

# Integrating Structural Biology

# WP6 – Expanding access provision and user services

Lead Beneficiary: P13 - Utrecht

Leader: Rolf Boelens (P13 – UU)

Deliverable D6.5: Implemented software package for remote access EM data

Contractual delivery date: 31 December 2020

Actual delivery date: 30 November 2020

Authors of this deliverable: Frederic Bonnet and Ludo Renault

## **Project objectives**

The objective for this project is to maximise the up-time and usage of cryo-electron microscopes operation 24 hours a day, 7 days a week. It comprises a trial of the remote control of cryo-electron microscopes (cryo-EM) by a user, expert technician or scientist. It also covers the implementation of a pipeline for fast pre-processing of the incoming EM data to allow optimisation of the data collection strategy including 24/7 availability of the equipment (including software for immediate image processing and quality control for data processing). Results of automated data collection will be made available via high- speed pipeline processing of cryo-EM data and on-cloud storage.

### **Executive summary**

We present two solutions for allowing remote control of cryo-electron microscopes at the Netherlands Centre for Electron Nanoscopy (NeCEN). These solutions have been tested with National partner laboratories and could be used by international partners as well. We also present a preliminary pipeline for fast live processing of incoming cryo-EM data providing feedback to the microscope operator as well as a preliminary schematic of a database to organise all datasets collected at the facility.

# 1. Introduction

Electron microscopy is a technique that supports a range of research across scientific disciplines enabling the characterisation and imaging of materials at the atomic level. Cryogenic Electron Microscopy (cryo-EM) is a structural biology technique that allows for the visualisation of biological samples in their near native states. This method is particularly suitable to study pathogens (viruses, bacteria) or develop novel therapeutics (*Peplow, 2017*). It is the method of choice for supporting the development of vaccines or image a particular type of medically relevant proteins (membrane proteins) that represent more than 50% of the drug targets (*Fernandez-Leiro & Scheres, 2016*).

The importance of cryo-EM in research is growing as shown by the rapid increase in publications over the past decade. Between 2012 and 2020, the number of cryo-EM publications has more than tripled while at the same time the number of x-ray crystallography publications (the traditional structural biology technique of choice) has decreased by half (*See reference - Electron Microscopy Databank*). During the last few years, cryo-EM methodology has improved performance dramatically. In 2015, the scientific journal *Nature Methods* voted cryo-EM the "Method of the year". In 2017, the Nobel Prize in Medicine was awarded to Richard Henderson, Joachim Frank and Jacques Dubochet for their contributions to the development of cryo-EM. Indeed, routine atomic scale structure determination of protein complexes by cryo-EM is now possible with high-end electron microscopes due to fast direct electron detectors and newly developed data processing software. Cryo-EM is also the key method to build the bridge between super-resolution life cell imaging and atomic structures of cell components.

Service facilities for cryo-EM such as those present at the different instruct centres are seeing a huge boost in demand for beamtime. It is increasingly important to be able to deliver those services 24h a day, 7 days a week.

Task 6.5 Implemented software package for remote access EM data has proven to be challenging and rewarding at the same time, allowing for continuous access provision during the 2020 covid-19 pandemic.

In this report we outline technical aspects of implementing remote controls of cryo-EM microscopes at NeCEN and a preliminary pipeline for fast live pre-processing, specifically designed to fit NeCEN's infrastructure.

# 2. Technical outline

The developed architecture allows for two possible ways to connect remotely to the cryoelectron microscopes at the NeCEN facility.

The first, is predominantly reserved for internal remote-control access, which allows the user to fully operate the microscope using the control pads while being physically in a different room at the NeCEN facility (and on some occasions, in a different city). This option allows for full control of the microscope as if the user was the operator sitting in front of the microscope including the control pads. This is reserved for advanced users and EM managers of related facilities such as Instruct-ERIC facilities. In the present case we have been testing this approach with a partner in Maastricht as part of the Dutch Electron Microscopy infrastructure.

The second, is a solution for external remote access using a virtual private network (VPN) server connection. One connects via a direct dedicated connector onto the server using a

certificate loaded onto the microscope. With this option, NeCEN controls what is available to the remote user (could be view only, or control over one application). It is thus available for a wider audience of users of the facility.

## 2.1 System requirements

The system requirements to set the connection is relatively simple and requires little effort from the users as shown in Table.1. This consists in installing the OpenVPN software and a VNC viewer in this case we chose OpenVPN and UltraVNC respectively.

Microscope	ιP	Software	Version GUI	Version VPN	Version VNC	Version SSL	Web link
Krios2	External access: 172.17.0.10 Internal access: 10.160.7.14 Network submask:255.255.0.0	OpenVPN	11.10.0.0	Oper/VPN 2.4.6 x86_64-mingw32 Windows version 6.1 (Windows 7)	UltraVNC viewer 1.2.4.0	OpenSSL 1.1.0h	https://openspn.net/community-downloads/ Download linic https://swupdate.openvpn.org/community/teleases/openvpn-install-2.4.6-802.exe
Krios1	External access: 172.17.0.6 Internal access: 10.160.7.14 Network submask:255.255.0.0	OpenVPN	11.10.0.0	OperIVPN 2.4.6 x86_64-mingw32 Windows version 6.1 (Windows 7)	UltraVNC viewer 1.2.4.0	OpenSSL 1.1.0h	https://open.pn.net/community-downloads/ Download link: https://owplate.openvpn.org/community/heleases/openvpn-install-2.4.6-1602.exe

Table 1: Software to setup the connection

The control pads are activated by a software called TARO Simple provided by Thermo Fisher company. On the hardware side, our microscopes are running on Windows 7 Professional. The ethernet must be at least at a 1Gb/s connection to render the frames at a usable rate.

# 2.1.1 The certificate

The two solutions for remote access available at NeCEN are based on the availability of a VPN as described below. To connect to a VPN server, one needs an account, associated with a certificate.

The certificates are issued by our host institution, Leiden university, as the IT department manages our VPN server. But if a facility manages its own VPN server, then it could control the certificate delivery and user accounts. In our case, we are limited by the University choice to manage accounts and VPN server. The certificate is an encrypted key which is readable by the VPN client software which then let the user connect and disconnect at will.

The header specifies the necessary information needed to establish the connection and is read in by the OpenVPN application. In this case a network tunnel which simulates a network layer device and operates in layer 3 carrying IP packets is used over a UDP protocol as a client to connect to the server. At NeCEN a typical header will look like as shown in the following code sample 1.

client dev tun proto udp #remote minla.vuw.leidenuniv.nl 1194 remote 132.229.104.196 1194 remote-cert-tls server auth-user-pass # Verifieer server certificate om MITM tegen te gaan verify-x509-name "C=NL, ST=Zuid Holland, L=Leiden, O=Leiden University, OU=NeCEN, CN=minla.vuw.leidenuniv.nl, emailAddress=NeCEN@biology.leidenuniv.nl" resolv-retry infinite nobind persist-key persist-tun tun-mtt 1500 fragment 1400 mssfix verb 6 key-direction 1 cipher AES-256-CBC auth SHA512 #comp-Izo yes < Encrypted keys and certificate>		
nobind persist-key persist-tun tun-mtu 1500 fragment 1400 mssfix verb 6 key-direction 1 cipher AES-256-CBC auth SHA512 #comp-lzo yes	dev tun proto udp #remote minla.vuw.leidenuniv.nl 1194 remote 132.229.104.196 1194 remote-cert-tls server auth-user-pass # Verifieer server certificate om MITM tegen te gaan verify-x509-name "C=NL, ST=Zuid Holland, L=Leiden, O=Leiden University, OU=NeCEN,	
< Encrypted keys and certificate>	nobind persist-key persist-tun tun-mtu 1500 fragment 1400 mssfix verb 6 key-direction 1 cipher AES-256-CBC auth SHA512	
	< Encrypted keys and certificate>	

Code sample 1. Code sample for a typical OpenVPN header

The header is combined with the certificate and encrypted keys into one configuration file which is then imported into OpenVPN and used to establish the connection.

A virtual networking computer (VNC) server needs to be set up and running at the microscope computer where one can configure a connection password. This password can be changed by the facility directly from the microscope Personal Computer (PC) interface. Once the connection is established a VNC viewer is then used to connect to a given VNC server IP address inside the private network. Then depending on the use of control pads or not, differences exist in the follow-up setup.

## 2.2 Remote connection

#### 2.2.1 From a remote location via the VPN

In this section we explicitly provide two different ways to remote connect to the microscope. Both methods are closely related from one another. In both cases one can also enable the use of the control pads. The setup for such capability is described Section 2.3.

The first case is a remote connection which is done from outside the University. The user can be in The Netherlands, but also in a different country altogether. The user only really needs internet connection and an account/certificate to log into our VPN network. This setup enables someone with authorized access to remotely view, monitor or even setup a basic electron microscopy session (such as single particle data acquisition using EPU software from Thermo Fisher).

The overall schema of the setup is shown in Figure 1.

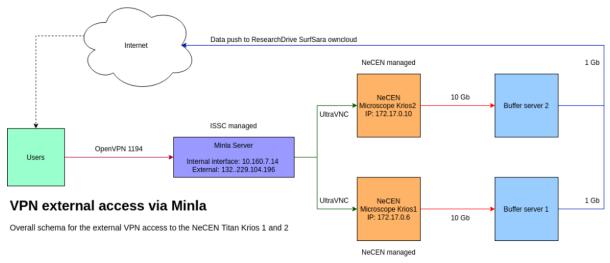


Figure 1. Schematic diagram of VPN access to microscopes at NeCEN

The principle is relatively simple. Using the same software setup outlined in the previous sections, one connects via OpenVPN to the VPN server. Once the connection is established, one then connects to the VNC server on the Microscope PC directly. The interface at the PC can also provide access to virtual control pads, but the facility can also restrict access to the software via user account on the microscope PC.

It is also possible to limit rights to view only, or to a specific window (such as an EPU window) that provides limited controls over the microscope. A local operator is on standby in case he/she would be needed. Once the session is ready and the data collection begins, the data is then transferred to local buffer servers via a 10Gb copper ethernet connection and then to our private partner SurfSara (using ResearchDrive SurfSara owncloud service) at the current available network speed. The data can then be downloaded by the user with their personal secure accounts onto a remote location of their choice.

#### 2.2.2 From a remote location onsite

Using a computer (desktop or laptop), one connects to the microscope PC via an internal IP address. This can be done from a different room at NeCEN (an office to allow social distancing for example) or from outside the University once the user is logged into our Virtual Private Network as described earlier. From there one then connects to the VNC server at the microscope PC and starts operating the electron microscope and setup the session.

Once this is ready and data collection begins the data is then transferred via a 10Gb ethernet copper cable to our storage servers which then transfers the data onto a commercial partner, Research Drive at SurfSara cloud service data centre in Amsterdam. Research Drive mirrors the data which allows for an additional backup. The users can then download the data *via* a created user account in a secure way at their destination of their choice. The overall layout of the setup is shown in Figure 2.

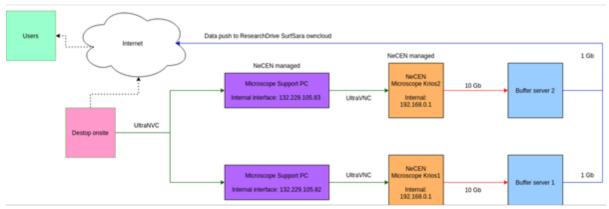


Figure 2. Schema for internal connection via local server

Both connection methods described in Sections 2.2.1 and 2.2.2 can make use of the control pads to control the microscope as if one was sitting in front of it. The setup of this capability is described in Section 2.3.

# 2.3 Remote connection with the control pads

The remote connection with control pads provides access to the microscopes remotely as if one would be physically located at the microscope in the NeCEN premises. The hand panels can be connected and operated remotely using an application already mentioned above called TARO Simple. This application allows full control of the microscope using the pads.

#### The way this works is as follows:

One need to run the software "*TARO Simple*" on the microscope, specify the remote-side IP, that is the one used by VPN to connect and establish the connection, then toggle Taro between the microscope's panels and the one used by the clients. The IP address is the one that is used by the client to connect, which is, in the same IP range as the one used by the microscope. Typically for NeCEN, in the range of 172.17.0.\*, which is not to be confused by the IP address used by Krios1 or krios2 which are 172.17.0.6 or 172.17.0.10 respectively, typically it would be 172.17.0.28 or similar.

#### In an explicit step by step procedure, one follows the following steps:

- One finds the IP address on the client VPN network in this session. This can be shown by hovering over the green VPN icon in the system bar on the bottom right corner of the screen. Typically, it would look like 172.17.0.30, and only the last number varies (in most cases it may be .28 or higher). Alternatively, one looks at the VPN connect log, or active network configuration)
- On the microscope side, using the Taro tray icon the client IP address (VPN) is then specified and has to be in the VPN range used by the microscope, which is, in the range 172.17.0.\*
- On the microscope side, one then switches to local (physically at NeCEN) or remote (physically at the remote location) the pads.

Taro is only deactivated once the VPN connection has been closed.

## 3. Security with the VPN

In any form of connectivity one of the first question that comes to mind is how secure the connection is. VPN offers a certain level of security which come with the VPN itself, the connection is done via an SSL/2 and it is an encrypted connection.

The main concern around this connection does not come from the connection itself, but instead concerns arise once the connection has been established.

Future setup would require managing the logged in user in such a way that it is not impacting their capabilities as a user but prevents intentional and/or accidental incidents such as, software deletion, manipulation of critical settings or data loss just to name a few. In short, the user over the VPN would have limited administrative rights or an administrator password protection on certain operations to limit any very damaging advert effects due to mishandling and/or accidental deletion.

## 4. Database and pre-processing

#### 4.1 Database and pre-pro

We are re-implementing our pre-processing pipeline and complementing it with an in-depth database where information is then stored which allows users to retrieve data and information about the data collection.

The database is designed to capture all the relevant information, ranging from user experimental details to data storage details, as well as storing pre- and post-processing details of the data collection. The general architecture of the database is shown in Figure 3.

SentDatSetzed • X DensTable E sentD onnene	🗿 ProjectsNameTable 🛞				
sersiD	ProjectsNameTable				
esD	ProjectsNameTable				
		ScopeNameTable	SPProjectsTable		
thlame all differs to be all d	Anayoto     Anayoto     Verso     Projectn/emeD     Projectn/emeD     Copeter     ComplementIn     DetAinStemp     DetAinStemp     DetAinStemp     DetAinStemp     DetAinStemp	Popjechlumker     Popjechlumker     Sophiand     Sop		MotionCorrDetailsTable MotionCorrDetailsTable SMotionCorrDetailsTXX MotionCorrDetailsTXX	9
taPostPreciD sensTableTableAdapter (8)	Anabice Software			MotionConDetailuD Coordinate_X Coordinate_V	
GetData ()	Sil ProjectsNameTableTableAdapter (3)	TomoProjectsTable	DataDetailsTable     S     DataDetailsINDK	IndividualDist AverageShift	
		TempProjectil0 ScapeName0 DataDeale0 Variat0 DataPetProc0 DataPetProc0 DataPetProc0 ScaPetProc0 \$2 FillCentData 0	DataOtaili0 DataFisiti0 MoisinCer(DataII0 CT(DestaII0 ProjectNumber Scopellume GridEquate GridEquate DisEguateAuthor PhetEise CulteringFored	MotionCorDetabl TableTableAdayter     St. MotionCorDetabl TableTableAdayter     St. MotionCorDetableTable     CTDetableCoX     CTDetableCoX     CTDetableCoX     CTDetableCoX     Defecutivitiens, X     Defecutivitiens, X     Defecutivitiens, X	PataPestPrec(NDX     DataPestPrec(NDX     DataPestPrec(D     DetaDetaND     PathSota     PathSota     PathSota     CassD     CassD     RefineD
			DoseRate DoseRate DataDetailsTableTableAdapter Science Fall,GetData ()	infoLimitCTF \$3 CTFDetailsTableTableAdapter (R) \$5 FillGetDeta ()	幻 DataPostProcTableTableAdapter 唐 名: Fil(GetData i)
				WarpTable     WarpToble     WarpToDX     WarpDD     PethToWarpDotput	
				記 WarpTableTableAdapter 原 処 Fill,GetData ()	

Figure 3. General design of live-processing database

The database is relational and captures all the details of the experiment. In addition to this backbone database, we have implemented an application interface which lets us manage and query this database. The first version of this application is shown in Figure 4.

								008										
4 0	are   F H   + ×																	
orra D.	121		UseraID	Festname	LastName	EnalAddress	Date Scope Start	Date ScopeEnd	E C									
mane:	Marc	1	21	Marc	antoL	m@gmail.com	03/01/2020 10:00	03/02/2020 10:00	8									
st Name:	Johns		22	Alana	Smth	a@hotnal.com	03/03/2020 10:00	03/04/2020 10:00	Б									
al Address:	ni@gmail.com		123	John	Black	j@five.com	03/03/2020 10:00	03/04/2020 10:00	6									
	Contraction of the local division of the loc		124	Martie	Bend	m@yahoo.com	03/05/2020 10:00	03/05/2020 17:00	Pr .									
e Scope Stat:	Sunday , March ~		25	Kate	Bush	k@bushi.com	03/06/2020 10:00	03/07/2020 10:00	Po I									
e Scope End:	Monday , March ~		126	Robert	Muller	réfisiology.com	03/07/2020 10:00	63/09/2020 10:00	M.									
r Group.	181,		27	Atane	Johns	a@bl.com		05/03/2020 10:00										
lation:	Leiden University		128	Tire.	Conscus	t@yahoo.com		05/05/2020 17:00										
Hator:	Fred		29	Ter	Corricus	t@yahoo.com		05/05/2020 12:00										
ernia ID:	121001	1	130	Gym	Faher	of@gnal.com	08/25/2020 10:00	09/25/2020	B.									
Pre Proc ID	121001	· .	-		-													
Post Proc ID																		
Post Proc ID	the second s	C	-					Projects Name	Form - NeCENDatabase	e								
	insert Entry in Database	Refe	h					14 4 0	d0 > H +	×								
Usen D	Contraction of the local data in the	Seat	Let Nat			rch User Group: (01.121001	8.	I 4 4 0 Analyse ID: Users ID: Projects Name ID Project Number:	001 121	×	AnalysisID 601: 002 003	UseniD 121 121 122	ProjectaNamelD 20200301 20200301 20200303 20200303	Project/Number 2001-278 2001-279 2001-200 2001-201	ScopeNameiD krice1 krice2 krice1	Operator Fred Wen Ludo	Complementinfo Sample 1 Sample 2 Sample 4 Sample 3	03/01/2020 03/02/2020 03/03/2020
	121	Seat	Let Nat				8.	Aralysis D: Users ID: Project Name ID: Project Namber: Scope Name ID: Operator: Date Min Stamp Date Min Stamp Date Min Stamp Date Scar: Ana Scar:	001 121 20200001 2001 278 knet 1 Sander , March 5under , March 10 2 10.1	•	601 602	121 121	20200301 20200301	2001-278 2001-279	krice1 krice2	Fred Wen	Sample 1 Sample 2	03/01/2020
	121	Seat	Let Nat				81	Analysis ID Liters ID Projects Name ID Project Number Scope Name ID Operator Complement Idro Date Mim Stamp Date Size Ana Size Schwaw	001 121 20200001 2001/278 Here 1 Find Sander , March Sunder , March 10-2 10.1 EPU	•	001: 002 003 004 005 006	121 121 122 122 123 124	20200301 20200301 20200303 20200304 20200304 20200303 20200305	2001-278 2001-279 2001-280 2001-281 2001-282 2001-283	kinos1 kinos2 kinos1 kinos1 kinos2 kinos1	Fred Wen Ludo Wen Fred Ludo	Sample 1 Sample 2 Sample 4 Sample 2 Sample 2 Sample 2 Sample 1	03/01/2020 03/02/2020 03/03/2020 03/03/2020 03/03/2020 03/05/2020
	121	Seat	Let Nat				8.	Aralysis D: Users ID: Project Name ID: Project Namber: Scope Name ID: Operator: Date Min Stamp Date Min Stamp Date Min Stamp Date Scar: Ana Scar:	001 121 20200001 2001 278 knet 1 Sander , March 5under , March 10 2 10.1	•	001: 002 003 004 005 006	121 121 122 122 123 124	20200301 20200301 20200303 20200304 20200304 20200303 20200305	2001-278 2001-279 2001-280 2001-281 2001-282 2001-283	kinos1 kinos2 kinos1 kinos1 kinos2 kinos1	Fred Wen Ludo Wen Fred Ludo	Sample 1 Sample 2 Sample 4 Sample 2 Sample 2 Sample 2 Sample 1	03/01/2020 03/02/2020 03/03/2020 03/04/2020 03/03/2020 03/05/2020
	121	Seat	Let Nat				8.	Analysis ID Liters ID Projects Name ID Project Number Scope Name ID Operator Complement Idro Date Mim Stamp Date Size Ana Size Schwaw	001 121 20200001 2001/278 Here 1 Find Sander , March Sunder , March 10-2 10.1 EPU	~ ~	001: 002 003 004 005 006	121 121 122 122 123 124	20200301 20200301 20200303 20200304 20200304 20200303 20200305	2001-278 2001-279 2001-280 2001-281 2001-282 2001-283	kinos1 kinos2 kinos1 kinos1 kinos2 kinos1	Fred Wen Ludo Wen Fred Ludo	Sample 1 Sample 2 Sample 4 Sample 2 Sample 2 Sample 2 Sample 1	03/01/2020 03/02/2020 03/03/2020 03/03/2020 03/03/2020 03/03/2020 03/05/2020 03/05/2020
	121	Seat	Let Nat				8.	Avalysis ID Liters ID: Projects Name ID Projects Name ID Coperator: Complement Info Date Min Starop Date Mas Starop Date Star Ava Star Schware Proj Type	001 121 2020001 2081 278 Junes 1 Fred Sampler, March 10,2 10,1 EPU SP	<ul> <li>✓</li> <li>✓</li> <li>✓</li> <li>Ref</li> </ul>	601 002 004 005 005 005 007	121 121 122 122 122 123 124 125	20200301 20200301 20200303 20200303 20200304 20200305 20200305 20200306	2001-278 2001-279 2001-279 2001-200 2001-201 2001-201 2001-202 2001-203 2001-204	kres1 kres2 kres1 kres1 kres1 kres2 kres1 kres2	Fred Wen Ludo Wen Ludo Wen	Sangle 1 Sangle 2 Sangle 4 Sangle 2 Sangle 2 Sangle 2 Sangle 3	03/01/2020 03/02/2020 03/03/2020 03/03/2020 03/03/2020 03/03/2020 03/05/2020 03/05/2020
	121	Seat	Let Nat				9.	Avalyses ID Liters ID Projects Name ID Project Names ID Congeneers Hrit Deter Min Starep Date Min Starep Date Min Starep Date Stare Schware Pro Type	001 121 20200001 2000 278 krise1 Fred Sunday , March Sunday , March 10.2 10.1 EPU SP 121	<ul> <li>✓</li> <li>✓</li> <li>✓</li> <li>Re</li> <li>S</li> </ul>	601 002 003 004 005 006 006 007	121 121 122 122 122 123 124 125	20200301 20200301 20200303 20200303 20200304 20200306 20200306 20200306	2001-278 2001-275 2001-275 2001-281 2001-281 2001-282 2001-283 2001-284	kinos1 kinos2 kinos1 kinos1 kinos2 kinos1	Fred Wen Ludo Wen Ludo Wen	Sample 1 Sample 2 Sample 4 Sample 2 Sample 2 Sample 2 Sample 1	03-01-2020 03-02-2020 03-03-2020 03-03-2020 03-03-2020 03-03-2020 03-05-2020 03-05-2020
	121	Seat	Let Nat				<b>4</b> .	Avalyses ID Liters ID Projects Name ID Project Names ID Congeneers Hrit Deter Min Starep Date Min Starep Date Min Starep Date Stare Schware Pro Type	001 121 20200001 2000 278 krise1 Fred Sunday , March Sunday , March 10.2 10.1 EPU SP 121	<ul> <li>✓</li> <li>✓</li> <li>✓</li> <li>Re</li> <li>S</li> </ul>	601 002 003 004 005 006 006 007	121 121 122 122 122 123 124 125	20200301 20200301 20200303 20200303 20200304 20200306 20200306 20200306	2001-278 2001-275 2001-275 2001-281 2001-281 2001-282 2001-283 2001-284	kres1 kres2 kres1 kres1 kres1 kres2 kres1 kres2	Fred Wen Ludo Wen Ludo Wen	Sangle 1 Sangle 2 Sangle 4 Sangle 2 Sangle 2 Sangle 2 Sangle 3	03/01/2020 03/02/2020 03/03/2020 03/04/2020 03/03/2020 03/05/2020
	121	Seat	Let Nat				4	Avalyses ID Liters ID Projects Name ID Project Names ID Congeneers Hrit Deter Min Starep Date Min Starep Date Min Starep Date Stare Schware Pro Type	001 121 2020001 2081 278 Junes 1 Fred Sampler, March 10,2 10,1 EPU SP	<ul> <li>✓</li> <li>✓</li> <li>✓</li> <li>Re</li> <li>S</li> </ul>	601 002 003 004 005 006 006 007	121 121 122 122 122 123 124 125	20200301 20200301 20200303 20200303 20200304 20200306 20200306 20200306	2001-278 2001-275 2001-275 2001-281 2001-281 2001-282 2001-283 2001-284	kres1 kres2 kres1 kres1 kres1 kres2 kres1 kres2	Fred Wen Ludo Wen Ludo Wen	Sangle 1 Sangle 2 Sangle 4 Sangle 2 Sangle 2 Sangle 2 Sangle 3	03/01/2020 03/02/2020 03/03/2020 03/03/2020 03/03/2020 03/03/2020 03/05/2020 03/05/2020
	121	Seat	Let Nat					Avalyses ID Liters ID Projects Name ID Project Names ID Congeneers Hrit Deter Min Starep Date Min Starep Date Min Starep Date Stare Schware Pro Type	001 121 20200001 2000 278 krise1 Fred Sunday , March Sunday , March 10.2 10.1 EPU SP 121	<ul> <li>✓</li> <li>✓</li> <li>✓</li> <li>Re</li> <li>S</li> </ul>	601 002 003 004 005 006 006 007	121 121 122 122 122 123 124 125 124 125	20200301 20200301 20200303 20200303 20200304 20200306 20200306 20200306	2001-278 2001-275 2001-275 2001-281 2001-281 2001-282 2001-283 2001-284	kres1 kres2 kres1 kres1 kres1 kres2 kres1 kres2	Fred Wen Ludo Wen Ludo Wen	Sangle 1 Sangle 2 Sangle 4 Sangle 2 Sangle 2 Sangle 2 Sangle 3	03/01/2020 03/02/2020 03/03/2020 03/03/2020 03/03/2020 03/03/2020 03/05/2020 03/05/2020
	121	Seat	Let Nat				8	Avalyses ID Liters ID Projects Name ID Project Names ID Congeneers Hrit Deter Min Starep Date Min Starep Date Min Starep Date Stare Schware Pro Type	001 121 20200001 2000 278 krise1 Fred Sunday , March Sunday , March 10.2 10.1 EPU SP 121	<ul> <li>✓</li> <li>✓</li> <li>✓</li> <li>Re</li> <li>S</li> </ul>	601 002 003 004 005 006 006 007	121 121 122 122 122 123 124 125 124 125	20200301 20200301 20200303 20200303 20200304 20200306 20200306 20200306	2001-278 2001-275 2001-275 2001-281 2001-281 2001-282 2001-283 2001-284	kres1 kres2 kres1 kres1 kres1 kres2 kres1 kres2	Fred Wen Ludo Wen Ludo Wen	Sangle 1 Sangle 2 Sangle 4 Sangle 2 Sangle 2 Sangle 2 Sangle 3	03-01-2026 03-02-2020 03-03-2020 03-03-2020 03-04-2020 03-03-2020 03-05-2020

Figure 4. Database NeCEN

The pre-processing process complements and populates this database, and the application provides a user interface for viewing the results.

#### 4.2. Pre-processing and post-processing

The link from the data to the database is done via the computational at the time of both the pre- and post-processing. The general idea is to provide a general fully automated mechanism for an on-the-fly evaluation and provide the user different with levels of data delivery starting from raw data to data delivery, with or without pre- or post-processing or both respectively. One of the key aspects is that users can start or stop any processes along the way at any time to allow full flexibility. The general overview of the data flow and how it is connected to the database, which stores information about the processing is shown in Figure 5.

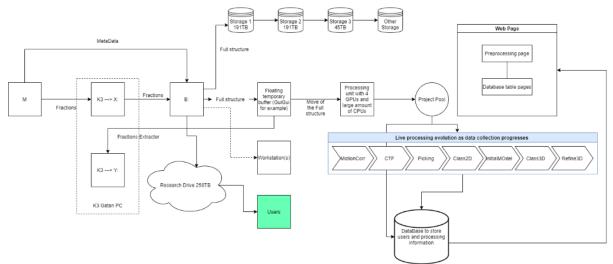


Figure 5. Overall data processing and data flow at NeCEN

The user is shown as a green box and shows how the data can be collected remotely using cloud service such as Research Drive (SurfSARA) for example.

In effect in order to accomplish this task we have designed and are currently implementing an application that allows the user to control data right from data acquisition. Although the user cannot change the data flow, the pre/post processing can be controlled by the user. In effect the data gets pushed onto the cloud service, called Research Drive (SurfSara). The user may start to download the data almost immediately as it is collected and may use the application to start the processing process, with the start and stop option, at any stage of the entire process.

The application is separated into two main components, the first one is the one that deals with the data acquisition, file transfer and handling, while the second being divided into two sub-components, one which deals with pre-processing and the other with post-processing, which starts from particle picking to initial modelling. In this instance one can think of this as real time processing as we are rendering results as we are collecting data.

The results can be summarized in a webpage like interface. This interface can then be expanded to allow data download directly from the database as if it was a portal.

In Figure 6 and 7 we show pictures of the two major components for the file transferring application and the pre-processing part

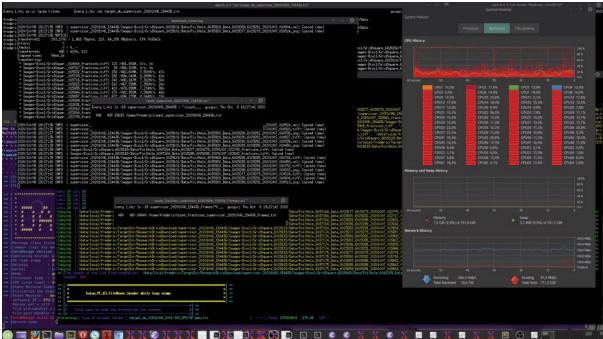
😑 💿 🕢 Gatan K3 (Fractions Handle	r) v1.0					
Project and Micrographs Input						Help
Inputs						
	t://home/frederic/Leiden/SourceCode/DataMana	age project				Set app path
Project Name (optiona						and off have
	y://data/local/frederic/TargetDir/ResearchDrivel	Download				Set Source Directory
Single Partic		C EPU-2.7	C SerialEM			
Tomograph		TOMOS	C SerialEM			
Movie Extensio	in: O mrc	<ul> <li>tiff</li> </ul>				
Fractions movement						
Warp Target Directo	y: //scratch/frederic/TargetDir/GatanScripts/War	pFolder				Set Warp Target Directory
How to handle file	es: 🗇 copy	move				
Multithreads copying (Default n	o): 🔿 yes	no no				
Gain Reference moveme	nt					
Gain Source Directo	y://data/local/frederic/GainRef_krios1					Set Gain Source Directory
Gain Target Directo	y: /scratch/frederic/TargetDir/GatanScripts/Gain	refFolder				Set Gain Target Directory
Gain Ref Typ	e: 🔿 x0.m0	<ul> <li>x0.m1</li> </ul>	x1.m0	x1.m1		
Download Research Driv	e					
cdi Nan	e: cdi-Frederic					
Research Drive Source Directo	y: krios1/20201006_2001-282_EPU-SP_Wen					
Download Target Directo	y: //data/local/frederic/TargetDir/ResearchDrive	Download				Set Target Directory
	Start Downloading From RD	Start Copy/Move Fractions	Start Copy Gain Reference	Stop ALL	Launch DataManager App	
Quit	Stop Downloading From RD	Stop Copy/Move Fractions	Stop Copy Gain Reference			

**Figure 6.** Transferring part of the application from remote location or servers to processing computing unit.

rographs and Gain Refs Input	Class 2D and 3D laun	ching Input					
Required Inputs							
Application Root:	/home/frederic/Leide	en/SourceCode/DataManage_project					Set app path
Project Name:	20201015_1910-272	_EPU-SP_Ludo					
Data Path:	/data/krios1buffer						Set Micrographs
Gain Refs :	/data/krios1buffer/g	ainrefs/2020-09-0809-43-44_K3-20050033_Gain_	Refx1.m1.mrc				Load Gain Reference
Target Directory:	/data/local/frederic/l	TargetDir					Set Target Directory
Single Particle:	© EPU-2.5	© EPU-2.6	<ul> <li>SerialEM</li> </ul>				
Tomography:	о томо4	томо5	<ul> <li>SerialEM</li> </ul>				
Movie Extension:	<ul> <li>mrc</li> </ul>	⊖ tiff					
Preprocessing Inputs							
Super resolution:	Factor of 1/2 from	XML reported value to compensate to true value					
MotionCorr Directory:	/opt/Devel_tools/Mo	tionCor2_v1.3.1/					Set Motioncorr path
Dose per Frame (e/A2):	1.0	Bfactor:	150 Binning facto	r. 1	Gain rotation: 2	Gain flip: 2	
CTFFind Directory:	/opt/Devel_tools/CT	Ffind/ctffind-4.1.18/bin_compat/				_	Set CTFFind path
Pre Processing details:	Live Process	MotionCorr	CTFfind				
File Pre-Handler:	Preprocess only h	ere at the moment					
Organise Files into Directories:	C Yes Organise	<ul> <li>Reorganise back into pool</li> </ul>	O Do Nothing				
Extract Data:	C From Files	<ul> <li>no</li> </ul>					
Computational Inputs							
BenchMarking:	Output the benchr	marking results to DataManages.log					
Use GPU:	CUDA C	DevID[Optional={0:1:2:3}} default=0]:	0:1:2:3	Create Project Pool:			

**Figure 7.** Pre-processing part of the application using data downloaded or transferred from remote location or servers to processing computing unit.

In Figure 8 we show a typical execution window and monitoring window of the process.



**Figure 8.** Console window for both the file transferring part and the Pre-processing part of the application using data downloaded or transferred from remote location or servers to processing computing unit.

Finally, in Figure 9 we show the post-processing application, where part of the preprocessing and all of the post-processing settings are set before the start of the application. This application computes on the fly for the different stages of the workflow.

😣 🗇 🕘 🛛 DataManage (Data Manag	ement from EM-Microscope) v1.0					
Micrographs and Gain Refs Input	Class 2D and 3D launching Input					Help
Particle Details						
Number of MPI Processors:	32					
Minimum diameter (Å):	160	Maximum diameter (Å):	250			
Particle Picking Inputs						
Input micrographs for autopick:	/data/local/frederic/TargetDir/2020	01015_1910-272_EPU-SP_Ludo_Pool/Relion	vapp_ctf_corrected_micrographs.star			CTF micrographs star file
Particle box size (pix):	400					
Class 2D Inputs						
Starting number of classes:	15	Starting number of iterations:	25			
Mask diameter:	200	Number of pooled particles:	3			
Number of MPI processors:	5	NUmber of threads:	4			
Class 3D Inputs						
Starting number of classes:	1	Starting number of iterations:	5			
Mask diameter:	200	Number of pooled particles:	3			
Mask diameter:	200	Symmetry:	C1			
Number of MPI processors:	5	Number of threads:	4			
Refine 3D Inputs						
Computational Inputs BenchMarking:	Output the benchmarking result	s to DataMagages log				
	CUDA C	DevID[Optional={0:1:2:3}} default=0]:	0:1:2:3			
	Start particle picking	Start RUN of Class2D	Start RUN of Class3D	Stop ALL	RUN Class 2D	
Quit	Stop particle picking	Stop RUN of Class2D	Stop RUN of Class3D			

**Figure 9.** The Post-processing part of the application using data downloaded or transferred from remote location or servers to processing computing unit.

The output from the computation of the applications shown in the above figure then gets pushed into the database mentioned earlier and displayed on the website.

# 5. Conclusions

- Remote control of cryo-electron microscopes has been implemented at NeCEN and tested with different sites
- Remote control can be restricted to a part of the microscope or full control can be provided. For full controls, the remote user needs the remote panels setup and associated software.
- Clear guidelines and restrictions need to be put in place.
- Future developments will provide access to collected data and preliminary processing results while the session is on-going.

The presented setup is a step toward automation. Microscopes are becoming easier to operate and some restricted controls can now be provided to a wider range of users than before. With the remote-control setup, it is envisioned a 'synchrotron' type of access can be supplied, where the facility personnel ensure the good working of the instrument, load the samples for the remote user and gives controls to the users than then setup their own experiment.

This setup has proven crucial during the COVID-19 pandemic. With lockdown and travel restrictions in place, remote access has become the norm and thank to the remote setup presented here, we continued to work without interruptions.

# 6. References

- Electron Microscopy Data Bank: https://www.ebi.ac.uk/pdbe/emdb/statistics\_pub.html/
- Fernandez-Leiro R., & Scheres S. H. Unravelling biological macromolecules with cryoelectron microscopy. *Nature*, 537, 339–346 (2016).
- Glaeser R. How good can cryo-EM become? Nat Methods 13, 28-32 (2016).
- Gruene T. *et al.*, & Pantelic R. Rapid Structure Determination of Microcrystalline Molecular Compounds Using Electron Diffraction. *Angew.Chem. Int.Ed.* 57,16313 – 16317 (2018)
- Jones C. G., Martynowycz M. W. *et al.* & Gonen T. The CryoEM Method MicroED as a Powerful Tool for Small Molecule Structure Determination. *ACS Cent Sci.* 2018 Nov 28; 4(11): 1587–1592.
- Nakane T., Kotecha A., Sente A. *et al.* Single-particle cryo-EM at atomic resolution. *Nature* (2020)
- Peplow M. Cryo-electron microscopy makes waves in pharma labs. *Nat Rev Drug Discov* 16,815–817 (2017).
- Subramaniam S. The cryo-EM revolution: fueling the next phase. IUCrJ 6, 1-2 (2019).