BY-COVID

Educational toolkit.

Where does this educational package come from?

This educational package is part of the BeYond-COVID (BY-COVID) project. The project is a scientific collaboration between 53 partner institutes in 19 European countries, who have joined together to improve the way scientific and health data is managed and shared in pandemics. Concepts from the project form a framework from which we can raise broader ethical questions about health data infrastructures in general.

The world has generated vast amounts of data in response to the COVID-19 pandemic, and is still generating more. The project aims to make COVID-19 data accessible to anyone who can use it, from scientists in laboratories to medical staff in hospitals and government officials. These data come from many different sources, and identifying, connecting and integrating it for effective analysis remains a challenge.

The project will also provide a framework for making data from other infectious diseases open and accessible to everyone. Data that is easier to access, aggregate and analyse will enable scientists to respond faster to new strains of SARS-CoV-2 or to new viruses. It will also help policy-makers assess the impact of the disease and take the most appropriate measures to protect people against new infectious diseases.

The 'data', however, are not merely abstract numbers in an abstract database. They often concern the information about individual citizens. Collecting these personal data enables scientific progress. Numerous ethical questions arise in this context. Where, how and by whom does the data need to be collected, stored or shared? How much information do we want to receive about what happens to our data? And how are different stakeholders and citizens involved in this data use? We believe that classrooms offer an ideal circumstance to inform young citizens and deliberate with them on important societal issues.

/ How to use this educational package?

As a teacher, what you will find here are resources to cover the topic in your lessons. It is a toolbox, meaning that you are free to be selective and adapt this educational package to suit your needs. The package consists of a suggestion for hooks, proposed topics and some introductory background information, all built around a number of visuals and cases that illustrate different elements of how data is used, or should be used, to prevent or manage pandemics. The package is built around questions rather than answers. We hope this material will inspire students to think about the ethics of health data systems and lead them to find their own answers. We invite them to share their views on each individual case here.

We suggest introducing the students to the material by emphasising these two points:

- **It already happened during the pandemic**:
  During the pandemic, when a person received a positive PCR test result, they might have needed to isolate, they may have received treatment and they could get proof of past infection. Long after the test, these data continue to serve other purposes to help understand more about the virus. BY-COVID aims to facilitate and improve the use of these data to prevent and manage current and future pandemics. However, it is important to know how these data are managed. Exploring the values behind this framework helps us to understand our own and others' views on data sharing.

- **It has an impact on you and on our society:**
  Since the beginning of the pandemic, more than 765.000.000 cases have been confirmed worldwide. To understand each of these cases, one needs to know if the person was vaccinated (for example: when, how often, which vaccine), hospitalised (for example: duration, type of treatment, in the ICU or not), if they had returned from abroad, been in contact with other people, and which variant of the virus they have. A researcher might also want to know if certain variants of the virus affect pregnant women under 25 differently compared to pregnant women over 35. All this requires linkage between many different datasets.

It is very important to establish links between different sources of data about COVID-19 and other infectious diseases to be able to inform policies, develop treatments and prevent the spread of viruses as much as possible. The links between these different types of data need to be made on an individual level. Data about you is information about you and information is power. Who can have access to this information? What can they do with it? What kind of information about you would you like to see protected?

A toolkit, ready to be adapted:
This educational package includes various cases, grouped into different topics. These illustrative cases can be used as a starting point for a class discussion and help give students the opportunity to explore their own values and principles concerning health data ecosystems. The topics can be used to guide students in forming ethical questions arising from the case. We also provide a set of questions related to the topics that can be used as a starting point for a more in-depth discussion on a specific topic. Classes can focus on one or more as a starting point, but the discussions are likely to touch on a range of different topics as they are all interlinked.

This short video can be used as an additional resource to illustrate the importance of health data reuse. What is Health Data? *(2:33 min)*

/ Topics

Numbers influencing policy

What ethical considerations arise when using scientific and health data to inform policy decisions? What are the potential biases and values embedded in the process of collecting and analysing data?

How do we address the ethical implications of data selection and exclusion in shaping policy decisions? How do we navigate the potential consequences and impacts of data-driven policies on different individuals and communities? Are there potential ethical conflicts between the pursuit of scientific knowledge and the well-being or rights of individuals or communities? To what extent can data justify a reorganisation of society?

How can we ensure transparency and accountability in the use of data to shape societal structures and institutions? What measures can be implemented to mitigate potential biases, discrimination or inequalities that may arise from data-driven decision-making? How do we incorporate diverse perspectives and ensure inclusivity in the process of using scientific data to improve public health?

Related cases: 2, 3, 6, 10

Privacy

How can we balance the benefits of collecting and processing data with the risk of potential privacy breaches? What are the potential risks and consequences of data collection? What measures can be implemented to mitigate the risks of privacy breaches associated with the collection and processing of data? What considerations should be taken into account when handling data that contains personally identifiable information? Does the way in which data is used influence the required level of protection?

Related cases: 1, 3, 4, 8

Aggregated data

How can data aggregation and anonymization practices ensure the protection of individual privacy while still allowing for meaningful analysis? How does the process of aggregating individual data into a larger dataset impact the inherent characteristics and properties of the original data? How does the aggregation of individual data influence the interpretation and analysis of the overall dataset?

Related cases: 3, 9

Data Security

How can we ensure that the combined data is securely stored and protected from unauthorised access or misuse? Which organisation do we grant the trust to store our data? Do we store our data in one place or does it need to be segregated? What technical measures are needed to ensure data security?

Who can access my data? How can health data be reused safely?

When you think of conditions and safeguards, what are you thinking of specifically? IT tools? Laws and regulations? Experts and committees? Other things? What might be the advantages and disadvantages of these? How might they be complementary or conflicting?

In which aspect of the health data reuse process are safeguards most important? Should they control who is accessing and reusing these data? Should they rather focus on the purposes and specific ways of reuse? What could be vulnerable spots in the process of health data reuse that demand specific protection?

How should we balance protecting privacy and promoting scientific progress? How can data federation ensure the protection of individuals' privacy and maintain the confidentiality of sensitive information?

Related cases: 1, 2, 4, 8

Trust

What ethical considerations arise when collecting and analysing personal data without individuals' explicit consent? How can we establish trust among citizens regarding the handling of health data through effective interventions or measures?

What are the ethical responsibilities of organisations and institutions in safeguarding individuals' personal information contained in their data?

Related cases: 2, 4, 7

Data quality

How can we ensure that the combined data are accurate and reliable? What measures can be taken to ensure that high-quality data is collected and used for research and decision-making? How can we ensure that data representation is diverse and inclusive?

Related cases: 2, 5

FAIR-principles

How can we strike a balance between promoting data findability, accessibility, reusability and interoperability (FAIR principles) while respecting individuals' privacy rights and maintaining data security? What are the ethical considerations when making decisions about the level of openness and accessibility of data, particularly when it involves sensitive or confidential information, to ensure responsible data sharing practices? How can we ensure that the benefits of combining health data are accessible to everyone and do not lead to further healthcare inequalities?

Related cases: 4, 5, 8

AI in health data

What ethical considerations arise when utilising artificial intelligence (AI) for health data reuse, particularly regarding the privacy, consent and security of individuals whose data are being utilised? How can the responsible and ethical use of AI be ensured, taking into account issues such as fair representation, and the potential for unintended biases or discrimination in the decision-making processes?

Related cases: 4

/ Introduction

Background information

The COVID-19 pandemic dominated our lives for
 more than two years. The battle against
the coronavirus forced us to make changes to
the way we work, have fun, educate our
children and interact with each other.
The justification for these changes is often found in 'the numbers'.
But where do these data come from? What is it used for exactly?
Which safeguards are in place?
And how are different stakeholders and citizens involved in this health data reuse?
All these questions are linked to the very first question, where and how is health data stored?
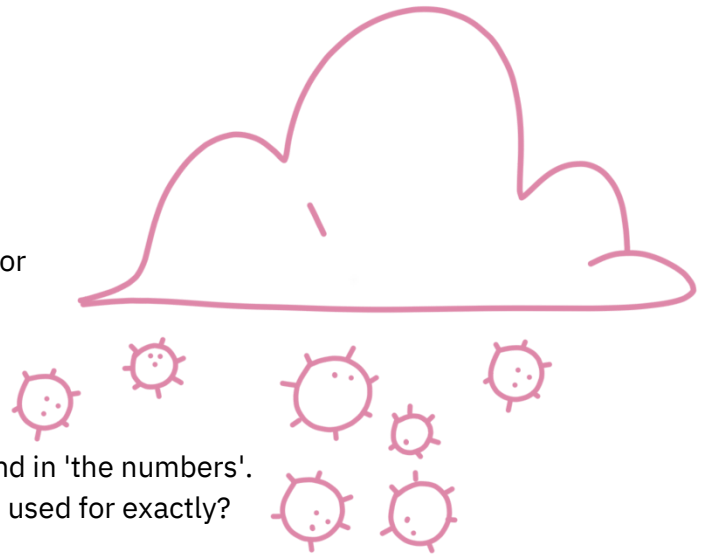
Currently, health data is stored in many separate silos. Imagine you have a lot of important information about your health, like your medical history, test results and medications. Now, picture this information being stored in separate places that don't talk to each other. It's like having different silos where each piece of information is kept isolated from the others. This is a problem because when your health information is scattered in these silos, it becomes difficult for your doctors and other healthcare providers to get a complete picture of your health. They may not have all the information they need to make the best decisions about your care. The silo problem also makes it hard for different doctors or departments to work together and share information. Another issue is that when health data is divided in silos, it's not easy to use that information for research or to find patterns that can help improve healthcare for everyone. It's like having puzzle pieces scattered in different places, and you can't put the whole puzzle together to see the bigger picture.

Data is what allows us to describe reality. A lack of data leads to a limited understanding of the world around us. We can illustrate the importance of a system to link data effectively, by imagining an alien named Xerox. He visits Earth to explore it as a potential market for his shoe company. Being unfamiliar with human anatomy, Xerox begins his investigation by entering a classroom where he finds only one student sitting in a corner. Based on the student's feet, Xerox makes some initial assumptions about the characteristics of human feet. He counts 10 toes, and measures the length and width. However, his understanding is limited as he has only observed one person.

Shortly after, a group of additional students enters the room, providing Xerox with a broader sample of human feet. As he compares the new individuals, he realises that his initial assumptions about foot length and width were incorrect. He discovers that human feet come

in various sizes and shapes, and the length of their feet may be shorter or longer than the initial 30cm he assumed. Furthermore, he notices that some people have narrower feet than others, prompting the need for adjustments to accommodate different foot widths.
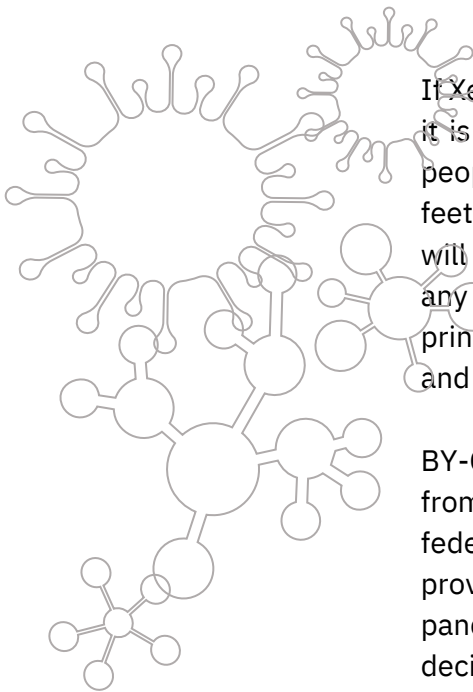
If Xerox wants to create a range of shoes that fits all humankind it is clear that he'll need to collect information from as many people as possible, the more, the better. Each additional pair of feet he can include in his research, adds a bit of information and will in some way change the final shoe collection. To answer any research question concerning health data, the same principle is applicable. Data is power, even more so when linked and combined.
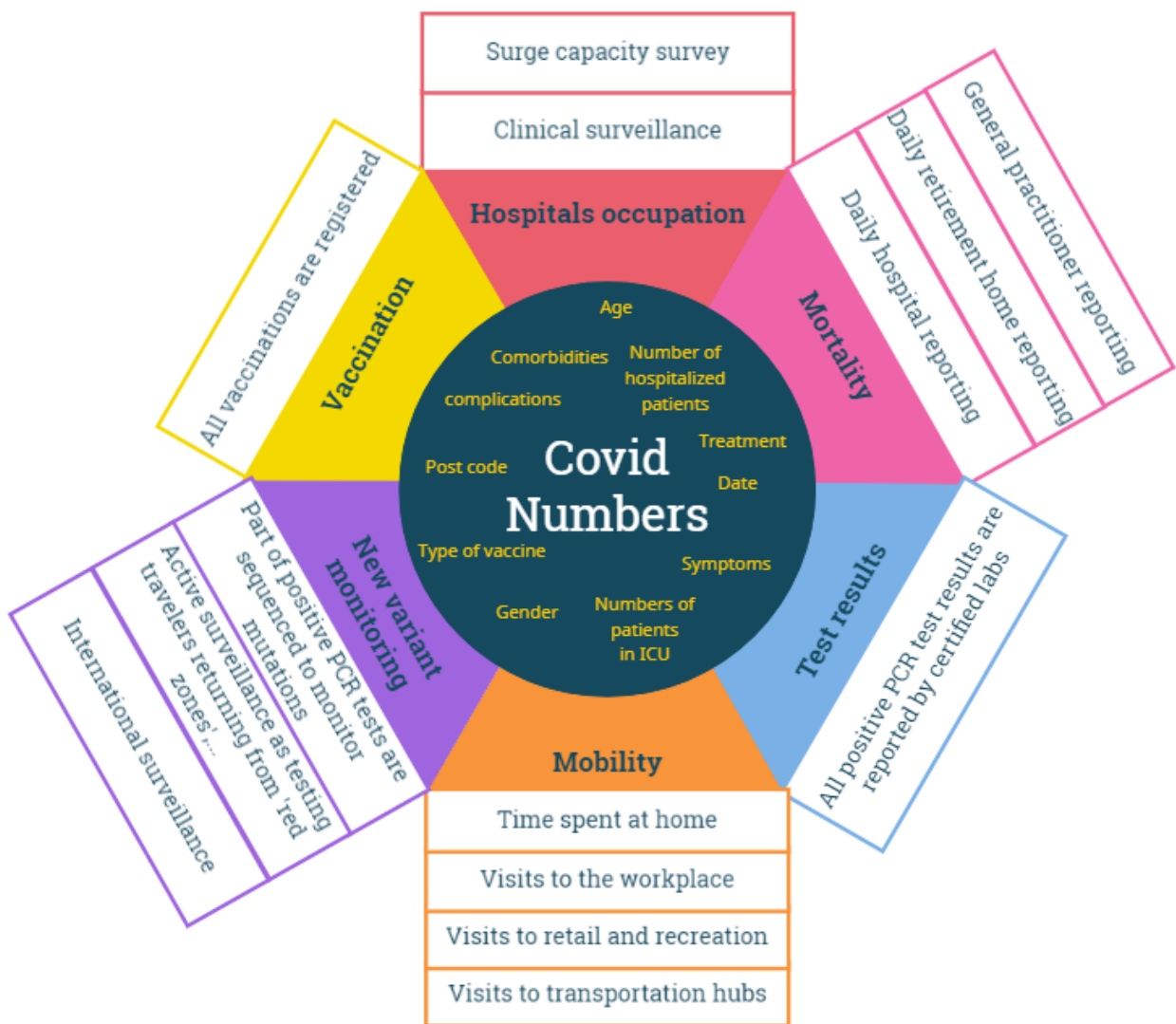
BY-COVID supports a federated structure to connect the data from these separate silos. Through the implementation of a federated data ecosystem, the BY-COVID project aims to provide a comprehensive and consolidated view of the pandemic's impact, enabling better analysis, tracking and decision-making to combat COVID-19, and possible future pandemics effectively.

Federation refers to an approach to data management where data remains in its original location but can be accessed and analysed by authorised parties. Different organisations or institutions collaborate and maintain control over their own silo of data while allowing others to securely access and utilise it. This way, silos containing various types of healthcare information can be linked together without having to store all data in one place. Let's say a public health researcher is collecting data to understand how the virus is spreading in different countries. Instead of requiring all these data to be sent to a central database, each individual instance representing the country provides access to the particular data to answer the research question. Each instance allows access to the required pieces of a puzzle they own and the researcher can build up an overall picture. Each research question may result in a different puzzle.

They can also decide whether they want to aggregate the data they provide. In the context of the BY-COVID project the health data will often be aggregated. Aggregated data refers to combining data collected from various sources related to the COVID-19 pandemic. This data includes information on infection rates, hospitalizations, vaccinations and other relevant metrics. When data are aggregated, atomic data rows, typically gathered from multiple sources, are replaced with totals or summary statistics.

None of this however would be possible without our personal data. In the visual provided you will find an overview of all the different data collected, linked and used during the COVID-19 crisis. It, for example, starts with an individual getting tested for COVID-19. In the first place, this information will be useful for the particular individual and his doctor. Only in a

second instance, when the data becomes a part of a bigger structure, does it become a powerful tool. The BY-COVID infrastructure is setting the foundations to make it possible to link the statistics of infected people in a region to the amount of visits to retail and recreation there. Consequently, it is partially my data that will guide policy makers to temporarily close music schools or relax social distancing guidelines in sports clubs.

/Cases

Case #1 / School of fish

United we stand, divided we fall

It can be difficult for a small fish to survive alone, swimming in an ocean of predators. Many species of fish have found a way to stay safe: the school of fish. Moving around in a group has many advantages: fish are stronger by swimming together, so less vulnerable to predators, who find it easier to attack an isolated prey. They also create a current, which saves energy and increases the speed they can swim. The group is created collectively, each fish adapts to the movements of its neighbor, and the life expectancy of each fish increases as a result.

OK, interesting, but what about data?

This fun fishy fact has parallels with our health data. Data from one individual is of little use to the collective, it is only when combined with a large pool of data that it can make a difference. For example, if we want to know the impact of Covid-19 on people with a specific disease, such as lupus, it is important to analyse data from as many patients as possible. The more datasets are combined, the more insightful the research will be, increasing the chances of finding effective treatments. The same goes for institutions, countries and researchers: collaboration makes the data stronger and the research faster. Combining data from different research fields also helps deepen understanding about health and disease.

How do we make a school of data?

To make this possible, we need a framework to collect, manage and allow access to the data across international borders that respects the data protection rules of each country. The BY-COVID project is making this pooling of data possible, by bringing together researchers and institutions from many EU countries, so that the data from each country can be combined to generate knowledge for everyone. Combining data increases the power of research and improves patient care. BY-COVID is building a facilitating system to store, share, link and use data. For example, to make sure that the combined pool of data can be analysed, BY-COVID is developing metadata, which tags each piece of data so it can be found and understood.

If you want to know more...

- Find out more about schools of fish: Fish swimming in schools save energy regardless of their spatial position - PMC
- Find out more about the value and importance of sharing data: Coronavirus accelerates drive to share health data across borders | Research and Innovation
- Find out more about the objectives of BY-COVID
- Why we need data for specific diseases: Impact of COVID-19 on people with cystic fibrosis - PMC
- This article shows the importance of linking data sets in the case of cancer: Linking data to improve health outcomes | The Medical Journal of Australia

Case #2 / Policy action

With help, it is easier **to make decisions**

Better data
Better knowledge **BY-COVID**

Good information leads to tasty outcomes

If you want to make a great birthday cake you need fresh ingredients and a good recipe to combine them. Without one or the other the cake is doomed.

Public health decisions in pandemics are much the same: you need good information and a way to bring the information together. For example, if you don't know how effective vaccines are or how severe the next variant is, it becomes impossible to formulate good policies. Data are the key ingredient for evidence-based policies, but they are useless without the framework of a good recipe. To be useful, scientific data needs to be correctly collected, stored and linked with other data in a controlled but accessible environment.

If policymakers are the cooks of public health policies to manage pandemics, the BY-COVID project is providing the best possible recipes to combine data to inform their decisions. The goal is to support policymakers and provide the tools required to make decisions that are rooted in reality and proportional to the risks that are faced.

Where would you like to eat?

Would you eat in a restaurant using expired ingredients, wild recipes that are never tested and no hygiene standards? In the same way, public health policies need good data, well-tested methods to combine the data and standards to ensure data protection. With the right raw materials, equipment and resources, everything is in place to make the best food possible.

Who do you trust to set policies in pandemic times?
What should their decisions be based on?
Which data are the most important to collect?

If you want to know more...

- About the collaboration between scientists and policy makers: Science advice in times of COVID-19 - OECD
- About the link between democracy, government action during the pandemic and citizen trust: Trust made the difference for democracies in COVID-19 - The Lancet
- About the objectives of BY-COVID: https://by-covid.org/pdf/BY-COVID_factsheet_dark.pdf
- About the impact of scientific uncertainty on public trust in science and support for science-based policy: Model uncertainty, political contestation, and public trust in science: Evidence from the COVID-19 pandemic

Case #3 / Wastewater

Did you know that flushing the toilet can help research?

Did you know that flushing your toilet might be contributing data for research? Wastewater from toilets has been used as a valuable tool for detecting disease hotspots in the COVID-19 pandemic. By surveying wastewater we can anticipate rates of COVID-19 related hospital admissions in specific areas or even predict future outbreaks of other diseases.

Our excrement can be seen as a dynamic record of events happening in our body, for example, it contains fragments of the COVID-19 virus when we are infected with COVID-19. This means we can measure which variants are circulating in particular areas and use this information to design a targetted public health response.

How is the data collected and stored?

The BY-COVID project is working with the Italian SARI initiative which collects data from wastewater approximately once or twice a week in each city in Italy. The data are then displayed on dashboards in real-time to provide up to date surveillance of pandemic waves. Researchers from the BY-COVID project ensure the data extracted from wastewater samples are available for re-use in other studies, for example to make comparisons between different countries, to answer research questions about viral spread or to inform public health decisions.

What are the benefits?

Wastewater analysis is useful for both the detection and surveillance of infectious disease outbreaks. Policy responses can be more effective if we know what pathogens are present and in which locations. The resources provided by BY-COVID (for example databases and data analysis software) are therefore useful for fighting the current pandemic and also for spotting future ones.

What are the risks when we use these data?

Personal data could be found in wastewater if the human DNA in the excrements is analysed. However, only data about viruses is collected, not data from humans. The BY-COVID infrastructure is set up to counter any risks of data misuse: all the data gathered for the wastewater analysis is aggregated, which means it would be impossible to precisely identify any information about individuals.

If you want to know more...

- Here is an example of the use of wastewater analysis on a university campus : https://www.youtube.com/watch?v=ysZsx5wS2YM
- This article presents a summary of the results of a first survey of COVID-19 variants in Italian wastewater : Flash survey on SARS-CoV-2 variants in urban wastewater in Italy 1st Report (Investigation period: 04 – 12 July 2021)
- Find out more about the use of wastewater : Water and wastewater digital surveillance for monitoring and early detection of the COVID-19 hotspot: industry 4.0 | SpringerLink
- Watch a short explanation on aggregated data: https://www.youtube.com/watch?v=PvohkMNV9Lo

Case #4 / Positive side effects

Can a pandemic have positive side effects?

Good infrastructures lead to good outcomes **BY-COVID**

Can a pandemic have positive side effects?

Have you ever been looking for something specific in a messy room and stumbled upon the unexpected? This is a common phenomenon in the world of research, where some studies carried out for a specific purpose turn out to be useful for something else. If scientific data are sufficiently numerous, accessible and well processed, it not only makes the research better, but also opens up possibilities for re-use of the data in new and different studies.

Some examples of (re)discoveries

The COVID-19 pandemic mobilised researchers, institutions and data from all over the world and led to advances in many areas. For example, existing methods, like PCR tests and messenger RNA vaccines, were validated on a massive scale during the pandemic, and are now easier to apply to other diseases (PCR tests for monkeypox and messenger RNA vaccines for the treatment of certain cancers).

Artificial intelligence (AI) and Open Science, a movement to make science widely accessible, both existed before the pandemic but were deployed in an unprecedented way leading to accelerated development. Scientists saved months of experimentation thanks to predictions of the virus structure generated by AI, which also helped with genome sequencing and predicting the evolution of the pandemic.

In terms of Open Science, the crisis has accelerated the move away from restricted and conditional access to knowledge. Essential information about the virus was made available to everyone, allowing a better understanding of the disease and the rapid manufacture of vaccines, diagnostics and treatments.

BY-COVID as a vector for scientific progress

These advances are only possible with a good data infrastructure that allows researchers and institutions to work together to provide readily available data to support discovery and decision making. One way the BY-COVID project is promoting Open Science is by contributing to the European Open Science Cloud (EOSC), which "aims to develop a Web of FAIR Data and services for science in Europe upon which a wide range of value-added services can be built."

By bringing together many countries, researchers and institutions, and by making the data easily accessible, BY-COVID is creating an infrastructure to enable positive things to come out of the pandemic.

If you want to know more...

- RNA vaccines: New cancer treatments may be on the horizon—thanks to mRNA vaccines
- AI: AI and control of Covid-19 coronavirus
- https://www.swissinfo.ch/eng/artificial-intelligence-helps-bring-about-record-fast-vaccines/46256752
- Open Science: Open science saves lives: lessons from the COVID-19 pandemic | BMC Medical Research Methodology
- How COVID-19 had an impact on research: How will COVID-19 reshape science, technology and innovation?
- The impact of COVID-19 on research - PMC

Case #5 / Metadata

Good standards, many connections

When we travel abroad we may have trouble charging our phone. Without the correct adapter, it is not possible to connect our devices and we will quickly run out of power. With a universal adapter, anyone can travel anywhere in the world. The same goes for metadata: no matter the type of data or the place it is stored, metadata standards allow anyone to find, access and use data for their research studies.

What is metadata? Why is it important to have standards?

Metadata is "data about data". However, not all metadata is useful, and standards need to be agreed by the research community and ideally follow a set of guidelines called the FAIR principles. Metadata standards assist with the creation of metadata catalogues to make data findable and accessible, and can serve also as adapters to make data interoperable and reusable.

To give a simple example, when you search for a film on a streaming platform, you will find information such as the year of release, the genre of the film, the director, the duration. The information describes the film and makes it easier to find a film you are interested in and decide whether you want to watch it.

It is the same with research data: if the information about the data are precise and detailed, it will be much easier for a researcher to discover that the data exists and to use it for their analysis. If a researcher finds two potentially useful datasets, but one refers to locations by name and another by postal codes, a metadata standard is necessary to combine the data.

We all use metadata standards every day: when we are using our GPS to drive to the Eiffel Tower, we should end up in the same place, regardless of the app we use. Good metadata standards allow the app developers to direct people to where they want to go, whether they entered 'Eiffel Tower', 'Tour Eiffel' or '75007'.

Data needs to be accompanied by metadata by design

The best book in the library will not be found if it is not indexed in a standard way. Different items in the library have different ways of being found. For example, a magazine has an issue number, a book can be a special edition, a comic book can be part of a larger series. The same is true for data: social data, medical data and biological data all require different metadata to describe them appropriately.

Metadata standards are necessary to find, link and use these data for research across different fields. Standards for data and their associated contextual and experimental metadata are also known as data standards, metadata standards or content standards, and can be classified in four standard subtypes: reporting guidelines or checklists, models/formats or syntax, terminology artefacts, and identifier schemata.

It is essential that metadata capture and standardisation are built into plans at the beginning of any research project, before the data are collected. This ensures that the data will find their correct place in a global ecosystem of information. Good metadata also improves the quality and reliability of data and confidence in the findings of the research.

Good standards, many connections

In the BY-COVID project there are many data sources (for example, databases, repositories and knowledge bases) from different research disciplines, including bioscience, clinical and epidemiological research, and social sciences and humanities. These data sources are being described in a FAIRsharing collection (in progress), along with the data and metadata standards used by each data source. A common metadata model has been developed to represent the metadata in each source, and to make it findable in one place: the Covid-19 Data Portal.

Developing a common metadata model is a major challenge, as the project involves a large number of researchers from different scientific fields, and each partner's data source uses different metadata standards. The approach is to map the key inter-relationships between the metadata in a way that makes sense and is practical to implement. This then opens up exciting possibilities for discovering more about how infectious diseases affect people and to inform evidence-based policymaking.

If you want to know more…

- FAIRsharing Educational: learn about standards for data and metadata, how FAIRsharing registry helps you if you are a consumer or a producer of data, metadata standards, databases and data policies.
- More details about metadata: Introduction to metadata management
- Find out how indexing is used to link data in the BY-COVID project: Release of indexing system to link COVID-19 data across research disciplines
- Learn more about the importance of having metadata standards (in general) : 5 Minute Metadata - What is a standard?
- Learn more about the importance of metadata standards against COVID-19 pandemic: COVID-19 pandemic reveals the peril of ignoring metadata standards | Scientific Data
- Find tools and guidelines to help you access, analyse and share infectious disease data, and respond quickly to disease outbreaks: Infectious Diseases Toolkit
- BY-COVID - D3.1 - Metadata standards. Documentation on metadata standards for inclusion of resources in data portal | Zenodo
- BY-COVID D2.1: Initial data and metadata harmonisation at domain level to enable fast responses to COVID-19 https://doi.org/10.5281/zenodo.7017728
- Learn more about recipes that help you make data FAIR in the FAIR Cookbook an online resource of hands-on recipes for "FAIR doers" in the Life Sciences. The FAIR Cookbook - pre-print: "The essential resource for and by FAIR doers", provided you with more information about its creation and content.
- FAIR, ethical, and coordinated data sharing for COVID-19 response: a review of COVID-19 data sharing platforms and registries | Zenodo
- Packaging research artefacts with RO-Crate - IOS Press
- Lightweight Distributed Provenance Model for Complex Real–world Environments | Scientific Data
- [2205.12098] COVID-19: An exploration of consecutive systemic barriers to pathogen-related data sharing during a pandemic

Case #6 / Future pandemics

Be innovative in the face of adversity...

...makes it **easier later**

BY-COVID

Creating paths for the future

The Covid-19 pandemic has upset our lives, but it has also transformed the world of research. Scientists from all over the world came together to understand the disease and develop treatments and vaccines as quickly as possible. But what happens to these advances once the crisis is over?

Science as exploration

Explorers go on expeditions to discover unknown areas of our world. They may be looking for a new species of fungus with medicinal properties or trying to observe novel animal behaviour. Most will find their way using tried and tested tools, such as GPS, along with the ability to innovate in the face of unknown situations. Once the pioneers have found a way through the wilderness, their experience allows future travelers to follow more easily in their footsteps. The danger of getting lost, getting stuck in dangerous situations or wasting time searching, is much lower when the terrain has already been mapped out. The more the path is used, the better it will be marked out and the easier it will be to access. It would be a pity to leave these routes abandoned when we know they can lead to interesting places.

Exploring infectious diseases

The pandemic is not the first, nor the last we will face. However, we will be better able to anticipate future pandemics thanks to the knowledge and experience gained throughout Covid-19. Just like exploration, prior research and experiences are very useful as a basis for further work.

Now that the pandemic is more under control, there is time to look back at lessons learnt and how these can be applied in future. We can use these learning to accelerate pandemic response efforts, for example optimising vaccine production. The research carried out, the results found, the procedures developed and the methods used can be used to prevent or manage future pandemics.

One of the most effective weapons against the Covid-19 virus was the rapid sharing of data from all over the world. The data showed how the disease affected the body, which measures were most effective in stopping transmission, how the virus was evolving and the best vaccines to prevent infections.

Beyond COVID

Like explorers, scientists led the way in collecting, linking, storing, using and sharing data to help combat the pandemic. Covid-19 caused a worldwide crisis, and the scientific pioneers had to work rapidly and innovatively, for example to ensure all ethical and legal frameworks were in place and find new ways to connect data. This work under pressure showed the benefit of making data across countries and research areas available to everyone.

Some data infrastructures were created specifically to resolve this crisis, but proved to be so useful that now it's important to make sure they are maintained. The BY-COVID project is following in the steps of the pandemic scientists to build maps, guides, bridges and stepping stones to mitigate and manage future pandemics.

Remember the toughest challenge you ever faced? Like studying for your exams or training for a marathon? Did you learn things from these achievements that proved useful for other challenges? What do you think the main lessons learned about data were during the pandemic?

If you want to know more...

- An example of a good health data infrastructure (Denmark): [Health data and registers - Sundhedsdata Styrelsen](#)
- Review of the molecular and cellular determinants of critical COVID-19 pneumonia : [Human genetic and immunological determinants of critical COVID-19 pneumonia | Nature](#)
- A look back at the lessons learned from the pandemic: [Ten lessons from the first two years of COVID-19 | McKinsey](#)
- An example of a lesson learned from understanding Covid-19 (knowing how to use all the data involved): [Why the WHO took two years to say COVID is airborne](#)
- [What COVID-19 has taught us: "Healthcare can no longer exist without technology" | Africa Renewal](#)
- The BY-COVID framework reused for monkeypox: [Repurposing COVID-19 pipelines for the monkeypox virus](#)
- Digital health during the pandemic and how to use it afterwards: [Data capture and sharing in the COVID-19 pandemic: a cause for concern - The Lancet Digital Health](#)
- What covid has shown us about the importance of data sharing and open science: [https://www.researchgate.net/profile/Kenneth-Zheng-2/publication/343892070_Data_sharing_during_COVID-19_pandemic_what_to_take_away/links/5f87cf9b458515b7cf81eda9/Data-sharing-during-COVID-19-pandemic-what-to-take-away.pdf](#)

Case #7 / Umbrella

Our data protect us.

⮞BY-COVID

In data we trust.

Throughout vaccine development and rollout, an array of data, from experiments in the lab to clinical trials to post-licence monitoring, helps make vaccines as effective and safe as possible. Access to data is important for experts making public health decisions but also for lay people, as their perceptions and attitudes strongly affect vaccination uptake.

Without trust in safety and effectiveness of Covid-19 vaccines it is hard to gain high (voluntary) vaccination rates in any population.

Data influences public opinion

Polls tell us that in EU countries hesitation to get vaccinated is often based on lack of access to reliable data, with almost one in two respondents agree that it is difficult to find information that they can trust about COVID-19 and vaccines.

The Covid-19 pandemic brought with it waves of data (for example numbers of tested, infected and diseased citizens) and information on how Covid-19 diagnostic tests and vaccines are working. For many people, not only lay people but for health experts too, it was hard to interpret all the data. But it is not possible to understand a pandemic without data. So what we need is not less data but ways and tools to make trustworthy data on Covid-19, and indeed other infectious diseases, easily accessible.

BY-COVID builds trust

The BY-COVID project has created a COVID-19 Data Portal which links data from many different research areas such as genomics and social science. It is a resource for scientists and health professionals, but it can also help reassure citizens that reliable data exists and is accessible to all.

Data doesn't always change public opinion easily. According to the EUROBAROMETER poll, trust in COVID-19 vaccination reflects general trust in institutions such as national governments. This trust reflects not only the current policies and actions of the institutions but their past record. This makes building trust a slow process, but one in which reliable and accessible data play an important role.

If you want to know more...

- Eurobarometer data on attitudes of Europeans towards vaccination:
- Eurobarometer (data collected in May 2021)
- Eurobarometer (data collected in February 2022)
- CESSDA Data Catalogue where some interesting COVID-19 related data located in European social science data archives can be found: CESSDA Data Catalogue
- Alternative way how to get to European data on social issues connected with COVID-19 is through CESSDA signpost page: CESSDA and COVID-19

Case #8 / FAIR data

Data are priceless, when you give them meaning

Data are generated all over the world to help respond to pandemics and infectious disease outbreaks. However, each isolated piece of data has limited value unless it can be linked with related data and made available in a meaningful way.

A common structure and approach is needed to link different types of data and to ensure the data are managed responsibly and efficiently. This makes data easier to share and reuse - and therefore even more useful and valuable. But how are the data managed and made accessible to others? There are a set of guidelines which point the way: the FAIR principles.

What are the FAIR principles?

The FAIR principles support open science, a global movement to make science accessible to all researchers and all levels of society. Each of the four letters represents a separate element:

Findable. Scientists must find the data they need. The more information about a dataset (metadata, which also need to be standardised) is available the easier it can be found.

Accessible. Access to the data should be as easy and open as possible. Data are sometimes locked away from the public eye. There can be valid reasons for this, such as protection of a person's personal data. However, to develop medicines or diagnostics tests scientists need to be able to access data in a safe and secure way.

Interoperable. Scientists often need to combine data for their work. To be able to do this they need to be available in well-defined and compatible formats so that data can be linked, compared and analysed effectively.

Reusable. F, A and I are set up to make the data reusable. Data, originally collected for one purpose, can improve health or advance scientific knowledge in other areas. For example, secure personal identifiers can ensure the usefulness of a dataset in the long term.

Highly sensitive and private data, for example vaccination status, can also be FAIR by placing controls on accessibility. This ensures the key parts of the data can be accessed without making available personal information.

FAIR principles are guidelines and not a standard. The principles describe a continuum of increasing reusability, full realisation of which can be hampered by regulations, budget and available technologies. Collaboration between all stakeholders (clinicians, researchers, public health officials or policymakers) is important to apply the principles as much as possible.

FAIR principles for making sense of data and for minimising errors

Applying the FAIR principles not only improves data sharing and accessibility, but also minimises the risk of bias and error in research. Indeed, the application of this framework leads to more datasets that are consistent, relevant to research and as reliable as possible. Having access to enough FAIR data allows the safety and efficacy of interventions to be openly assessed based on the evidence.

The contribution of BY-COVID to the FAIR principles

The application of FAIR principles brings benefits to the BY-COVID project and, by extension, to the research community. By applying these principles, including making research results accessible, the BY-COVID project enables future researchers to reuse infectious disease data in order to make new discoveries. Providing access to data and results also demonstrates the transparency of the research, as it allows people to check how the research was carried out, what data were used and how.

The BY-COVID project contributes to the FAIR principles in many ways. For example, by carrying out inter-domain metadata mapping (a process of mapping metadata from different specialities so they can be used together), or by improving data discovery, integration and citation, using tools based on the European Covid-19 Data Portal and FAIRsharing. These tools enable researchers to easily find the data they need, to integrate them into their work and to properly acknowledge their sources.

To better respond to future pandemics, it is essential that projects such as BY-COVID apply the FAIR principles to enable a shift from fragmented, competitive data sharing to a network of interconnected and accessible data.

If you want to know more…

If you want to know more about FAIR data :
- FAIR data — Ghent University
- How to make your data FAIR through Data Standardization – HERAX
- Les principes FAIR de gestion et d'intendance des données de recherche (FR)
- FAIR, ethical, and coordinated data sharing for COVID-19 response: a review of COVID-19 data sharing platforms and registries | Zenodo
- The FAIR Guiding Principles for scientific data management and stewardship

EOSC
- The European Open Science Cloud (EOSC) is an environment for hosting and processing research data to support EU science.

If you want to know more, especially about interoperability
- Q&A: Interoperability and COVID-19, Part 1 - Watson Health Perspectives
- Interoperability in Healthcare | IBM.
- Harmonising clinical data to facilitate large-scale health research | News | CORDIS | European Commission
- How COVID-19 has fundamentally changed clinical research in global health

Case #9 / Federated system

Data visiting is the new data sharing

During the COVID-19 pandemic, online meetings became part of working life for many people. We stayed safely at home and just like any other meeting, we shared ideas to move our projects forward. The difference: we stayed at home. However, the ideas we shared left our house to join up with others. Our thoughts and opinions, put together, enriched the discussion and enabled progress.

What is the link between data and online meetings?

Research data are stored in different ways in various places and access is not always easy. With a federated system, the data remains in place, rather like the people working from home, but researchers and other people who benefit from the data can gain access. A federated system is composed of several data providers, such as hospitals or universities who collaborate together, while maintaining control over their data management.

Federation can be understood as similar to online meetings: each person represents the data (which doesn't move) whereas the knowledge derived from the data is represented by the ideas of each person (which is free to be shared). The alternative to federation is known as centralisation: meeting participants (data) all travel to the same location to meet in person and share ideas.

Data visiting is the new data sharing

Federation facilitates the reuse of data, following the FAIR principles of findability, accessibility, interoperability and reusability. There are many political, ethical, administrative, regulatory and legislative barriers to data leaving its home location, federation allows the data to stay where it is but still be reused to advance research.

BY-COVID as an online meeting

The BY-COVID project uses both centralised and federated systems, depending on the data type. In centralised systems, the data is shared in common databases which reside in a different location from the data. Centralised systems are used for data which is not sensitive, for example, data that isn't about people.

However, to provide answers to interesting policy questions, access to more sensitive data is also useful. This data, found in a variety of different places, can be accessed and analysed using federated systems, following national governance rules and regulatory restrictions. The results can be widely shared in scientific papers without the data moving location, like sharing the outcome of an online meeting while the participants stay at home.

If you want to know more...

- [What is federation (English)](#) ([French](#))
- [The way forward is not data sharing but data visiting](#)
- [Coping with interoperability in the development of a federated research infrastructure](#)
- [BY-COVID data management plan](#)

Case #10 / Collaboration

Collaboration for innovation

During a public health emergency like the COVID-19 pandemic, it is essential that experts, from both academia and industry, unite for the common good even if their interests differ. To safeguard public health, as much data as possible must be open and accessible, so everyone can contribute their expertise and benefit from the knowledge generated. Keeping science open promotes transparency and collaboration, which then stimulates innovation and accelerates the speed of research.

Given the volume and complexity of data produced across national borders, all data must be carefully managed to remain meaningful and useful. Data management is a role played by publicly funded data infrastructures working for the benefit of all. These infrastructures develop tools and make available data that can be widely used not only by academics, but also by the private sector and policymakers. This enables productive collaboration between the public and private sectors, and allows the private sector to translate scientific discoveries into innovations to benefit society.

In the COVID-19 pandemic, thanks to the large amount of data available and openly shared, companies were able to create innovative solutions, such as new vaccines, diagnostics and therapeutics at a speed never before seen. This was only possible due to the wealth of data available in publicly funded infrastructures.
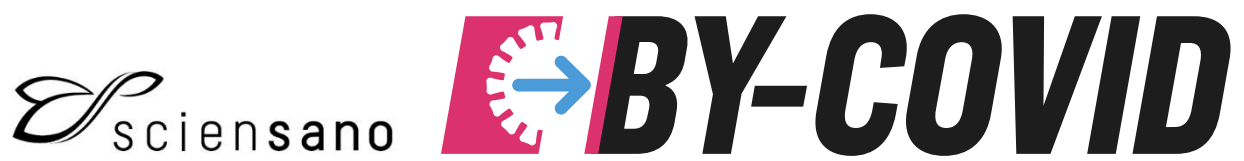
BY-COVID as a model for innovation

The BY-COVID project works to build the capacity of public infrastructures to meet the need of future pandemics. The speed of innovation during COVID-19 was enabled by data infrastructures that predated the start of the pandemic. Despite being existing resources, the infrastructures needed to rapidly adapt to accommodate the deluge of data produced. The BY-COVID project is building on the tools and resources created during the pandemic to provide support for future emergencies and to enable ongoing research into infectious diseases. This work will ensure that the resources created will be optimised to fuel future innovation, and that whatever we may face, we will face together.

If you want to know more…

- [Accelerated innovation in crises: The role of collaboration in the development of alternative ventilators during the COVID-19 pandemic - ScienceDirect](#)
- [What COVID-19 taught us about collaboration – 7 lessons from the frontline](#)
- [Collaboration and innovation to overcome COVID-19 challenges – Dr Ir Siti Hamisah Tapsir – INGSA](#)
- [Partnership and collaboration: The new normal in the fight against COVID-19](#)