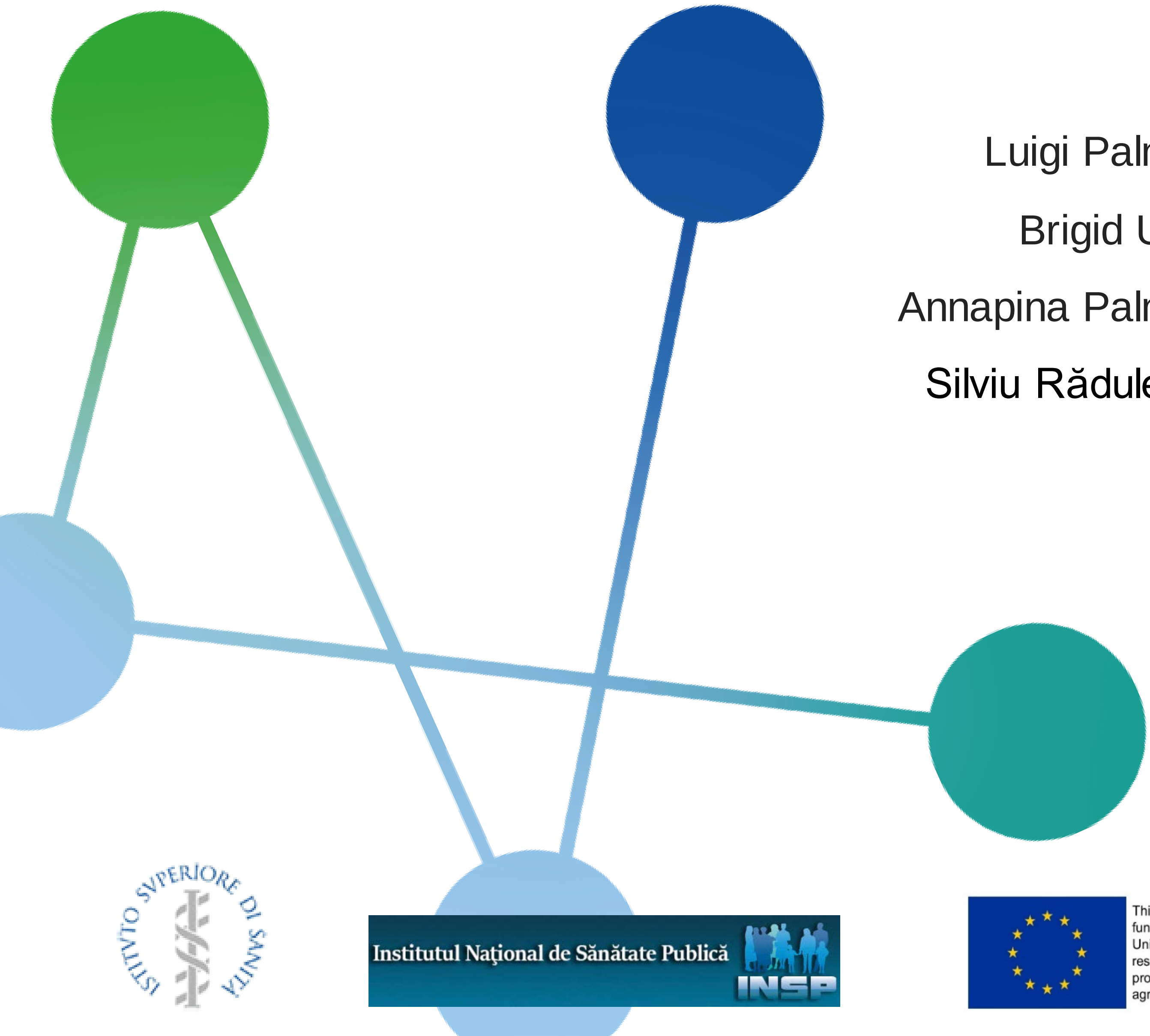
PHIRI
Population Health Information
Research Infrastructure

Metadata template for COVID-19 population health studies

Deliverable 4.2, 2021-02-25



Luigi Palmieri

Brigid Unim

Annapina Palmieri

Silviu Rădulescu

Istituto Superiore di Sanità

Institutul Naţional de Sănătate Publică
INSP

This project has received
funding from the Europea
Union's Horizon 2020
research and innovation
programme under grant
agreement No 10101831

Table of Contents

Executive summary

This paper is the first in a series supporting the achievement of the aim of task 4.3 (T4.3) of Work Package 4 of PHIRI, namely, to prepare a catalogue of population health studies on the impact of COVID-19 in Member States. This paper proposes a metadata template for T4.3 taking also inspiration from the metadata guidelines of the Consortium of European Social Science Data Archives (CESSDA [2]). This template will be the foundation for setting up a catalogue containing metadata information on data sources, catalogue that will be integrated with other PHIRI deliverables in the PHIRI Portal.

The template allows the collection of information on population health studies on the impact of COVID-19 in Member States using a systematic and comprehensive approach, which will facilitate analyses attempting meaningful comparisons of studies with similar methodologies and pooling of data from multiple sources.

The elements of the “COVID-19 population health studies metadata template” are linked to the other metadata templates that will be elaborated and provided in the framework of the PHIRI project focusing on the collection of information on National Nodes, Research Networks, FAIR (Findable, Accessible, Interoperable, Re-usable) data sources (focusing on routine data collections), Training materials, Guidelines, Publications, and ELSI (Ethical, Legal and Social Implications)

The reporting standard established by the catalogue using a common metadata template to describe the large variety of population health data sources retrieved through the PHIRI research activities will facilitate the effective use of health information for policy-making and research in and across Member States through systematic comparisons of similar national or local data collection and analyses.

Key points

- COVID-19 population health studies
- metadata template developed
- use of systematic and comprehensive health information catalogues as a foundation for analyses supporting policy-making and research development

PHIRI: WP4, Metadata template for COVID-19 population health studies

List of contributors to the report:

Luigi Palmieri, Brigid Unim, Annapina Palmieri, Silviu Rădulescu, Hanna Tolonen, Mari Makinen, Mariken Tijhuis, Jane Idavain, Asuncion Diaz, Richard Pentz, on behalf of all PHIRI WP4 partners.

I. Introduction

The aim of task 4.3 (T4.3) is to prepare a catalogue of population health studies on the impact of COVID-19 in Member States; we define population health as “the health outcomes of a group of individuals, including the distribution of such outcomes within the group,” and we argue that the field of population health includes health outcomes, patterns of health determinants, and policies and interventions that link these two [1]; epidemiological studies on Health Monitoring and Health System Performance Assessment; clinical trials and surveillance systems will not be considered. The catalogue can include information on national population health studies, but also about multinational collaborative studies. The studies will be identified through a web-based search of EU public health institutions and research networks. Additional information will be obtained from project partners through country visits organized in WP3, to include studies not available on the web, e.g., those published only locally or in national languages. A catalogue containing metadata information (see section II. Metadata template) will be provided for the PHIRI Portal. The metadata template for T4.3 was developed according to the metadata guidelines of the Consortium of European Social Science Data Archives (CESSDA) [2], that classifies metadata components in 4 groups: a) Descriptive (provides information on the intellectual content of a data collection); b) Administrative (includes information that helps archives and repositories ingest and manage data for preservation); c) Structural (concerns physical or logical links between objects or between parts of a complex object); and d) Technical metadata (gives information about file formats, file size, encoding, or storage).

II. Metadata template

Nr.	Element	Description	Proposed taxonomy/ ontology/ control vocabulary/ defined list of terms
1	Last update	Date when this information was last updated	ISO 8601[1]
2	Title	Name of the study	Free text
3	Abbreviation	Abbreviation of the study	Free text
4	URL of the study	URL of the study web site	url
5	Description	Description should describe the purpose, nature, and scope of the data collection, special characteristics of its contents, major subject areas covered, etc. The length may vary from study to study.	Free text

---

[1] https://www.iso.org/iso-8601-date-and-time-format.html

6	Governance and legal framework	Brief description of the steering committee, scientific committee, study coordination (if available)	Free text
7	Funding	Brief description of funding sources (public/private)	Free text
8	Keywords	List of the key words identifying the study	Defined list of terms: MeSH, ELSST, etc. lists (with the possibility of adding terms not included in the lists)
9	Research network	If this population health study is part of multinational research network/project, please define here which one.	List of research networks/project and possibility to add new one if not on the list
10	Country	Country where data are collected	ISO-3166[2]
11	Geo-coverage	Geographical coverage of the data	NUTS 1, 2, 3[3], Multinational
12	Target population	Brief description of the studied population	Defined list of terms: general population, specific groups (i.e., young, elderly, workers, migrants/minorities), patients, hospitalized patients, deceased persons (with the possibility of adding clarification as a free text)
13	Age range	Max and Min ages of the studied population	Text (Age-range), 'No age limit'
14	Sex	Sexes distinguished in the studied population	Defined list of terms: M/F/both
15	Sample size	Number of persons included in the sample of the study	Numeric
16	Type	Type of the study	HIS, HES, cohort study, register based studies, administrative health data elaboration (HDR, Mortality), Population health monitoring, HSPA, health policies, interventions, preventive actions, survey, Biobank/sample/specimen data, Customer record data, other[4]
17	Topic	Topic of the study	Defined list of terms: conceptual framework (see Annex 1)
18	Aggregated data	If aggregated data are available	Yes/No
19	Individual level data	If individual data are available	Yes/No

---

[2] https://www.iso.org/iso-3166-country-codes.html

[3] https://ec.europa.eu/eurostat/web/nuts/national-structures

[4] Examples:

- Survey/interview data: health surveys, interview surveys, online panels
- Administrative data: (electronic) health care records, hospital discharge data, reimbursement data
- Population data: mortality, population structure
- Registry data: patient registries for cancer, diabetes, HIV
- Biobank/sample/specimen data: 'omics'-data, saliva, urine, blood
- Customer record data: mobility data, food consumption
- Observational study data: cohorts, examination surveys

20	Data collection period	Starting year/month and ending year/month of data collection or Starting year/month of data collection and 'ongoing' (if data collection is still active)	ISO 8601
21	Date of last update	Date of the last update of data collection	ISO 8601
22	Updating periodicity	Periodicity of new data collection	Weekly, Monthly, Annually, Biannually, Every 2+ years, Irregularly, one time collection
23	Data owner	Name of the institute which owns the data	Free text
24	Data owner English	Name of the coordinating institute in English (if available)	Free text
25	Acronym of the data owner	(if available)	Defined list of terms/Free text
26	URL of the data owner	URL of the website of the coordinating Institute	url
27	Contact person	Name of the contact person	Free text
28	Contact info	e-mail address	Free text
29	Terms of access	Brief description of rules and procedures needed to have access to data (if available) (i.e., if data are freely accessible as aggregate/individual data, if a formal request is needed for approval, if a financial charge is requested, etc.)	Free text
30	Dataset	Description of the variables included in the dataset or link to the codebook which provides this information	Free text and/or url
31	Language	Language of the data documentation	ISO 639-1[5] Code or name of the language in English
32	Publication	List of main publications of the study (description, protocols, results) *(If possible, they can be drawn from the PHIRI publications/guidelines catalogue using for example the acronym of the study as the keyword for linkage)* or Website link to the list of the project/ study publications (if available)	Free text / url
33	COVID-19	Brief description of the relation between the study objectives to the COVID-19 pandemic	Free text

III. Example of one filled data entry

A collaborative pilot study to monitor and optimize access to urgent surgery during the period of reduced resources in Genoa (Italy) caused by the COVID-19 pandemic [3], was used as an example to fill-in the proposed metadata template with relevant information.

Nr.	**Element**	**Output**
1	Last update	26/05/2020
2	Title	Monitoring the introduction of a model to audit, prioritize and

---

[5] https://www.iso.org/iso-639-language-codes.html

		optimize access to cancer and urgent surgery to all patients in the metropolitan area of Genoa during and after the COVID-19 pandemic.
3	Abbreviation of the study (ID study)	CoV-2-GOASUR-SWALIS2020-FP1
4	URL of the study	https://www.isrctn.com/ISRCTN11384058
5	Description	This study includes a 2-week feasibility and a 4-week pilot implementation of a newly designed, software-aided, inter-hospital, collaborative surgical pathway. Main aims are to monitor and optimize access to urgent surgery during the period of reduced resources in Genoa (Italy) caused by the COVID-19 pandemic, using a bespoke referral process and the SWALIS 2020 model to prioritize surgery by clinical urgency and waiting time.
6	Governance and legal framework	Multicenter study coordinated by Policlinico S. Martino University Research Hospital
7	Funding	Funded by Liguria Health System (public funds)
8	Key words	urgent surgery, COVID-19
9	Research network	
10	Country	Italy, IT
11	Geo-coverage	Italy, Liguria Region
12	Target population	Surgery patients, hospitalized patients
13	Age range	18+
14	Sex	Both sexes
15	Sample size	300
16	Type of study	Interventional
17	Topic	Health systems, Health system performance, Access, Health care coverage (late effects on other diseases)
18	Aggregated data	Yes
19	Individual level data	Yes (Available, but accessible according to reported rules)
20	Data collection period	From March 2020 to May 2020
21	Date of last update	May 2020
22	Updating periodicity	No
23	Data owner	Ospedale Policlinico S. Martino IRCCS, Azienda Liguria Sanità (ALiSa) (Italy)
24	Data owner English	S. Martino Polyclinic, University Research Hospital
25	Acronym of the data owner	AliSa
26	URL of the data owner	https://www.alisa.liguria.it/; https://www.ospedalesanmartino.it/
27	Contact person	Roberto Valente
28	Contact info	email: roberto.valente@hsanmartino.it (Address: Largo Rosanna Benzi 10, Genoa 16122 Italy; Phone +39 (0)105552549;
29	Terms of access	The datasets generated during and/or analysed during the current study are/will be available upon request from Mr Roberto Valente (roberto.valente@hsanmartino.it) and Dr Stefano Di Domenico (Stefano.didomenico@hsanmartino.it). The dataset is in an MS Access TM format, and can be anonymized to any level required, as allowed by the Liguria Ethics Committee. The dataset will be made available 3 months after study completion, for 12 months,

		extendable. Any sharing request will be submitted to the Liguria Ethics Committee for approval.
30	Dataset	https://www.isrctn.com/ISRCTN11384058 Primary outcome measure: Measured weekly, i.e. every Tuesday following the just-finished week: feasibility (weeks 1 and 2), pilot (weeks 3, 4, 5, 6): 1. Clinical complications and adverse events reportedly caused by the pathway (safety), measured by recording and numbering single reports by surgeons responsible for patient care on identified proforma at weekly MDT meetings 2. According to the SWALIS model, the waiting list is measured weekly through purposely designed performance indexes, including cross-sectional (for patients currently on the list at an index day) and retrospective views (for patients who received treatment during a given period T): 2.1. Priority (as average and SD) on the list (new for SWALIS 2020) (for each and all urgency categories) 2.2. Waiting time in days (as median and range) (for each and all urgency categories) 2.3. Waiting list length (as average and SD) (for each and all urgency categories). Secondary outcome measures Measured weekly, i.e. every Tuesday following the just-finished week: feasibility (weeks 1 and 2), pilot (weeks 3, 4, 5, 6): 1. Appropriate Performance Index (API) (for each and all urgency categories): 1.1. Retrospective: percentage of patients receiving treatment within their respective MTBT in a given period T 1.2. Cross-sectional: percentage of patients currently on the list at an index day having waiting time less than their respective MTBT 2. Deviation events (total and percentage). These include updates in urgency during wait, number of postponements (prior to the day of admission) or cancellations (on the day)
31	Language	English
32	Publications	1. The study protocol and the statistical analysis plan will be available at publication; 2. International journals and meetings, study website (in progress); 3. https://www.isrctn.com/ISRCTN11384058
33	COVID-19 related material	The stydy aims at prioritising, monitoring, and optimising access to surgery, balancing the needs of COVID-19 against cancer and urgent patients during the pandemic.

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

The list of metadata proposed for the collection of information and descriptions of population health studies/projects related to the SARS-CoV-2 pandemic (focused on population based studies such as HES, HIS, cohorts etc.) and conducted in European countries, present some synergies with the lists of metadata proposed for other metadata templates regarding other topics related to COVID-19.

The other metadata templates that will be elaborated and provided in the framework of the PHIRI project focus on the collection of information on National Nodes, Research Networks, FAIR (Findable, Accessible, Interoperable, Re-usable) data sources (focused on routine data collections), Training materials, Guidelines, Publications, and ELSI (Ethical, Legal and Social Implications)

Here following, specific elements or group of elements of Population health studies/projects description that link with the other mentioned metadata templates.

- Descriptive metadata (e.g., Country/ list of participating countries, Geo-coverage, Coordinating institute/data owner, URL of the coordinating institute, Acronym of the coordinating institute, Contact person for the study, Contact details for contact person) are basically common to quite all the other metadata templates (details in the table below);
- URL of the study can be linked to the National Node and Research Network metadata template
- Governance and legal framework can be linked to the National Node metadata template;
- Description (target population, objectives, etc.), if organised a summary, can be linked to the same metadata in Training materials, Guidelines, Publications, ELSI;
- Name of the study can be linked to the National Node, Research Network, FAIR catalogue, Publications metadata templates;
- Abbreviation of the study can be linked to the Research Network metadata template;
- Topic of the study can be linked to the similar metadata in FAIR catalogue, Training materials, Guidelines, Publications, ELSI metadata templates;
- Language can be linked to the same metadata in FAIR catalogue, Training materials, Guidelines, Publications, ELSI metadata templates;
- Access rights/procedures can be linked to the same metadata in Research Network and FAIR catalogue metadata templates;
- Data collection period can be linked to the same metadata in FAIR catalogue and Publications metadata templates;
- Publication date can be linked to the same metadata in FAIR catalogue, Training materials, Guidelines, Publications metadata templates;
- Date of last update and Updating periodicity elements can be linked to the same elements in FAIR catalogue metadata template;
- COVID-19 related material element can be linked to National Node, Research Network, Training materials, Guidelines, Publications
- Key words for study element can be linked to the same elements in all the other metadata templates;
- Aggregated data and Individual level data elements can be linked to the same elements in Research Network and FAIR catalogue metadata templates;

- URL of the study data description (if available) element Research Network, FAIR catalogue metadata templates;
- Main publication reference or URL of the study publications references element can be linked to the Unique publication identifier (doi) and to the Title of the publication elements in the Publications metadata template.

Table 1 - Summary table reporting possible links between the elements of ‘COVID-19 population health studies/projects’ metadata template and the other metadata templates

Element of ‘COVID-19 population health studies’ metadata template	**Links with other metadata templates**
Last update	All metadata templates
Country	National Node, Research Network, FAIR catalogue, Training materials, Publications
Data owner	All metadata templates
Geo-coverage	Research Network, FAIR catalogue, Guidelines, Publications, ELSI
Contact	All metadata templates
URL of the data owner	National Node, Research Network
Governance and legal framework	National Node
Description	Training materials, Guidelines, Publications, ELSI
Title	National Node, Research Network, FAIR catalogue, Publications
Acronym of the data owner	Research Network
Type	National Node, FAIR catalogue
Topic	FAIR catalogue, Training materials, Guidelines, Publications, ELSI
Language	FAIR catalogue, Training materials, Guidelines, Publications, ELSI
Terms of access	Research Network, FAIR catalogue
Data collection period	FAIR catalogue, Publications
Date of last update	FAIR catalogue
Updating periodicity	FAIR catalogue
COVID-19	National Node, Research Network, Training materials, Guidelines, Publications
Keywords	All metadata templates
Aggregated data	Research Network, FAIR catalogue
Individual level data	Research Network, FAIR catalogue
Target population	-
Sex	-
Age range	-
Sample size	-
Dataset	Research Network, FAIR catalogue
Publications	Publications

V. Conclusions and recommendations

The Health Information Portal will adopt a common metadata reporting standard to describe population health data retrieved through the PHIRI research activities. This will facilitate the use of health information for policy-making and research in and across Member States.

References

1. David Kindig D and Stoddart G. What Is Population Health?. Am J Public Health. 2003 March; 93(3): 380–383. doi: 10.2105/ajph.93.3.380;
2. CESSDA: Documentation and Metadata. https://www.cessda.eu/content/download/498/4475/file/CESSDA%20User%20Guide%20for%20digital%20preservation4_Documentation_Metadata.pdfAppendices;
3. https://www.isrctn.com/ISRCTN11384058

Appendices

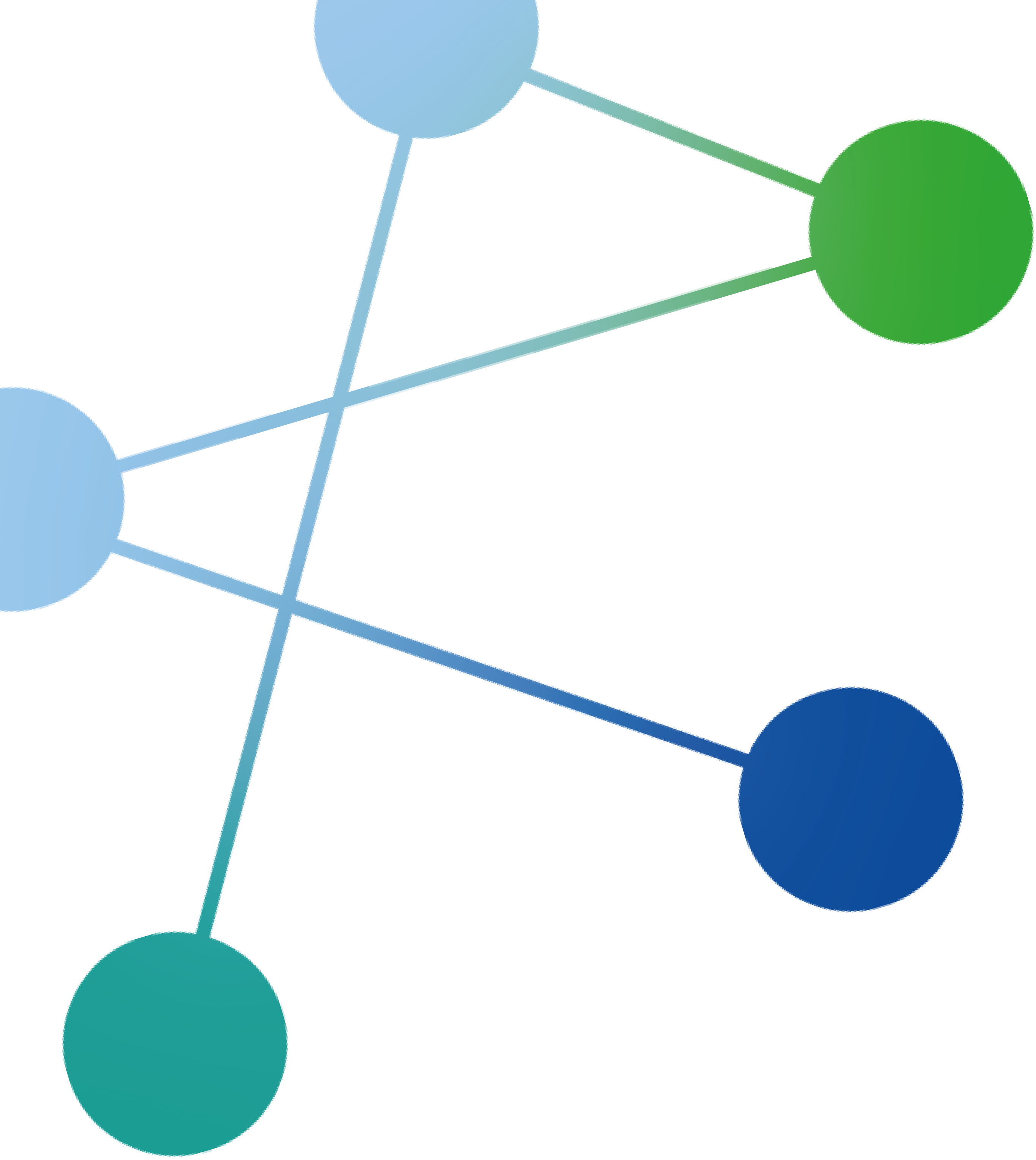
Annex 1: Conceptual Framework (including COVID-19)

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	
		Individual characteristics	Birth weight	
			Blood pressure	
			Cholesterol	
			Overweight/obesity	
	Physical and social environment	Physical environment	Housing and sanitation	
			Pollution	
		Social environment	Social network	
			Work-related	
	Socioeconomic and demographic factors	Demographic factors	NA (Demographic factors)	
		Socioeconomic factors	Education	
			Employment/occupation	
			Income/poverty	
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	
			Workforce other	
		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)	



Istituto Superiore di Sanità-ISS
Via Giano della Bella, 34
00161 – Rome, Italy
luigi.palmieri@iss.it
www.phiri.eu
@PHIRI4EU