Tools for the Process of Data Harmonization

Dany Doiron and Isabel Fortier

Maelstrom Research group
Research Institute of the McGill University Health Centre

mælstrom

Centre universitaire de santé McGill
Institut de recherche

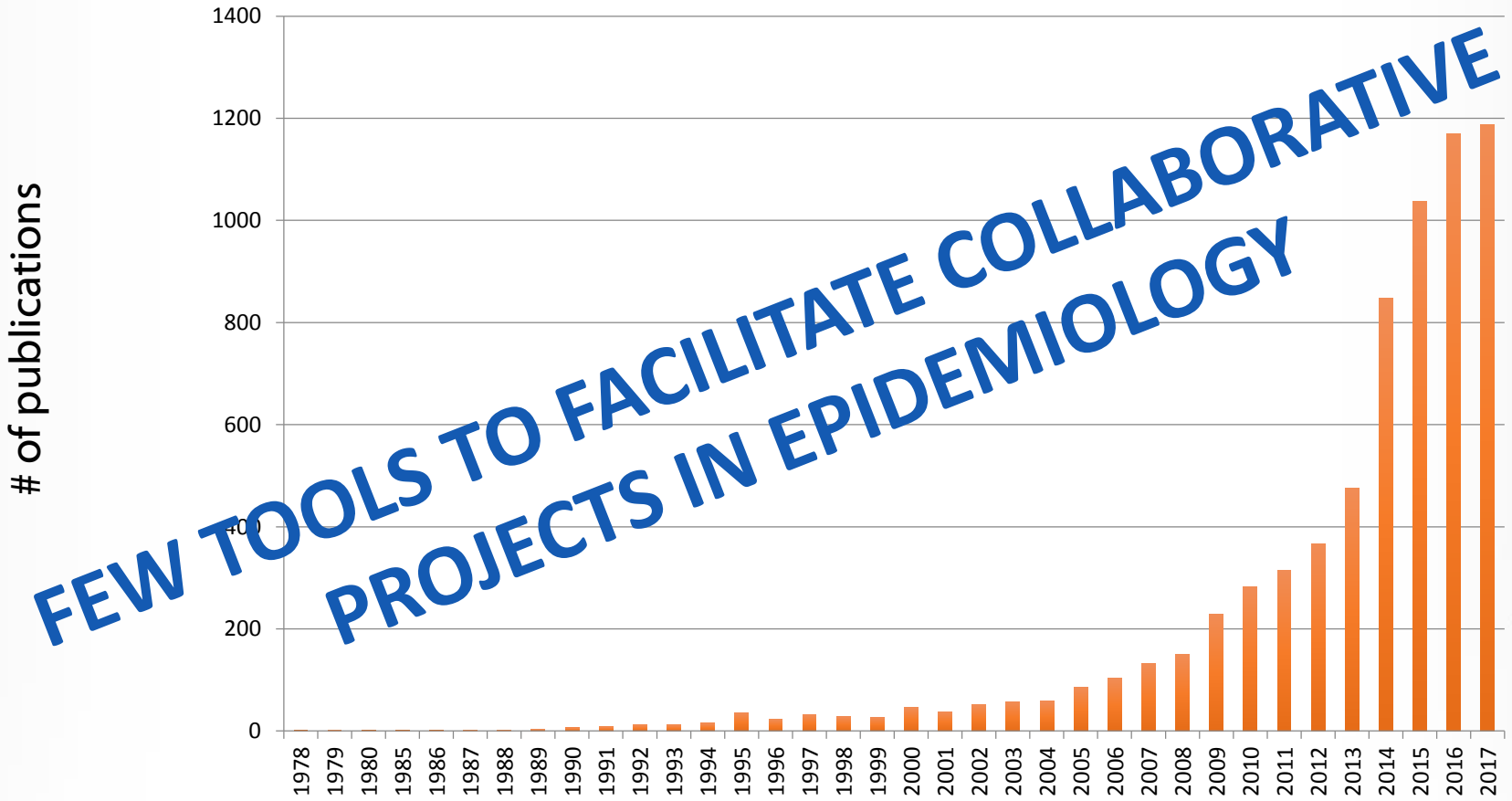
McGill University Health Centre
Research Institute

Presentation overview

- Maelstrom Research
- Tools for **data discovery**
  Who is collecting what?
- Tools for **data transformation**
  Deriving data into a common format
- Tools for **data analysis**
  Cross-cohort analyses

Increasing popularity of cross-cohort projects

PubMed search number of manuscripts matching "epidemiology" and "consortium"



Maelstrom Research overview

Mission: *To facilitate collaborative epidemiological research through rigorous data documentation, harmonization, and co-analysis*

Activities:

Methodological guidelines/support
for data cataloguing, harmonization, and co-analysis

Open-source software
for data cataloging, harmonization, and co-analysis

Web-based catalogues and harmonization platforms
searchable and scalable metadata catalogues and platforms to generate common-format variables for co-analysis

Tools for **data discovery**

Who is collecting what?

Why catalogue study metadata?

To conduct pooled or comparative analyses, investigators need to know which study collects what data.

However, there are problems:

Information unavailable

Study metadata often not publicly available

- Few epidemiological studies on the web
- Direct contact with PIs necessary to enquire about data collected
- Time-intensive for researchers and cohorts

Information hard to find

Study metadata publicly available but

- Dispersed on different websites
- Presented in different formats

RESULT: underexploited resources

Existing data collected by epidemiological studies are not being used to their full potential

mælstrom

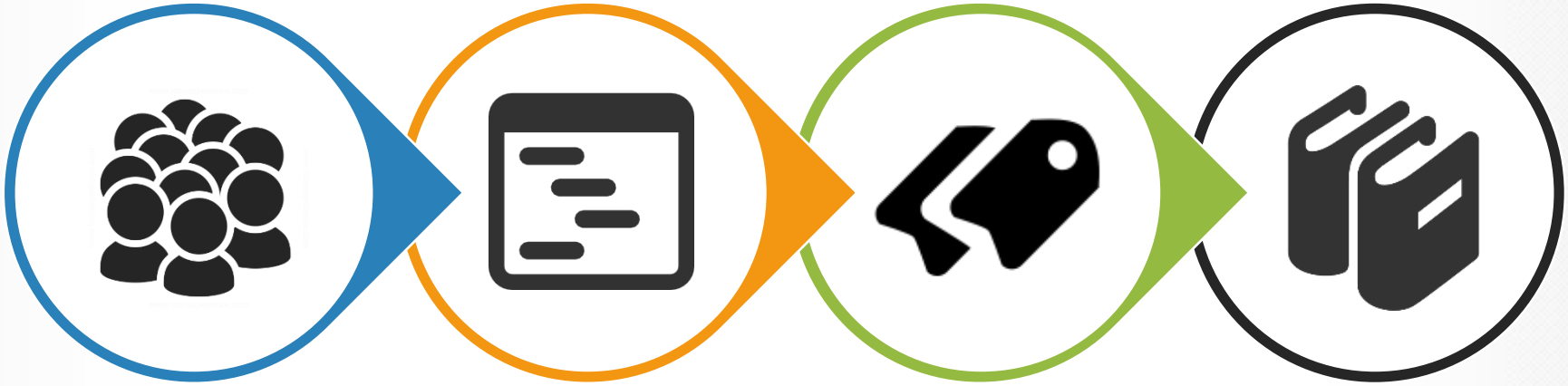
mica in a nutshell

A software application to **create web-based metadata catalogues** for individual studies or networks of studies

Main features:

- **Study cataloguing**: document study characteristics such as design, collection sweeps, data access rules
- **Variable cataloguing**: document variables collected by studies
- **Metadata search engine:** find information you are looking for!
- **Data access requests**: online application form, reviewing
- **Communication tools**: news, events, calendar, forums

Metadata cataloguing with mica



Study metadata

Document:

- General study design
- Participant selection criteria
- Data collection events
- Information on access to data and biosamples

Variable metadata

Document:

- Variable name/label
- Categories
- Units
- Assessment items
- SOPs

Areas of information

Classify:

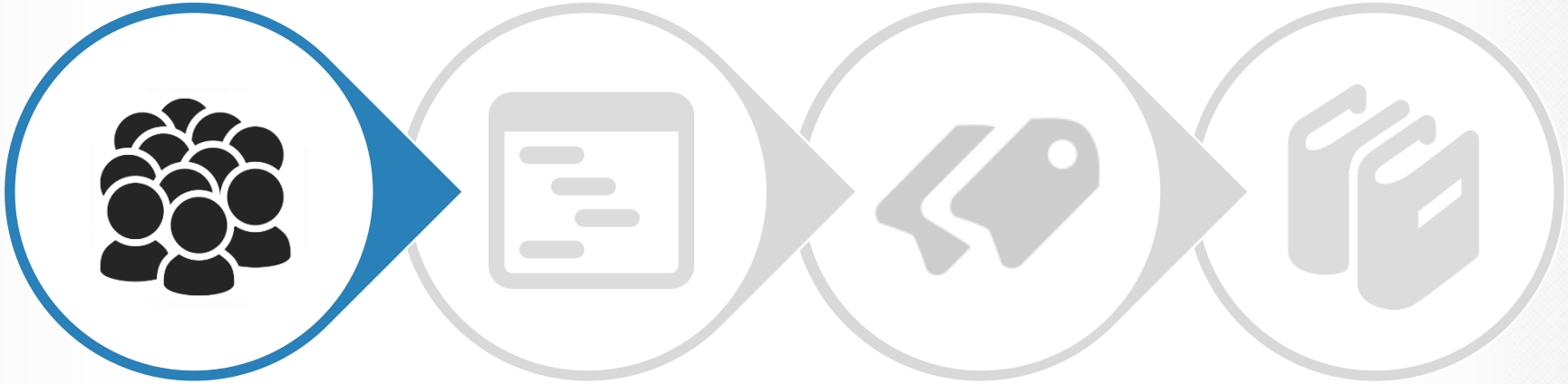
- Variables according to area of information (e.g. tobacco, neoplasms, anthropometry)

Metadata catalogue

End product:

- Fully searchable metadata portal to facilitate data discovery

Metadata cataloguing with mica



Study metadata

Document:

- General study design
- Participant selection criteria
- Data collection events
- Information on access to data and biosamples

Variable metadata

Document:

- Variable name/label
- Categories
- Units
- Assessment items
- SOPs

Areas of information

- Classify variables according to area of information (e.g. tobacco, neoplasms, anthropometry)

Metadata catalogue

End product:

- Fully searchable metadata portal to facilitate data discovery

Study metadata

The Canadian Longitudinal Study on Aging (CLSA) is a large, national, long-term study that will follow approximately 50,000 men and women between the ages of 45 and 85 for at least 20 years. The study will collect information on the changing biological, medical, psychological, social, lifestyle and economic aspects of people's lives. These factors will be studied in order to understand how, individually and in combination, they have an impact in both maintaining health and in the development of disease and disability as people age.

Overview

Acronym	CLSA
Website	CLSA website
Contact	Dr. Ine Wauben (McMaster University) Roxanne Cheeseman (McMaster University) Canadian Longitudinal Study on Aging (National Coordinating Centre) Canadian Longitudinal Study on Aging (Statistical Analysis Centre)
Investigator	Dr. Parminder Raina (McMaster University) Roxanne Cheeseman (McMaster University) Dr. Susan Kirkland (Dalhousie University)
Study Start Year	2002

Design

Study Design	cohort study
Recruitment Target	individuals
Target Number of Participants	50,000
Target Number of Participants with Biological Samples	30,000

Access

Access to external researchers or third parties provided or foreseen for:

Data (questionnaire-derived, measured...)	✔
Biological Samples	✔

Marker Paper

Raina PS, Wolfson C, Kirkland SA, Griffith LE, Oremus M, Patterson C, Tuokko H, Penning M, Balion CM, Hogan D, Wister A, Payette H, Shannon H, and Brazil K, The Canadian longitudinal study on aging (CLSA). Can J Aging, 2009. 28(3): p. 221-9.

PUBMED 19860977

General overview/study design

Timeline

Each colour in the timeline graph below represents a separate Study Population, while each segment in the graph represents a separate Data Collection Event. Clicking on a segment gives more detailed information on a Data Collection Event.

Study timeline

Study metadata

The Canadian Longitudinal Study on Aging (CLSA) is a large, national, long-term study that will follow approximately 50,000 men and women between the ages of 45 and 85 for at least 20 years. The study will collect information on the changing biological, medical, psychological, social, lifestyle and economic aspects of people's lives. These factors will be studied in order to understand how, individually and in combination, they have an impact in both maintaining health and in the development of disease and disability as people age.

Populations

- CLSA sub-population (Telephone interview)
- CLSA sub-population (In-depth)

CLSA sub-population (Telephone interview)

Representative sample of the Canadian population.

Sources of Recruitment

Participants from Existing Studies	Canadian Community Health Survey (CCHS) - Healthy Aging
Supplementary Information	CCHS cycle 4.2 would be used as the recruitment vehicle for the telephone interview cohort.

Selection Criteria

Age	Minimum 45, Maximum 85
Country	Canada
Health Status	Exclusion of cognitive impaired individuals
Other	Language: Individuals who are able to respond in either French or English. The CLSA uses the same exclusion criteria as the Statistics Canada Canadian Community Health Survey – Healthy Aging. Excluded from the study are: • Residents of the three territories • Full-time members of the Canadian Forces • Individuals living in long-term care institutions (i.e., those providing 24-hour nursing care). However, those living in households and transitional housing arrangements (e.g., seniors' residences, in which only minimal care is provided) will be included. CLSA cohort participants who become institutionalized during the course of the study will continue to be followed either through personal or proxy interview. • Persons living on reserves and other Aboriginal settlements. However, individuals who are of First Nations descent who live outside reserves are included in the study.

Sample Size

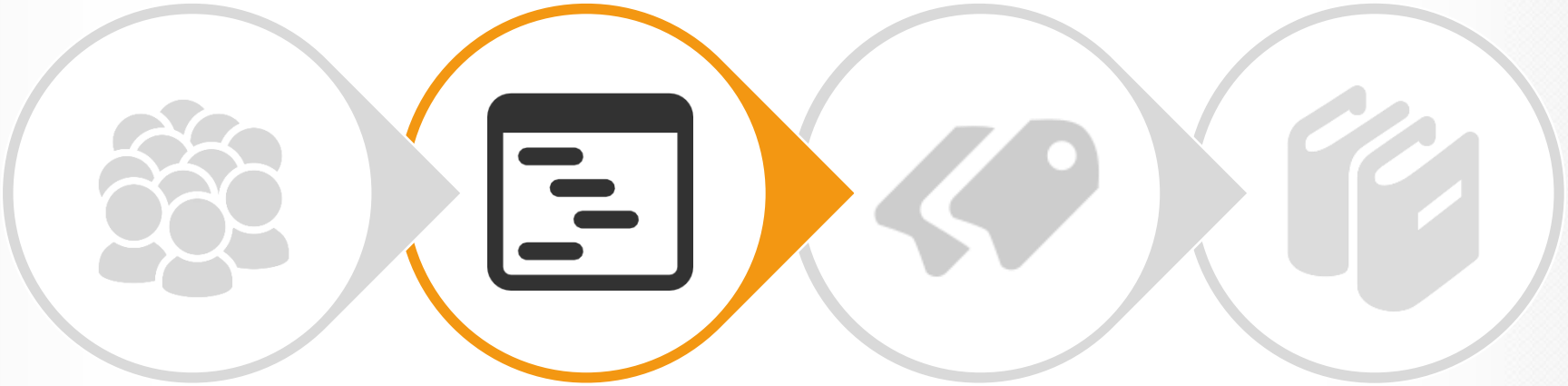
Number of Participants	20,000

Data Collection Events

Name	Description	Start	End
Baseline Recruitment	The first selection of the participants of the...	2008 (January)	2013 (December)
Follow-up one	The first follow-up of the CLSA participants...	2015 (January)	2018 (December)

Sub-populations (sources of recruitment, inclusion/exclusion criteria, number of participants)

Metadata cataloguing with



Study metadata

Document:

- General study design
- Participant selection criteria
- Data collection events
- Information on access to data and biosamples

Variable metadata

Document:

- Variable name/label
- Categories
- Units
- Assessment items
- SOPs

Areas of information

- Classify variables according to area of information (e.g. tobacco, neoplasms, anthropometry)

Metadata catalogue

End product:

- Fully searchable metadata portal to facilitate data discovery

Variable metadata

The Canadian Longitudinal Study on Aging (CLSA) is a large, national, long-term study that will follow approximately 50,000 men and women between the ages of 45 and 85 for at least 20 years. The study will collect information on the changing biological, medical, psychological, social, lifestyle and economic aspects of people's lives. These factors will be studied in order to understand how, individually and in combination, they have an impact in both maintaining health and in the development of disease and disability as people age.

Name	Label	Type	Study/Network	Dataset
AGE_NMBR_COM	Age (years)	Study	CLSA	CLSA_Comprehensive
AGE_NMBR_TRM	Age (years)	Study	CLSA	CLSA_tracking
CAG_GNDR_COM	Gender of person who participant provided most care giving assistance	Study	CLSA	CLSA_Comprehensive
CAG_GNDR_TRM	Gender of person who participant provided most care giving assistance	Study	CLSA	CLSA_tracking
CAG_MOST_COM	Dwelling location of person who participant provided most care giving assistance	Study	CLSA	CLSA_Comprehensive
CAG_MOST_TRM	Dwelling location of person who participant provided most care giving assistance	Study	CLSA	CLSA_tracking
CR2_AGE_NB_COM	Age of person who provided most non-professional assistance	Study	CLSA	CLSA_Comprehensive
CR2_AGE_NB_TRM	Age of person who provided most non-professional assistance	Study	CLSA	CLSA_tracking
CR2_GNDR_COM	Gender of person who provided most non-professional assistance	Study	CLSA	CLSA_Comprehensive
CR2_GNDR_TRM	Gender of person who provided most non-professional assistance	Study	CLSA	CLSA_tracking
CR2_PERS_COM	Dwelling location of person who provided most time for non-professional assistance	Study	CLSA	CLSA_Comprehensive
CR2_PERS_TRM	Dwelling location of person who provided most time for non-professional assistance	Study	CLSA	CLSA_tracking
ED_ELHS_COM	Education highest elementary or high school grade	Study	CLSA	CLSA_Comprehensive
ED_ELHS_TRM	Education highest elementary or high school grade	Study	CLSA	CLSA_tracking
ED_HIGH_COM	Education highest degree	Study	CLSA	CLSA_Comprehensive
ED_HIGH_OTSP_COM	Education highest degree other, Specify	Study	CLSA	CLSA_Comprehensive
ED_HIGH_OTSP_TRM	Education highest degree other, Specify	Study	CLSA	CLSA_tracking
ED_HIGH_TRM	Education highest degree	Study	CLSA	CLSA_tracking
ED_HSGR_COM	Education high school graduated	Study	CLSA	CLSA_Comprehensive
ED_HSGR_TRM	Education high school graduated	Study	CLSA	CLSA_tracking
ED_OTED_COM	Education other degree	Study	CLSA	CLSA_Comprehensive
ED_OTED_TRM	Education other degree	Study	CLSA	CLSA_tracking
ED_UDR04_TRM	Highest Level of Education - Respondent, 4 Levels	Study	CLSA	CLSA_tracking
ED_UDR11_TRM	Highest Level of Education - Respondent, 11 Levels	Study	CLSA	CLSA_tracking
INC_FRST_TRM	Highest source of household income	Study	CLSA	CLSA_tracking
INC_FRST_TRM	Highest source of household income	Study	CLSA	CLSA_Comprehensive

Variable metadata

The Canadian Longitudinal Study on Aging (CLSA) is a large, national, long-term study that will follow approximately 50,000 men and women between the ages of 45 and 85 for at least 20 years. The study will collect information on the changing biological, medical, psychological, social, lifestyle and economic aspects of people's lives. These factors will be studied in order to understand how, individually and in combination, they have an impact in both maintaining health and in the development of disease and disability as people age.

ED_HIGH_COM

Overview

Label	Education highest degree
Description	What is the highest degree, certificate, or diploma you have obtained?
Study	CLSA
Dataset	CLSA_Comprehensive
Value Type	text
Variable Type	Study Variable

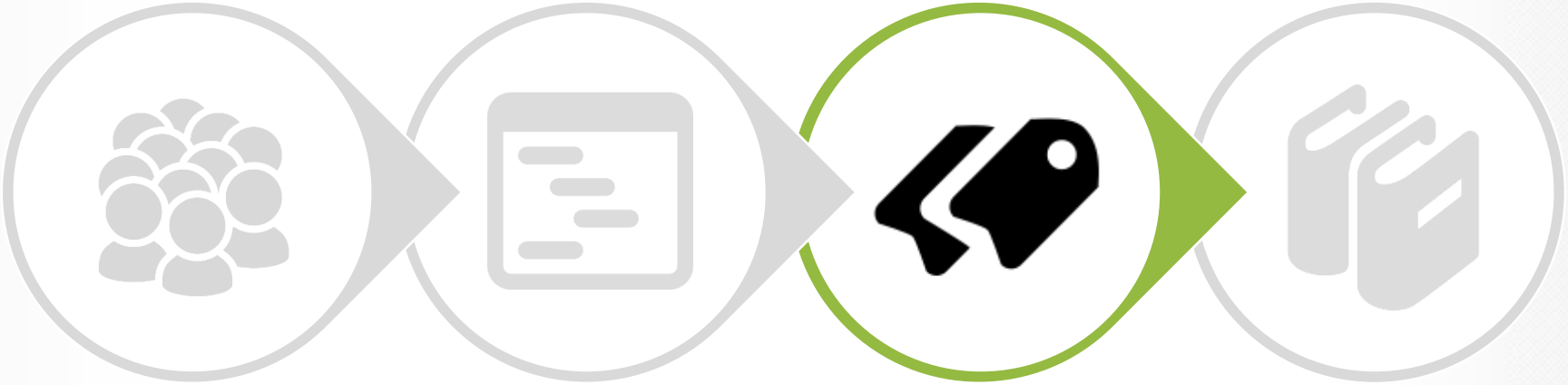
Classification

Additional information	
Source	Questionnaire
Target	Participant
Areas of Information	
Socio-demographic and economic characteristics	Education

Categories

Name	Label	Missing
01	No post-secondary degree, certificate, or diploma	
02	Trade certificate or diploma from a vocational school or apprenticeship training	
03	Non-university certificate or diploma from a community college, CEGEP, school of nursing, etc.	
04	University certificate below bachelor's level	
05	Bachelor's degree	
06	University degree or certificate above bachelor's degree	
97	Other (please specify)	
98	Don't know/ No answer	✔
99	Refused	✔

Metadata cataloguing with mica



Study metadata

Document:

- General study design
- Participant selection criteria
- Data collection events
- Information on access to data and biosamples

Variable metadata

Document:

- Variable name/label
- Categories
- Units
- Assessment items
- SOPs

Areas of information

- Classify variables according to area of information (e.g. tobacco, neoplasms, anthropometry)

Metadata catalogue

End product:

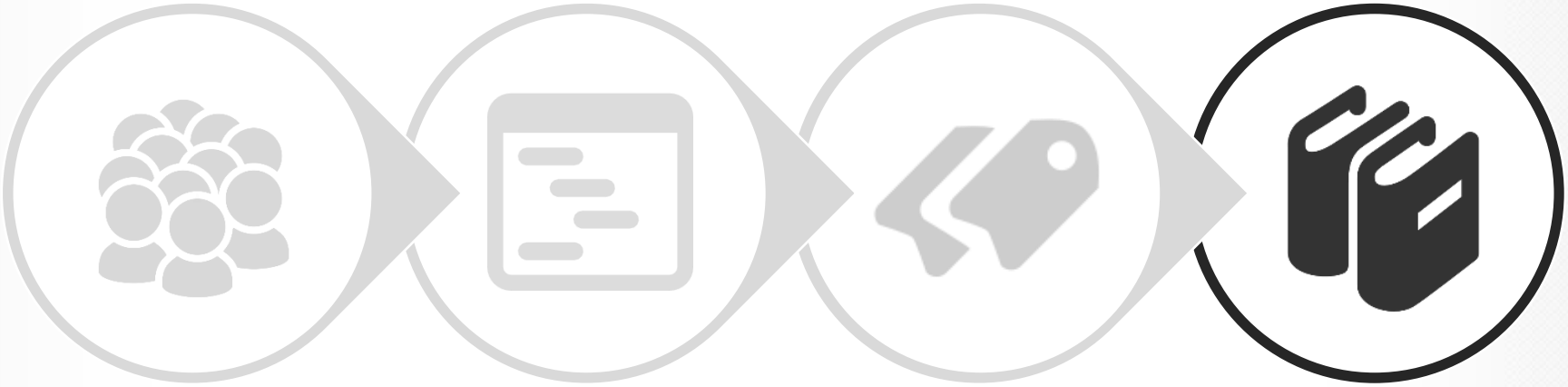
- Fully searchable metadata portal to facilitate data discovery

Making variables 'discoverable': Areas of information

17 sections 132 categories

- Socio-demographic and economic characteristics
- Health status and functional limitations
- Diseases (ICD-10)
- **Lifestyle and health behaviours**
- Symptoms and signs (ICD-10)
- Medications and supplements
- Non-pharmacological interventions
- Health and community care utilization
- Reproduction
- Birth, infancy and childhood
- End of life
- Physical measures
- Cognition, personality and other psychological measures
- Laboratory measures
- Social environment and life events
- Physical environment
- Administrative information

Metadata cataloguing with Mica



Study metadata

Document:

- General study design
- Participant selection criteria
- Data collection events
- Information on access to data and biosamples

Variable metadata

Document:

- Variable name/label
- Categories
- Units
- Assessment items
- SOPs

Areas of information

- Classify variables according to area of information (e.g. tobacco, neoplasms, anthropometry)

Metadata catalogue

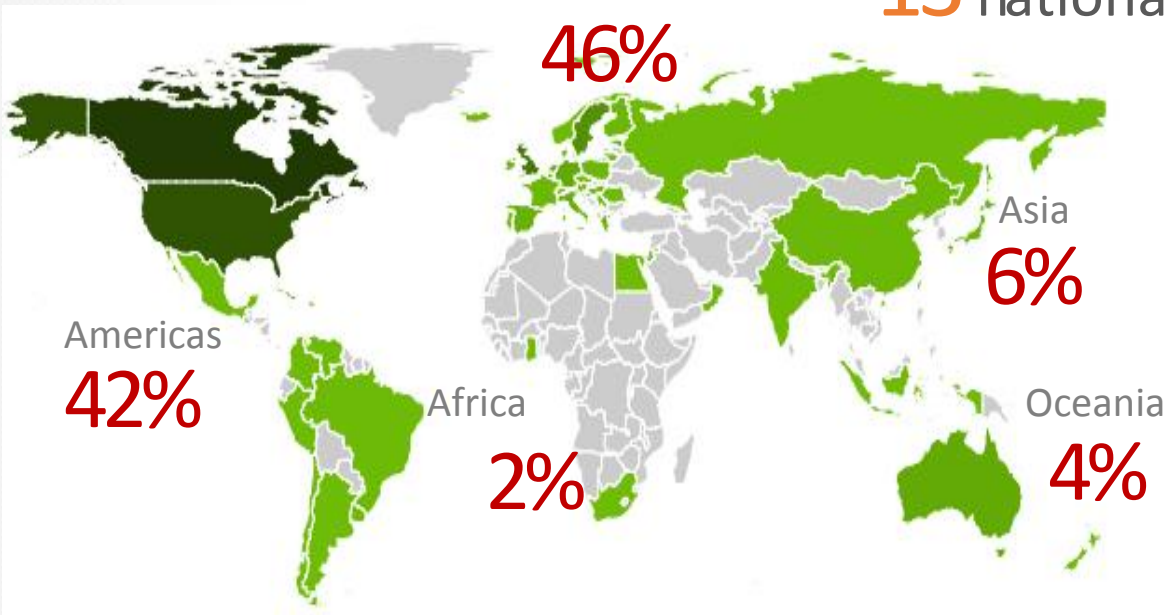
End product:

- Fully searchable metadata portal to facilitate data discovery

Maelstrom metadata catalogue

mælstrom

www.maelstrom-research.org/maelstrom-catalogue



13 national and international networks

182 studies (**102** with variables)

6,241,758 participants

762,959 variables

182 studies, including...

clsa élcv
Canadian Longitudinal Study on Aging
Étude longitudinale canadienne sur le vieillissement

biobank uk

Nurses' Health Study

Canadian Partnership for Tomorrow Project

LifeLines

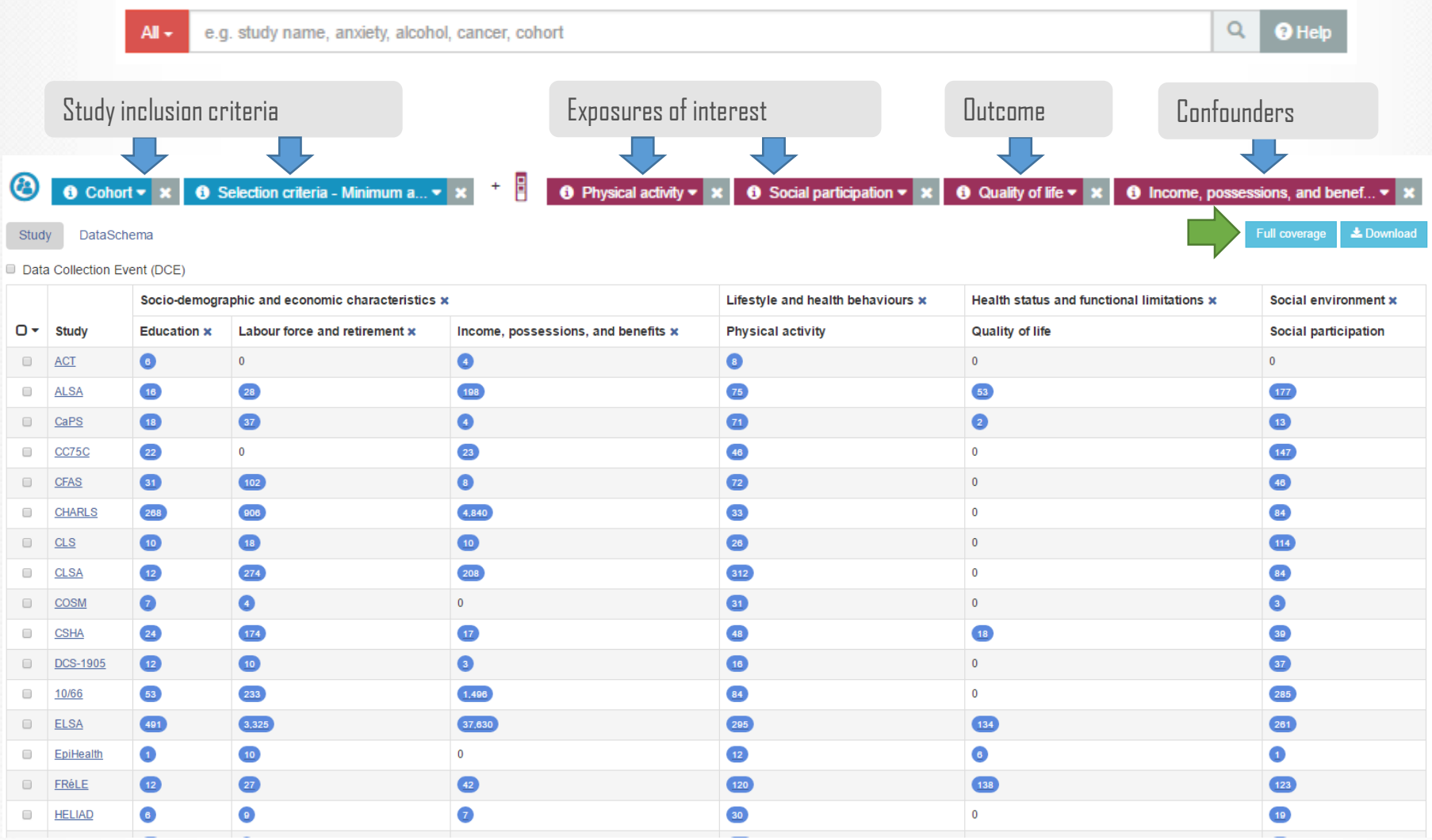
NTNU
HUNT Research Centre

epic
European Prospective Investigation of Cancer

CONSTANCES

Metadata search

"I am interested in pooling data across cohort studies to explore the effect of physical activity and social participation on quality of life in older adults, adjusting for SES"



		Socio-demographic and economic characteristics ×			Lifestyle and health behaviours ×	Health status and functional limitations ×	Social environment ×
	Study	Education ×	Labour force and retirement ×	Income, possessions, and benefits ×	Physical activity	Quality of life	Social participation
	ACT	6	0	4	8	0	0
	ALSA	16	28	198	75	53	177
	CaPS	18	37	4	71	2	13
	CC75C	22	0	23	46	0	147
	CFAS	31	102	8	72	0	46
	CHARLS	268	906	4,840	33	0	84
	CLS	10	18	10	26	0	114
	CLSA	12	274	208	312	0	84
	COSM	7	4	0	31	0	3
	CSHA	24	174	17	48	18	39
	DCS-1905	12	10	3	16	0	37
	10/66	53	233	1,496	84	0	285
	ELSA	491	3,325	37,630	295	134	261
	EpiHealth	1	10	0	12	6	1
	FRéLE	12	27	42	120	138	123
	HELIAD	6	9	7	30	0	19

Metadata search

"I am interested in pooling data across cohort studies to explore the effect of physical activity and social participation on quality of life in older adults, adjusting for SES"

Displays only studies collecting all variables of interest for the research project

Study | DataSchema

Download

Data Collection Event (DCE)

	Study	Socio-demographic and economic characteristics ×			Lifestyle and health behaviours ×	Health status and functional limitations ×	Social environment ×
		Education ×	Labour force and retirement ×	Income, possessions, and benefits ×	Physical activity	Quality of life	Social participation
	ALSA	16	28	198	75	53	177
	CaPS	18	37	4	71	2	13
	CSHA	24	174	17	48	18	39
	ELSA	491	3,325	37,630	295	134	261
	FRéLE	12	27	42	120	138	123
	NuAge	1	5	14	430	256	93
	OATS	34	43	6	121	40	59
	PATH	156	228	77	150	261	44
	TILDA	16	99	374	21	28	46
	VETSA	21	28	9	24	54	1
	All	789	3,994	38,371	1,355	984	856

Metadata search

"I am interested in pooling data across cohort studies to explore the effect of physical activity and social participation on quality of life in older adults, adjusting for SES"

Displays variables collected per population and data collection event

	Data Collection Event (DCE)			Socio-demographic and economic characteristics ×			Lifestyle and health behaviours ×	Health status and functional limitations ×	Social environment ×
☐ ▾	Study	Population	DCE	Education ×	Labour force and retirement ×	Income, possessions, and benefits ×	Physical activity	Quality of life	Social participation
☐	CaPS	Caerphilly Cohort	Phase I 1979-07 to 1983-09	0	11	0	6	0	0
☐			Phase II 1984-07 to 1988-06	0	13	0	59	0	0
☐			Phase III 1989-11 to 1993-09	16	6	0	0	2	5
☐			Phase IV 1993-10 to 1997-02	0	2	0	4	0	2
☐			Phase V 2002-01 to 2004-12	2	4	4	2	0	6
☐			Follow-up research 1997-03 to 2016-12	0	1	0	0	0	0
☐	CSHA	CSHA Caregivers	CSHA-1 Caregivers 1991-02 to 1992-05	2	17	0	0	0	0
☐			CSHA-2 Caregivers 1996-01 to 1997-12	2	25	2	13	0	3
☐			CSHA-3 Caregivers 2001-01 to 2002-12	0	23	4	0	13	0
☐		CSHA Institutional sample	CSHA-1 1991-02 to 1992-05	6	49	0	7	0	1
☐			CSHA-2 1996-01 to 1997-12	0	0	3	0	0	0
☐			CSHA-3 2001-01 to 2002-12	3	1	0	0	0	1

Metadata search

"I am interested in pooling data across cohort studies to explore the effect of physical activity and social participation on quality of life in older adults, adjusting for SES"

Displays variables included in a category part of a data collection event

Name	Label	Type	Study	Dataset
cgedyrs	9. How many years of education?	Collected	CSHA	CSHA1_Caregiver
cgedlev	9. Level of education	Collected	CSHA	CSHA1_Caregiver

Metadata search

"I am interested in pooling data across cohort studies to explore the effect of physical activity and social participation on quality of life in older adults, adjusting for SES"

Displays variable details

cgedlev

Overview

Label	9. Level of education
Individual Study	CSHA
Dataset	CSHA1_Caregiver
Value Type	Integer
Variable Type	Collected Variable

Classifications

Additional information	
Source	Questionnaire
Target	Proxy
Areas of information	
Socio-demographic and economic characteristics	Education

Categories

Name	Label	Missing
88	Don't know	✔
99	Missing	✔
6666	NA/Skipped	✔
1	No formal school	
2	Some primary school	
3	Finished primary	
4	Some high school	
5	Finished high school	
6	Some technical college	
7	Finished technical college	
8	Some University	
9	Bachelor degree	
10	Master degree	
11	PhD	
12	Other	

Tools for **data transformation**

Deriving data into a common format

mælstrom

opal in a nutshell

A database application for **storing, managing and transforming** study data from multiple sources

Main features

- **Import data** from different formats (CSV, SPSS, SAS, Stata, R)
- **Store data** on an unlimited number of variables using standardized data dictionaries
- **Transform data** into common (i.e. harmonized) formats

Harmonization potential evaluation

***Target variable*: Highest level of education attained**

Study A data

1	Primary school 7-10 years, continuation school, folk high school
2	High school, intermediate school, vocational school, 1-2 years high school
3	University qualification examination, senior high school
4	University or other post-secondary education, less than 4 years
5	University / college, 4 years or more
9	Missing

Study B data

1	Primary school
2	Lower vocational school
3	Lower secondary education
4	Secondary vocational education and training
5	Higher secondary education
6	Higher professional education
7	University
98	Not applicable
99	Missing

Study C data

1	No education
2	Primary education completed
3	Lower or pre-vocational
4	Junior general secondary education
5	Secondary vocational or apprenticeship
6	Higher general and pre-university
7	Higher professional
8	University education
9	Other

UNESCO International Standard Classification of Education **ISCED 2011**

ISCED level 0	Early childhood education
ISCED level 1	Primary education
ISCED level 2	Lower secondary education
ISCED level 3	Upper secondary education
ISCED level 4	Post-secondary non-tertiary education
ISCED level 5	Short-cycle tertiary education
ISCED level 6	Bachelor's or equivalent level
ISCED level 7	Master's or equivalent level
ISCED level 8	Doctoral or equivalent level

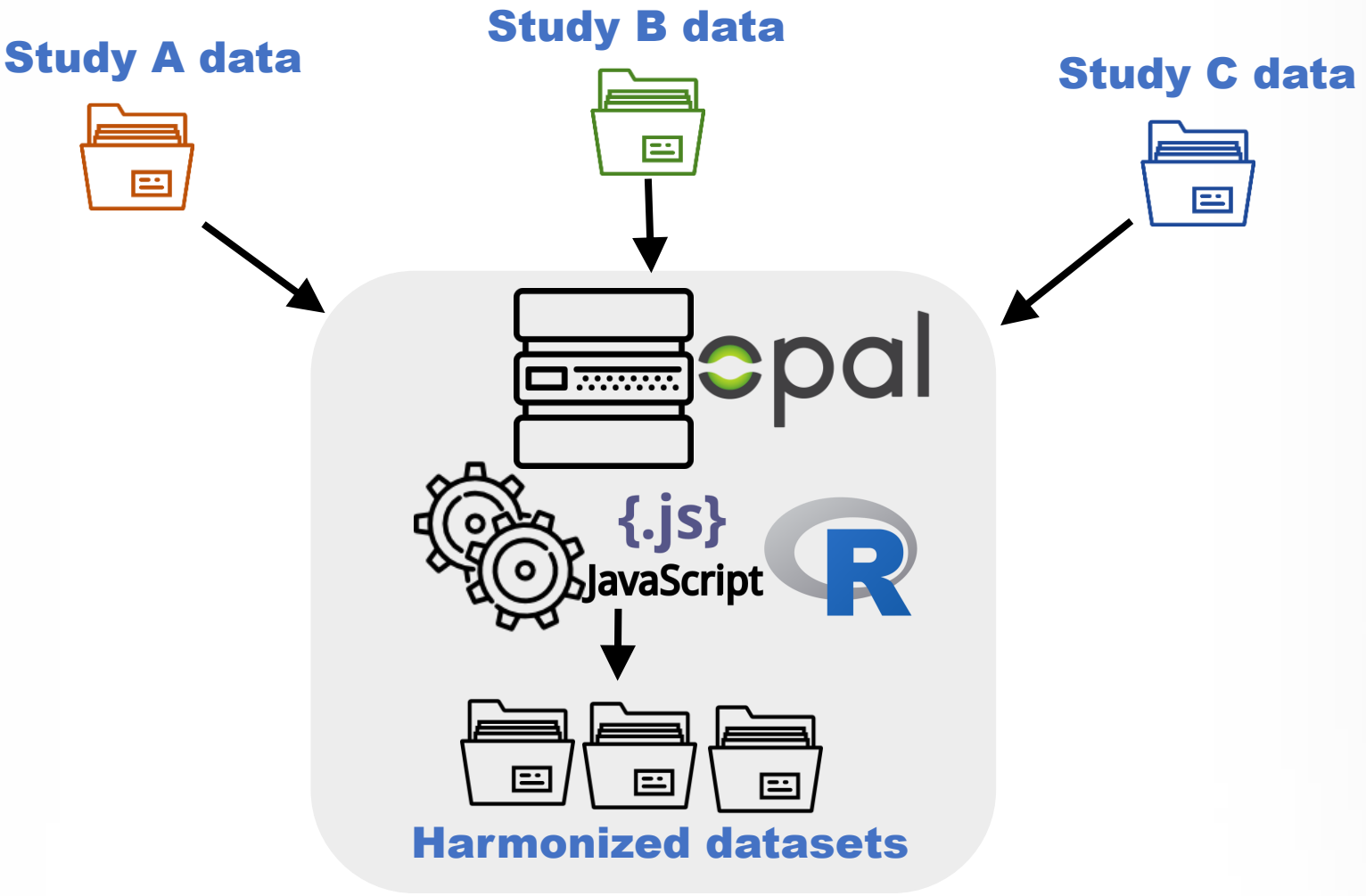
Final harmonized variable:

Highest level of education attained

0	No education/Primary education (ISCED 0-1)
1	Secondary education (ISCED 2-3)
2	Higher education (including vocational/professional education, college and university) (ISCED 4-8)
9	Missing

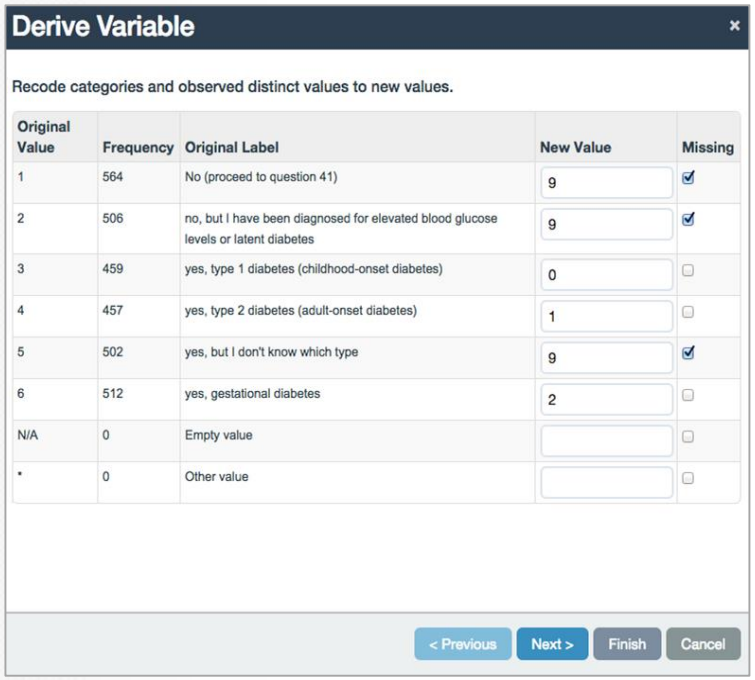
Data transformation in Opal

Opal allows executing R or JavaScript code to transform study-specific data



Data transformation in Opal

Graphic interface and JS or R scripting interface



User-friendly interfaces for recoding variables

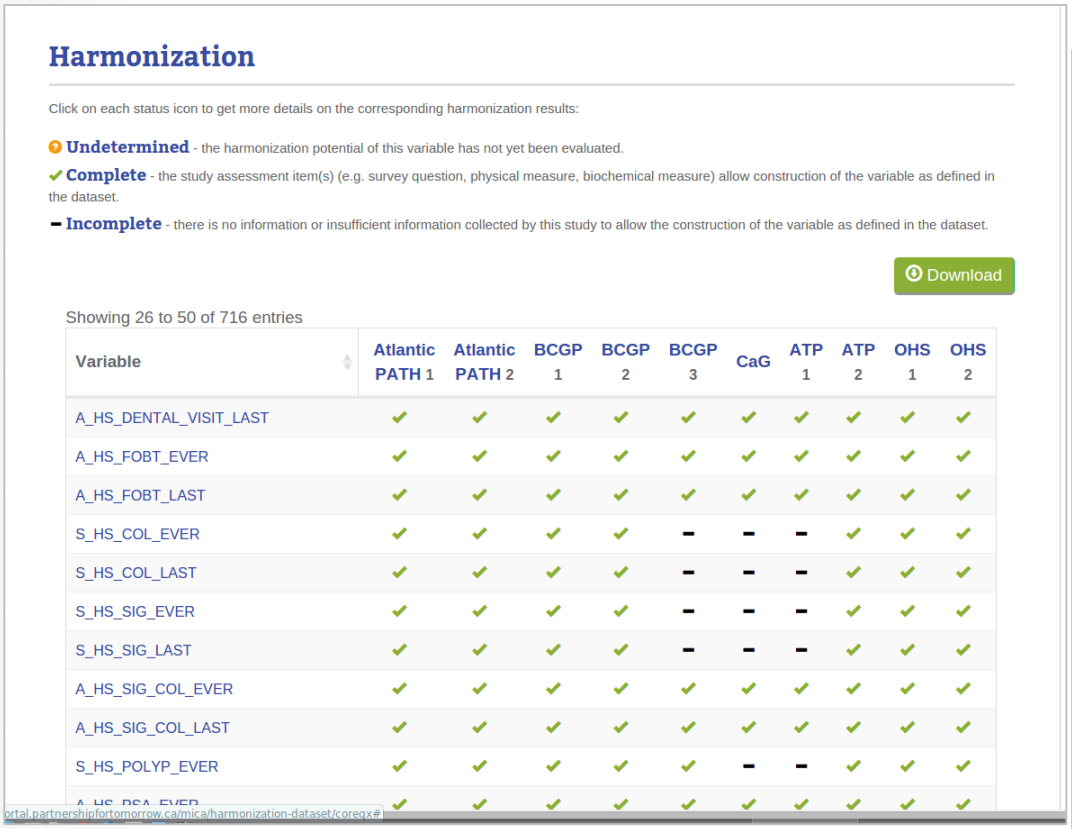
JavaScript or R interface for more advanced transformations

{.js}
JavaScript

R

Harmonization results documentation

Opal communicates with Mica



Harmonization

Click on each status icon to get more details on the corresponding harmonization results:

Undetermined - the harmonization potential of this variable has not yet been evaluated.

Complete - the study assessment item(s) (e.g. survey question, physical measure, biochemical measure) allow construction of the variable as defined in the dataset.

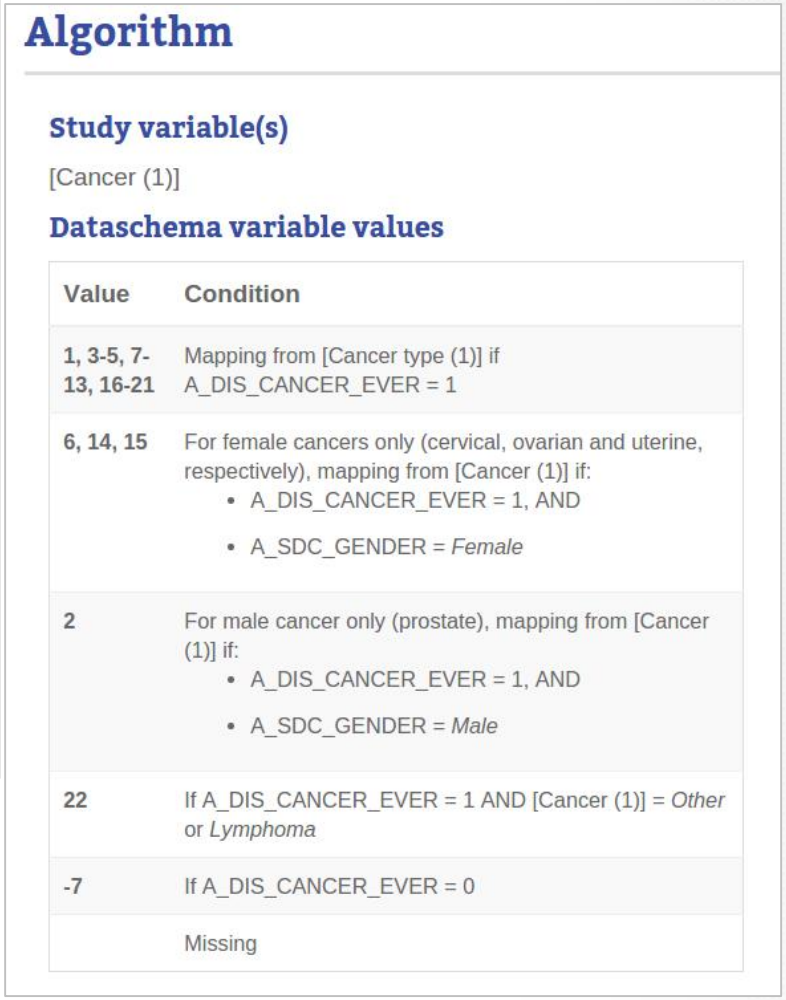
Incomplete - there is no information or insufficient information collected by this study to allow the construction of the variable as defined in the dataset.

Download

Showing 26 to 50 of 716 entries

Variable	Atlantic PATH 1	Atlantic PATH 2	BCGP 1	BCGP 2	BCGP 3	CaG	ATP 1	ATP 2	OHS 1	OHS 2
A_HS_DENTAL_VISIT_LAST	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
A_HS_FOBT_EVER	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
A_HS_FOBT_LAST	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
S_HS_COL_EVER	✓	✓	✓	✓	–	–	–	✓	✓	✓
S_HS_COL_LAST	✓	✓	✓	✓	–	–	–	✓	✓	✓
S_HS_SIG_EVER	✓	✓	✓	✓	–	–	–	✓	✓	✓
S_HS_SIG_LAST	✓	✓	✓	✓	–	–	–	✓	✓	✓
A_HS_SIG_COL_EVER	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
A_HS_SIG_COL_LAST	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
S_HS_POLYP_EVER	✓	✓	✓	✓	✓	–	–	✓	✓	✓
A_HS_PSA_EVER	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

ortal.partnershipfortomorrow.ca/mica/harmonization-dataset/coreqx#



Algorithm

Study variable(s)

[Cancer (1)]

Dataschema variable values

Value	Condition
1, 3-5, 7-13, 16-21	Mapping from [Cancer type (1)] if A_DIS_CANCER_EVER = 1
6, 14, 15	For female cancers only (cervical, ovarian and uterine, respectively), mapping from [Cancer (1)] if: • A_DIS_CANCER_EVER = 1, AND • A_SDC_GENDER = *Female*
2	For male cancer only (prostate), mapping from [Cancer (1)] if: • A_DIS_CANCER_EVER = 1, AND • A_SDC_GENDER = *Male*
22	If A_DIS_CANCER_EVER = 1 AND [Cancer (1)] = *Other* or *Lymphoma*
-7	If A_DIS_CANCER_EVER = 0
	Missing

Researchers have an overview of which variables are harmonized across studies

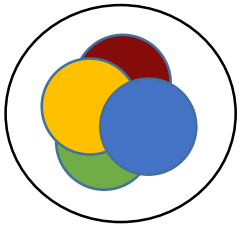
Tools for **data analysis**

Co-analyzing harmonized data across cohorts

Data infrastructures/analysis

Pooled analysis

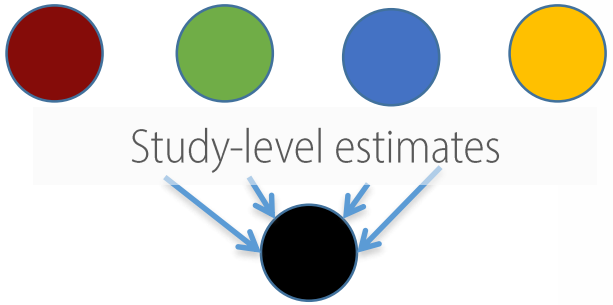
Data pooled and analyzed in a central location



Sharing of IPD

Summary data meta-analysis

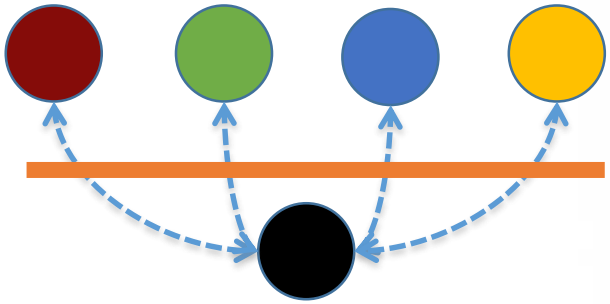
Study-specific data analyses done locally locally followed by a meta-analysis combining the study-level estimates



No sharing of IPD

Federated analysis

Analyses done on a central computer, but but individual participant data remain on remain on local servers



Sharing or not of IPD

Data infrastructures/analysis

Pooled analysis

Data pooled and analyzed in a central location

Sharing of IPD

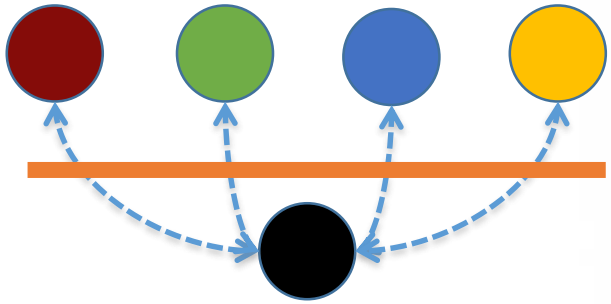
Summary data meta-analysis

Study-specific data analyses done locally locally followed by a meta-analysis combining the study-level estimates

No sharing of IPD

Federated analysis

Analyses done on a central computer, but computer, but individual participant data data remain on local servers



Sharing **(or not)** of IPD

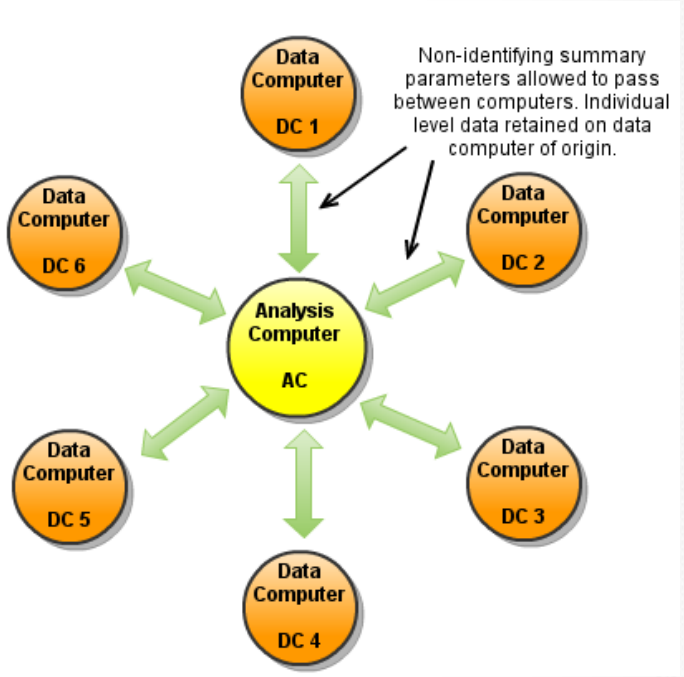
DataSHIELD Secure Bioscience Collaboration in a nutshell

A statistical method and software to **perform pooled data analysis without sharing individual-level data**

"take analysis to data … not data to analysis"

Main features:

- **Remote analyses:** investigators analyse data at their own convenience (via secure web connections)
- **Iterative analyses:** parallel processes linked together by non-identifying summary statistics – *e.g.* for glm = score vectors and information matrices
- Limited to functions in *DataSHIELD R package*



Real-time federated analyses

mælstrom

Researcher's computer

Analysis tools

mica

Descriptive statistics, histograms

or

DataSHIELD
Secure Bioscience Collaboration

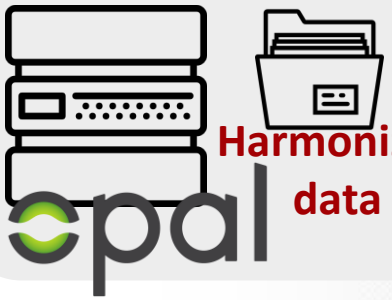
Functions in DataSHIELD R package

or

RStudio®

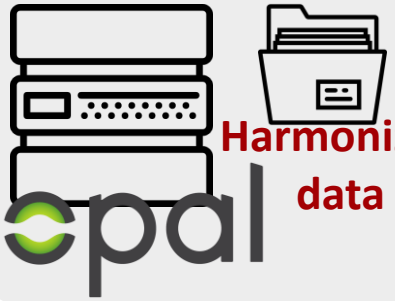
Any R function in any R package

Study A server



Harmonized data

Study B server



Harmonized data

Study C server

Harmonized data

Some projects making use of our tools

mælstrom

10 adult cohorts
Metabolic, enviro. exposures
Harmonization, Federated analyses (DataSHIELD)

12 adult cohorts
Cancer, menopause
Federated analysis

109 adult studies
Healthy aging
Cataloguing, Meta-analyses

4 mother/child cohorts
Developmental origins
Prospective harmonization

24 mother/child cohorts
Developmental origins
Cataloguing, harmonization

10 aging cohorts
Healthy aging, urban form
Harmonization, federated analysis

19 aging studies
Healthy aging
Harmonization, federated analysis (DataSHIELD)

260 adult studies
Diabetes
Cataloguing, harmonization, federated analysis(DataSHIELD)

5 adult cohorts
Cancer, chronic diseases
Cataloguing, harmonization

More information….

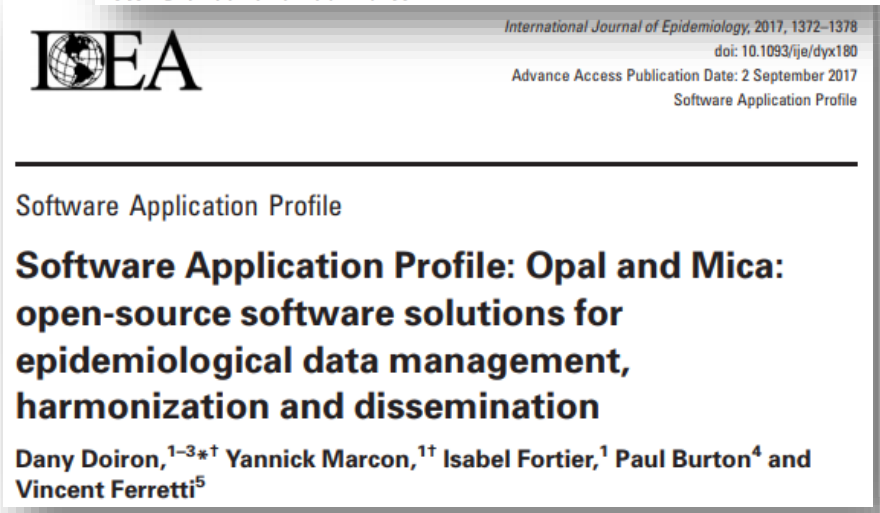
www.maelstrom-research.org

International Journal of Epidemiology, 2016, 1–13
doi: 10.1093/ije/dyw075
Original Article

Original Article

Maelstrom Research guidelines for rigorous retrospective data harmonization

Isabel Fortier,[1]* Parminder Raina,[2] Edwin R Van den Heuvel,[3] Lauren E Griffith,[2] Camille Craig,[1] Matilda Saliba,[1] Dany Doiron,[1] Ronald P Stolk,[4] Bartha M Knoppers,[5] Vincent Ferretti,[6] Peter Granda[7] and Paul Burton[8]

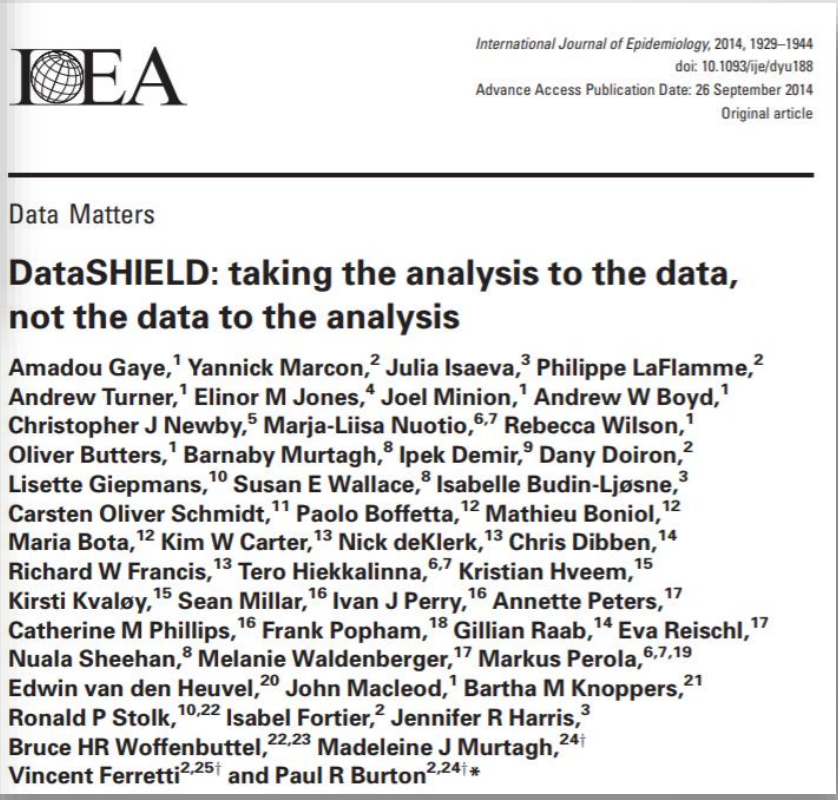


International Journal of Epidemiology, 2017, 1372–1378
doi: 10.1093/ije/dyx180
Advance Access Publication Date: 2 September 2017
Software Application Profile

Software Application Profile

Software Application Profile: Opal and Mica: open-source software solutions for epidemiological data management, harmonization and dissemination

Dany Doiron,[1–3]*† Yannick Marcon,[1†] Isabel Fortier,[1] Paul Burton[4] and Vincent Ferretti[5]



International Journal of Epidemiology, 2014, 1929–1944
doi: 10.1093/ije/dyu188
Advance Access Publication Date: 26 September 2014
Original article

Data Matters

DataSHIELD: taking the analysis to the data, not the data to the analysis

Amadou Gaye,[1] Yannick Marcon,[2] Julia Isaeva,[3] Philippe LaFlamme,[2] Andrew Turner,[1] Elinor M Jones,[4] Joel Minion,[1] Andrew W Boyd,[1] Christopher J Newby,[5] Marja-Liisa Nuotio,[6,7] Rebecca Wilson,[1] Oliver Butters,[1] Barnaby Murtagh,[8] Ipek Demir,[9] Dany Doiron,[2] Lisette Giepmans,[10] Susan E Wallace,[8] Isabelle Budin-Ljøsne,[3] Carsten Oliver Schmidt,[11] Paolo Boffetta,[12] Mathieu Boniol,[12] Maria Bota,[12] Kim W Carter,[13] Nick deKlerk,[13] Chris Dibben,[14] Richard W Francis,[13] Tero Hiekkalinna,[6,7] Kristian Hveem,[15] Kirsti Kvaløy,[15] Sean Millar,[16] Ivan J Perry,[16] Annette Peters,[17] Catherine M Phillips,[16] Frank Popham,[18] Gillian Raab,[14] Eva Reischl,[17] Nuala Sheehan,[8] Melanie Waldenberger,[17] Markus Perola,[6,7,19] Edwin van den Heuvel,[20] John Macleod,[1] Bartha M Knoppers,[21] Ronald P Stolk,[10,22] Isabel Fortier,[2] Jennifer R Harris,[3] Bruce HR Woffenbuttel,[22,23] Madeleine J Murtagh,[24†] Vincent Ferretti[2,25†] and Paul R Burton[2,24†*]

mælstrom

www.maelstrom-research.org

Funding and support:

CANADIAN PARTNERSHIP AGAINST CANCER
PARTENARIAT CANADIEN CONTRE LE CANCER

Centre universitaire de santé McGill
Institut de recherche

McGill University Health Centre
Research Institute

Économie, Innovation et Exportations
Québec

National Institute on Aging

SEVENTH FRAMEWORK PROGRAMME

INNOVATION.CA
CANADA FOUNDATION FOR INNOVATION | FONDATION CANADIENNE POUR L'INNOVATION

CIHR IRSC
Canadian Institutes of Health Research | Instituts de recherche en santé du Canada

MUHC
McGILL UNIVERSITY HEALTH CENTRE
FOUNDATION