

# Challenges and Perspectives of Open Data in Modelling Infectious Diseases

Francesco Branda <sup>1,\*</sup>  and Giorgia Lodi <sup>2,†</sup>

<sup>1</sup> Department of Computer Science, Modeling, Electronics and Systems Engineering (DIMES), University of Calabria, 87036 Rende, Italy

<sup>2</sup> Institute of Cognitive Sciences and Technologies, Italian National Council of Research (CNR), 00185 Rome, Italy

\* Correspondence: francesco.branda@unical.it

† These authors contributed equally to this work.

The pandemic challenged the scientific community and governments around the world, who were looking for real-time answers but lacked the data or evidence to guide decision-making. This inevitably led to diverse and fragmented policies and responses. In this context, a positive note was the need, recognized by all, for an acceleration of several innovations that had already been introduced but had not yet taken hold: from decentralized clinical trials and digitization, to new ways of leveraging real-world data (RWD). However, RWDs is only the starting point: once the raw data is collected, it is necessary to structure all the information, thus obtaining Real-World Evidence (RWE); this was later defined in 2016 by the Food and Drug Administration (FDA) as “the clinical evidence relating to the use and potential benefits, or risks of a medical product derived from the analysis of Real World Data” [1].

At the beginning of the pandemic, the scientific community knew little to nothing about SARS-CoV-2 and a patient’s clinical course. For instance, RWDs showed that steroids or anticoagulants, administered at the right time in the progression of the disease, would have significantly improved a patient’s health condition. They were also instrumental for vaccines, which were developed rapidly and came to market through emergency approvals based on limited clinical data. As more people received administration, a growing body of evidence suggested that these vaccines were actually safe and effective [2].

Epidemiologists and public health authorities are used to take critical decisions based on available evidence, and many of the prevention measures implemented during the pandemic (e.g., social distancing and frequent handwashing) were based on RWD. In this context, it was demonstrated how important open data—the usefulness of which has increased in recent years in all fields—is for informing health policy makers, taking better decisions, and improving clinical trials. The term open data refers to any non-personal data (e.g., categorical data, numerical data, etc.) that can be freely used, re-used, re-distributed, made accessible, and available in an open, standardized format that can be processed by machines to facilitate data consultation and incentivize data reuse, according to the rights and obligations defined in specific open licenses [3].

In general, in an epidemiological context, open data are crucial for (i) conducting real-time situational analysis, implementing predictive models that can provide effective and timely responses for the effective containment of a disease [4]; (ii) estimating key epidemiological parameters, such as incubation period, reproduction numbers, etc. [5]; (iii) ensuring community trust in the government through increased transparency and better communication; and (iv) countering misinformation. Importantly, open data are not only valuable for improving drug development and enabling the confirmation of efficacy and safety of post-marketing products, but are necessary for the development of effective policies and guidelines. For example, the use of masks is a good example to demonstrate this. Without clear, supportive, transparent and open data that can show the



**Citation:** Branda, F.; Lodi, G.

Challenges and Perspectives of Open Data in Modelling Infectious

Diseases. *Data* **2023**, *8*, 27.

<https://doi.org/10.3390/data8020027>

Received: 17 January 2023

Accepted: 23 January 2023

Published: 26 January 2023



**Copyright:** © 2023 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

effectiveness of face masks in preventing airborne disease transmission, the result can be non-acceptance, as we witnessed at the beginning of the pandemic, and which still persists in some geographical areas and within some groups of people.

However, for RWDs to play a key role in the epidemiological understanding of the origins and transmission dynamics of emerging infectious diseases, it will be critical to optimize data quality, further strengthen data analysis capacity in public health facilities, and develop new models for epidemiological surveillance, based on a range of traditional [6] and non-traditional approaches [7–10]. Several lessons have been learned from the COVID-19 pandemic that can be applied in the short term:

1. Clinical trials could be integrated with RWD on participating patients. Data capture in all electronic medical records would provide a more complete picture of patients beyond the data collected during the trial.
2. The medical community may want to emphasize the importance of RWDs in vaccine launches. In the United States, because COVID-19 vaccines were largely administered in community settings and were not linked to electronic medical record (EMR) systems, the ability to collect data was greatly hampered. Israel, on the other hand, was able to quickly and robustly track and report large amounts of data on vaccine safety and efficacy because the vaccines were administered through a centralized, nationwide effort with a substantial organizational, IT, and logistical infrastructure.
3. The need for greater diversity in clinical trials arises. Typically, incentives are given to recruit participants as quickly as possible and to reduce costs, not to ensure a diverse patient pool. The pharmaceutical industry needs to develop better ways to collect more high-quality RWD.
4. Drug development and health care delivery will be greatly improved when we are able to fully capture patient lifestyle data, known as “patient 360.” This includes collecting and aggregating information on how long they sleep, how much exercise they get, what groceries they buy, what environment they live in, and their attitudes toward factors affecting lifestyle and health. With these data, drug developers will become more knowledgeable, reduce the noise level and be able to better explain disease outcomes. It will also enrich our understanding of the diseases themselves, which will feed into new approaches to drug discovery and development.
5. Although rigorous quality controls exist for the collection, cleaning, and processing of clinical trial data, RWDs can be messy, as researchers depend on data acquired or reported by individuals. “Wearable devices” and biosensors offer exciting tools for additional data collection and enable real-time data capture in a way that was not previously possible. If these technologies could be incorporated more often into clinical trials, the additional data could potentially help explain certain responses to a given drug.
6. The affirmation of innovative artificial intelligence and semantic technologies can help to even more rapidly and massively analyze and reason about the products of the scientific community, thus accelerating incremental research results.

In addition, the intent to create collaborative data spaces such as those defined in the context of the European data strategy [11] and the American GovLab Data Collaboratives [12], where any data, not only open, flow seamlessly between the public- and private-sector entities, makes clear the need for a focused effort to increase collaboration and invest in our collective capacity to identify and understand public health risks in order to be better prepared for future pandemics and epidemics. Promoting transparency and data sharing is more important than ever in a global context where openness, reliability, and trustworthiness of data will be critical to accelerating the global response to health emergencies. The value of data to society is at a critical juncture, and its utility will improve the ability of public health institutions to synthesize contextual information for risk assessment and decision making.

**Author Contributions:** Conceptualization, F.B.; writing—original draft preparation, F.B.; writing—review and editing, G.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Mahendraratnam, N.; Mercon, K.; Gill, M.; Benzing, L.; McClellan, M.B. Understanding Use of Real-World Data and Real-World Evidence to Support Regulatory Decisions on Medical Product Effectiveness. *Clin. Pharmacol. Ther.* **2022**, *111*, 150–154. [[CrossRef](#)] [[PubMed](#)]
2. Branda, F. Impact of the additional/booster dose of COVID-19 vaccine against severe disease during the epidemic phase characterized by the predominance of the Omicron variant in Italy, November 2021–March 2022. *medRxiv* **2022**. [[CrossRef](#)]
3. Open Knowledge Foundation. “Open Definition”. Available online: <https://opendefinition.org/od/2.1/en/> (accessed on 16 January 2023).
4. Branda, F.; Abenavoli, L.; Pierini, M.; Mazzoli, S. Predicting the Spread of SARS-CoV-2 in Italian Regions: The Calabria Case Study, February 2020–March 2022. *Diseases* **2022**, *10*, 38. [[CrossRef](#)] [[PubMed](#)]
5. Branda, F.; Pierini, M.; Mazzoli, S. Monkeypox: Early estimation of basic reproduction number  $R_0$  in Europe. *J. Med. Virol.* **2023**, *95*, e28270. [[CrossRef](#)]
6. Lynfield, R.; Van Beneden, C.A.; M’ikanatha, N.M.; de Valk, H. *Infectious Disease Surveillance*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
7. Brownstein, J.S.; Freifeld, C.C.; Reis, B.Y.; Mandl, K.D. Surveillance Sans Frontieres: Internet-based emerging infectious disease intelligence and the HealthMap project. *PLoS Med.* **2008**, *5*, e151. [[CrossRef](#)] [[PubMed](#)]
8. Salathé, M. Digital epidemiology: What is it, and where is it going? *Life Sci. Soc. Policy* **2018**, *14*, 1–5. [[CrossRef](#)] [[PubMed](#)]
9. Kostkova, P.; Saigí-Rubió, F.; Eguia, H.; Borbolla, D.; Verschuuren, M.; Hamilton, C.; Azzopardi-Muscat, N.; Novillo-Ortiz, D. Data and digital solutions to support surveillance strategies in the context of the COVID-19 pandemic. *Front. Digit. Health* **2021**, *3*, 89. [[CrossRef](#)]
10. Morgan, O.W.; Abdelmalik, P.; Perez-Gutierrez, E.; Fall, I.S.; Kato, M.; Hamblion, E.; Matsui, T.; Nabeth, P.; Pebody, R.; Pukkila, J.; et al. How better pandemic and epidemic intelligence will prepare the world for future threats. *Nat. Med.* **2022**, *28*, 1526–1528. [[CrossRef](#)] [[PubMed](#)]
11. European Commission. The European Data Strategy—Shaping Digital Future. Available online: <https://digital-strategy.ec.europa.eu/en/policies/strategy-data> (accessed on 16 January 2023).
12. The GovLab. Data Collaboratives—Creating Public Value by Exchanging Data. Available online: <https://datacollaboratives.org/> (accessed on 16 January 2023).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.