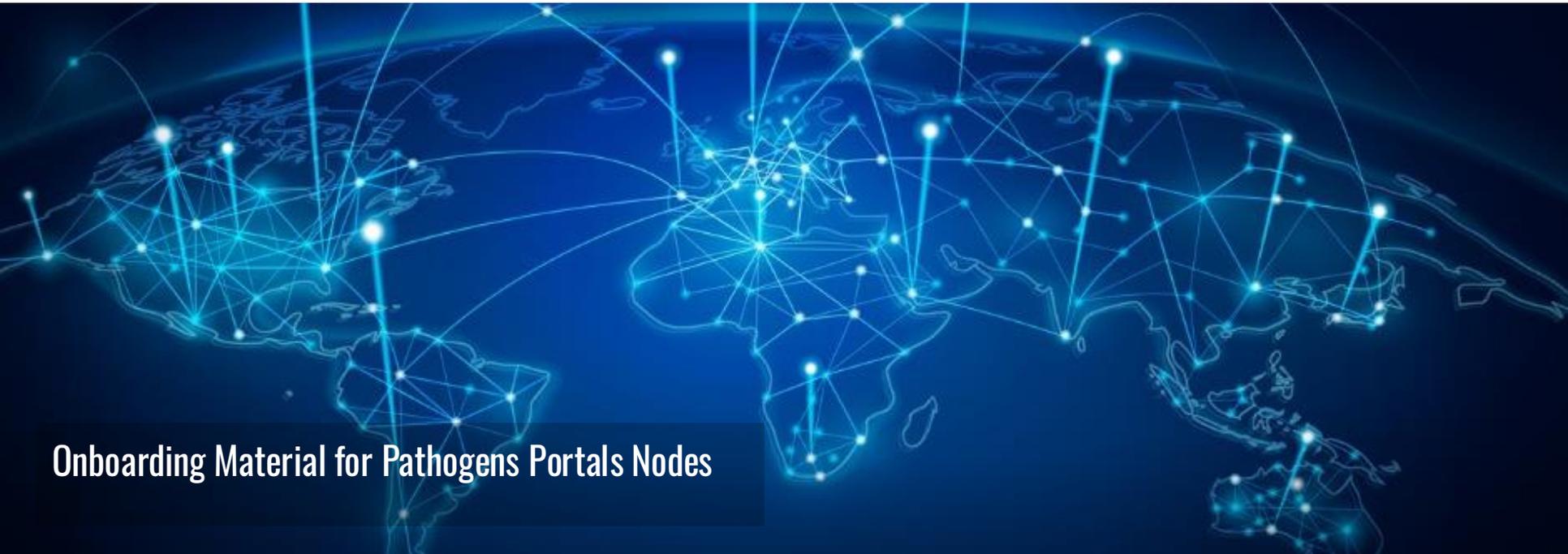




# Pathogen Data Network



Onboarding Material for Pathogens Portals Nodes

# Pathogen Data Network

## Overview



# Pathogen Data Network

An ecosystem of resources to support  
FAIR infectious-diseases data sharing and reuse





It all starts with data. PDN will cover diverse biodata types, including host and pathogen genomics, transcriptomics, proteins, pathways and networks, imaging and cohorts.

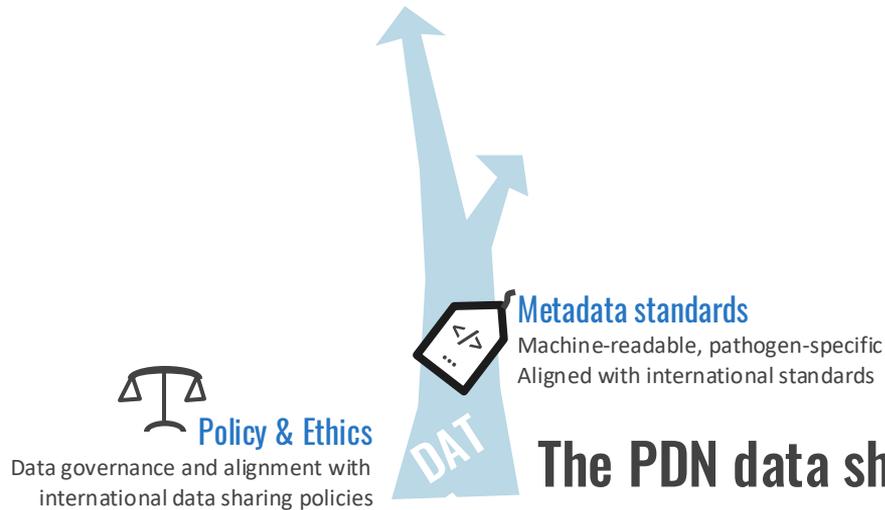
## The PDN data sharing journey

 **Policy & Ethics**  
Data governance and alignment with  
international data sharing policies



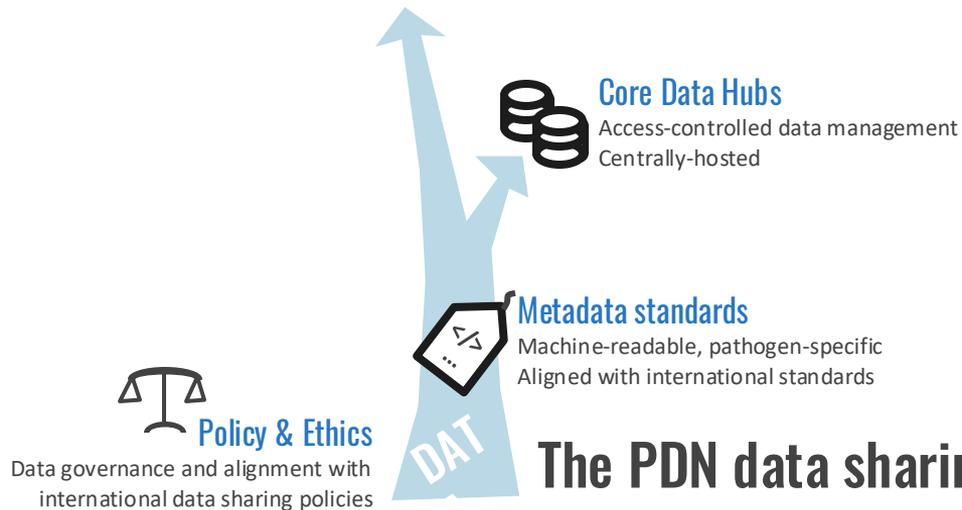
## The PDN data sharing journey

Data sharing cannot be uncoupled from ethics and international policy. We need to carefully consider these aspects throughout the data journey, to ensure benefit sharing and trust.

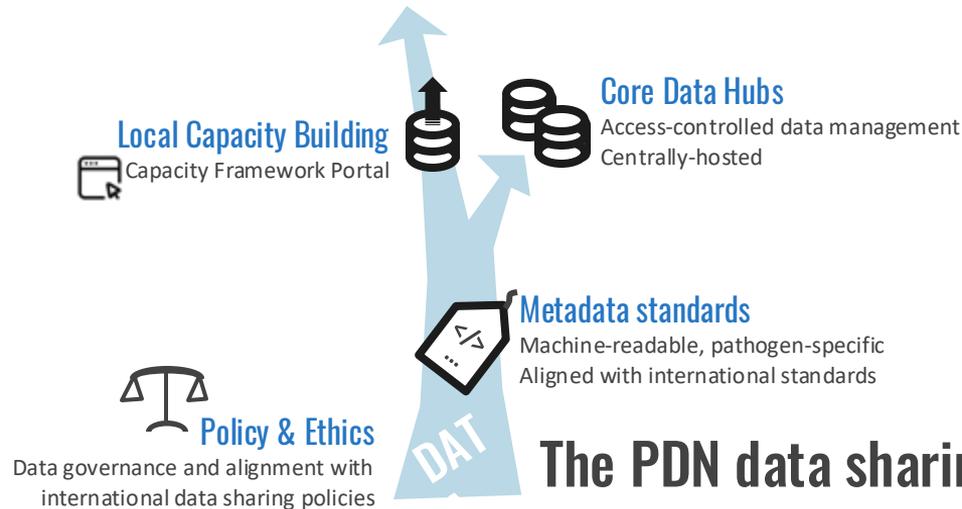


Data need to be standardized against international standards to ensure interoperability. These standards should be usable by machines, to support automated data validations.

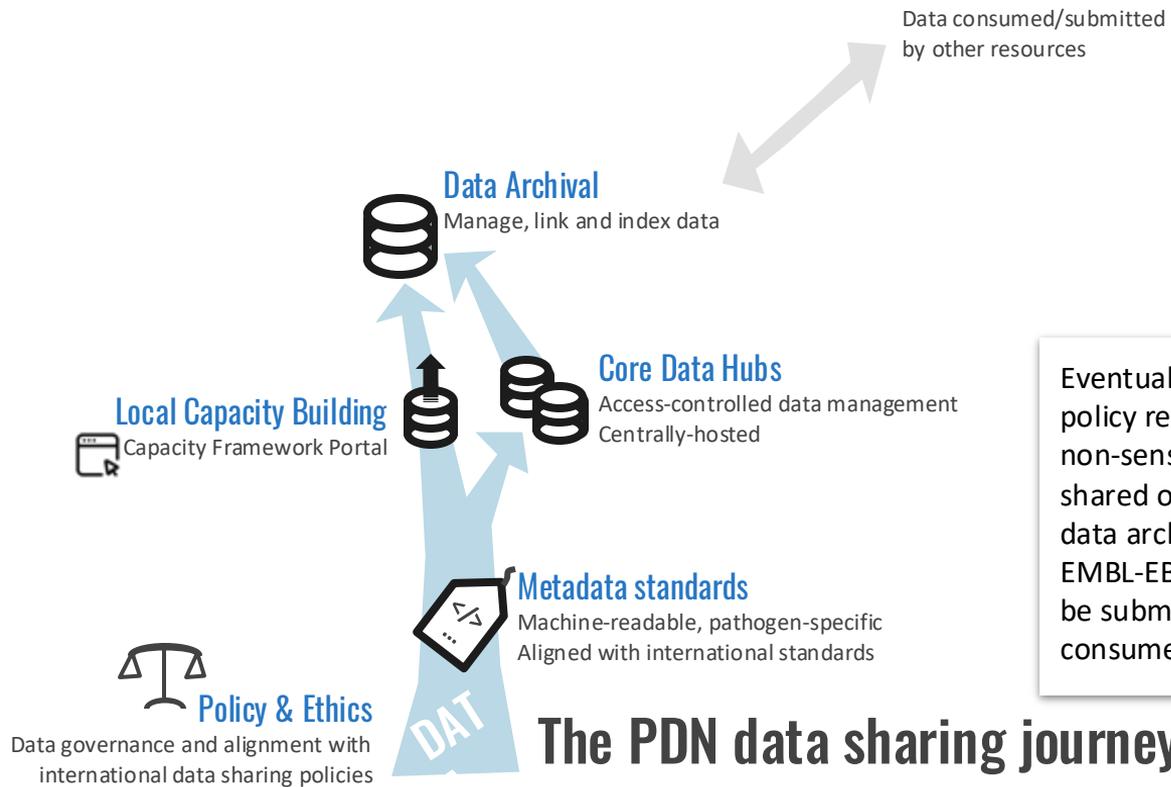
## The PDN data sharing journey



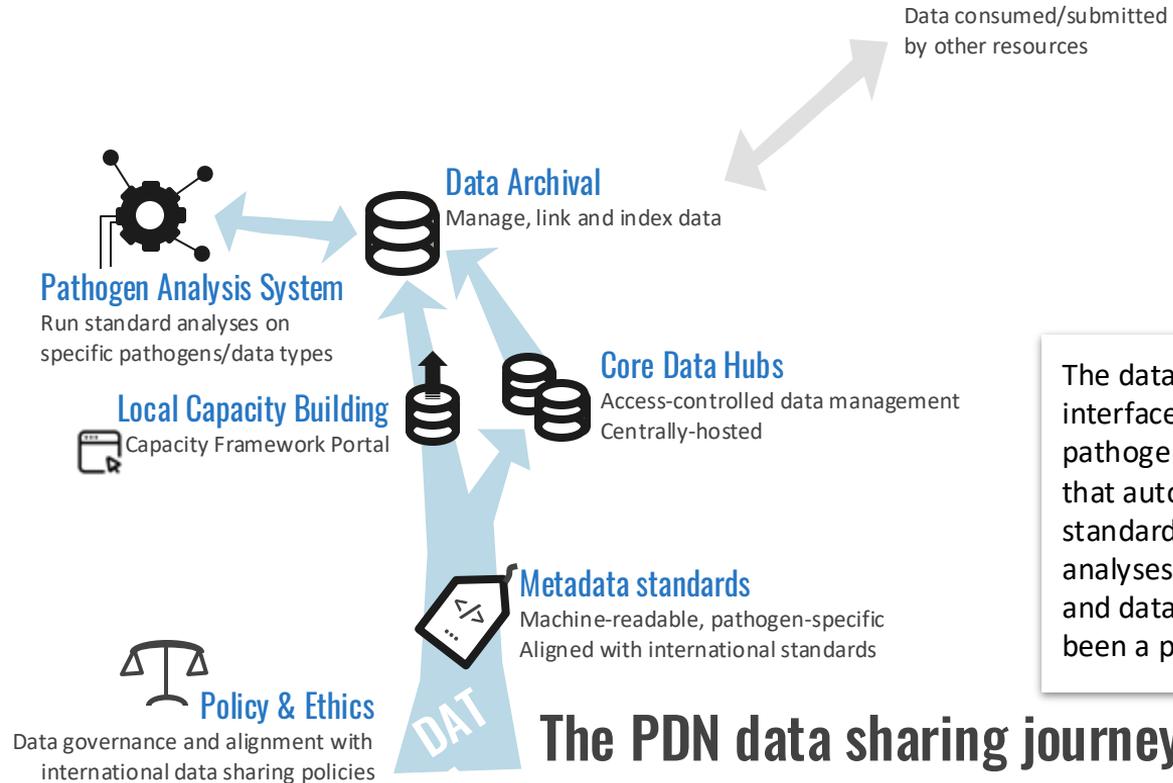
Data may transit through a data hub with time-limited controlled data access. EMBL-EBI will provide central hosting of pathogen data hubs with access control, analyses and visualizations.



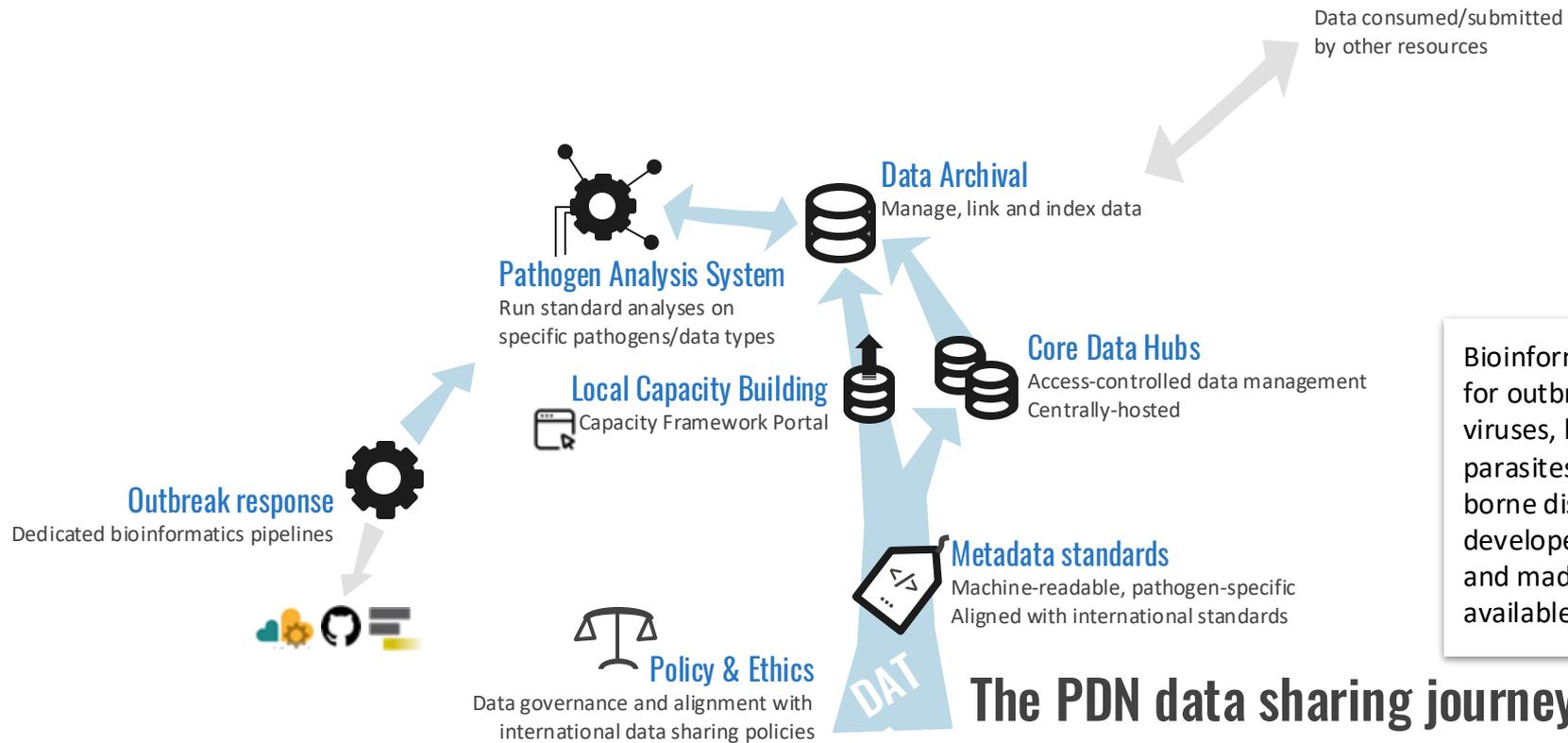
Data may transit through local data hubs (e.g. at national level), that centralize data and may host personal data not to be openly shared. We will support capacity building for local hubs.



Eventually (according to policy recommendations), non-sensitive data are shared on the central data archive hosted by EMBL-EBI. Data may also be submitted directly or consumed by others.

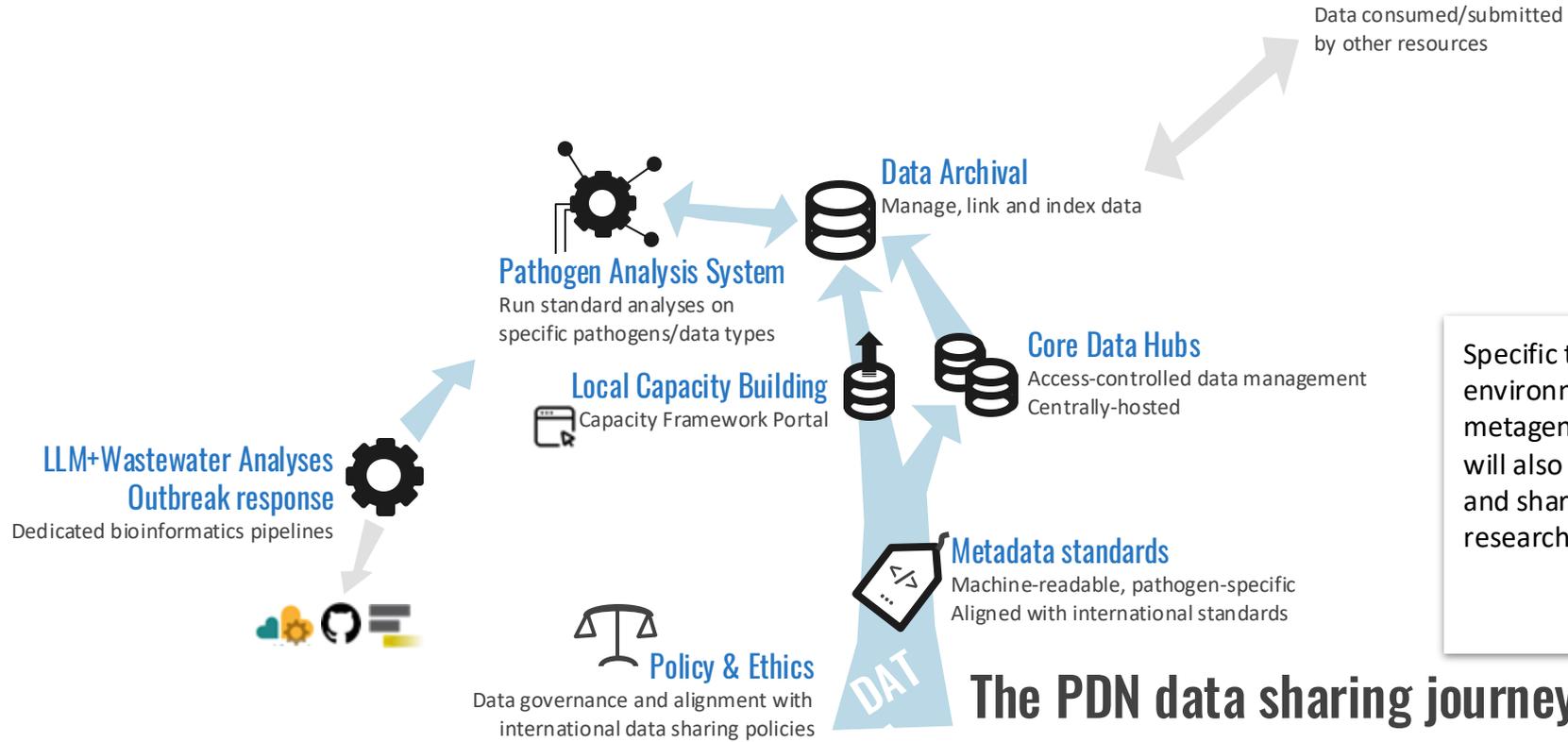


The data archival is interfaced with the pathogen analysis system that automatically runs standard bioinformatics analyses on pathogens and data types that have been a priori specified.

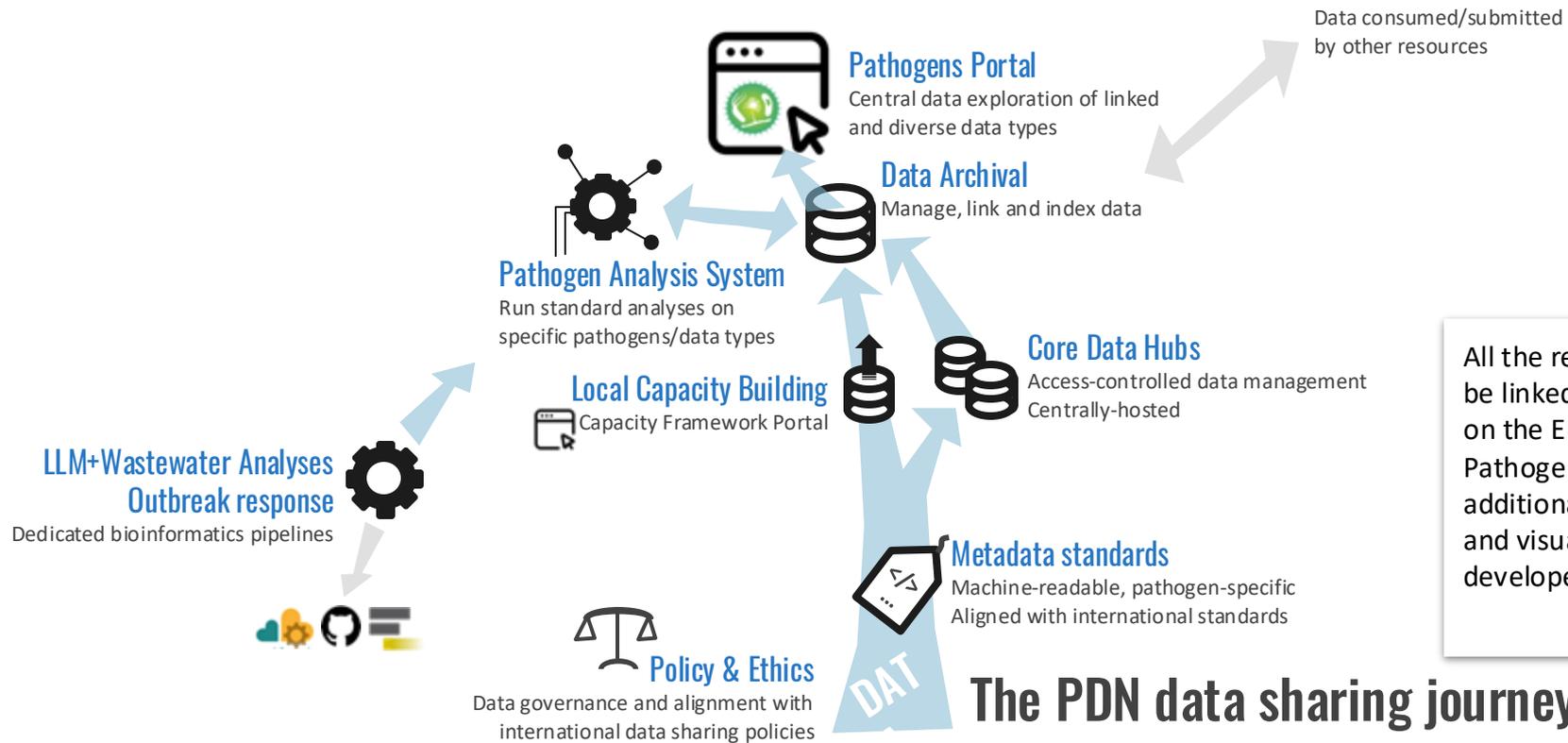


Bioinformatics pipelines for outbreak analyses in viruses, bacteria, fungi, parasites, and vector-borne diseases will be developed, integrated, and made publicly available.

# The PDN data sharing journey



Specific tools to advance environmental metagenomics analyses will also be developed and shared with the research community.



All the resulting data will be linked and searchable on the EMBL-EBI hosted Pathogens Portal, with additional dashboards and visualizations to be developed.

# The PDN data sharing journey

### Pathogens Portals Nodes

Network of community driven pathogens portals  
Based on reference software implementation



### Pathogens Portal

Central data exploration of linked and diverse data types

Data consumed/submitted by other resources



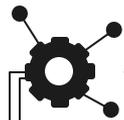
### Data Archival

Manage, link and index data



### Pathogen Analysis System

Run standard analyses on specific pathogens/data types



### Local Capacity Building

Capacity Framework Portal



### Core Data Hubs

Access-controlled data management  
Centrally-hosted



### Metadata standards

Machine-readable, pathogen-specific  
Aligned with international standards



### Policy & Ethics

Data governance and alignment with international data sharing policies



### LLM+Wastewater Analyses Outbreak response

Dedicated bioinformatics pipelines



To support local communities, the Portal will be interfaced with a network of local pathogens portals, to support local researchers in data discovery and good FAIR practices.



## The PDN data sharing journey

### Pathogens Portals Nodes

Network of community driven pathogens portals  
Based on reference software implementation



### Community of Practice

Share knowledge & implementation ideas



### Pathogens Portal

Central data exploration of linked and diverse data types



### Data Archival

Manage, link and index data

### Pathogen Analysis System

Run standard analyses on specific pathogens/data types

### Local Capacity Building

Capacity Framework Portal

### Core Data Hubs

Access-controlled data management  
Centrally-hosted

### LLM+Wastewater Analyses Outbreak response

Dedicated bioinformatics pipelines



### Policy & Ethics

Data governance and alignment with international data sharing policies



# The PDN data sharing journey

Data consumed/submitted by other resources

Local Portals will be based on a reference software implementation to be provided. Ideas of features and implementation issues will be shared within a Community of Practice.

### Pathogens Portals Nodes

Network of community driven pathogens portals  
Based on reference software implementation



### Community of Practice

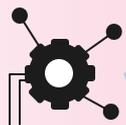
Share knowledge & implementation ideas



### Pathogens Portal

Central data exploration of linked and diverse data types

Data consumed/submitted by other resources



### Pathogen Analysis System

Run standard analyses on specific pathogens/data types



### Data Archival

Manage, link and index data



### Training & Outreach

Maximizing impact and usability of all resources (tools, databases, portals...)

### Local Capacity Building

Capacity Framework Portal



### Core Data Hubs

Access-controlled data management  
Centrally-hosted

### LLM+Wastewater Analyses Outbreak response

Dedicated bioinformatics pipelines



### Policy & Ethics

Data governance and alignment with international data sharing policies



### Metadata standards

Machine-readable, pathogen-specific  
Aligned with international standards

DAT

# The PDN data sharing journey

As data portals and tools are being developed, the Training and Outreach stream will support the development of training material for its users, in particular where gaps are identified.

### Pathogens Portals Nodes

Network of community driven pathogens portals  
Based on reference software implementation



### Community of Practice

Share knowledge & implementation ideas



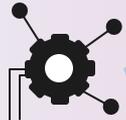
### Open Community Forum

Helps identify priority pathogens  
Roundtable for resources' roadmap



### Pathogen Analysis System

Run standard analyses on specific pathogens/data types



### Local Capacity Building

Capacity Framework Portal



### LLM+Wastewater Analyses Outbreak response

Dedicated bioinformatics pipelines



### Policy & Ethics

Data governance and alignment with international data sharing policies



### Pathogens Portal

Central data exploration of linked and diverse data types

### Data Archival

Manage, link and index data



### Core Data Hubs

Access-controlled data management  
Centrally-hosted



### Metadata standards

Machine-readable, pathogen-specific  
Aligned with international standards



Data consumed/submitted by other resources



### Training & Outreach

Maximizing impact and usability of all resources (tools, databases, portals...)

PDN is community-driven. By joining the Open Community Forum, anyone can contribute ideas and feedback on the tools. Don't miss the opportunity to be part of the design roadmap!

DAT

## The PDN data sharing journey

# Pathogens Portal

## Overview





## Pathogens Portal

Central data exploration of  
linked and diverse data types



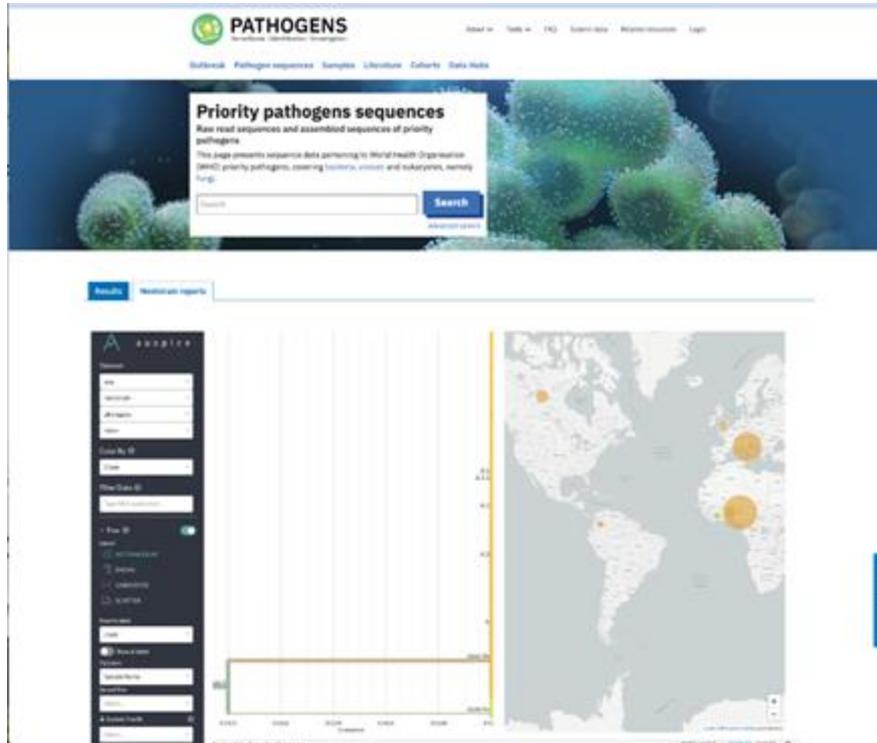
# Pathogens Portal <https://www.pathogensportal.org/>

- Current focus on sequence data
- Extend to diverse biodata types (based on Covid-19 Data Portal):
  - host and pathogen genomics,
  - Transcriptomics
  - Proteins
  - pathways and networks
  - imaging data
  - cohorts
- 1,000s of pathogen species
- Outbreak section to focus specifically on priority pathogens.
- Emerging user base

year	requests	uniq hosts	pages
2023	163,382	5,960	139,077
06/09/2024	665,884	8,629	478,017



# Develop additional visualizations and dashboards



- Auspice-based phylogenomic and geospatial mapping tool integrated
  - <https://docs.nextstrain.org/projects/auspice/en/stable/>
- Additional visualizations and dashboards will be developed based on feedback from Covid-19 Data Portal users and from the PDN Open Community Forum.



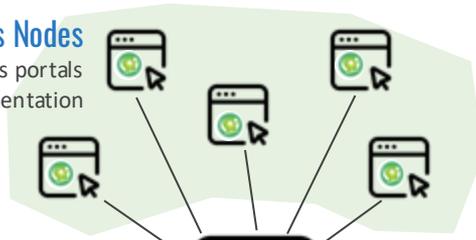
# Pathogens Portals Nodes

## Overview & Examples from Sweden, Norway and Switzerland



## Pathogens Portals Nodes

Network of community driven pathogens portals  
Based on reference software implementation



## Pathogens Portal

Central data exploration of  
linked and diverse data types



# Pathogen Portal Nodes (PPN)

The main Pathogens Portal website features a search bar at the top with the text "Pathogens Portal: Pandemic preparedness through open biodata". Below the search bar, there are several data sections: "Outbreak" with 2,346,479 records, "Sequences" with 17,713,372 records, "Samples" with 3,634,702 records, "Literature" with 136,302 records, "Cohorts" with "The Pathogens Portal Cohort browser presents infectious disease cohort", and "Data hubs" with "Browse and search Data Hubs". A "Latest news" section highlights "Swiss and Norwegian national pathogen portals launched" from 4 Nov 2016.

pathogensportal.ch

The Swiss Pathogens Portal dashboard includes a "GSP Dashboard" and "viSPECTRUM" section. It also features a "Swiss Pathogens Portal: supporting researchers for pandemic preparedness" header and a "Dashboards" section with various data visualizations and links to resources like "Pathogens Portal: Pandemic preparedness through open biodata" and "Swiss Pathogens Portal: supporting researchers for pandemic preparedness".

pathogens.se

The Swedish Pathogens Portal dashboard includes a "CRUSH" section and a "CRUSH-COVID-19" app. It also features a "Dashboards" section with various data visualizations and links to resources like "CRUSH-COVID-19" and "CRUSH-COVID-19 app".

pathogens.no

The Norwegian Pathogen Portal dashboard includes a "Pathogen Portal Norway" section and a "One Health" section. It also features a "Dashboards" section with various data visualizations and links to resources like "Pathogen Portal Norway" and "One Health".

- Discovery of local data and services to support data management, analysis, and sharing.
- PDN will provide a reference software implementation for rapid deployment.
- Led by local communities.
- Aims to serve local research communities.
- Can be deployed in local language.
- Interfaced with the central Pathogens Portal (e.g. to allow data catalogues to be discoverable).

# PPN Examples

## Data management

The Swedish Pathogens Portal / Data management

"Research data management concerns the organization, storage, preservation, and sharing of data that is collected or analysed during a research project. Proper planning and management of research data will make project management easier and more efficient while projects are being performed. It also facilitates sharing and allows others to validate as well as reuse the data. As funding agencies are recognizing the importance of research data management and some now request Data Management Plans (DMP) as part of the grant application process." (NIH)

Here we give an overview of available resources regarding research data management relevant Swedish researchers working on COVID-19 and Pandemic Preparedness research.

## Available datasets

- Biochemistry (34)
- Drug Discovery (15)
- Health (98)
- Imaging (7)
- Protein (43)
- Public Health (63)
- Social Science (59)

## PATHOGENS

Swedish

### Swedish COVID-19 Sample Collection Database

Filtering tools

Name	Type	Size	Material
AxioV-19	Respiratory sample collection	1,000 - 10,000	Ready to use
The Svalbard archipelago COVID-19	Respiratory sample collection	1,000 - 10,000	Ready to use

## Pathogen sequences

60,364 results

View all 31,467 results in Sequences | View all 5,188 results in Analysis | View all 16,259 results in Raw reads | View all 11,450 results in Assembly

## Samples

17,507 results

View all 17,507 results in Samples

## Outbreak

12,246 results

View all 236 results in Sequences | View all 338 results in Analysis | View all 5,880 results in Raw reads | View all 5,792 results in Samples

## Datasets available from research groups in Switzerland

The table below lists the latest publications from research groups in Switzerland that are on Europe PubMed Central.

### Data sources

#### All data (1920)

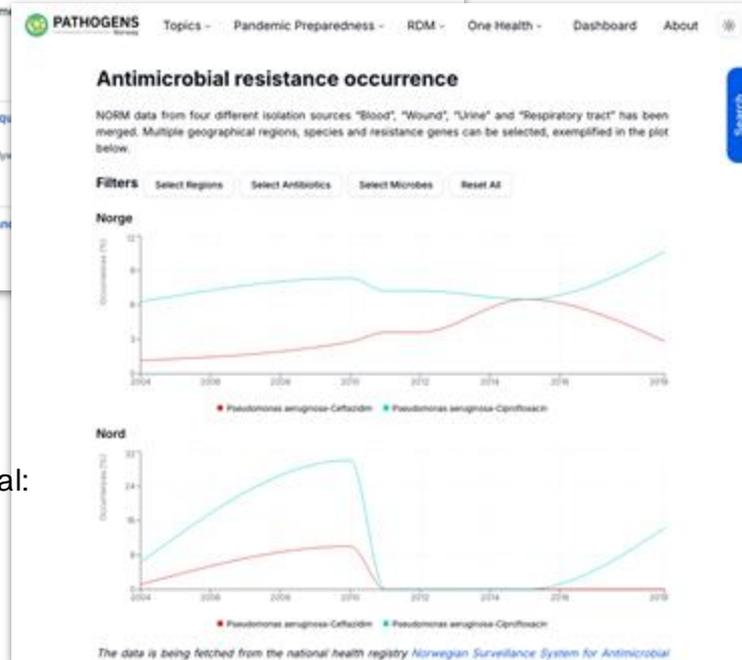
Figshare (65)  
Zenodo (235)  
GitHub (722)  
Dryad (177)  
Gene Expression Omnibus (151)  
Protein Data Bank (296)  
Proteome Xchange (15)  
SASBDB (3)  
Electron Microscopy DB (76)  
ENA (72)  
Open Science Framework (265)

10 entries per page

### Publication

"I will take part in the revolution with our people": a quack resistance after the 2021 Myanmar coup d'état.  
Haar RJ, Crawford K, Fast L, Win TH, Rubenstein L, Blanchet K, Liley undefined  
DOI: 10.1186/s13031-024-00611-7

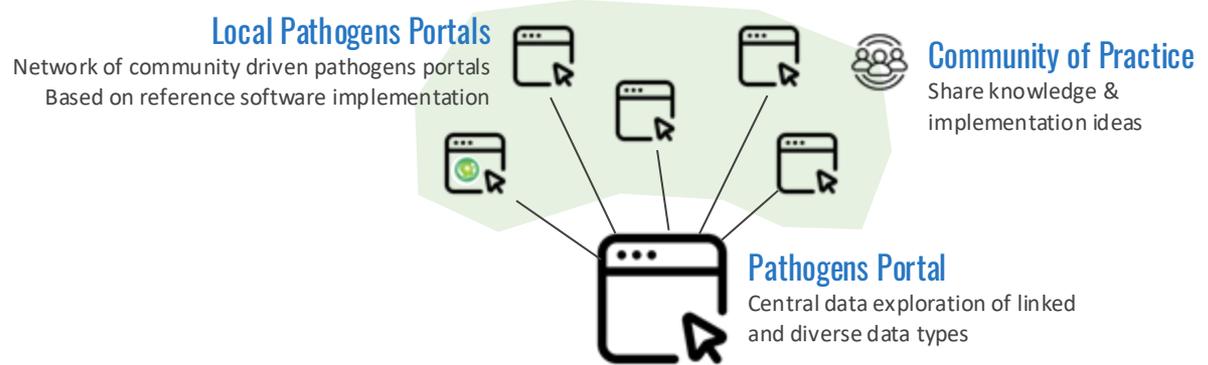
"The Tragedy of the Commons": How Individualism and...  
Hassani Y, Levy A, Gur T, Confino D, Segal S, undefined  
DOI: 10.3389/fpubh.2021.627559



- Swedish Pathogens Portal: <https://www.pathogens.se/>
- Norwegian Pathogens Portal: <https://pathogens.no>
- Swiss Pathogens Portal: <https://pathogensportal.ch/>

# Node portals have dual roles

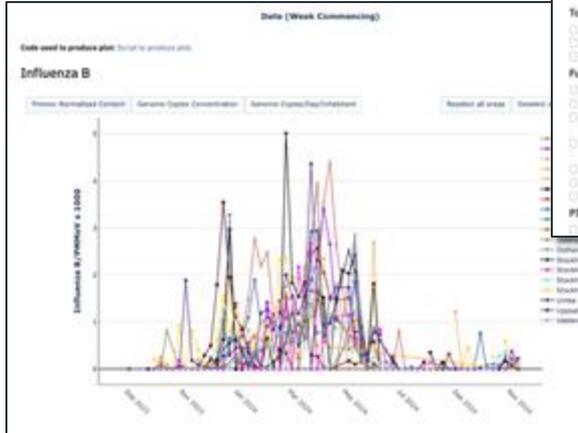
- Center for local data sharing and local community driven content
- Linking local community to international data resources (Pathogens Portal)
- Facilitating access to international data sharing resources (Pathogens Portal)
- Providing services, tools, guidelines, information to local community
- Hosting community driven content in a lightweight framework
- Supporting local users for data sharing and research data management support
- Local focal point in outbreaks



# PPN Examples

The screenshot shows the 'PATHOGENS Sweden' website. The main heading is 'Ongoing research projects'. Below this, there is a list of projects with details on their topics, funders, and PI affiliations. The website also features navigation menus for 'Data Dashboards', 'Share & Access Data', and 'Research & Funding'.

Local data in international Pathogens Portal



Dashboards for local wastewater surveillance

Databases for local research projects

The screenshot shows the 'PATHOGENS Switzerland' website. The main heading is 'Available Datasets'. Below this, there is a section for 'Access-controlled datasets' with a search bar and a list of datasets. The website also features navigation menus for 'Data Dashboards', 'Share and Access Data', and 'Analyze data'.

A grid of logos for various Norwegian health and research institutions, including HelseDirektoratet, FHI (Folkehelseinstituttet), Veterinærinstituttet, and the Norwegian University of Life Science. Each logo is accompanied by a brief description of the institution's role in public health and research.

Facilitate access to locally relevant data resources, research and public health



# Portal development

- PDN will provide base software repositories, provide development guidelines and recommendations
- Built on lightweight Hugo framework
- PPN software roadmap to support local developers in planning ahead
- Network of pathogen portals nodes share common software components, for example visualisation frameworks
- PPNs can align and share locally developed content proving useful for others
- Software components developed also together with Pathogens Portal
- Freedom to fork and develop in direction needed for local community
- Network of PPN developers

The screenshot displays the GitHub interface for the repository `ScilifelabDataCentre/pathogens-portal`. The repository is public and has 4 branches, 67 tags, and 3,193 commits. The main content area shows a list of files and folders with their commit history:

File/Folder	Commit Message	Time Ago
<code>.github</code>	Modify sitemap.xml's URL in patly_test.yaml	2 weeks ago
<code>CONTRIBUTING</code>	adjustments to image documentation	7 months ago
<code>Dockerfiles</code>	Prevent from being iframed	2 months ago
<code>archetypes</code>	Move mockup into a new Hugo site	4 years ago
<code>content</code>	Moved and deleted Biobank forms from portal (#1203)	last week
<code>data</code>	Merge pull request #1185 from ScilifelabDataCentre/Ha...	2 months ago
<code>layouts</code>	Moved and deleted Biobank forms from portal (#1203)	last week
<code>static</code>	Moved and deleted Biobank forms from portal (#1203)	last week
<code>.gitignore</code>	Add patly dependency via devbox and npm	3 months ago
<code>CITATION.cff</code>	Finalising with orcid and minor adjustments	9 months ago
<code>LICENSE</code>	add license	4 years ago
<code>README.md</code>	Fix URLs in accord to visualisations repo refactoring (#1...	3 months ago

The right sidebar shows the repository's metadata, including the name "Swedish pathogens portal", the website `www.pathogens.se`, and tags for `bootstrap`, `markdown`, `hugo`, and `covid-19`. It also lists the license (MIT), activity, and release information (67 releases, latest: `Release 2024-11-15`).

PPN Github repository





# PDN | Pathogen Data Network

[www.pathogendatanetwork.org](http://www.pathogendatanetwork.org)

Please keep in touch, share ideas, feedback!

Contact everyone: [pdn-all@sib.swiss](mailto:pdn-all@sib.swiss)

Contact executive board (PIs): [pdn-eb@sib.swiss](mailto:pdn-eb@sib.swiss)



# THANK YOU!