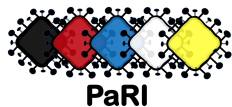# PaRI Final Report

The report is drawn-up in agreement between NeIC as the project owner represented by Tomasz Malkiewicz and the project manager Abdulrahman Azab. It is verified through a steering group decision.

	Name	Partner/Activity	Date
From	Abdulrahman Azab	PaRI	2021-11-01
Reviewed by	Tomasz Malkiewicz	NeIC	2021-11-20
Approved by	PaRI SG	PaRI	2021-11-29

PaRI

◈neic

Edition History:

Issue	Date	Comment	Author/Partner
0.1	2020-10-29	First draft by the project manager for review by the project team	Abdulrahman Azab
0.2	2020-11-01	Second draft by the project team for review by the project owner	Abdulrahman Azab, NO, Project Manager Björn Grüning, DE Christine O. Rasmussen, DK Fabian Leonardo Martinez Bernal, NO Giani Conzett, DK Johan Viklund, SE Kjell Petersen, NO Wolfgang Maier, DE Ott Eric Oopkaup, EE Ulvi Gerst Talas, EE Wolmar Nyberg Åkerström, SE

Abstract:

This report summarizes the achievements of the NeIC PaRI project (2020-2021) towards the objectives, and documents experiences and recommendations for improvements in the working methods of future projects.
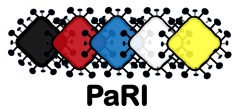
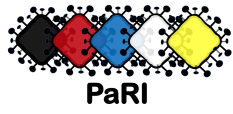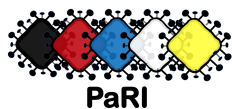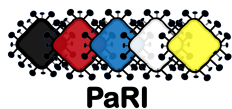| Comprehensive information about the project		
Type of project	Collaboration project	
Scope	Result	0,35
	Time	0,25
	Cost	0,4
Documentation location	https://neic.no/pari (external) https://drive.google.com/drive/u/0/folders/1A3Ns_WcQHdcQ7vuEQCgi_2CY5xpUZJTb (internal)	

Table of Contents
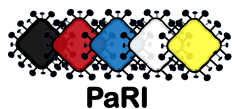
1. Basic information

1.1. The project

By embracing and aligning with other European initiatives who contribute to the struggle against COVID-19, such as ELIXIR and EOSC-Life, PaRI facilitates pandemic research for Nordic researchers. The one-year project has built a secure platform for managing and executing tools and workflows for cross-border pandemic data processing. Data processing means collecting, quality controlling, analysing and securely storing data. Making it possible for multiple organisations in the Nordic region to have the same processes enables more seamless and efficient collaboration in further research. One of the PaRI project's objectives was to implement Nordic pilots and best practices for FAIR data that benefit from these technical results. The project benefits greatly from the outcomes of NeIC's Tryggve project.

NeIC PaRI project contributed to NeIC Strategic objectives, which are aligned with PaRI partners objectives, as follows :

- Facilitate collection and storage of sequence and other human related COVID-19 data
- Facilitate Nordic analysis of Nordic pandemic data using Galaxy and a research e-infrastructure in the Nordics
- Deploy Secure cloud infrastructure for sharing and presentation of pandemic research datasets and results
- Contribute to building the evolutionary tree of viral strains, based on RNA sequences of patient samples infected with coronavirus as they become known and shared.
- Intend to make the viral and host data collected during pandemic findable, accessible, interoperable and reusable (FAIR) by supporting seamless cross-border access, and adopting data representation standards by existing federated sharing platforms, e.g. EGA.
- Work in close collaboration with other nordic e-Infrastructure projects in addition to EU projects to support the use cases.

PaRI partners are: University of Oslo (UiO), University of Bergen (UiB), Technical University of Denmark (DTU), Estonian Scientific Compute Infrastructure (ETAIS), German Network for Bioinformatics Infrastructure (de.NBI), National Bioinformatics Infrastructure Sweden (NBIS).

Among the most important achievements of the project and its collaborators is the Galaxy Nordic COVID-19 portal. The portal monitors the public output of viral genome sequencing projects that is submitted to the ENA and makes the data more accessible for further use. The portal also demonstrates the feasibility of tracking the output of national genome surveillance projects. The Nordic COVID-19 portal is part of the international Galaxy network. Galaxy helps researchers perform computational analyses with their data without having to have expertise in informatics,

programming or data visualization. The platform is open source and enables accessible, reproducible and transparent computational research.

Another output of the PaRI project is the PaRI dashboard, a visualisation tool designed to help stakeholders such as epidemiologists, state institutions and researchers monitor the pandemic locally, down to regional and municipality levels. The dashboard is based on the NextStrain tool and visualizes Nordic COVID-19 data (raw sequencing data brokered into ENA or consensus sequences from the Galaxy pipeline). The dashboard visualizes the data as well as any connected metadata, allowing for filtering of created visualisation. Furthermore, it even allows adding sensitive metadata to the sequences, with the key feature that this additional metadata will only be stored in the user's browser and thus alignes with any local legislation about sensitive data sharing.

1.1.1 Work Package description

The project work is divided into 3 activities that each have one delivery. Thus in the remaining we use the work deliverable to refer to these activities. Each of the deliverables are responsible for preparing a set of delivery objects as results of their work. The activities of the project are described below:
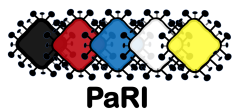
WP1 Data Management

This work package targeted facilitating collection and storage of sequence and other human related COVID-19 data. It included two tasks:

Task 1.1 Mirroring public viral data to local infrastructures: We endorse all researchers to submit their research samples to public archives, like SRA and ENA for sequencing data or PRIDE for mass-spec data and EGA for sensitive data. Having all SARS-CoV-2 data freely available, allows for real-time analysis and tracking of the virus, but also enables an unprecedented turnaround in analysing the genome and discovering new drugs or vaccines. However, the pure amount of data can be challenging to deal with and downloading 3000 (at time of submission) sequences can be time consuming and error-prone on its own. To make it very convenient for Nordic researchers to access all public resources at once we will set up a system similar to the one of the European Galaxy server[1] and mirror all public data locally in our Nordic clouds. This will also address the issues of Findability and makes the data FAIR.

Task 1.2 Sensitive patient data: There is also COVID-19 host data that is sensitive. Here, the Federated EGA provides a solution, and NBIS (ELIXIR-SE) has set up the EGA-SE node that is technically ready and able to receive sensitive COVID-19-related data for GDPR-compliant controlled access. The Nordic countries have long-term collaboration within ELIXIR that will be important for fast and agile systems development and COVID-19-related additions. The work will be performed in close collaboration with ELIXIR-CONVERGE extension). According to the

---

[1] https://bit.ly/usegalaxy-eu-covid19-data

current information, multiple teams in Estonia have filed requests for access and ethics committee approval to analyse SARS-CoV-2 clinical diagnostic sample frozen material and couple this information with patient data in genbank, if available for those individuals. Computerome (ELIXIR-DK) has set up a secure cloud using Cloudera platform to acquire multiple datasets for pandemic related data and will to make them FAIR and available in a controlled GDPR-complaint access

WP Deliverable (D-1): Infrastructure components providing 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.

WP2 Facilitating Nordic analysis of Nordic pandemic data using Galaxy and a research e-infrastructure in the Nordics

The target of this work package has been to Obtain access to all Nordic sequences, collected in WP1 and establish a flexible and functional interface with existing tools set up for analyses of COVID-19 genomic data and undertake evolutionary mapping of all available data. Use that as a backbone to facilitate further samples to be mapped and analyzed in the context of Nordic strains, and to also map Nordic interchange[2]. It included two tasks:

Task 2.1 Nordic Pandemic Research Infrastructure: Deploy a Nordic Galaxy pandemic portal. This will use the Open Infrastructure based deployment by the European Galaxy team. For HPC backend interaction, a pulsar server will be deployed on at least two Nordic clusters, section 2, and on virtual clusters on NREC cloud. ELIXIR-Belgium designed a software for uploading data from Galaxy to ENA [3]

Task 2.2 Analysis of Nordic COVID-19 genomic sequence data: Deploy all the workflows that have been developed by the Galaxy community as a response to the COVID-19 threat into a Nordic Galaxy pandemic portal. Some steps of the workflows can be run using Galaxy interactive environments, similar to https://live.usegalaxy.eu, where interactive tools are secured in containers.

Deliverable (D-2): A Nordic platform to analyze all COVID-19 genomic data using Galaxy within a research e-infrastructure in the Nordics, to develop a resource to undertake evolutionary mapping of all available data, and use the results as a backbone to facilitate further samples to be mapped and analyzed in the context of Nordic strains, to also map Nordic interchange, and to perform a virtual screening of the SARS-CoV-2 main protease.

WP3 Secure cloud infrastructure for sharing and presentation of pandemic research datasets and results

The objective of this work package was To deploy a cloud environment that would be treated in principle as a federated system for sharing both public and sensitive information as well as

---

[2] https://nextstrain.org
[3] https://github.com/ELIXIR-Belgium/ena-upload-container

providing the possibility to compare multiple epidemiological programs in the course of data analysis. Help users to FAIRify data so that they become usable, and help users to publish data in international repositories.

T3.1 Secure cloud: Design a blueprint for a cloud solution for use in the current COVID-19 outbreak that will be available for future outbreaks regardless of their size, using trusted data and secure standardised tool sets to enhance the analyses with the user's own data and customized methods.

T3.2 FAIR data support: Provide a set of tools and a support channel to help users FAIRify data before submission to the European COVID-19 portals[45] (ENA for viral nucleotide sequences) and federated EGA nodes for sensitive host data. In addition, harmonise data from WHO

T3.3 Nordic strains dashboard: Deploy a dashboard for Nordic strains, similar to GISAID[6,7] but with improved precision of location mapping (municipality of residence would be desirable), date stamped and annotated with age and gender of the affected individual. Ability to trace existence or lack of transmission of a strain between specific regions would be of future value when evaluating adequacy of lockdown measures.

Deliverable (D-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 da

1.2.   Background and Business Case

The ongoing COVID-19 outbreak has induced an unprecedented near real-time exchange of information with an exceptional number of individual researchers performing multiple analyses in parallel, using publicly available, as well as their own, data. It has, however, also become clear that there is an unmet demand for services based on the real-time comparison of different epidemiological analyses of the same data, as well a major need for exchange of more person-sensitive information, such as geolocalisation, patient data, etc., between authorities, researchers and different countries. The ideal solution, described by the domain experts working with the data, will allow integration of patient-specific and population-wide levels of epidemiological data and combine it with sequencing and later serological information. There is a need for facilitating collection and storage of sequence and other human related COVID-19 data, and facilitating analysis of Nordic pandemic data.

On the EU level, EMBL-EBI and partners have set up the COVID-19 Data Portal, which will bring together relevant datasets submitted to EMBL-EBI and other major centres for biomedical data. The aim is to facilitate data sharing and analysis, and to accelerate coronavirus research. ELIXIR is maintaining a

---

[4] https://www.covid19dataportal.org/

[5] https://www.ebi.ac.uk/ena/pathogens/covid-19

[6] https://www.gisaid.org/epiflu-applications/next-hcov-19-app/

[7] https://nextstrain.org

COVID-19 portal[8] and contributing through its member states to a Europe-wide distributed compute network. Moreover, the ELIXIR communities are providing expert knowledge, which is fundamental in workflow development and driving the pan-european infrastructures. ELIXIR CONVERGE project 2020--2023 aiming at improved data management across Europe. ELIXIR-SE is leading WP1 aiming at establishing a pan-European data management network. ELIXIR-NO is co-leading WP3 aiming at development of a data management toolkit. Within ELIXIR-CONVERGE additional activities aiming at FAIRifying COVID-19-related data are planned.

Nordic collaboration should speed up developments needed and make these faster and more cost-efficient. This would be through speedy collection and high performance analysis of viral data, during the pandemic. The parties have had close collaboration with ELIXIR, especially on human data and data management, and the activities in this project will be coordinated with similar actions in ELIXIR and EOSC-life.

Today biomedical scientists are facing the divide between GISAID[9,10] with multiple data sharing and limited information content restrictions (no raw data stored and no cross-submissions of the same sample data allowed to other archives) or ENA (full raw data submissions welcome and cross referencing to more sensitive patient data in federated EGA, but no graphic dashboard available for immediate display of processed viral strain data geographical distribution) for submissions of fresh viral sequence data. This kind of divide makes a stumbling stone in negotiations between a researcher and the local health authority for release of the COVID-19 diagnostic material for research purposes. The sequencing results submitted to ENA only disappear from the health authorities' radar for an extended period, which might be critical for current pandemic management decision making.

An important achievement has been deploying a dashboard for Nordic strains, similar to GISAID, but with improved precision of location mapping (municipality of residence would be desirable), date stamped and annotated with age and gender of the affected individual, where the non-sensitive datasets can be exposed as soon as they are available. Ability to trace existence or lack of transmission of a strain between specific regions would be of future value when evaluating adequacy of the lockdown measures.

1.3. Summary of PaRI Activities

Project activities have been organised in three work packages:

WP1 Data Management

The work on data management in WP1 can be categorised into three parts - High quality submission guidelines working group, supporting users in general to FAIRifing data and infrastructure components. More details are also provided in the D-1 report delivered earlier in the project.

---

[8] https://elixir-europe.org/services/covid-19
[9] https://www.gisaid.org/epiflu-applications/next-hcov-19-app/
[10] https://nextstrain.org
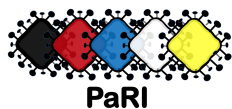
Task Force on High Quality Submissions to ENA

At a team meeting in march 2021, we saw the need to form a subgroup to discuss experiences related to supporting data submissions to ENA and issues with data quality in existing submissions. The contribution to the PaRI deliverables would be to 1) describe the data that would be required to power the PaRI dashboard and 2) Provide guidance to support FAIRification of non-sensitive viral genome data. The idea was that the discussions could be summarised as a document with guidance on SARS-CoV-2 genome submission to Nordic partners and that the document could then be shared as a page on the PaRI website.

The task force was active from March to October and conducted a survey of resources and experiences for supporting data submissions, produced a set of data dictionaries and a mapping from GISAID EpiCoV to ENA. Members from the ELIXIR network and PaRI communities were invited to comment and the experiences were summarised in a document called PaRI Members' Guide to SARS-CoV-2 Genome Data Sharing. The data dictionaries, the mapping and the guide can all be found at the PaRI Project's website.
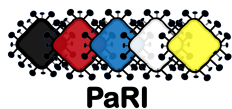
Supporting users FAIRifing data

The main online entry points for users to get in contact with support staff related to PaRI, are the two Nordic Elixir node portals in the ELIXIR Covid19 Data Portal network, https://covid19dataportal.se/ , https://covid19dataportal.no/ and https://covid19dataportal.ee/ . Here both concrete advice on localizing public as well as annotating and submitting own Covid19 related data is provided, as well as contact information on how to get live support from helpdesk staff.

All partners have been contributing to the working group High Quality Submission with focus on harmonizing meta-data content on Covid19 datasets that we are supporting end users to annotate, both for mandatory and optional meta-data fields. This work ranges from annotation of samples themselves through experimental data generation and processing of raw data to consensus sequence. This work has informed and been informed by local activities among the PaRI partners as well as broader collaborations across ELIXIR Converge and has resulted in a list of shared references to existing guidelines, a draft recommendation for extensions to sample metadata to support data integration across the Nordic and a draft metadata mapping between GISAID's EPICoV and EMBL-EBI's ENA databases. Work on the drafted documents continued after the deliverable, while synchronizing with the tasks in ELIXIR Converge WP9 until the project ended.

Outside of the PaRI project all the nodes have done systematic outreach to research groups that have been granted special Covid19 research grants, to inform about services and support that can be provided.

In Norway the main contact for Covid19 data has been towards the National Institute of Public Health, where initial data sequenced by NIPH were shared relatively early with Elixir Norway to test and prepare for submission to ENA after appropriate curation of meta-data annotations. After continuous contact for more than 6 months, a formal collaboration agreement and associated data processing agreements have been finalized. The agreements mandate ELIXIR Norway to designing data flows to remove human host DNA from reads and converting data to formats that can be widely shared with the scientific community, and broker the resulting data to ENA. The first 205 Covid19 samples datasets including raw reads has

been submitted to ENA with Elixir Norway acting as a data broker. The agreement also opens up for possibly more samples from the health regions in Norway, being shared with NIPH, can be submitted to ENA through the same brokering mechanism and infrastructure.

ELIXIR Sweden has been collaborating with SciLifeLab's Data Centre to establish a data brokering service for viral genome data. The main contacts have been towards the National Pandemic Centre (NPC) and Genomic Medicine Sweden (GMS) resulting in a one-time deposit of approximately 800 samples to ENA scheduled for early autumn and a pipeline for continuous submissions of samples with consensus sequences to ENA. A handful of (4) Covid-19 related research projects with sensitive data have applied to be placed on the waiting list for submissions to the [Swedish Sensitive Data Archive](#). While waiting to have their data submitted, the projects have been encouraged to register metadata records with information that can be used to find and access their data in the [SciLifeLab Data Repository](#).

ETAIS and ELIXIR Estonia have taken a long step further from supporting the local Covid-19 initiatives with infrastructure access and know-how and in May entered a tight co-operation with the University of Tartu [KoroGenoEst-3](#) consortia, which is sequencing SARS-CoV-2 for variant detection from preselected high priority Covid-19 positive samples upon request from Republic of Estonia Health Board. The consortia is tackling samples in the need of fast-turnover time and reporting back to the public health authority, ministries and epidemiology contact tracing team where as a up to a few hundred samples a week (depending on the total number of Covid-19 samples detected) are sent for sequencing to ECDC contractor laboratories. The raw sequences of the around 1700 samples sequenced by the KoroGenoEst-1 and KoroGenoEst-2 have already been submitted to ENA. ETAIS is fully in charge of the routine bioinformatics in the current consortia and brokering rights of the samples sequenced under KoroGenoEst-3 agreement have been transferred to ETAIS and ELIXIR Estonia and we will be submitting the raw data along-side with the consensus fasta sequences from the Galaxy pipeline as soon as complete metadata has been assembled. Apart from some backlog samples, the current brokering speed for raw data is within 48 hours from completion of the sequencing run in the wetlab. In addition to the SARS-CoV-2 samples sequenced in the local lab, we have requested and granted download access also to the raw sequence files generated from Estonian samples in the ECDC contractor labs and requested access to the relevant high quality metadata and ENA brokering rights also for this data (so far submitted to GISAID only and arriving with about a months delay). Only the two datasets combined can provide a balanced overview of the COVID-19 situation in the country. The total amount of Estonian SARS-CoV-2 samples sequenced so far must be over 10000.

There are no host genomic samples sequenced alongside with the viral data, but having on fourth of the adult population genotyped by Estonian Biobank, provides an excellent opportunity for matching genedonor data with the Covid-19 samples. Since Estonia has no FEGA instance yet, the human will remain in its current silo.

Infrastructure Components

PaRI related infrastructure components have been built on top of the PaRI partners own infrastructure services.
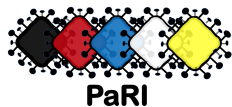
To provide researchers with a data collection of Nordic Covid-19 sequence data (all PaRI partners), the new PaRI Galaxy service has been used as the front end to serve both ENA consensus sequences and direct access to raw reads from ENA submissions. This facilitates easy access for users to directly load data into Covid-19 analysis pipelines developed by the Galaxy Covid19 community, that has been installed on the PaRI Galaxy server, as well as generic download opportunities for researchers that just want to benefit from the data collection and data access mechanisms itself. The consensus sequences has been downloaded by GAS country code from the SARS-CoV-2 database hosted by Elixir Norway (https://covid19.sfb.uit.no/) containing all ENA submitted sequences with curated annotations. The raw read data files are mounted from CRG in Spain's ftp site using the Galaxy feature Pluggable File Source (PFS), that allows the users to extract the exact subset they need for their analysis.

For researchers to organize their own data and collaborate with research partners and support scientists, including PaRI associated staff, other infrastructures have also been utilized. The Norwegian e-infrastructure for Life Sciences - NeLS (https://nels.bioinfo.no) has been central for this in Norway as well as TSD for sensitive data (see next section).

In Sweden data analysis and preparation is (mostly) done in the HPC systems offered by SNIC, Rackham for non-sensitive data and Bianca for sensitive data. Sensitive data will then be stored in the Swedish Federated EGA node hosted in the Sunet cloud which is operated by Safespring. While in Norway the same underlying infrastructure, TSD, is used both by the researchers when doing their analysis and as an infrastructure platform to the Federated EGA node.

Federated EGA - virus sequence data linked with corresponding human genetic data

Very few human datasets with linked Covid19 virus sequence have been identified among the PaRI partners. So far 4 research groups in Sweden have data that may lead to a linked submission before the end of the project.

The FEGA nodes of the Nordic countries of Sweden, Norway and Finland have progressed towards production state, and are among the 5 first countries now preparing for launch as the FEGA federation is being formally established and are members of the FEGA Founding Committee that will become the FEGA Strategic Committee. Preparation of

Technically the nodes have also progressed. Performing tests needed for verifying API communications with CEGA after necessary upgrades of CEGA infrastructure before first nodes can join in production. All three nordic nodes are ready to perform end-to-end coordinated testing of technical and organisational features with CEGA in the second half of this year.

Milestones

All milestones were achieved during the project lifetime:
- M-1-1: All available viral genome datasets in ENA from the partners will be mirrored and made available in local infrastructure.
- M-1-2: Multiple sensitive datasets from at least two different partners, deposited in the federated EGA.
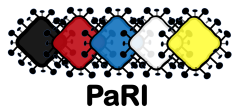
WP2 Facilitating Nordic analysis of Nordic pandemic data using Galaxy and a research e-infrastructure in the Nordics

The work in this Work Package can be described as two major parts, I) the development of Galaxy workflows and automatic execution of these on PaRI partner countries datasets, and II) the work to develop a Nordic Galaxy instance to host workflows and ready processed virus datasets/collection from PaRI partner countries.

At the beginning of the pandemic, Galaxy was used to re-analyze and to assess the reproducibility of the first COVID-19 genome papers and the results were published in the article 'No more business as usual: Agile and effective responses to emerging pathogen threats require open data and open analytics'. The project focuses on underscoring the importance of access to raw data and to demonstrate that existing community efforts in curation and deployment of biomedical software can reliably support rapid reproducible research during global crises. All the workflow descriptions and versions of the software used are detailed in the article.

Later during the PaRI project we have been shifting our focus in monitoring public datasets and analysing them in an automatic way: From public RAW data, to variant calling, consensus building, report generation and populating dashboards and other end-user databases.
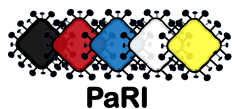
At the end of the projects we could track all data from 4 different European countries, and other countries have been using the develop workflows and infrastructures on its own. Knowledge transfer was initiated with internatinal partners lik VEO and NIH.

The work of the Nordic Galaxy server is described in more detail in the D-2 deliverable report. As the PaRI project was of relatively short duration, it was early on decided to focus on making an easily re-deployable solution, which was very well aligned with the existing set of ansible playbooks in the Global Galaxy project. These were extended with specific code to adapt the front page to a PaRI context, retrieve a list of covid-19 workflows from usegalaxy.eu, install the tools and workflows in the PaRI Galaxy instance, and finally to extract relevant PaRI partner countrie virus sequence datasets from the curated Sars-Cov-2 DB in Elixir Norway to be made available on the server for the end users. As the code was made from the start to be re-usable in case of new settings/situations, the knowledge

Milestones

All milestones were achieved during the project lifetime:
- M-2-1: Establish a galaxy instance with access to covid relevant tools and reference data
- M-2-2: All workflows available in the covid19.galaxyproject.org portal are available in the Nordic Galaxy pandemic portal.

WP3 Secure cloud infrastructure for sharing and presentation of pandemic research datasets and results

The work in this work package consists of three individual parts. Firstly, two independent secure environments have been created, one in TSD in Norway and one in Computerome in Denmark. Both the TSD and Computerome environments are designed for storing and processing person sensitive data. In each of the environments, the Covid-19 flavours of Galaxy has been installed, such that researchers and health institutions can safely process their data with the tools and workflows developed by the Galaxy team. Secondly, the work package aimed to help users fairify their data before submitting to ENA and EGA. This work was done in close collaboration with WP1, through the high-quality submissions (HQS-ENA) work group. Lastly, the Nordic Covid-19 dashboard was deployed using the NextStrain software packages, aiming to give researchers and health authorities a better and more detailed view of the viral strains in the Nordic regions and bringing an interactive tool to their table.

The deliverable D-3 contained all of the major tasks of the work package. The Covid-19 Galaxy software and tools were offered within secure environments in both Denmark and Norway, as well as Estonia who had a head start in having a Galaxy usable for also sensitive data set up on the local infrastructure before onset of the pandemic. The users could get assistance with FAIRifying their data by following the metadata requirements proposed in the HQS-ENA group and by sending their data to ENA. The dashboard extracted Nordic data (primarily Estonian, as other countries' samples were delayed) directly from ENA and offered a more detailed view than existing dashboards (e.g. coming from GISAID).
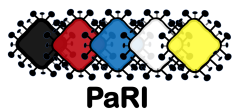
Milestones

All milestones were covered during the project lifetime. Tools and workflows were made available in secure environments (M-3-1), data submission to ENA were facilitated by the Galaxy tools and the support offered by the HQS-ENA group (M-3-2) and the Nordic Covid-19 dashboard was deployed in Estonian infrastructure.

Software services developed by the PaRI team

The PaRI team of systems developers developed a list of software services which are available on public repositories and portable for deployment by other parties.

- **https://github.com/pari-neic/galaxy-playbook**: Ansible playbooks to deploy a PaRI Galaxy instance, with selected analysis workflows from usegalaxy.eu and the tools used by the workflows.
- **https://github.com/pari-neic/pari-neic.github.io**: Adopted web pages to serve as the front page of the PaRI Galaxy server.

.

1.4. Personnel Allocation

Each of the six Partner nodes is asked to allocate 0.5 FTE of personnel into the project. This adds up to 3 FTEs, and thus is co-funded in even parts (50% / 50%) by NeIC and the partner nodes. In addition NeIC funds in total 0.4 FTE for the Project manager. Thus the full volume of the project is 3.4 FTEs.

The distribution of funding and thus work force to organizations in each country will depend on the national situation and division of tasks, and will be set by a service contract between NeIC and the employing organization.
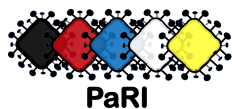
2. Achievement of objective

This section describes the achievements of the project according to the three viewpoints of Results and their quality, Time and schedule, and Cost. These are the same categories that are currently followed in the project portfolio report reviewed by the NeIC director and submitted for the NeIC Board. The PaRI project plan gave the following priorities for these two viewpoints: Result 0.35, Time 0.25, and Cost 0.4.

2.1. Result, delivery objects

Concrete outcomes from the project include the following:

- WP1:
    - DK was stalled as the Government had a different strategy for sequencing and making data public within GISAID
    - EE ETAIS & ELIXIR-EE assist brokering of the SARS-CoV-2 sequence data to ENA following the high quality submission guidelines drafted in the PaRI project.
    - NO established a fruitful collaboration with NIPH resulting in the start of raw sequence reads being submitted to ENA, and contributed to the High Quality Submission guidelines work.
- WP2:
    - EE assists research community and the health authorities with full automation SARS-CoV-2 sequence data processing on the local Galaxy instance
    - NO developed infrastructure as code software to deploy a Galaxy instance with Covid19 specific workflows, the necessary tools and nordic datasets.
- WP3:
    - DK setup a Galaxy server within a secure cloud environment
    - DK contributed to the creation of the dashboard and the data preparation

○ EE all the SARS-CoV-2 quality sequence data from Estonian samples has been integrated into PaRI NextStrain based dashboard with metadata of appropriate resolution.

The Galaxy community, in collaboration with PaRI, organised a workshop on SARS-CoV-2 Data Analysis and Monitoring with Galaxy. The goal was to build capacity, both in the analysis and data management sides of COVID-19, based on the expertise acquired during the pandemic. On a broader scope, sharing the experience in COVID-19 data analysis as well as fostering the principles of open data, open science and open infrastructure are key aspects in the current and global public health situation.

The four-day workshop consisted of multiple demos, hands-on exercises and lectures with topics varying from general information on Galaxy to more specific ones, such as analysis of public datasets and data export to public and open archives. The workshop was conducted asynchronously, with pre-recorded videos, but also providing live support and Q&A sessions. The workshop was attended by researchers in all career stages and different roles, and in total, there were 767 registrants from 106 countries. The PaRI project is helping both Nordic researchers and researchers around the world to better understand and fight the COVID-19 virus by providing the e-infrastructure – platforms, interoperability, tools and training – that facilitates their work. The infrastructure also makes it easier for researchers to benefit from each other's work and work more seamlessly across projects and countries.
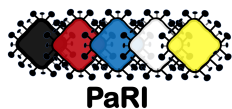
2.2. Time

The project submitted the planned deliverables according to the schedule with very few exceptions. Also the implementations for use cases were done in the agreed schedule. So in summary the project has kept the agreed schedules very well.

2.3. Cost

The project costs were mainly due to the personnel work. Other costs such as travel and meetings were relatively small. NeIC was able to fund a few additional costs that the project made proposals, namely external service used for security penetration testing, and some legal consultation for the use cases program

Personnel cost was followed through tracking the personnel commitments by the partners, to spot any overspending or underspending well in advance. At the end, the project partners committed

resources according to the plan so that the personnel costs were within the budget very accurately.

3. Project execution

The project included tasks on technical development, legal topics, knowledge transfer, and implementation of use cases. The project adopted a pragmatic approach for development of the Nordic platform for sensitive data, so that the services would fulfill the demands of the user communities whether it is about pooling data to one location or about moving the analysis pipelines to where the data is. This is complemented with allowing access to secure services from all Nordic countries, so that the end users can select the best suited platform for their research.

Project execution was organised through deliverables that were defined within the project. Each deliverable has a defined lead, a team, targets and milestones. Follow-up of the deliverables and milestones were done using Trello web tool. This allowed the project to productively work on a wide area of topics by distributing the responsibility of advances to the task owner and teams. In addition to deliverable leaders, the project management has local team leadership, sub-project managers, to on site management level personnel, who helped with allocating tasks and also in reporting of the project results.

4. Transferral of results

As mentioned above, the project produced a number of deliverable reports which detail the results of the project. These were reviewed and approved by the steering group. These are published on the project website along with other reports and documents.

The software that the project developed is mostly available from public repositories, so that they are continuously available to both project partners and the wider community. The updates that were made to the services were taken into the continuous operation of the services.

5. Collected experiences

5.1. PaRI team

The experience from the PaRI team has been very positive. It was highlighted that the team consisting of members representing institutions from various countries enabled efficient knowledge sharing.
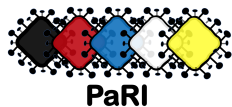
The team spirit has been up-kept through regular online meetings and it proved to work despite the fact that it wasnät possible to organize any face-to-face meetings, except the one at the end of the project.

5.2.   Steering Group

The project overall is considered a success by the Steering Group, significantly contributing to consolidating the Pandemic research community in the Nordics and beyond. Due to a short duration of the project, i.e., 1 year, the Steering Group met every second month, which was considered appropriate, allowing us to discuss and make necessary adjustments to the project plan.

5.3.   The reference group

An efficient reference group of 37 members has been composed and led by the project manager. The RG has worked efficiently through several meetings and interactive sessions throughout the project. The RG members have been several stakeholders representing research communities, EU/Nordic projects with shared interests, and Nordic/Baltic service providers. The Reference Group highlighted that the project achievements were remarkable given the short period of the project execution phase.

6. Recommendations

Such short efforts pose challenges in terms of delivering the results. It is recommended that project calls are for a minimum of 2 years. Having a very active, contributing, diverse, and efficient reference group has been very useful overcoming the short time and small funding obstacle, getting very useful recommendations and even active in kind contributions at times