



PHIRI

Population Health Information
Research Infrastructure

Metadata template for Health Information (HI) sources

Deliverable 4.1, 2021-02-09

Petronille Bogaert

Nienke Schutte

Miriam Saso

Table of Contents

Executive summary.....	2
Key points.....	2
I. Introduction.....	3
II. Metadata template	4
III. Link with other metadata templates to be used in the Health Information Portal.....	6
IV. Conclusions and recommendations.....	9
References	9
Appendices.....	10

Executive summary

PHIRI implements a one-stop shop to facilitate access to and (re-)use of European (meta-)data for research on the health of populations and on health care systems. Building on the work in the Joint Action for Health Information (InfAct, www.inf-act.eu), metadata templates are presented to organise and describe (national or regional) Health Information (HI) sources in Europe. This metadata template will be implemented in the Health information Portal (www.healthinformationportal.eu).

Key points

- The Health Information Portal implemented in PHIRI offers a FAIR¹ catalogue with overviews of (COVID-19) population health information sources, to facilitate identification, access to and use of high quality and standardized (meta-)data
- The metadata template presented here contains 30 elements to describe HI sources of routinely collected data
- By means of the InfAct network of National Nodes, the country visits in PHIRI WP3 and web searches, these metadata templates can be used to implement the one-stop shop in the Health Information Portal
- In next steps, metadata catalogues could be linked with metadata templates governed by international organisations, to come to a more comprehensive cumulative catalogue of data resources to research or policy derived queries

PHIRI: WP4, Metadata template for Health Information (HI) sources

Created by Sciensano (Petronille Bogaert, Nienke Schutte, Miriam Saso), DGS/UNL, UZIS, ISCIII/CNE, CDPC, NIJZ, ISS, SU, OKFO, NIHD

Reviewed by RIVM, THL

Approved by PHIRI Steering Committee

I. Introduction

The Population Health Information Research Infrastructure (PHIRI) is a research infrastructure on population health information, implemented to facilitate and generate the best available evidence for research on health and well-being of populations as impacted by COVID-19. PHIRI will allow for better coordinated European efforts across national and European stakeholders to generate the best COVID-19 population health knowledge.

Work package (WP) 4 of PHIRI aims to implement the Health Information (HI) portal on health and health care data for COVID-19 to support structured information exchange between EU countries. The portal offers a FAIR catalogue with overviews of COVID-19 population health data sources, studies and capacity building exercises in EU countries in the framework of PHIRI and beyond. This FAIR catalogue is building on the work in the Joint Action for Health Information (InfAct, www.inf-act.eu, Milestone 26: European Health Information Web Based Platform).

In InfAct, a metadata template was developed to map HI sources and actors (see Table below). Within PHIRI WP4, we extend these elements to come to an updated version to describe HI sources of routinely collected data.

Metadata template for HI sources and actors as developed in InfAct

Nr	Element	Proposed taxonomy/ ontology/ control vocabulary/ defined list of terms
1	Name of HI source	Free text
2	Type of HI source	Defined list of terms + option to add new source
3	Institution	Free text
4	Topic covered	Defined list of terms, based on conceptual framework
5	Website	url
6	Access information	Free text
7	Key words	Free text

This metadata template presented in this report is developed within PHIRI Task 4.1 and aims to facilitate identification, access to and use of high quality and standardized COVID-19 relevant population health (health status and determinants of health) and health care (meta-)data available to the public health bodies represented in the PHIRI consortium.

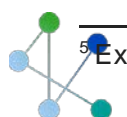
In order to harmonize metadata, free text fields are avoided as much as possible, and the template controls the information by limiting it to a set of predefined terms or categories. The information is (ideally) provided by the data owner and will be screened by the administrator of the Health Information portal before it is published. Finally, this metadata template will be accompanied by a manual for the data owners, to guide them through the template and will provide examples. The

technical set-up and ICT support for the Health Information portal for COVID-19 is foreseen in PHIRI Task 2.3.

II. Metadata template

Nr.	Element	Description	Proposed taxonomy/ ontology/ control vocabulary/ defined list of terms	Remarks
1	Last update	Date when this information was last updated	ISO 8601 ²	
2	Title	Name of the HI source	Free text	
3	Alternative title	Alternative/parallel name of the HI source	Free text	
4	Abbreviation	Abbreviation of the HI source	Free text	
5	URL HI source	URL of the HI source description	URL	
6	Description	Description: e.g. the purpose, nature, and scope of the data collection, special characteristics, major subject areas and topics covered	Free text	The length may vary from study to study.
7	Keywords	Keywords for HI source contents	Free text	Element nr 14 provides a fixed set of topics, based on the conceptual framework (Annex 1). Keywords can be freely chosen.
8	Country	Country in which HI source is registered	ISO-3166 ³	
9	Geo coverage	Geographical coverage of the data	NUTS 1, 2, 3 ⁴ , multinational	
10	Target population	Brief description of the covered population	General population, specific groups (i.e., young, elderly, workers, migrants/minorities), patients, hospitalized, deaths	Multiple answers possible (tick boxes)
11	Age range	Age range of the target population (starting age and max age)	Number	Will also require alternative 'no age limit' (tick box)
12	Sex	Which sexes are covered by the data	Men only/Women only/both/unknown	
13	Type	Type of HI source	Survey/interview data, Administrative data, Population data, Registry data, Biobank/sample/speci	On the Health Information Portal, different examples can be shown for

			men data, Customer record data, Observational study data, Data from other records ⁵	each data type (see footnote)
14	Topic	Topic of HI source	Defined list of terms: conceptual framework	See Annex 1 for conceptual framework
15	Aggregated data	Availability of the data in aggregated level	Yes/no	
16	Individual level data	Availability of the data in individual level	Yes/no	
17	Data collection period	Years in which data was collected (starting year and ending year)	ISO 8601	Will also require an option: 'ongoing' (tick box)
18	Updating periodicity	How frequently data collection is updated	Weekly, Monthly, Annually, Biannually, Every 2+ years, Every 5+ years, Irregularly, one time collection	
19	Data owner	Name of the HI source data owner. Full name in national language	Defined list of terms: build on existing list	Will also require the option to add new data owner
20	Data owner English	Name of the HI source data owner in English	Free text	
21	Data owner acronym	Acronym of the HI source data owner	Free text	
22	URL data owner	General URL of the HI source data owner	URL	
23	Contact	Contact person for HI source	Free text	
24	Contact info	Contact details for contact person HI source	Email address Phone number	Phone number is optional
25	Terms of data access	Describe who can get access to the data, how the access rights are provided, and where and how the access rights should be requested	Free text	
26	Regulations for data sharing	Description how GDPR/national binding regulations guide/possibly limit data sharing	Free text	
27	Linkage possible	Is it possible to link this data source with other HI sources in the country	Yes/Only to some/No	
28	Personal identifier	Personal identifier used in the dataset	National identifier, patient identifier, participant number, other	Identifier in the dataset can be based on national identification or



PHIRI

Survey/Interview data:
Administrative data:

- Population data:
- Registry data:
- Biobank/sample/specimen data:
- Customer record data:
- Observational study data:

health surveys, interview surveys, online panels
(electronic) health care records, hospital discharge data, reimbursement data
mortality, population structure
patient registries for cancer, diabetes, HIV
'omics'-data, saliva, urine, blood
mobility data, food consumption
cohorts, examination surveys

				registry number, or based on patient number (in hospital records), etc.
29	Dataset description	Description of variables included to the dataset or link to codebook which provides this information	URL	If URL is unavailable, a free text field can be provided to fill out a description of the variables.
30	Language of data source	Language of the HI source (language of the variables)	ISO 639-1 ⁶	

III. Example of one filled data entry

To show the relevance of the metadata template presented in part II, here the cause of death registry in Belgium was described as routine data source.

Nr.	Element	Output
1	Last update	2021-02-22
2	Title	Mortality - Cause of death statistics
3	Alternative title	
4	Abbreviation	CoD
5	URL HI source	https://statbel.fgov.be/en
6	Description	Mortality and population data are collected by the National Register. The causes of death statistics are compiled using the statistical death forms (Model IIIC and Model IIID) which are filled in by a certified doctor for each death occurring in Belgium and then completed by the local authority of the place of death. These forms are forwarded to the Communities, which review, code and input the information in order to compile their own statistics. The databases are then sent to Statbel, which aggregates them to produce statistics at the federal level. To this end, Statbel links the forms with the deaths registered in the national register. This allows for checking and completing the available information; to excluding the deaths in Belgium of people who are not legally resident in the country; and finally including in the statistics the deaths abroad of people legally resident in Belgium for whom no civil status form has been filled out.
7	Keywords	Keywords for HI source contents
8	Country	Belgium
9	Geo coverage	NUTS 1
10	Target population	General population Deaths
11	Age range	No age limit
12	Sex	Both
13	Type	Population data
14	Topic	Health status – Mortality – Age-and cause-specific mortality – All causes Health status – Mortality – Age-and cause-specific mortality – infectious diseases – COVID-19
15	Aggregated data	Yes

16	Individual level data	Yes
17	Data collection period	2010-01-01 ongoing
18	Updating periodicity	Yearly
19	Data owner	Statistics Belgium
20	Data owner English	Statistics Belgium
21	Data owner acronym	StatBel
22	URL data owner	https://statbel.fgov.be/en
23	Contact	Statistics Belgium
24	Contact info	statbel@economie.fgov.be
25	Terms of data access	<p>Aggregated data are available here: https://statbel.fgov.be/en/open-data</p> <p>Pseudonymized study data are available for scientific or statistical research within a public institution or research institute:</p> <ul style="list-style-type: none"> • Contact Statbel via statbel@economie.fgov.be or contact directly the statistician in charge of your study field, if you have his contact details. • After consulting the statistician, you should submit an official data request with the request form and a conformity declaration. • You should send the signed documents, together with a pre-contract, by e-mail to statbel.datarequests@economie.fgov.be • Statbel's Data Protection Officer team will evaluate the request and issue an opinion. Finally, as data controller, Statbel's director-general will decide if the data can be delivered. • If the evaluation is positive, Statbel draws up the final contract, that you need to sign. <p>After receipt of the original signed contract, Statbel will deliver the requested microdata within the set timeframe.</p>
26	Regulations for data sharing	Please contact statbel.datarequests@economie.fgov.be for more information on data sharing regulations
27	Linkage possible	Yes
28	Personal identifier	National identifier
29	Dataset description	https://statbel.fgov.be/en/open-data
30	Language of data source	EN

IV. Link with other metadata templates to be used in the Health Information Portal

The other metadata templates that will be provided in the PHIRI-framework will focus on the collection of information on National Nodes, Research Networks and projects (national and international), Population Health Studies, Training Materials, Publications and reports (scientific and policy) and ELSI (Ethical, Legal and Social Implications). These are all part of PHIRI WP4 and WP8. Several items in these metadata catalogues link with the metadata template for HI sources (presented in part III):

Nr.	Element	Related metadata template
1	Last update	National Nodes, Research networks and projects, Population Health Studies, Training materials, Publications and reports
2	Title	National Nodes, Research networks and projects, Population Health Studies, Training materials, Publications and reports
3	Alternative title	National Nodes, Research networks and projects, Population Health Studies
4	Abbreviation	National Nodes, Research networks and projects, Population Health Studies
5	URL HI source	National Nodes, Research networks and projects, Population Health Studies, Training materials, Publications and reports
6	Description	National Nodes, Research networks and projects, Population Health Studies, Training materials, Publications and reports
7	Keywords	National Nodes, Research networks and projects, Population Health Studies, Training materials, Publications and reports
8	Country	National Nodes, Research networks and projects, Population Health Studies, Training materials, Publications and reports
9	Geo coverage	National Nodes, Research networks and projects, Population Health Studies, Publications and reports
10	Target population	National Nodes, Research networks and projects, Population Health Studies, Publications and reports
11	Age range	Population Health Studies
12	Sex	Population Health Studies
13	Type	Population Health Studies
14	Topic	National Nodes, Research networks and projects, Population Health Studies, Training materials, Publications and reports
15	Aggregated data	Research networks and projects, Population Health Studies
16	Individual level data	Research networks and projects, Population Health Studies
17	Data collection period	Research networks and projects, Population Health Studies
18	Updating periodicity	Research networks and projects, Population Health Studies
19	Data owner	Research networks and projects, Population Health Studies, Training materials, Publications and reports
20	Data owner English	Research networks and projects, Population Health Studies, Training materials, Publications and reports
21	Data owner acronym	Research networks and projects, Population Health Studies, Training materials, Publications and reports
22	URL data owner	Research networks and projects, Population Health Studies, Training materials, Publications and reports
23	Contact	National Nodes, Research networks and projects, Population Health Studies, Training materials, Publications and reports
24	Contact info	National Nodes, Research networks and projects, Population Health Studies, Training materials, Publications and reports
25	Terms of data access	Research networks and projects, Population Health Studies, Training materials
26	Regulations for data sharing	Research networks and projects, Population Health Studies
27	Linkage possible	Research networks and projects, Population Health Studies
28	Personal identifier	Research networks and projects, Population Health Studies
29	Dataset description	Research networks and projects, Population Health Studies
30	Language of data source	Research networks and projects, Population Health Studies, Training materials, Publications and reports

V. Conclusions and recommendations

Metadata templates facilitate the identification, access to and use of high quality and standardized COVID-19 relevant data from routine data sources. The national COVID-19 data portals can be searched through a web search to pre-fill these metadata templates and in preparation for the country visits organised in PHIRI Task 3.1. Through the national nodes and country visits, EU countries will be supported to complete these FAIR metadata catalogues and provide inputs on optimisation of the portal.

It is recommended for the PHIRI consortium to explore how metadata catalogues on data sets are compatible with metadata templates developed by international organisations such as the ECDC⁷ that curate (meta-)data themselves. Linking these metadata templates could offer a more comprehensive cumulative catalogue of data resources to research or policy derived queries.

Furthermore, if there is an overlap in template elements for HI sources and other metadata templates (for example for Population Health Studies - PHIRI Task 4.3), it would be advised to merge the two templates to improve searchability on the Health Information Portal.

References

MS26: European Health Information Web Based Platform. WP7 partners JA for Health Information – InfAct, August 2019

Appendices

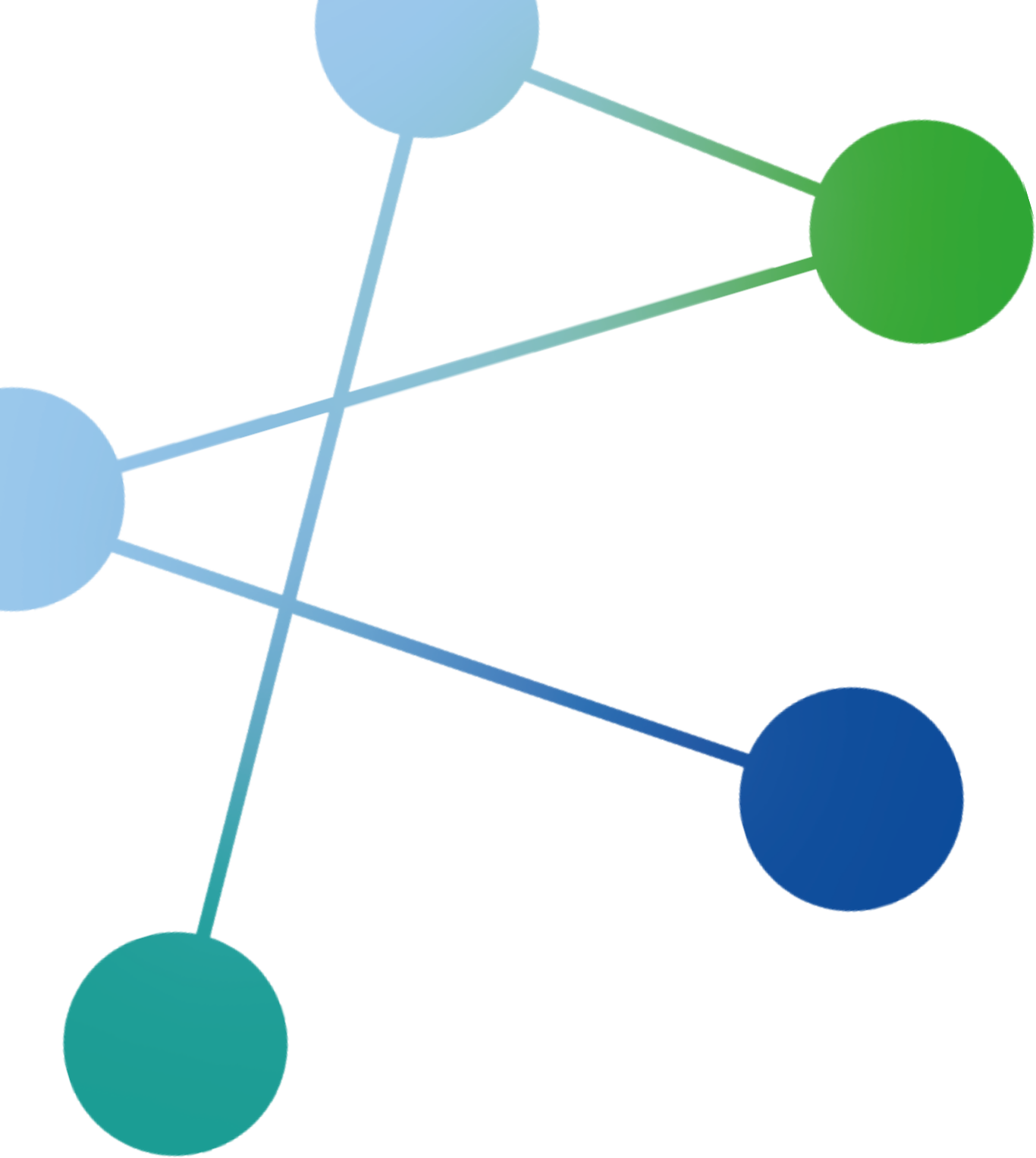
Annex 1: Conceptual Framework (including COVID-19)

This conceptual framework is based on the work in InfAct WP7 and updated to include topics on COVID-19.

Level 1	Level 2	Level 3	Level 4	Level 5
Determinants of health	Individual characteristics and behaviors	Behaviors	Alcohol consumption	
			Food consumption (or proxy)	
			Physical activity	
			Reproductive and maternal	
			Tobacco use	
			Use of psychoactive substances	
			Isolation and quarantine	
			Attitude towards vaccination	
	Physical and social environment	Physical environment	Birth weight	
			Blood pressure	
		Social environment	Cholesterol	
			Overweight/obesity	
	Socioeconomic and demographic factors	Demographic factors	Housing and sanitation	
			Pollution	
		Socioeconomic factors	Social network	
			Work-related	
			NA (Demographic factors)	
			Education	
Health status	Morbidity/disability	Accidents & injuries	Home/leisure	
			Self-injury	
			Traffic	
			Work	
		Communicable diseases	Airborne and/or vaccine-preventable diseases	
			Airborne and/or vaccine-preventable diseases	COVID-19
			Food and water borne	
			Sexually transmissible and/or blood borne	
			Zoonotic	
		Disability	NA (Disability)	
	Non-communicable diseases	Cancer		
		Cardiovascular diseases		
		Dental diseases		
		Diabetes		
		Mental diseases		
		Reproductive, maternal and newborn health		
		Respiratory diseases		
	Self-reported health status	Self-perceived health/morbidity		
Mortality	Age- and cause-specific mortality	All causes		
		Cancer		
		External causes		

			Infectious diseases				
			Infectious diseases	COVID-19			
			Non-communicable diseases (excluding cancer)				
			Life expectancies	Health expectancy			
				Life expectancy			
				Reduction of life expectancy (PYLL)			
			Maternal, perinatal and newborn mortality	Infant mortality			
				Maternal mortality			
			Wellbeing	NA (Wellbeing)	NA (Wellbeing)		
			Health systems	Health resources and activities	Care utilization	Consultations	
						Diagnostic exams and tests	
Diagnostic exams and tests	Nucleic acid or antigen test						
Diagnostic exams and tests	Seroprevalence						
Hospital utilization							
Hospital utilization	ICU						
Long-term care							
Reproductive, maternal and newborn health							
Surgical procedures							
Health employment and education	Education						
	Health workforce migration						
	Nurses and/or midwives						
	Physicians						
	Remuneration						
Pharmaceutical sales & consumption	Generic market						
	Pharmaceutical consumption						
	Pharmaceutical consumption	Post-marketing surveillance					
	Pharmaceutical sales						
Physical and technical resources	Hospitals and beds						
	Hospitals and beds	ICU					
	Medical technology						
	Other care units/beds						
Health system performance	Access	(Un)met needs or their causes					
		Health care coverage					
		Delivery and supply chain					
	Costs/expenditure	Assets					
		Financing scheme					
		Function					
		Provider					
		Provision factors					
		Revenues					
	Quality	Autopsy					
		Cancer screening					
		Cancer survival rates					
		Care					
		Patient experience					
		Patient safety					
Premature/avoidable mortality							

			Reproductive, maternal and newborn health	
			Vaccination coverage	
Policy	Response plan			
	Adherence/ compliance to measure			
	Interventions	Face Mask		
		Physical distancing		
		Mass gathering		
Contact tracing				
NA (Policy)	NA (Policy)	NA (Policy)		
Communication	Misinformation and disinformation			
Unclassified	NA (Unclassified)	NA (Unclassified)	NA (Unclassified)	



Sciensano

Rue Juliette Wytsmanstraat 14
1050 Brussels, Belgium
Phiri.coordination@sciensano.be
www.phiri.eu

 @PHIRI4EU