NBIS and SciLifeLab
- national Research Infrastructures in life science

Jessica Lindvall
Deputy Head of Node (ELIXIR-SE)
Head of Training @SciLifeLab

23-11-21

Definition - Research Infrastructure

"Research Infrastructures are **facilities, resources and services** that are **used by the research communities** to conduct research and foster innovation in their fields. They include: major scientific equipment (or sets of instruments), knowledge-based resources such as collections, archives and scientific data, e-infrastructures, such as data and computing systems and communication networks and any other tools that are essential to achieve excellence in research and innovation.*"

Research Infrastructures enables research!

https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/european-research-infrastructures_en

https://www.esfri.eu/

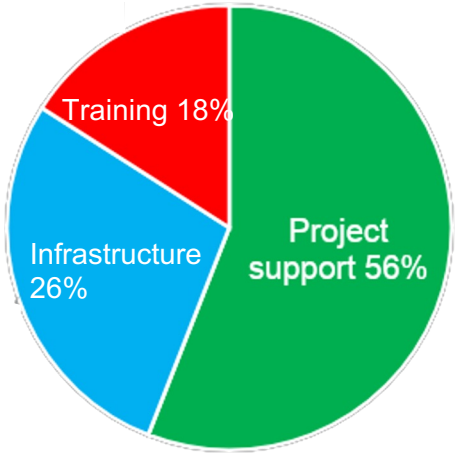
**Article 2 (6) of the Regulation (EU) No 1291/2013 of 11 December 2013: `Establishing Horizon 2020 - the Framework Programme for Research and Innovation (2014- 2020)`*

NBIS/SciLifeLab Bioinformatics Platform

Vision and Mission

Enable world-class life science research and maximise scientific and societal impact of collected data by:

- Providing **expert knowledge**, innovative data integration, advanced training, efficient data publication for open science, and access to high-performance data analysis methods
- Coordinating bioinformatics support within Sweden and **making bioinformatics easily accessible** for life science researchers
- **Swiftly responding to changes** in support needs as new techniques are developed and utilised
- Forming the **Swedish ELIXIR node** and participating in relevant international projects



Analysis of biological data

- **Support**
  - Genomics/NGS/Metagenomics
  - Genome annotation and assembly
  - MS-Proteomics and Protein bioinformatics
  - Systems biology and Metabolomics
  - Bioimage informatics and spatial omics
  - Integrative bioinformatics
  - Biostatistics
  - Data publication and FAIRification of data
  - Data management and Data stewardship (collaboration with SciLifeLab Data Centre) https://nbis.se/support
- **Infrastructure**
  - **Services, computational resources, data management, tools and guidelines**
  - https://nbis.se/infrastructure
- **Training**
  - https://nbis.se/training
- The Swedish node in ELIXIR – the European infrastructure for biological information

Distributed infrastructure with nodes at each of the 6 large university towns and in total **~120 staff**

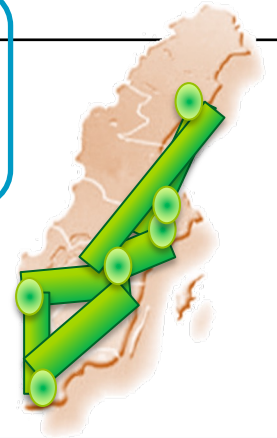
Marcus Lundberg @NBIS/Uppmax, marcus.lundberg@nbis.se

NBIS activities in short

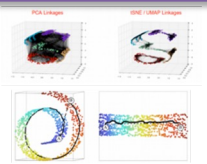
NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

SciLifeLab

Training Hub/DDLS

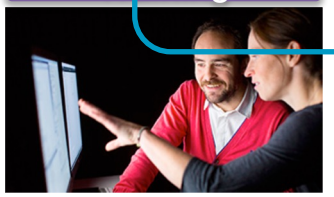


Support (incl. BIIF)



Weekly drop-in
100+ consultations
200+ support projects

Training



20 national courses
5 international courses
600 students
30 PhD mentors

Data producing platforms

Embedded staff
Dedicated contacts
Joint study design

Research project

Infrastructure/Data (incl. AIDA DH)

EUROPEAN GENOME-PHENOME ARCHIVE
B1MG
THE HUMAN PROTEIN ATLAS
Mr Bayes Bayesian Inference of Phylogeny
elixir SWEDEN

Development of systems
International collaborations
Maintenance of tools

Compute & Storage

700 research groups
1000 software
50 databases
5000 tickets

SNIC NAISS

UPPMAX

Berzelius

Data Management

DM plans
FAIR publishing
Sensitive data

Data Centre/DDLS

SciLifeLab - a national research infrastructure

- is a national institution for the advancement of molecular biosciences in Sweden
- is funded as a national research infrastructure by the Swedish government
- provides access to the cutting-edge instrumentation and deep scientific expertise necessary to be internationally competitive in bioscience research
- has a strong research community, including internationally recognized experts in life science and technology

Dimensions of SciLifeLab

Infrastructure

10 technology platforms, 40 units

>1400 users and 3000 projects per year

~ 500 technology experts

- Bioinformatics
- Genomics
- Clinical Genomics
- Clinical Proteomics and immunomonitoring
- Metabolomics
- Spatial biology
- Cellular and Molecular Imaging
- Integrated Structural Biology
- Chemical Biology and Genome Engineering
- Drug Discovery and Development

Training Hub

Training resources, services, support

Consolidates and coordinates training efforts across the SciLifeLab ecosystem, giving the life science community easy access to the SciLifeLab infrastructure knowledge, skills and expertise.

Research environment

Approx. 190 affiliated research groups

- Environment and climate change
- Farming and forestry
- Evolution and biodiversity
- Gene editing
- Biofuels and biomaterials
- Microbiology and microbiome
- Drugs and biomedicine
- Healthcare and aging

Data-driven life science

3.1 billion SEK, 12-year-program

Accelerating the data-driven life science paradigm shift

- Four strategic research areas
- Recruiting talent from across the globe
- Academic and industry PhD and postdoc programs
- Sparking collaborations, innovation and interdisciplinary team science
- Building a strong computational and data science base for open, real-time data

EU DIGITAL health data collaboration in Sweden

European Genomic Data Infrastructure

Bengt Persson
Uppsala University

EDIH Health Data Sweden

Sebastiaan Meijer
KTH

European Federation for Cancer Images

Katrine Riklund
Umeå University

Testing and Experimentation Facility for Health

Daniel Lundqvist
Karolinska Institutet

DIGIfor1healthSE (2024-2027)

Leveraging strategic capacity within:

- ELSI
- Technical infrastructure
- Semantic interoperability

Coordinated by SciLifeLab

Funded by VINNOVA
Sweden's Innovation Agency

NBIS

Compute and storage at SciLifeLab

- SciLifeLab Data Centre (Head of Data, Johan Rung) has a national coordinating role for SciLifeLab's Compute and Storage issues
- Close collaboration with NAISS and the local compute centers, in particular UPPMAX, NSC, C3SE, HPC2N, LUNARC and PDC
- Long term investment in staff and equipment, and integration in SciLifeLab DC national data platform
- Purpose is to maximise the access and useability of national e-infrastructure services for life science research
- Provides services hosting as a service on SciLifeLab Kubernetes based data platform

Johan Rung @scilifelab Data Centre, johan.rung@scilifelab.se

SciLifeLab Compute and storage

UU / UPPMAX

KTH / Campus Solna

LiU / NSC / Berzelius

Uppsala Multidisciplinary Center for Advanced Computational Science UPPMAX

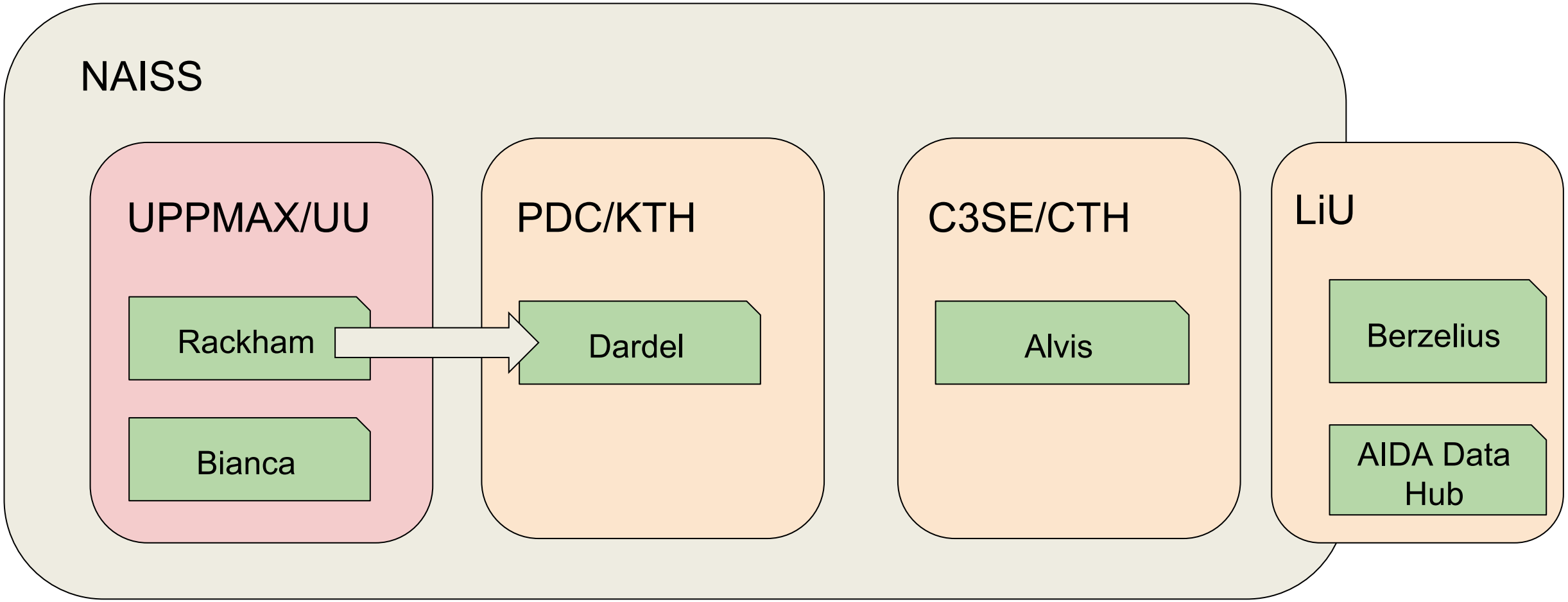
KTH VETENSKAP OCH KONST

- Bianca cluster
- Sensitive data, primarily genomics
- Storage
- High-mem CPU nodes and some GPU
- NBIS C&S Support team

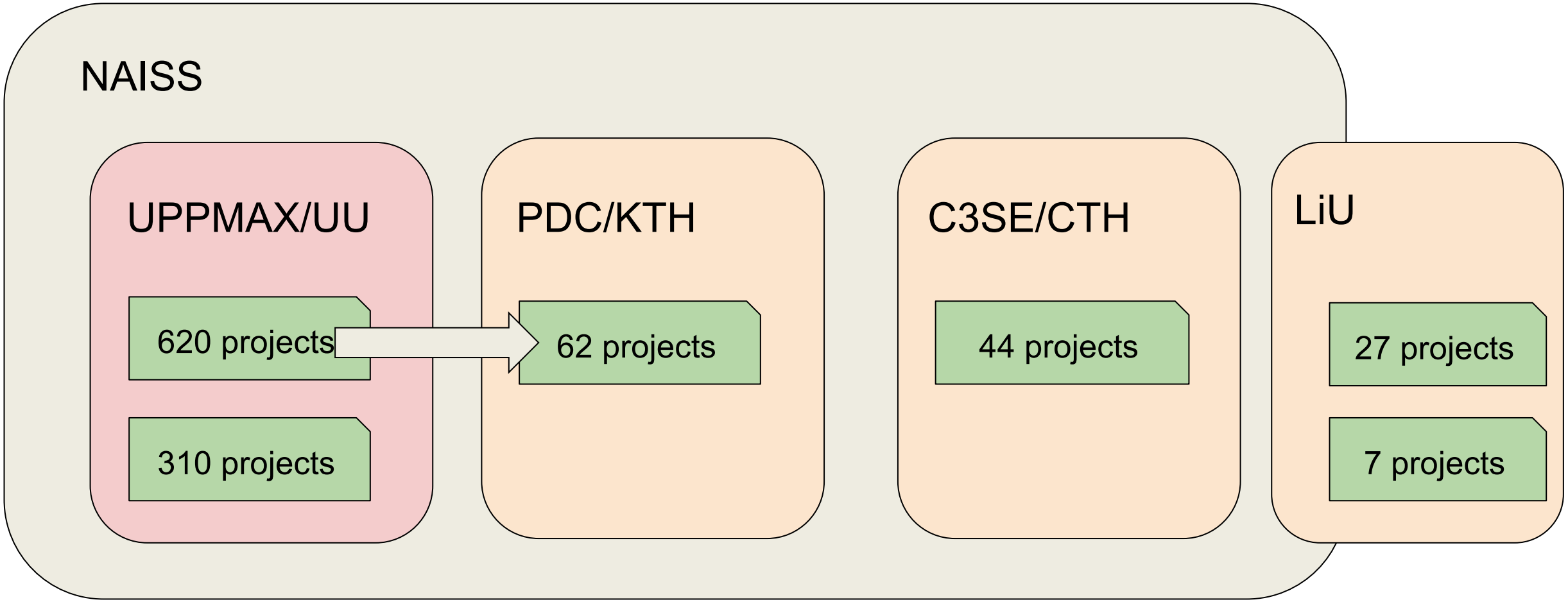
- High speed link to infrastructure
- Kubernetes service platform with CPU and GPU
- Storage
- Capacity for sensitive data building

- Berzelius AI resource
- 60 x DGX A100 nodes
- Kubernetes platform planned
- Storage
- Sensitive data resources planned

HPC Systems Used by Bioinformaticians



HPC Systems Used by Bioinformaticians



HPC environment

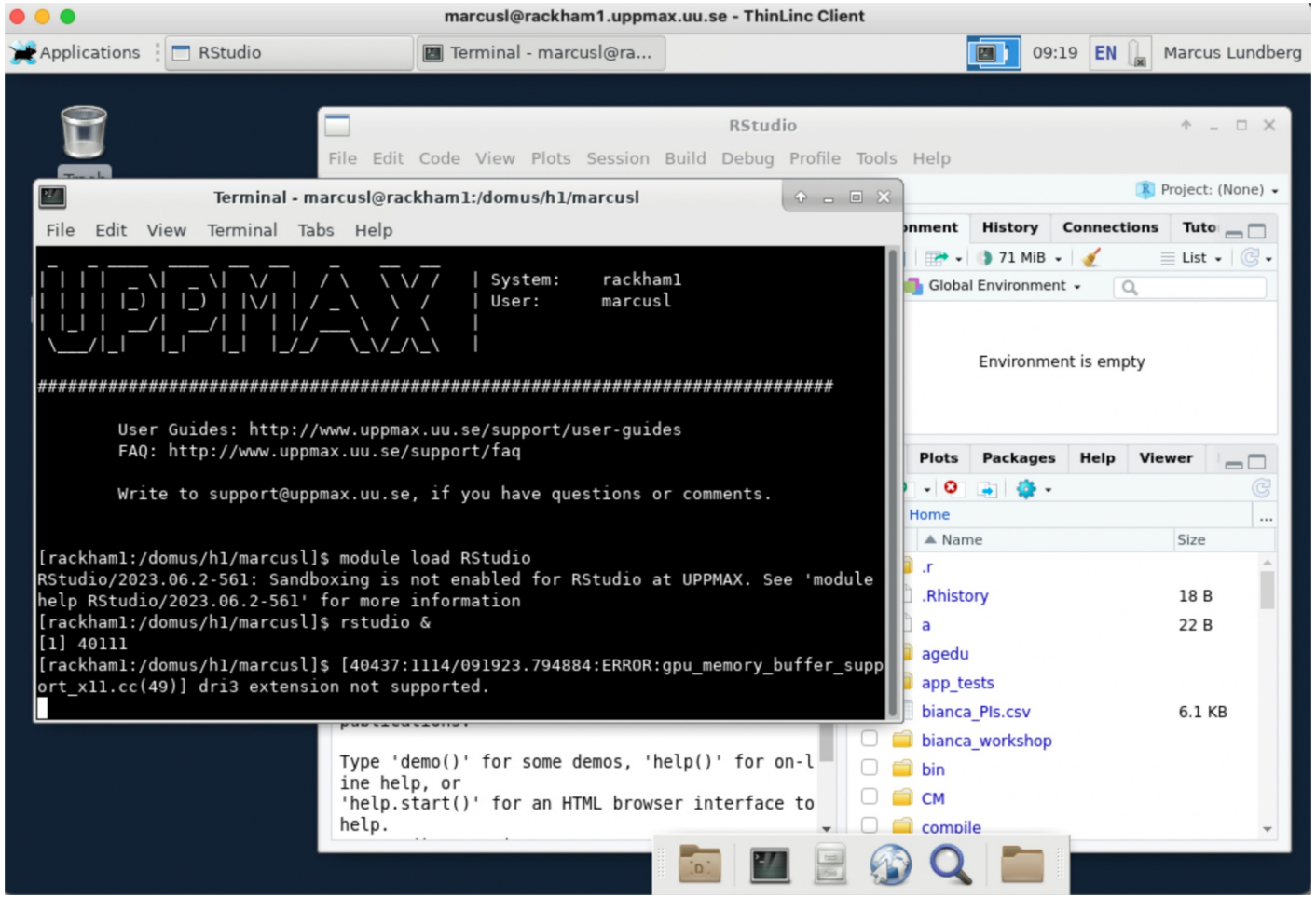
OS: Linux

Login: Desktop or command-line

Features:

- Free to use
- Large software library
- Slurm job queue

Environment developed by NBIS in collaboration with providers



Support and Training

- Current support landscape fractionated, NAISS is assembling a new structure
  - Currently: Email or support form to system support, each site handles issues within their local team
  - Future: National coordination over each discipline/tech area
- Training activities nationally coordinated and funded
  - Free to all users
  - Hundreds of participants per year
  - Introductory courses held regularly, but also more advanced workshops on particular topics

https://www.uppmax.uu.se/support/courses-and-workshops/

Impact

- 30-40% of SciLifeLab publications benefit from NAISS systems with NBIS-supported environment
- >50% of SNIC Nature publications from life sciences
- ~55% of current NAISS projects in life sciences

2023 - “Behovsanalys” sensitive data and life science


Analysis report (spring 2023) from SciLifeLab to SNIC/NAISS

- Compute and storage capacity of importance to life science research
- The need from life science researchers differ somewhat from general HPC needs from other disciplines
- Todays resources (towards the life sciences) need replacement (due to age)
- Support to users is critical (to secure efficient use of resources)
- services, tools etc needs more development

Marcus Lundberg @NBIS/Uppmax,marcus.lundberg@nbis.se
Johan Rung @scilifelab Data Centre, johan.rung@scilifelab.se

NBIS
NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

SciLifeLab

Thank you!