

PReSTO in Support of Swedish Integrated Structural Biology

Alexandra Ahlner¹, Martin Moche², Oskar Aurelius³, Stefan Fleischmann⁴, Ana Gonzalez³, Michael Hall⁵, Anton Höglund⁶, Karl Hörnell⁶, Daniel Larsson⁷, Zdenek Matej³, Jose Miguel de la Rosa Trevin^{***}, Anders Sjöström³, Sebastian Thorarensen, Emil Tykesson³, Torben Rasmussen⁶ and Maria Sunnerhagen¹

¹Linköping University, ²Karolinska Institutet, ³Lund University, ⁴Stockholm University, ⁵Umeå University, ⁶National Supercomputer Centre, ⁷Uppsala University

PReSTO

PReSTO is a software stack for integrated structural biology adapted to high performance computing resources at Swedish National Infrastructure for Computing (SNIC) and the local MAX IV compute cluster. Our aim is to support integrative structural biologists evaluating their data from Macromolecular X-ray crystallography (MX), X-ray Free Electron Lasers (XFEL), Nuclear Magnetic Resonance (NMR), cryo Electron Microscopy (cryo-EM), Neutron scattering (NMN) and Small-Angle X-ray Scattering (SAXS) to determine 3D structures of macromolecules such as proteins, DNA and RNA.

Integrated Structural Biology

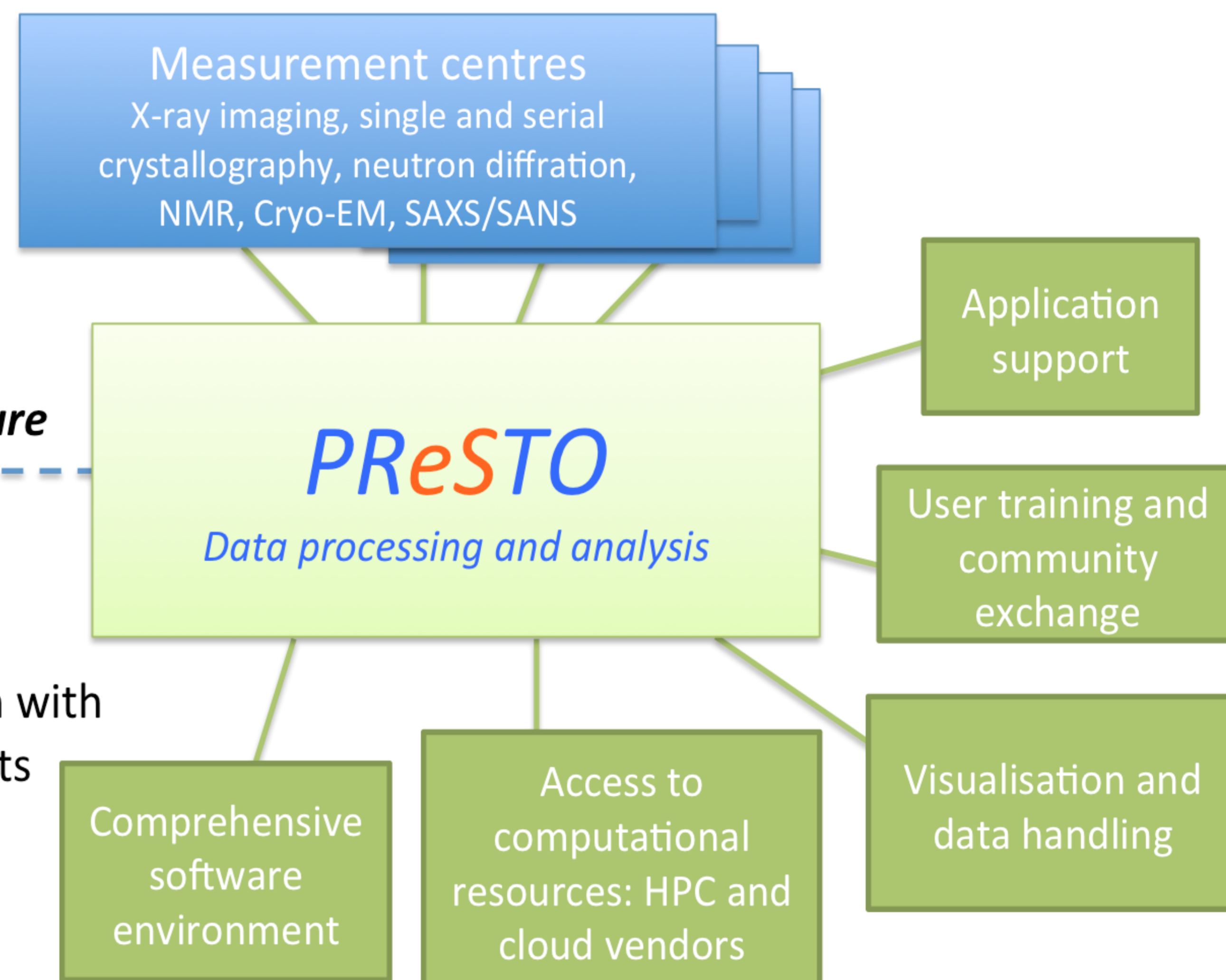
The multi-purpose and multi-technique approach to characterize how macromolecules and their assemblies interact in space and time is known as integrated structural biology. Swedish SciLifeLab recently launched a new Integrated Structural Biology Platform (ISB) platform and PReSTO will connect data generating structural biology infrastructures with the large-scale Swedish National Infrastructure of Computing (SNIC). Structural biology is a subset of the large Swedish effort known as Data Driven Life Science (DDL) that will benefit protein scientist when addressing complex structural biology questions involving machine learning and artificial intelligence when stretching towards cellular length scales and dynamic events of life via synchrotron X-ray tomography, correlative microscopy and molecular dynamics simulations.

Data collection, initial processing and temporary storage

Experimental infrastructure

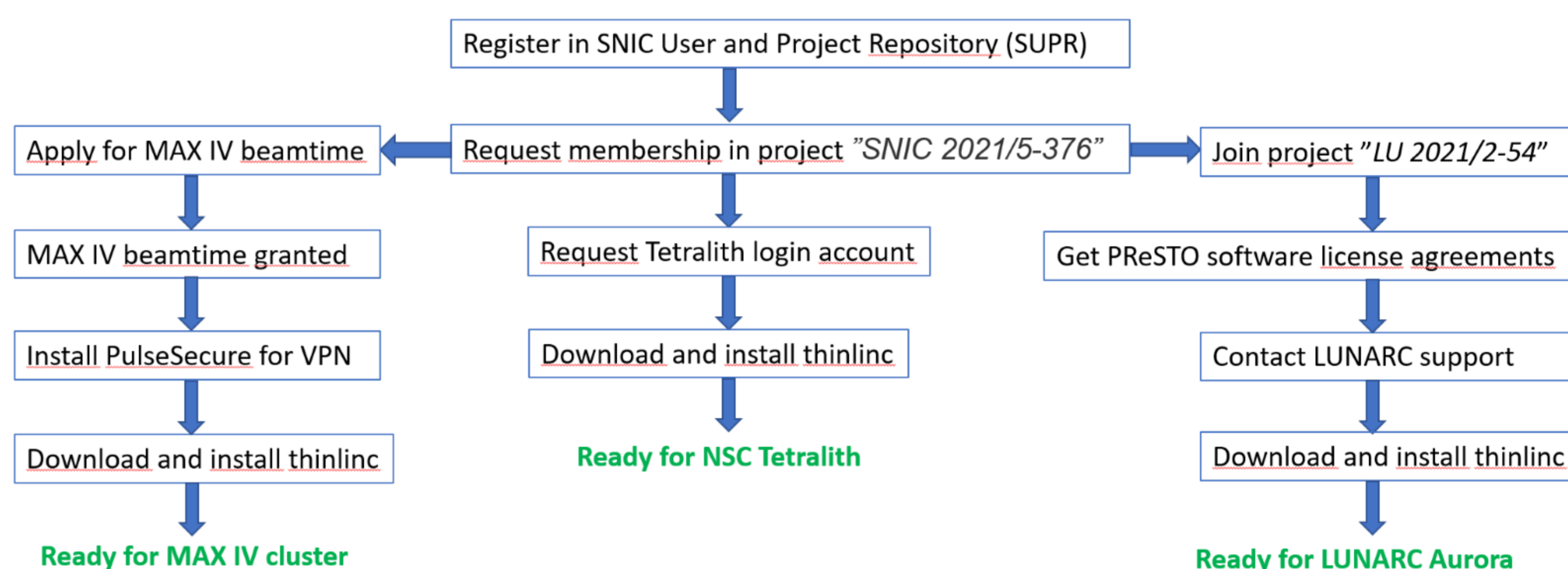
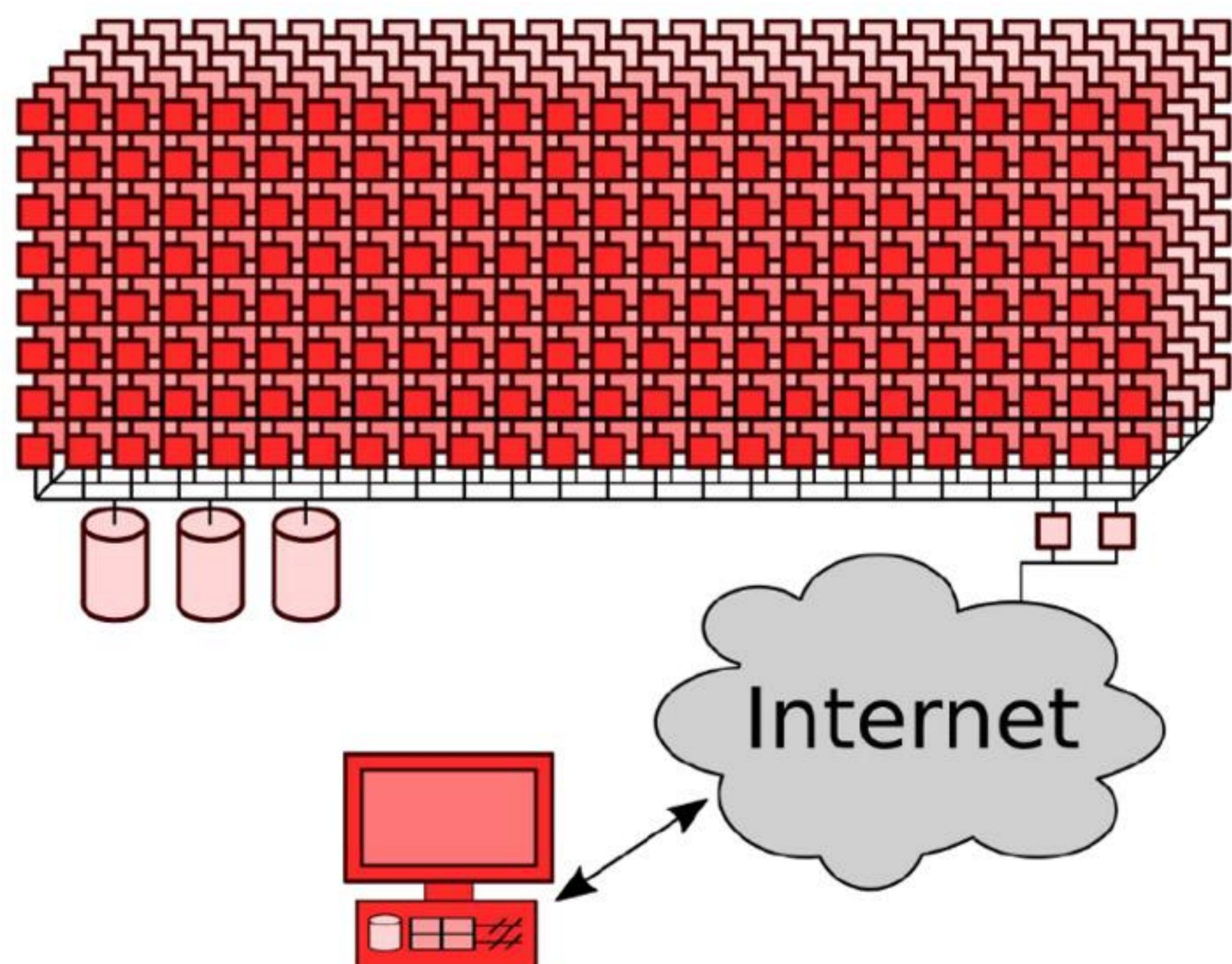
Home lab

Further data processing, refinement, co-evaluation with other softwares and results from other methods



Accessing PReSTO

Access to the PReSTO installation is via Swedish National Infrastructure of Computing (SNIC) funded by the Swedish Research Council. Thinlinc is a remote desktop server that supports the integrated structural biology workflow by enabling graphic applications including coot/chimera/ccp4mg/pymol to run smoothly from a remote computer. The homepage (<https://www.nsc.liu.se/support/presto/>) is written for structural biologists new to HPC environments where we discuss how to; manage software modules, login and compute nodes, edit sbatch scripts, schedule and monitor jobs etc. We also describe how to configure PHENIX for slurm scheduling and guide readers towards software developer manuals.



NSC	Nodes	CPU type	CPUs	RAM	GPU	/scratch/local
Tetralith	170	2x Intel Xeon Gold 6130	32	96 GiB	1 NVIDIA® T4	2 TB NVMe SSD
Berzelius	60	2 AMD Epyc™ 7742	128	1 TB	8 NVIDIA® A100 Tensor Core	15 TB NVMe SSD

Available software

MX: XDS, CCP4, PHENIX, GlobalPhasing, PyMOL, ChimeraX, USF

cryo-EM: Scipion, Relion, CistEM, EMAN2, Gctf, Gautomatch, Motion2Cor, CTFFIND, crYOLO, Xmipp

NMR: NMRPipe, MddNMR

XFEL: CrystFEL, Cheetah, nXDS, NanoPeakCell

Compute:

Help PReSTO!

1. Use PReSTO! Report BUGs to PReSTO team members!
2. Suggest software to be added into PReSTO!
3. Perform software updates! Free up time for PReSTO staff to deal with new installations and issues
4. Add new software to PReSTO