Milestone 7.3 - Report

Industry sector mapping report

Aim: overviewing the industrial ecosystem of openly available COVID-19 data, its enablers and users as well as the core value chain provided by the BeYond-COVID (BY-COVID) project.

Authors: Lauer K.B., Smith A., Sousoni D., Westwick E., Martin C., Alloza Anguiano E., David R., Blomberg N.

BY-COVID Project Summary

BeYond-COVID (BY-COVID) aims to provide comprehensive open data and analysis tools on SARS-CoV-2, and other infectious diseases across scientific, medical, public health and policy domains. The project has a strong emphasis on mobilising raw viral sequences, helping to identify and monitor the spread of SARS-CoV-2 variants. It further accelerates access to and linking of data and metadata on SARS-CoV-2 and COVID-19 (e.g., public health data, data on patient outcomes), enables federated data analysis to conform with data protection regulations, and harmonisation and management of meta-data and sample-identifiers, as well as long-term cataloguing to ensure interoperability of national and global efforts.

Open data enabling an ecosystem for innovation and economic value creation in health-, bio-sciences and affiliated industries

Predictions see the market size of open data calculated as a percentage of Gross Domestic Product (GDP, total market value of all goods and services produced within a country's border) amounting to €199.51 - €334.20 billion in 2025 (Huyer and Van Knippenberg, 2020). Particularly for the life science and health sector, easily accessible and reusable open data fuel scientific discovery and the data-driven economy, through efficiency gains, cost savings, new products and services or job and venture creation. This has been confirmed through studies such as the quantitative assessment of the value of open databases using citations of these in patents (Abbasi *et al.*, 2016; Bousfield *et al.*, 2016a; Schulte-Althoff *et al.*, 2021).

Funded by the European Union

BY-COVID is funded by the European Union's Horizon Europe research and innovation programme under grant agreement number 101046203.

The role open science plays in stimulating innovation and creating long-term value for industry and academia has been recognised by the scientific community as well as governments and funders such as the European Commission, as stated in the Directive on Open Data and the Re-use of Public Sector Information: "*The vision underlying the Commission's strategy on open data and knowledge circulation is that information already paid for by the public purse should not be paid for again each time it is accessed or used, and that it should benefit European companies and citizens to the full. This means making publicly-funded scientific information available online, at no extra cost, to European researchers and citizens via sustainable e-infrastructures, also ensuring long-term access to avoid losing scientific information of unique value*" (Commission of the European Communities, 2009).

The benefits outlined in the above statement have most recently been demonstrated by the COVID-19 pandemic. An example of this is the mRNA-based COVID-19 vaccine developed by Pfizer and BioNTech, the first ever approved mRNA-based vaccine for usage in healthy humans (Polack, 2020). A phase 1/2/3 study to evaluate the safety, tolerability, immunogenicity, and efficacy or RNA vaccine candidates against COVID-19 in healthy individuals references the publicly available viral sequence in the European Nucleotide Archive with the sequence number MN908947.3 (BioNTech SE and Pfizer Inc., 2020).

The immense amount of openly shared data and the change in pace and regulations has enabled entrepreneurs and companies to innovate at an accelerated rate. Processes that would normally take months or years particularly on the regulatory and logistical front were decreased to weeks or months (Garud *et al.*, 2011; Gkeredakis *et al.*, 2021; van de Ven, 2017). Drawing on the example outlined above, it is important to understand who the players in the ecosystem are that, for example, contributed to the successful and timely delivery of vaccines against SARS-CoV-2? Moving beyond the boundaries of a pharmaceutical company, public health agencies, academic researchers, data infrastructure providers and public funders played an essential role.
In this report we will try to overview the industrial ecosystem of openly available COVID-19 data, its enablers and users as well as the core value chain provided by the BY-COVID project.

Summary of resources provided by BY-COVID - the BY-COVID Value Chain

As recently highlighted in a Eurosurveillance editorial on supporting better preparedness for future health emergencies *data fragmentation and non-interoperable information systems, limited integration of health information with other contextual information, little*

Funded by the European Union

BY-COVID is funded by the European Union's Horizon Europe research and innovation programme under grant agreement number 101046203.

standardisation of analytical tools and methods, difficulties in assessing the effectiveness of public health interventions in real time, sub-optimal use of data-driven insights for public health decision-making and a weak global trust architecture for data and information sharing have been identified as the main bottlenecks preventing effective disease surveillance (Morgan and Pebody,2022).

BY-COVID is a truly interdisciplinary project bringing together stakeholders from the biomedical field, hospitals, public health, social sciences and humanities addressing many of the above mentioned limitations which ultimately will increase European readiness for future pandemics, enhance surveillance and rapid-response capabilities.

BY-COVID serves as a foundation and demonstrator for interdisciplinary work across country borders. The outputs of the project will allow scientists across multiple domains, including SMEs and industry, to access a range of data that will generate new knowledge on infectious disease. On a more granular level the technical and scientific output of BY-COVID can be summarised in the value chain depicted in figure 1.

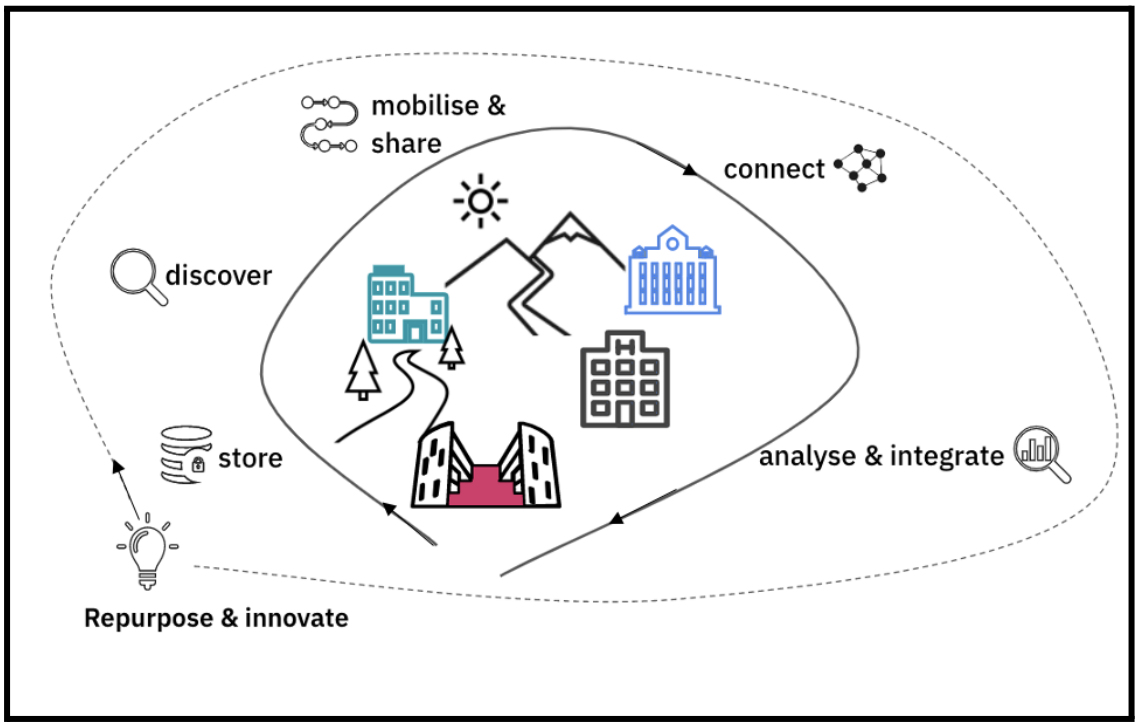

Figure 1: The data value chain provided through BY-COVID.

Core activities of the project include data storage, discovery, mobilisation, linkage, analysis, integration. This value chain enables innovation based on openly available data as well as research activities. In this report we will particularly focus on the industries building on, either currently or potentially in future, from resources provided through BY-COVID partners and project outcomes. Services provided under the BY-COVID project umbrella

Funded by the European Union

BY-COVID is funded by the European Union's Horizon Europe research and innovation programme under grant agreement number 101046203.

can broadly be categorised into data, training, data management services and software/tools/workflows.

Concrete resources for usage:

- COVID-19 Data Platform: A platform that enables the collection and comprehensive data sharing of available COVID-19 data.
- List of data resources : A comprehensive list of data resources provided under the BY-COVID umbrella (socio-economic, health, clinical and biomolecular data).
- Infectious diseases data management toolkit (IDTk): An online guide containing good data management practices applicable to research projects from the beginning to the end.

Data driven industries - potential users of BY-COVID resources

A review of secondary literature and conversations/interviews with opinion leaders such as the ELIXIR Industry Advisory Committee, as well as a survey shared through the project channels have revealed the following industry sectors as potential users of services under the BY-COVID umbrella, or providers of products/services that enable interaction with the main value chain for consortium members and other users (non exhaustive list):

- Diagnostics
- Pharmaceutical companies (e.g., Drug or vaccine development)
- Biotech-/bioinformatics companies (e.g. Data as a service DaaS platform providers, ontology providers...)
- Contract Research Organisation/Contract Manufacturing Organisation
- Consulting
- Medical Devices
- Education industry
- Publishers/media/journalists
- Hospitals/healthcare providers
- Hardware (e.g. sequencing machines) and other tech providers
- Cloud providers
- Software as a service SaaS providers e.g. data analysis software providers
- Compute power providers
- European/Global public-private partnerships, consortia & projects

Funded by the European Union

BY-COVID is funded by the European Union's Horizon Europe research and innovation programme under grant agreement number 101046203.

According to the open systems perspective, companies are often resource-constrained and rely on externally maintained resources, such as data, technology or code as part of their business model. Co-evolution of providers of those external resources is therefore essential. Innovation styles such as open innovation are particularly common in the life and health science industries and are partly fueled by publicly funded infrastructure as the basis for technological and business innovation ultimately driving economic performance (Yang and Yan, 2019; Lauer, Smith, Blomberg *et al.*, 2021). The open infrastructure provided through BY-COVID is an example of a positive driver in the open innovation ecosystem overviewed in figure 2.

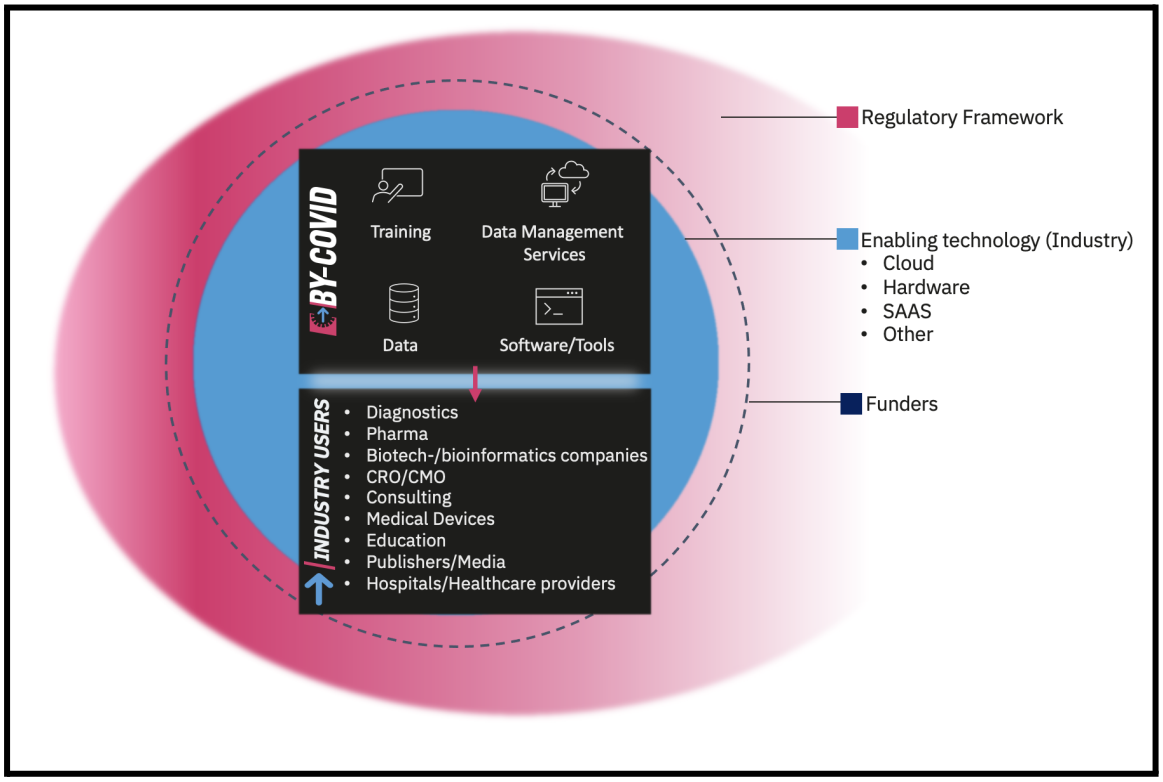

Figure 2: Graphical representation of the BY-COVID relevant industry sectors. Regulators and funders provide the underlying framework for public and private players to interact and innovate in the ecosystem BY-COVID is embedded in. Services centred around data, software/tools, data management services as well as training are provided within this framework to users, such as industry.

Public-private partnership example:

Collaboration between Dieter Maier (Biomax informatics) and Marek Ostaszewski (BY-COVID WP5) on a COVID-19 disease map.

Funded by the European Union

BY-COVID is funded by the European Union's Horizon Europe research and innovation programme under grant agreement number 101046203.

Biomax Informatics: Biomax Informatics is a Munich-based software company specialising in knowledge management and the creation of knowledge networks using openly available published literature (e.g., in Europe PMC).

Dieter and Marek initially met through PREPARE (Platform for European Preparedness Against (Re-) emerging Epidemics), an European Commission Funded partnership, which is building large networks of research organisations, hospitals, primary care sites and laboratories across Europe to be prepared for rapid research during infectious disease outbreaks and pandemics threatening Europe. Biomax's AILANI literature-mining algorithm was used for data mining to collect information for a knowledge network on Zika virus.

At the start of the COVID-19 pandemic Biomax joined the Disease Map initiative to support the creation of a knowledge network on SARS-CoV-2 using the above mentioned algorithm (AILANI), and made the result freely available as AI driven question and answering system for COVID-19 focussed projects (https://ailani.ai).

The COVID-19 disease map is a computational repository of SARS-CoV-2 virus-host interaction mechanisms based on openly published literature and is one of the use cases in BY-COID WP5. The collaboration has helped to identify relevant literature on COVID-19 and SARS-CoV-2 and create a comprehensive knowledge network that will be further developed in BY-COVID WP5. In the future the technology and learning developed in the COVID-19 disease map can be applied to other infectious diseases.

References:

Abbasi, A., Sarker, S. and Chiang, R. (2016), "Big Data Research in Information Systems: Toward an Inclusive Research Agenda", Journal of the Association for Information Systems, Vol. 17 No. 2.

Bousfield, D., McEntyre, J., Velankar, S., Papadatos, G., Bateman, A., Cochrane, G., Kim, J.-H., et al. (2016), "Patterns of database citation in articles and patents indicate long-term scientific and industry value of biological data resources", F1000Research, Vol. 5, p. 160

BioNTech SE and Pfizer Inc. (2020), A PHASE 1/2/3, PLACEBO-CONTROLLED Title: A Phase 1/2/3 Study to Evaluate the Safety, Tolerability, Immunogenicity, and Efficacy of RNA Vaccine Candidates Against COVID-19 in Healthy Individuals, available at:

Funded by the European Union

BY-COVID is funded by the European Union's Horizon Europe research and innovation programme under grant agreement number 101046203.

https://cdn.pfizer.com/pfizercom/2020-11/C4591001_Clinical_Protocol_Nov2020.pdf (accessed 9 February 2022).

Garud, R., Gehman, J. and Kumaraswamy, A. (2011), "Complexity Arrangements for Sustained Innovation: Lessons From 3M Corporation", Organization Studies, Vol. 32 No. 6, pp. 737–767.

Gkeredakis, M., Lifshitz-Assaf, H. and Barrett, M. (2021), "Crisis as opportunity, disruption and exposure: Exploring emergent responses to crisis through digital technology", Information and Organization, Vol. 31 No. 1, p. 100344.

Huyer, Esther. and European Union. Publications Office. (n.d.). The Economic Impact of Open Data Opportunities for Value Creation in Europe.

Lauer, K.B., Smith A., Blomberg N. et al. (2021) "Open data: A driving force for innovation in the life sciences", F1000Research, 10(ELIXIR):828 (document) (https://doi.org/10.7490/f1000research.1118745.1)

Morgan, O., and Pebody, R. (2022) "The WHO Hub for Pandemic and Epidemic Intelligence; supporting better preparedness for future health emergencies", Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin, Vol. 27

Schulte-Althoff, M., Fürstenau, D. and Lee, G.M. (2021), "A Scaling Perspective on AI Startups", available at: https://doi.org/10.24251/HICSS.2021.784.

van de Ven, A.H. (2017), "The innovation journey: you can't control it, but you can learn to manoeuvre it", Innovation, Vol. 19 No. 1, pp. 39–42.

Yang, T. K. and Yan, M. R. (2019) "Exploring the Enablers of Strategic Orientation for Technology-Driven Business Innovation Ecosystems", Sustainability 2019, Vol. 11, Page 5779

Funded by the European Union

BY-COVID is funded by the European Union's Horizon Europe research and innovation programme under grant agreement number 101046203.