Software publication in public repository

Deliverable 6.5, 07.12.2022

Javier González-Galindo

Francisco Estupiñán-Romero

Natalia Martinez Lizaga

Juan González-García

Carlos Tellería-Orriols

Martin Thissen

Enrique Bernal-Delgado

Table of Contents

PHIRI
Population Health Information
Research Infrastructure

1. Executive summary

[Population Health Information Research Infrastructure](#) (PHIRI) aims to facilitate and support open, interconnected, and data-driven research by sharing cross country COVID-19 population health information, and exchanging best practices related to data collection, curation, processing, use, and reuse following Ethical, Legal, and Social (ELSI) Issues and Findability, Accessibility, Interoperability, and Reusability (FAIR) principles. It has the objective to: 1) provide a [Health Information Portal/HIP](#) for COVID-19 with FAIR catalogues on health and health care data, 2) provide structured exchange between countries on COVID-19 best practices and expertise, and 3) promote interoperability and tackle health information inequalities.

The aim of PHIRI Work Package (WP) 6 is to demonstrate how a broad variety of data (e.g., administrative and survey data) can be pooled and/or reused in a distributed way across Europe. WP6 looks at COVID-19 impacts in specific subgroups by conducting research through real life use cases of immediate relevance. Furthermore, these use cases represent pilot activities for the benefits and added value of a research infrastructure by mobilizing data from different European countries.

In conjunction with WP7 which aims to create and validate a federated research infrastructure that overcomes data reuse and data sharing hindrances, WP6 has established the governance mechanism and implemented the technological solutions to respond to four use cases carried out in multiple sites (local nodes at country/region level) [1]:

1. Impact of the COVID-19 on health care in more vulnerable populations (Use case A)
2. COVID-19 related delayed care in breast cancer patients (Use case B)
3. Effects of the COVID-19 pandemic on maternal and newborn health (Use case C)
4. COVID-19 related changes in mental health care (Use case D)

The small-medium scale prototype of the federated research infrastructure has demonstrated the suitability of this federated research approach in the production of accurate and timely research outputs for a rapid policy response on COVID-19.

The data models, analytical pipelines and reporting solutions, a total of 13 digital objects within a technological solution (Docker Image) that is a secure environment for the implementation of the Federated Research Infrastructure (FRI) across the participant nodes, have been published and are persisted in Zenodo, an OpenAIRE initiative, under a creative common international attribution 4.0 license at https://zenodo.org/record/6936063#.Y0IV4HZBzDc. These outputs will be integrated in the HIP developed in WP4.which will include, among other FRI services, example of use case results as an illustration of the methods, in a tutorial format.

---

[1] Use Case A: Use case A: Austria (AT), Croatia (HR), Finland (FI), Aragon (ES), Wales (UK); Use case B: Belgium (BE), Marche (IT), Latvia (LV), Aragon (ES), Wales (UK); Use case C: Euro-Peristat Network (27 countries and 4 UK nations); Use case D: Austria (AT), Croatia (HR), Finland (FI), Romania (RO), Aragon (ES), Wales (UK).

PHIRI
Population Health Information
Research Infrastructure

2. Introduction

In the design of a federated infrastructure that aims at reusing sensitive data distributing scripts to the data holders, WP7 has designed and implemented a workflow that is general to all the use cases in WP6 (figure 1).
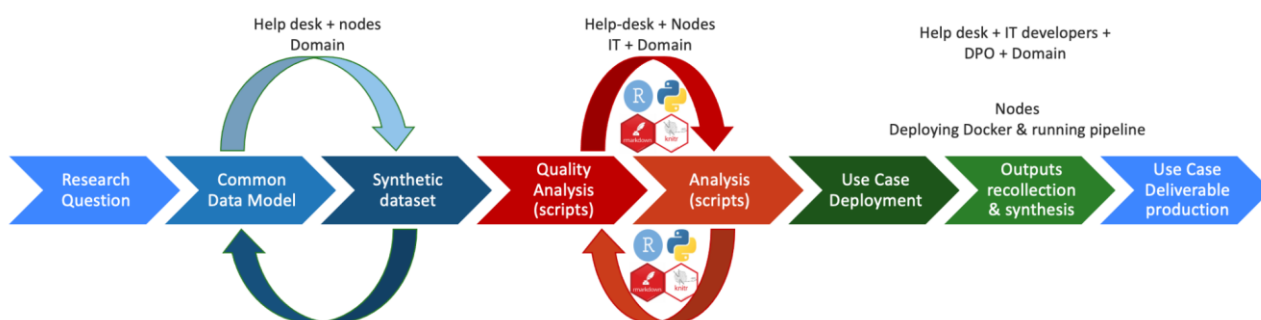
Each use case materializes in one or more research queries that are carried out in an interoperable way by formalizing data models ie. agreeing on a Common Data Model (CDM[2]) per use case that materializes the research query.

A synthetic dataset that represents the variables of each use case is created.

The data quality/data assurance scripts that allows the data quality analysis of each of the participants' original data summarizing information on missing data, anomalous distributions, impossible values or information gaps that helped in the interpretation of the outputs.

Along the workflow, more different digital objects are produced: the analytical scripts specific to the research queries, the container that includes all the digital objects (a Docker image) and the results per node.

All digital objects are published and persisted in Zenodo, an Open Aire initiative under a creative commons international attribution 4.0 license at 10.5281/zenodo.5729310

3. Zenodo

Zenodo is a publishing service infrastructure, provided by the OpenAIRE initiative, supporting open scholarly communication, an Open Science for European research. Zenodo is an open source, open access and open data platform supported by the European Commission that eliminates barriers to adopting data sharing practices without imposing any requirements on format, size, access restrictions or license.

Most valuable features we considered for sharing WP6 research outputs are:

a) lack of restrictions in digital object formats and sizes,
b) capacity to index more than ten different research outputs as digital objects (i.e. publications, datasets, software, workflow, and other) across any field,
c) facilitation of discoverability by providing advanced search features and assigning to all published research outputs a Digital Object Identifier (DOI) to make the research outputs uniquely citable and univocally identifiable,
d) enabling researchers to create specific or thematic collections and communities enhancing dissemination and reusability,

---

[2] A Common Data Model (CDM) is a way of organizing data into a standard structure

e) encouraging scientist to share their research as openly as possible by providing several Open Source licensing options, under a variety of different licenses and access levels,

f) providing support for semantic versioning and continuous update of the research outputs published,

g) providing secure object persistence through safe storage in same cloud infrastructure as research data from CERN's Large Hadron Collider and using CERN's battle-tested repository software Invenio, and

h) being integrated into reporting lines for research funded by the European Commission.

PHIRI
Population Health Information
Research Infrastructure

4. Content of the Open Source publication

Specifically, the following digital objects have been published in Zenodo:

Five common data models: one for the demonstration pilot and four referred to the WP6 use cases.

The common data models are the consequence of the formalization of the research queries included in these use cases in a way that is semantically interoperable for all the participant nodes. The research queries are:

a) Use case A, Vulnerable populations, inequalities and risk factors with direct or indirect impact on health care outcomes during the COVID-19 pandemic: To clarify whether health care utilization patterns in vulnerable populations vary between settings and over time and is linked to the COVID-19 epidemiological situation, using individual level health record, administrative and research data combined with ecological/group level contextual data;

b) Use case B, COVID-19 related delayed care in breast cancer patients "Was there any delay in the treatment of breast cancer patients because of the COVID-19 stringency measures?": Demonstrate whether there has been any increase in surgical and/or co-adjuvant (i.e. radiotherapy, chemotherapy, hormonotherapy and immunotherapy) treatments delay because of the COVID19 crisis in eligible women diagnosed of breast cancer using individual level health record, administrative and research data combined with ecological/group level contextual data;

c) Use case C, Effects of the COVID-19 pandemic on maternal and newborn health: Investigating the pandemic's direct (infection by SARS-CoV-2) and indirect effects on perinatal health using routine population birth data, and assess whether effects differ by socioeconomic context;

d) Use case D, COVID-19 related changes in population mental health: Measure changes in population mental health associated with the COVID-19 pandemic and, in particular, demonstrate whether there has been any increase in healthcare utilisation of mental health services and treatments as a consequence of the COVID19 crisis in eligible patients diagnosed of depression and/or anxiety using individual level health record, administrative and research data combined with ecological/group level contextual data;

e) Demonstration pilot for rapid-cycle federated analysis. Piloting the federated analyses of the evolution of a number of indicators for the monitoring and surveillance of the COVID-19 pandemic (7-day reproductive number, 7- and 14-day incidence rates, regular and ICU admissions, bed occupancy rates) and 7- and 14-day predictions, using individual data from a number of data sources -COVID-19 monitoring and surveillance registries, hospital data, and administrative data.

The common data models can be found here:

WP6 - Use Case A Common Data Model https://doi.org/10.5281/zenodo.5148013; (v3.0.0)

WP6 - Use Case B Common Data Model https://doi.org/10.5281/zenodo.5148021; (v2.0.0)

WP6 - Use Case C Common Data Model https://doi.org/10.5281/zenodo.5148031; (v2.0.0)

WP6 - Use Case D Common Data Model https://doi.org/10.5281/zenodo.5148039; (v2.0.0)

WP7 – Pilot Common Data Model https://doi.org/10.5281/zenodo.5148466; (v1.0.0)

PHIRI
Population Health Information
Research Infrastructure

Five algorithms for data analysis scripts and HTML interactive reporting solution for the PHIRI use cases A to D (plus the Pilot in WP7). Once the dataset has been accepted by the researchers (i.e., the data quality assessment provided recommendations to follow), the analytical scripts are run to produce the actual research results. Analytical algorithms can be found here:

WP6 – Use Case A scripts (R Markdown) https://doi.org/10.5281/zenodo.6359850; (v1.1.1)

WP6 – Use Case B scripts (R Markdown) https://doi.org/10.5281/zenodo.6359893; (v4.0.1)

WP6 – Use Case C scripts (R Markdown) https://doi.org/10.5281/zenodo.6380733; (v2.0.1)

WP6 – Use Case D scripts (R Markdown) https://doi.org/10.5281/zenodo.6359904; (v1.1.1)

WP7 – Pilot scripts (R Markdown) https://doi.org/10.5281/zenodo.7092522; (v1.0.0)

PHIRI
Population Health Information
Research Infrastructure

# 5. Next steps in the integration of the PHIRI FRI in the PHIRI HIP

As use cases evolve (new nodes, pooled analyses, etc.), the digital objects should reflect any changes. The help desk setup for the orchestration of the use cases (so far 39 tickets, 253 emails and 19 meetings to support use case´s deployment) is in charge of the publication of any update until the end of the PHIRI project. The current design should evolve towards a more peer-to-peer structure where the Help desk function to support use cases' deployment will be part of the services provided by the PHIRI HIP.

Moreover, while in Zenodo there is a publication that allows the persistence of the versions, to build capacity in the community, there is a need to include all the workflows as a part of the knowledge base of PHIRI, using the PHIRI HIP as a landing page. Along the upcoming months, in addition to the Publication of FAIRified use cases analysis results on the HIP, a team from WP4, WP6 and WP7 will design and include the workflows in the HIP using WP6 results as an illustration of the federated research methods, in a tutorial style.

PHIRI
Population Health Information
Research Infrastructure

6. Disclaimer

Disclaimer excluding Agency and Commission responsibility

The content of this document represents the views of the author only and is his/her sole responsibility. The European Research Executive Agency (REA) and the European Commission are not responsible for any use that may be made of the information it contains.
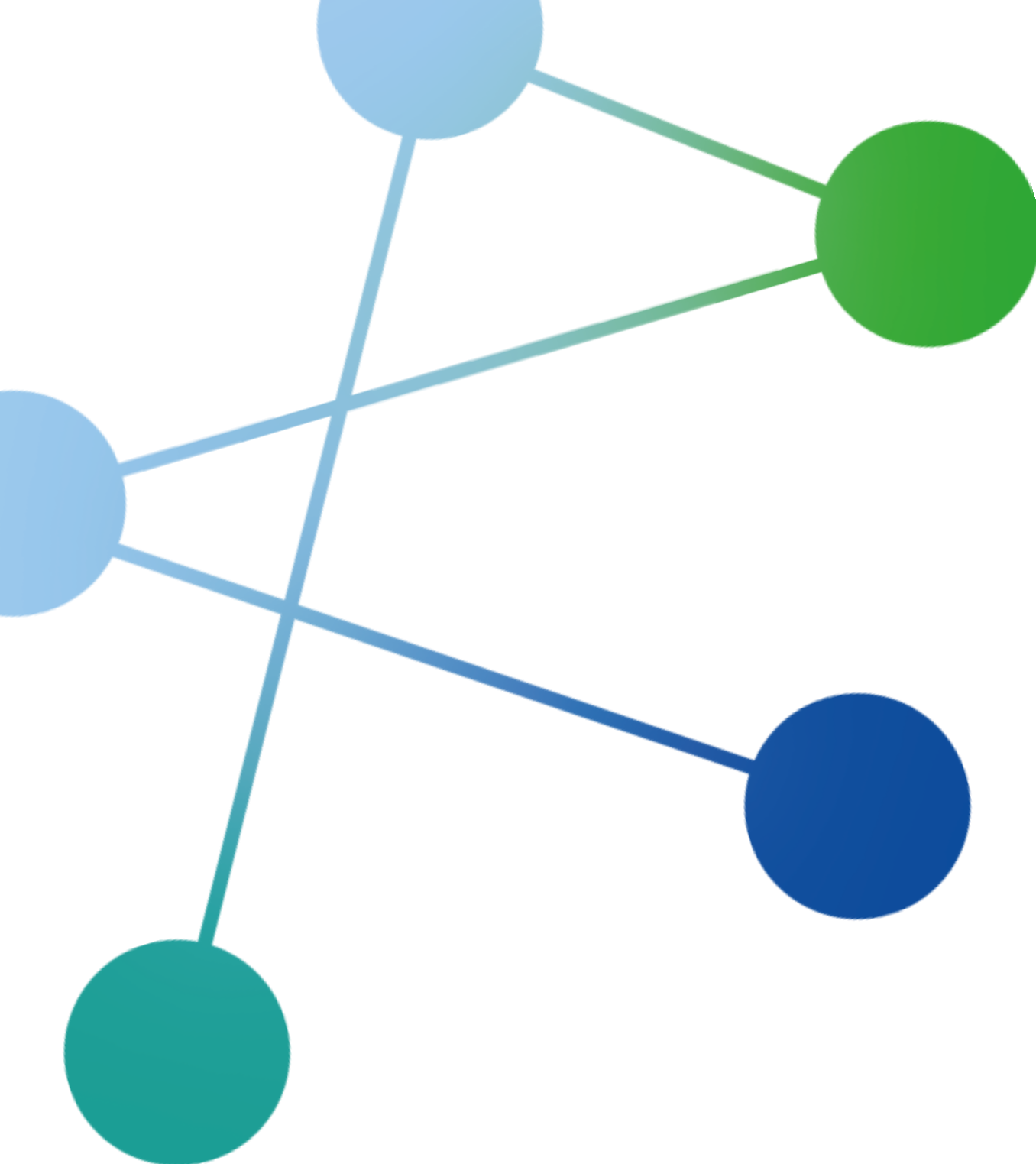
IACS

Centro de Investigación Biomédica de Aragón (CIBA)
Avda. San Juan Bosco, 13, planta 0
50009 Zaragoza, Spain
Ebernal.iacs@aragon.es
www.phiri.eu
@PHIRI4EU

© 2022| published by IACS