Roadmap Pathogens Portal Node toolbox 2025-2026

Contributors

The document was created as part of the Pathogens Data Network (PDN) project. It was drafted by Liane Hughes (project leader for the Swedish Pathogens Portal, the Swedish PPN), at SciLifeLab Data Centre, Uppsala University, in collaboration with the PDN Community of Practice (CoP).

Glossary

Term	Definition
PDN	Pathogen Data Network - the consortium conducting the work to, among other things, create a network of Pathogens Portal nodes (PPNs). Funded by NIH (award number U24AI183840).
CoP	Community of Practice formed as part of the PDN. IT comprises individuals from multiple Pathogen Portal Nodes (PPNs).
PPN	Pathogen Portal Node - a local node of the Pathogens Portal, typically developed and maintained by a group/organisation representing a nation. For example, the Swedish PPN (the Swedish Pathogens Portal), is developed and maintained by SciLifeLab Data Centre in Sweden.
Central Pathogens Portal	Pathogens Portal developed and maintained by EMBL-EBI (https://www.pathogensportal.org/)
EMBL-EBI	European Bioinformatics Institute (EBI), part of the European Molecular Biology Laboratory (EMBL)

Summary

The Pathogens Portal Node toolbox comprises minimal code for a PPN, written in Hugo (a static site generator). It was designed to minimise the time and technical expertise needed to launch a

PPN. It also facilitates greater consistency between PPNs, and provides more opportunity for collaboration between them (as it is easier to exchange features developed for any given PPN). It uses a theme inspired by the Swedish and Swiss Pathogens Portals and applies the visual identity outlined by the Central Pathogens Portal.

This roadmap outlines the work to be completed for 1 year, from July 2025.

Stakeholders

Stakeholder	**Involvement**	**RACI**
SciLifeLab Data Centre	Responsible for generating and maintaining the code, as well as the associated documentation	Responsible
Liane Hughes	Project leader for the work	Accountable
PDN	Work together to link PPNs and to onboard new PPNs.	Consulted
PDN CoP	Provides input for the toolbox and on launching a PPN	Consulted
PPNs	Often make use of the toolbox to generate a PPN. Should be aware of changes, and may add/request features.	Informed

Themes/objectives

A. Provision of support

The toolbox and associated documentation are sufficient to enable groups/organisations to launch a PPN with minimal intervention. However, when launching new PPNs, it is not unusual for users to require some support, typically in the form of advice. Existing PPNs also reach out with feature requests and for advice on more advanced features.

B. Clear means of citation

It is important that the toolbox is easy to cite, so that credit can be given when PPNs make use of it.

Motivation

It is important that the use of the toolbox is tracked. To facilitate this, PPNs should be able to cite it in a consistent and traceable manner.

Success criteria

- Get an RRID for the toolbox.
- Link the GitHub repository for the toolbox to Zenodo, so that versions can be tracked and cited.
- Addition of documentation to the GitHub repository on how to cite the toolbox.

Risks

- This must be done quickly, or PPNs will launch without citing the toolbox in a traceable manner.

C. Create links between PPNs and Central Pathogens Portal

One of the aims of the PDN is to ensure that data and information related to pathogens is discoverable throughout the network of Pathogens Portals.

Motivation

Discoverability of data and other research resources is key in accelerating research on a given pathogen. Data will be more discoverable if it can be accessed via multiple resources. In order to aid this, there should be a process in place to enable data shown on a PPN to also show on the Central Pathogens Portal, which is likely to have more visitors than any given PPN. Ideally, this process should be automated in order to minimise the investment needed by both PPNs and the Central Pathogens Portal.

Success criteria

- Identification of a workflow enabling data on a PPN to be shown on the Central Pathogens Portal.
- Documentation and/or a code in the toolbox that would enable PPNs to provide information about their data on the Central Pathogens Portal.

Risks

- PPNs have access to different levels of resources, and it might be difficult to find a way to automate workflows to show data on PPNs on the Central Pathogens Portal.
- There will be a need to standardise how data is provided by PPNs, and there are many different types of data and ontologies used, which might make standardisation difficult.

D. Information on how to remove unwanted features

Each PPN represents a different national community. These communities have different needs, and so each PPN is unique. This means that PPNs will not necessarily use all of the features included in the toolbox.

Motivation

Some PPNs have expressed that they do not want to include particular sections (e.g. news and data highlights). The PPNs want a means to easily hide/remove those features.

Success criteria

- Addition of documentation on how to hide/remove features in the toolbox.

Risks

- Need feedback on the clarity of the documentation from PPNs.

E. Automation for publications

It is useful to have access to relevant publications about a given pathogen. By showing relevant publications from a given country, PPNs can promote national research on that pathogen. This can facilitate national collaborations.

Motivation

Since 2020, the Swedish PPN has maintained a curated database of SARS-CoV-2/COVID-19 publications involving at least one researcher affiliated with a Swedish research institution. This resource is incredibly valuable, and promotes Swedish publications and preprints on SARS-CoV-2/COVID-19 (as well as the data and code shared within it). Other PPNs have expressed interest in generating a similar database. However, this currently represents a significant investment of time and resources. It would be beneficial for all PPNs to be able to automatically collate national research on a given pathogen.

Success criteria

- Production of code that can identify national research on a given pathogen.
- Inclusion of a means of automation in the toolbox, so that PPNs can easily show both publications and preprints on particular pathogens involving researchers from the relevant nation.
- Documentation on automation in the toolbox.

Risks

- It may be difficult to find a way that PPNs can automate the code, as they have access to different computational tools.

- The code will rely on external publication resources (e.g. EuroPMC).

F. Incorporate other technologies and software

Software and technologies are being developed that could be useful for PPNs. For example, the Central Pathogens Portal develops software related to data visualisations for that portal. This could be incorporated into PPNs. Further, there are technologies and software that could be used to generate dynamic data dashboards, which attract a considerable amount of visitors, and effectively promote data, but can be technically difficult to establish.

Motivation

PPNs have limited resources to create features that would draw in more visitors, and make the resource more useful to the research community. This includes dynamic data dashboards, which require technical expertise to create and maintain, and represent a significant investment of resources of PPNs. It would be beneficial for PPNs to be able to make use of other software and technologies that could enable them to create beneficial features without requiring the same level of expertise and investment. If this is possible, it could lead to more visitors to different PPNs, which could fortify the Pathogens Data Network, and ensure more collaborations over a greater period of time.

Success criteria

- Investigation of software and technologies that could be suitable for producing features such as dashboards, that can draw in a significant number of visitors..
- Documentation on the best way to incorporate the selected software and technologies.

Risks

- There is a reliance on other technologies and software, so things should be selected that have long-term sustainability.
- It may not be possible for PPNs to incorporate the software and technologies.. If this is the case, ‘best practice’ advice on using some software/technology to produce desired features can be given instead.

(stretch goal) G. Implement contact form

It can be useful to have a contact form on the site, so that users can easily reach out.

Motivation

Contact forms often make it easier for users to get in touch, compared to an email address alone. The toolbox does not currently contain any contact form functionality.

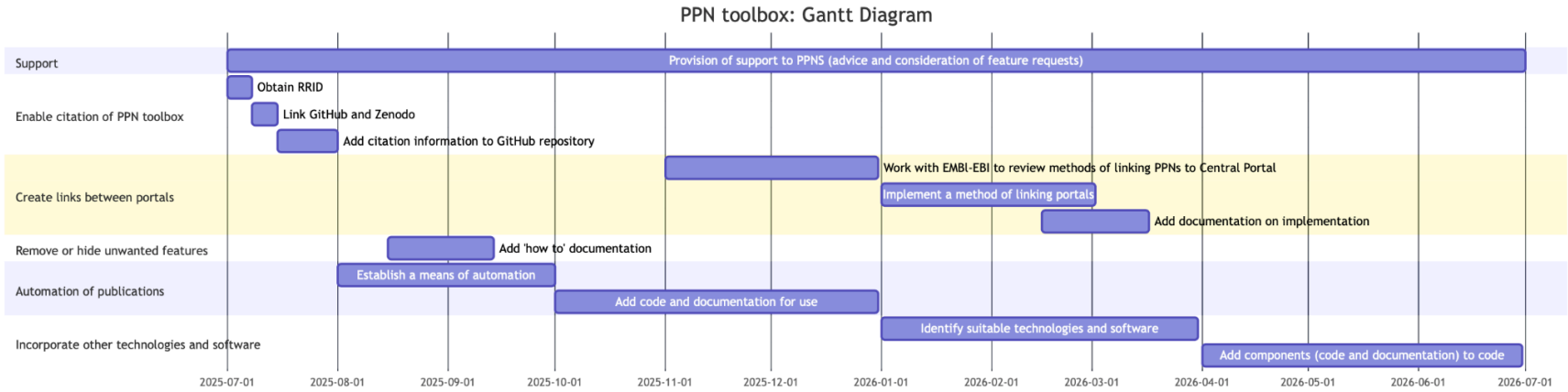
Success criteria

- Addition of a contact form functionality to the toolbox

Risks

- Forms can require some dynamic functionality, and it can be difficult for PPNs to implement in a static framework given limited resources.

Phases & milestones



Source: [Mermaid Markdown]

Resources

People

- This work will be undertaken by the SciLifeLab Data Centre.
- It will require 1 FTE to complete, and should involve a developer familiar with the code.

Risks & dependencies

- Dependent on feedback and usage from PPNs and other stakeholders from the PDN
- Dependent on the acquisition of funding to continue PDN
- Risk that insufficient funding will be acquired to enable this work to be conducted as planned
- Dependent on other technologies for dashboard and publication automation.

Stakeholder communication

Stakeholder	**Communication**
SciLifeLab Data Centre	Private Slack channel and regular meetings to ensure that this work is completed
PPNs	Email communication, and plan to establish regular meetings to get feedback. Potential new PPNs can be established in contact via GitHub/Swedish Pathogens Portal
PDN	Email communication and regular meetings planned with e.g. Swiss and Norwegian nodes