Towards FAIR Human Biomonitoring Data

Development of a tool to enhance HBM data harmonization

Ruben Peeters

21/03/2024 – Berlin, ISES Europe 2024

PARC

- Background
- Objective
- Method & Results
- Conclusion

Background

Heterogeneous data in Human Biomonitoring (HBM) studies

- Need for harmonization
  - Improve comparability between studies
  - Draw more correct conclusions
  - Move towards FAIR data
- Our solution
  - Standardized protocol for harmonization
  - Workable tool

FAIR data - why?

- Enhances the **visibility** and **easy access** of research data, **increasing citations** and **use**. (F & A)
- Promotes **collaboration** and **integration** with other datasets and tools across disciplines. (I)
- Ensures data can be effectively reused in future research, maximizing **long-term value and impact**. (R)
- Supports compliance with research standards, fosters **innovation**, and drives **scientific discovery** by making data more usable for the global research community. (FAIR)

FAIR data - how?

- Findable & Accessible:
  - Assign a **persistent identifier** (e.g., DOI) and create **rich metadata** for easy discovery.
  - Choose a **reputable repository** that supports PIDs and clearly state **access conditions**.
- Interoperable:
  - Use **standard, widely recognized formats** and adopt **controlled vocabularies** for data description.
- Reusable:
  - Provide comprehensive **documentation** and apply clear, appropriate **licenses** to inform users about data reuse conditions.

FAIR data - how?

- Findable & Accessible:
  - Assign a persistent identifier (e.g., DOI) and create **rich metadata** for easy discovery.
  - Choose a reputable repository that supports PIDs and clearly state access conditions.
- **Interoperable:**
  - **Use standard, widely recognized formats and adopt controlled vocabularies for data description.**
- Reusable:
  - Provide comprehensive documentation and apply clear, appropriate licenses to inform users about data reuse conditions.

Objective

Goals

- Optimizing harmonization
- Enabling future FAIRification,
- Creation of derived variables and
- Conducting summary statistics.

While

- Using flexible templating,
- Ensuring data integrity,

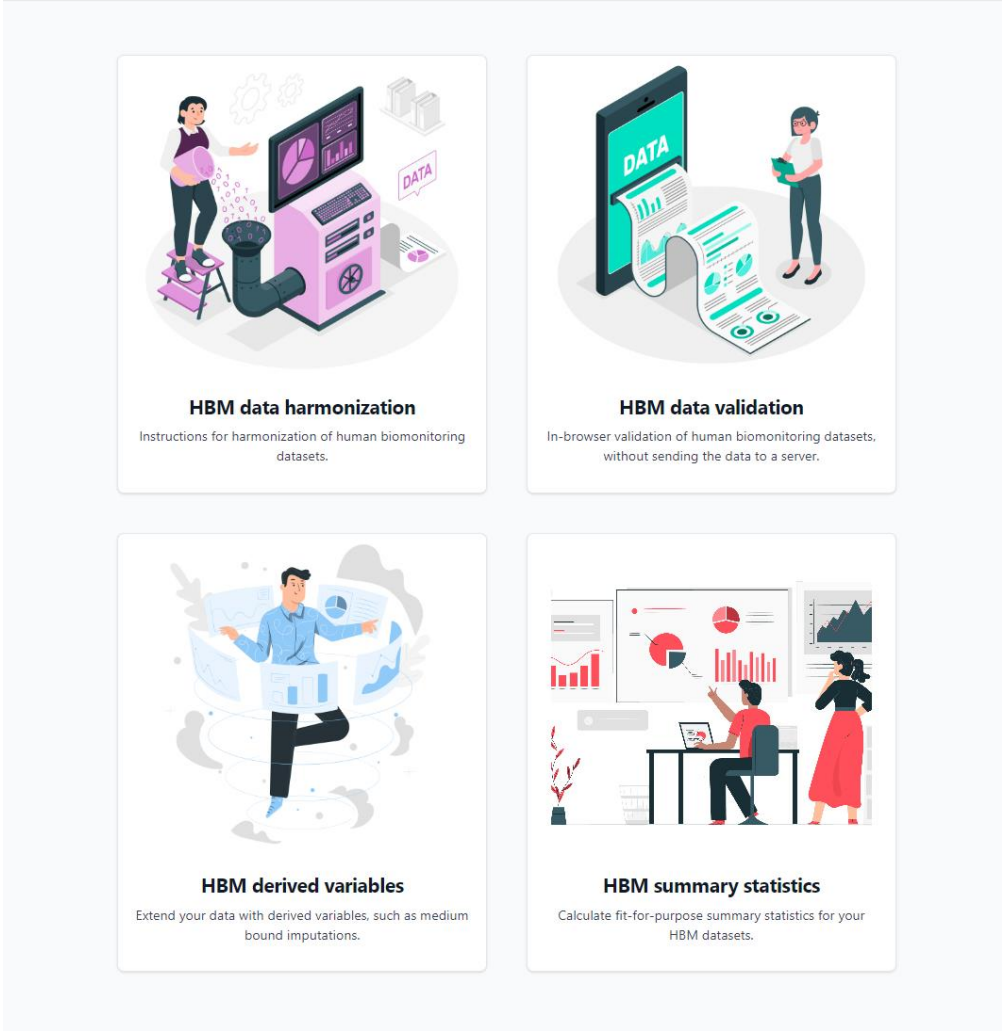
Method & Results

Overview

- Standardized protocol for harmonization
- HBM Tools
- Considerations
  - Usability
  - Data integrity
- FAIRification through Harmonization

Method & Results

- **HBM Tools**
- Personal Exposure & Health (PEH) Platform Storage
- HBM Metadata Workshops Standards
- Compatibility with MCRA platform



https://tools.hbm.vito.be/

source

Data Harmonization

HBM data harmonization

- Basis for harmonization
- List of **Standardized variables**
- Extendable with project-specific variables into **Project Specific Configuration**
- Definition of **protocol**

https://hbm.vito.be/tools/data-harmonization

source

Aligning with standards and formats

- HBM4EU
- PARC
- DCAT
- ISA
- FDO
- molgenis

Data Validation (Quality Control)

HBM data validation

- Automated Data Quality checks
- Individual Level Data
- *Available*

https://tools.hbm.vito.be/validation

source

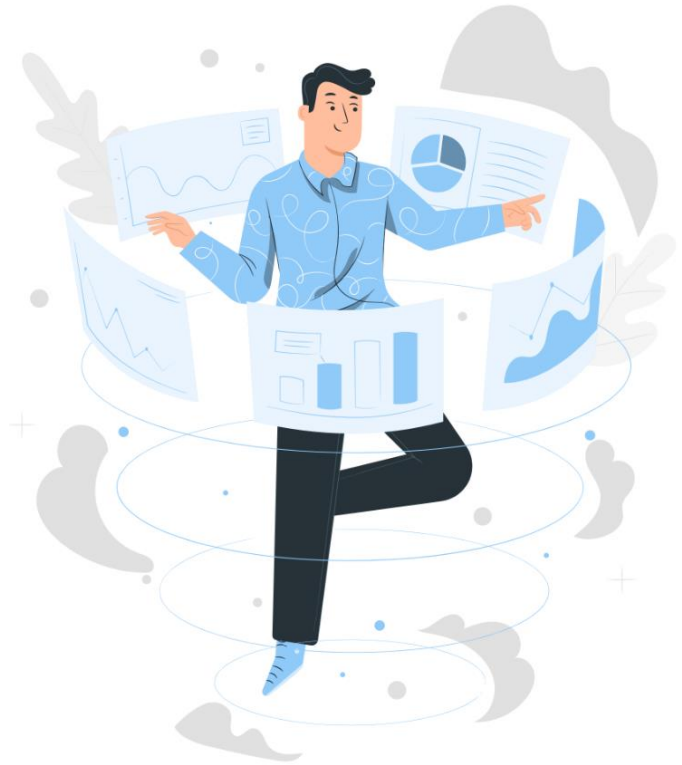
Data Validation (Quality Control)

!	Values are required under condition: matrix ValueIn US;UD;UM but not given.	uvolume	SAMPLETIMEPOINT_US	428 affected values
⚠	Values are extreme outliers. Please check if these values are correct with the lab.	ag	SAMPLETIMEPOINT_BWB	74, 88, 125, 145, 171, 239, 349

Excel rows	Values	Details
74, 88, 125, 145, 171, 239, 349	0.638065075033993, 0.638065075033993, 0.638065075033993, 0.638065075033993, 0.638065075033993, 0.638065075033993, 0.638065075033993	Based on the range defined by [p25-3*IQR;p75+3*IQR] of the natural log transformed values.

⚠	Values are extreme outliers. Please check if these values are correct with the lab.	al	SAMPLETIMEPOINT_BWB	9, 53, 62, 117, 133, 222, 352, 407

Derived Variables



HBM derived variables

- Automated calculation of derived variables
- e.g.: CRT corrected data, LIPID corrections, etc.
- Individual Level Data
- *Available*

https://tools.hbm.vito.be/derivedvariables

source

Derived Variables

DE	DF	DG
bpa	bpa_lod	bpa_loq
0.644	0.15	0.3

DE	DF	DG	DH	DI	DJ	DK	DL	DM	DN
bpa	bpa_lod	bpa_loq	bpa_bin	bpa_meb	bpa_imp	bpa_meb_crt	bpa_imp_crt	bpa_meb_sg	bpa_imp_sg
0.644	0.15	0.3	≥ LOD/LOQ	0.644	0.644	1.4	1.4	0.736	0.736

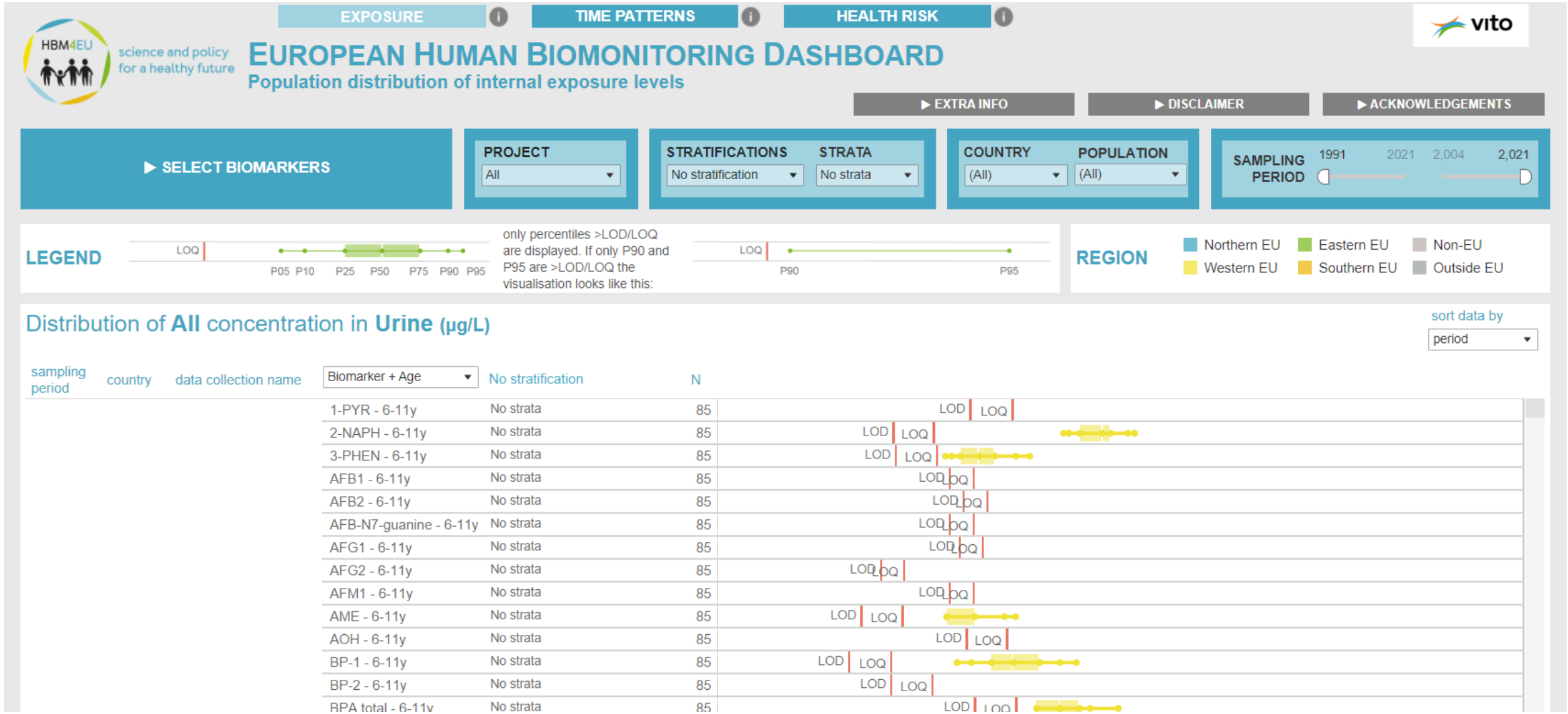
Summary Statistics

HBM summary statistics

- Automated calculation of summary statistics
- Used for the European HBM dashboard.
- e.g., Geometric mean, Percentiles
- *Not yet available*

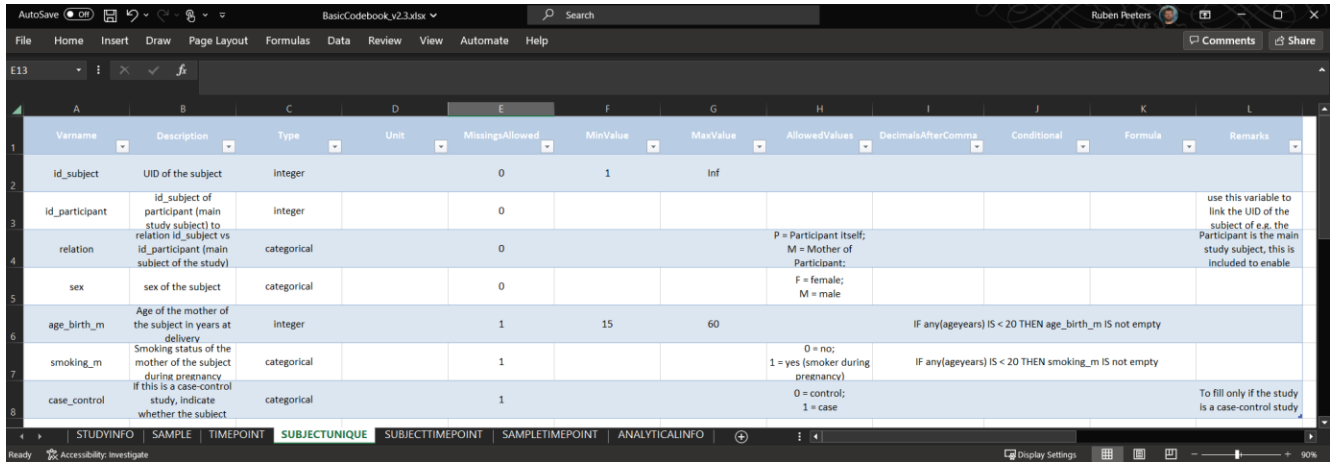
source

Summary Statistics

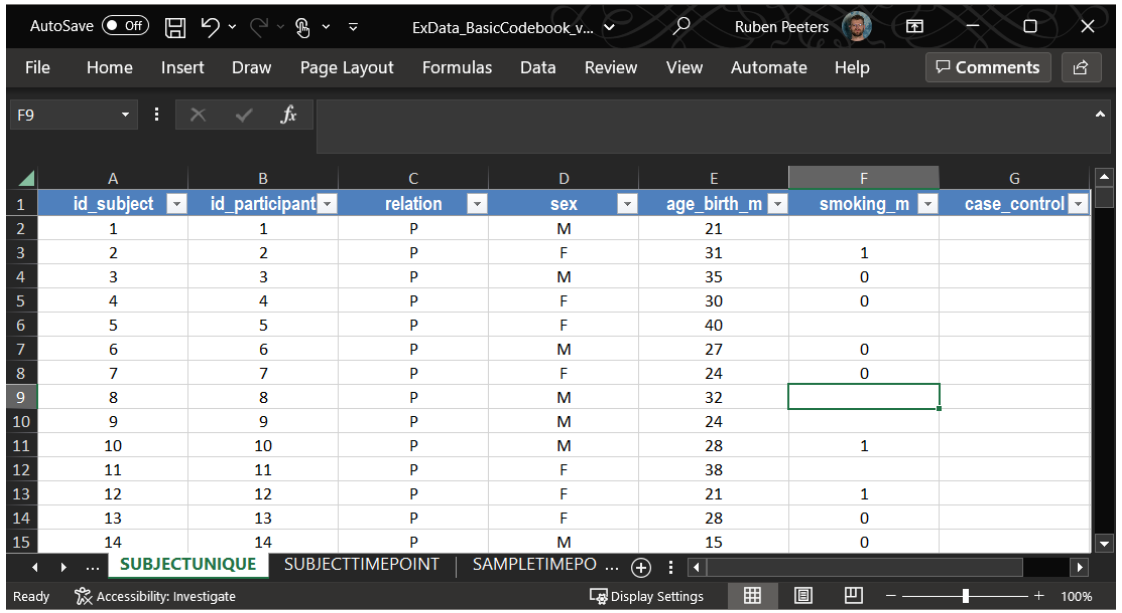


Flexible templating

Excel logo / Alpár-Etele Méder / CC BY 3.0



Varname	Description	Type	Unit	MissingsAllowed	MinValue	MaxValue	AllowedValues	DecimalsAfterComma	Conditional	Formula	Remarks
id_subject	UID of the subject	integer		0	1	Inf					
id_participant	id_subject of participant (main study subject) to	integer		0							use this variable to link the UID of the subject of e.g. the
relation	relation id_subject vs id_participant (main subject of the study)	categorical		0			P = Participant itself; M = Mother of Participant;				Participant is the main study subject, this is included to enable
sex	sex of the subject	categorical		0			F = female; M = male				
age_birth_m	Age of the mother of the subject in years at delivery	integer		1	15	60		IF any(ageyears) IS < 20 THEN age_birth_m IS not empty			
smoking_m	Smoking status of the mother of the subject during pregnancy	categorical		1			0 = no; 1 = yes (smoker during pregnancy)	IF any(ageyears) IS < 20 THEN smoking_m IS not empty			
case_control	If this is a case-control study, indicate whether the subject	categorical		1			0 = control; 1 = case				To fill only if the study is a case-control study

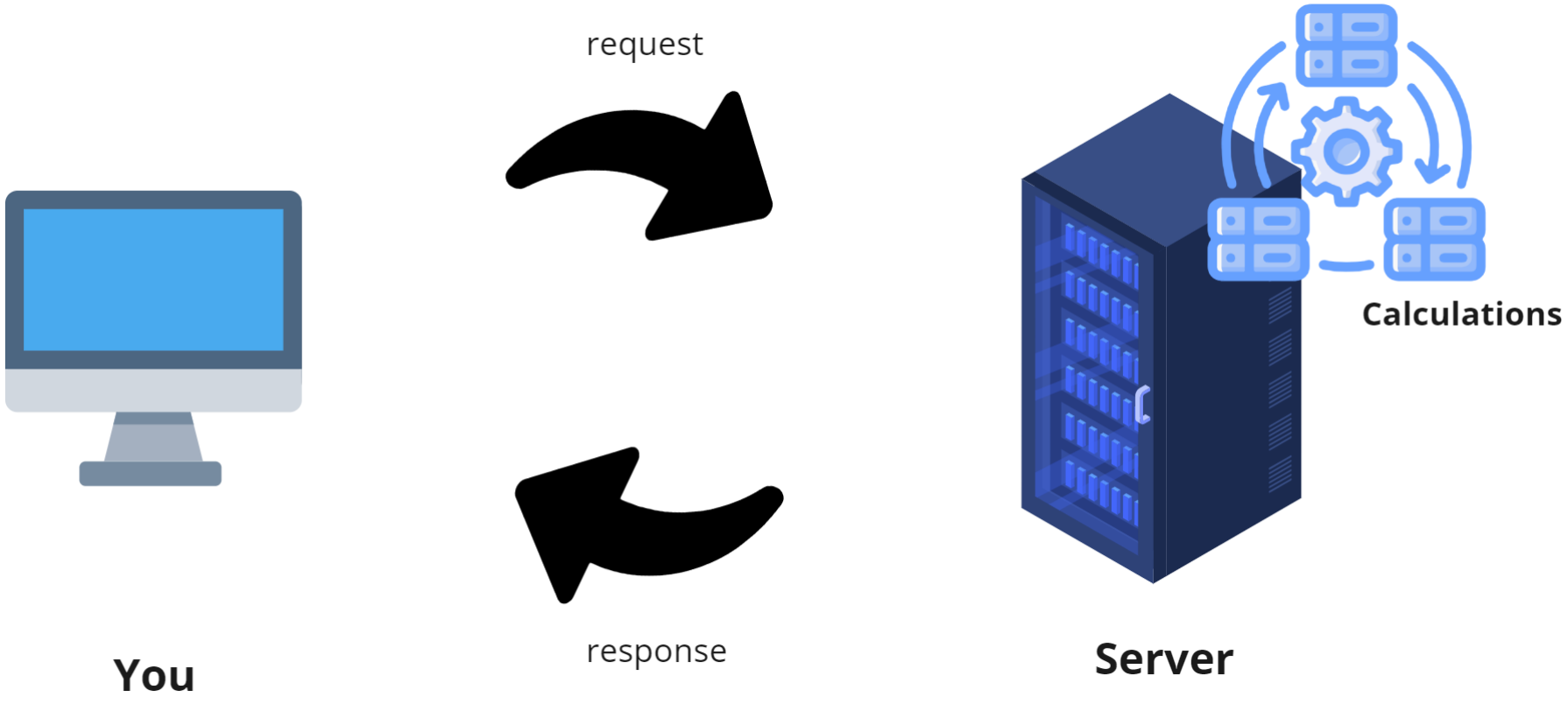


	A	B	C	D	E	F	G
1	id_subject	id_participant	relation	sex	age_birth_m	smoking_m	case_control
2	1	1	P	M	21		
3	2	2	P	F	31	1	
4	3	3	P	M	35	0	
5	4	4	P	F	30	0	
6	5	5	P	F	40		
7	6	6	P	M	27	0	
8	7	7	P	F	24	0	
9	8	8	P	M	32		
10	9	9	P	M	24		
11	10	10	P	M	28	1	
12	11	11	P	F	38		
13	12	12	P	F	21	1	
14	13	13	P	F	28	0	
15	14	14	P	M	15	0	

Ensuring data integrity

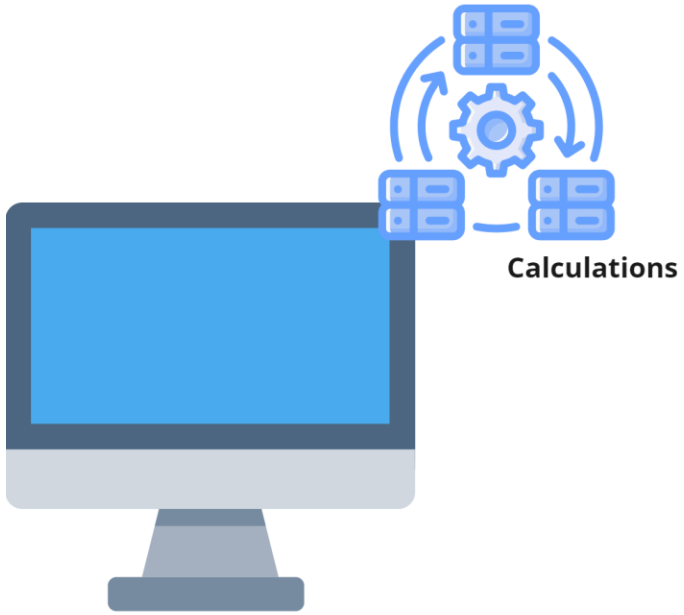
- Python package to
- Pyodide and WebAssembly
  - Ensure data confidentiality and availability

Ensuring data integrity



Computer Icon / Just Icon / CC BY 3.0
Arrow Icon / Artyom Khamitov / CC BY 3.0
Gear icon / aficons studio

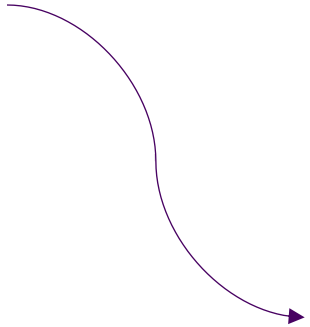
Ensuring data integrity



Computer Icon / Just Icon / CC BY 3.0
Gear icon / aficons studio

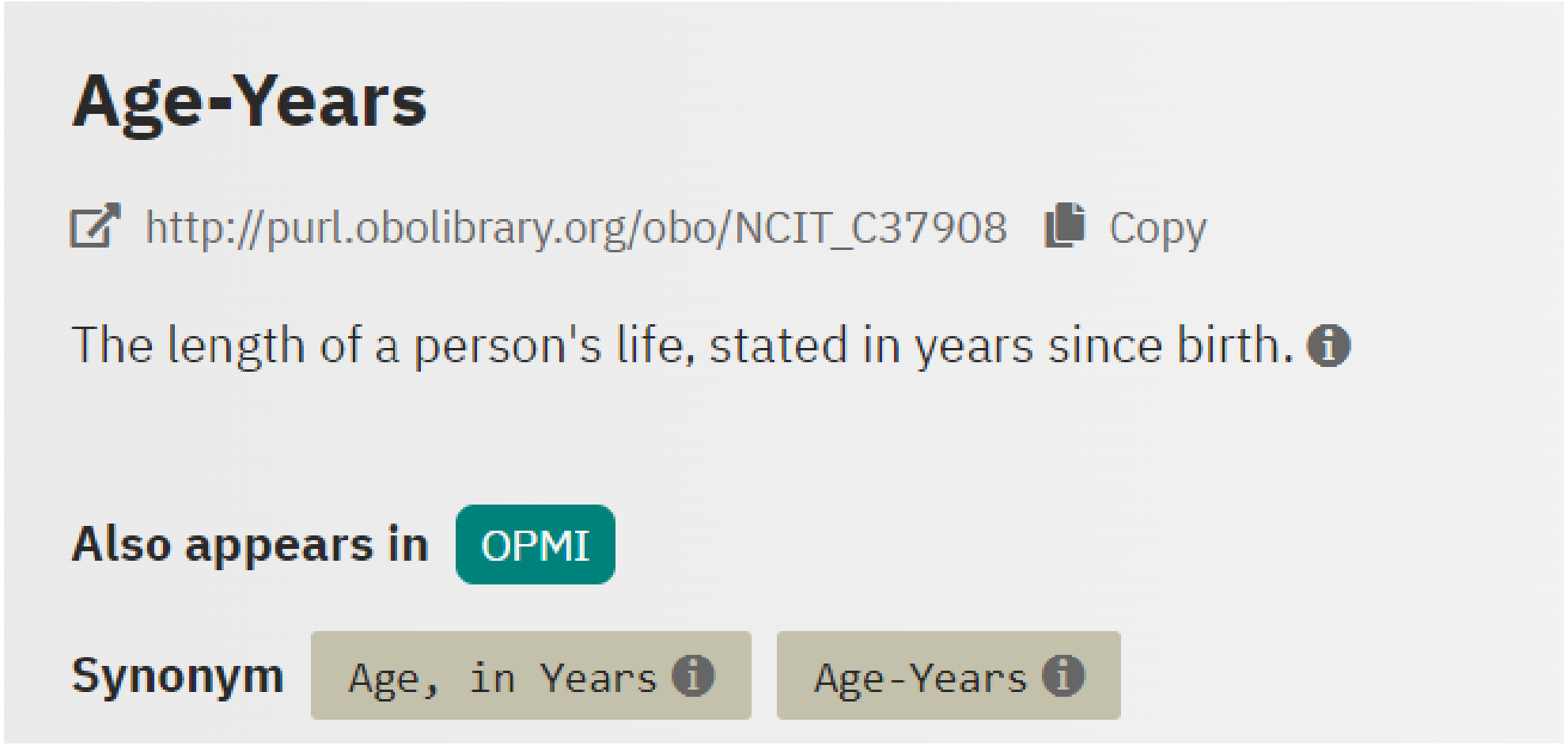
FAIRification through harmonization

12	degurba	Degree of urbanization of residence at sampling of the subject.
13	ageyears	age in years of the subject at sampling
14	agemonths	age in months of the subject at sampling
15	ageweeks	age in weeks of the subject at sampling
16	agedays	age in days of the subject at sampling



12	hbm:degurba	Degree of urbanization of residence at sampling of the subject.
13	hbm:ageyears	age in years of the subject at sampling
14	hbm:agemonths	age in months of the subject at sampling
15	hbm:ageweeks	age in weeks of the subject at sampling
16	hbm:agedays	age in days of the subject at sampling

FAIRification through harmonization



http://purl.obolibrary.org/obo/NCIT_C37908

Summary

- **Available** for stakeholders (https://hbm.vito.be/tools).
- Harmonized data storage in the Personal Exposure and Health (**PEH**) data platform.
- **Standards** defined and relevant HBM data validated.
- Made compatible with the Monte Carlo Risk Assessment (**MCRA**) platform to allow the use of HBM data in the context of risk assessment.
- Calculate **derived variables** (e.g., imputed censored data, standardization and normalization,)
- Conduct **summary statistics** (e.g., geometric mean, percentiles) for e.g. visualization in the European HBM dashboard (TBI).
- Towards **FAIR HBM Data**!

Acknowledgements

Authors:
Ruben Peeters *,
Fen Zhang,
Hanny Willems,
Laura Rodriguez Martin,
Liese Gilles,
Jan Theunis,
Jos Bessems ,
Caio Mescouto Terra de Souza,
Dirk Devriendt,
Eva Govarts

Funding:

This work was carried out in the framework of the European Partnership for the Assessment of Risks from Chemicals (PARC) and has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101057014, as well as the Research Foundation – Flanders (FWO) for International Research Infrastructure EIRENE I001923N. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.