

# Remote on-the-fly analysis and cloudification of the cryo-electron microscopy data

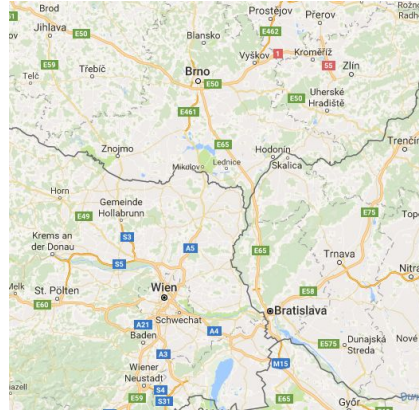
Jiri Novacek

*CEITEC, Masaryk University*

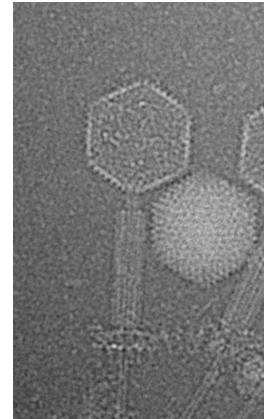
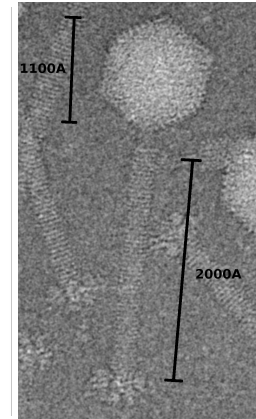
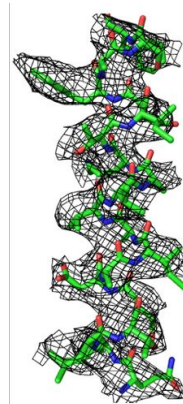
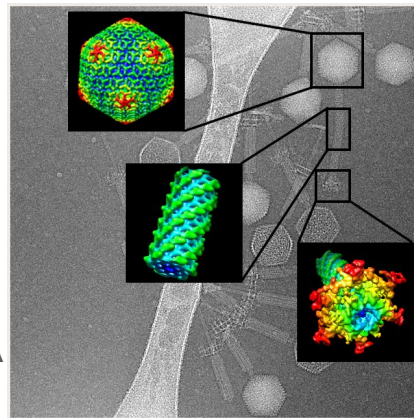


# CEMCOF

- 6 partners
- 500 researchers
- 7 research programs
- 63 research groups
- 25,000 m<sup>2</sup> of lab area
- 13 core facilities



- **Staphylococcus Aureus**
- **Bacteriophage phi812**
- **Active against 75% MRSA and 95% MSSA**

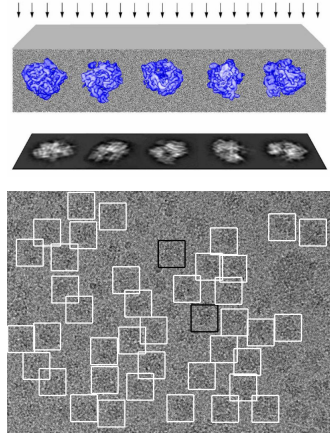


# Goals

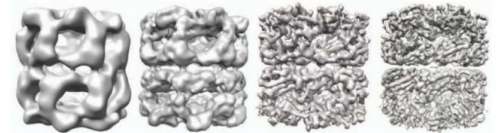
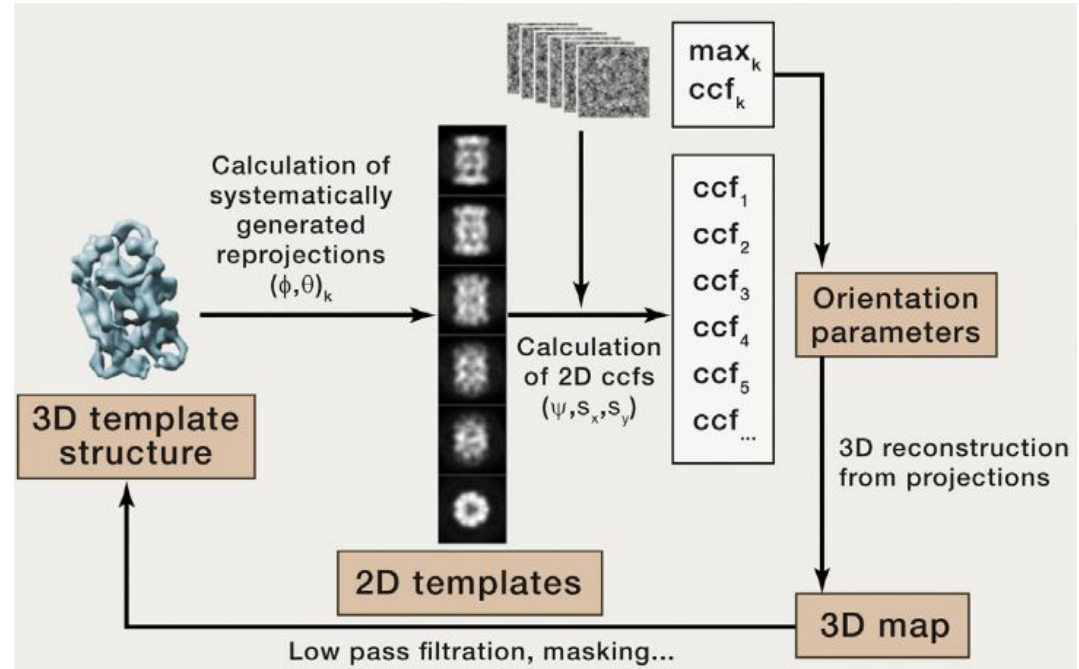
- **Facilitate management of the raw cryo-electron microscopy data**
  - **Data annotation directly on the instrument**
  - **Deployment of the data to the federated cloud solution**
  - **Automated archival and publication of the raw data**
- **Real-time data analysis on the remote HPC resources**

# Motivation

## Single particle cryo-EM workflow description

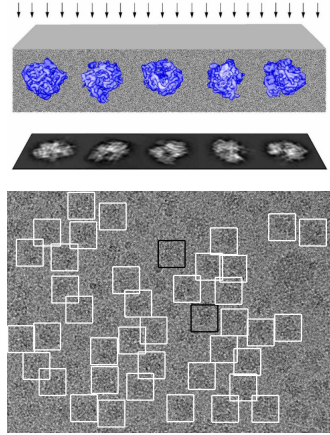


- micrograph (image), particle (ROI)
  - 2D projection of a 3D object -> projection orientations not known
  - very low signal to noise -> large number of images
  - sample drift and beam induced motion during imaging
  - convolution with microscope point spread function
- > data intensive



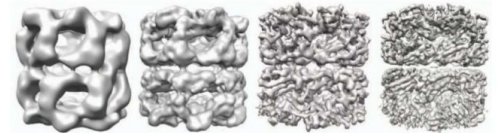
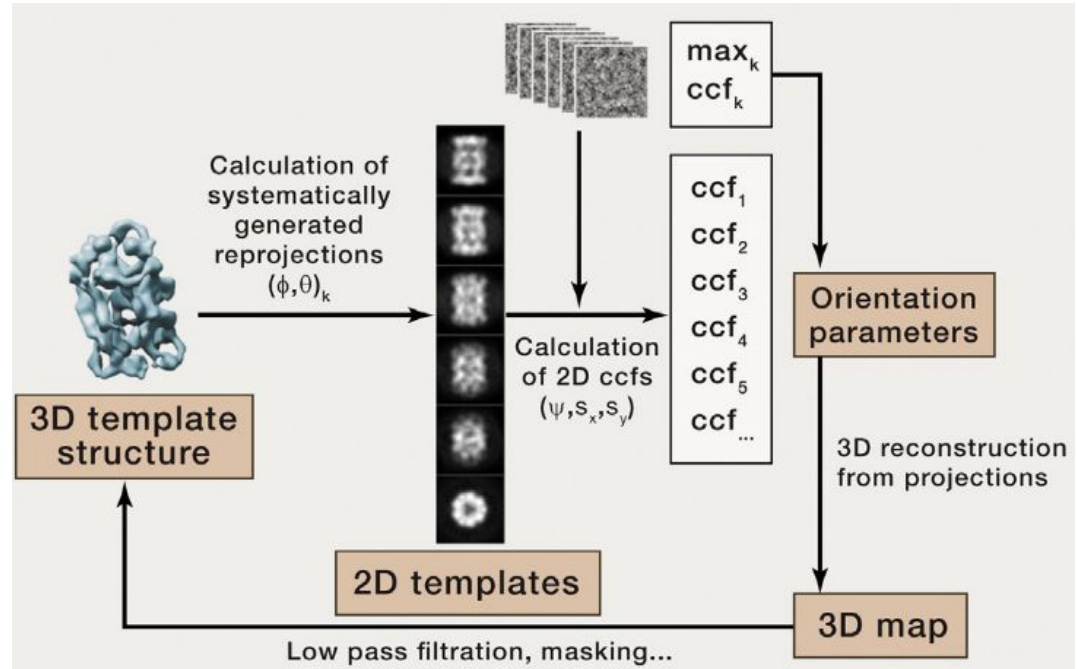
# Motivation

## Single particle cryo-EM workflow description



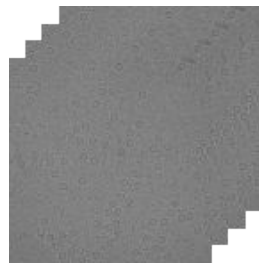
- compare experimental images to all reference projections
- assign the projection angles of the best matching reference to the particle
- calculate 3D model
- project 3D model and use it as a new reference
- iterate

-> computationally intensive



# Motivation

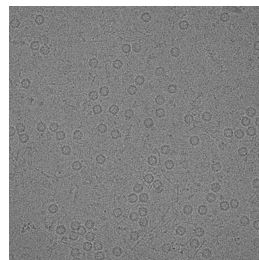
2-4 TB



raw data



~400GB



preprocessed data



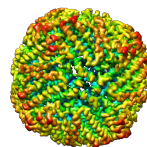
~40GB



ROIs/particles



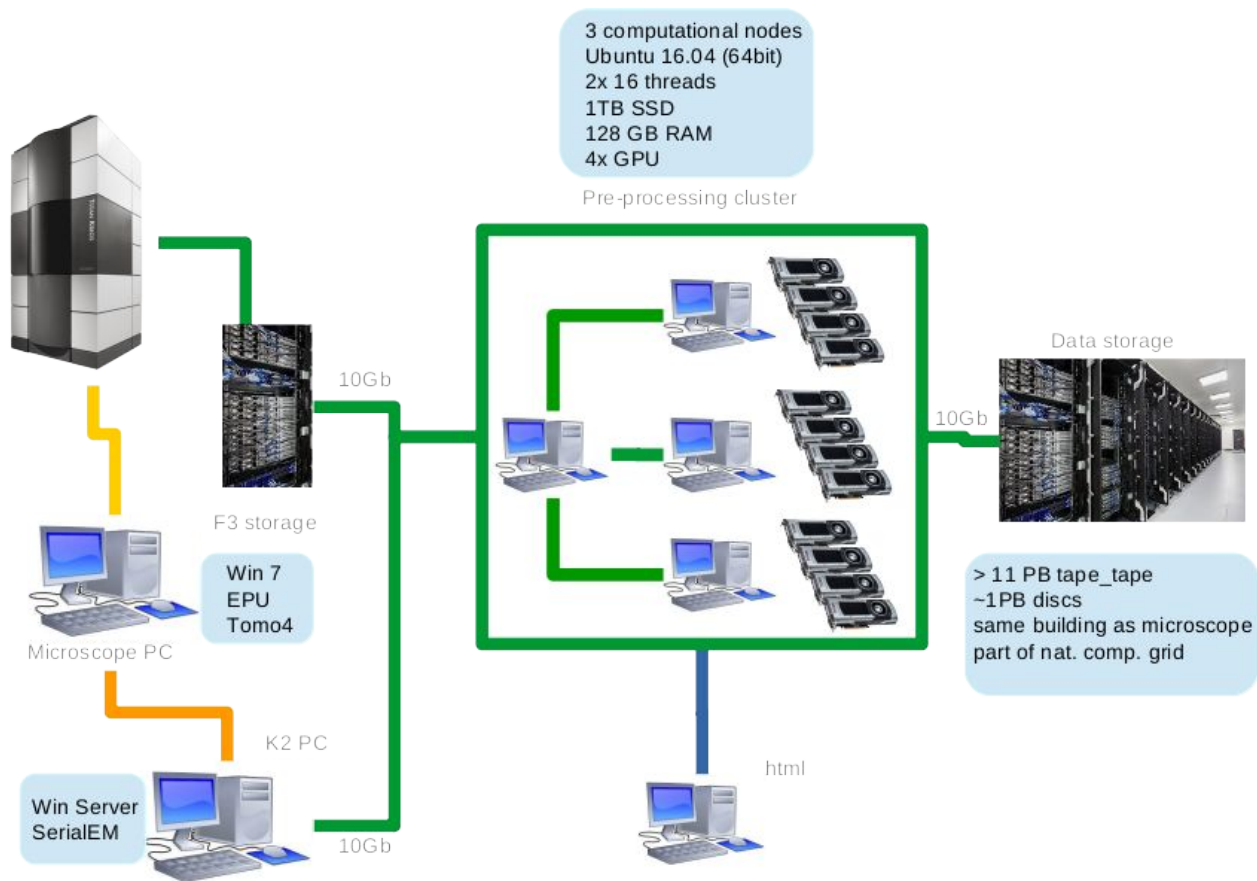
~40MB



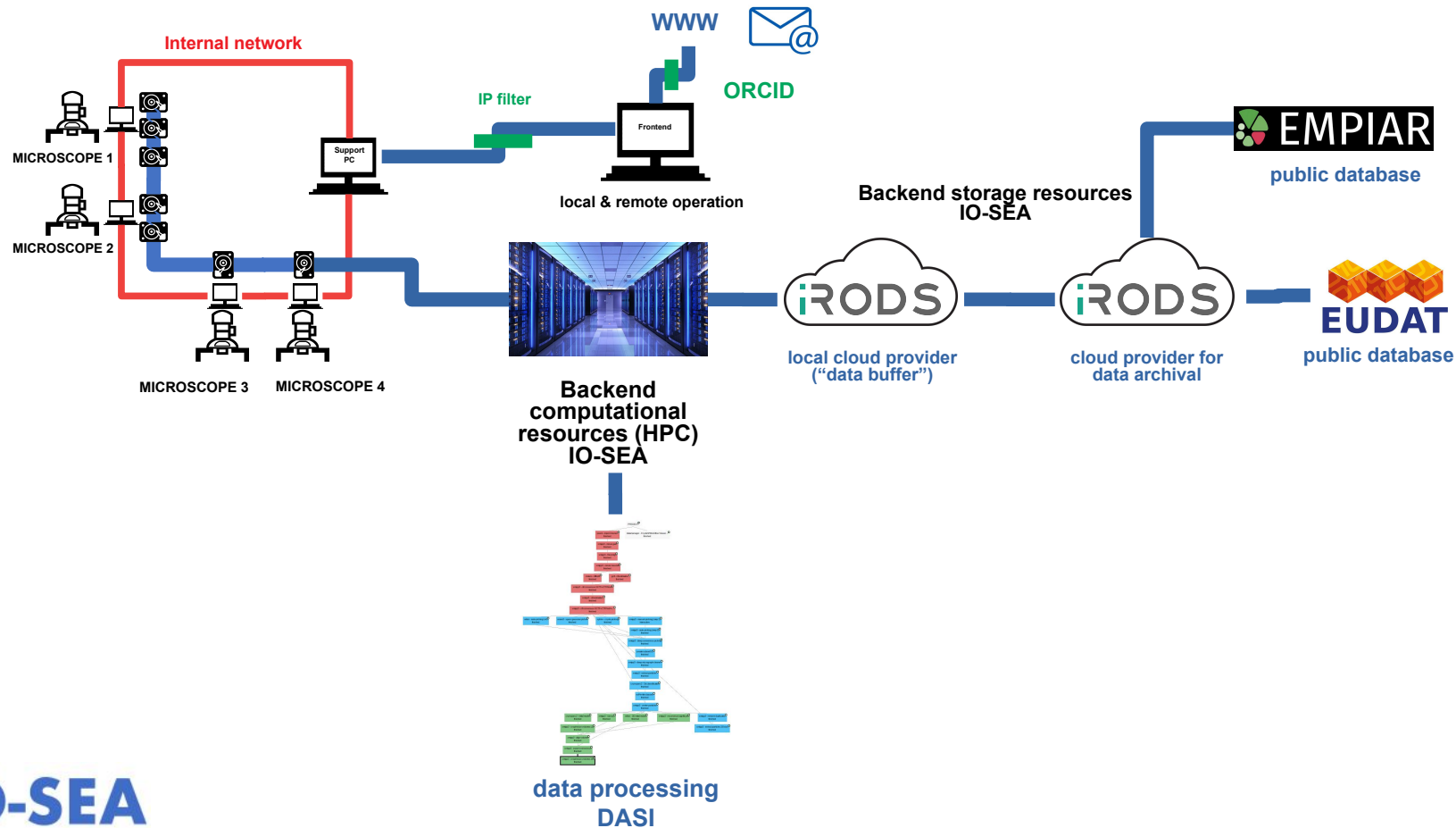
Final 3D structure



# Motivation



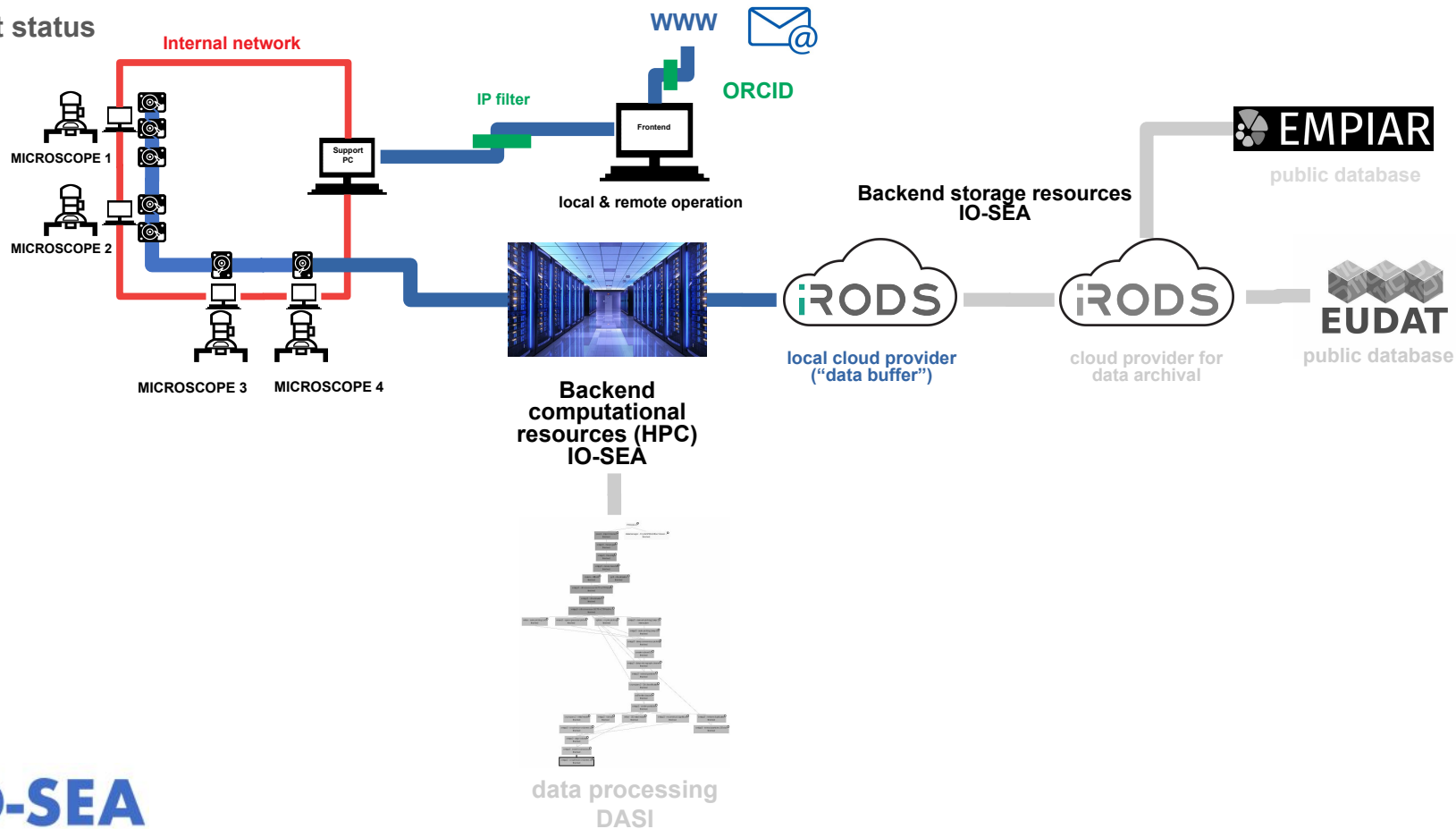
# Proposed solution





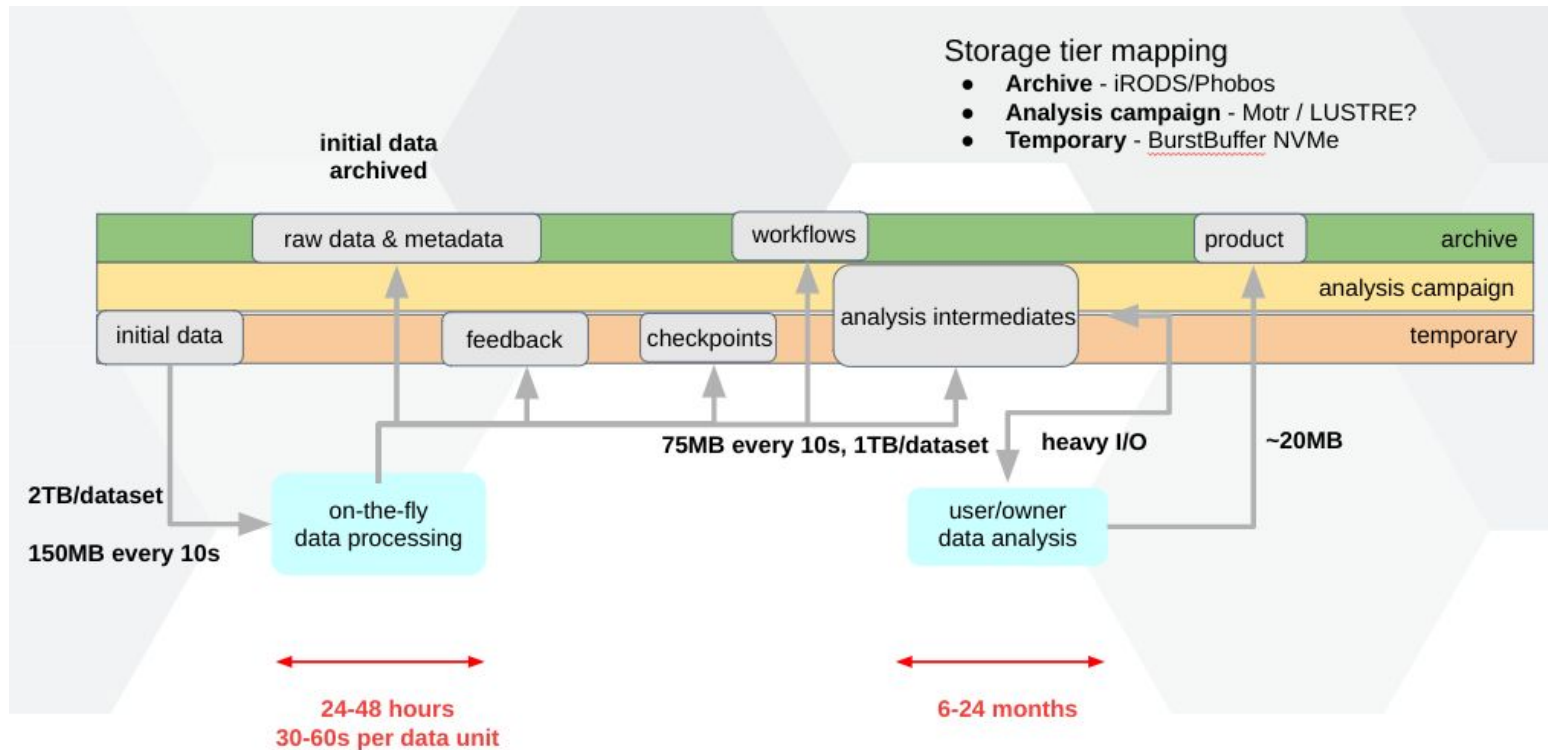
# Proposed solution

## Current status



# Proposed solution

## Data cycle



# Proposed solution

## Front end

- web-browser based operation
- accessible directly from the instrument/microscope
- secure access from outside the microscope network:
  - Google
  - eduID
  - ORCID
- instrument specific setup
- Implementation of different data analysis workflows
- data-model with four different components:
  - the laboratory (center)
  - the user (data owner)
  - the sample
  - the data
- real-time job status
- automated user notification (email)
- local database of the projects

Remote access  
(admin)

> TalosArctica

> TitanKrios

▼ TalosF200C

Transfer

SPA

Tomo

> Versa3D

**Operator** - contact information of the person setting up the experiment

 Daniel Pinkas <daniel.pinkas@ceitec.muni.cz>  
CEITEC, Masaryk University, Kamenice 5, 62500 Brno, Czech Republic

**User** - contact information of the data owner

[Same as operator](#)

 Alžbeta Dikunová <alzbeta.dikunova@ceitec.muni.cz>  
CEITEC, Masaryk University

**User Type** ☒ Internal

☐ External

☐ Private

☐ Eyen

**Access Route** ☒ Ciisb

☐ Instruct

☐ Inext discovery

☐ Direct access

Project

**Sample**

Rai1+Rat1+Rtt103

Job

**Data  
path**

☐ Remove all data from the microscope after job termination

☒ Archive data

Run job

**DEMO**

# Summary

- automated annotation of the raw cryo-EM data
- real-time data transfer to the remote storage resources (iRODS)
- on-the-fly data analysis on the remote HPC resources
- database entry for data publication (FAIR)

# Acknowledgements

- Radek Veverka
- Radek Furmanek
- Martin Golasowski
- Jan Martinovic

