About Pathoplexus

Who we are: Pathoplexus is a non-profit, non-governmental, independent database founded by and for the pathogen data-sharing community: academics, scientists, and public-health professionals. Pathoplexus has members from 14 countries and an executive board with five members from five different continents.

Why we started Pathoplexus: Our members have decades of experience generating, analysing, and sharing pathogen data: we track outbreaks, sequence viruses, provide open analyses and tools, and inform public health. We experienced how data-sharing systems could create inequity, delays, blame, and fear of sharing - and built a system specifically designed to address these challenges.

Who funds us: Nobody. Pathoplexus has no dedicated funding and is supported entirely by volunteer effort and donated IT resources. The underlying software receives small academic grants, but all management and development of Pathoplexus itself is donated time - because building a community-driven system is the right thing to do. We are pursuing grant and donation support to sustain our work, but any funding must respect that Pathoplexus is governed by and for its community. Our decisions and policies are not for sale.

What our values are: Learning from the shortcomings of other databases, we are committed to **transparency** (public minutes and monthly community meetings), **clear governance** (active, powerful, diverse membership and board), **interoperability** (persistent identifiers, metadata standards, machine-readable data), and **equity** (clear data-use rules that include publishing embargos and co-authorship requirements). Read more in our Statutes and Values. Our focus so far has been on academic inequity ("scooping") and benefits for public health, but we are open to being part of a system that implements broader benefit sharing principles.

How Pathoplexus works: Users can share data as 'Open' or 'Restricted-Use'. Open data does not have restrictions on use and is passed to INSDC databases (GenBank, ENA, DDBJ). Restricted-Use data cannot be used in academic publications without submitter consent until protections expire (currently up to 1 year), but is available at any time for public health use, such as government reports and open tool dashboards. We also make it easy to upload data, correct errors, view quality metrics, search records, and create DOIs to credit data generators.

How 'open-access, restricted-use' works: Pathoplexus has an open system - all data is available. *Whenever users access Restricted-Use data*, they agree to our Data Use Terms, which restricts how data can be used. Academic use is not private: publications are public and linked to authors. Journals also require data sources to be transparent, so published use of Pathoplexus data is immediately visible and attributable. Thus, if someone uses Pathoplexus data in a publication we know who they are, and whether they are conforming to our Data Use Terms or not.

How enforcement works: We work directly with journals. If a publication violates our Data Use Terms, we seek solutions that support the original data generator - for example, inclusion in analyses/authorship, or paper retraction.

Is this an attractive system? Yes. Since launching in August 2024, Pathoplexus has seen a steady rise in submissions and submitting countries. Uganda and DRC chose Pathoplexus to share the first sequences of their recent Ebola outbreaks, and Pathoplexus now hosts the most direct mpox submissions in 2025 (many as 'Restricted-Use'), largely from West and Central Africa. Our users are testament to their belief that the system works.

The Challenges of Closed Systems

Closed or registration-based databases are often proposed as a way to protect genetic data and ensure fair benefit-sharing. This instinct is understandable: if we could tightly control who accesses data, it might seem easier to ensure that obligations are met. However, in practice, the situation is more complex, especially in a system that will be needed by large numbers of diverse users, and closed systems can unintentionally create loopholes rather than close them.

1. **Practical challenges in controlling access**

Registration systems could range from a simple sign-up procedure for any valid email address to complex systems requiring institutional affiliation or verified government issued IDs. The former provides little control over access, while the latter is complex and can be costly and time consuming for databases to implement, and risks excluding citizen scientists, trainees, and researchers in settings where formal IDs or official email addresses are less common. During outbreaks, these challenges are amplified. When analyses must proceed quickly, delays or complex registration push data sharing into informal channels, often making use harder to track rather than easier.
Moreover, access control cannot reliably prevent all downstream redistribution of data through technical means without severely limiting usability and public health response. The result can be an illusion of control without true enforceability.

2. **Pathogens don't respect borders**

Closed systems tie obligations to the provenance of data, but the COVID-19 pandemic has shown that pathogens can spread around the world within weeks. By using data from other sources, e.g. non-signatories of PABS, obligations could be circumvented. A system where obligations apply *regardless* of where data came from is simpler, easier to comply with, and harder to avoid.

3. **Outbreak response rests on many shoulders**

Pathogen data is used by far more than public health agencies. Effective outbreak response depends on researchers who develop innovative tools, investigate pathogen origins or patterns of spread, develop vaccines or therapeutics, characterize immune or therapy escape, and provide open guidance for in-house diagnostics or sequencing protocols. These analyses come from many parts of the global research ecosystem - including academic groups and public health scientists - and also from citizen scientists and trainees who often contribute key checks and innovations. It is difficult to predict before an outbreak who is going to make crucial contributions, and restricting access risks shutting out critical contributors. Limiting access can limit insight - and in turn reduce the world's ability to protect public health.

4. **Control must be complete, or risk being ineffective**

Technical reports, features of variants, and primer schemes are central to outbreak response and public health communication. Yet they inherently disclose key features of the data, such as mutations, relevant for commercial product development. For closed systems to truly prevent downstream use, they would need to restrict such outputs - an approach that is difficult to implement in practice, and risks slowing outbreak response.

Access controls are costly and risk harming public health responses through a poorer analysis ecosystem. The potential benefits of closed-access databases may be difficult to realize, since tying benefit sharing obligations to data access creates incentives to by-pass the system.

Strong, equitable benefit sharing is essential. Realising a system that can deliver it effectively requires careful attention to the tradeoffs in fairness, accountability, and timely public health action in each potential solution.

How “Open-Access, Restricted-Use” Can Support PABS

Pathoplexus currently enforces benefits at the point where value is gained. In the academic system, **benefit** is realised when researchers publish: publications boost reputation and lead to funding opportunities. At the point of publication, researchers are never anonymous. Journals also require data sources to be declared. This means:

- we can identify who used the data
- we can verify whether they comply with our Data Use Terms
- we can ensure the *benefit* (authorship, collaboration, credit) is shared with the data generators

These data-use terms maximize impact of the data generated by making these data available to all, while ensuring that data generators can publish on these data first and/or are directly involved in research on their data. This goes beyond unclear promises of acknowledgements and instead delivers concrete benefits.

The same principle can apply to commercial benefits. Like academic publications**,** products like vaccines, diagnostics, and therapeutics are also not anonymous: they are tied to named companies that bring products to the market. Avoiding tying benefit sharing to specific pathogen genomes is even more critical. Products are typically built on *knowledge* rather than *primary data*. The specific sequence or protein underpinning a vaccine or a diagnostic test could come from many sources and the relevant information will quickly become common knowledge.

We believe an open access database with clear data use terms, alongside a PABS system that puts benefit sharing obligations on commercial actors independently of where data was obtained from and how products were generated, will maximize benefit sharing across the domains of public health, academic credit, as well as access to and profit from pharmaceutical and commercial products.