

To	Board
From	Director
Executive officers	Abdulrahman Azab and Tomasz Malkiewicz
Regarding	PaRI Project Midway Report
Date	23.02.2021

Background

The one year PaRI project was launched on 2020-11-01. PaRI contributes to NeIC Strategic objectives, aligned with PaRI partners objectives, as follows:

- Beneficial collaborations: enhancing pandemic research collaboration among Nordic and Estonian National Providers and possibly among other European countries
- Nordic influence: PaRI aims to enable easy and seamless access to datasets for researchers and possibly non-academic users. PaRI is not tied to COVID-19, but will work on preparing the grounds and expertise for future pandemics
- Motivated people: collaboration of six countries being also the NeIC members (Finland collaborates as an observer) and Germany. At the core of PaRI is also its consortium's expertise
- Effective processes: contributing to improvement of the processes at partners organizations, e.g., dialogue between the researchers and national providers.

PaRI aims at aligning with other COVID-19 European initiatives, e.g. ELIXIR and EOSC-Life, thereby contributing to their strategic objectives. Efficient response to pandemic (COVID-19 and future ones) is a key part of National Strategies of the PaRI partners.

The expected project outcomes are:

1. Repositories suitable for sensitive data enabling FAIR and secure data storage
2. Secure platform for managing and executing cross-border workflows for sensitive data processing
3. Implementation of Nordic pilots that benefit from the technical results.

Progress and Next Steps

The Nordic collaboration on sensitive bio- and medical data facilitated through NeIC is now widely recognized as a unique fit-for-purpose competence network of software developers.

Progress to date:

1. Infrastructure components provide storage both for sensitive data (federated EGA) and non-sensitive data (European COVID-19 data portal; ENA) and support to users enabling them to FAIRify data so that the new data becomes useful and visible.

Status: A collection of FAIR tools is supported for submitting data to ENA. Pilot Federated EGA instances are deployed based on the development and sustainability work in the Tryggve project. In progress: deployment of the operational instances.

Issue: Access to non-sensitive virus data, still a struggle to get all GISAID (initiative promoting the rapid sharing of data from all influenza viruses and the coronavirus causing COVID-19) available data (genome consensus sequence only) deposited into ENA with raw data.

2. A Nordic platform to analyze all COVID-19 genomic data using Galaxy within a research e-infrastructure in the Nordics.

Status: A pilot COVID Galaxy up and running, equipped with COVID analysis tools.

3. A *cloud* in two different national e-infrastructures to scale the analysis. To help users FAIRify data before submission to the European COVID-19 portals, a set of tools and a support channel will be provided. For data presentation a dashboard of Nordic strains will be deployed

Status: Docker mobile Galaxy has been tested in TSD and Computerome. The development of the nordic dashboard is ongoing.

The project partners aim to submit a proposal for an ambitious follow-up project to current NeIC's open call. To reach the ambitious goals set up for the new project 2022-2024, these partners estimate that a budget of around 8 FTE annually will be needed, all funding sources included. Such a budget may be necessary to follow up on Nordic opportunities, providing a Nordic pandemic research e-Infrastructure.

Outlook and issues

Sustainability of PaRI services

The project will produce the following services:

- WP2: Pandemic Galaxy portal with backend cloud HPC, equipped with genomic viral analysis tools.
- WP2: Collection of tools for genomic analysis of pandemic data
- WP3: Nordic Dashboard
- WP3: Mobile pandemic Galaxy for analysing sensitive data (viral and host)

In case of no extended funding for PaRI after October 2021, it is not clear how these "NeIC" services will be sustained. Partners are interested in contributing. But if NeIC does not provide "any" contribution, there can be a risk of these services not maintained. One discussed option is to include these services in the Nordic NT1 service portfolio

Insufficient time to get sensitive data into FEGA before PaRI ends

Part of the project plan is to provide storage and analysis tools for sensitive pandemic data, i.e. host and host+viral data. An outstanding issue is getting the legal approvals to get access to the sensitive data. The process is progressing but there is a risk for not getting access to the actual data before the end of the project, October 2021. In the Tryggve1 and Tryggve2 project we have not faced such a situation since the time period has been three years for each while it is just 1 year for PaRI, due to the "urgent small funding" call context.

Recommendation

To receive the Midway report.