

Containing Future Epidemics with Trustworthy Federated Systems for Ubiquitous Warning and Response

Dick Carrillo, *Member, IEEE*, Pedro H. J. Nardelli, *Senior Member, IEEE*, Evangelos Pournaras, Plinio Morita, *Member, IEEE*, Demóstenes Z. Rodríguez, *Senior Member, IEEE*, Merim Dzaferagic, Harun Siljak, *Member, IEEE*, Alexander Jung, *Member, IEEE*, Laurent Hébert-Dufresne, Irene Macaluso Mehar Ullah, *Student Member, IEEE*, Gustavo Fraidenraich, *Senior Member, IEEE*, Petar Popovski, *Fellow, IEEE*

Abstract—In this paper, we propose a global digital platform to avoid and combat epidemics. It leverages the pervasiveness of wireless connectivity while being trustworthy and secure. The proposed system is conceptualized to be decentralized yet federated, based on ubiquitous public systems and active citizen participation. Its foundations lie on the principle of informational self-determination. We argue that only in this way it can become a trustworthy and legitimate public good infrastructure for citizens by balancing the asymmetry of the different hierarchical levels within the federated organization while providing highly effective detection and guiding mitigation measures towards graceful lockdown of the society.

Index Terms—COVID-19, epidemiological warning system, wireless, privacy-preserving data collection, communication networks

I. INTRODUCTION

The COVID-19 pandemic has clearly shown that, in many senses, the world as a whole was not prepared for dealing with a disease of such magnitude. Although a final assessment is infeasible at this point (April, 2020), the current statistics available about COVID-19 indicate that the most successful policies for monitoring and controlling the virus propagation are employing various digital technologies and connectivity [1], [2]. These technologies can be used in two principal ways: (1) to provide status and predictions of the epidemiological spread and facilitate actions, such as administration of diagnostic tests or preparation of medical equipment; (2) to implement active policies that facilitate societal processes and safe citizen movement, such as shop schedules, organized delivery of goods, and similar. The success of such informed policies is also strongly related to the *timing of their implementation* for the respective place, as virus contagion is a spatiotemporal phenomenon that usually leads to exponential growth in the number of infected individuals [3].

Unfortunately, even the most successful timely interventions are in some way overreaching, shutting down abruptly most of the economic activities and creating social distancing between

citizens. While acknowledging that this is our current societal need to contain the propagation of this pandemic, our main hypothesis is that, in the near future, digital technologies and wireless connectivity can enable a *graceful lockdown*. This will be achieved based on ubiquitous and privacy-preserving data acquisition, explainable predictive methods for contagion risk assessment, and digitized policies for selective lockdown and scheduling of societal activities. It is foreseen that an effective global solution shall be decentralized, yet hierarchical (system-of-systems), to provide fine-tuned coordination among the federated entities to orchestrate actions and then avoid both the human rights cost of overly stringent solutions [4] and the human life cost of “hands-off” approaches. Beyond more traditional centralized policy-making, such a decentralized platform should also provide citizens with tools for more direct participation in mitigation measures based on tailored individual incentives.

The proposed platform aims at achieving the minimal use of personal data, processed with the maximum possible privacy-preservation to eliminate global health risks, while protecting the economy [5]. The COVID-19 pandemic has demonstrated that nations and governments need to be better prepared and coordinated to detect and react to global threats focusing on *early detection, early response*, moving beyond traditionally passive disease monitoring based on voluntary reporting systems. Hence, a rapid, participatory responsive detection system must be in place to support public health officials. A synergistic relationship with public health officials, policymakers, and citizens’ active participation will be critical to align the mandate for public health with the protection of privacy, freedom and democracy [6], [7]. In this context, one main factor is to design a special set of incentives that would allow the citizens to provide secured anonymized access to their data while actively participating in the crowd platform to support early disease detection, a public information system, and possible mitigation measures. Building up such an incentive structure that maps those trade-offs is a key aspect (beyond purely technical ones) for the success of the proposed federated architecture. Furthermore, this system should be constructed and evaluated rigorously following ethical guidelines as in [2], [8].

DC, PHJN and MU are with LUT University, Finland. EP is with University of Leeds, Leeds, UK, PM is with University of Waterloo, Canada, DZR is with Federal University of Lavras, Brazil, MD, HS and IM are with Trinity College Dublin, Ireland, AJ is with Aalto University, Finland, LHD is with University of Vermont, USA, GF is with State University of Campinas, Brazil, PP is with Aalborg University, Denmark. This paper is partly supported by Academy of Finland via ee-IoT project n.319009. Contact: pedro.nardelli@lut.fi.

II. EPIDEMIOLOGICAL MODELS AND DATA COLLECTION

Biomedical data alone do not contain the information required for preventing and mitigating pandemics. In a highly interconnected globalized society, social interactions, environmental data, spatiotemporal events, and collective nonlinear phenomena observed on complex infrastructures such as transport systems, require the mining of heterogeneous pandemic big data that are a result of a complex system of systems processes. In fact, modern disease monitoring systems, as the Global Public Health Intelligence Network (GPHIN) [9] and the Global Outbreak Alert and Response Network (GOARN) [10], and event-based approaches that use a combination of web crawlers, artificial intelligence, and public health expertise in the detection of indicators [11], [12] provide good—although limited—examples of successful platforms to identify latent indicators of an outbreak.

To go beyond those existing solutions, data scale, contextualization, and granularity are key requirements for data quality, which is usually orthogonal to privacy preservation, timely processing and analysis as well as storage and processing cost. In the following, more details will be provided about existing epidemiological models and epidemic-related data collection.

A. Models and their limitations

Models for emerging epidemics come in various forms, all relying on differently coarse-grained individuals' data. State-of-the-art population-level models for COVID-19 rely on metapopulation models with an underlying Susceptible-Exposed-Infectious-Recovered disease dynamics [13]. The metapopulations represent regions surrounding travel hubs and are interconnected through mobility data, for example, data from airlines, public transportation systems, and traffic control systems. The disease dynamics itself accounts for the fact that exposed individuals can be presymptomatic and unaware that they infect others, as suggested by early case data [14]. Unfortunately, while this model is effective at providing large-scale forecasts and predicting importations at aggregated levels, it is less informative at the individual level. However, this more granular information is necessary to accurately estimate the probability of such epidemics given a single importation event, which is rare given the population size. This particular event has the potential to become the initial point of a widespread contagion in a region not reached by epidemics until that point. Since this event basically depends on one individual, aggregated level statistics provide poor description of the heterogeneity of individual level behavior and mobility.

Given this, most governments around the world are introducing *stay-at-home* policies to reduce the mobility of their citizens, which should result in the now famous *flattening of the curve*. This expression refers to the process of slowing down the virus spread to keep the need for hospital care below the health care capacity. Different variations of the *stay-at-home* policy have been introduced, e.g. allowing citizens to move in a predefined radius to walk with their pets and exercise, restricting the number of times per day or the number of people per household that can go out. As previously argued, this is, understandably, a necessary abrupt solution for now,

but it can become more graceful by properly utilizing crowd-sourced data and active citizens' participation.

B. Data collection: current activities and potential existing sources

At higher levels, mining existing data sets collected by different types of service providers is an obvious and powerful way to quantify to which extent citizens comply with the *stay-at-home* policy, thereby providing an essential means of assessing their effectiveness. For example, smart grid systems can provide information about the changes in the energy consumption in different areas suggesting that people, on average, spend more time at home. The network usage information from broadband operators can also be used as an indicator of change in human behavior related to their mobility. Another source of information are mobility data sets from cellular network operators that are already present on the network. Smart appliances, like smart TVs, smart fridges, and smart light bulbs can be used to provide information about the overall time spent at home versus the time spent outside. Finally, information collected by several smart city applications, such as public transport usage and traffic, could also be leveraged to estimate the change in mobility patterns.

At a lower granularity, proximity-tracing platforms have been proposed to collect data at the individual level, for instance the Pan-European Privacy-Preserving Proximity Tracing (PEPP-PT) [15], often in partnership with mobile service providers and operators. By aggregating proximity data over time, these approaches can follow data protection regulations. Among others, two of the largest Big Tech companies Apple and Google proposed a joint solution that appears to focus on the protection of user privacy, for example by keeping people anonymous in the central servers and making data submission voluntary. Different technologies are being proposed to automate and extend contact tracing (e.g. shared databases, GPS traces of confirmed cases, contact tracing through Bluetooth). The use of Bluetooth-based solutions, either in the form of the third-party dedicated apps or as a feature built into the mobile devices' operating systems, seems to be the most promising solution to date. Although these solutions are a new source of granular data, they cannot cover all relevant forms of infection, namely asymptomatic infections, and infections occurring out of proximity through shared surfaces (e.g. doorknobs). In this sense, our hope is to combine the metapopulation and individual tracing models in a citizen science framework to consider more possible routes of infection as well as both proximity (e.g. close contact) and mobility data (e.g. location visited). In addition to this, machine learning methods are expected to combine expert knowledge (e.g. from virologists and sociologists) with global data collected from different sources for fitting powerful predictive models to high-dimensional data. In doing so, the target is to leverage the different strengths of different data sources at different granularity for privacy protection and allow for more thorough probabilistic forecasts. Ultimately this will allow for graceful lockdowns, tailored incentives for citizens' participation, and fine-tuned legitimate interventions.

C. Data privacy and trustworthiness

Fine-grained data collection approaches raise two major concerns as highlighted in [16]: (1) user privacy; and (2) trustworthiness of the shared information. Clearly, sharing location and health information with third-party entities can lead to the misuse of the data set for instance through unwanted advertising and health insurance implications. The impact of fake health information on the contact tracing system should be analyzed to assess its implications in terms of access to testing. All these applications rely on the premise that users will trust and adopt these applications. This brings us to an additional problem that could hamper these approaches. The adoption of new technologies is hard to predict, and it might introduce bias for example related to age, wealth, level of education, and state of country development. Citizens' participation in warning and response of epidemic outbreaks requires new incentives that reward the responsible use of citizens' personal data for protecting public health, while penalize and prevent citizens' profiling actions, manipulative nudging and power misuse.

D. Challenges and the way forward

Designing new data fusion schemes tailored to fuel data analytics processes for prevention and mitigation of pandemics is an open research question: Which smart phone sensor data can model compliance of *stay-at-home* policies? How such models can be enhanced with social media, smart grid, or transport data? Pandemics require that we revisit and potentially reinvent how data should be managed and how systems should be designed to manage data in a more responsible way [8]. More specifically, discovering new ways to turn private sensitive data of citizens into public good, while preventing massive surveillance, profiling, and discriminatory analytics, becomes imperative. To pioneer this capability, new transparent and accountable data commons [17], for instance personal data stores [18]–[20], to which citizens have incentives to contribute data, are required. The citizens retain the right for their data to be "forgotten" [21]. Ultimately, they have control and give their consent to how these data are used. These data commons are designed to maximize the use of techniques for privacy preservation [5], [22], for example homomorphic encryption, differential privacy, anonymity, and obfuscation, while citizens are made aware of the privacy risks they experience.

III. WIRELESS CONNECTIVITY AND EPIDEMIC-RELEVANT INFORMATION

The previous section covered different aspects related to the data required to support a fine-grained epidemiological model. It is argued that a ubiquitous system can be employed to manage such data in a privacy-preserving manner. Here, we identify wireless connectivity and Internet of Things (IoT) devices as a way to collect data following those principles.

Wireless connectivity and IoT devices come in many flavors, but can, in general, be classified into (1) personal devices, such as mobile phones and earbuds; and (2) unattended connected machines and things, such as surveillance cameras and motion

sensors. There are three principal sources of epidemic-relevant data acquired through wireless connectivity: (1) online social networks; (2) personal smart phone and mobile data; and (3) sensory and Internet of Things (IoT) devices.

A. Online social networks

In critical events, people tend to use online social networks to post comments about the emergency and learn from other users' comments. As a result, online social networks become a rich source of diverse information that could help to understand the main characteristics of the crises and their potential magnitude in early stages. For instance, in [23], authors collected comments posted in the Chinese social media channel Weibo during the first month of the COVID-19 epidemic in Wuhan (China) to understand the evolution trend of the public opinion. A similar approach was used on Twitter in [24]. Government entities can use this approach to give more attention to the needs of the public during the beginning of the epidemic and adjust their responses accordingly. Another form of data content, normally provided by explicit consent from the citizens, is the GPS location data from personal devices. At the application level, the Twitter platform offers various user information, geographical location being one of the most important for a wide range of applications. However, only a few users enable their geolocation as public. To discover the user's geolocation (geographical region or exact geocoordinates), there are mainly two approaches called content-based and network-based algorithms. The first one uses textual contents from tweets, and the latter uses the connections and interactions between users. In [25], a neural network model based on multiview learning by combining knowledge from both user-generated content and network interaction is proposed to infer users' locations.

B. Personal smart phone devices

This source of information comes in many different forms, ranging from the use of metadata in cellular mobile networks for user localization up to the metadata associated with different applications, such as Twitter. The use of data¹ collected in mobile networks during the COVID-19 pandemic has received significant attention [26]. Agglomerated mobile network data can be used to verify if interventions, such as school closures, are effective, and help to understand the geographical spread of an epidemic. Different types of mobile network metadata are collected at different levels of the communication system, and they offer varying levels of information. At the lowest level, localization of mobile network users is possible by evaluating the strength of the wireless signal received at base stations. Depending on how many base stations are connected to the device, the location can be determined at the level of entire cells or down to a few meters using triangulation.

The resolution offered by cellular mobile localization maybe too coarse for detecting citizens' proximity and potential

¹They are referred to as metadata due to its role played in the communication system to which it belongs; here we will also refer to it as a "data" when there is no danger of ambiguity.

infections by face-to-face interactions. Nevertheless, even low-resolution location data can give insight into behavioral patterns (how much time is spent at home, office, and events) of individuals. Proximity detection can be enhanced by using the metadata from Bluetooth devices, such as beacons and discovery messages (e.g., [27]). Wi-Fi is also a technology that is used ubiquitously, and its metadata can provide information on user proximity. For example, proximity can be inferred by comparing the lists of access points that each device can see within a given time interval [28].

Fusing the metadata from these different sources, along with the context information (how many family members are at home, which events) is very relevant for monitoring the epidemic.

C. Sensory and IoT devices

Many applications and devices become an important tool to provide epidemic-relevant data. These can be related to surveillance/thermal cameras, drones and even wearable devices. For example, surveillance cameras can be employed to count the number of people entering and exiting a specific area. In other scenarios, thermal cameras are used with specialized settings to focus on human skin temperature range, and infrared spot sensors for individual temperature scanning. A similar application to identify individual or group activities in a given place is based on motion sensors as infrared lasers, or ultrasonic sensors. In this approach, the motion sensors are installed and deployed in a specific environment to recognize different human activities.

Other important source of IoT data are the smart health care devices, for instance, remote monitoring systems to check body temperature, which is a key sign in the support of homeostasis. Other popular applications are oxygen saturation monitoring based on beat oximetry, electrocardiogram monitoring with a specialized framework to estimate the heart rate, elderly monitoring using doppler radar to identify risk movements of elderly people, sugar level monitoring, and blood pressure monitoring. These IoT devices provide information that is analyzed by the IoT data to support specific applications.

IV. FEDERATED GLOBAL EPIDEMIOLOGICAL WARNING SYSTEM

We envision a federated and decentralized coordination system for epidemiological warning and response. This system is federated by citizen communities that crowd-source data, personal smart phone devices, community-level IoT devices (e.g. LoRaWAN networks) and other computational resources on the edge. It can scale organically and bottom-up to city-level, national-level and ultimately at global-level to coordinate in a socially responsible way the international actions of public health organizations and governments. This scaling requires tailored incentives that align public health policy-making with citizens' privacy and autonomy. Distributed ledgers with secure crypto-economic incentive models running on edge servers and scaling-up on-demand at a global level using wireless connectivity and public cloud infrastructure are the means to support the federated nature of this proposed system. Figure 1 illustrates the our vision.

A. Key components

The proposed federated global epidemiological warning and response system is built upon epidemic-relevant information obtained through wireless connectivity introduced in Section III, and the internet infrastructure including fiber-optic and satellite links. Thus, in addition to the devices that are the data sources (e.g., smart phones, wearables, smart appliances, and cameras) with their applications and existing communication networks, there is a need for a computer infrastructure dedicated to store and process the epidemic-relevant data in a secure and privacy-preserving manner. It is argued that this infrastructure also needs to follow a similar federated organization based on the geographic locations. Each municipality, city, county, or larger neighborhood will rely on edge servers to process their respective data contents. The premise is to keep the computations as local as possible.

However, larger-scale computations related to the interrelations between locations (e.g. mobility, commuting, and traveling) requires data from these different places. Once again, the federated organization supports this collaborative sharing (up and downstream between the federated entities), but it will probably require computationally more demanding algorithms. It is possible that those computational tasks require more powerful cloud servers, or collaborative parallel computing at the edges. For security reasons, Distributed Ledger Technologies shall be employed to store data from the edge servers at the global level. At the technical level, the following aspects deserve special attention.

1) *Privacy*: In this paper, a privacy-by-design approach is proposed with which several critical operations of the federated global epidemiological system can be performed; for instance, decentralized data analytics [29], [30], social interactions analysis [31], [32], decentralized planning and resource allocation [33] and federated learning [34], [35]. Such operations integrate in a smart way state-of-the-art techniques, for instance, informational self-determination, homomorphic encryption, differential privacy, obfuscation, and anonymity. Often, privacy may limit the quality of service known as the *system utility* as the accuracy and quality of data are deteriorated to hide information content. For instance, predicting the risk of infections for individuals based on an epidemiological network model is a graph-based semi-supervised learning problem. Privacy-preserving semisupervised learning over graphs has been considered in [36]. However, the precise trade-off between (lowering of) privacy protection and learning accuracy is largely open. Pareto optimal trade-offs can be configured and regulated by tuning the parameters of the privacy techniques as previously shown in [5]. Monetary and other incentives can be used to coordinate data sharing choices in a crowd. However, new types of social incentives are required for such an epidemiological system; for instance, incentives related to well-being, receiving solidarity, and long-term payoffs.

2) *Interoperability*: The big data required by the proposed solution have to be interoperable, i.e. the several applications that are providing data to be used to accomplish a specific task have to operate together and share a common "understand-

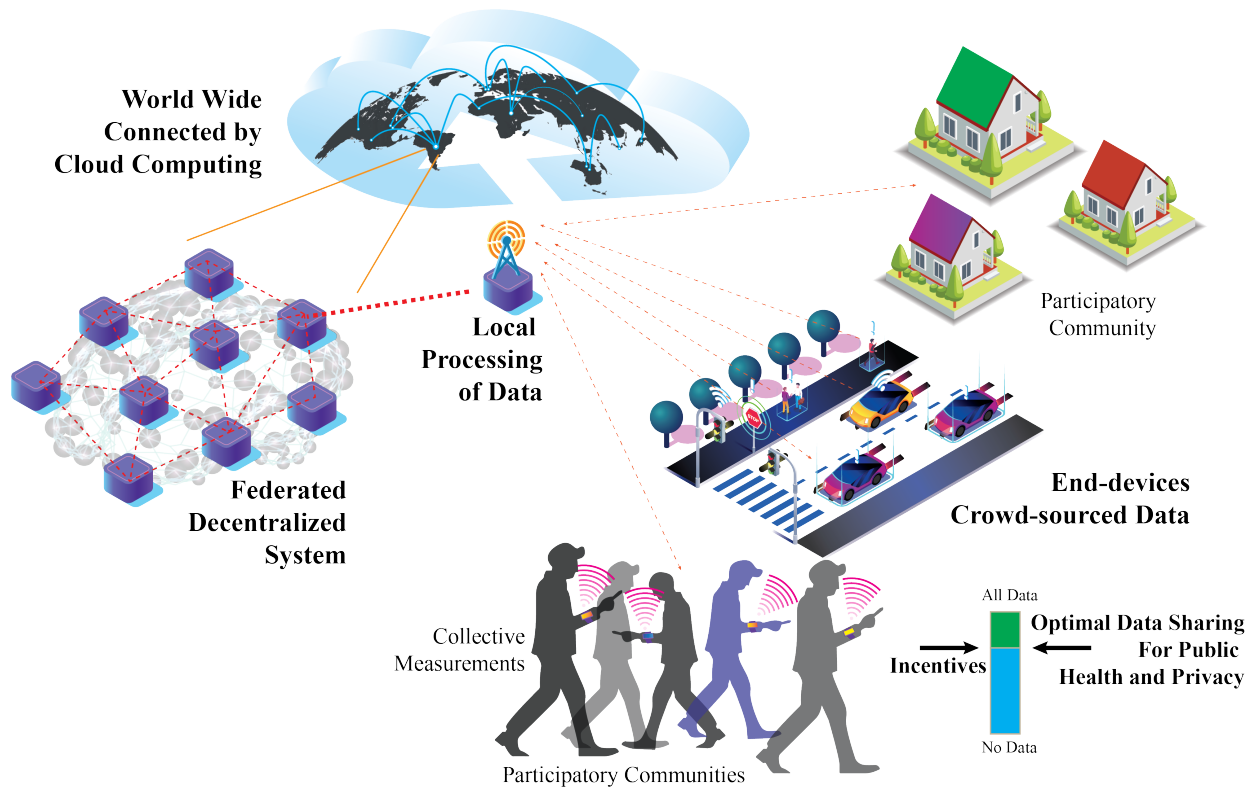


Fig. 1. A federated ubiquitous systems for epidemiological warning and response. An organic and bottom-up scaling at global level is envisioned based on active citizens' participation. Decentralized privacy-preserving computations are performed from the the edge to the cloud based on crowd-sourced obfuscated and anonymized data managed with distributed ledgers to empower trust. Incentive mechanisms for responsible data sharing align the public health mandate with citizens' privacy and autonomy.

ing" [37]. This can be achieved by employing standards for health informatics such as ISO TS22220:2011, TS21090:2011, or TS13606. For instance, the data collected from different sources can be used to predict refinements to patient care or new drug contraindication [38]. This key issue, though, has deserved little attention in large-scale epidemiological studies [39], [40]; it is usually assumed that heterogeneous data sources are compatible with each other. In practice, though, the highly heterogeneous data sources lead to poor interoperability, which creates barriers to effectively combat pandemics like COVID-19, as indicated by [41].

3) *User interfaces*: A successful platform also involves a suitable end-user interface [42]. In this sense, data consumption by public health officials and global health agencies will require user-friendly web-based interfaces, using common dashboarding techniques. However, beyond this, the proposed federated platform has to consider citizen participation and the heterogeneity of end-users. Therefore, the following characteristics should be taken into consideration in the design of the platform: it should be (a) **explainable/accountable** to improve for instance awareness and engagement [43]; (b) **gamified** to engage and incentivize participation [44], [45]; and (c) **customizable** for different user groups at an international level.

B. Proposed architecture

The proposed architecture has to articulate the key elements from data acquisition to analytics, following the best practices related to privacy and cyber-security. At the acquisition level, in addition to the existing data retrieval from the web, wireless-connected devices will send data to edge nodes that will be associated with specific regions. Data will be anonymously preprocessed at the edge, including some intelligent detection of anomalies or event detection, and then sent to regional (cloud) servers to run a more complete model that will fuse the geolocated timestamped data and run detection models based on explainable AI approaches combined with mathematical computational methods. In this case, crowd-sourcing models are applied based on hardware as a public good in an approach similar to the DIASPORA social network, also extending this approach to software and data. Regional models will be associated in *federations* resembling the governance structure of the actual regions under interest, which is customizable to local policies/governance models.

Following this federated and collaborative organization, it will be possible to build organically a *fine-tuned early detection* system at the global level in a decentralized yet hierarchical manner to support *graceful lockdowns*. This would turn the proposed federated architecture into a holorarchy, where each level is independent and self-sustained, but it can also be encapsulated at a level above to capture new goals [46].

This system will rely on Distributed Ledger Technologies

to guarantee a trustful system without a responsible third party and minimize the risk of data manipulation. Distributed ledgers could also be used to facilitate different (crypto-economic) incentive mechanisms, such as token curated registries. Its main objective is to build a crowd-sensing trustworthy platform based on privacy-preserving methods to detect potential harmful symptoms in a specific region in almost real-time and flag them to relevant authorities. This alarm needs to be accurate, reliable, and explainable. It will also require a user-tailored interface that could empower citizens by providing detailed explanations and easy-to-access information, while being a tool for policymakers (from the city level to global organizations) designing the correct interventions (a variety of incentives, sanctions, and other persuasive measures) given the specific context of the epidemic in some specific region in a given period of time.

V. DISCUSSION AND FUTURE PERSPECTIVE

Although the proposed architecture is designed to be fair among all citizens, few important practical points still deserve attention. Good quality data usually follow the wealth distribution over all scales of the proposed federation. In other words, good quality data is very likely to be less available in poorer regions of a city. The same is valid for poor regions within a country and poor countries. This existing gap needs to be considered when building incentives for participation, otherwise the proposed solution has the potential of reinforcing inequality [47].

Another current issue we cannot neglect is the public perception of wireless technologies. In this sense, part of the general public perceives IoT as insecure [48] and fifth generation (5G) as a health hazard [49] to the point of claiming an astonishing causal relation between 5G and COVID-19, which has caused destruction of base stations across UK. Beyond these conspiracy theories that are hard to combat, there exist legitimate concerns of anonymized data not being anonymous [50], and of novel surveillance techniques introduced in times of crisis that are maintained for monitoring (legally or illegally) populations, e.g. post-9/11 surveillance in the USA [51].

All these have to be carefully addressed from the early stages of the system development. Furthermore, it has to be clear that neither *technology* nor *data* can prevent another outbreak on their own, but can only provide the extremely valuable tools to enable the holy grail of controlling epidemics: *early detection*, *early response* along all relevant actors within the federated organization. In other words, technology and data detect and identify potential harms and suggest actions and reactions, but the final diagnostic and further interventions are due to the responsible institutions within the federated structure and citizens' active participation. Our proposal answers those challenges based on informational self-determination as the way to build trustworthy and secure public infrastructure that shall enable *graceful lockdowns* as advocated here.

REFERENCES

- [1] D. S. W. Ting, L. Carin, V. Dzau, and T. Y. Wong, "Digital technology and covid-19," *Nature Medicine*, pp. 1–3, 2020.
- [2] M. Ienca and E. Vayena, "On the responsible use of digital data to tackle the covid-19 pandemic," *Nature Medicine*, pp. 1–2, 2020.
- [3] D. Adam, "Special report: The simulations driving the world's response to covid-19," *Nature*, 2020.
- [4] K. Kupferschmidt and J. Cohen, "Can china's covid-19 strategy work elsewhere?" 2020.
- [5] T. Asikis and E. Pournaras, "Optimization of privacy-utility trade-offs under informational self-determination," *Future Generation Computer Systems*, 2018.
- [6] D. Helbing, B. S. Frey, G. Gigerenzer, E. Hafen, M. Hagner, Y. Hofstetter, J. Van Den Hoven, R. V. Zicari, and A. Zwitter, "Will democracy survive big data and artificial intelligence?" in *Towards digital enlightenment*. Springer, 2019, pp. 73–98.
- [7] F. Amat, A. Arenas, A. Falcó-Gimeno, and J. Muñoz, "Pandemics meet democracy. experimental evidence from the covid-19 crisis in spain," 2020.
- [8] T. Hagendorff, "The ethics of ai ethics: An evaluation of guidelines," *Minds and Machines*, pp. 1–22, 2020.
- [9] E. Mykhalovskiy and L. Weir, "The global public health intelligence network and early warning outbreak detection," *Canadian journal of public health*, vol. 97, no. 1, pp. 42–44, 2006.
- [10] S. E. Davies and J. R. Youde, "Gphin, gone" the role of the world health organization in global disease surveillance and response," in *The Politics of surveillance and response To disease outbreaks*. Routledge, 2016, pp. 119–132.
- [11] D. Carter, M. Stojanovic *et al.*, "Revitalizing the global public health intelligence network (gphin)," *Online journal of public health informatics*, vol. 10, no. 1, 2018.
- [12] R. Bernard, G. Bowsler, C. Milner, P. Boyle, P. Patel, and R. Sullivan, "Intelligence and global health: assessing the role of open source and social media intelligence analysis in infectious disease outbreaks," *Journal of Public Health*, vol. 26, no. 5, pp. 509–514, 2018.
- [13] M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. P. y Piontti, K. Mu, L. Rossi, K. Sun *et al.*, "The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak," *Science*, 2020.
- [14] W. E. Wei, "Presymptomatic transmission of sars-cov-2—singapore, january 23–march 16, 2020," *MMWR. Morbidity and Mortality Weekly Report*, vol. 69, 2020.
- [15] Pan-european privacy-preserving proximity tracing. [Online]. Available: <https://www.pepp-pt.org/>
- [16] H. Cho, D. Ippolito, and Y. W. Yu, "Contact tracing mobile apps for covid-19: Privacy considerations and related trade-offs," *arXiv preprint arXiv:2003.11511*, 2020.
- [17] S. A. Wright, "Privacy in iot blockchains: with big data comes big responsibility," in *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019, pp. 5282–5291.
- [18] M. Alessi, A. Camillo, E. Giangreco, M. Matera, S. Pino, and D. Storelli, "Make users own their data: A decentralized personal data store prototype based on ethereum and ipfs," in *2018 3rd International Conference on Smart and Sustainable Technologies (SpliTech)*. IEEE, 2018, pp. 1–7.
- [19] M. J. M. Chowdhury, A. Colman, M. A. Kabir, J. Han, and P. Sarda, "Blockchain as a notarization service for data sharing with personal data store," in *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (Trust-Com/BigDataSE)*. IEEE, 2018, pp. 1330–1335.
- [20] N. B. Truong, K. Sun, and Y. Guo, "Blockchain-based personal data management: From fiction to solution," in *2019 IEEE 18th International Symposium on Network Computing and Applications (NCA)*. IEEE, 2019, pp. 1–8.
- [21] J. Burkell, "The future of privacy lies in forgetting the past," *Eur. Data Prot. L. Rev.*, vol. 3, p. 435, 2017.
- [22] I. J. Vergara-Laurens, L. G. Jaimes, and M. A. Labrador, "Privacy-preserving mechanisms for crowdsensing: Survey and research challenges," *IEEE Internet of Things Journal*, vol. 4, no. 4, pp. 855–869, 2016.
- [23] L. Li, Q. Zhang, X. Wang, J. Zhang, T. Wang, T. Gao, W. Duan, K. K. Tsoi, and F. Wang, "Characterizing the propagation of situational information in social media during covid-19 epidemic: A case study on weibo," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 2, pp. 556–562, 2020.
- [24] T. Alshaabi, J. Minot, M. Arnold, J. L. Adams, D. R. Dewhurst, A. J. Reagan, R. Muhamad, C. M. Danforth, and P. S. Dodds, "How the world's collective attention is being paid to a pandemic: Covid-19

- related 1-gram time series for 24 languages on twitter,” *arXiv preprint arXiv:2003.12614*, 2020.
- [25] T. H. Do, D. M. Nguyen, E. Tsiligiani, B. Cornelis, and N. Deligiannis, “Twitter user geolocation using deep multiview learning,” in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 6304–6308.
- [26] N. Oliver, E. Letouzé, H. Sterly, S. Delataille, M. De Nadai, B. Lepri, R. Lambiotte, R. Benjamins, C. Cattuto, V. Colizza, N. de Cordes, S. P. Fraiberger, T. Koebe, S. Lehmann, J. Murillo, A. Pentland, P. N. Pham, F. Pivetta, A. A. Salah, J. Saramäki, S. V. Scarpino, M. Tizzoni, S. Verhulst, and P. Vinck, “Mobile phone data and COVID-19: Missing an opportunity?” *arXiv e-prints*, p. arXiv:2003.12347, Mar. 2020.
- [27] L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, and C. Fraser, “Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing,” *Science*, 2020.
- [28] P. Sapiezynski, A. Stopczynski, D. K. Wind, J. Leskovec, and S. Lehmann, “Inferring person-to-person proximity using wifi signals,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 2, pp. 1–20, 2017.
- [29] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth, “Practical secure aggregation for privacy-preserving machine learning,” in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017, pp. 1175–1191.
- [30] E. Pournaras, J. Nikolic, A. Omerzel, and D. Helbing, “Engineering democratization in internet of things data analytics,” in *2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA)*. IEEE, 2017, pp. 994–1003.
- [31] F. Musciotto, S. Delpriori, P. Castagno, and E. Pournaras, “Mining social interactions in privacy-preserving temporal networks,” in *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 2016, pp. 1103–1110.
- [32] L. Bahri, B. Carminati, and E. Ferrari, “Decentralized privacy preserving services for online social networks,” *Online Social Networks and Media*, vol. 6, pp. 18–25, 2018.
- [33] E. Pournaras, P. Pilgerstorfer, and T. Asikis, “Decentralized collective learning for self-managed sharing economies,” *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, vol. 13, no. 2, pp. 1–33, 2018.
- [34] Q. Yang, Y. Liu, T. Chen, and Y. Tong, “Federated machine learning: Concept and applications,” *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 2, pp. 1–19, 2019.
- [35] V. Smith, C.-K. Chiang, M. Sanjabi, and A. S. Talwalkar, “Federated multi-task learning,” in *Advances in Neural Information Processing Systems*, 2017, pp. 4424–4434.
- [36] H. Arai and J. Sakuma, “Privacy preserving semi-supervised learning for labeled graphs,” in *ECML PKDD 2011. Lecture Notes in Computer Science, vol 6911*. Springer, Berlin, 2011.
- [37] M. Jacoby, A. Antonic, K. Kreiner, R. Łapacz, and J. Pielorz, “Semantic interoperability as key to IoT platform federation,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10218 LNCS, no. March, pp. 3–19, 2017.
- [38] S. Schulz, R. Stegwee, and C. Chronaki, “Fundamentals of Clinical Data Science,” *Fundamentals of Clinical Data Science*, no. January, 2019.
- [39] M. Roos, E. L. Martin, and M. D. Wilkinson, “Preparing data at the source to foster interoperability across rare disease resources,” in *Rare Diseases Epidemiology: Update and Overview*. Springer, 2017, pp. 165–179.
- [40] J. Moreira, M. J. van Sinderen, L. F. Pires, and P. D. Costa, “Improving semantic interoperability of big data for epidemiological surveillance,” in *International Conference on Interoperability for Enterprise Systems and Applications, I-ESA 2016*. International Society for Technology in Education, 2016, pp. 314–324.
- [41] H. Legido-Quigley, N. Asgari, Y. Y. Teo, G. M. Leung, H. Oshitani, K. Fukuda, A. R. Cook, L. Y. Hsu, K. Shibuya, and D. Heymann, “Are high-performing health systems resilient against the covid-19 epidemic?” *The Lancet*, vol. 395, no. 10227, pp. 848–850, 2020.
- [42] R. Maciejewski, P. Livengood, S. Rudolph, T. F. Collins, D. S. Ebert, R. T. Brigantic, C. D. Corley, G. A. Muller, and S. W. Sanders, “A pandemic influenza modeling and visualization tool,” *Journal of Visual Languages & Computing*, vol. 22, no. 4, pp. 268–278, 2011.
- [43] L. D. Geneviève, A. Martani, T. Wangmo, D. Paolotti, C. Koppeschaar, C. Kjelsø, C. Guerrisi, M. Hirsch, O. Woolley-Meza, P. Lukowicz *et al.*, “Participatory disease surveillance systems: ethical framework,” *Journal of medical Internet research*, vol. 21, no. 5, p. e12273, 2019.
- [44] O. L. Neto, O. Cruz, J. Albuquerque, M. N. de Sousa, M. Smolinski, E. Â. P. Cesse, M. Libel, and W. V. de Souza, “Participatory surveillance based on crowdsourcing during the rio 2016 olympic games using the guardians of health platform: Descriptive study,” *JMIR Public Health and Surveillance*, vol. 6, no. 2, p. e16119, 2020.
- [45] K. Wazny, “Applications of crowdsourcing in health: an overview,” *Journal of global health*, vol. 8, no. 1, 2018.
- [46] A. Diaconescu, B. Porter, R. Rodrigues, and E. Pournaras, “Hierarchical self-awareness and authority for scalable self-integrating systems,” in *2018 IEEE 3rd International Workshops on Foundations and Applications of Self* Systems (FAS* W)*. IEEE, 2018, pp. 168–175.
- [47] M. A. Gianfrancesco, S. Tamang, J. Yazdany, and G. Schmajuk, “Potential biases in machine learning algorithms using electronic health record data,” *JAMA internal medicine*, vol. 178, no. 11, pp. 1544–1547, 2018.
- [48] A. Zubiaga, R. Procter, and C. Maple, “A longitudinal analysis of the public perception of the opportunities and challenges of the internet of things,” *PLoS one*, vol. 13, no. 12, 2018.
- [49] S. F. Hasan, “The 5g debate in new zealand—government actions and public perception,” in *International Conference on Ad Hoc Networks*. Springer, 2019, pp. 3–12.
- [50] I. S. Rubinstein and W. Hartzog, “Anonymization and risk,” *Wash. L. Rev.*, vol. 91, p. 703, 2016.
- [51] S. Landau, “Making sense from snowden: What’s significant in the nsa surveillance revelations,” *IEEE Security & Privacy*, vol. 11, no. 4, pp. 54–63, 2013.