COVINFORM

CORONAVIRUS VULNERABILITIES AND INFORMATION DYNAMICS RESEARCH AND MODELLING

D2.1 Database containing different data sources

Project

Acronym	**COVINFORM**
Title	COronavirus Vulnerabilities and INFOrmation dynamics Research and Modelling
Coordinator	SYNYO GmbH
Reference	101016247
Type	Research and Innovation Action (RIA)
Programme	HORIZON 2020
Topic	SC1-PHE-CORONAVIRUS-2020-2C Behavioural, social and economic impacts of the outbreak response
Start	01 November 2020
Duration	36 months
Website	https://covinform.eu

Consortium	**SYNYO GmbH (SYNYO)**, Austria
	Magen David Adom in Israel (MDA), Israel
	Samur Proteccion Civil (SAMUR), Spain
	Università Cattolica del Sacro Cuore (UCSC), Italy
	SINUS Markt- und Sozialforschung GmbH (SINUS), Germany
	Trilateral Research LTD (TRI UK), UK
	Trilateral Research LTD (TRI IE), Ireland
	Kentro Meleton Asfaleias – Center for Security Studies (KEMEA), Greece
	Factor Social Consultoria em Psicossociologia e Ambiente LDA (FS), Portugal
	Austrian Red Cross (AUTRC), Austria
	Media Diversity Institute (MDI), UK
	Societatea Naţionalǎ de Cruce Rosie Din România – Romanian Red Cross (SNCRR), Romania
	University of Antwerp (UANTWERPEN), Belgium
	Sapienza University of Rome (SAPIENZA), Italy
	University Rey Juan Carlos (URJC), Spain
	Swansea University (SU), UK
	Gotenborg University (UGOT), Sweden

Deliverable

Number	**D2.1**
Title	**Database containing different data sources**
Lead beneficiary	TRI
Work package	WP2
Dissemination level	Public (PU)
Nature	Report (RE)
Due date	31.08.2021
Submission date	31.08.2021
Authors	**Niamh Aspell**, TRI
	Ilaria Bonavita, TRI
	Su Anson, TRI
	Mark Paine, TRI
	Kush Wadhwa, TRI
Reviewers	**Hayley Watson**, TRI
	Toby Fenton, TRI
	Diotima Bertel, SYNYO

Document history

Version	Date	Comments
0.1	20.08.2021	First full draft completed and shared with reviewer team (SYNYO).
0.2	23.08.2021	Review SYNYO
1.0	27.08.2021	Final version

Executive Summary

The COVINFORM project investigates the impact of the COVID-19 pandemic and national, regional and local responses with a particular focus on vulnerable and marginalised groups. The project will develop various outputs, including an interactive risk assessment dashboard to provide insights for various stakeholders (e.g., policy makers, first responders and researchers) to inform their response to further COVID-19 waves or future pandemics. In addition, guidelines and recommendations will be developed to ensure that those most vulnerable are considered in future policy responses.

This report provides an overview of the desk-based research undertaken to identify relevant COVID-19 studies and indicators to map the responses and impact across four levels: 1) national level across the EU27 MS and the UK; 2) regional and 3) local levels in 15 target sub-national research sites (Austria, Belgium, Cyprus, France, Germany, Greece, Italy, Ireland, Israel, Portugal, Romania, Spain, Sweden, Switzerland, UK, and the US), dependent on data availability[1]; 4) local levels pertinent to the case study sites. For the national level, we have gathered and reviewed relevant COVID-19 data sets including the 'Oxford COVID-19 Government Response Tracker' (OxCGRT), and baseline data identified by Eurostat and National Statistics Organisations. In addition, numerous COVID-19 risk assessment models and vulnerability indexes have been reviewed, including the INFORM Epidemic Risk Index and the 'British Red Cross – Vulnerability Index' (BRC-VI). Furthermore, supplementary survey data was collected and reviewed, for example, the 'Measuring Worldwide COVID-19 Attitudes and Beliefs' project. Each dataset has been reviewed based on a standard list of requirements for inclusion: geography (to date at Level 1: national level 27 MS and the UK), data access, data quality and indicators that accurately represent the objectives of the COVINFORM risk assessment model. In addition, data attributes were assessed to determine an appropriate database architecture including the data characteristics (i.e., volume, type, frequency of updates and availability).

A summary of identified databases containing data on COVID-19 impacts (e.g., deaths, cases and housing) and responses (e.g., ban on mass gatherings and curfews) is described, including a justification for inclusion or exclusion in the COVINFORM risk assessment model. Eligible databases will be ingested to the STRIAD application of the research partner, TRI and stored in a cloud-based data lake (central online repository hosted by Amazon Web Services S3). The interactive dashboard will provide important insights on the COVID-19 status at various geographic levels with consideration to varying levels of vulnerability as defined by the COVINFORM vulnerability index.

---

1 Initial data collection activities have focused on Level 1) national level across the EU27 MS and the UK. Inclusion of local level research sites, Israel, and the US will depend on data availability, access, and suitability of mapping data with the chosen risk assessment model based on country-specific economic, social and legal factors.

Contents

Figures

Tables

Acronyms & Abbreviations

Term	Description
API	Application Programming Interface
AWS	Amazon Web Service
BRC-VI	British Red Cross – Vulnerability Index
CCVI	COVID-19 Community Vulnerability Index
DoA	Description of Action
ECDC	European Centre for Disease Prevention and Control
GDPR	General Data Protection Regulation
NPI	Non-pharmaceutical intervention
RT-PCR	Reverse transcription polymerase chain reaction
RI	Resilience Index

1 Introduction

This deliverable reports the identified COVID-19 surveys and data sources to map the responses and impacts of the coronavirus pandemic across four levels 1) national level across the EU27 MS and the UK; 2) regional and 3) local levels in 15 target sub-national research sites (Austria, Belgium, Cyprus, France, Germany, Greece, Italy, Ireland, Israel, Portugal, Romania, Spain, Sweden, Switzerland, UK, and the US), dependent on data availability[2]; 4) local levels pertinent to the case study sites Through desk-based research, we reviewed numerous datasets with consideration of the countries included, the relevance of the indicators, the quality of the data and the available Application Programming Interfaces (APIs), which serves as an access point to the database. Data sources deemed suitable are stored in a central database for data processing and cleaning (Task 2.2). The data sources will be used to map the government's economic and public health responses and their impacts across Europe. First at national level across the EU27 MS and the UK, then regional and local levels in the 15 target countries, dependent on data availability and finally at a local level in 5-10 case study sites.

The project will conclude in the development of an interactive dashboard and visual toolkit for stakeholders in government, public health, and civil society integrating data streams, indices and indicators, maps, and models, including primary research conducted by the COVINFORM project partners and integrated qualitative case study findings (such as interview transcripts, photos, drawings, and verbal stories). This report describes the process for data identification and review.

In Section 2, the data identification strategy is described, covering the data source search strategy and eligibility criteria. This is followed by a description of how the initial indicators were identified in Section 3. This section also includes an analysis of various COVID-19 response and impact data sources for inclusion. Section 4 presents several COVID-19 risk assessment models, a description of the attributes of each and a justification for choosing the 'British Red Cross – Vulnerability Index' (BRC-VI) as an appropriate methodology to pursue. An overview of the BRC-VI methodology is presented, as well as the ongoing process for development in the COVINFORM target countries. This process is described for our first target country of focus, Spain. Section 5 and Section 6 describe how the data will be ingested and stored in project partner's, TRI's, proprietary STRIAD application with consideration of data privacy and data access. The report concludes with an overview of the next steps in data collection activities, data processing and the development of the risk assessment model.

---

2 Initial data collection activities have focused on Level 1) national level across the EU27 MS and the UK. Inclusion of local level research sites, Israel and the US, will depend on data availability, access and suitability of mapping data with the chosen risk assessment model based on country-specific economic, social and legal factors.

2 Identification of data sources

2.1 Background

Open data plays an important role in the response to global emergencies, as demonstrated in previous viral outbreaks, such as Ebola in 2016 (Cora et al. 2017) and Zika in 2017 (Bragazzi et al. 2017). In the present coronavirus pandemic, vast amounts of new data have been and continue to be generated. A key aim of the COVINFORM project is to review and analyse the available data in the development of a risk assessment model that will support the evaluation of government responses and the impacts, particularly on vulnerable groups.

The aim of the data collection task in the COVINFORM project is to perform desk-based research to identify and review relevant COVID-19 studies, indicators, data sources and models, for inclusion in the COVINFORM risk assessment model. The model is built in Task 2.3, 'Development of risk assessment tools to map the response, impact and consideration of vulnerability in each sub-national research site' and reported in Deliverable 2.3, 'Technical Report'. A data collection plan was established at the outset of the project and is described below. The data collection plan initially focused on exploring the key indicators based on the project themes through review of desk-based research conducted by project partners in their target countries (Baseline Reports: D4.1 Governmental Response, D5.1 Public Health Responses, D6.1 Community and Citizen Responses, and D7.1 Communication and Information). As part of the work undertaken in WP2, we have reviewed global and domain specific (e.g., economic vulnerability index) vulnerability indexes for their appropriateness in describing the characteristics of vulnerable groups as they are described in the project. A number of indexes have been proposed during the COVID-19 pandemic (Macharia et al. 2020; DeCaprio et al. 2020; Tiwari et al. 2021; Daras et al. 2021) and important features of each were considered for the development of the COVINFORM risk assessment model. Several online repositories, databases, models and software tools have been critically reviewed for identification of appropriate indicators at different administrative levels across the EU27 MS and the UK. As the COVID-19 pandemic continues to generate a large volume of data, data collection activities will be ongoing over the lifecycle of the project to ensure insights generated are relevant and up-to-date.

2.2 Data Collection Methodology

Prior to identification of relevant data sources and indicators, a thematic lens was applied to group the domains of interest, where insights would result. This required the 'population' of thematic groupings which can be described by multiple relevant indicators, as shown in Table 1 and Figure 1.

2.2.1 Data source search strategy

The data collection for this first deliverable focused on quantitative structured datasets (datasets in tabular form where values of each variable are numbers or categories). Since the project's primary data collection process has not yet started and the format that the data and associated indicators will take is unclear, the focus has been on investigating the availability of secondary data (i.e., already existing data collected by other organisations and institutions for other research purposes) and their coverage of the topics and geographical areas of interest to the COVINFORM project. For regional and local level data, the technical and social scientists will further interpret the qualitative data available (e.g., interview transcripts, verbal stories, pragmatic narratives, lockdown diaries, blogs, short films

such as BBC's Lockdown Diaries) and the COVINFORM primary data to further determine the indicators that emerge.

Online searches were conducted from January 2021 on and are ongoing as new data becomes available. Based on the project requirements the following eligibility criteria for inclusion were as follows:

2.3 Data inclusion criteria

Participant Location:

- Across four geographical levels: 1) national level across the EU27 MS and the UK; 2) regional and 3) local levels in 15 target sub-national research sites (Austria, Belgium, Cyprus, France, Germany, Greece, Italy, Ireland, Israel, Portugal, Romania, Spain, Sweden, Switzerland, UK, and the US), dependent on data availability; 4) local levels pertinent to the case study sites.
- Data should be available in English language

Population Characteristics:

- All age groups, where possible
- COVID-19 impact (e.g., deaths, cases, vaccinations, employment, housing etc.)
- COVID-19 response (e.g., public health and economic measures implemented by governments or other relevant actors)
- Indicators of vulnerability such as health and wellbeing, socio-economic factors, access (i.e., facilities, services and infrastructure) and environmental data
- Indicators of vulnerable groups (e.g., ethnics minorities, clinically and/or economically vulnerable) at national and sub-national level
- Specific communities (e.g., healthcare workers, migrants, and ethnic minorities) for 5-10 case studies sites

Themes:

- The variables collected are mapped to the main themes identified in the project: WP4 Government responses and impact assessment, WP5 Public Health responses and impact assessment, WP6 Citizen and Community responses and impact assessment and WP7 Inclusive COVID-19 communication for behaviour change and addressing misinformation (see Table 1).

Table 1. Main themes for the assessment of responses and impacts of the COVID-19 pandemic

Government (WP4)	Public Health (WP5)	Citizen & Community (WP6)	Communication (WP7)
Pandemic Planning and Preparedness	Impact of social factors (family, gender, age, cultural norms, trust)	Local impacts of government response	Communication strategies and practices
Approaches to Vulnerability	Organisational Procedures	Voluntary and citizen-led response	Digital communication and exclusion
Response at different geographical levels	Public Health Communication		Malinformation

Economic Response	Impact on healthcare workers		Disinformation
Social Response			Misinformation
Legal response			Behaviour Change
Ethical Factors			

Access:

- Open-source data, data that can be freely downloaded, stored, corrected, and used by external organizations, will be primarily leveraged but closed access data is also considered on a case-by-case basis.
- If closed access data is identified and considered valuable for inclusion in the COVINFORM database, a request will be submitted to the data-owner or institution/organization to request data access. Typically, sharing agreements for closed data include restrictions on the use, storage, correction of erroneous data and how the data may be displayed (i.e., allowing geodata to be displayed at different levels and on different maps).

3 Identification of relevant indicators

Prior to data collection activities, a number of relevant indicators to map the responses and impacts were identified through literature review and COVID-19 dashboards and risk assessment models. As the pandemic continues, new indicators will likely be deemed more relevant to the present situation and may be incorporated in the risk assessment model, as data becomes available. For example, at the outset of the COVINFORM project little was known regarding vaccinations and vaccination programmes, such as, expected population uptake and impacts on deaths and hospitalisations. These indicators are most relevant to the COVID-19 impacts, both in terms of health and economic impacts, and were highlighted as important in the COVINFORM baseline reports and supplementary literature review, as described below.

3.1 COVINFORM Baseline Report

In total, four country level, baseline reports are outputs of the COVINFORM project, as follows:

- Deliverable 4.1 Baseline report: Governmental responses [Submitted April 2021]
- Deliverable 5.1 Baseline report: Public health responses [Submitted May 2021]
- Deliverable 6.1 Baseline report: Community and citizen responses [Submitted August 2021]
- Deliverable 7.1 Baseline report: Communication and information [Submitted March 2021]

Following the review and analysis of Deliverables 4.1, 5.1,6.1 and 7.1, several relevant indictors were identified and deemed necessary in understanding the government, public health and community/citizen-led response and impacts, as well as indicators for effective communication and information practices (Figure 1). Figure 1 also illustrates the primary data collection activities planned for each Work Package (WP), which will be conducted from M12 of the project. A summary of the data collected from the case studies and empirical research will be included in Deliverable 2.8, "Database containing different data sources – update M30" [April 2024].
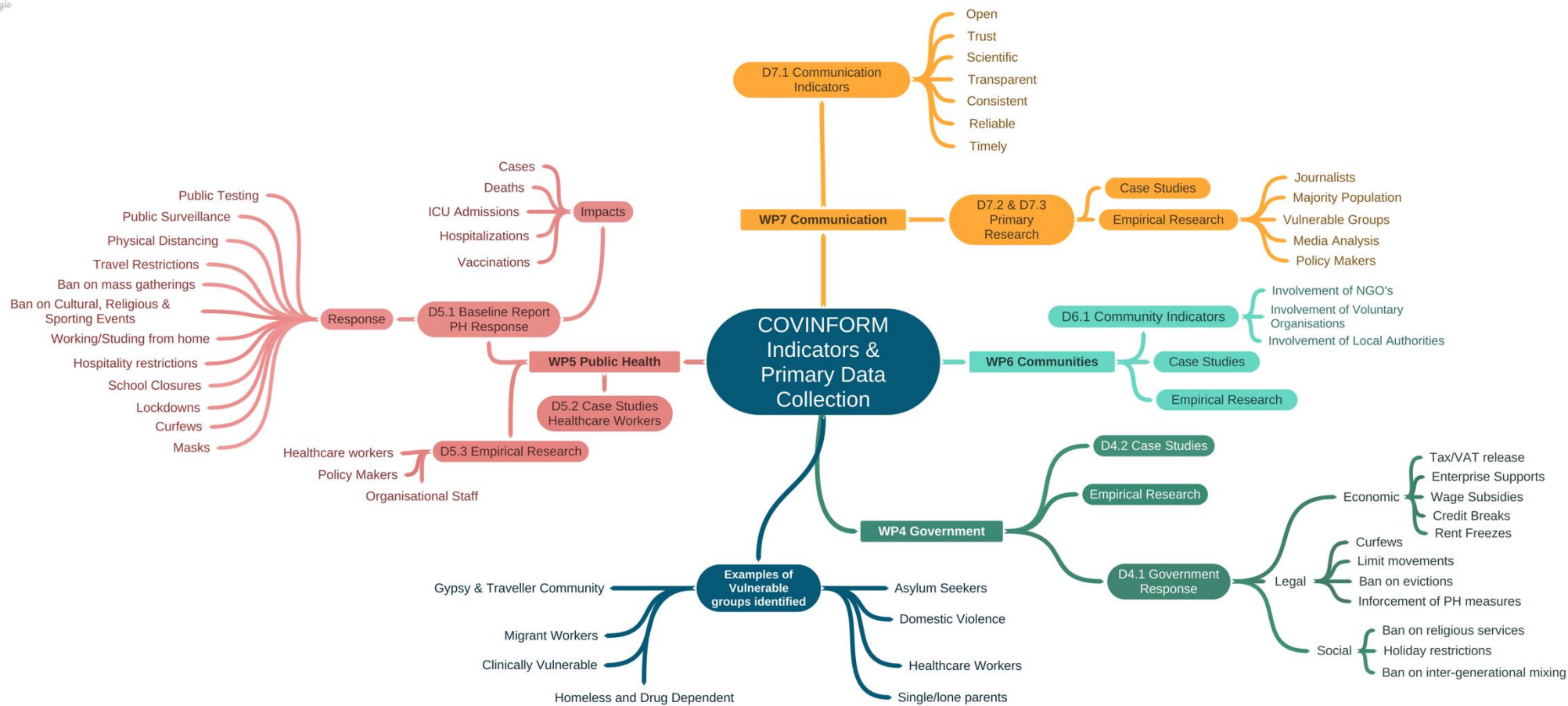
Figure 1. COVINFORM Indicators for COVID-19 impacts and vulnerable groups

At the start of 2020, in response to an increasing number of admissions to hospital and intensive care units across Europe, government decisions were made to implement response measures to decrease the spread of coronavirus. The purpose of response measures, or non-pharmaceutical interventions (NPIs) such as physical distancing and wearing of masks, was to reduce the level of contact between susceptible and infectious individuals[3]. NPIs have proven effective at reducing COVID-19 cases but have social and economic consequences, which are exacerbated by periods of lifting and re-implementing measures. As shown in Figure 1, several public health and government responses have been identified by the COVINFORM project. In addition, the public health and economic impacts are also listed. To gain insights from the COVINFORM dashboard, data for COVID-19 responses at a national level in the EU27 MS and the UK is required. Additionally, understanding the changes overtime, from peaks of virus transmission to times of recovery and the 'stringency' (i.e., strictness) of the measures is required to strengthen the understanding of the impacts throughout the pandemic.

Note: An overview of local level data at case study site level will be included in the updated report, D2.8 in M30 [April 2024].

3.2 COVID-19 Response databases

Several online databases have been developed since the start of the pandemic to track government responses globally. To evaluate the response measures implemented during the COVID-19 pandemic across the EU27 MS plus the UK, several online databases were evaluated for suitability as secondary data sources for inclusion in the project dashboard. Inclusion criteria followed the requirements described above in Section 3.2, 'Data set inclusion criteria'. In tandem, D4.1/D5.1/D7.1[4] provided a cross comparison of the response measures implemented in the 15 target countries and the representation of these measures in the databases of interest, as described in the Description of Action (DoA).

In the following text, we review three COVID-19 response databases representing the EU27 MS and the UK for inclusion in the COVINFORM model. These databases were included due to their geographical coverage, continuous data review and update practices, data quality procedures and access features (i.e., open access to download in tabular format or direct sharing via API). Multiple databases were chosen as no single database included all response measures identified in the COVINFORM baseline reports.

Oxford COVID-19 Government Response Tracker (OxCGRT)

The OxCGRT project collects information on government policy measures that have been implemented in response to the COVID-19 pandemic[5]. The tracker includes data collected since the 1 January 2020 and continues to be updated as new policies are implemented. The data collection is based on news articles and government press releases and briefings (ibid.). It includes policy response data from 180 countries (including the EU27 MS and the UK), with the aim to compare government responses, their effect on the rate of infection and to identify independent factors that produce varying response outcomes. The tracker includes a standardised list of indicators and creates an overall score to measure the impact of these responses. The complete dataset is available open access via API[6] with various

---

[3] https://www.ecdc.europa.eu/sites/default/files/documents/Framework-for-tuning-COVID-19-response-measures.pdf
[4] D6.1 was submitted later (August 2021).
[5] https://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker
[6] https://covidtracker.bsg.ox.ac.uk/about-api

resources including technical reports, images, and workflows available and regularly updated on GitHub[7].

Indicators for policy response will continue to be updated as COVID-19 is an on-going crisis, and a number of new indicators have been identified since the completion of the COVINFORM Baseline reports, such as 'vaccine prioritisation', 'vaccine eligibility/availability' and 'vaccine financial support'. Data for new indicators will be added overtime and as relevant, for example, as businesses and industries re-open across Europe, government measures regarding aviation or hospitality sectors or changes to economic supports may result in new impacts. At present, the indicators are aggregated into four policy indices:

1. Overall government response index
2. Stringency index (Cheng et al., 2020)
3. Health index
4. Economic support index

The indices do not provide information on enforcement of measures or characteristics of the target populations that may affect the rate of infection. At present, data is available at national level, regional and local level data will be included in future iterations to enhance the analysis of government responses within the target country. Of the indicators identified in this tracker, data is not reported for some indicators evidenced in the Baseline Reports (i.e., wearing of masks and closure of restaurants and bars), however, additional economic measures are described ('investment in healthcare', 'investment in vaccines' and 'vaccine policies'). These indicators have been documented in the COVINFORM baseline reports based on project partners local knowledge and are therefore deemed important in the representation of country specific response measures in the final interactive dashboard.

Response2covid19

Response2covid19[8] is another online tracker comparing governments' responses across the globe, to COVID-19 (Porcher 2020). The dataset of government responses includes worldwide coverage for measures described in Table 2.

Table 2. Response2covid19 response measures

Public health measures	▪ masks mandates, school closures, domestic lockdowns, curfews, international travel bans, domestic travel bans, bans on mass gatherings, cancellation of sporting and large events, elections postponing, restaurants closures, public testing policies, enhanced surveillance via mobile apps or bracelets, and the declaration of the state of emergency
Economic measures	▪ direct cash transfers, wage support, credit schemes, tax cuts, tax delays, support to exporters or importers, and interest rate cuts from the Central Bank

Like the OxCGRT tracker, the response2covid tracker also includes a measure of stringency or 'rigidity'. In this instance, public health measures are based on implementation: no measures, partial or regional scale (e.g., social gatherings limited to outdoors and groups of six from no more than two households, within a specified region) or strict/national (e.g., no social gatherings allowed across a country). For economic measures, the coding is: implemented / not implemented. Data are available open access,

---

[7] https://github.com/OxCGRT/covid-policy-tracker
[8] https://response2covid19.org

are updated monthly and the technical descriptions and methodologies are freely available on GitHub[9]. Additionally, data has been merged in respect to COVID-19 impacts, specifically daily cases and deaths from the European Centre for Disease Prevention and Control (EDCD). Unlike the OxCGRT tracker, the response2covid19 dataset considers additional important indicators, such as elections and enhanced surveillance.

CORONANET

The CORONANET project has generated a database of government response to coronavirus worldwide[10]. The dataset currently includes over 20,000 separate policy notifications or announcements by government at a global scale since 31 December 2019. In total, 260 research assistants, in 18 time zones, update and release the data daily. Daily data includes the type of government policy (e.g., closure of schools), level of authority implementing the measure (e.g., national or local government), geographical area affected, the direction of the policy action and transport mode (for example, inbound flights), if measures were mandatory or voluntary, who was responsible for enforcing the measure and for how long they were or would be in place. CORONANET also includes a machine-learning component which collects more than 200,000 news articles from around the world related to COVID-19.

As illustrated in Figure 1 above, the response indicators identified in the COVINFORM Baseline reports will be the primary response indicators of interest in the dashboard development and data will be extracted from each of the response databases described within this section. Project partner consultation for each target country, COVINFORM empirical data collection and case study analysis will inform tailored requirements for representation of responses at regional and local level.

3.3 COVID-19 Impacts

As illustrated in Figure 1, several COVID-19 impacts were identified in terms of public health and economic circumstances. For the first iteration of data collection, we will focus on national level reported COVID-19 related deaths and cases. A review of the determinants of COVID-19 deaths are presented in Table 3 as supplementary data for consideration in the development of the risk assessment model vulnerability index. The ECDC dataset has been reviewed for relevant data and will be utilised to provide insights into the public health impacts of COVID-19 at a national level[11]. The main purpose of the ECDC is to conduct activities including surveillance for infectious disease across Europe. COVID-19 data is collected and available open access by API for deaths, cases, vaccination rates, testing and new variants. The data is updated on a weekly basis from 196 countries, following a data screening process by ECDC's standard epidemic intelligence process for which each data entry is validated and reported in an ECDC database[12].

3.3.1 COVID-19 Deaths

To fulfil the development of the risk assessment model (Task 2.3), an understanding of the factors contributing to COVID-19 impacts is required. The first impact of importance is COVID-19 deaths, the

---

[9] https://github.com/simonporcher/COVID-19-Governments-Responses
[10] https://www.coronanet-project.org/
[11] https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases
[12] https://www.ecdc.europa.eu/en/publications-data/online-tutorial-how-detect-and-assess-public-health-threats

increased risk of death due to COVID-19 was explored by means of a literature review. As reported on the ECDC web portal, the WHO definition for surveillance of a COVID-19 death is;

> 'a death resulting from a clinically compatible illness in a probable or confirmed COVID-19 case, unless there is a clear alternative cause of death that cannot be related to COVID disease (e.g., trauma). There should be no period of complete recovery between the illness and death. A death due to COVID-19 may not be attributed to another disease (e.g. cancer) and should be counted independently of pre-existing conditions that are suspected of triggering a severe course of COVID-19'[13].

However, summary data differs across Europe making cross country comparison difficult. In general, data is reported in two ways, in real-time by official government sources via national broadcast and by national statistics offices that obtain data from death registries which is most accurate but requires longer processing time. The definition of a COVID-19 death does not always follow the WHO definition; for example, in Austria, Italy, the Netherlands, Spain and the UK, a COVID-19 death is primarily based on a positive laboratory sample.[14] For example, case-fatality statistics in Italy are based on defining COVID-19–related deaths as those occurring in patients who test positive by real-time reverse transcription polymerase chain reaction (RT-PCR) analysis, independently from pre-existing diseases that may have caused death (ibid.). The differences in COVID-19 death definitions between European countries has been reported and should be considered when interpreting COVID-19 data on deaths[15]. For the COVINFORM project, data from the ECDC, who endorse the WHO definition, will be included.

Several authors have attempted to determine the factors associated with increased risk of death from COVID-19 using worldwide or national level data. As shown in Table 3, re-occurring determinants of death identified include international travel, prevalence of cancer, prevalence of non-communicable disease (diabetes, hypertension, obesity), intensity of COVID-19 testing, case numbers, income, age (older ages being at greater risk) and ethnicity. As anticipated, vaccination is preventative and reduces the risk of COVID-19 related deaths. These indicators were cross-referenced with the COVID-19 vulnerability model development framework described in the following section of this report.

Table 3. Determinants of COVID-19 deaths

Author (year)	Country of analysis	Statistically significant determinants of COVID-19 reported deaths
Pana et al. (2020)	Thirty-seven countries[16]	International arrivals (increase)
		Prevalence of cancers (age-standardised for both sexes) (decrease- lacked data)
		Prevalence of hypertension (age-standardised >=140 or >=90) (decrease -lacked data)
		GDP per capita (Purchasing power parity)
		BCG vaccination coverage (decrease)
		Testing capacity (total COVID-19 tests per 1000 pop performed) (increase)
Kenyon, C. (2020)	Worldwide	Intensity of COVID-19 testing
		Cumulative COVID-19 cases

---

[13] https://www.ecdc.europa.eu/en/covid-19/surveillance/surveillance-definitions

[14] https://eurohealthobservatory.who.int/monitors/hsrm/analyses/hsrm/how-comparable-is-covid-19-mortality-across-countries

[15] https://analysis.covid19healthsystem.org/index.php/2020/06/04/how-comparable-is-covid-19-mortality-across-countries/

[16] Algeria, Argentina, Austria, Belgium, Brazil, Canada, Chile, Colombia, the Dominican Republic, Ecuador, Egypt, Finland, France, Germany, Hungary, India, Indonesia, Ireland, Italy, Japan, Mexico, the Netherlands, Peru, the Philippines, Poland, Portugal, Romania, the Russian Federation, Saudi Arabia, South Africa, Spain, Sweden, Switzerland, Turkey, Ukraine, the UK and the USA.

Seligman et al. (2021)	US – NHANES	Non-white race/ethnicity (increase)
		Income below median (increase)
		Low education (increase)
Fielding-Miller et al. (2021)	US – rural areas	% in farm work
		Household with 1 fluent English speaker
		% uninsured individuals under 65
		% living in poverty (below federal poverty line)
	US-urban areas	Higher population density
		Insurance
Karmakar et al. (2021)	US	Social Vulnerability Index (SVI) - 0.1-point increase in the overall SVI score
Velasco et al. (2021)	Worldwide analysis	Total tests
		Test/Case
		Age
		Rural Population (increase deaths)
		Temperature (environmental)
Zheng et al. (2020)	Meta-analysis-China	Males
		Older Adults (>65)
		Smoking
		Diabetes
		Hypertension
		Cardiovascular disease
		Clinical manifestation - shortness of breath
		Clinical manifestation – fever
		Lab analysis – AST, Creatinine, hypersensitive cardiac troponin, procalcitonin
Lai et al. (2020)	57 countries including the EU	Country healthcare resources (Health Care Index (HCI) and Healthcare Access and Quality Index
Goh et al. (2020)	Top 50 countries based on case fatality rate	Proportion of people aged 65 above
		Diabetes prevalence
Alberca et al. (2020)	Other	Obesity
Cowger et al. (2020)	US	Black and Latinx individuals

4 COVID-19 Vulnerability Index

The underlying objective of the COVINFORM project is to describe the response and impact at different geographical levels with a focus on vulnerability and vulnerable groups. The development of the risk assessment tool will be undertaken as part of Task 2.3 'Development of risk assessment tools to map the response, impact and consideration of vulnerability in each case study sites. However, data collection activities will consider the indicators and variables that best describe vulnerability in the case study sites.

Several risk assessment models have been created since the beginning of the COVID-19 pandemic. A particular focus was given to studies attempting to explicitly model vulnerability. Indicators of vulnerability in the COVID-19 context were assessed to evaluate the different definitions and quantification of vulnerability used in current research. For example, the COVID-19 Economic Vulnerability Index by European Investment Bank[17] was reviewed for relevant indicators for economic vulnerability. This index may be used as a reference point for proxy indicators; however, the index does not model other relevant vulnerabilities such as clinical, health and wellbeing and social vulnerabilities. Other international indexes have been created, such as the COVID-19 Community Vulnerability Index (CCVI) by Surgo Ventures[18], a vulnerability index for COVID-19: spatial analysis at the subnational level in Kenya (Macharia et al. 2020), the novel COVID-19 Vulnerability Index (Tiwari et al. 2021) and the Cornell University COVID-19 Vulnerability Index (DeCaprio et al. 2020). Each of these indexes provided valuable references for modelling vulnerability through various domains in the context of their target countries, mainly the US. Whilst the framework for these models is applicable to the development of the COVINFORM risk assessment model, indicators are based on the social and political context of the US and Africa, which is not comparable to the EU landscape in many ways (e.g., social systems, healthcare systems, access to services and geographical characteristics).

Therefore, several dashboards with global or European coverage were reviewed in greater detail for indicators of importance and to drive data collection activities (see Table 4).

---

[17] https://www.eib.org/attachments/thematic/the_eib_covid-19_economic_vulnerability_index_en.pdf
[18] https://precisionforcovid.org/ccvi

Table 4. Online COVID-19 vulnerability dashboards

Dataset	Brief description	Categories	Geographic Level	Includes EU27_UK?	Data Access
INFORM COVID-19 Risk Index[19]	The INFORM COVID-19 Risk Index is an experimental adaptation of the INFORM Epidemic Risk Index and aims to identify: "countries at risk from health and humanitarian impacts of COVID-19 that could overwhelm current national response capacity, and therefore lead to a need for additional international assistance".	Hazards and Exposure Vulnerability Lack of Coping Capacity	Country-(Worldwide)	Yes	Open
COVID-19 Pandemic Vulnerability Index (PVI)[20]	The PVI was created to communicate data to policy makers and the public. The dashboard creates risk profiles described as PVIs which provides an overall risk scores for each county in the US.	Infection Rate Population Concentration Intervention (Response) Health & Environment	County-level (United States)	No	Open
British Red Cross COVID-19 Vulnerability Index[21]	"To help focus help on the most vulnerable people whose needs aren't being met, we have developed a COVID-19 Vulnerability Index."	Clinical vulnerability Other health/wellbeing needs Economic/financial vulnerability Social vulnerability (including physical/geographical isolation)	County-level (UK)	No	Open
International Health Regulations Index[22]	"The revised International Health Regulations (IHR) are a set of legal instruments designed to ensure and improve the capacity of all signatories or States Parties to prevent, detect, assess, notify, and respond to public health risks and acute events. "	Prevent, Detect, Respond, Enabling function and Operational readiness	Country-level (Worldwide)	Yes	Open

---

[19] https://drmkc.jrc.ec.europa.eu/inform-index/inform-covid-19#:~:text=The%20INFORM%20COVID%2D19%20Risk,need%20for%20additional%20international%20assistance"
[20] https://covid19pvi.niehs.nih.gov
[21] https://docs.google.com/document/d/1aWpzgvLKGEF5Ay_xVps17nnbT1zIEki7RGIIJXL5APo/edit#heading=h.a5ipgbuvnul8
[22] https://extranet.who.int/e-spar/

As part of Task 2.5, 'Evaluation', each dashboard was presented to the project's practitioner partners (Austrian Red Cross, Romanian Red Cross, Magen David Adom in Israel, Universita Cattolica Del Sacro Cuore in Italy and SAMUR Proteccion Civil in Spain) during the first user requirements workshop on the 25 June 2021. The British Red Cross COVID-19 Vulnerability Index (BRC-VI) presented similar attributes to the goals of the COVINFORM risk assessment model, such as tailored, local level geographic representation of vulnerability and flexible model development in terms of enhancing vulnerability indexes based on new evidence. The BRC-VI hosts open access data sources with API functionality, and the ability to create bespoke models for each target country to map vulnerabilities, coping capacity and threats[23]. The domains and relevant indicators have been reviewed for the UK (including England, Scotland, Wales and Northern Ireland) and ongoing data identification is underway for each case study location. This is described in more detail below for the first target country of interest, Spain. Spain was chosen as the first target country as the project consortium includes two partners in Spain, emergency responders, SAMUR and the Universidad Rey Juan Carlos. Both partners attended the first user requirement workshop as part of Task 2.5 'Evaluation' on the 25 May 2021, and shared knowledge of potential data sources and barriers for data collection, such as standardised data collection.

The BRC-VI was developed to identify vulnerable people whose needs are not being met. The Index is comprised of four types of vulnerability:

1. Clinical vulnerability (comprised of thirty-three unique variables[24] including proportion of people over 70 and underlying health conditions, prevalence of obesity, cancer, asthma etc)
2. Health and Wellbeing needs (comprised of thirteen unique variables including prevalence of dementia, drug-related hospital admissions, serious mental illness and loneliness risk etc.)
3. Economic/financial vulnerability (comprised of thirty unique variables including adults receiving social care benefit, proportion of people receiving employment and support allowance, proportion working in the arts and hospitality etc.)
4. Social vulnerability (comprised of thirty-two unique variables including longest mean road distance to GP surgery, supermarket etc., housing and digital access e.g., broadband speed, buying online, online banking etc.)

In addition to vulnerability indicators, a Resilience Index (RI) has also been created[25]. The RI includes vulnerability (as described above), capacity to cope, and risk/history of shocks. Capacity to cope considers fourteen unique variables and nine indicators of risk of shocks calculated for Local Authorities.

Capacity to cope includes indicators such as;

- Local Authority spending power (£m per person) in the current year
- Charities per 1,000 people
- Volunteer capacity
- Community engagement (voter turnout at local elections)

Whilst risk of shocks is measured by indicators including;

- Fire and Rescue Service response times
- Percentage of people in flood risk areas

---

[23] https://britishredcrosssociety.github.io/covid-19-vulnerability/
[24] Count of unique variables is accurate to the 1 August 2021.
[25] https://britishredcross.shinyapps.io/resilience-index/

Another factor deemed essential in emergency responses, particularly when considering vulnerable groups, is the level of community engagement. This has been completed for England and is based on the Oxford Consultants for Social Inclusion[26].

> *"The Community Needs Index aims to bring together data on the social and cultural factors that can lead to poorer outcomes in communities. For example, challenges around a lack of places to meet, poor facilities, poor connectivity, and low participation. The index attempts to bring all the various sources of information on these themes together into a single index, so that we can identify which areas have the greatest challenges on those factors. It is a very different index from the Index of Multiple Deprivation (IMD) as it's more about the social needs and the social gaps that contribute towards challenges in a community – rather than the economic needs and challenges." (pg. 16)[27].*

This index includes indicators on connectedness, engagement, and community assets.

4.1.1 Mapping BRC-VI indicators to available data in each target country – example Spain

Following extensive review of the BRC-VI methodology and data sources, data collection activities through desk-based research and consultation with project collaborators in the target country have commenced. Several data sources were identified for Spain and provide coverage at national, regional and local levels. The two following data sources have been identified as containing most comparable data for the indicators described in the BRC-VI: European Union Statistics Office and National Statistics Institute. Additional data sources have been identified and may be considered for inclusion also; in total, thirty-two COVID-19-related databases have been identified, an overview of the five most applicable (data coverage and relevance of indicators) are described in Table 5.

European Union Statistics Office (EUROSTAT) (National level across the 27MS)

Eurostat is the statistical office of the European Union which provides high quality statistics and data in Europe[28]. In addition, Eurostat produces statistics in partnership with National Statistical Institutes and other national authorities in the EU27 MS and therefore is a valuable resource to the COVINFORM data collection activities. The dissemination policy is open and free access to data (some microdata are restricted access i.e., data regarding individual respondents or businesses) and demonstrates compliance in terms of confidentiality and accessibility, as described in the Regulation No223/2009[29] on European statistics and in the European Statistics Code of Practice[30]. Data is updated regularly (twice a day), and data can be extracted manually from the database or connected via API. Metadata including information on definitions, classifications, methodologies, and code lists are also available[31]. An overview of the domains of data collected, and relevant to the COVINFORM risk assessment model, can be found in Table 6 and are available open access.

---

[26] https://ocsi.uk/2020/01/15/community-needs-index-your-questions-answered/
[27] https://docs.google.com/document/d/1aWpzgvLKGEF5Ay_xVps17nnbT1zIEki7RGIIJXL5APo/edit#
[28] https://ec.europa.eu/eurostat/web/main/home
[29] http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32009R0223
[30] http://ec.europa.eu/eurostat/web/quality/european-statistics-code-of-practice
[31] http://ec.europa.eu/eurostat/data/metadata

National Statistics Institute (Spain)

The National Statistics Institute (INE) is an independent institution within the Ministry of Economic Affairs and Digital Transformation in Spain[32]. Its mandate is to regulate statistical activity for the State and it plays an important role in population statistical activity (e.g., demographic and economic censuses, national accounts, demographic and social statistics, economic and social indicators). The INE has created an open access data space with API functionality[33]. Additionally, the INE demonstrates compliance with relevant quality standards and frameworks, namely the Common Quality Framework for European Statistics[34]. Furthermore, the quality management strategy follows the principles of the European Statistics Code of Practice[35].

The INE designs the National Statistics Plan with the collaboration of the Ministerial Departments and the Bank of Spain. Agreement on concepts, statistical units, classifications, and codes are determined with specialised International Institutions and with the European Union Statistics Office (EUROSTAT). This process provides standardisation of data collection and reporting within Spain and across the EU member states. An overview of the INE dataset indicators of interest are presented in Table 5.

---

[32] https://www.ine.es/en/
[33] https://datos.gob.es/en/aplicaciones/api-json
[34] https://www.ine.es/ss/Satellite?L=en_GB&c=Page&cid=1259943453642&p=1259943453642&pagename=MetodologiaYEstandares%2FINELayout#id1
[35] https://ec.europa.eu/eurostat/documents/4031688/8971242/KS-02-18-142-EN-N.pdf/e7f85f07-91db-4312-8118-f729c75878c7

Table 5. European and national data sources containing indicators of interest to the COVINFORM target country of Spain

EUROSTAT	COVID-19 Data Portal Spain - Health[36]	DatAC (Data Against COVID-19)[37]	COVID-19 Flow-Maps[38]	National Statistics Institute (INE)
National and regional	**Regional - Catalan only**	**Provincial and autonomous communities**	**Provincial**	**National and regional**
Agriculture	Anxiety/depression and other mental illness	Activity, occupation and unemployment	COVID-19 cases	Commerce, Hostel and Tourism and Transport
Agriculture, energy, transport and tourism	Asthma attack last 12 months	Culture and education information	Mobility	Health and mortality statistics regarding COVID-19
COVID-19 related data	Asthma ever	Demographic information	Mobility-based risk	Life conditions survey
Deaths and mortality	Autoimmune disease	Economic accounts		Social Protection
Demography	Cancer	Cumulative hospitalized		Sociodemographic data regarding COVID
Digital economy and society	Chronic bronchitis, COPD or emphysema	Cumulative ICU		Spain: Economic and financial indicators
Economy	Chronic Kidney disease	Daily discharged		Standard and living conditions (CPI)
Energy	Chronic Liver disease	Daily hospitalized		
Family and households	Current asthma medication	Daily ICU		
GDP	Diabetes	Deaths rate		
Government finance	Gastro-intestinal diseases	Vaccination rate		
Health	Doctor Diagnosis of asthma	Hospitals occupancy		
Household accounts	Gynaecological diseases	ICU occupancy		

[36] https://covid19dataportal.es
[37] https://covid19.genyo.es
[38] https://flowmaps.life.bsc.es/flowboard/

Income and living conditions	Heart disease	Incidence rate		
Inflation and prices	HIV and other immunodeficiency problems	Environmental data (temperature, wind speed)		
International trade	Hypertension	**Provincial and autonomous communities**		
Labour indicators	Joint problems, rheumatism	Cumulative cases (PCR+)		
Labour market, economic and business	Nasal allergy (e.g. rhinitis)	Cumulative cases (Total)		
Population and health	Skin allergies	Cumulative deaths		
Poverty		Daily cases (PCR+)		
Production and trade		Daily cases (Total)		
Society and work		Daily deaths		
Tourism				
Transport				
Unemployment rates				

4.1.2 Current status of data collection activities for Spain

As presented above, a number of key data sources have been identified for the target country of Spain, through desk-based research and consultation with project partners (SAMUR and Universidad Rey Juan Carlos). A gap analysis is ongoing to highlight missing data for the key indictors as described in the BRC-VI methodology. The findings will be presented to both partners in Spain for a final review and an agreement on potential proxy indicators. From here, the technical team will continue their data search activities to identify sources for the proxy indicators as agreed. Once all indicators are identified and the data is obtained, statistical techniques will be employed to determine the relative importance of each indicator and proxy indicator in relation to COVID-19 responses and impacts (e.g., COVID-19 related deaths). This process will also include stakeholder engagement for 'ranking' the importance of each indicator when describing the model's component (e.g., clinical vulnerability) for each target country. Primary research conducted by URJC and SAMUR will focus on 'newcomer' migrants in Madrid. Findings of which will be mapped to the Spanish vulnerability index under sub-themes, for example, economic activity, social vulnerability, community assets and health and wellbeing, to contextualise the insights generated by the risk assessment model.

4.1.3 Data collection challenges

When mapping the indicators of interest in the BRC-VI to the 15 target countries, identifying comparable indicators will require critical review from project partners and the technical team to ensure appropriate representation of each domain of interest. In the event similar data are not available, proxy indicators will be proposed and their contribution will be ranked based on expert knowledge of the domain of interest. The final list of indicators will be statistically evaluated for their relative contribution to the vulnerability index. Where data pertaining to indicators of interest are not available, this will be clearly stated within the vulnerability index methodology and referenced in the insights generated by the dashboard for full transparency.

4.1.4 Other surveys and data sources

Beyond data collection for policy response and related impacts, a large database of worldwide open access surveys (n=72) was identified for the review of indicators. In total of 40 surveys represent geographic regions within each EU MS and the UK, or partial coverage, as shown in Table 6. These data sources will be utilised in the development of the Risk Assessment Model in T2.3, as supplementary material for refined models at sub-national research site and case study (local) level based on the BRC-VI model as the baseline model, enhanced with primary data collection activities and qualitative data.

Table 6. Additional COVID-19 surveys of interest

Survey Name	Domains	URL	EU MS_UK
SHARE	Employment and financial situation; infection and health; mental health; social networks; healthcare	http://www.share-project.org	Yes
SOEP-CoV	Employment and financial situation; children and family; COVID-related behaviours and perceptions; mental health; political attitudes	https://www.soep-cov.de/	Partial
Mannheimer Corona-Studie	Employment and financial situation; children and family; COVID-related behaviours and perceptions; mental health; political attitudes	https://www.uni-mannheim.de/gip/corona-studie/	Partial
gesis panel	Employment and financial situation; children and family; COVID-related behaviours and perceptions; mental health; political attitudes	https://www.gesis.org/gesis-panel/gesis-panel-home	Partial
LISS Panel	Employment and financial situation; children and family; COVID-related behaviours and perceptions; mental health	https://www.lissdata.nl/	Yes
Norwegian Citizen Panel	Employment and financial situation, political behaviours; health; social cohesion	https://www.uib.no/en/digsscore/122162/methodology-and-field-periods	Partial
Understanding Society	Employment and financial situation; children and family; COVID-related behaviours and perceptions; mental health; political attitudes	https://www.understandingsociety.ac.uk/research/themes/covid-19	Yes
Covid19 Behaviour Tracker	Perceptions and COVID19-related health behaviours	https://github.com/YouGov-Data/covid-19-tracker	Yes
YouGov international COVID19 tracker	Public attitudes on a broad range of topics from brands to politics	https://yougov.co.uk/topics/international/articles-reports/2020/03/17/YouGov-international-COVID-19-tracker	Yes
Attitudes Towards COVID-19 - A comparative study	Public attitudes on health; economy; civil liberties; government; policies and institutions	https://www.sciencespo.fr/cevipof/attitudesoncovid19/	Yes
COVID Inequality Project	Impact on workers and inequality across many dimensions, such as age; gender; occupation; work arrangements; and education.	https://sites.google.com/view/covidinequality/home	Partial
intergen-COVID	Intergenerational contact during COVID and mental health	https://sites.google.com/unifi.it/intergen-covid	Partial
COVID-19 Snapshot Monitoring (COSMO)	COVID-related behaviours and perceptions; mental health; political attitudes	https://www.psycharchives.org/handle/20.500.12034/2397	Partial

Austrian Corona Panel Project	Employment and financial situation; children and family; COVID-related behaviours and perceptions; mental health; political attitudes	https://viecer.univie.ac.at/en/projects-and-cooperations/austrian-corona-panel-project/	Partial
International Survey on Coronavirus	Worldwide COVID-19 Attitudes and Beliefs	https://covid19-survey.org/	Yes
COVIDiSTRESS Global Survey	Psychological stress; compliance with behavioural guidelines; trust in governmental institutions and their preventive measures	https://osf.io/z39us/wiki/home/	Yes
Living working and COVID-19	Work and well-being	https://www.eurofound.europa.eu/data/covid-19	Yes
Global Drug Survey	Alcohol and other drug use; mental health and coping mechanisms	https://www.globaldrugsurvey.com/	Yes
Opinion and Lifestyle Survey	Employment and financial situation; children and family; COVID-related behaviours and perceptions; mental health; political attitudes	https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandwellbeing/bulletins/coronavirusandthesocialimpactsongreatbritain/21august2020	Partial
COVID-19 Survey in Five National Longitudinal Cohort Studies	Physical health (including COVID-19); Time use; Family and household; Financial situation and benefits; Employment and education; Health behaviours; Mental health and social connectedness	https://cls.ucl.ac.uk/covid-19-survey/	Partial
Avon Longitudinal Study of Parents and Children	Health; Travel; COVID- related behaviours; mental health;	http://www.bristol.ac.uk/alspac/	Partial
Six-country Survey on Covid-19	Work and living situations; health; COVID related behaviours; beliefs about the Covid 19 pandemic; exposure to the virus	https://osf.io/gku48/	Partial
COVID19 impact survey	Economic and labour situation of people; prevalence of symptoms; quarantine capabilities; testing and contact tracing availability; personal protection behaviours; NPI support	https://covid19impactsurvey.org	Partial
International Study on Compliance Behaviour of Students	Behavioural response; compliance measures governments; personality	https://www.eur.nl/en/news/interview-phd-student-initiates-global-study-impact-covid-19-measures	Partial
Childcare issues throughout the COVID-19 pandemic	Usage of public childcare services and other care arrangements; children and family; parental employment	https://www.dji.de/KiBS	Partial

European Social Survey	Employment and financial situation; COVID-related behaviours and perceptions	https://www.europeansocialsurvey.org/	Partial
National Educational Panel Study	Everyday school; work; and family life	https://www.neps-data.de/Mainpage	Partial
Coping with Covid-19 (CoCo)	Social practices; social cohesion; household arrangements; socio-political attitudes	https://www.sciencespo.fr/osc/fr/content/faire-face-au-covid-19.html	Partial
pairfam COVID-19 survey	Employment and financial situation; COVID-related behaviours and perceptions; health; mental health; partnership and division of housework	https://www.pairfam.de/	Partial
NEPS - National Educational Panel Study	Work and employment; family life; schools and learning	https://www.neps-data.de/Mainpage	Partial
The German National Cohort (GNC)	Employment; health behaviours; health; mental health; social contacts	https://nako.de/studienteilnehmer/das-untersuchungsprogramm/gesundheitsfragebogen/sonderfragebogen-covid-19/	Partial
European COvid Survey (ECOS)	Information policy and trust in information sources; Agreement with introduction and lifting of restrictions; financial and economic worries; Attitudes towards vaccination	https://www.hche.uni-hamburg.de/en/corona.html	Partial
COGIS-NL: Covid19 Gender (In)equality Survey Netherlands study	Work, care and well-being between men and women in households with at least one child under 18	https://osf.io/mu7jg/	Partial
COVID-19 Social Study	Psychological and social experiences	https://www.covidsocialstudy.org/	Partial
COVID-19 Beliefs, Behaviours & Norms Survey	Beliefs; behaviours; norms	https://covidsurvey.mit.edu/	Partial
COVID-19 World Symptoms Survey	Symptoms; social distancing behaviours; mental health issues; financial constraints	https://covidmap.umd.edu/index.html	Partial
Confined childhood	Child confinement and well-being; family and children; financial situations; mental health	https://infanciaconfinada.com/	Partial

Living, Working and COVID-19	Well-being; work-life balance; employment and working from home; trust in institutions	https://www.eurofound.europa.eu/data/covid-19	Partial
Health Care and Social Survey (ESSOC)	Housing condition; physical and mental health; health behaviours; financial situations; employment	https://juntadeandalucia.es/organismos/saludyfamilias/areas/salud-vida/vigilancia/paginas/encuesta-andaluza-salud.html	Partial
Welfare at a (Social) Distance	Employment and financial situation; benefit claims and stigma	https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8689#!/details	Partial

5 Database Architecture

In order to identify the most suitable architecture for the COVINFORM database, two complementary activities were carried out by TRI: 1) assessing the volume, format and access modality of the data to be used by the COVINFORM risk and vulnerability assessment model and 2) collecting data requirements and the COVINFORM project needs by engaging end users and other partners in the consortium (1st user requirement workshop held on the 25th June 2021). Trilateral's general use application for data analysis and machine learning, STRIAD, will be applied in the COVINFORM project. Within the workshop, representatives from Trilateral demonstrated the use of the STRIAD architecture in relation to a specific use case on human security (marketed externally as the HAMOC application for military and humanitarian use). Demonstration through a specific use case was a useful mechanism to aid the co-design process, to demonstrate what is possible and to guide user requirements gathering and prioritisation. Attendees to the workshop felt that a similar approach from a use case perspective was favourable. Accordingly, the team will use the non-use-case specific application, STRIAD, and apply the architecture to COVINFORM. The STRIAD application has a flexible and modular structure than can adapted on the COVINFORM domain. COVINFORM will use STRIAD data management component for efficient and secure ingestion, indexing and harmonisation of datasets of different formats and sizes. The figure below shows the proposed architecture for COVINFORM.
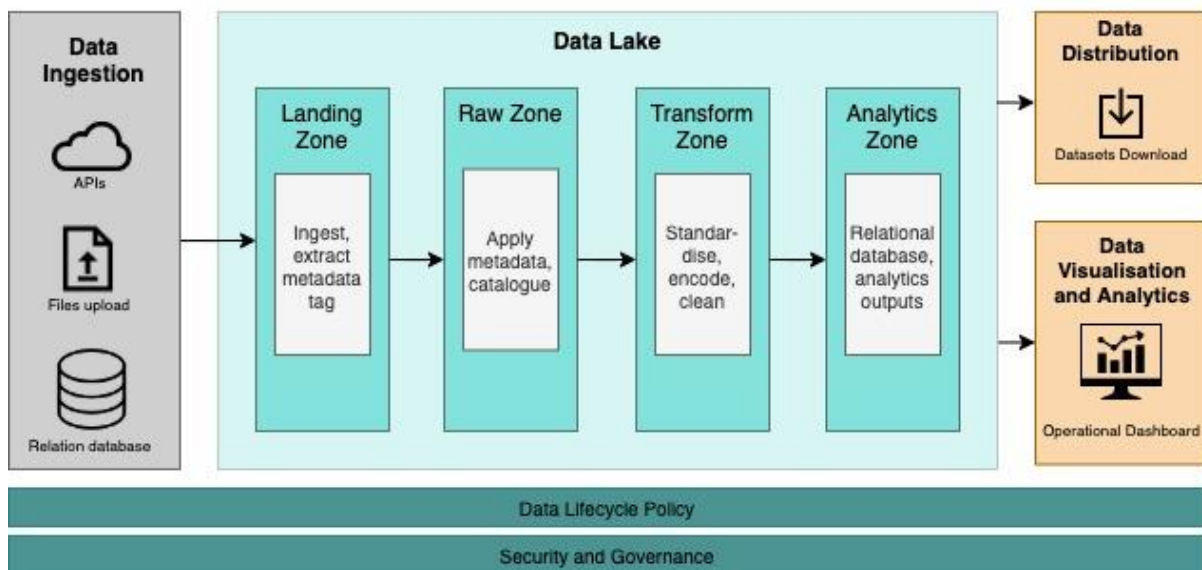

Figure 2. COVINFORM data architecture

5.1 Data ingestion

There exist three modalities of data acquisition in STRIAD:

1. Data are uploaded by STRIAD administrators into the data storage;
2. Data are uploaded by STRIAD end users;
3. Data are gathered by external APIs.

In COVINFORM, data will be ingested into the system primarily according to the first modality. The data collected in Task 2.1 and collected or generated by other tasks, will be gathered by TRI and uploaded into the designed storage. Depending on the modelling choice and final user requirements,

the option of automatically gathering data from external APIs will be considered. This may be the case for data (e.g., impacts, policies, vaccination coverage) that are updated on a daily or weekly basis by the data providers for which manual upload would be time-consuming and error prone. The possibility to allow end users to upload their own data has been discussed during the first end-users workshop and, though technically feasible, the utility of this requirement with respect to COVINFORM objectives is currently being evaluated.

5.2 Data lake

All the datasets needed by the COVINFORM interactive dashboard and models are stored in a *data lake* (built using AWS S3). A data lake is a centralised, curated, and secured repository that allows users to store structured and unstructured data at any scale. In a data lake, data can be uploaded and stored as they are (raw data), without the need of previous structuring; processed data resulting from the analytics performed on the data can also be stored, at any stage of processing, in the data lake. Due to its flexibility, a data lake can adequately meet COVINFORM data requirements such as:

- Collecting data from various sources over an extended period of time;
- Gathering qualitative and quantitative data;
- Allowing to easily upload data without strict requirements on the format.

Files stored in the data lake move across four *processing zones*. The first zone, the *Landing zone*, takes the files loaded into the system as they are. Metadata from these files are extracted, added as tags and moved into the second zone, the *Raw zone*. Here, metadata are applied to the datasets and data are categorised (categories will be defined based on COVINFORM needs). Further, catalogues files are dumped into the third zone, the *Transform zone*. Here, different transformations (standardisation, encoding, cleaning, etc.) are performed and the resulting structured data are added to a relation database (deployed with AWS RDS using PostgreSQL instances) hosted in the final *Analytics zone.* In this zone, all the outputs generated by the analytics are stored and ready to be retrieved or visualised on the dashboard.

6 Considerations for Database Privacy and Access

6.1 Data lifecycle policy

STRIAD implements a specific data lifecycle policy that will be applied and, if necessary, adapted, to COVINFORM. Data lifecycle policies help to manage and control the flow of data through the data lake during all the processing stages. In terms of security, the highest security and encryption industry protocols are implemented (e.g., data resides in Virtual Private Clouds, data masking and tokenisation is performed) and server-side encryption will be enabled.

Additionally, to prevent any accidental deletion of data corruption, replication rules for datasets in the data lake will be provided.
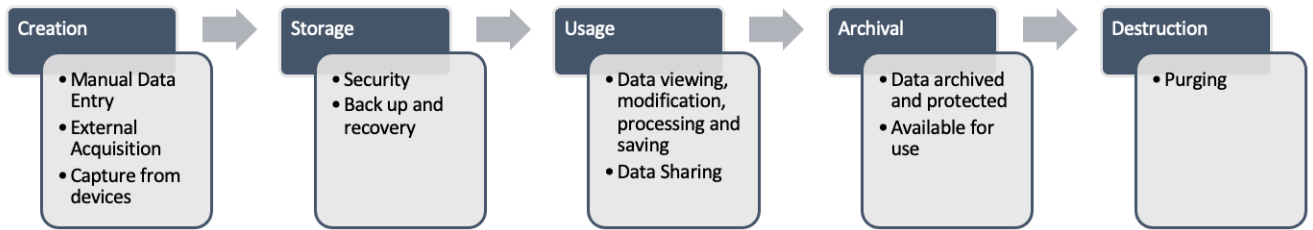
Figure 3. Data lifecycle policy

6.2 Data security and governance

Access policy rules for COVINFORM end users will be set up along with account management using secure protocols provide by services such as AWS Cognito and AWS IAM. Data related to an end-user organisation (End-user ID, Name, Users, etc.) will be securely stored. Each organisation will have an allocated user pool which will not be shared with other client organisations. Each user pool will have groups created using AWS IAM Roles, mainly - Admin, Privileged Users and Users - which defines the resources that a user will have access to and the operations they may perform. For added security, multi-factor authentication is enabled for all users.

All the data stored in the data lake, to which an organisation has access to, can be manually downloaded by authorised users from the application.
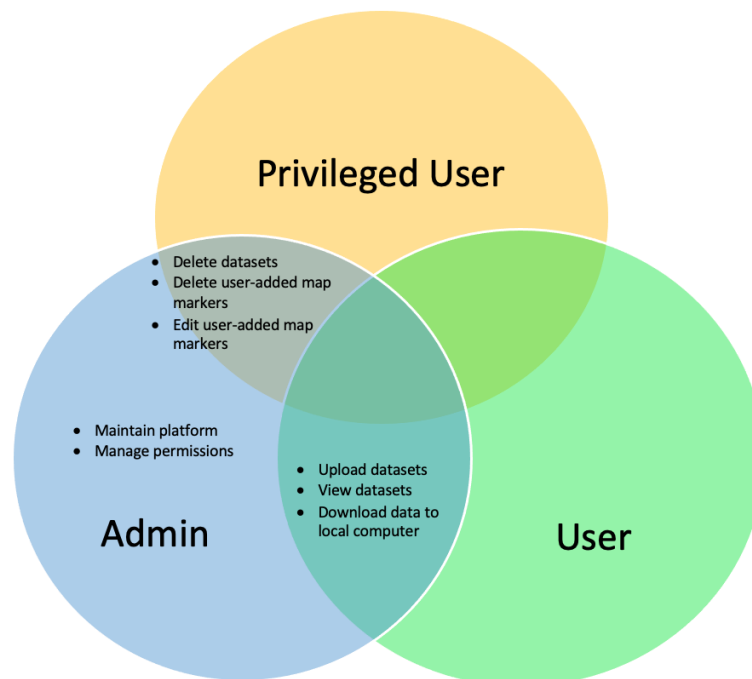

Figure 4. Example of STRIAD user permissions

7 Next Steps

The next steps for Work Package 2 include:

- With the chosen vulnerability model framework, we will continue to implement a co-design approach to identify suitable data sources and indicators that can be mapped to each framework dimension for each of the 15 target countries.
- Present and discuss indicators identified through the projects baseline reports with project partners based on Figure 1.
- Complete the pilot model for Spain and present findings to the project partners for feedback and refinement. This process will be repeated for the remaining target countries with continual review as the COVID-19 pandemic continues.
- Conduct the second 'Evaluation' workshop as described in T2.5 with project partners (SINUS) by year-end 2021.
- Specify the metadata and tags to associate with each dataset for easy retrieval and use.
- Specify organisation data access rules.
- Begin the collection of qualitative data and establishing the procedure for inclusion in the data lake and cataloguing.

8 Conclusion

This report provides an overview of the data collection activities conducted to date in fulfilment of Task 2.1, 'Data Collection and Review'. Data sources have been identified and reviewed with consideration to COVID-19 impacts and responses across the EU27 MS and the UK. It provides an overview of the data collection strategy, identification of indicators of interest and a summary of desk-based research conducted. Through this desk-based research we reviewed numerous datasets with consideration of the countries included, the relevance of the indicators, the quality of the data and the available Application Programming Interface (API), which serves as an access point to the database.

As described, data deemed suitable are stored in a central database for data processing and cleaning (Task 2.2). An overview of published COVID-19 risk assessment models to demonstrate COVID-19 and other vulnerabilities have been presented and based on the model of interest, BRC-VI, an overview of data collection activities for Spain, as the first target country, has been described. These activities will continue for the duration of the project to ensure a co-design approach is implemented in the generation of the COVINFORM risk assessment dashboard. A follow-up report will be presented in M30 of the project.

References

Literature

Alberca, R. W., Oliveira, L. D. M., Branco, A. C. C. C., Pereira, N. Z., & Sato, M. N. (2020). Obesity as a risk factor for COVID-19: an overview. *Critical Reviews in Food Science and Nutrition*, 1-15.

Bragazzi, N. L., Alicino, C., Trucchi, C., Paganino, C., Barberis, I., Martini, M., ... & Orsi, A. (2017). Global reaction to the recent outbreaks of Zika virus: Insights from a Big Data analysis. *PloS one*, 12(9), e0185263.

Cheng, C., Barceló, J., Hartnett, A. S., Kubinec, R., & Messerschmidt, L. (2020). COVID-19 government response event dataset (CoronaNet v. 1.0). *Nature human behaviour*, *4*(7), 756-768.

Cori, A., Donnelly, C. A., Dorigatti, I., Ferguson, N. M., Fraser, C., Garske, T., ... & Blake, I. M. (2017). Key data for outbreak evaluation: building on the Ebola experience. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1721), 20160371.

Cowger, T. L., Davis, B. A., Etkins, O. S., Makofane, K., Lawrence, J. A., Bassett, M. T., & Krieger, N. (2020). Comparison of weighted and unweighted population data to assess inequities in coronavirus disease 2019 deaths by race/ethnicity reported by the US Centers for Disease Control and Prevention. *JAMA network open*, 3(7), e2016933-e2016933.

DeCaprio, D., Gartner, J., Burgess, T., Garcia, K., Kothari, S., Sayed, S., & McCall, C. J. (2020). Building a COVID-19 vulnerability index. *arXiv preprint arXiv:2003.07347*.

Fielding-Miller, R. K., Sundaram, M. E., & Brouwer, K. (2020). Social determinants of COVID-19 mortality at the county level. *PloS one*, 15(10), e0240151.

Goh, H. P., Mahari, W. I., Ahad, N. I., Chaw, L., Kifli, N., Goh, B. H., ... & Ming, L. C. (2020). Risk factors affecting COVID-19 case fatality rate: A quantitative analysis of top 50 affected countries. *medRxiv*.

Karmakar M, Lantz PM, & Tipirneni R. (2021). Association of Social and Demographic Factors With COVID-19 Incidence and Death Rates in the US. *JAMA Netw Open*; 4(1):e2036462. doi:10.1001/jamanetworkopen.2020.36462

Kenyon, C. (2020). Flattening-the-curve associated with reduced COVID-19 case fatality rates-an ecological analysis of 65 countries. *Journal of Infection*, *81*(1), e98-e99.

Lai, C. C., Wang, C. Y., Wang, Y. H., Hsueh, S. C., Ko, W. C., & Hsueh, P. R. (2020). Global epidemiology of coronavirus disease 2019 (COVID-19): disease incidence, daily cumulative index, mortality, and their association with country healthcare resources and economic status. *International journal of antimicrobial agents*, 55(4), 105946. https://doi.org/10.1016/j.ijantimicag.2020.105946

Macharia, P. M., Joseph, N. K., & Okiro, E. A. (2020). A vulnerability index for COVID-19: spatial analysis at the subnational level in Kenya. *BMJ global health*, *5*(8), e003014.

Pana, T.A., Bhattacharya, S., Gamble, D.T., et al (2021). Country-level determinants of the severity of the first global wave of the COVID-19 pandemic: an ecological study. *BMJ Open,* 11:e042034. doi: 10.1136/bmjopen-2020-042041

Porcher, S. (2020). Response2covid19, a dataset of governments' responses to COVID-19 all around the world. *Scientific data*, 7(1), 1-9.

Seligman, B., Ferranna, M., & Bloom, D. E. (2021). Social determinants of mortality from COVID-19: A simulation study using NHANES. *PLoS medicine*, 18(1), e1003490.

Tiwari, A., Dadhania, A. V., Ragunathrao, V. A. B., & Oliveira, E. R. (2021). Using machine learning to develop a novel COVID-19 Vulnerability Index (C19VI). *Science of The Total Environment*, *773*, 145650.

Velasco, J. M., Tseng, W. C., & Chang, C. L. (2021). Factors Affecting the Cases and Deaths of COVID-19 Victims. *International Journal of Environmental Research and Public Health*, 18(2), 674.

Zheng, Z., Peng, F., Xu, B., Zhao, J., Liu, H., Peng, J., ... & Tang, W. (2020). Risk factors of critical & mortal COVID-19 cases: A systematic literature review and meta-analysis. *Journal of infection*, 81(2), e16-e25.