COVINFORM

CORONAVIRUS VULNERABILITIES AND INFORMATION DYNAMICS RESEARCH AND MODELLING

D2.3 Technical Report

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 101016247.

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 Nationalặ 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

Acknowledgement: This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 101016247.

Deliverable

Number	**D2.3**
Title	**Technical Report**
Lead beneficiary	TRI
Work package	WP2
Dissemination level	Public (PU)
Nature	Report (RE)
Due date	30.06.2022
Submission date	30.06.2022
Authors	**Ilaria Bonavita,** TRI **Niamh Aspell,** TRI **Emily Maitland,** TRI **Ejay Nsugbe,** TRI **Nanditha Plakazhi,** TRI **Kush Wadhwa,** TRI
Reviewers	**Su Anson**, TRI **José M. Palma**, Factor Social **Igor Linkov**, Factor Social

Document history

Version	Date	Comments
0.1	06.05.2022	ToC draft
0.2	10.06.2022	Version 0.2 for internal TRI review
0.3	13.06.2022	Version 0.3 for updates
0.4	14.06.2022	Version 0.4 updated and sent for internal project review
0.5	28.06.2022	Version 0.5 updated after internal project review
0.6	28.06.2022	Version 0.6 updated after internal TRI review
0.7	29.06.2022	Final draft version for final coordinator review
1.0	30.06.2022	Final version submitted

Executive Summary

The COVINFORM project investigates the impact of the COVID-19 pandemic across the 27 EU member states (MS) and the UK. National, regional, and local responses varied in success due to the changeable nature of the public health crisis and the level of preparedness and adaptability of governing institutions. Whilst the general population in the EU and UK experienced disruptive and life changing consequences, the impacts were not equally experienced, and notable differences were evidenced in certain social clusters. In the COVINFORM project we pay particular focus on those vulnerable and marginalised groups. One of the projects outputs includes an interactive risk assessment dashboard that centres on various types of vulnerability (i.e., physical, social, economic and information) and the related consequences, within the context of COVID-19. Extensive work was undertaken to appropriately define the risk assessment framework, this will ultimately provide a risk score for each country and will be visualised on the geospatial dashboard. This included engagement with all project stakeholders to inform the various types of indicators that are best used to quantitatively represent key domains, such as social vulnerability, within the context of COVID-19. This was after a review of the literature and finally the identification of such data routinely collected in the EU and UK. Following data collection, the secondary data requires pre-processing for in-depth modelling that will provide insights for various stakeholders (e.g., policy makers, first responders and researchers). For example, the contribution of certain features of vulnerability which are highly correlated to the consequences of COVID-19, such as poorer physical health or dependency on social supports. These insights will offer stakeholders the opportunity to explore their assumptions of risk, particularly for vulnerable groups, during public health crisis and ultimately inform their future responses.

This report provides an overview of the technical work undertaken in the development of the risk assessment model and the interactive dashboard. It includes an update on activities relating to the collection and processing of national level quantitative data across the EU27 MS and the UK. Various qualitative data will also be integrated into the dashboard to allow for both consideration of the impacts and response of COVID-19 on vulnerable groups as defined by the COVINFORM case studies. A summary of identified data sources and indicators to be integrated into the risk assessment model under the domains of threat, vulnerability, consequences, and resilience is also included in this report. In addition, we describe the process followed to download, clean, and reformat data for ingestion into our data model and in preparation for storage in a database. This includes standardising dates and region names as well as removing unexpected characters from numerical data and discarding redundant or unnecessary data. We also describe the risk framework and how this is mapped to our modelling procedure, as well as provide technical details of this modelling and how we compute preliminary scores from the data we have collected so far. A preview of these results is given alongside some commentary. Finally, we describe the software used to carry out this work, as well as the plans for the continuation of this data collection and data analysis work and development of the dashboard tool.

Contents

Figures

Tables

Terminology

Term	Description
Data frame	Data structure that organises data into a 2D arrangement of named rows and columns, like a spreadsheet.
Data lake	A centralised repository that hosts multiple data types in both structured and unstructured formats.
Data pipeline	A set of processes and actions that converts raw data into interpretable knowledge or results.
Feature	A measurable piece of data used for analysis or computation.
Indicator	An indicator is the representation of data for a specified time, place or any other relevant characteristic, corrected for at least one dimension (usually size) so as to allow for meaningful comparisons.[1]
Values	Individual data, numbers or labels that make up features.

[1] https://ec.europa.eu/eurostat/statistics-explained/

Acronyms & Abbreviations

Term	Description
API	Application Programming Interface
CSV	Comma separated values (a text file for storing tabular data, delineated by commas)
ECDC	European Center for Disease prevention and Control
ECAD	European Climate Assessment & Dataset
EEA	European Environment Agency
TSV	Tab separated values (a text file for storing tabular data, delineated by tabs)
MS	Member State
TVCR	Threat, vulnerability, Consequence, Resilience
WHO	World Health Organization

1 Introduction

This deliverable describes the technical progress in relation to data collection and processing, model development and performance, and dashboard development as part of WP2 'Risk assessment model to evaluate the response and impact at different geographical levels'. It briefly provides an update of data collection activities and the selection of relevant indicators to inform the risk assessment framework. As described in *D2.1 Database containing different data sources*[2], we reviewed numerous datasets and risk assessment frameworks and finalised the first iteration of the risk assessment methodology which is described in Section 4 of this report. This framework includes COVID-19 related data (i.e., public health surveillance data such as COVID-19 case and deaths) and data describing the wider risk environment (e.g., hospital resources and poverty). These indicators were selected based on literature review (D2.1.) and stakeholder consultation, for quantification of vulnerabilities that were pre-existing or as a direct or indirect result of COVID-19. Due to the evolving nature of the pandemic, indicators to reflect the degree to which a country was able to recover and adapt as the crisis changed (i.e., during and after waves of high transmission) were also identified. The identified indicators can then be mapped to the responses and impacts of the coronavirus pandemic at the national level across the 27 EU member states (MS) and the UK. Extending on this work, mapping activities for qualitative data sources (e.g., interview transcripts and thematic results) at local levels in the ten case study sites will add a layer of contextual validation to the risk assessment model.

The project will conclude in the development of an interactive dashboard and visual toolkit for stakeholders in government, public health, and civil society organisation. The dashboard will integrate 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 risk modelling and performance, including generation of the features that are fed to the model.

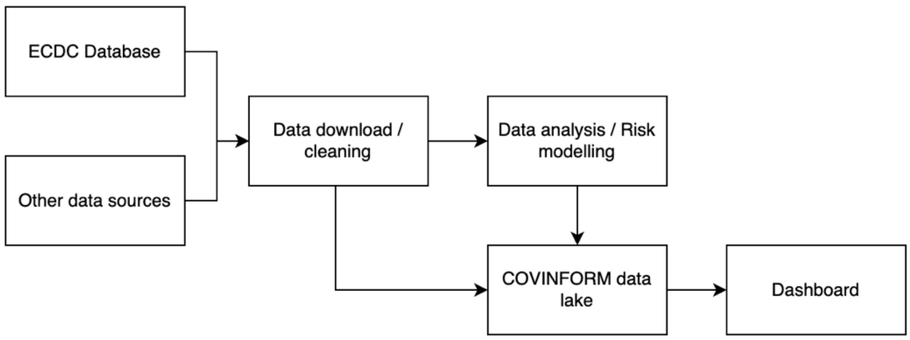

Figure 1. COVINFORM Risk Assessment dashboard pipeline

A flow diagram showing the full process, from data collection to visualisation in the dashboard is pictured above (Figure 1). Data for the indicators is collected from both the ECDC database and other data sources (where necessary), downloaded, and cleaned to be stored in the COVINFORM data lake (see Appendix 1). Analysis of the data and computation of the risk score is performed on the data and

[2] Available online at https://www.covinform.eu/project-outputs/technical-reports/

these results are also stored in the data lake alongside the indicator data. The dashboard will then retrieve these data from the data lake for visualisation and interactive exploration by the user.

It is important to note that the features used for the risk modelling are *quantitative;* however, the dashboard will be also used to visualise and allow interaction with *qualitative features* coming from data which are not included in the risk model but have been collected or produced during the project and can provide additional contextual information to the risk score. This additional data, which will include, among others, interview transcripts, deliverables, documents related with case studies, will effectively constitute the COVINFORM knowledge repository (WP8, Task 8.6). The dashboard will offer a user-facing interface to this repository. However, this report focuses on the progress in the COVINFORM risk modelling; therefore, only data useful for the modelling approach will be discussed. The approach taken for the COVINFORM Knowledge Repository will be further outlined in D8.7, 'COVINFORM COVID-19 Knowledge Repository' due M36.

In Section 2, the updated data identification and collection strategy is described, covering the data source search strategy and eligibility criteria. This is followed by an overview of the risk assessment methodology in Section 3 and how it relates to the ongoing activities in the case study research design (WP3) in Section 4.

Section 5 describes the pipeline followed to clean and pre-process the data and Section 7 describes the adopted risk modelling approach alongside its internal architecture as well as preliminary results and findings for the sub-criteria investigated.

Section 7 provides details of the software used as part of the data modelling process

The report concludes with an overview of the next steps in data collection activities, model development, dashboard development, integration of qualitative data from case study research and the evaluation process.

2 Data collection

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, government responses and impact, public health responses and impact, citizen and community responses and inclusive COVID-19 communication. These can be described by multiple relevant indicators as outlined in Table 1 below and has been described in more detail in D2.1.

2.1 Updated data source search strategy

Data collection activities continue to focus on quantitative structured datasets (datasets in tabular form where values of each variable are numbers or categories). As described above, the indicators identified to describe various forms of vulnerability within a population and the built environment were identified through literature review as part of D2.1 and presented at various stakeholders' consultations, as follows:

- COVINFORM Virtual Workshop 1: Practitioner user requirements 25th May 2021, (Task 2.5, Evaluation)
- COVINFORM Case Study Meeting: Vienna 2-3rd March 2022
- COVINFORM Consortium Meeting: Lisbon May 11-13th 2022

- Re-occurring weekly case study meetings with Factor Social and case study leads commencing Jan 4th, 2022 - ongoing

Under guidance of risk assessment experts of partner organisation, Factor Social, and feedback from the ten-case study leads, pertinent to their case study populations, the 'baseline' framework was finalised in Spring 2022 (Table 1 and outlined below). The models indicators will be tested for their contribution to the vulnerability index using statistical techniques described in Section 6. Further to this, the indicators will be validated against the findings from the case study research. Therefore, the current framework is deemed a 'baseline' model as its attributes may alter based on best fit validations as the project continues to gather findings.

Following the finalisation of the COVINFORM Risk Assessment Framework, extensive data collection activities were required to obtain the appropriate data for each EU MS and the UK. The relationship between WP2 and WP3 Case Study Research, are described in Section 4 of this report and in D3.4. 'Comparative Analysis'.[3] The data acquired from the local level case study research will be interpreted, processed, and ingested into the dashboard as the qualitative data becomes available (from autumn 2022).

Based on the projects developing requirements, the eligibility criteria have been updated since D2.1 [M10]. The inclusion criteria are as follows:

1) As previously reported, **national** level across the EU27 MS and the UK.
2) Modified, **regional** in a single MS. Following extensive data review, it became evident that due to varying data collection practices or incomplete/absent data *within* EU member states the current application of a tailored regional risk modelling would be inappropriate, in terms of efficiencies, and comparability and usability. To demonstrate the application, challenges, and usability of the framework at regional level, for tailored sub-national a-priori risk assessment, a single MS will be nominated for development and testing. This will be based on the MS having, within reason, harmonised data collection practices across regions and open data access policies. The level of granularity and coverage of indicators will be considered and reported in D2.8, 'Database containing different data sources - update M30'.
3) As previously reported**, local levels** pertinent to the case study sites (more details are provided in section 3)

Open-source data continues to be the primary data source to be used (See Appendix 1. Risk Assessment Framework Indicator Description and Source). As the WP activities continue, closed access data may also be 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, as mentioned in D2.1.

[3] Available at https://www.covinform.eu/project-outputs/technical-reports/

	OBJECT AT RISK (HEALTH)												
	THREAT (RISK OBJECT)			VULNERABILITIES				CONSEQUENCES (IMPACTS)				RESILIENCE	
Time period	**Virus (cases)**	**Variant of concern** [1]	**Likelihood of dev.**	**Physical**	**Social**	**Economic**	**Information**	**Health**	**Social**	**Economic**	**Environmental**	**Ability to recover**	**Ability to Adapt**
Baseline (2019)	-	-	Mobility, international trade, migration, housing concentration, pollution, temperature, age of population	No. hospitals, ICU beds, No. of LTCF beds, medical frontline staff, pre-existing health conditions, ports and airports	Education level, rural vs urban, gender - % female, migrant population	GDP, % living in poverty income inequality, unemployment rates	literacy, digital access, digital skills	covid-related death rate, excess deaths (% change in weekly mortality compared to average mortality), hospital admissions, ICU admissions	food insecurity, loss of education, rates of violence	job losses, reduced income, % receiving social support, disruption to supply chains	exposure to outdoor air pollution, GHG emissions	Spread of economic activity, emergency investment in healthcare, investment in vaccines healthcare workforce debt by sector, investment, international support, income support, vaccine financial support	rebuilding vulnerable industries, digitisation, innovation and fiscal measures
First Phase (2020)	Weekly case rate, positivity (%), ratio test/cases, no. of detections of variants, % variant	Beta (sept 2020)	Mobility, international trade, migration, housing concentration, pollution, temperature, age of population	No. hospitals, ICU beds, No. of LTCF beds, medical frontline staff, weekly testing volume, pre-existing health conditions, ports and airports	Education level, rural vs urban, gender - % female, migrant population	GDP, % living in poverty income inequality, unemployment rates	literacy, digital access, digital skills	covid-related death rate, excess deaths (% change in weekly mortality compared to average mortality), hospital admissions, ICU admissions	food insecurity, loss of education, rates of violence	job losses, reduced income, % receiving social support, disruption to supply chains	exposure to outdoor air pollution, GHG emissions	Spread of economic activity, emergency investment in healthcare, investment in vaccines healthcare workforce debt by sector, investment, international support, income support, vaccine financial support	rebuilding vulnerable industries, digitisation, innovation and fiscal measures
Post-vaccine (2021)	As above, no. of detections of variants, % variant	Gamma, Delta, Omicron	As above	As above, + No. of vaccination centres	As above	As above, equity of vaccination programmes	As above	As above	As above	As above	As above	As above +, % first dose and fully vaccinated, and first booster. number of vaccine doses distributed by manufacturers	As above

Table 1. COVINFORM Risk Assessment Framework

2.2 Data Quality

Geographic coverage of most indicators is good at the national level. Across all the indicators collected so far, 25% of the data is missing over the duration of its entire reporting period, mostly owing to data not being reported or unavailable rather than data quality problems. It is anticipated that the rate of missing data will be significantly reduced for the time periods of interest (2019-2022), as data collection practices improved across MS and within the national statistics reporting frameworks. Only indicators pertaining to demographic data such as age and gender are disaggregated at subnational levels so far. At the national level, the UK has a large fraction of missing data, most likely because of the changing relationship between the UK and the other 27 MS; however, this mostly affects indicators under the physical sub-criteria, data under other sub-criteria of the framework are not missing.

Coverage across time varies (see Appendix 1). For indicators where there is data across multiple dates recorded (e.g., migration flows or proportion of people living in poverty) and the majority of these are recorded on a yearly basis, rather than weekly or monthly. Data pertaining to public health such as number of cases, deaths and testing rates are recorded on a daily and/or weekly basis, whereas indicators relating to infrastructure and environment are recorded on a yearly basis. Based on the risk assessment framework, and the importance of considering the evolution of the pandemic, data will be representable from 2019-2021 and the frequency of the data will be dependent on the indicator in question.

2.3 Data sources

The specific indicators and their data source are listed in Appendix 1. The predominant sources of data utilised up to this moment are:

- European Climate Assessment & Dataset (https://www.ecad.eu/)
- European Centre for Disease Prevention and Control (https://www.ecdc.europa.eu/en)
- European Sea Ports Organisation (https://www.espo.be/)
- Eurostat (https://ec.europa.eu/eurostat)
- Google Mobility data (https://www.google.com/covid19/mobility/)
- WHO - Global Health Observatory (https://www.who.int/data/gho)

3 Risk Framework Methodology

As an update to D2.1 and to allow for understanding on the data currently being processed and modelled, we provide a brief update on the risk framework methodology. Risk management during crisis situations, such as COVID-19, is challenging. In Europe, we observed significant divergence in (viral) threat levels and the subsequent responses and impacts due to the constantly changing nature of the pandemic (Wolff, & Ladi, 2020; Żak, & Garncarz, 2020). The distribution of the impacts, in both time and space, and the response of governments, public health and local communities were experienced differently across populations based on their unique attributes. For example, their level of physical health and wellbeing, economic stability, social inclusion and environmental conditions. As described in D3.2 'Case Study Research Methods' and D3.4. 'Case Study Comparative Analysis'[4], the complex and dynamic nature was a result of multiple bi-directional relations between these systems and their interdependencies. This

[4] Available at https://www.covinform.eu/project-outputs/technical-reports/

coupled with the relatively uncertain trajectory of the virus, put stress on the systems to recover and adapt overtime. The risk assessment model developed by the COVINFORM project is not intended to predict future waves, or to forecast the end of the pandemic, neither predict an emerging pandemic. Rather, its aim is to allow for evaluation a-priori of the COVID-19 pandemic, to test hypothesis, explore assumptions, offer unique community-centred insights[5], coupled with national level analytics to strategize policy development for potential future waves or other health crisis situations.

Traditionally, risk assessment methodologies have focused on analysing the given threat while planning and reducing the vulnerabilities of a given system. Traditional risk assessment methodologies categorise risk into threats, vulnerabilities, and consequences (TVC). To properly determine the consequences of a given system's disruptions, we must also determine the system's resilience when exposed to shocks or stressors. Policy decisions adopted over the course of the COVID-19 pandemic influenced the capacity for a country to recover, with some taking the disruptions of COVID-19 to adapt, while others, did so to a lesser degree (Roloff, 2020). Building resilience (i.e., ability to respond, recover and adapt to disruption, Linkov & Trump, 2019) is of crucial importance in overcoming COVID-19 and future pandemics.

If risk is defined as a product of TVC and used to prevent degradation of system functions, resilience is defined as ability of the system to recover and adapt following such disruptions (Galaitisi et al., 2020; Linkov & Trump, 2019). The risk imposed by COVID-19 in all EU MS and the UK in March 2020 varied depending on the characteristics of the system and were transformed from uncoordinated and disorganised policy decisions to several broadly implementable measures demonstrating varying degrees of adaptability to the crisis[6]. Governments across Europe and the UK experienced initial challenges in supporting health and economic sectors but quickly altered common practice with various emergency measures, including the limitations on freedom of movement, physical distancing, and wage subsidies, to name a few[7]. Whilst vaccine development and procurement, and subsequent population level immunisation marked a significant turning point, the efficacy of vaccination with the detection of new, more transmissible variants, presented one of many uncertainties (Andrews et al., 2022). To that extent, a risk assessment methodology for the COVID-19 pandemic requires delineating by TVC and Resilience (TVCR) over time, which we have divided into three periods: pre-covid (2019), initial response and pre-vaccine (2020) and post vaccine (2021). As WHO are yet to declare the end to the emergency, there is potential that new events or waves could be experienced. However, given the time limitation of the COVINFORM project, our focus will be to demonstrate the methodology and its application from 2019-2021.

As reported in D2.1 'Database containing different data sources', several relevant indicators were identified through literature review, evaluation of various COVID-19 dashboards and both related and unrelated risk assessment models. Central to the objectives of the COVINFORM project, the response and impacts on vulnerable and marginalised groups have been prioritised and will be validated against the WP3 case study findings. In our risk model we define threat as the risk object, "the virus, or variant of, and the level of transmission due to the built and natural environment" (e.g., pollution, temperature, and housing concentration). The leading 'object at risk' of any health crisis is population

[5] Community' as defined by the COVINFORM case studies 'communities of practice' or 'characteristic', not geographic (e.g., migrant frontline workers).

[6] https://blogs.lse.ac.uk/europpblog/2020/03/26/assessing-the-european-unions-performance-in-the-covid-19-pandemic/. Accessed on 03 June 2022.

[7] https://ec.europa.eu/commission/presscorner/detail/it/statement_20_567. Accessed on 03 June 2022.

health. As observed, the risk exposure was disproportionally experienced and reports of 'clustering' within certain groups were reported widely from the early stages (e.g., BAME, low socioeconomic status, females, migrants, healthcare workers). To ensure inclusive risk modelling, factoring the level of vulnerability of a country and its population, a vulnerability index including factors linked to physical, social, economic and information vulnerabilities and their consequences (impacts) are central to the risk framework. Resilience is defined by the COVINFORM project as a countries ability to recover (e.g., economic activity) and the ability to adapt (e.g., fiscal measures such as income support) overtime.

4 WP2 relation to Case Study Research (WP3)

The quantitative methods applied in the risk assessment model allows for generalisability; however, it may omit the complete picture of risk for marginalised or vulnerable sub-groups in the population. The findings of the ten COVINFORM case studies (WP3) are due in October 2022 [M24] and August 2023 [M34] of the project. The findings will provide an opportunity to test various assumptions of the risk framework. This validation process offers an extra layer of contextual insights which are commonly overlooked or reduced by quantitative methods alone. This will allow for deeper underlying meanings and explanations of the risk profiles during COVID-19. For example, the risk framework may indicate that women are more vulnerable to the consequences of COVID-19, but it does not explain the reasons of the effect and their meanings in certain contexts. 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 2023].

To map the case study attributes to the framework, data was collected from each case study lead to determine 1) the characteristics of their study population and, 2) the factors which make these groups more vulnerable to COVID-19. Several comparable characteristics and vulnerabilities were identified across the case studies allowing for mapping between our risk framework and the case studies (Figure.2)[8].

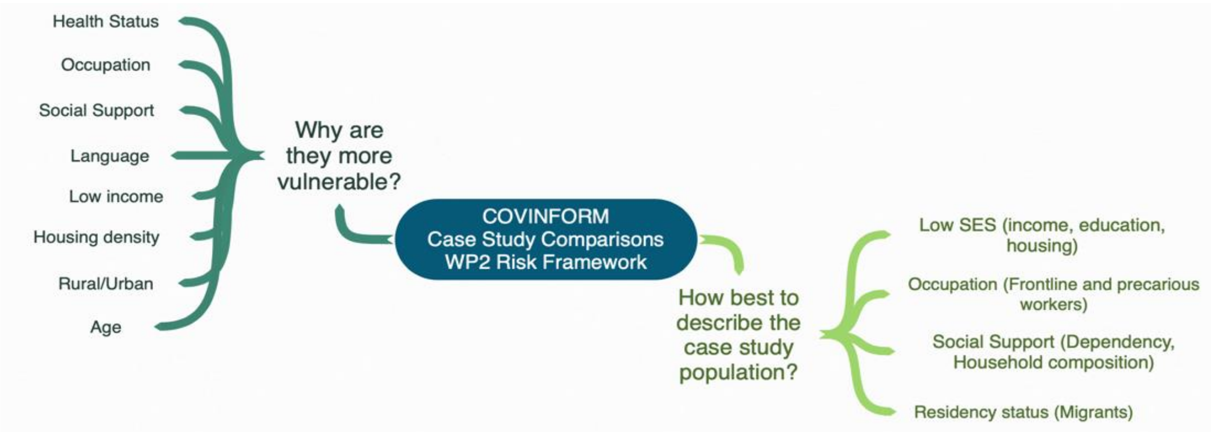

Figure 2. WP3 Case Study Populations and WP2 Risk Framework Indicators and Features of comparison

[8] As per the qualitative nature of the case study design, the final sample sizes are not intended for statistically powered analytical testing.

5 Data cleaning and processing

Since most of the data is sourced from the Eurostat database, the process for collecting and cleaning the data follows a similar procedure of downloading, cleaning, reshaping and then saving the data. So far, we have produced 24 cleaned data files corresponding to 24 indicators in our framework that cover data at the national level. Each file contains yearly, weekly, or daily recordings of the corresponding indicators across the EU27+UK MS. Tables of the data coverage for collected indicators is shown below in Figures 3 and 4. These tables shows how much data is present from each file, allowing us to quickly identify gaps in data collection or systematic issues with the data source, indicating a need to identify alternative data sources. Figure 3 shows the UK is not represented in the ECDC data sources relating to COVID-19 cases, testing, or variant data, which will require us to find an alternative source. We can also see which indicators have poorer coverage and again, which countries or regions are missing data, and which indicators they are missing. Figure 4 shows that some of the indicators under the 'vulnerability', namely 'ICU beds', 'Frontline medical staff' and 'Nursing staff' heading have much poorer coverage than is typical, owing to large gaps in the data for 2020, as well as all years prior to 2008.

	Virus (cases)			Variant of concern				Likelihood of dev.						
Country	Weekly case rat	Weekly Positivity Rate (%)	Weekly Testing	varian	number_detections_variant 2	valid_denominato	percent_variant	Mobility inde	international trad	migration	housing concentration	pollution	temperature	age of population
Austria	92.73183	87.5	87.5	100	100	100	86.696901	100	100	100	100	100		100
Belgium	92.73183	92.85714286	92.857143	100	100	100	89.806401	100	100	90.9091	100	100		85.71429
Bulgaria	92.73183	80.35714286	80.357143	100	100	98.882236	73.652695	100	100	40.9091	100	100		100
Croatia	92.73183	96.42857143	96.428571	100	100	100	66.613098	100	100	90.9091	91.6666667	100		100
Republic of Cyprus	92.73183	91.07142857	91.071429	100	100	100	43.368592	0	100	100	100	100		100
Czech Republic	92.73183	96.42857143	96.428571	100	100	100	87.845734	100	100	86.3636	100	100		100
Denmark	92.73183	96.42857143	96.428571	100	100	100	89.036679	100	100	100	100	100		100
Estonia	92.73183	94.64285714	94.642857	100	100	99.728261	61.956522	100	100	90.9091	100	100		85.71429
Finland	92.73183	100	100	100	100	100	83.359014	100	100	100	100	100		100
France	92.73183	92.85714286	92.857143	100	100	100	92.733434	100	100	63.6364	100	100		100
Germany	92.73183	91.96428571	91.964286	100	100	100	92.598304	100	100	100	100	100		100
Greece	92.73183	92.85714286	92.857143	100	100	100	78.106739	100	100	100	100	100		100
Hungary	92.73183	82.14285714	82.142857	100	100	100	37.03247	100	100	100	100	100		100
Ireland	92.73183	96.42857143	96.428571	100	100	100	87.953668	100	100	100	100	100		85.71429
Italy	92.73183	79.46428571	79.464286	100	100	100	93.282015	100	100	100	91.6666667	100		85.71429
Latvia	92.73183	86.60714286	86.607143	100	100	100	75.287356	100	100	100	100	100		100
Lithuania	92.73183	91.96428571	91.964286	100	100	100	74.889691	100	100	100	100	100		100
Luxembourg	92.73183	92.85714286	92.857143	100	100	100	86.168959	100	100	100	100	100		100
Malta	92.73183	94.64285714	94.642857	100	100	100	31.275386	100	100	68.1818	100	100		100
Netherlands	92.73183	91.07142857	91.071429	100	100	100	90.962213	100	100	100	100	100		100
Poland	92.73183	91.96428571	91.964286	100	100	100	80.622292	100	100	100	100	100		100
Portugal	92.73183	92.85714286	92.857143	100	100	100	79.323747	100	100	100	100	100		100
Romania	92.73183	91.07142857	91.071429	100	100	97.328244	82.519084	100	100	54.5455	100	100		100
Slovakia	92.73183	95.53571429	95.535714	100	100	100	74.013551	100	100	100	100	100		100
Slovenia	92.73183	92.85714286	92.857143	100	100	100	84.38247	100	100	100	100	100		100
Spain	92.73183	83.92857143	83.928571	100	100	100	90.006868	100	100	100	100	100		100
Sweden	92.73183	88.39285714	88.392857	100	100	100	90.990291	100	100	100	100	100		100
UK	0	0	0	0	0	0	0	100	0	95.4545	83.3333333	100		100

Figure 3. Table showing data coverage of indicators under the THREAT heading for different regions.

	Physical								Social				Economic				Information		
Country	ICU	med	nurs	weekly	ratio test/case	pre-existing health	sea passenge	No. airp	Education lev	rural vs urba	gender - % female	migrant population	GDP	% living in pove	income inequality	unemployment rates	language	digital literac	digital acce
Austria	93.1308	100	48.7805				0	100		100	100	100	100	100	100	68.4211		100	100
Belgium	95.3765	95.082	2.43902				100	100		100	100	100	100	100	100	100		100	85
Bulgaria	92.4703	80.3279	14.6341				83.3333333	100		100	100	100	100	80	100	55.2632		100	85
Croatia	0	65.5738	48.7805				87.5	78.9474		100	100	100	100	53.33333333	91.6667	50		100	75
Republic of Cyprus	96.1691	62.2951	0				79.1666667	100		100	100	100	100	86.66666667	100	55.2632		100	100
Czech Republic	97.2259	70.4918	26.8293				0	100		100	100	100	100	86.66666667	100	63.1579		100	95
Denmark	91.6777	62.2951	65.8537				100	100		100	100	100	100	100	100	100		100	100
Estonia	97.8864	91.8033	53.6585				83.3333333	100		100	100	100	100	93.33333333	100	63.1579		100	90
Finland	71.2021	22.9508	29.2683				100	100		100	100	100	100	93.33333333	100	68.4211		100	100
France	98.679	13.1148	0				100	100		100	100	100	100	93.33333333	100	100		100	90
Germany	94.716	49.1803	24.3902				87.5	100		100	100	100	100	86.66666667	100	100		100	100
Greece	100	0	0				100	100		100	100	100	100	100	100	100		100	100
Hungary	0	85.2459	48.7805				100	100		100	100	100	100	86.66666667	100	65.7895		100	90
Ireland	93.6592	9.83607	0				100	100		100	100	100	100	100	100	100		100	95
Italy	0	13.1148	17.0732				100	100		100	100	100	100	93.33333333	100	100		100	100
Latvia	100	45.9016	48.7805				83.3333333	100		100	100	100	100	86.66666667	100	60.5263		100	95
Lithuania	0	45.9016	48.7805				83.3333333	100		100	100	100	100	86.66666667	100	60.5263		100	100
Luxembourg	98.0185	91.8033	21.9512				0	100		100	100	100	100	100	100	100		100	100
Malta	87.8468	16.3934	17.0732				79.1666667	100		100	100	100	100	86.66666667	100	55.2632		100	85
Netherlands	97.6222	8.19672	31.7073				100	100		100	100	100	100	86.66666667	100	94.7368		100	100
Poland	0	90.1639	4.87805				83.3333333	100		100	100	100	100	86.66666667	100	63.1579		100	100
Portugal	97.6222	0	0				100	100		100	100	100	100	93.33333333	100	92.1053		100	100
Romania	49.0092	32.7869	31.7073				83.3333333	100		100	100	100	90.90909091	73.33333333	100	63.1579		100	85
Slovakia	95.9049	8.19672	0				0	100		100	100	100	100	86.66666667	100	60.5263		100	90
Slovenia	96.037	36.0656	41.4634				83.3333333	100		100	100	100	100	86.66666667	100	65.7895		100	90
Spain	50.9908	44.2623	0				100	100		100	100	100	100	93.33333333	100	92.1053		100	95
Sweden	97.6222	39.3443	0				100	100		100	100	100	100	93.33333333	100	68.4211		100	85
UK	0	88.5246	17.0732				95.8333333	100		100	100	85.7143	90.90909091	0	83.3333	97.3684		0	95

Figure 4. Table showing data coverage of indicators under the VULNERABILITIES heading for different regions.

The data are collected and processed according to the following procedure. Data is downloaded from the Eurostat database through a custom API that uses the dataset code to retrieve and read the corresponding data file as a web response. This is then written as an archived file and unarchived locally to produce a file of tab separated values (TSV). These TSV contain the numerical data. Colons are use as placeholders for missing or unavailable data. Special characters and alphabetic characters are used as flags to indicate data that has been estimated, forecast, or used as a provisional figure. Indicators in each data file are usually given codes, and sometimes disaggregated across age, gender, and other demographic subdivisions. Country names are not always given in full and are often supplied as two or three letter ISO 3166-1 codes, which can be converted into full names. Finally, dates can be supplied in a 'Year-week' format, for instance "2021-W35" indicating the 35th week of 2021, the week from August 30th to September 5th.

To clean and standardise the data for ingestion, we first remove flags and codes from cells containing numerical data which can be achieved by using regular expressions to capture only numeric characters and discard other characters. Country names are being standardised by using a lookup table to replace the country codes. For data that is associated with named geo-spatial administrative regions (nations, counties, districts etc) we ensure that the location names are first standardised (removing special characters and updating region names that may have changed). In cases where geo-spatial regions do not correspond to standard administrative or political boundaries, we can include custom polygon data so that the corresponding region can be drawn on a map. Next, we discard all the data in the file that does not correspond to one of the indicators we have defined, as well as making sure we have access to totals across demographic subdivisions. That is, if we have an indicator disaggregated by gender but no total, we sum across the gender categories to produce a total. We ensure that dates are specified as a year, or as year-month-day. When only a year and week are specified, we choose the first day of that week so that each date has a day, month, and year. Finally, we reshape the data frame so that countries are listed as columns, dates as rows, and the values are the indicator values. These data frames are then saved in a file of comma separated value (CSV).

For the air pollution and mobility data files that are not sourced from Eurostat, a similar process is carried out with affordances made for the idiosyncrasies of the data source, with the intention of achieving the same result of a cleaned and formatted CSV files. For mobility, we use Google's COVID-19 community mobility reports. Google uses data from users to record the average number of visitors and time-spent across different sectors. The change from 'baseline' levels (i.e., pre-pandemic) were recorded. Downloading just the relevant data, we have 28 CSV files corresponding to each EU27+UK MS, each containing daily records of the percentage change from baseline in movement across different six different sectors - retail, transit, grocery and pharmacy, workplace, residential and parks. This data covers three years, 2020-2022, with the first 2 months of 2020 deemed as the 'baseline'.

Once the datafiles have been processed, they can be stored in the COVINFORM database, and then retrieved for data modelling and visualisation.

6 Data modelling

6.1 The weighted naïve risk scoring system

The concept of risk scoring typically provides a score-based method towards the assignment of values which reflect the degree of risk classification of a particular group or entity. In societal endeavours, politics and economics, the risk scoring of a particular group can help provide insights on the nature of the group itself alongside their susceptibility to external circumstance (Robnik-Sikonja et al., 2003). In the context of the COVINFORM project, the risk scoring system serves as a data driven auxiliary source of information poised towards helping inform decision making and policy formation. The risk model devised as part of this report is an empirical risk scoring system which is based on a series of heuristics and is referred to as a *weighted "naïve" risk scoring system* due to a combination of the transparency afforded as part of the model in addition to some of the assumptions made as part of the heuristic scoring system as will be discussed subsequently.

Building on the risk framework defined in Ganin et al. (2020), we built a risk scoring model where scores are hierarchically computed (Figure 5). We define the risk score as the weighted sum of *criteria* (threat, vulnerability, consequences, and resilience) scores.

$$\text{Risk score} = w_1 * \text{Threat score} + w_2 * \text{Vulnerability score} + w_3 * \text{Consequences score} + w_4 * \text{Resilience score}$$

Where w_i (for i=1,..,4) is the weight of each criteria. The criteria scores are in turn obtained by a weighted average of the sub-criteria of the criteria; for instance, for the vulnerability criteria the score is computed as follow:

$$\text{Vulnerability score} = z_1 * \text{Physical score} + z_2 * \text{Social score} + z_3 * \text{Economic score} + z_4 * \text{Information score}$$

Where z_i (for i=1,..,4) is a weight for each sub-criteria. Finally, the sub-criteria score is obtained by combination of the single features composing the sub-criteria using a heuristic and feature ranking process that is detailed later in this section.

Note that the weights w_i and z_i are by default equal but can be adjusted to reflect the importance of their associated (sub-)criteria in the formula. The weights are defined in such a way that the final risk score as well as the criteria and sub-criteria scores take continuous values between 0 and 3. The value can then be grouped into three categories indicating low (1), medium (2) and high (3) score (the higher the score, the higher the risk/vulnerability/threat etc. of a country).

Note also that scores are computed separately for each country and for each time period in our framework, therefore a risk score for baseline, first-phase and post-vaccine periods will be obtained, allowing for drawing insights on the change of the risk score over time.

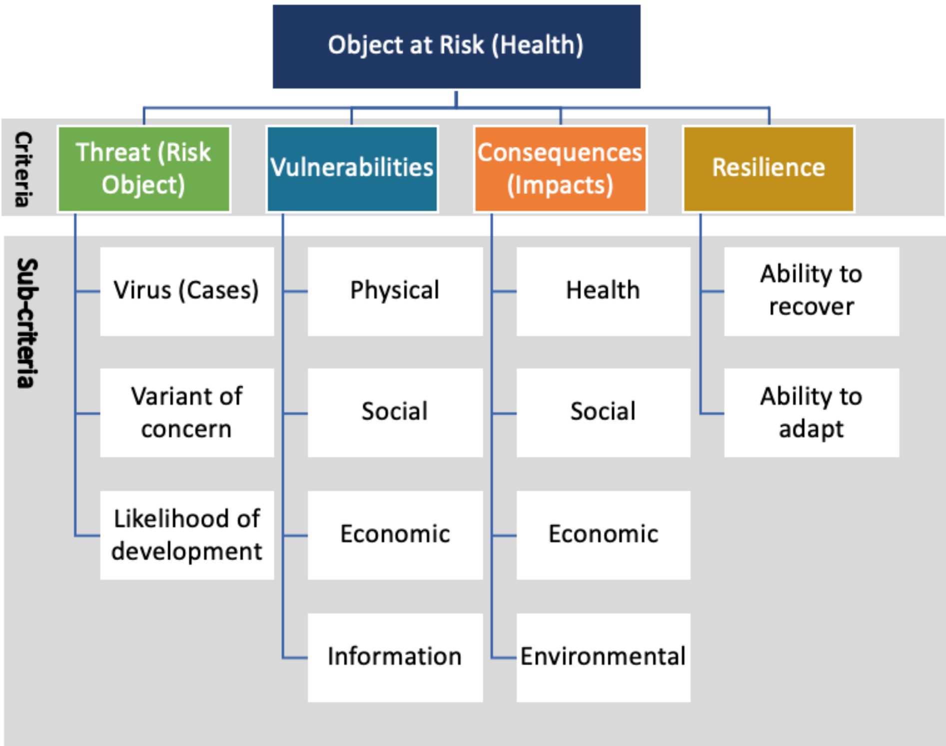

Figure 5. Hierarchic representation of the risk framework. Note that each Sub-criteria can comprise one or more features, as indicated in Table 1.

6.2 Sub-criteria scoring method

The final risk score is built from the bottom to the top of the hierarchy where at the bottom lay the more granular information available for a country (i.e., the features extracted from the gathered data) that are then aggregated to represent higher-level information enclosed by sub-criteria and criteria (i.e., the headings and subheadings of the risk framework described in Table 1). Therefore, the first and core step of the risk modelling involves processing of the features to obtain a sub-criteria score. We describe here the method used to obtain weights and combine features that made up each sub-criterion. As it is believed that all indicators contribute unevenly towards the risk scoring, a feature ranking process is carried out to effectively evaluate and assign weightings to the various features underlying each sub-criterion. The feature weighting exercise was done using the ReliefF algorithm, belonging to the class of ReliefF algorithms which are a statistically driven means towards the ranking of features (Robnik-Sikonja et al., 2003). An interesting characteristic of Relief algorithms is that they do not assume independence among the features, so they are suitable for a problem with strong dependencies between features, like ours. The algorithm provides scores where the higher the magnitude of the assigned value, the greater the feature weight, where feature weights which are assigned a negative score are deemed to be negatively impacting the prediction model and are therein discarded where necessary.

Problem formulation: Given a set of instances (in our case, a set of countries), a set of features (in our case, the features composing a sub-criteria) and a label for each instance (in our case, a sub-criteria score for that country) we want to assess the weight that each feature has in determining the label (in our case, we want to assess the extent with which each sub-criteria features determines the sub-criteria score; for instance, we want to quantify the contribution of income inequality, poverty, GDP and unemployment towards the definition of the 'Economic-vulnerability' score of a country).

The main idea behind ReliefF algorithms is to estimate the quality of a feature based on how well that feature allows to distinguish between instances that are near to each other. More precisely, if a feature value difference is observed in a neighbouring instance pair with the same class (e.g., if two countries with similar value for a feature have the same Economic-vulnerability 1-2-3 score), the feature score decreases. Alternatively, if a feature value difference is observed in a neighbouring instance pair with different class values (e.g., two countries have similar value of that feature, but they have a different Economic-vulnerability 1-2-3 score), the feature score increases.

Therefore, in order to be applied, the algorithm requires knowledge of class value (or label) for each instance which, in our case, would be the sub-criteria score/class for each country. As we do not know this value a priori, before applying the ReliefF algorithm to obtain the feature weights, we derive initials labels using a Naïve scoring approach. The complete method to obtain a sub-criteria score follows these steps:

1. For each country, we convert each feature of a sub-criteria into a category (1,2 or 3) obtained by dividing the range of the feature into three bins and assessing in which bin the feature value falls.
2. We then sum all the features values, each of which is now expressed as a category and can be 1, 2 or 3, and we divide them by the maximum score attainable, e.g., for a 4-features group, the maximum score obtainable would 12 (i.e., 3 which is the maximum score attainable as described in step 1, multiplied by 4-for all the features in question).
3. We convert the expressed fraction from step 3 into a 1-2-3 category by rounding the faction value. In this way, for each country, a score (the assigned category) for each sub-component is obtained. This score is referred to as "data-derived label" and the scoring approach described (steps 1 to 3) is referred to as naive scoring approach.
4. As we now have, for each country, a set of features and a class, we can apply the ReliefF algorithm which produces weights for each feature.
5. Given the feature weights from step 4, we sum them and convert them to 1-2-3 categories in a similar way as done in step 2. This is the final sub-criteria score that is then used aggregated to produce criteria scores and the final risk score.

A complete pipeline of the sub-criteria scoring process is provided in Figure 6. In summary, we compute an initial score for each country using the Naïve score heuristic, we then feed the features and the Naïve scores to the ReliefF algorithm to obtain weights for each feature, we then use these weights to compute an adjusted weighted score which updates the Naïve scores.

*Each block involves a multitude of steps.

Figure 6. High-level pipeline of the scoring process where the Feature Weighting block is carried out with the ReliefF algorithms. Scores (Naïve and final) are expressed as categories 1 (low), 2 (medium) and 3 (high).

6.3 Preliminary experiment and results

While the computation of the final risk score will be possible only upon processing of the features comprised in each component of the framework, which is currently on going, we were able to design a pilot of the sub-criteria scoring system using the features of the Economics subheadings of the Vulnerability component (a hierarchic diagram of which is shown in Figure 7).

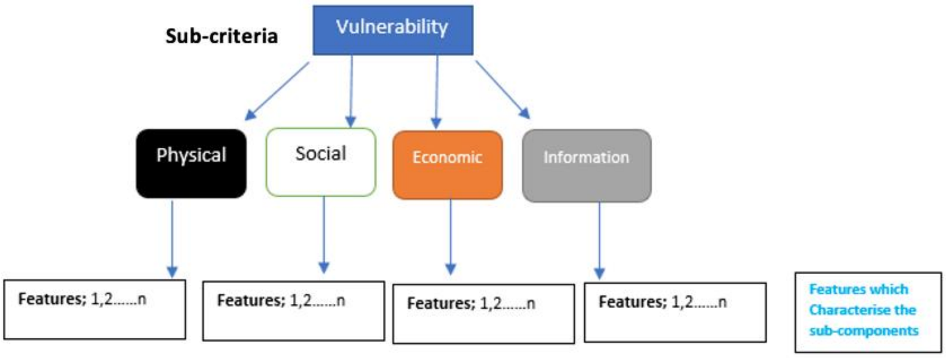

Figure 7. Hierarchical diagram showing the various sub-components and features which can be used to quantitatively establish a countries level of vulnerability

The table below outlines the results for the initial scoring (1=Low Risk, 2=Medium Risk, 3=High Risk), obtained with the Naïve approach, as well as the weighted scoring for the Economics sub-heading for the year 2020, where on this occasion there exists only a marginal change in the score for the initial and weighted scores.

Table 2. Results of the sub-criteria scoring model

	Initial Score (expressed as a factor of 3)	Weighted Score
Austria	2	2
Belgium	2	2
Bulgaria	3	3
Croatia	2	2
Czech Rep	2	1
Denmark	2	2
Estonia	2	2
Finland	2	2

France	2	2
Germany	3	3
Hungary	2	2
Ireland	2	2
Italy	3	3
Latvia	3	3
Lithuania	2	2
Luxembourg	2	2
Malta	2	2
Netherland	2	2
Poland	2	2
Portugal	2	2
Rep of Cyprus	2	2
Romania	2	2
Slovakia	2	2
Slovenia	2	2
Spain	3	3
Sweden	2	2

It should be noted that the list of countries used for the modelling exercise represent those with data for the various sub-components used to compute scores, surrogate figures may be used as part of further work for countries who are missing aspects of data. Subsequent work in this area includes the presentation of the scores which would involve the use of statistical charts for the visualisation of the risk scoring system and it is anticipated to be completed by August 2022 [M22].

7 Software implementation

The data post-processing and modelling was done in both Microsoft Excel and Python programming software, using standard libraries for data processing. The software is securely stored on TRI's cloud code repository which uses a Git-based code management solution for assuring version controlling of the code. The raw data are stored on TRI's secured Microsoft SharePoint. Raw and processed data are stored on the COVINFORM database cloud system, built leveraging STRIAD, TRI's in-house cloud data platform, solutions (the details of the database have been outlined in D2.1).

8 Next steps

Data collection activities will continue to develop in an iterative process as required. As model development continues and the model performance is tested, if new additional data is required based on data characteristics (format, coverage, reporting periods), this will be reported back to the social scientists within partner TRI, to identify alternative data sources. If required, following the validation of the model based on case study research findings, new indicators may be added to the Risk

Assessment Framework, if deemed important features for describing the risk environment. This will be reported in D2.8, 'Database containing different data sources - update M30'.

We have a collection of scripts for gathering and cleaning data from the ECDC database. These will be refined and adapted to download data for the remaining indicators. For other databases and data sources, the parts of the script that fetch and reshape data have been modified to collect and clean mobility data from Google and air quality data from the EEA. Similar modifications can be made on a case-by-case basis to collect and clean data from the remaining sources.

As far as the data modelling is concerned, the subsequent steps would now involve the computation of the risk scores for the various other sub-components (i.e., physical, social and information) which will be algorithmically combined to yield a final 'Vulnerability Index' as per the devised risk scoring model.

Regarding the validation of the risk scoring model, after completion of the Vulnerability Index computation for each country, the model's performances will be assessed using quantitative and qualitative approaches. These will involve correlation analysis between the Vulnerability Index and measures of impact and response, quality metrics of the scoring system and empirical evaluations based on literature review findings and case studies outcomes. The model performances and validation approaches will be detailed in the April 2023 [M28] update of this deliverable.

In terms of the COVINFORM dashboard, we originally planned to build the COVINFORM dashboard by adapting the dashboard provided by TRI's HAMOC solution[9], tailored to the analysis of human security environment. However, we realised that, due to the specific requirements of COVINFORM and the different level and type of user interaction required by the two use cases, building a visualisation tool from scratch, specifically for COVINFORM, would allow a faster and more efficient deployment together with higher flexibility. After analysing different alternatives on the market, we decided to use the Python library 'Dash'[10] to create an interactive dashboard as it offers relatively easy integration with the data processing pipeline and allows web-based dashboards to be produced that are easy to access and interact with. More information on the dashboard will be provided in D2.4, 'Cloud-based interactive dashboard for displaying geospatial layers' in August 2022 [M22]. In brief, data frames containing the indicator data will be fetched from the COVINFORM database (built on TRI's data platform, STRIAD, as detailed in D2.1) for visualisation. The dashboard itself will be a browser-based visualisation platform that allows the user to explore the indicator data, toggling data displays and selecting indicators to be included in the visualisation. The main functionality of this dashboard will be to display indicator data as layered choropleths (coloured geographical maps), alongside line-graphs, bar charts and scatter plots, with aggregated risk scores generated by the data science component also available for visualisation. The user will be able to select which indicators are visualised on a choropleth map, as well as choose indicators to be visualised as line graphs, or select pairs of indicators to display on a scatter plot.

[9] HAMOC solution is built on TRI's STRIAD platform. A demo of HAMOC was delivered to the consortium in previous meetings to provide an example of interactive dashboard that TRI could build and gather initial end-user requirements. Although the technology underlying the new dashboard will differ from the one used by HAMOC, it will provide the same type of visuals (map-based plots, bar/line charts, scatter plots, graph visualisations from text).

[10] https://dash.plotly.com/

Due to the complexity of the risk framework design, and its relationship with WP3 case study research, further collaboration with partners was required prior to the development of a physical dashboard. Following an in-person workshop between WP2 and WP3 in Vienna (March 2022) model testing commenced and was presented for feedback during the Spring Consortium Meeting in Lisbon (May 2022). This delayed the conduct of the second 'Evaluation' workshop as described in T2.5 with project partners (SINUS). Therefore, this will be completed in autumn 2022.

9 Conclusions

This report provides an overview on the data processing and cleaning activities carried out in fulfilment of Task 2.2 and on the progresses made towards development of the COVINFORM risk assessment tool to map response and impact level in the EU27 MS and the UK (Task 2.3). Following further research and discussions with WP3 partners on the risk assessment methodology and its relationship with the case studies, in the months following submission of D2.1, the risk assessment framework has been further refined and new data requirements have been identified accordingly. Hence this report also provides an update on the data collection activities (Task 2.1) that have been carried out after submission of D2.1 and details on the final risk framework adopted. The mapping of the relevant indicators (extracted from the datasets collected) to the risk framework model have also been defined together with a data-driven approach to assign an importance score (or weight) to each indicator of each component of the model. Performance of the scoring approach to one component of the risk framework model are here reported. The technology and deployment strategy for building the geo-spatial visualisation dashboard (Task 2.3) have been identified and preliminary dashboard prototyping is underway. In the following months, WP2 activities will focus on finalising the data collection (the final database will be reported in the M30 update of D2.1), deploying the risk model following the defined scoring approach and theoretic framework together with validating the model's outcome (both activities will be reported in the M28 update of this deliverable) and implementing a preliminary version of the cloud-based interactive visualisation dashboard, which will be demonstrated in D2.4, due at M22 (the final dashboard will be provided at the M30 update of this deliverable).

References

Andrews, N., Stowe, J., Kirsebom, F., Toffa, S., Rickeard, T., Gallagher, E., ... & Lopez Bernal, J. (2022). Covid-19 vaccine effectiveness against the Omicron (B. 1.1. 529) variant. *New England Journal of Medicine*, *386*(16), 1532-1546.

Ganin, A. A., Quach, P., Panwar, M., Collier, Z. A., Keisler, J. M., Marchese, D., & Linkov, I. (2020). Multicriteria decision framework for cybersecurity risk assessment and management. *Risk Analysis*, *40*(1), 183-199.

M. Robnik-Sikonja and I. Kononenko. Theoretical and empirical analysis of Relief and ReliefF. Machine Learning, 53:23–69, 2003.

Roloff, R. (2020). COVID-19 and No One's World. *Connections*, *19*(2), 25-37.

Trump, B. D., & Linkov, I. (2020). Risk and resilience in the time of the COVID-19 crisis. *Environment Systems and Decisions*, *40*(2), 171-173.

Wolff, S., & Ladi, S. (2020). European Union responses to the Covid-19 pandemic: Adaptability in times of permanent emergency. *Journal of European Integration*, *42*(8), 1025-1040.

Żak, M., & Garncarz, J. (2020). Economic policy towards the challenges of the COVID-19 pandemic in selected European Union countries. *International Entrepreneurship Review*, *6*(4), 21-34.

ANNEX I.

Table 3. COVINFORM Data indicators and sources for the risk framework domains of Threats, Vulnerability and Consequences

Indicator	Indicator Name	URL	Source	Format	Geographic Coverage	Time Coverage	Frequency	Indicator
Weekly Case Rate	Data on the daily number of new reported COVID-19 cases by EU/EEA country	https://www.ecdc.europa.eu/en/publications-data/data-daily-new-cases-covid-19-eueea-country	ECDC	XLSX, CSV, JSON, XML	27 MS	From Early March 2020 to Present	Daily	Weekly Case Rate
Positivity %	Data on testing for COVID-19 by week and country (indicator - "positivity rate")	https://www.ecdc.europa.eu/en/publications-data/covid-19-testing	ECDC	XLSX, CSV, JSON, XML	27 MS	Around Week 05-09, 2020 to Week 21, 2022	Weekly	Positivity %
Testing rate	Data on testing for COVID-19 by week and country (indicator - "testing rate")	https://www.ecdc.europa.eu/en/publications-data/covid-19-testing	ECDC	XLSX, CSV, JSON, XML	27 MS	Around Week 05-09, 2020 to Week 21, 2022	Weekly	Testing rate
Number of detections	SARS-CoV-2 variants of concern as of 25 May 2022	https://www.ecdc.europa.eu/en/covid-19/variants-concern	ECDC	XLSX, CSV, JSON, XML	27 MS	From week 1, 2020 to Present	Weekly	Number of detections
% Variant	SARS-CoV-2 variants of concern as of 25 May 2022	https://www.ecdc.europa.eu/en/covid-19/variants-concern	ECDC	XLSX, CSV, JSON, XML	27 MS	From week 1, 2020 to Present	Weekly	% Variant
Mobility Index	Google Mobility Data	https://www.google.com/covid19/mobility/	Google	CSV	26 MS, No Republic of Cyprus + UK	Jan 3–Feb 6, 2020. February 2020 to Present	Daily	Mobility Index
International Trade	Intra and Extra-EU trade by Member State and by product	https://ec.europa.eu/eurostat/databrowser/view/ext_lt_intratrd/default/table?lang=en	Eurostat	XLSX, CSV,	27 MS	2012 - 2021	Yearly	International Trade

	group			TSV				
Housing Concentration	Overcrowding rate by household type - EU-SILC survey	https://ec.europa.eu/eurostat/web/products-datasets/-/tessi175	Eurostat	XLSX, CSV, TSV	27 MS + UK	2010-2021	Yearly	Housing Concentration
Air Quality - Pollution	PM10 annual average of the mean daily concentration and relative % change	https://www.eea.europa.eu/data-and-maps/explore-interactive-maps/up-to-date-air-quality-data	European Environmental Agency	CSV	27 MS + UK	2019, 2020 and 2021	Daily	Air Quality - Pollution
Temperature	Mean of daily mean temperature (TG)	https://www.ecad.eu/download/millennium/millennium.php#temp	ECAD	TXT	27 MS + UK	2019, 2020 and 2021	Daily	Temperature
Age of Population	Population on 1 January by age group, sex and NUTS 3 region	https://ec.europa.eu/eurostat/web/products-datasets/-/demo_r_pjangrp3	Eurostat	XLSX, CSV, TSV	27/28 EU Countries and NUTS3 Regions	2014 - 2021	Yearly	Age of Population
Hospital Beds	Hospital beds per 100,000 inhabitants	https://ec.europa.eu/eurostat/web/products-datasets/-/tps00046	Eurostat	XLSX, CSV, TSV	27 MS + UK	2008-2019	Yearly	Hospital Beds
ICU Beds	Data on hospital and ICU admission rates and current occupancy for COVID-19	https://tradingeconomics.com/ireland/icu-beds	Trading Economics	XLSX, CSV, JSON, XML	27 MS + UK	Around Week 12-14, 2020 to Week 21, 2022	Monthly	ICU Beds
Medical Frontline Staff	Physicians by medical speciality	https://ec.europa.eu/eurostat/web/products-datasets/-/hlth_rs_spec	Eurostat	XLSX, CSV, TSV	26 MS + UK, No Slovakia	1986-2020	Yearly	Medical Frontline Staff
	Nursing and caring professionals	https://ec.europa.eu/eurostat/web/products-datasets/-/hlth_rs_prsns	Eurostat	XLSX, CSV, TSV	27 MS + UK	1980-2020	Yearly	Nursing and caring professionals
	Health personnel (excluding nursing and caring professionals)	https://ec.europa.eu/eurostat/web/products-datasets/-/hlth_rs_prs1	Eurostat	XLSX, CSV, TSV	27 MS + UK	1976-2020	Yearly	Health personnel (excluding nursing and caring professiona

								ls)
Pre-existing health conditions	Population by type of longstanding health problem, sex and age	https://ec.europa.eu/eurostat/web/products-datasets/-/hlth_dp020	Eurostat	XLSX, CSV, TSV	27 MS + UK	2011	Yearly	Pre-existing health conditions
	Births with Down's syndrome per 100 000 live births	https://gateway.euro.who.int/en/indicators/hfa_603-7120-births-with-downs-syndrome-per-100-000-live-births/visualizations/#id=19698	WHO	XLSX, CSV	27 MS + UK	Latest data available upto 2017-2019 depending on country	Yearly	
	Beds in nursing and residential care facilities, per 100 000	https://gateway.euro.who.int/en/indicators/hlthres_23-beds-in-nursing-and-residential-care-facilities-per-100-000/	WHO	XLSX, CSV	27 MS + UK	Latest data available up to 2017-2019 depending on country	Yearly	
Ports	Number of ports by country	https://www.espo.be/fact-and-figures	ESPO	XLSX, CSV	27 MS + UK	2019-2022	Yearly	Ports
	Relative change in % total cargo	https://ec.europa.eu/eurostat/databrowser/view/mar_mt_am_csvi/default/table?lang=en	ECDC	XLSX, CSV	27 MS + UK	1997 — 2020	Yearly	
Commercial Airports	Number of commercial airports	https://ec.europa.eu/eurostat/databrowser/view/avia_tf_acc/default/table?lang=en	ECDC	XLSX, CSV	27 MS + UK	2013-2019	Yearly	Commercial Airports
	Aircraft traffic data by country	https://ec.europa.eu/eurostat/databrowser/view/avia_tf_acc/default/table?lang=en	ECDC	XLSX, CSV	27 MS + UK	1993 — 2022-Q1	Monthly	
Education Level	Population by sex, age and educational attainment level (1 000)	https://ec.europa.eu/eurostat/web/products-datasets/-/lfsa_pgaed	Eurostat	XLSX, CSV, TSV	27 MS + UK	1986-2021	Yearly	Education Level
Rural vs Urban	Population by sex, age, citizenship, labour status and degree of urbanisation	https://ec.europa.eu/eurostat/web/products-datasets/-/lfst_r_pgauwsn	Eurostat	XLSX, CSV, TSV	27 MS + UK	1995-2021	Yearly	Rural vs Urban
	Population density by NUTS 3 region	https://ec.europa.eu/eurostat/web/products-datasets/-/demo_r_d3dens	Eurostat	XLSX, CSV, TSV	27 MS + UK + NUTS3 Regions	1990-2019	Yearly	
Gender - % female	Population on 1 January by age and sex	https://ec.europa.eu/eurostat/web/products-datasets/-/demo_pjan	Eurostat	XLSX, CSV, TSV	27 MS + UK	1960-2021	Yearly	Gender - % female

Migrant Population	Immigration by age group, sex and level of human development of the country of citizenship	https://ec.europa.eu/eurostat/web/products-datasets/-/migr_imm9ctz	Eurostat	XLSX, CSV, TSV	27 MS + UK	2013-2021	Yearly	Migrant Population
	Immigration by age group, sex and country of previous residence	https://ec.europa.eu/eurostat/en/web/products-datasets/-/MIGR_IMM5PRV	Eurostat	XLSX, CSV, TSV	27 MS + UK	2013-2020	Yearly	
	Population on 1 January by age group, sex and level of human development of the country of birth	https://ec.europa.eu/eurostat/en/web/products-datasets/-/MIGR_POP8CTB	Eurostat	XLSX, CSV, TSV	27 MS + UK	2013-2021	Yearly	
GDP	Real GDP per capita	https://ec.europa.eu/eurostat/web/products-datasets/-/sdg_08_10	Eurostat	XLSX, CSV, TSV	27 MS + UK	2000-2021	Yearly	GDP
% living in poverty	At-risk-of-poverty rate	https://ec.europa.eu/eurostat/web/products-datasets/-/tespm010	Eurostat	XLSX, CSV, TSV	27 MS + UK	2010-2021	Yearly	% living in poverty
Income Inequality	Inequality of income distribution	https://ec.europa.eu/eurostat/web/products-datasets/-/tespm151	Eurostat	XLSX, CSV, TSV	27 MS + UK	2010-2021	Yearly	Income Inequality
Unemploy ment Rates	Unemployment by sex, age and duration of unemployment (1 000)	https://ec.europa.eu/eurostat/web/products-datasets/-/lfsa_ugad	Eurostat	XLSX, CSV, TSV	27 MS + UK	1985-2021	Yearly	Unemploy ment Rates
Digital Literacy	Individuals' level of digital skills (until 2019)	https://ec.europa.eu/eurostat/web/products-datasets/-/isoc_sk_dskl_i	Eurostat	XLSX, CSV, TSV	27 MS + UK	2015-2019	Yearly	Digital Literacy
	Individuals' level of digital skills (from 2021 onwards)	https://ec.europa.eu/eurostat/web/products-datasets/-/isoc_sk_dskl_i21	Eurostat	XLSX, CSV, TSV	27 MS	2021	Yearly	
Digital Access	Households - level of internet access	https://ec.europa.eu/eurostat/web/products-datasets/-/isoc_ci_in_h	Eurostat	XLSX, CSV, TSV	27 MS + UK	2002-2021	Yearly	Digital Access
COVID-related death rate	Data on the daily number of new reported COVID-19 deaths by EU/EEA country	https://www.ecdc.europa.eu/en/publications-data/data-daily-new-cases-covid-19-eueea-country	ECDC	XLSX, CSV, JSON,	27 MS	From Early March 2020 to Present	Daily	COVID-related death rate

				XML				
Excess deaths	% Change in weekly mortality compared to average mortality)	https://ec.europa.eu/eurostat/databrowser/view/demo_mexrt/default/table?lang=en	Eurostat	XLSX, CSV, TSV	27 MS + UK	2020-2022	Monthly	Excess deaths
Hospital admissions	Data on hospital admission rates and current occupancy for COVID-19	https://www.ecdc.europa.eu/en/publications-data/download-data-hospital-and-icu-admission-rates-and-current-occupancy-covid-19	ECDC	XLSX, CSV, JSON, XML	27 MS	Around Week 12-14, 2020 to Week 21, 2022	Weekly	Hospital admissions
ICU admissions	Data on ICU admission rates and current occupancy for COVID-19	https://www.ecdc.europa.eu/en/publications-data/download-data-hospital-and-icu-admission-rates-and-current-occupancy-covid-19	ECDC	XLSX, CSV, JSON, XML	27 MS	Around Week 12-14, 2020 to Week 21, 2022	Weekly	ICU admissions
Food security	Inability to afford a meal with meat, chicken, fish (or vegetarian equivalent) every second day - EU-SILC survey	https://ec.europa.eu/eurostat/databrowser/view/ILC_MDES03/default/table?lang=en&category=livcon.ilc.ilc_md.ilc_mdes	Eurostat	XLSX, CSV, JSON, XML	27 MS and UK (Survey data)	2003-2022	Yearly	Food security
Loss of education	Countries that implemented digital and broadcast remote learning policies during lockdowns	https://data.unicef.org/topic/education/remote-learning-and-digital-connectivity/	UNICEF	CSV	27 MS and UK (Survey data)	2020-2021	Yearly	Loss of education
Violence	Police-recorded violent offences per hundred thousand inhabitants	https://ec.europa.eu/eurostat/cache/metadata/en/crim_off_cat_esms.htm	Eurostat	XLSX, CSV, JSON, XML	27 MS and UK (Survey data)	2003-2022	Yearly	Violence
Jos losses	Employment and activity by sex and age - annual data	https://ec.europa.eu/eurostat/databrowser/view/lfsi_emp_a/default/table?lang=en	Eurostat	XLSX, CSV, JSON, XML	27 MS and UK (Survey data)	2003-2022	Yearly	Jos losses
Social protection	Total expenditure on social protection benefits by type	https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Social_protection_statistics_-_social_benefits#Expenditure_on_social_protection_benefits_by_function	Eurostat	XLSX, CSV, JSON, XML	27 MS and UK	2003-2022	Yearly	Social protection
Unemployment, remote	To be determined	https://ec.europa.eu/eurostat/web/lfs/data/database	Eurostat	XLSX, CSV, JSON,	27 MS and UK (Survey data)	2003-2022	Yearly	Unemployment, remote

working and unemploy ment by sector				XML				working and unemploy ment by sector
Exposure to pollution	Fine particulate matter (PM2.5)	https://data.oecd.org/air/air-pollution-exposure.htm	OECD	XLSX, CSV, JSON, XML	27 MS and UK	2003-2022	Daily	Exposure to pollution
Greenhous e gas emissions	Greenhouse gas emissions	https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions	Our World Data	XLSX, CSV, JSON, XML	27 MS and UK	Variable	Yearly	Greenhous e gas emissions