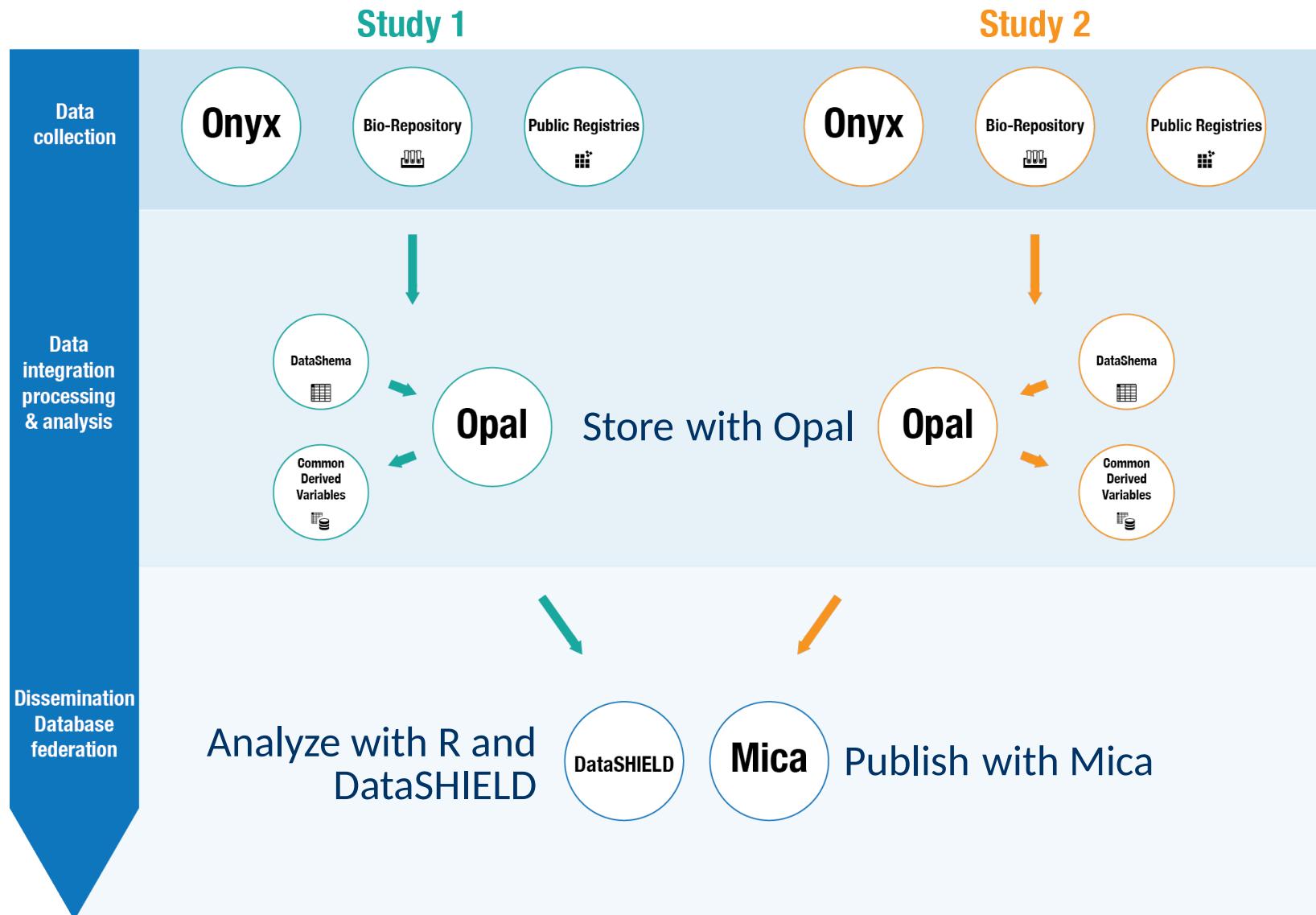
Mica, Mica Drupal Client, Mica Python Client

Harmonization



CC BY-SA 4.0

ObiBa applications suite



Installation prerequisites

- Ubuntu 14.04
- 3 Virtual machines (XenServer) in private network
 - 2 opal instances
 - 1 mica instance
- HTTP only, no HTTPS
- I'm a software developer, no sysadmin



CC BY-SA 4.0

Introduction

Opal

DataSHIELD

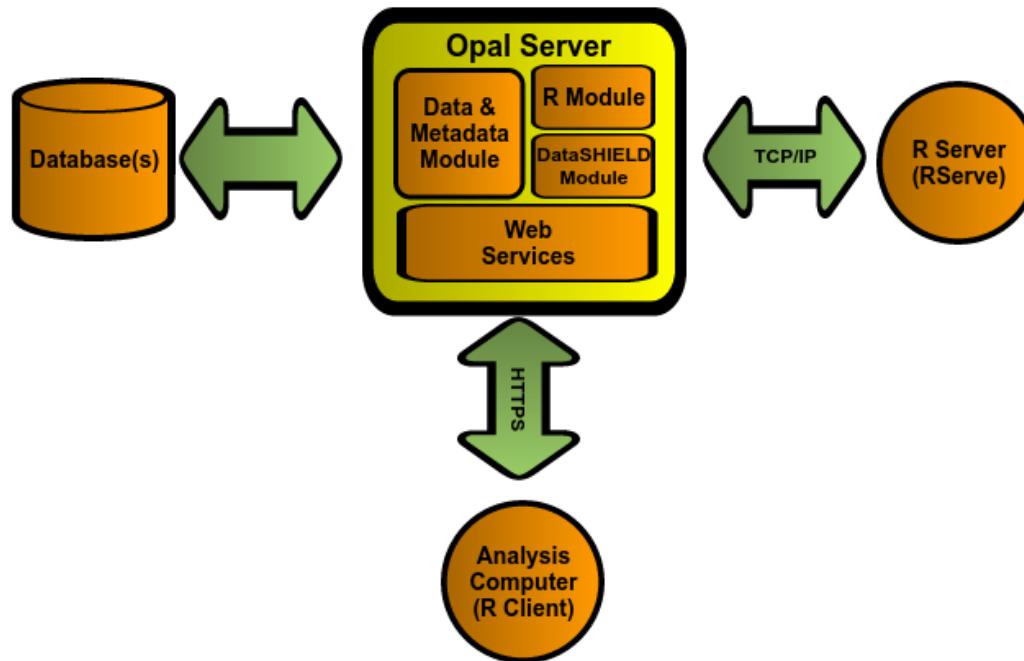
Mica, Mica Drupal Client, Mica Python Client

Harmonization



CC BY-SA 4.0

Opal schematic overview



<http://wiki.obiba.org/display/OPALDOC/Opal+R+and+DataSHIELD+User+Guide>

Opal database installation

- <http://wiki.obiba.org/display/OPALDOC/Databases+Administration>
- MongoDB or MySQL/MariaDB, PostgreSQL

Database Engine	Data Schema	Storage	Import	Export
MySQL	Opal SQL	✓	✗	✗
MySQL, PostgreSQL	Tabular SQL	✓	✓	✓
MySQL, PostgreSQL	Limesurvey	✗	✓	✗
MongoDB	Opal Documents	✓	✗	✗

Some MongoDB installation remarks

- MongoDB on Ubuntu:
<https://docs.mongodb.org/master/tutorial/install-mongodb-on-ubuntu/>
- Create a MongoDB “root” user according to
<http://stackoverflow.com/questions/20117104/mongodb-root-user>
- Follow the Opal wiki documentation
- Security authorization in the /etc/mongod.conf file

Some MySQL installation remarks

- Database creation
 - CREATE DATABASE opal_ids CHARACTER SET utf8 COLLATE utf8_bin;
 - CREATE DATABASE opal_data CHARACTER SET utf8 COLLATE utf8_bin;
 - CREATE USER 'opal' IDENTIFIED BY '<opal-user-password>';
 - GRANT ALL ON opal_ids.* TO 'opal'@'localhost' IDENTIFIED BY '<opal-user-password>';
 - GRANT ALL ON opal_data.* TO 'opal'@'localhost' IDENTIFIED BY '<opal-user-password>';
 - FLUSH PRIVILEGES;
- Data schemas:
 - Opal SQL: flexibility over performance
 - Tabular SQL: SQL limitations, import/export

Opal web application demo

- <http://opal-demo.obiba.org/ui/index.html#!dashboard>

Projects

A project is a repository of data and dictionaries. In a project, data can be imported, exported, analysed and transformed. Projects also offer access controls to the dictionaries, the summaries and the individual data.

+ Add Project		Filter projects...	Total 21
▲ Name	Title	Description	Last updated
CLS	CLS		6 months ago
CLSA	CLSA		6 months ago
CNSIM2	Testing DataSHIELD functions 2	Simulated data 2	8 months ago
CNSIM3	Testing DataSHIELD functions 3	Simulated data 3	2 years ago
CPTP	CPTP		7 months ago
FNAC	FNAC		7 months ago
FRELE	FRELE		7 months ago
GEEtest1	Testing DataSHIELD GEE function	Simulated data	2 months ago
GEEtest2	Testing DataSHIELD GEE	Simulated data	a year ago
HELIAD	HELIAD		7 months ago
LASA	LASA		7 months ago
LBLS	LBLS		7 months ago
NuAge	NuAge		7 months ago
PATH	PATH		7 months ago

Introduction

Opal

DataSHIELD

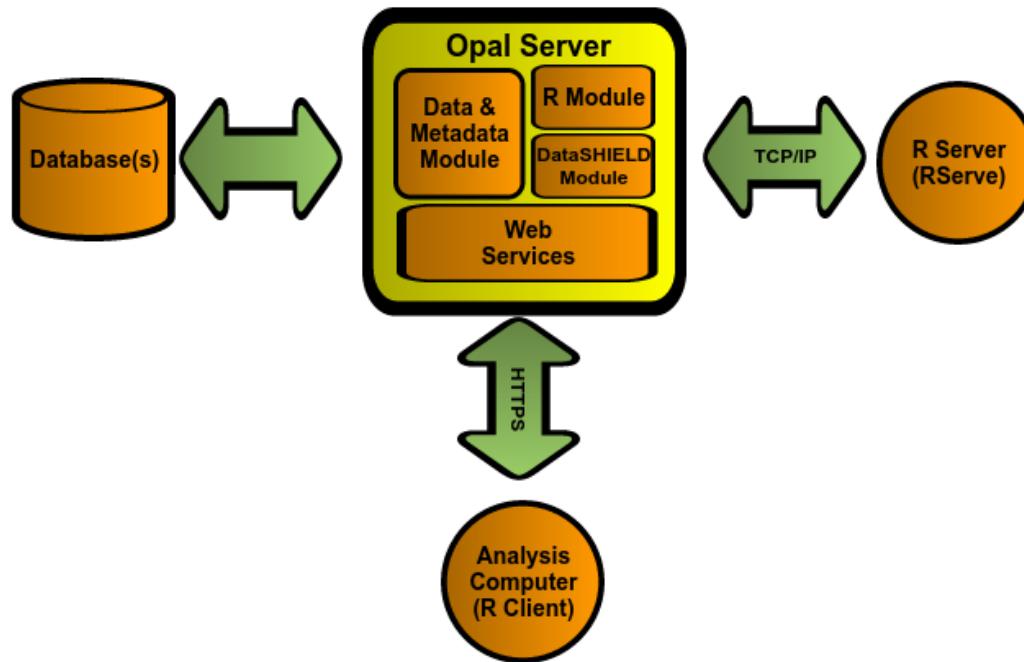
Mica, Mica Drupal Client, Mica Python Client

Harmonization



CC BY-SA 4.0

Opal schematic overview



<http://wiki.obiba.org/display/OPALDOC/Opal+R+and+DataSHIELD+User+Guide>

Opal R server side note

- Can be accessed without dataSHIELD!

```
# Load Opal R library
require('opal')

# Then, create an opal object with the login information
# Change the login credentials and url with the appropriate values!
o <- opal.login(username='administrator', password='password', url='https://localhost:8443')

# To verify if the connexion with Opal works,
# get the list of all projects
opal.datasources(o)

# Assign the content of the LifeLines table (from the test project)
# into a data frame in a R session of the R server
opal.assign(o,'D','test.LifeLines', missings = TRUE)

# Get the summary of this data frame from the remote R session
opal.execute(o,'summary(D)')

# Terminate the remote R session
opal.logout(o)
```

DataSHIELD installation and usage

- <http://wiki.obiba.org/display/OPALDOC/How+to+install+and+use+Opal+and+DataSHIELD+for+Data+Harmonization+and+Federated+Analysis>
- On Opal server (web application):
 - install dataSHIELD packages in admin section
 - create a user and set the correct permissions on both dataSHIELD and project/table level
- On client side:
 - install necessary R packages
 - connect for example with Rstudio (demo?)

Introduction

Opal

DataSHIELD

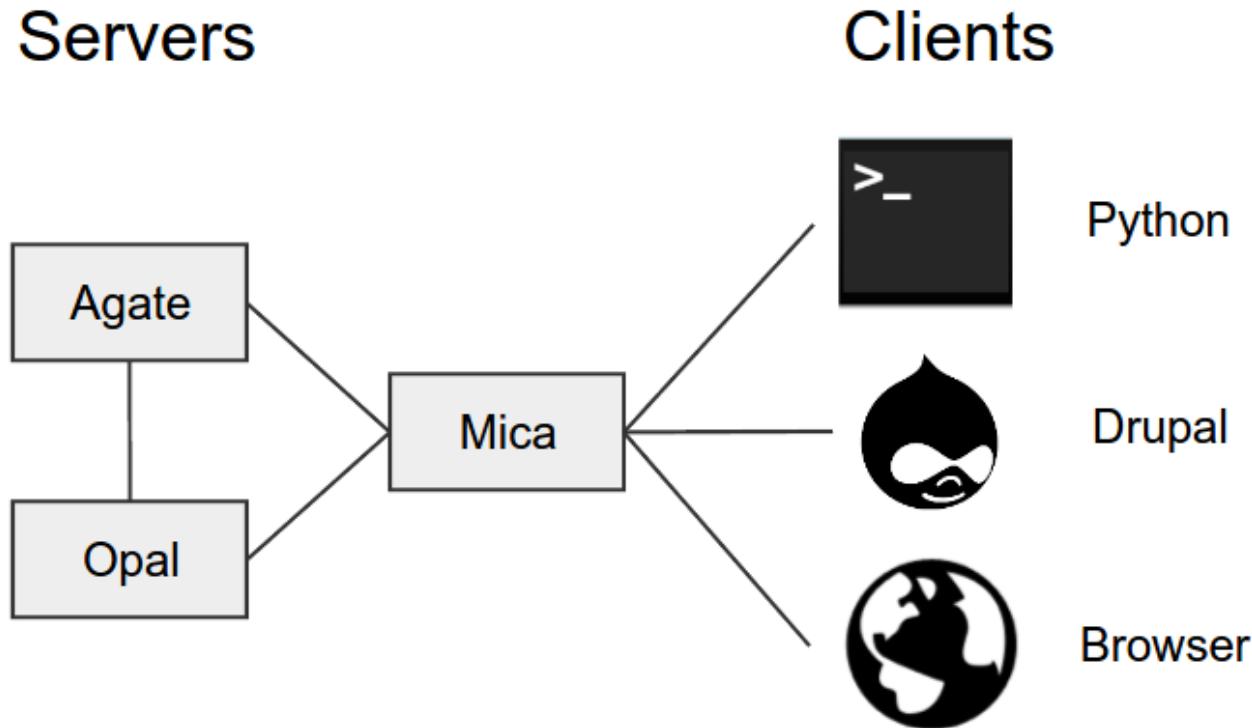
Mica, Mica Drupal Client, Mica Python Client

Harmonization



CC BY-SA 4.0

Mica schematic overview



<http://wiki.obiba.org/display/MICADOC/Overview>

Mica modules

- Mica server: RESTful Java server for managing, storing and searching Mica content
- Mica Web Application: user interface for administration and configuration of a Mica server
- Mica Drupal Client: Mica CMS providing tools to create custom web portals
- Mica Python Client: command-line interface for administration and automation of a Mica server

Some Mica-server installation remarks

- Install MongoDB first!
- <http://wiki.obiba.org/display/MICADOC/Mica+Server+Installation+Guide>
- Connection with Opal:
 - On Opal server: create Mica user and set the necessary project/table permissions
 - On Mica server: add (multiple) Opal credentials

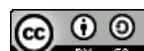
Mica demo's

- Mica Web Application demo
 - <http://mica-demo.obiba.org/>
- Mica Drupal Client demo
 - <http://demo.obiba.org/mica2/>
- Mica Drupal Client example
 - <https://www.bioshare.eu/>

Mica-python-client

- Command line scripting tool written in Python, enables automation of tasks in a Mica server
- Provides access to the Mica REST-API

```
compomics@compomics-xubuntu:~$ mica rest -mk http://mica-demo.obiba.org -u administrator -p password -j /study/finrisk-2007
{
  "acronym": [
    {
      "lang": "en",
      "value": "FINRISK 2007"
    }
  ],
  "contacts": [
    {
      "email": "helena.kaariainen@ktl.fi",
      "firstName": "Helena",
      "institution": {},
      "lastName": "Kaariainen",
      "phone": "+358 (0) 20 610 6000"
    },
    {
      "email": "markus.perola@ktl.fi",
      "firstName": "Markus",
      "institution": {
        "address": {
          "city": [
            {
              "lang": "en",
              "value": "Helsinki"
            }
          ],
          "country": {
            "iso": "FI",
            "name": [
              {
                "lang": "en",
                "value": ""
              }
            ]
          },
          "zip": "251"
        }
      }
    }
  ]
}
```



Introduction

Opal

DataSHIELD

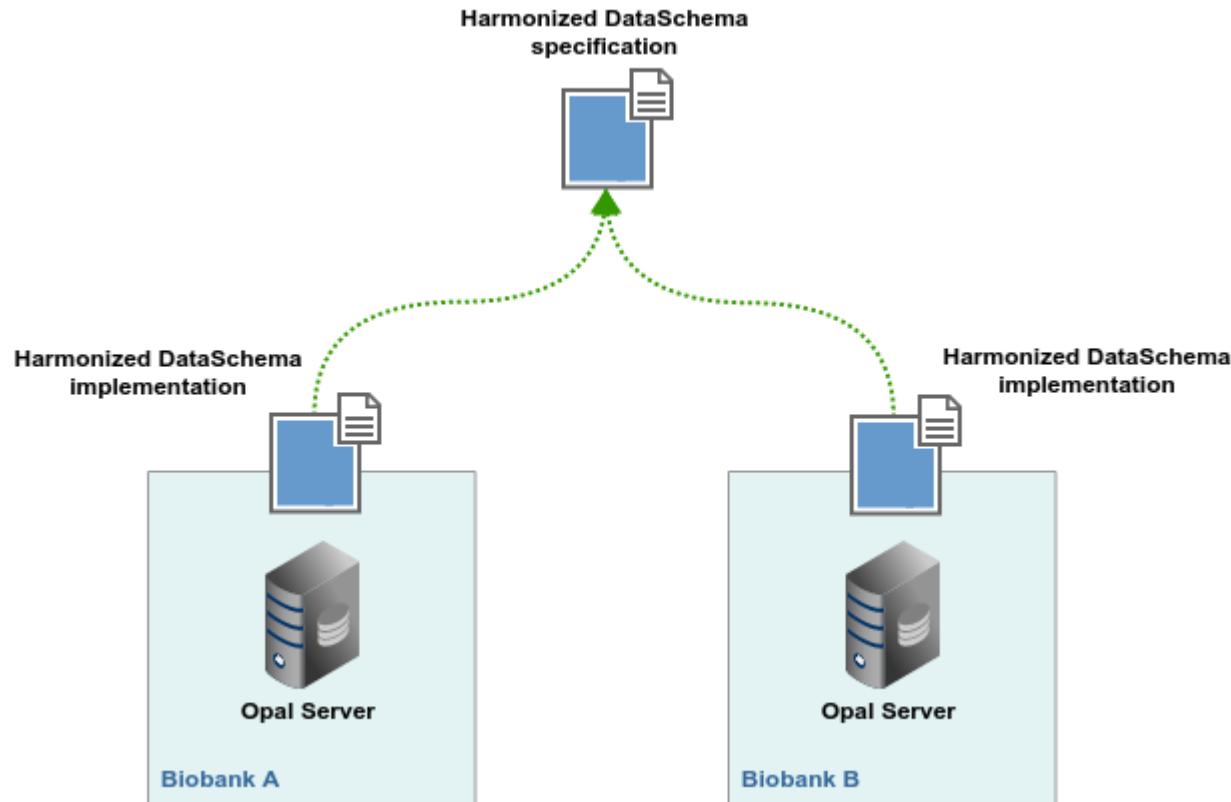
Mica, Mica Drupal Client, Mica Python Client

Harmonization



CC BY-SA 4.0

Harmonization across Opal instances



<http://wiki.obiba.org/display/OPALDOC/Data+Harmonization+with+Opal#DataHarmonizationwithOpal-DataHarmonizationacrossBiobanks>

Harmonized DataSchema

- Create on Opal (or Mica Drupal Client) web application
- Export/Import as XML to other Opal instances
- Extensive tutorial:

<http://wiki.obiba.org/display/MLSTRM/Data+Harmonization+and+Database+Federation+Tutorial>

	Name	Label	Value Type	Categories
<input type="checkbox"/>	Gender		integer	0, 1, 9
<input type="checkbox"/>	PM_BMI_CAT		integer	1, 2, 3, 9
<input type="checkbox"/>	DIS_DIAB		integer	0, 1, 9
<input type="checkbox"/>	PM_DIASTOLIC		integer	999
<input type="checkbox"/>	SMK_CIG_CURRENT		integer	0, 1, 9

Harmonized DataSchema variable

- Has categories (from taxonomy if available)
- Project/table data processing algorithms
 - Javascript
 - Wizard for simple transformations

Categories

[Edit Categories](#) Total 3

Name	Label	Missing
0	Male en Male	
1	Female en Female	
9	Missing	✓

Attributes

[+ Add Attribute ▾](#)

<input type="checkbox"/>	Name	Value	Actions
<input checked="" type="checkbox"/>	script	<pre>\$('SUKUP').map({ '1': '0', '2': '1' }, null, '9');</pre>	Edit Remove

More complex algorithm example

Categories

Edit Categories

Total 4

Name	Label	Missing
1	Less than 25 kg/m2 en Less than 25 kg/m2	
2	25 to 30kg/m2 en 25 to 30kg/m2	
3	Over 30 kg/m2 en Over 30 kg/m2	
9	Missing	<pre>/* If Q69 = 1 or 2 and TUPI1 is greater than 0, then SMK_CIG_CURRENT = 1*/ function cig_smoker() { var cig_day = \$('#TUPI1'); if (cig_day.gt(0).value()) return 1; else if (cig_day.eq(0).value()) return 0; else return 9; } \$('Q69').map({ '1': cig_smoker(), '2': cig_smoker(), '3': 0 }, 9, 9);</pre>



CC BY-SA 4.0

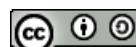
Harmonization dataset example

- <http://demo.obiba.org/mica2/mica/datasets/harmonization-datasets>

Statistics

Crosstabs

Value	Atlantic PATH 1	Atlantic PATH 2	BCGP 1	BCGP 2	BCGP 3	CaG	OHS 1	OHS 2	TPP 1	TPP 2	All
Valid Values											
0 None	17 12,90% (12,90%)	15 12,50% (12,50%)	13 12,30% (12,30%)	14 12,70% (12,70%)	13 12,50% (12,50%)	15 11,90% (11,90%)	23 12,40% (12,40%)	14 12,20% (12,20%)	20 12,70% (12,70%)	24 12,70% (12,70%)	168 12,50% (12,50%)
1 Elementary school	16 12,10% (12,10%)	15 12,50% (12,50%)	13 12,30% (12,30%)	14 12,70% (12,70%)	13 12,50% (12,50%)	16 12,70% (12,70%)	23 12,40% (12,40%)	14 12,20% (12,20%)	20 12,70% (12,70%)	23 12,20% (12,20%)	167 12,40% (12,40%)
2 High school	17 12,90% (12,90%)	15 12,50% (12,50%)	13 12,30% (12,30%)	13 11,80% (11,80%)	13 12,50% (12,50%)	16 12,70% (12,70%)	24 12,90% (12,90%)	15 13,00% (13,00%)	19 12,00% (12,00%)	23 12,20% (12,20%)	168 12,50% (12,50%)
3 Trade, technical or vocational school, apprenticeship training or technical CEGEP	16 12,10% (12,10%)	15 12,50% (12,50%)	13 12,30% (12,30%)	14 12,70% (12,70%)	13 12,50% (12,50%)	15 11,90% (11,90%)	23 12,40% (12,40%)	15 13,00% (13,00%)	19 12,00% (12,00%)	23 12,20% (12,20%)	166 12,30% (12,30%)
4 Diploma from a community college, pre-university CEGEP or non-university certificate	16 12,10% (12,10%)	15 12,50% (12,50%)	14 13,20% (13,20%)	14 12,70% (12,70%)	13 12,50% (12,50%)	16 12,70% (12,70%)	24 12,90% (12,90%)	14 12,20% (12,20%)	20 12,70% (12,70%)	24 12,70% (12,70%)	170 12,60% (12,60%)
5 University certificate below bachelor's level	17 12,90% (12,90%)	15 12,50% (12,50%)	13 12,30% (12,30%)	14 12,70% (12,70%)	13 12,50% (12,50%)	16 12,70% (12,70%)	23 12,40% (12,40%)	15 13,00% (13,00%)	20 12,70% (12,70%)	24 12,70% (12,70%)	170 12,60% (12,60%)



CC BY-SA 4.0



Comp omics



www.compomics.com
@compomics



CC BY-SA 4.0