



Annual Report 2021

Introduction

NBIS (National Bioinformatics Infrastructure Sweden) is a distributed national research infrastructure, hosted by Uppsala University. NBIS constitutes the SciLifeLab Bioinformatics Platform. NBIS is formed by 11 partners: Uppsala universitet (hosting NBIS), Chalmers tekniska högskola AB, Göteborgs universitet, Karolinska Institutet, Kungl. Tekniska högskolan, Linköpings universitet, Lunds universitet, Naturhistoriska riksmuseet, Stockholms universitet, Sveriges lantbruksuniversitet, and Umeå universitet.

Funding is provided from the Swedish Research Council (VR), Knut and Alice Wallenberg Foundation, Science for Life Laboratory (SciLifeLab), Swedish universities, and user fees. In addition, we have funding from NordForsk (NeIC), ELIXIR, Chan Zuckerberg Initiative (CZI) and EU.

NBIS expanded 1 Jan 2021 with the SciLifeLab BioImage Informatics Facility (BIIF) and the AIDA Data Hub in Linköping joining us. These are funded outside the current VR grant with resources from SciLifeLab and other sources.

The number of supported users has increased and is now close to 300. This increase is partly dependent upon the new units BIIF and AIDA Data Hub and partly reflects a general increase in needs for bioinformatics support in life sciences.

Also 2021 was a year affected by the corona virus pandemic. Most drop-in sessions and training events were organised online but a few events could be arranged physically. As in 2020, much of our international work within ELIXIR has been run via digital meeting systems – even the ELIXIR All Hands Meeting was held online.

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

Support

One of the major activities is support, where our staff helps researchers with bioinformatics tasks in various projects. Currently, NBIS provides expertise in many areas within bioinformatics: genome assembly, genome annotation, genetic variation, comparative genomics, phylogenomics, transcriptomics, proteomics, metabolomics, systems biology, single-cell biology, biostatistics, and multi-omics integration.

In the current funding landscape, the universities need to contribute a substantial part of infrastructure funding, in proportion to the utilisation of NBIS by the respective university. This is reflected in the increased university contributions from 2018 and onwards. NBIS obtains user fees on direct project-related support, with the exception of the KAW-funded long-term support (LTS; cf. below). It is not feasible to apply user fees on the infrastructure, outreach and training activities, staff education and project management.

NBIS serves users at all major universities. In 2021, NBIS provided support to **297 PIs** (117 female, 180 male). The distribution of the supported PIs is shown in Figure 1. The top three universities using NBIS are Karolinska Institutet, Uppsala University and Lund University. The total number of active support projects were 276, and in addition to those NBIS has also provided 79 study design consultations. This is an increase of 30 PIs, 64 support projects and 25 study design consultations, compared to 2020.

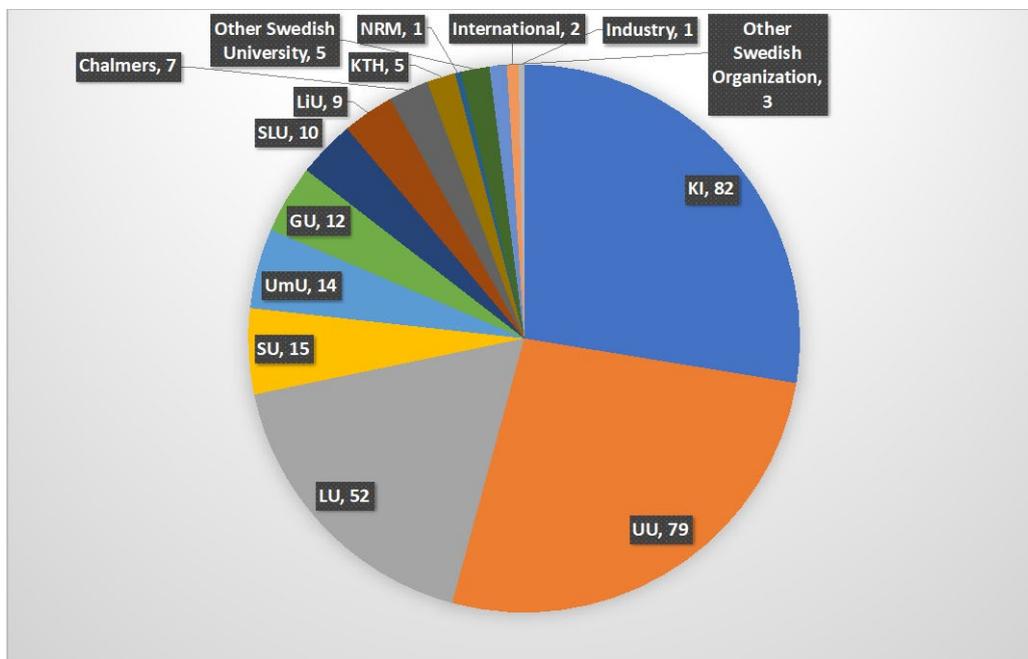


Figure 1 Distribution of PIs that have received support during 2021

In 2021, short- and medium-term support (SMS) was provided to 132 projects and long-term support (LTS) to 96 projects. In addition, NBIS was engaged in 9 partner projects, and the BioImage Informatics Facility that joined NBIS in 2021 provided support to 39 projects. A wide variety of projects were supported, with gene expression studies being the single biggest category.

About half of all support is provided from one site to another, emphasising the importance of a national infrastructure.

Analysed data were from a range of sequencing-based methods, and also other large-scale data, such as proteomics, metabolomics and cell imaging. Projects including single-cell RNA sequencing (scRNA-seq) and human whole genome sequencing data remain in high demand of support, while emerging areas include ancient DNA (aDNA) and cell image analysis.

Examples of important achievements from NBIS support projects

NBIS has a total of **93 publications** in 2021 from our direct project support work (excluding Compute & Storage and AIDA Data Hub), listed in the Annex. A few publications are shortly highlighted here.

The BioImage Informatics Facility (BIIF) and the AIDA Data Hub joined NBIS in 2021. BIIF is actively involved in the Network of European Bioimage Analysts (NEUBIAS), and became part of the EuroBioimaging in 2021, via its connection to the National Microscopy Infrastructure (nmisweden.se).

BIIF in collaboration with the Wählby lab (UU) and Östman lab (KI) developed an automated approach to classify multiplexed image data of cells into different cell types and analyse spatial relationships between cell types in brain tumour tissue slices (L. Solorzano et al., *Cytometry* 2021, doi:10.1002/cyto.a.24467). Another project addressed the need of annotated data sets in the context of deep learning. Creating manual annotations in a large number of images is a tedious bottleneck that often limits image analysis by deep learning. BIIF showed that the output from a classical image analysis pipeline for cell nuclei segmentation can be used when training a neural network, not only reducing the time spent annotating images but it may also lead to an improvement of results when compared to the output from a classical pipeline. (Matuszewski *et al.*, *WISCG*, 2021, doi: 10.24132/esm.2021.3002.37). BIIF was involved in exploring how a pre-trained neural network from a previous publication on grading of prostate cancer biopsies (Ström et al., *Lancet Oncology* 2020) can be applied as a feature extractor to prostatectomies, followed by clustering to explore tumour heterogeneity. In this project the approach was applied to tissue samples that had also been analysed by spatial transcriptomics in collaboration with the Lundeberg lab (KTH), and it was shown that morphologically distinct regions also showed differences in gene expression (Chelebian et al., *Cancers* 2021, doi:10.3390/cancers13194837). BIIF is closely involved with the international bioimage informatics community, and published three joint papers with the developers of software such as ImageJ and CellProfiler (Dobson et al., *Curr Protoc* 2021, doi:10.1002/cpz1.89; Martins et al., *F1000Res* 2021, doi:10.12688/f1000research.25485.1; and Paul-Gilloteaux et al. *F1000Res* 2021, doi: 10.12688/f1000research.52569.1).

NBIS has supported many studies with analysis of genome sequencing data. For example, in a project with Ulrika Norin and Rikard Holmdahl (KI), we helped establish Endophilin A2 (EA2) as a potential target for treatment of autoimmune diseases. The KI group had discovered a rat colony resistant to autoimmune arthritis, but the genetic cause of the resistance was unknown. By analysis of whole-genome sequencing data, we found the resistant rats to carry a structural mutation decreasing EA2 protein levels. Further experiments established a novel role for EA2 in T cell activation and autoimmunity (Norin, U. et al.: Endophilin A2 deficiency protects rodents from

autoimmune arthritis by modulating T cell activation. *Nat Commun* 12:610 (2021); <https://doi.org/10.1038/s41467-020-20586-2>).

We have seen an increasing demand for genome sequence analysis in studies of evolution and biodiversity. For example, in this area we supported a project investigating the genomes of *Sylvioidea* songbirds to understand sex chromosome evolution. These birds are interesting because a chromosome fusion event ~24 million years ago formed “neo-sex chromosomes” consisting of an added (new) and an ancestral (old) part. With researchers Bengt Hansson and Helena Westerdahl (LU), we have assembled and analysed the female genome of the great reed warbler (*A. arundinaceus*), showing that the new region has been added to both of the two bird sex chromosomes (Z and W), but with major differences in gene maintenance and degradation after the fusion event (Sigeman, H. et al.: Avian Neo-Sex Chromosomes Reveal Dynamics of Recombination Suppression and W Degeneration. *Mol Biol Evol* 38:12 (2021); <https://doi.org/10.1093/molbev/msab277>).

Many scientists approach NBIS for help with analysis of gene expression data, and we have been involved in several studies pushing the boundaries in this area, e.g. to investigate gene expression at the single-cell level in greater detail, and with parallel measurements of other molecular events. For example, we have assisted Caroline Gallant (UU) with developing a method to simultaneously measure global mRNA and a panel of intracellular proteins in individual cells. We showed that mRNA expression alone failed to accurately reflect protein abundance at the time of measurement and that protein levels of transcription factors better predict downstream effects than do their corresponding transcripts. The new method overcomes previous limitations in throughput and protein localisation, and does not require cell fixation (Reimegård, J. et al.: A combined approach for single-cell mRNA and intracellular protein expression analysis. *Commun Biol* 4:624 (2021); <https://doi.org/10.1038/s42003-021-02142-w>).

In recent years, we have seen an increasing number of projects using multiple types of omics data. In a project with Christofer Juhlin (KI), we helped to chart molecular aberrations in follicular thyroid carcinomas. The group applied whole-genome sequencing to study the genomes of the tumours, and other techniques to measure global mRNA and microRNA levels. Interestingly, the genome analyses revealed that two tumours contained mutations in a gene involved in microRNA processing (DGCR8). The same tumours displayed differences in microRNA levels, demonstrating the utility of multi-omics profiling and suggesting a role for DGCR8 and microRNAs in thyroid carcinomas (Paulsson et al., Whole-genome Sequencing of Follicular Thyroid Carcinomas Reveal Recurrent Mutations in MicroRNA Processing Subunit DGCR8. *J Clin Endocrinol Metab* 106:3265 (2021); <https://doi.org/10.1210/clinem/dgab471>).

In the systems biology area, the results of an NBIS Partner Project was published, presenting novel data on biomarkers for early diagnosis of Alzheimers’s disease (Wang et al.: Genome-scale metabolic network reconstruction of model animals as a platform for translational research. *Proc Natl Acad Sci USA* (2021); <https://doi.org/10.1073/pnas.2102344118>). The authors used functionality of Metabolic Atlas and involved researchers from Gothenburg University, Chalmers University of Technology and the Wallenberg Center for Molecular and Translational Medicine among others.

During 2021, NBIS has been engaged in multiple projects related to the COVID-19 pandemic. One such study investigated the hypothesis that higher prevalence of severe COVID-19 among men might be influenced by androgen hormones, since men have higher levels of such hormones.

This was investigated using multiple approaches, including a trial on COVID-19 patients and experiments in cell lines, but there was no evidence for a beneficial effect of anti-androgen treatment. Here, NBIS contributed biostatistical analyses to reach the important conclusion that anti-androgen treatment should not be used against COVID-19. This project involved research groups from Gothenburg, Umeå, Linköping and Lund universities (Welén et al.: A Phase 2 Trial of the Effect of Antiandrogen Therapy on COVID-19 Outcome: No Evidence of Benefit, Supported by Epidemiology and In Vitro Data. *Eur Urol* (2021); <https://doi.org/10.1016/j.eururo.2021.12.013>).

User fees

NBIS has five support tracks:

- Study design consultation ($\leq 3h$)
- Short- and Medium-term Support (SMS)
- Partner Projects (PP)
- Long-term Support (LTS)
- BioImage Informatics (BIIF)

For SMS, PP and BIIF, NBIS charges user fees, while LTS is provided for free according to the funding requirements by the Knut and Alice Wallenberg foundation. In recent years, the total income from user fees has increased and in 2021, the amount was 10.3 MSEK, indicating better awareness of costs for data management and advanced bioinformatics support. The BioImage Informatics Facility joined NBIS in 2021 and may later be integrated with the other tracks.

Infrastructure

A fundamental part of NBIS is the formation of a sustainable bioinformatics infrastructure for life sciences, consisting of access to tools and data. The infrastructure is typically constructed as domain-specific supporting layers utilising resources from the national e-infrastructure providers SNIC and SUNET. Our computational and storage needs are set up in close collaboration with these infrastructures, in order to avoid duplication of efforts and to benefit from the competences within SNIC and SUNET.

Data management

During 2021, NBIS has expanded its data management efforts by hiring two new data stewards to form a data management team currently consisting of six data stewards and one data manager.

The overarching vision for the NBIS Data management activities is that Swedish life science researchers apply good data management practices so that the research outputs produced are available to the global research community, and to society at large, according to the principles of Open Science, Reproducible Research, and FAIR (Findable, Accessible, Interoperable, Reusable).

The mission of the data management team to work towards this vision is to ensure that Swedish life science researchers have access to data management **support** and **training** to be able to manage their research data according to international best practice, and that these data management practices, solutions and training for Swedish life science are established in **collaboration** with relevant national and international stakeholders. The primary collaborator in this regard is the SciLifeLab Data Centre, with which the collaboration has strengthened further this year as new

initiatives have been initiated, particularly in response to the COVID-19 pandemic. The Data management team has continued to contribute to the content of the national Swedish COVID-19 data portal set up by the Data Centre, and to support SciLifeLab and KAW-funded COVID-19 initiatives. In relation to this, joint work has also been initiated to establish support routines and guidelines for submitting COVID-19 related data to ELIXIR's core deposition databases. The collaboration has also included contributions to the Nordic Pandemic Research Infrastructure (NeIC PaRI) and ELIXIR resources focusing on FAIRifying pathogen genome data.

NBIS has previously, in close collaboration with the SciLifeLab Data Centre, established an online resource for users to create Data Management Plans (DMP). This tool, called the Data Stewardship Wizard (DSW), originating from the ELIXIR collaboration, is provided as a service to the community. In the DSW, we have continued to improve and deepen the life science-focused DMP template model that adheres to the national guidelines.

A key factor in good management of research data is to make generated datasets discoverable and accessible to society, and the research community in particular. This is done most efficiently by submitting the datasets to trusted international public repositories, especially the core Deposition Databases identified by ELIXIR. We have continued supporting users in making such submissions. As the FAIRification of research data is one of the main objectives of the NBIS data management team, one focus activity has been to work out how this can be achieved for the projects that engage the NBIS bioinformatics analysis support. Working together with the NBIS bioinformaticians, the data management team started launching procedures to ensure that data management issues are considered in NBIS support projects, and/or that support projects get the help of a Data Steward when needed. These procedures aim at promoting good data management practices in all NBIS support projects, e.g. that datasets in these projects are published in suitable international repositories whenever possible.

The ethics and laws around personal data present challenges when it comes to making datasets containing e.g. human genetic data available to the research community. The strategy taken is to provide a solution for this under the umbrella of the European Genome-phenome Archive (EGA), and in collaboration with other ELIXIR nodes to establish a Federated EGA landscape. The NBIS systems development team has made considerable progress in establishing an operational Swedish EGA node that should become available as a service for archiving human genetic and phenotypic research data for the Swedish biomedical research community in 2022 (see *Systems Development* below). To complement the technical development efforts, the Data management team has during 2021 established user-facing submission support helpdesk routines, to enable a full-fledged data deposition and sharing service. This work will facilitate high quality data and metadata data depositions, aiming at high FAIRness of the data deposited.

Data Management training efforts have been focused on providing guidelines documentation, and on giving an entry level course on data management practices. There is currently a lack of data management competence and know-how in the Swedish research community. To meet this, a practical course *Introduction to Data Management Practices*, targeted at researchers to take their first steps towards good and reproducible data management practices, was given on two occasions (April and October 2021).

Much effort has been spent on taking on a leadership role in the pan-European ELIXIR-CONVERGE project focused on data management. ELIXIR-CONVERGE aims to connect and align ELIXIR Nodes to deliver sustainable FAIR life science data management services. NBIS as the Swedish ELIXIR node has led the work in work package *WP1 Data Management Expert Network*. A competence network for data management issues has been established across all 23 participating national nodes, with focus on knowledge sharing and capacity building. During 2021, the network has continued to improve and extend the *Data Management best practices guidelines for the Life Sciences*,

which is available at the ELIXIR RDMKit web resource (<https://rdmkit.elixir-europe.org>). In particular, efforts have been made to deepen the guidelines by linking RDMkit with the Data Stewardship Wizard knowledge base. The idea is that RDMkit will provide the context while the DSW will provide the details, when developing good data management strategies. The Data Management team and other NBIS staff have also actively contributed to activities in other work packages *WP2 Training*, *WP3 RDM Toolkit*, *WP5 Demonstrators*, *WP7 Federated EGA*, and *WP9 Mobilisation of SARS-CoV-2 variant surveillance data tracking services and tools*.

The Data Management team has also contributed to EOSC's advisory group on metadata and data quality by representing NBIS in a task force focusing on semantic interoperability.

Compute & Storage

High throughput biomedical science depends on high-performance computers for bioinformatics analysis. The hardware is maintained by SNIC but NBIS provides expertise needed to efficiently access the computational and storage resources. Notably, roughly 1000 bioinformatics-related software packages are installed and maintained on a general-purpose SNIC compute cluster and the SNIC compute cluster for sensitive data, Bianca. Compute & Storage also administers resource allocations and participates in help-desk support at the major national compute clusters, handling 2266 omics research projects with 922 unique PIs in 2021. These projects have used a total of close to 6 million core hours each month and ~10.5 PB of storage.

This year, a focused effort has been made to make storage system usage more efficient. This has allowed roughly 200 more projects and 30 more PIs to share the limited resources this year, as well as providing the necessary motivation for a SNIC decision to invest in a large increase in available non-sensitive data storage and make continued growth possible.

The addition of new staff in the last years has allowed Compute & Storage to improve the level of support in a couple of key areas – workflow managers and container-based software deployment. These are important technologies that dramatically improve the productivity of users. Additionally, a significant effort is spent on supporting users in navigating the judicial issues related to sensitive data processing, e.g. performing Data Protection Impact Assessments and writing Personal Processing Agreements.

Systems Development & Tools

Our systems development team provides support in deploying tools so that they can be used by the entire life science community, and not just by bioinformaticians. We also facilitate large scale analyses by the development of stable and user-friendly pipelines. NBIS works strategically with Reproducible Research, actively promoting and deploying stable workflow systems and reproducible compute environments, both for internal tools development and for supported research projects.

The combination of bioinformatics experts, who can appraise the scientific value and usefulness of tools, with developers who have the capability of making tools stable and accessible, allows us to develop projects for the scientific community in a better way, since we can understand the needs of the user and provide our unique set of knowledge to help them. The development team also assists in internal development projects, e.g. for integrating data services with other initiatives.

Furthermore, the team is active in the Global Alliance for Genomic and Health (GA4GH), working with technical standards and frameworks. The system development teams also serve NBIS infrastructure internally. During the year, this included upgrading and migration of internal infrastructure to new hardware.

NBIS maintains a software repository for internal purposes and also public repositories for programs and scripts that are developed by our staff in order to get them to spread to the scientific community. For the latter, we have an organisational account on GitHub. We are also in charge of registration of relevant tools developed by Swedish research groups into the ELIXIR Tools Registry.

NBIS has worked on several projects during the year, and the largest project is the EU- and NeIC-funded Federated EGA. In this project, several of our developers work together with developers from other ELIXIR nodes.

NBIS is taking charge of the development of Metabolic Atlas, a web platform integrating open-source genome scale metabolic models (GEMs) for easy browsing and analysis. Some tools, such as Data overlay, are developed in the context of quantitative data from the Human Protein Atlas project.

NBIS is also developing a range of pipelines for other data types. For example, our pipelines for single-cell RNA-seq and metagenomics are in advanced stages of development, openly available and already applied in multiple projects (see <https://www.nbis.se/infrastructure/tools/> and <https://www.nbis.se/support/ltsprojects.html>).

Examples of important infrastructure achievements during 2021

Federated EGA

NBIS has in our capacity as the Swedish ELIXIR node worked in the EU projects ELIXIR-CONVERGE and B1MG, and the Nordic NeIC project Heilsa Tryggvedottir to establish a system for handling sensitive human data across the ELIXIR nodes: the federated EGA (European Genome-phenome Archive).

The federated EGA is instrumental to enable the **European 1+ Million Genomes initiative (1+MG)**¹. In April 2018, a joint declaration of cooperation regarding giving access to at least 1 million sequenced genomes in the European Union by 2022 was signed by Sweden and 12 other member states; now a total of 24 member states have signed. To fulfil this vision in the short time-span outlined, it is necessary to build on existing research infrastructures. This involves transfer of existing and emerging technology and know-how between the research and diagnostic areas. Prime examples of technology to support this European effort are the Beacon and federated EGA activities that NBIS is already engaged in.

The federated EGA enables a sustainable, secure and legally correct storage of sensitive human genetic data. This **facilitates data publication and open science** (but with controlled access since the data are sensitive) according to the FAIR principles, which in turn enables data sharing of benefit for scientists, healthcare providers and industry.

¹ <https://ec.europa.eu/digital-single-market/en/european-1-million-genomes-initiative>

During 2021, NBIS has provided knowledge in systems development in order to create important building blocks for data submission at the local EGA nodes. The EGA-SE node is technically ready and legal agreements are negotiated between Sweden and central EGA. In collaboration with our Nordic partners we have added functionality to run a standalone sensitive data archive.

As part of B1MG, NBIS has participated in a proof of concept of sharing human genome data between two countries, ranging from the discovery of data at another site to accessing the data through an approval process and encrypted data transfer.

Metabolic Atlas

Metabolic Atlas is a web platform integrating open-source genome scale metabolic models (GEMs) for easy browsing and analysis. Building on the initial version in 2015, version 1.0 was prepared for the publication of a human GEM in Science Signalling (Robinson et al., 2020). This year marks the 2.0 release with the addition of 5 more GEMs through the publication in PNAS (Wang et al., 2021). With a dedicated Scrum team, Metabolic Atlas will continue the collaboration established in 2019 with the Human Protein Atlas and establish new collaborations. Reaching 135 countries and serving visitors for over 1600 hours per year, Metabolic Atlas is aiming to become an ELIXIR Core Data Resource as envisioned in the NBIS Strategic Plan.

BigPicture

During 2021, the large EU IMI project BigPicture started, aiming at creating the future European digital pathology platform. NBIS provides systems development support to improve the Sensitive Data Archive (SDA, formerly Local EGA), and we have adapted the software stack for the BigPicture project (<https://bigpicture.eu/>) to enable access to sensitive pathology images. In BigPicture we have piloted services for data mirroring between archives and also a federated discovery service.

Trainity

One of the obstacles when delivering training events is the administration of registrations, typically handled manually by course leaders. We are setting up a new system, Trainity, to automate much of the process, saving considerable administration time. Trainity will be fully integrated with our Learning management system Canvas, our website, and the Elixir training portal TeSS, and can be used across all SciLifeLab platforms through the future SciLifeLab Training Platform, also enabling aggregation of course statistics, measuring impact, as well as providing the students an overview of their courses and certificates.

TissUUmeps – a web platform for spatial transcriptomics data

BIIF contributed to the release of a GPU-empowered version of TissUUmeps in 2021. TissUUmeps is a browser-based tool for fast visualisation and exploration of millions of data points overlaying a tissue sample. TissUUmeps can be used as a web service or locally in your computer, and allows users to share regions of interest and local statistics. Tutorials and a gallery of shared projects can be found at <https://tissuumaps.github.io/>. It is used for data from the Human Developmental Cell Atlas (to be released 2022).

Training

Training is one of the pillars of NBIS and the training programme offered mirrors the Bioinformatics knowledge and competence of NBIS staff. NBIS provides individual training of researchers in the majority of our project-related support in order to teach them new bioinformatics tools and to help them utilise bioinformatics tools more efficiently. The unique critical mass of bioinformatics knowledge and skills within NBIS presents us with the opportunity to establish NBIS advanced training at the forefront of Swedish bioinformatics training for life-wide learning in the life sciences. Training as an activity constitutes approximately 15% of NBIS time and involves the majority of NBIS staff. The training includes the delivery of our own NBIS training program as well as collaborating NBIS training, e.g. supporting the VR-funded research school in bioinformatics (MedBioInfo) or as invited lecturer or speaker in training events arranged by others. In addition to our national efforts, NBIS is engaged in various types of training activities and efforts in ELIXIR, both in delivering training and in creating structures and guidelines to both educators and learners with regards to FAIR and OPEN training in the life-wide learning from research infrastructure.

In 2021, the majority of NBIS training events were delivered in an online format and the training programme is tailored to graduate students, post-docs and researchers. There are pros and cons with the online format. Depending on the type of course the online delivery option might continue for specific courses as it increases inclusiveness and accessibility for the participants. We had over 1300 applicants to the NBIS courses, where we accept on average 25 students per course. In total, NBIS collaborated on 36 training events, where 18 courses are part of the NBIS training program and 18 are collaborative efforts with others. The table below lists the NBIS-organised courses, dates and delivery format. The PhD advisory mentor programme in bioinformatics is a long-term initiative in NBIS that is highly appreciated in the life science community where PhD students get mentorship guidance from NBIS experts over a period of 2 years. As a continuation of the transition to online course delivery done in 2020, NBIS courses have been centralised in a Learning Management System, similar to how e.g. universities in Sweden manage courses. The Learning Management System enables all course material to be open and accessible to all. During 2021, a new course on Data management practices for researchers has been launched, and held at two occasions, once online and once on site.

Course	From date	To date	Place
Courses organised by NBIS			
Single cell RNA sequencing analysis	25.01.2021	29.01.2021	Online
Introduction to Data Management	13.04.2021	15.04.2021	Online
Omics integration and Systems biology	19.04.2021	23.04.2021	Online
Tools for reproducible research	19.04.2021	23.04.2021	Online
Quick and Clean - Advanced python	03.05.2021	07.05.2021	Online
Introduction to bioinformatics using NGS	17.05.2021	21.05.2021	Online
RaukR	14.06.2021	24.06.2021	Online
Omics integration and Systems biology	06.09.2021	10.09.2021	Online
Plotting in R	08.09.2021	10.09.2021	Online

Snakemake BYOC (bring your own code)	29.09.2021	01.10.2021	Online
Introduction to biostatistics and machine learning	04.10.2021	08.10.2021	Online
Introduction to Data Management	12.10.2021	14.10.2021	Uppsala
Python programming with applications to bioinformatics	18.10.2021	22.10.2021	Online
Epigenomics data analysis	25.10.2021	29.10.2021	Online
RNaseq data analysis	15.11.2021	19.11.2021	Online
Tools for reproducible research	15.11.2021	19.11.2021	Online
Introduction to bioinformatics using NGS	22.11.2021	26.11.2021	Online
NBIS/SIB Seasonal school: single cell omics analysis	30.08.2021	03.09.2021	Online
Courses co-organised by NBIS			
Smart Microscopy - Image Analysis to Improve Remote Image Acquisition	17.05.2021	21.05.2021	Online
EMBO practical course Advanced Methods in BiImage Analysis	26.06.2021	02.07.2021	Online
ELIXIR Train-the-Trainer	22.10.2021	25.10.2021	Online
ELIXIR Train-the-Trainer	23.11.2021	26.11.2021	Online
ELIXIR Train-the-Trainer	17.12.2021	20.12.2021	Online
NBIS/SIB Seasonal school: single cell omics analysis	30.08.2021	03.09.2021	Online
Zidas2021: ImageJ/Fiji Macro Language	27.09.2021	28.09.2021	Online
QuPath workshop within STorM retreat	12.10.2021	12.10.2021	Online
Internal NBIS Training			
NBIS internal RMarkdown workshop	20.01.2021	20.01.2021	Online
NBIS internal Jupyter workshop	03.02.2021	02.02.2021	Online
NBIS internal Snakemake workshop	15.10.2021	15.10.2021	Online
NBIS internal Container workshop	21.10.2021	21.10.2021	Online

The Training co-production is a model where NBIS reach out to other training providers, both nationally and internationally, in order to collaboratively meet the increasing demand of bioinformatics knowledge and skills. The training co-production model enhances the capacity and knowledge building for researchers as well as empowers the networking aspects across different research infrastructures, organisations etc. for infrastructure staff and course participants alike. Internal training is an important aspect of NBIS training, to transfer knowledge within NBIS and from 2021, we are arranging internal courses in “Tools for reproducible research” for new staff or for staff in need of such upskilling. Open and FAIR as well as reproducible research is a strategic mission, where NBIS actively promotes and deploy stable workflow systems and reproducible compute environments, both for internal tools development and for supported research projects.

We will continue to follow the vision for NBIS Training, which is to offer high quality training programmes aimed at the Swedish Life Science research community in order to grow bioinformatics capacity and competence in Sweden.

Since January 2021, Jessica Lindvall is Executive Committee member of ELIXIR Training Platform, positioning NBIS even further as a training provider for advanced bioinformatics in the life-wide learning trajectory. Further strengthening NBIS and Swedish research infrastructures position with regards to the life-wide learning is the recent involvement in EOSC Association Task Forces (TFs) that were established in the fall of 2021. Here, Jessica Lindvall was appointed chair for the “Upskilling of countries to engage in EOSC”, which is one of three TFs part of the Advisory group “Research Careers and Curricula”. NBIS is also involved in the TFs “Semantic Interoperability” (Wolmar Nyberg Åkerström), “AAI architecture” (Jonas Söderberg), “Technical Interoperability” (Lars Eklund).

Outreach

Outreach activities have proven important to inform the scientific community about the support that NBIS can provide, to increase collaborations and to increase the number of users and quality of projects. The activities consist of involvement in local community initiatives that provide the possibility to meet staff representing our wide variety of competences, presentations at events and conferences and weekly Bioinformatics Drop-in sessions enabling face-to-face contact between researchers and our experts, which many times is the first contact in a support case. During 2021, we arranged 44 site-wide drop-ins via Zoom, due to the pandemics.

NBIS has a community coordinator to oversee and plan NBIS outreach activities and so far two local site coordinators. Collaboration continues between NBIS and NGI to reach out and let life scientists know that we can help strengthen their grant applications using our bioinformatics, sequencing and data management expertise during VR application season.

NBIS also participates in local bioinformatics networks, such as GOTBIN in Gothenburg which held 10 seminars and an Annual meeting in 2021, all virtual. The NBIS-sponsored “Pangenomes: Evolution and Computation” conference, a virtual event, had an average of 300 participants, with scientific presentations and a presentation of NBIS. The Swedish Bioinformatics Workshop, or SBW, was in Lund in 2021 and sponsored in part by NBIS as well as having several NBIS employees in the organising committee. In Umeå, NBIS participated in the UCMR day, KBC days and helped organise two Bioinformatics Mingles and the Machine learning in Life Science event. NBIS has also interacted with LUBI in Lund.

The first SciLifeLab Infrastructure Outreach Week, an outreach and community effort to highlight the infrastructure platforms, started planning in 2021 with the NBIS community coordinator on the organizing task force. The event is planned to be recurring and will feature projects that have received SciLifeLab infrastructure support, such as from NBIS.

During 2021, SciLifeLab BiG Talks! continued with 5 presentations. BiG Talks! is a collaborative effort between NBIS and other SciLifeLab platforms to increase the networking possibilities for staff-scientists and affiliated researchers. The seminars are given by internationally renowned speakers within the fields of Bioinformatics and Genomics. The seminars are, as far as possible, live broadcasted via the SciLifeLab YouTube channel in order to enable all SciLifeLab sites to participate.

Dr. Pedro Beltrao (EMBL-EBI), Mihail Anton (Chalmers University) and Dr. Lars Juhl Jensen (Novo Nordisk Foundation Center for Protein Research) were all highly appreciated open and live streamed seminars, which were part of the NBIS course “Omics integration and Systems biology” as well as Associate Professor Molly Gale Hammell (Cold Spring Harbor laboratory), part of the Swedish Bioinformatics Workshop.

In collaboration with the Netherlands ELIXIR node and the SciLifeLab Data Centre, the NBIS Data management team organised an online workshop on *FAIR principles in life science research practice*, with keynote talks by Erik Schultes (GO-FAIR Foundation) and Rob Hooft (ELIXIR-NL/DTL).

Finally, NBIS started its internal Outreach group meetings again which maintains a github repository of outreach materials and presentations as well as discusses NBIS outreach strategy, visual profile and more.

Collaboration with industry

NBIS continues our efforts to increase contacts with life science companies, which can utilise our resources at a full-cost basis, enabling them access to tools, expertise, and training.

Contributions by the Swedish ELIXIR node (ELIXIR-SE)

- ELIXIR-SE has continued maintaining and updating the Human Protein Atlas – which is officially named as an ELIXIR Core Data Resource – in the ELIXIR landscape.
- ELIXIR-SE has been very active in the systems development for and coordination and support of Federated EGA in collaboration with other Nordic ELIXIR nodes, ELIXIR-Spain and ELIXIR-EBI.
- We have continued our engagement in the two EU-funded projects ELIXIR-CONVERGE for data management and B1MG for the European 1+ Million Genome project.
- In 2021, the EU IMI-project BigPicture started, aiming at developing the European infrastructure for digital pathology.
- As of January 1st 2021, Jessica Lindvall (Deputy Head of Node) was elected Executive Committee member to the ELIXIR Training Platform.

Staff

The table shows the number of staff in FTE (full time equivalents) during 2021 for the different functions in NBIS. Gender balance is 72% male and 28% female.

NBIS	Staff (FTE)
1 Central functions, incl. management	9.7
2A Support (SMS)	14.9
2B Support (LTS)	16.4
2C Support (Sysbio)	1.1
3 Data management	3.9
4 Systems development	8.9
5 Compute & Storage	3.9
6 Training	11.0
7 ELIXIR	5.1
8A BioImage Informatics	2.4
8B AIDA Data Hub	1.5
Other	0.5
TOTAL	79.4

Economy report for 2021

NBIS	2021	Result	Budget
Incomes	VR	20 000 000	20 000 000
	SciLifeLab National	24 550 000	24 050 000
	Universities + SciLifeLab SFO	15 815 000	15 815 000
	KAW	21 867 955	19 200 000
	EU + ELIXIR	4 961 210	5 615 000
	NeIC NordForsk	850 400	1 300 000
	User fees	10 300 000	8 700 000
	SUM Incomes	97 744 565	94 680 000
Expenses	Personnel	66 805 048	70 063 200
	Equipment	2 827 667	
	Travel	821 617	
	Other costs	1 415 137	
	Office space	4 313 611	24 616 800
	Indirect costs	18 403 808	
	SUM Expenses	94 586 888	94 680 000
	Surplus forwarded to 2022	3 157 677	

The surplus from 2021 is due to outstanding requisitions for staff costs, parental leaves and delayed recruitments; the surplus will be used for additional staff recruited in 2022.

Distribution of costs on the different NBIS activities

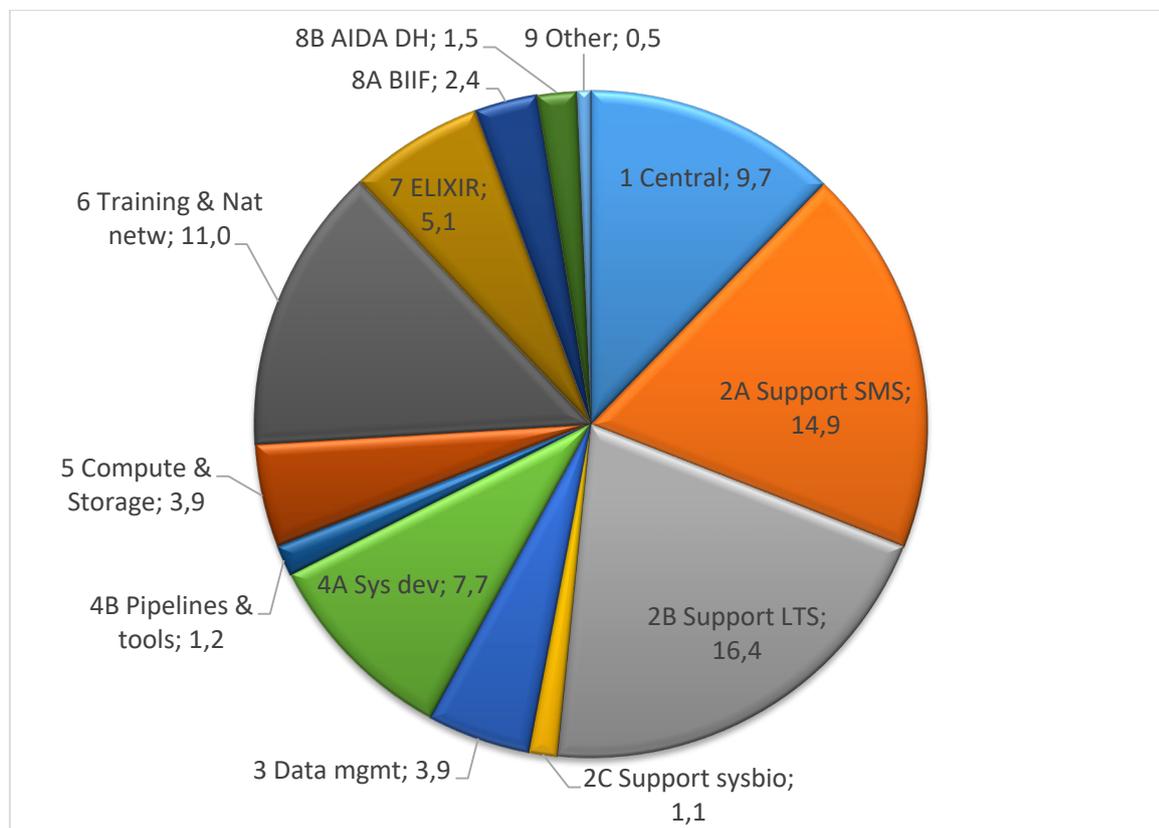
Activity	kSEK	Budget ¹
1 - Central functions ²	12 348	9 300
2A - Support SMS	15 629	22 250 ³
2B - Support LTS	19 654	20 000 ³
2C - Support Sysbio	1 403	5 000 ³
3 - Data management	6 659	8 280 ³
4A - Systems development	8 906	12 200 ³
4B - Pipelines & tools	1 376	
5 - Compute & Storage	4 099	6 000 ³
6 - Training & Nat networking	12 927	³
7 - ELIXIR	6 676	6 600
8A - Biolmage Informatics	2 628	
8B - AIDA data hub	1 681	
9 - Other	601	
SUM	94 587	
SUM (excluding new modules 8A 8B)	90 278	89 630

¹Budget version 9 Nov 2019 submitted to VR

²Central functions also include project management for Support

³Training costs included in the modules 1–5

Distribution of FTEs on the different areas



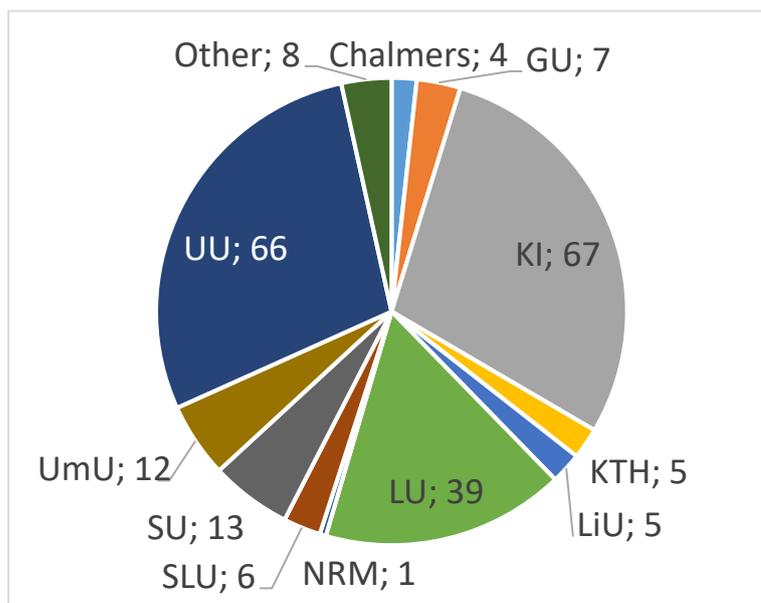
Annex – Key performance indicators

1. Number of projects

During 2021, we have worked on 276 support projects and provided 79 consultations. The 276 support projects involve 233 unique PIs (88 female, 145 male).

2. Number of PIs distributed on universities

Univ	# PI
Chalmers	4
GU	7
KI	67
KTH	5
LIU	5
LU	39
NRM	1
SLU	6
SU	13
UmU	12
UU	66
Other	8



3. Number of projects distributed on SCB codes

SCB code and subject	Number of Projects
102 Computer and Information Sciences	6
105 Earth and Related Environmental Sciences	2
106 Biological Sciences	89
107 Other Natural Sciences	1
205 Materials Engineering	1
209 Industrial Biotechnology	1
301 Basic Medicine	65
302 Clinical Medicine	73
303 Health Sciences	6
304 Medical Biotechnology	8
305 Other Medical and Health Sciences	13
401 Agricultural, Forestry and Fisheries	6
403 Veterinary Science	3
404 Agricultural Biotechnology	1
501 Psychology	1

4. Gender balance

Total 297 PIs that have received support, of which 117 female (39%) and 180 male (61%).

5. Publications

A total of 93 publications in 2021 have been published from our direct project support work (excluding Compute & Storage and AIDA Data Hub); DOIs provided in the table below.

10.3390/cancers14010077	10.1038/s41598-021-95109-0	10.1021/acs.jproteome.1c00018
10.3390/cancers14010009	10.1021/acs.jproteome.1c00387	10.1038/s41422-020-00445-x
10.1016/j.eururo.2021.12.013	10.1016/j.gene.2021.145723	10.1038/s42255-021-00391-x
10.1002/1878-0261.13158	10.1186/s40478-021-01234-2	10.1002/cpz1.89
10.1093/rheumatology/keab912	10.1093/molbev/msab086	10.1136/bmjdr-2020-002076
10.12688/f1000research.73825.1	10.1073/pnas.2102344118	10.1055/s-0040-1720980
10.1038/s41597-021-01090-8	10.1080/15476286.2021.1952758	10.12688/f1000research.25485.1
10.1016/j.tox.2021.153028	10.1016/j.cell.2021.05.030	10.1128/MRA.00175-21
10.1016/j.apsoil.2021.104140	10.1038/s41467-021-24549-z	10.1073/pnas.2011574118
10.1111/mec.16291	10.1002/dvdy.399	10.1038/s41467-021-22386-8
10.1016/j.gim.2021.09.003	10.1183/13993003.03468-2020	10.12688/f1000research.52569.1
10.1016/j.yexcr.2021.112880	10.1002/jev2.12128	10.1186/s12931-021-01724-w
10.3390/ijms222112033	10.1016/j.ygeno.2021.05.021	10.1186/s12864-021-07454-z
10.1183/13993003.00142-2021	10.1126/sciadv.abh2169	10.1186/s13148-021-01054-0
10.1111/mec.16252	10.3389/fmicb.2021.667718	10.1016/j.neuron.2021.01.018
10.3390/cancers13215341	10.1007/978-1-0716-1503-4_2	10.1080/17453674.2020.1846956
10.1210/clinem/dgab471	10.1038/s41398-021-01481-w	10.1371/journal.pcbi.1008898
10.1186/s12859-021-04407-x	10.1016/j.celrep.2021.109286	10.1186/s12859-021-04043-5
10.1186/s12864-021-08021-2	10.3390/cells10061543	10.1002/acn3.51296
10.1080/10428194.2021.1913147	10.1186/s13059-021-02386-z	10.1038/s41598-021-83541-1
10.3389/ffunb.2021.716385	10.1002/cyto.a.24467	10.1128/mSystems.00884-20
10.1038/s41598-021-97491-1	10.1186/s13068-021-01975-1	10.1002/ijc.33270
10.1093/molbev/msab277	10.1093/nargab/lqab045	10.1182/bloodadvances.2020003709
10.1016/j.immuni.2021.08.017	10.1007/s10126-021-10033-8	10.1093/jn/nxaa391
10.1111/evo.14234	10.1038/s42003-021-02142-w	10.1016/j.chemosphere.2020.128467
10.1111/1365-2745.13550	10.3390/jpm11060448	10.1002/dvdy.253
10.1002/lno.11896	10.3389/fvets.2021.642085	10.1111/mec.15787
10.1038/s41467-021-25186-2	10.1038/s41467-021-23027-w	10.1186/s12974-021-02082-6
10.1002/JLB.5A0321-162R	10.1038/s41598-021-89986-8	10.1038/s41467-020-20586-2
10.1371/journal.pone.0255953	10.1093/rheumatology/keab441	10.1111/1755-0998.13252
10.1093/g3journal/jkab214	10.1016/j.ijcard.2021.01.059	10.1080/15592294.2020.1788325

6. Number of users that have applied for access but not being prioritised

Here we show numbers from the WABI part of NBIS, since other users have the option of paying user fees as long as their project is technically feasible and NBIS has capacity. In 2021, a total of 57 applicants of which 18 were granted LTS support (32%). Female: 28 applicants of which 6 were granted (21%). Male: 29 applicants of which 12 were granted (41%). NBIS has observed that the gender distribution in 2021 was skewed, and we have therefore started to investigate explanations in order to avoid any future unfair bias.

In addition, we can mention that during consultations or drop-in sessions we identify if a project is technically suitable. If a project is not technically sound, that project never enters our support system.