Conceptualizing a Pandemic Early Warning System Using Various Data: An Integrative Approach

Sebastian Sterl
Technische Universitaet Braunschweig
sebastian.sterl@tu-berlin.de

Nirmin Almalla
Technische Universitaet Berlin
nirmin.almalla@campus.tu-berlin.de

Lars Gerhold
Technische Universitaet Braunschweig
lars.gerhold@tu-braunschweig.de

ABSTRACT

Covid-19 demonstrated the vulnerability of various systems and showed, however, that digital tools and data can serve not only to stop infections but also to detect viruses before or immediately after a zoonosis has occurred, thus preventing a potential pandemic. Although several pandemic early warning systems (P-EWS) and German pandemic-related projects (G-PRP) exist, they often use a limited data range or rely on third-party data. Here, we present a concept of an integrative pandemic early warning system (*IS-PAN*) applied to Germany using various data such as health data (e.g., clinical/syndromic) or internet data (e.g., social media/apps). Based on a systematic literature research of P-EWS and G-PRP on scientific and public health platforms, we derived indicators that help to detect virus threats with a system consisting of modules monitored in parallel. By integrating various pre-collected digital data, this approach can help to identify a potential health threat efficiently and effectively.

Keywords

Pandemic data management, epidemiological situation picture, early warning system, data fusion and integration, vulnerable communities

INTRODUCTION

"Viruses move fast. But data can move even faster. With the right information, countries and communities can stay ahead of emerging risks, and save lives". (Ghebreyesus, 2021)

Covid-19 demonstrated the vulnerability of, e.g., social, health, and economic systems and the importance of non-pharmaceutical interventions (NPI), such as social distancing or wearing masks, for fighting the pandemic before vaccines were available. A growing world population (United Nations, 2023) and thus greater interconnection as well as, e.g., increased global travel, the integration of natural environments or urbanization increase the likelihood of pandemics (Jones et al., 2008; Madhav et al., 2017; Morse, 1995). However, there are systems and tools, developed during or before Covid-19, that use epidemiological or health data to detect health issues, outbreaks, or virus spreads with a high pandemic potential. In this way, they can either prevent epidemics from becoming pandemics (*pre-pandemic*) or predict the next wave (*peri-pandemic*) and thus provide decision-makers with evidence-based data. For instance, there are already several (inter)national pandemic early warning systems (abbreviated P-EWS), such as the "Program for Monitoring Emerging Diseases" (ProMed-mail; ISID, 2023), the "HealthMap" (Boston Children's Hospital, 2023), or the "Signal" system (Robert Koch Institute (RKI), 2023) that monitor the epidemiological situation using different data sources, analytical and visualization approaches. Furthermore, German pandemic-related projects (abbreviated G-PRP*)*, were developed during or before Covid-19 to, e.g., track/break infection chains (Corona-Warn-App; German Government, 2023), measure changes of body functions (Data Donation App; RKI, 2022), or monitor the real-time infection situation in emergency rooms (Project ESEG; RKI, 2021a) in Germany. Both approaches, P-EWS and G-PRP, are based on digital data sources,

such as clinical data, application/Internet of Things (IoT) data, or (social) media data. These can be used to construct an integrative approach with data already in use to exploit synergic effects, address possible disadvantages of P-EWS and G-PRP (e.g., a limited range of used data, dependency on third-party data from other P-EWS or lack of public availability) to avoid unnecessary redundancies, and prepare the system, as Hodson (2022) put, for possible next pandemics.

Therefore, this paper asks

(1) how a successful P-EWS can be structured and (2) what elements it should include.

To address the questions above, we first present the state of the art by describing both (inter)national P-EWS and G-PRP using a comprehensive literature research. Subsequently, we explain our integrative P-EWS, the *IS-PAN*, by deriving indicators from these data sources (explained above) to structure them into different modules and integrate them in an entire concept considering virus spreads only. Then, we illustrate our concept with a practical use case and discuss further challenges and limitations. We conclude the paper with an outlook.

We fill a research gap by systematically researching pandemic-related data (partly relatively new as produced since Covid-19 only) used to derive indicators, to find synergy effects between these P-EWS and G-PRP, to address the limited range of data per system/project, and to identify signals in a parallelly structured modular system which can be less prone to missing data.

STATE OF THE ART AND RELATED WORK

To research (inter)nationally active P-EWS and G-PRP, we conducted an extensive literature research on scientific internet databases such as GoogleScholar, Web of Science, PubMed or ResearchGate using strings like "(pandemic/epidemic) early warning system, pandemic/epidemic, Corona/Covid* projects" etc. Also, we researched on public health platforms like the German Robert Koch Institute, Federal Ministry of Health, Federal Ministry of Education and Research, and German universities/research institutions to include as many G-PRP as possible and to finally derive indicators from the data of all researched P-EWS and G-PRP. We classified all the indicators using an inductive category scheme we developed from the data itself (e.g., intensive care beds belong to main category "clinical data", syndromic data from hospitals belong to main category "syndromic data").

Pandemic Early Warning Systems (P-EWS)

In general, an early warning system is "[t]he set of capacities needed to generate and disseminate timely and meaningful warning information to enable individuals, communities and organizations threatened by a hazard to prepare and to act appropriately and in sufficient time to reduce the possibility of harm or loss (United Nations, 2009, p. 12)". These threats or hazards can be natural (e.g., biological, zoonotic) or man-made (e.g., nuclear, wars) and divided into short-, middle- and long-term hazards (Glantz, 2004). We define a P-EWS as one that warns its users about "infectious diseases [by analyzing] surveillance data with specialized technologies for early detection and warning of notable aberrations" (Yang et al., 2017, p. 3). Meaning, it observes the situation with (1) indicators to (2) predict a probable event and (3) send warnings that a disastrous event may be imminent in order to (4) start emergency measures operated by (inter)national emergency authorities (León et al., 2006, p. 23).

Table 1 illustrates a small selection of some P-EWS, their related institutions/operators, data sources, current status, and disadvantages we identified. We focus on data sources to derive indicators necessary for the *IS-PAN* further below. Various types of P-EWS integrate data sources, such as media/official reports or news sites as well as clinical indicators or unstructured data like social media. For instance, the ProMED-mail system (ISID, 2023) is an internet-based information and communication system which uses the "one medicine" concept[1], by recognizing animal diseases that may become zoonotic (Madoff, 2004, p. 228). ProMED is publicly available and verified by experts. The system "HealthMap" (Boston Children's Hospital, 2023) is an internet-based global map that searches, e.g., for articles on health threats in different languages (Lyon et al., 2012, p. 224). Then, these articles are analyzed by relevance to aggregate publicly available data (Brownstein et al., 2008, p. 1021). In Germany, the information system "Signale" (or Signals), operated by the RKI (2023), automatically analyzes specific "signals"[2] in the data to provide information on potential domestic outbreaks of transmissible diseases. The first version of "Signale" was based on the clinical and medical case data of the "Infection Protection Act" (RKI, 2021b). The second version plans to expand the data pool to include structured and unstructured data as

[1] The "one medicine" concept means, "that there is no difference of paradigm between human and veterinary medicine and both disciplines can contribute to the development of each other". It is strongly connected to the "ecosystem health", meaning, there are rather global and interdependent connections between health and the ecosystem as a whole (Zinsstag et al., 2011, p. 148).

[2] In this project, epidemiological case data on a daily basis are analyzed and stratified by population subsets (e.g., location, age group, or sex) and pathogen/symptoms. Then, using certain algorithms, the stratified data subsets are parallelly analyzed and compared to historic data to obtain a threshold where a signal occurs when the case numbers exceed the threshold (Salmon et al., 2016).

well as data from social media.

Table 1. Selection of P-EWS

Name / (citation)	Operated by	Data sources (e.g.)	Status	Disadvantages (e.g.)
The Global Public Health Intelligence Network (GPHIN) / (PHAC, 2023)	Public Health Agency of Canada (PHAC) + The Center for Emergency Preparedness and Response (CEPR) Canada	Media reports by health care organizations, product recalls, curated RSS feeds, news sites	Active since 1997	Not publicly accessible, (only) based on reports
The Program for Monitoring Emerging Diseases (ProMED-mail) / (ISID, 2023)	International Society for Infectious Diseases (ISID) USA	Media reports, official reports, local observers, other sources	Active since 2004	(Only) based on reports
HealthMap / (Boston Children's Hospital, 2023)	Boston Children's Hospital USA	Expert-curated discussions such as ProMED mail, verified official reports from organizations such as WHO, online news sources via aggregators such as Google News/ Eurosurveillance/"Wildlife Information Node"	Active since 2006	Based on third-party data
Semantic Processing and Integration of Distributed Electronic Resources for Epidemiology (EpiSPIDER) / (Tolentino et al., 2007)	US Centers for Disease Control and Prevention (CDC)	Daylife, Google, Moreover, ProMED, Twitter, WHO	Active since 2006	Only English articles, based on third-party data
DEMIS, Signale 2.0 (Signals 2.0) /(RKI, 2023)	RKI Germany	IfSG (Infection Protection Act) case data (number of cases in specific region and population group), structured data (e.g., syndromic monitoring), curated unstructured data (e.g., professional publications), social media data (e.g., search queries, sentiment analysis)	Previous project Signale 1 active since 2018	Signale 2 not active yet
Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE II) / (Burkom et al., 2021)	US CDC	Clinical indicators of human behavior (including syndromes from the emergency department to syndrome-grouped billing codes from private practices and veterinary syndromes), nonclinical indicators of human behavior (including absenteeism, calls to the nurse hotline, prescription medications and over-the-counter self-medication)	Active	Not publicly accessible

Note: This is only a small extract for this WiP paper. The complete list can be requested from the authors.

In sum, at both the national and international level, there are several P-EWS focused on certain diseases, regions, populations or interest groups. However, these systems are unable to monitor every critical situation (e.g., virus/outbreak threat situation) globally, as a potential pandemic would require and are dependent on either third-party data from other P-EWS (and therefore partly redundant), only use a narrow data range or are not publicly accessible. However, the various data sources and structures can be used to derive indicators for an integrative P-EWS. An integrative system is crucial because, as Glantz puts it (2004, p. 9), "while certain indicators observed by one person might suggest that a warning is warranted, another person using different indicators might not believe that a warning is warranted". Thus, an integrative approach can monitor several indicators to detect a

possible virus threat and to reduce a false-negative, which presents a risk for the population or false-positive classification, which decreases trust, is costly, and binds important resources.

German Pandemic-related Projects (G-PRP)

Covid-19 has given rise to numerous research and development projects in Germany. *Table 2* shows some examples of G-PRP, along with the operators, data sources, time point of warning and an evaluation of its effectiveness if applicable. "Pre-pandemic" means that the projects/data can be used to detect a possible virus threat before it becomes an epidemic/pandemic, whereas "peri-pandemic" means that an epidemic/pandemic has already begun. Therefore, data help to predict, e.g., next waves or outbreak hotspots. In the pre-pandemic phase, projects such as the "agent-based modeling of epidemics" (National Research Platform for Zoonoses, 2023) generate simulation data of possible zoonoses and affected areas, while the ESEG project operated by the RKI (2021a) uses real-time data from emergency rooms to help detect infections. Data such as bed occupancy/capacity (RKI & DIVI, 2023) as well as body data (e.g., temperature, fever; RKI, 2022) are useful in both the pre- and peri-pandemic phases. Projects like the Corona-Warn-App (German Government, 2023) or MODUS-COVID (Technical University Berlin, 2021) track infection chains or use simulation data (based on R-values, etc.) to, e.g., inform decision-makers about the necessity of NPIs and to provide reasons for measures such as lockdowns or mask mandates.

Table 2. Selection of G-PRP

Project name / (citation)	Operated by	Data sources	Time point of warning	Evaluation / effectiveness
Pilot project: Agent-based modeling of epidemics / (National Research Platform for Zoonoses, 2023)	National Research Platform for Zoonoses	Agent-based simulation of epidemics (zoonoses)	Pre-pandemic	Improve quality and effects of epidemic/pandemic response, without requiring explicit input of epidemiologists/other experts (Ciunkiewicz et al., 2022)
Erkennung und Sicherung Epidemischer Gefahrenlagen ESEG (Detection and Protection of Epidemic Situations) / (RKI, 2021a)	Robert Koch Institute (RKI), Public Health Department Frankfurt	Real-time infection data from emergency rooms	Pre-pandemic	N/A
DIVI-Intensive registry (German Interdisciplinary Association for Intensive and Emergency Medicine) / (RKI & DIVI, 2023)	RKI, DIVI	Bed occupancy/capacity, intensive care beds	Pre/Peri-pandemic	Data match with data from other surveillance systems (DIVI, 2021)
Datenspende-App (Data Donation App) / (RKI, 2022)	RKI	Body data using wearables/smartwatches	Pre/Peri-pandemic	Unusual amplitudes of temperature chart detectable (Wiedermann et al., 2022)
Corona-Warn-App / (German Government, 2023)	German Government, RKI	Data on possible infection risk, tracking of infection chains	Peri-pandemic	74% got test results within a day; 61% of positively tested users warned others by sharing results; 87% of warned users got tested; 6% of tests made after red warning were positive; positive results warned other users (CWA-Team, 2021)
MODUS-COVID / (Technical University Berlin, 2021)	TU Berlin	Simulation data on NPI and their effectiveness	Peri-pandemic	Mathematical modeling used since beginning of Covid-19-

		pandemic helped to contain spread of Covid-19 (Federal Ministry of Education and Research, 2021)

Note: This is only a small extract for this WiP paper. The complete list can be requested from the authors.

All in all, G-PRP developed during or before Covid-19 can deliver valuable data for an integrative system consisting of several elements. For one, the combination of elements makes it less likely that the entire system fails in case of a threat, and for another, it exploits synergy effects that a single project cannot guarantee. The next section collects different types of data from both P-EWS and G-PRP to derive indicators that can be used in the integrative P-EWS *IS-PAN* introduced below.

Data Sources from P-EWS and G-PRP

Table 3 sums up both the P-EWS shown in *Table 1* and the G-PRP displayed in *Table 2*. The P-EWS and G-PRP produce, among other things, clinical data with sub data, including bed occupancy and capacity, intensive care beds or real-time data from emergency rooms (see column "type of data" and "sub data"). Furthermore, recently developed digital tools can enrich the database by providing applications or technologies such as the "Internet of Things" (IoT) or social media. Also, media and official reports or reports by local observers provide the system with useful information. On the basis of the data types shown in *Table 3*, we derive the final indicators for the integrative P-EWS *IS-PAN* in the next section and then conclude with an explanation of the entire concept.

Table 3. Selection of data sources derived from table 1 and table 2 for *humans* as object of study

G-PRP | P-EWS / (citation)	Type of data	Sub data
DIVI-Intensive registry / (RKI & DIVI, 2023)	Clinical Data (inpatient)	Bed occupancy/capacity, intensive care beds
ESEG (Detection and Protection of Epidemic Situations) / (RKI, 2021a)	Clinical Data (inpatient)	Real-time data from emergency rooms
Corona-Warn-App / (German Government, 2023)	Applications	Data on possible infection risk, tracking of infection chains
Datenspende-App (Data donation app) / (RKI, 2022)	IoT	Body data (height, weight), measured values (sleep patterns, heart rate, number of steps), sociodemographic information, survey data, fitness wristbands/smartwatches
HealthTweets / (Dredze & Broniatowski, 2014)	Social media	Twitter data
ProMED-mail / (ISID, 2023)	Data from P-EWS	Media/official reports, local observers, and other sources

Note: This is only a small extract for this WiP paper. The complete list can be requested from the authors.

INTEGRATIVE PANDEMIC EARLY WARNING SYSTEM (*IS-PAN*)

Derived Indicators

Table 4 shows the complete list of indicators derived from all researched P-EWS and G-PRP, a selection of which is presented in *Table 1*, *Table 2*, and *Table 3* in the previous sections whereas a full list can be requested from the authors. Taken together, the indicators, including sub data, encompass clinical, syndromic, and medical data, laboratory (lab) and wastewater data as well as to survey, social media, and IOT/application data. All of these indicators can be integrated into a single system.

Table 4. Grouping and classification of the sub data related to humans as object of study

Indicator	Full set of sub data
Clinical data	Bed occupancy, intensive care beds, severity and course of the disease, disease burden, syndromic data from hospitals/emergency rooms, Infection Protection Act data
Syndromic data	Syndromic data from hospitals/emergency rooms/community shelters
Medical data	Data from primary medical care, Infection Protection Act data, telemedicine data, health-related web data
Demographic data	Mortality rate, place of death, cause of death, type of death
Laboratory data	Infection Protection Act data, virus species determination by wastewater, infection level due to waste water, genomic information on mutants in wastewater
Wastewater data	Virus species determination by wastewater, infection level due to waste water, genomic information on mutants in wastewater
Autopsy data	Cause of death, pathomechanisms, therapy options, routes of spread of Covid-19 within the human body, effects of Covid-19 on the organism/organs
Simulation data	Intelligent simulation and evaluation routine of existing data and intersections, agent-based modelling of epidemics, predicting the expected number of intensive care Covid-19 patients, 6-day forecasts of Covid-19 case numbers by state
Survey data	Covid-19 perception and behavior data, information from the population on respiratory diseases
Social media	Twitter data
IOT-devices	Fitness wristbands/smartwatches
Applications	Body data, measurements, sociodemographics, survey data, fitness wristbands/smartwatches, potential infection risk data, person-related infection data, symptom data
Data from P-EWS	Media/official reports, local observers, ProMed-Mail, WHO, Eurosurveillance, Google News, Wildlife Information Node
Other types of data	Absence data (daycare centers/schools/work), sales data (pharmacy), mobility data

Note: This is the complete list of indicators derived from all researched P-EWS and G-PRP where only a small extract for this WiP paper is given in tables 1, 2, and 3.

The indicators from *Table 4* are divided into different domains in *Table 5*. Here, we also differentiate between data collected from two main objects of study, namely *humans* and *animals*. However, for reasons of clarity, we only explain the classification of indicators from humans. “Medical/health data” refer to verbal or text data or data related to numerical measures, recorded signals, drawings or pictures (Shortliffe & Cimino, 2014, pp. 48–49). These data are collected by authorized personnel including doctors, lab workers, or hospital staff (Shortliffe & Cimino, 2014, pp. 52–54). A second indicator class is called “internet data”. Due to the increasing amount of digital internet data and the ability to process them, this domain constitutes a highly relevant data source apart from pure medical/health data and can greatly enrich *IS-PAN*. Subdomains are process-produced data (e.g., social media or queries during web searches), active data donations (e.g., survey data, apps, IoT) or web data from projects, and professional/academic publications and research institutes (e.g., simulation data). “Other types of data” that may be relevant for integration into the *IS-PAN* include sales data from pharmacies, absence data, or mobility data. These indicators can enlarge the picture of the epidemiological situation and help detect a possible virus threat as well as help predict next infection waves/hotspots.

Table 5. Classification of indicators into object of study and subdomains

Collected from humans					**Collected from animals**			
Medical/health data	**Internet data**			**Other types of data**	*Clinical data*	*Simulation data*	*Internet data (citizen science data)*	*Other types of data*
	Unintended data generating/process produced data	**Active data donations**	**Data from projects/ professional publications /research institutes**		*Lab data*			

Medical data *Clinical data* *Syndromic Data* *Demographic data* *Autopsy data* *Laboratory data* *Wastewater data*	*Social media* *Queries during web search* *Health-related web data*	*Survey data* *Applications* *IOT-devices*	*Simulation data* *Data from P-EWS*	*Sales data (pharmacy)* *Absence data (daycare centers/ schools/work)* *Mobility data*			

Entire Concept: *IS-PAN*

Figure 1 displays the entire concept of the *IS-PAN* to show the processes' different steps and to highlight that data collection must be embedded in the whole system. The figure is structured as an ER-diagram, showing entities in rectangles and their relationships in circles. Beginning at the top left, a pandemic presents a risk/threat for, e.g., the health, social, or economic system. Therefore, it requires an early-warning mechanism, using data-based technologies, either to detect a virus threat before a pandemic occurs (pre-pandemic) and so to prevent a zoonosis or to predict next waves/hot spot regions (peri-pandemic) and justify NPIs. Data-based technologies rely on data collected from humans, animals and the environment, all of which can carry pathogens. Data from humans – this paper's sole focus – are composed of the main domains "medical/health data", "internet data", and "other types of data. Following successful data collection for several pandemic indicators, the data have to be stored (e.g., in data warehouses), processed and analyzed (e.g., by predictive analyses, AI), and visualized (e.g., in maps, graphs) to serve as a decision support system for, e.g., authorities such as public health institutes as main users that collect epidemiological data, consult the government and politicians and communicate public-health related information (such as protection recommendations) to the public.

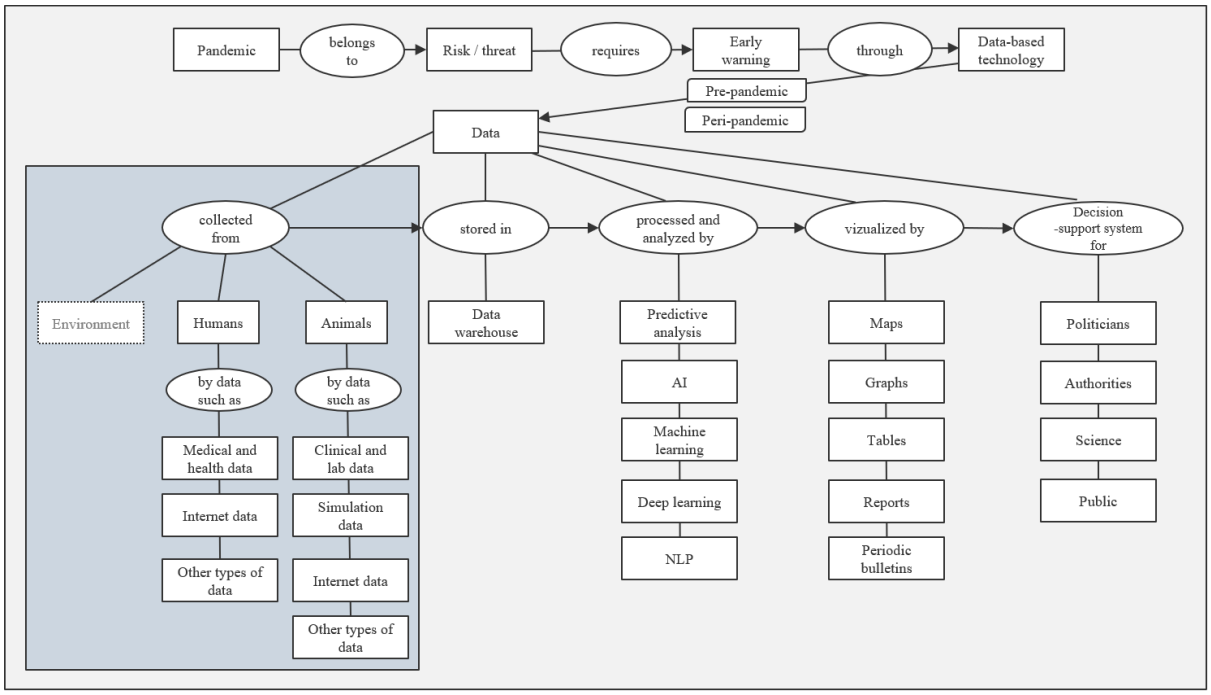

Figure 1. Concept of the IS-PAN using an ER-diagram

In sum, the concept of the *IS-PAN* comprises all steps, beginning with data collection (the focus of this paper) to storing, processing, analyzing, and visualizing the data as a way to provide more evidence-based support for decision-makers. Also, the modular system makes it possible to monitor different indicators simultaneously, thus increasing the likelihood of finding signals in the data and reducing false-negative and false-positive results. Here, we focus on presenting P-EWS and G-PRP as well as the inductive derivation of pandemic indicators and their integration into an entire concept, and not the data analysis itself addressing this in the challenges section. The next section presents a use case to explain the functionality of the *IS-PAN* more detailed.

USE CASE

Here, we present a completely fictional use case that explains the functionality of the *IS-PAN* using a system of modules monitored in parallel. We assume two different regions A and B consisting of two different age groups (see *Table 6*). Only young people aged 18 to 30 years live in region A, whereas only older, already retired people from 65 to 80 years live in region B. First, the source data stored in three different places (by clinical, social media and absence data) have to be harmonized by time, region, age group and indicator each to obtain three single target data sets ready to be appended into a final target data set for further analyses. Now, an unknown pathogen occurs in both populations equally. As older people might be more prone to more serious symptoms due to comorbidities, hospitalizations, and the severity and course of the disease, the indicator "clinical data" applies to region B only. Following another simple assumption, younger people use the indicator "social media" more frequently, and they are actively employed compared to the retired population B (indicator: "absence data"). Signals in region A, provided by, e.g., social-media notifications ("I feel ill", etc.) or absence data from work, can be used to identify virus spread in this population.

Table 6: Fictional example of an integrative P-EWS: different signals from two different populations

Time	Region	Age group	Indicator: clinical data	Indicators: social media / absence data	Number of infection cases due to indicator	Number of non-infected
0	A	18-30	0	0	0	100
1	A	18-30	0	1	2	98
2	A	18-30	0	1	4	96
3	A	18-30	0	1	8	92
4	A	18-30	0	1	16	84
5	A	18-30	0	1	32	68
0	B	65-80	0	1	0	100
1	B	65-80	1	0	2	98
2	B	65-80	1	0	4	96
3	B	65-80	1	0	8	92
4	B	65-80	1	0	16	84
5	B	65-80	1	0	32	68

The example is constructed in such a way that the indicators (which can be updated with new incoming data by time/location) produce different signals in either region. In other words, although the pathogen spreads in the same way in both regions, it can be detected using different indicators dependent on the region. This modular system functions like a "parallel connection", ensuring that each signal can be located, even though other indicators do not work for this specific area due to, e.g., various demographic distributions. Further, the system would monitor each location on different levels (states, communities) on a regular basis as time and space are key variables and necessary to merge differently structured data after their harmonizing to make it comparable in a first step. The advantage of this system is that it is less likely to fail, since it is possible to detect signals even if other indicators show no signals and/or lack information. In this example, we did not weigh signals yet as it is in a work in progress status. All in all, we used this very fictional assumption to show how the *IS-PAN* can work and identify signals in different populations through different data that can only be identified when all data of the P-EWS and G-PRP are considered. The assumptions given in the beginning of this section does not necessarily reflect reality, but show how the indicators can work dependent on time, location and sociodemographic characteristics.

CHALLENGES AND LIMITATIONS

Challenges

The *IS-PAN* concept implies several challenges, that are briefly described. Regarding data collection, on which

this paper focuses, it is difficult to access all the data discussed above due to restrictions in public use licenses, thus allowing for proprietary use only (e.g. GPHIN). Also, some data sources comprise data from third parties, such as Twitter or ProMED-mail, as well as primary data from labs/hospitals, which follow several levels of data protection. A semi-public, integrative P-EWS, which is only partly accessible to the public and where only certain user roles can see/process sensitive data, could address this problem. Where the concept as a whole is concerned, it is a challenge to link all steps with one another and to store, process (e.g., data fusion), analyze, and visualize data in a consistent and comprehensive way. For instance, all the data have to be stored, harmonized and merged before analyzing the results. Data structures that differ in, e.g., time, space, or object of study must be harmonized and merged to make data consistent. Furthermore, all the entities described in the entire concept have to be connected efficiently to make all steps work and to produce easily readable, automated, and approved results that prove useful for decision-makers. As mentioned in section "use case", we have not weighed the indicators yet since it is in a work in progress status. Thus, we will work on a weighing scheme in the future to rank signals which are more important than others to identify potential virus threats. Furthermore, reliability and credibility of social media is another issue to work on by both developing and applying filter systems checking and processing social media data. Another challenge is, that companies such as Google might have global search data being highly valid. However, Google services such as Google flu trends was prone to bias in an overestimation of case numbers (Kandula & Shaman, 2019).

Limitations

Although this concept has several limitations, we briefly present only a few. One limitation of this integrative approach is the focus on data collected from humans (presented in this paper) and animals, which have been the main source of outbreaks and subsequent epidemics/pandemics (Pike et al., 2010). While we have left out other objects of study and indicators related to, e.g., foodborne disease outbreaks (FDA, 2022) or environment-borne diseases (e.g., airborne infectious diseases containing bacteria; Bango et al., 2021), these must be considered in future research from a holistic integrative perspective. For foodborne disease outbreaks, lab data from laboratories of "Consumer Protection" and food analyses could be helpful. Regarding environment-related indicators, Bango et al. (2021) offer a concept of decision analyses for a P-EWS using a distributed network of air samplers via electrostatic air precipitation. Additionally, the present paper only considers viruses. Other pathogens such as bacteria and their ways of infections do not play a role in this concept and could be therefore analyzed more deeply in future research. Another limitation is the credibility and reliability of social media as well as the individually different accessibility to users by locations and sociodemographics.

CONCLUSION AND OUTLOOK

In this paper, we presented a concept of an integrative P-EWS, the *IS-PAN*, explaining first its structure and then its constitutive elements. As discussed above, the *IS-PAN* includes P-EWS and G-PRP, the data sources derived from them as well as their functioning and addresses their advantages and disadvantages. Thereafter, we developed relevant indicators based on P-EWS and G-PRP and assigned them to the objects of study "humans" and "animals" for pre- or peri-pandemic forecasts. However, for reasons of clarity, we presented only "humans" here. Next, we elaborated on the concept by integrating all the single elements and levels into one process, including the steps of collecting, analyzing, and visualizing data. In doing so, we highlighted that different digital data sources were available during the collection step, including those available before and those generated during Covid-19. This work has three particular strengths: the comprehensive research on P-EWS and G-PRP, the systematic derivation of indicators related to the objects of study, and the modular composition of the system from heterogeneous elements, which brings together signals that occur only in certain subgroups with other signals from other data sources in other subgroups explained with a use case.

Next, the concept should be elaborated further by developing ways to obtain, process, and analyze the data to visualize them for an evidence-based decision support system. This includes, e.g., a weighting scheme to rank indicators and their signals by importance as well as filter systems processing, e.g., social media data as well as developing efficient data fusion steps including harmonizing and merging the differently structured data into one target data set. As the *IS-PAN* is an integrative approach including a full range of indicators and sub data derived by the extensive literature research explained above, we plan to reduce the indicator set to keep it more feasible according to the limitations section.

All in all, the concept offers a valuable way to integrate different pre-collected digital data sources, thus efficiently and effectively identifying potential health threats in the context of disaster public health and healthcare informatics and providing targeted and systematic pandemic data management, analysis, fusion, and visualization.

ACKNOWLEDGMENTS

We would like to thank our research intern Sophie Teigler from Technische Universitaet Braunschweig for her invaluable support in writing this article.

REFERENCES

Bango, J. J., Agostinelli, S. A., Maroney, M., Dziekan, M., Deeb, R., & Duman, G. (2021). A Pandemic Early Warning System Decision Analysis Concept Utilizing a Distributed Network of Air Samplers via Electrostatic Air Precipitation. *Applied Sciences*, *11*(11), 5308. https://doi.org/10.3390/app11115308

Boston Children's Hospital. (2023). *HealthMap | Flu Map | Contagious Disease Surveillance | Virus Awareness*. https://www.healthmap.org/en/

Brownstein, J. S., Freifeld, C. C., Reis, B. Y., & Mandl, K. D. (2008). Surveillance Sans Frontières: Internet-based emerging infectious disease intelligence and the HealthMap project. *PLoS Medicine*, *5*(7), e151. https://doi.org/10.1371/journal.pmed.0050151

Burkom, H., Loschen, W., Wojcik, R., Holtry, R., Punjabi, M., Siwek, M., & Lewis, S. (2021). Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE): Overview, Components, and Public Health Applications. *JMIR Public Health and Surveillance*, *7*(6), e26303. https://doi.org/10.2196/26303

Ciunkiewicz, P., Brooke, W., Rogers, M., & Yanushkevich, S. (2022). Agent-based epidemiological modeling of COVID-19 in localized environments. *Computers in Biology and Medicine*, *144*, 105396. https://doi.org/10.1016/j.compbiomed.2022.105396

CWA-Team. (2021). *SB01: Über die Wirksamkeit und den Nutzen der Corona-Warn-App*. https://www.coronawarn.app/de/science/2021-06-15-science-blog-1/#welches-ziel-hat-die-evaluation-der-corona-warn-app-und-welche-aspekte-spielen-dabei-eine-rolle

Deutsche Interdisziplinäre Vereinigung für Intensiv- und Notfallmedizin. (2021). *Stellungnahme: Das DIVI-Intensivregister hat zur Bewertung der Lage in der Pandemie stets belastbare Zahlen geliefert!* https://www.divi.de/presse/pressemeldungen/stellungnahme-das-divi-intensivregister-hat-zur-bewertung-der-lage-in-der-pandemie-stets-belastbare-zahlen-geliefert

Dredze, M., & Broniatowski, D. (2014). HealthTweets.org: A Platform for Public Health Surveillance using Twitter. *AAAI Workshop on the World Wide Web and Public Health Intelligence*. https://www.academia.edu/7100377/HealthTweets_org_A_Platform_for_Public_Health_Surveillance_using_Twitter

FDA (2022). Outbreaks of Foodborne Illness. *FDA*. https://www.fda.gov/food/recalls-outbreaks-emergencies/outbreaks-foodborne-illness

Federal Ministry of Education and Research. (2021). *Karliczek: Modellierungen verbessern, um künftige Pandemien gezielter bekämpfen zu können: BMBF fördert Verbundvorhaben mit bis zu 10 Millionen Euro*. https://www.bmbf.de/bmbf/shareddocs/downloads/files/2021-05-28_114-pm_modellierung.pdf?__blob=publicationFile&v=1

German Government. (2023). *Corona Warn App: Unterstützt uns im Kampf gegen Corona | Bundesregierung*. https://www.bundesregierung.de/breg-de/themen/corona-warn-app

Ghebreyesus, T. A. (2021, September 1). *Tedros Adhanom Ghebreyesus on Twitter: „Viruses move fast, but data can move even faster. With the right information, countries & communities can stay ahead of emerging risks, & save lives. I'm honoured to open the @WHO Hub for Pandemic & Epidemic Intelligence together with #Germany Chancellor Merkel in Berlin today." / Twitter* [Tweet]. Twitter. https://twitter.com/drtedros/status/1433159081259413511?lang=de

Glantz, M. (2004). Usable Science 8: Early Warning Systems: Do's and Don'ts. https://www.academia.edu/31773841/Usable_Science_8_Early_Warning_Systems_Dos_and_Donts

Hodson, R. (2022). Preparing the world for the next pandemic. *Nature*, *610*(7933), S33. https://doi.org/10.1038/d41586-022-03353-9

ISID. (2023). *ProMED-mail*. https://promedmail.org/

Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., & Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature*, *451*(7181), 990–993. https://doi.org/10.1038/nature06536

Kandula, S., & Shaman, J. (2019). Reappraising the utility of Google Flu Trends. *PLoS Computational Biology*, *15*(8), e1007258. https://doi.org/10.1371/journal.pcbi.1007258

León, J.V.D., Bogardi, J., Dannenmann, S., & Basher, R. (2006). Early warning systems in the context of disaster risk management. *Entwicklung and Landlicher Raum*. https://www.unisdr.org/2006/ppew/info-resources/docs/ELR_dt_23-25.pdf

Lyon, A., Nunn, M., Grossel, G., & Burgman, M. (2012). Comparison of web-based biosecurity intelligence systems: Biocaster, EpiSPIDER and HealthMap. *Transboundary and Emerging Diseases*, *59*(3), 223–232. https://doi.org/10.1111/j.1865-1682.2011.01258.x

Madhav, N., Oppenheim, B., Gallivan, M., Mulembakani, P., Rubin, E., & Wolfe, N. (2017). Pandemics: Risks, Impacts, and Mitigation. In N. Madhav, B. Oppenheim, M. Gallivan, P. Mulembakani, E. Rubin, & N. Wolfe (Eds.), *Disease Control Priorities: Improving Health and Reducing Poverty. 3rd edition.* The International Bank for Reconstruction and Development / The World Bank. https://doi.org/10.1596/978-1-4648-0527-1_ch17

Madoff, L. C. (2004). Promed-mail: An early warning system for emerging diseases. *Clinical Infectious Diseases*, *39*(2), 227–232. https://doi.org/10.1086/422003

Morse, S. S. (1995). Factors in the emergence of infectious diseases. *Emerging Infectious Diseases*, *1*(1), 7–15. https://doi.org/10.3201/eid0101.950102

National Research Platform for Zoonoses. (2023). *Pilotprojekt: Agentenbasiertes Modeling von Epidemien | Nationale Forschungsplattform für Zoonosen*. https://www.zoonosen.net/pilotprojekt-agentenbasiertes-modeling-von-epidemien

PHAC. (2023). *Global Public Health Intelligence Network*. https://gphin.canada.ca/cepr/listarticles.jsp?language=en_CA

Pike, B. L., Saylors, K. E., Fair, J. N., LeBreton, M., Tamoufe, U., Djoko, C. F., Rimoin, A. W., & Wolfe, N. D. (2010). The origin and prevention of pandemics. *Clinical Infectious Diseases*, *50*(12), 1636–1640. https://doi.org/10.1086/652860

Robert Koch Institute. (2021a). *RKI - ESEG-Projekt*. https://www.rki.de/DE/Content/Infekt/Ausbrueche/ESEG/ESEG_node.html

Robert Koch Institute. (2021b). *RKI - Infektionsschutzgesetz (IfSG)*. https://www.rki.de/DE/Content/Infekt/IfSG/ifsg_node.html

Robert Koch Institute. (2022). *RKI - Coronavirus SARS-CoV-2 - Datenerhebung der Corona-Datenspende-App endet zum 31.12.2022*. https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Corona-Datenspende-allgemein.html

Robert Koch Institute. (2023). *RKI - Signale-Frühwarnsystem*. https://www.rki.de/DE/Content/Infekt/IfSG/Signale/Signale_Fruehwarnsystem_node.html

Robert Koch Institute, & DIVI. (2023). *DIVI Intensivregister*. https://www.intensivregister.de/#/index

Salmon, M., Schumacher, D., Burmann, H., Frank, C., Claus, H., & Höhle, M. (2016). A system for automated outbreak detection of communicable diseases in Germany. *Euro Surveillance : Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin*, *21*(13). https://doi.org/10.2807/1560-7917.ES.2016.21.13.30180

Shortliffe, E. H., & Cimino, J. J. (2014). *Biomedical Informatics*. Springer London. https://doi.org/10.1007/978-1-4471-4474-8

Technical University Berlin. (2021). *MODUS-COVID: Modellgestützte Untersuchung von Massnahmen zur Eindämmung von COVID-19 - Senozon AG*. https://senozon.com/modus-covid/

Tolentino, H.D., Kamadjeu, R., Fontelo, P., Ms, F., Matters, M., Pollack, M. & Madoff, L. (2007). Scanning the Emerging Infectious Diseases Horizon - Visualizing ProMED Emails Using EpiSPIDER. https://www.semanticscholar.org/paper/Scanning-the-Emerging-Infectious-Diseases-Horizon-Tolentino-Kamadjeu/932ce547df4e88c3f2615279587d7fa70cc4c16b

United Nations. (2009). *2009 UNISDR terminology on disaster risk reduction*. https://www.unisdr.org/files/7817_UNISDRTerminologyEnglish.pdf

United Nations. (2023). *Population | United Nations*. https://www.un.org/en/global-issues/population

Wiedermann, M., Rose, A. H., Maier, B. F., Kolb, J. J., Hinrichs, D., & Brockmann, D. (2022, April 6). *Evidence for positive long- and short-term effects of vaccinations against COVID-19 in wearable sensor metrics -- Insights from the German Corona Data Donation Project*. http://arxiv.org/pdf/2204.02846v1

Yang, W., Lan, Y., Sun, Q., Wang, J., & Li, Z. (2017). Introduction. *Early Warning for Infectious Disease Outbreak*, 3–22. https://doi.org/10.1016/B978-0-12-812343-0.00001-1

Zinsstag, J., Schelling, E., Waltner-Toews, D., & Tanner, M. (2011). From “one medicine” to “one health” and systemic approaches to health and well-being. *Preventive Veterinary Medicine*, *101*(3-4), 148–156. https://doi.org/10.1016/j.prevetmed.2010.07.003