Annual Report 2022

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.

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 2022, NBIS provided support to **283 PIs** (117 female, 166 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 62 study design consultations. These numbers are at about the same levels as 2021.
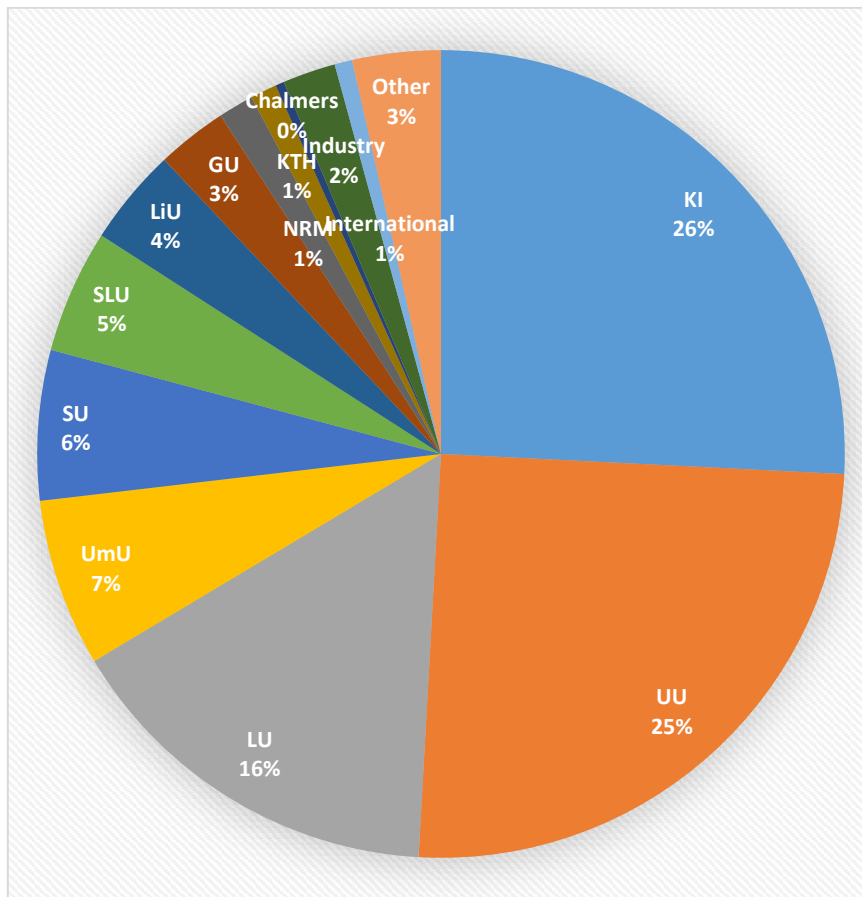

Figure 1 Distribution of PIs that have received support during 2022

In 2022, short- and medium-term support (SMS) was provided to 135 projects and long-term support (LTS) to 91 projects. In addition, NBIS was engaged in 6 partner projects, and the BioImage Informatics Facility provided support to 44 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 **79 publications** in 2022 from our direct project support work (excluding Compute & Storage), listed in the Annex. A few publications are shortly highlighted here.

Metatranscriptomics of Nasopharyngeal Microbiota and Host Distinguish between Pneumonia and Health. Lower respiratory tract infections, including community-acquired pneumonia, are major contributors to morbidity and mortality world-wide especially in children. With the Henriques-Normark's group (KI), we developed a novel metatranscriptomic pipeline with ultrasensitive sequencing of nasopharyngeal aspirates to map the expression of the nasopharyngeal microbiota and associated host responses, including a Machine Learning approach for identifying potential new biomarkers. DOI: 10.1164/rccm.202203-0463LE

Clonal relations in the mouse brain revealed by single-cell and spatial transcriptomics. With the Frisén's lab (KI), we have developed a method to simultaneously profile cell phenotypes and clonal relations in vivo, based on inserting DNA barcodes into early progenitors, followed by single cell RNA-sequencing. In the mouse brain, thousands of clones could be reconstructed, which enabled us to discover fate-restricted progenitor cells in the mouse hippocampus. We also showed that microglia are derived from few primitive myeloid precursors that massively expand throughout the brain. This method enables high-throughput reconstruction of cell phenotypes and clonal relations at the single-cell and tissue level in individual animals. DOI: 10.1038/s41593-022-01011-x

High-throughput muscle fiber typing from RNA sequencing data. Together with the Hansson group (LU) and the 2022 Nobel laureate Svante Pääbo (Max Planck Institute), we have investigated muscle samples from humans and apes, which is of evolutionary interest and has implications for human health. We developed a new deconvolution method for accurate prediction of muscle fibre composition from low-coverage bulk RNAseq in biopsies, which is a radically more labor- and cost-efficient approach than traditional staining, making it feasible to perform well-powered muscle fibre type studies. DOI: 10.1186/s13395-022-00299-4

The COVID-19 pandemic The app-based COVID Symptom Study was launched in Sweden in April 2020 to contribute to real-time COVID-19 surveillance. NBIS has collaborated with this national and international consortium, for example showing ability to help predict the next week COVID-19 hospital admissions (Kennedy et al., 2022). The NBIS Data Management team has also continued to collaborate with the Data Centre within the Pandemic Laboratory Preparedness (PLP) initiative, e.g. supporting the Swedish Environmental Epidemiology Center (SEEC) in sharing genomic reads and associated data from waste water samples at the COVID-19 portal.

Biodiversity: Pollinator crisis Our expertise has been valuable for biodiversity studies, e.g. regarding *Anthophora plagiata* (Humlepälsbi), a threatened species and thought extinct in Sweden, but a few populations were recently discovered. During 2022, NBIS has assembled and annotated the complete DNA sequence of the species, which will be used to infer genetic diversity and allow

the nature conservation experts in the county board of Värmland (Länsstyrelsen) to study and decrease the risks of population inbreeding.

BioImage Informatics

The BioImage Informatics Unit (BIIF) provides support and training to perform state-of-the-art analysis on image data. With the innovative progress and quickly increasing interest in spatially resolved omics techniques such as in situ sequencing, spatial transcriptomics, and multiplexed immunohistochemical staining, BIIF has taken a major role in supporting users working with this type of data. Our GPU-accelerated viewer TissUUmaps (https://tissuumaps.github.io/), developed in close collaboration with the Wählby lab, has been pivotal in forwarding biomedical research on development and cancer. One example is the open release of the Human Developmental Lung Cell Atlas, which is now possible to view via TissUUmaps at https://tissuumaps.dckube.scilifelab.se/web/private/HDCA/index.html.

Furthermore, BIIF together with other NBIS staff organized a workshop on Spatial Omics Data Analysis on August 29 – September 3, 2022, providing training in advanced tools for analysis of spatial datasets via a series of online lectures and computer practicals. See more info here: https://www.scilifelab.se/event/spatial-omics-data-analysis/ The course had 120 applicants, showing the great interest for these techniques. We also organised an internal workshop on Spatial Omics tools for the Single cell and Spatial Biology platform in November 2022, strengthening the connection between the facilities for improved user support.

We continuously develop new computational tools in collaboration with the Wählby lab, and our graph neural network-based methods for mapping of individual cells and developmental processes in the heart were released at
https://tissuumaps.dckube.scilifelab.se/human_heart.tmap?path=demo
in collaboration with the Mats Nilsson group, and published in PLOS Computational Biology (https://doi.org/10.1371/journal.pcbi.1010366)

We have continued our collaboration with the new DDLS fellow, Wei Ouyang (KTH) in using the ImJoy server to bring deep-learning tools to our users, through the BioEngine solution in connection with SciLifeLab Serve (platform from SciLifeLab Data Centre). From the perspective of imaging facilities, the BioEngine is a more scalable and multi-user architecture. Image analysis pipelines and workflows can then be implemented as web applications via ImJoy and Napari plugins or Jupyter notebooks as well as other generic APIs for image analysis.

We collaborate with the AIDA Data Hub, and see great potential for development by closer integration and physical placement of expertise at LiU. This will help us provide the best possible support for projects with a strong clinical component and special requests for secure data handling close to computational power.

Furthermore, we have been active in the Nordic and European Bioimage Analysis community and plans are to further shape these communities, expanding our network of experts and keeping up with the latest developments in the field. Gisele Miranda from BIIF is funded as Imaging Scientist by the Chan-Zuckerberg Initiative (CZI) and is actively involved in their network worldwide. CZI has developed into a major funding resource in the last years with a special focus on funding open-source tools and networks connected to bioimaging.

In parallel with the analysis of SpatialOmics, we have also focused on quantitative analysis of the morphology of the underlying tissue samples, with methods based on AI in the form of deep convolutional neural networks proving to be powerful in classifying tissue structures, exemplified with a BIIF-support project on tumour draining lymph nodes, published in Journal of Pathology (https://doi.org/10.1002/path.5981). In this project, led by Maria Ulwmar, we developed an AI tool that can detect and classify lymph nodes as normal or dilated, which has previously been shown to be relevant for predicting tumour aggressiveness. Automated lymph node detection and classification enables large scale studies and extension of the analysis to other cancer types.

Although no longer directly supported by the national SciLifeLab funding, we would also like to highlight the progress on developing zebrafish tools for automated large scale screening, by Amin Allalou. Together with Johan Ledin and Beata Filipek Gorniok (previously Genome Engineering Zebrafish at SciLifeLab) they have established a new service facility at Uppsala called DanioReadout. The service facility provides genetic disease modeling and substance exposure services in the zebrafish embryo, integrated with unique methods for imaging and analysis of image data. This has led to four publications in 2022, in eLife (PMID: 36377467), PLoS Genetics (PMID:35192612), Developmental Dynamics (PMID: 34242444) and Cell Reports (PMID: 35732122).

We have also published several overview articles together with the international bioimage analysis community, e.g. a glossary of current bio-image analysis tools (https://doi.org/10.1002/1873-3468.14451), which contribute to the outreach in the field.

AIDA Data Hub

AIDA Data Hub has engaged in the Swedish Government cooperation group for Health and Life Sciences, working group Health Data, which delivered its report on reuse of clinical health data to the Swedish government offices in November.

The AIDA Data Hub DGX-2 Service (https://datahub.aida.scilifelab.se/services/#dgx-2) is host to a VINNOVA funded AIDA-SCAPIS Data lab, for secure and private sharing and processing of large volumes of SCAPIS medical imaging data in a trusted research environment. The SCAPIS Data lab storage has been extended to enable providing read-only access to all SCAPIS imaging datasets to approved research groups.

The AIDA Nextcloud service (https://datahub.aida.scilifelab.se/services/#nextcloud) is operated in collaboration with the SciLifeLab Data Centre, and is currently being used to facilitate integration of a large pathology dataset, whose addition is expected to double the amount data shared through the AIDA Data Hub.

The AIDA Data Hub DGX-2 Service (https://datahub.aida.scilifelab.se/services/#dgx-2) is the current demonstrator implementation of the LiU part of the RÖ–-LiU common vision for double data lakes, to increase accessibility to clinical data for research, analogous to European Health Data Spaces for primary- and secondary use of health data. As a next step toward achieving this vision, a more featureful successor to the AIDA DGX-2 Service is being planned. Procurement of hardware is currently carried out in cooperation with the NSC Berzelius Sensitive expansion to support the implementation of the DDLS data platform. The system design of the successor service will be kept close to that of Berzelius / Berzelius sensitive, in order to facilitate data- and user mobility. The AIDA Data Hub DGX-2 Service is currently tailored to suit the needs of leading

edge national expert AI researchers. The planned feature additions are intended to make the service more usable by further user competence profiles (such as clinicians) and other usage patterns (such as through a locked-down remote desktop, or a project-private PACS). The intention is to keep a low-level access model open, in order to appropriately support needs of the current user base. The project-private PACS functionality is pursued in collaboration with Sectra.

The Bigpicture repository has received its first three clinical datasets, from Region Östergötland, Medical University of Vienna, and University Medical Center Utrecht.

The research projects utilising AIDA Data Hub have resulted in many publications, cf. annex. One publication (in The Journal of Machine Learning for Biomedical Imaging, https://www.melba-journal.org/papers/2022:023.html) concerned self-supervised contrastive learning, bringing new insights to how these methods should be adapted when applied in histopathology. Another example is a paper (in Scientific Reports, https://www.nature.com/articles/s41598-022-11826-0) investigating if uncertainty assessments could contribute to precision or robustness of AI predictions in histopathology. Moreover, experiences from the AIDA Data Hub development contributed to a mapping of care providers quality assurance methods for AI validation and monitoring (published in Journal of Digital Imaging, https://link.springer.com/article/10.1007/s10278-022-00731-7).

User fees

NBIS has five support tracks:
- Study design consultation (≤ 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 2022, the income from user fees was 9.5 MSEK, which is slightly lower than previous year.

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.

Human data

NBIS contributes to major European efforts for handling sensitive human data, and has since long been engaged in building these infrastructures. To support the European 1+ Million Genomes Initiative (1+MG), the 40 MEUR EU project GDI (Genomic Data Infrastructure) was launched in 2022. A major part in this project is construction of the infrastructure, led by NBIS as ELIXIR-

SE together with ELIXIR-FI. The work builds upon previous initiatives where we have been involved (B1MG, NeIC Tryggve, NeIC Heilsa Tryggvedottir). Sweden is one of the vanguard nodes in enabling genomic and phenotypic data sharing in Europe. In B1MG, we successfully demonstrated cross-border access in a Proof-of-Concept (https://www.youtube.com/watch?v=6MtIJA4xXdU)

During 2022, NBIS signed the Federated European Genome-phenome Archive (EGA) Collaboration Agreement, thereby formally joining the Federated EGA for GDPR-compliant sharing of large volumes of sensitive research data. Early 2022, NBIS, the Centre for Genomic Regulation Spain and the European Bioinformatics Institute successfully finished an end-to-end test of the Swedish node. The Federated EGA Sweden is an NBIS service and the deployed software is a result of our long standing collaborative development with the Nordic ELIXIR Nodes, EMBL-EBI and CRG.

Bigpicture is a flagship EU-IMI project 2021–2027 to establish a Petabyte platform for European digital pathology AI. NBIS leads the repository infrastructure development. The repository is now in operation, based on ELIXIR technologies developed for genetics: Federated EGA, Life Science AAI, REMS, and ELIXIR Beacon. The repository has received its first three clinical datasets, from Region Östergötland, Medical University of Vienna, and University Medical Center Utrecht. Service integration was demonstrated in Jan 2023, hands-on using live services, in the opening plenary session at the Bigpicture annual meeting in Vienna, by going through an end-to-end platform user story showcasing data submission, discovery, access request and approval, download, and on-platform data use. Large-scale archive operations started Feb 2023.

Data management and Open Science support

During 2022, NBIS has expanded its data management efforts by hiring two new data stewards to form a data management team currently consisting of eight data stewards and one data manager.

NBIS has continued its mission to provide data management support and training in collaboration with key national and international stakeholders to enable Swedish life science researchers to 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).

On the support side, we have continued providing Data Management Planning (DMP) support, as well as FAIRification and data publication support to research projects. Together with the SciLifeLab Data Centre, activities have been focused on developing and improving tools functionality for data management, such as the Data Stewardship Wizard for DMPs and sample metadata handling. NBIS has also laid the groundwork for a user helpdesk for the EGA-SE node to enable a straightforward and efficient submission process for researchers.

Together with the SciLifeLab Data Centre we have launched a joint support portal and guide-line knowledge-base for research data management for Swedish life science (https://data-guidelines.scilifelab.se/), as well as contributed to the Swedish COVID-19 data portal.

NBIS is leading one work package in the EU project ELIXIR-CONVERGE focused on data management and has now created a pan-European competence network for data management issues across all 23 participating national nodes, with focus on knowledge sharing and capacity

building, and training for data management. We have also contributed to the ELIXIR RDMKit (https://rdmkit.elixir-europe.org).

Computational support

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 2462 omics research projects with 861 unique PIs in 2022. These projects have used a total of close to 6 million core hours each month and ~10.5 PB of storage.

During 2022, it became clear that the advanced user support organisation within SNIC will not continue into 2023 with NAISS. NBIS has therefore since 2023 pledged to provide this support for the Life Science community in collaboration with other support providers in the NAISS context, as well as SciLifeLab Data Centre. We expect a continuing high number of projects as well as increasing sizes due to decreasing costs for sequencing and the increased capacity of NGI (National Genomics Infrastructure).

Systems Development & Tools

The systems development team at NBIS is dedicated to supporting the life science community by deploying tools and facilitating large scale analyses. They develop custom-made tools for researchers and user-friendly pipelines for stable and efficient analysis. The team works towards promoting reproducible research through the deployment of stable workflow systems and compute environments. They follow the scrum methodology in managing both internal and external projects and aim to make cutting-edge bioinformatics accessible to all.

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.

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.

Metabolic Atlas is a web platform integrating open-source genome scale metabolic models (GEMs) for browsing and analysis. In the 2022 release we implemented GotEnzymes (Li et al., 2022), an extensive database of predicted enzyme turnover numbers. The functionality to overlay scaled values onto components was also extended to allow multiple simultaneous overlays, custom uploads and inclusion of reaction data. Metabolic Atlas is now part of the Identifiers.org Central Registry, and during 2022 reached users in 130 countries, serving visitors for over 2700 hours (80% increase compared to 2021). Metabolic Atlas is an ELIXIR-SE service, aiming to become an ELIXIR Core Data Resource. https://metabolicatlas.org/

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).

The AIDA Data Hub provides compute services for AI research on sensitive personal data in diagnostic imaging. Among many resulting publications, one example is an effort on self-supervised learning for pathology. Moreover, the infrastructure is providing policy support and other guidance for the domain. Current developments include plans to be a strategic collaborator to the SCAPIS project for data sharing of their unique imaging data collections.

In 2022, NBIS has launched **Chemical Space** – a tool for selection of compounds in a hit-to-lead optimisation process in drug development. They also contributed to the development of **Grnbenchmark.org**, a benchmarking platform for gene regulatory network inference methods.

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 2022, about half of the NBIS training events were delivered in an online format, while half were organised on-site as before the pandemics. Our training programme is tailored to graduate students, post-docs and researchers. This year we have three new courses: Neural networks, Spatial omics, and Train the Trainer, while we did not give these three courses: Omics integration (will instead run early 2023), Advanced Python, and SIB single cell course. In total, NBIS had 43 training events, where 19 courses are part of the NBIS training programme and 24 are collaborative efforts, cf. table below. During 2022, NBIS had 424 participants, not counting the collaborative efforts. We

had close to 900 applications, so the application rate was 48%. The gender distribution is close to equal with 54% females accepted.

During 2022, NBIS delivered our Omics integration and Systems Biology workshop to the Chan Zuckerberg Initiative community reaching beyond the European life science community as a more targeted international user community was approached, the majority being American-based researchers.

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.

Course	From date	To date	Place
Courses organised by NBIS			
Neural nets and deep learning	17.01.2022	21.01.2022	Uppsala
Single cell RNAseq analysis	02.02.2022	04.02.2022	Stockholm
Introduction to bioinformatics using NGS data	28.03.2022	01.04.2022	online
Introduction to Data Management Practices	05.04.2022	07.04.2022	Uppsala
Tools for Reproducible Research	25.04.2022	29.04.2022	online
Data Visualization in R	27.04.2022	29.04.2022	Lund
Nordic ELIXIR-GOBLET Train-the-Trainer	10.05.2022	13.05.2022	online
AI course for clinicians	16.05.2022	18.05.2022	Linköping
RaukR, Advanced R for Bioinformatics, Summer course	13.06.2022	17.06.2022	online
Spatial omics data analysis	29.08.2022	02.09.2022	online
Snakemake bring-your-own-code (BYOC)	31.08.2022	02.09.2022	online
Introduction to biostatistics and machine learning	12.09.2022	16.09.2022	Uppsala
Introduction to bioinformatics using NGS data	26.09.2022	30.09.2022	Uppsala
Introduction to Python with applications to bioinformatics	10.10.2022	14.10.2022	Uppsala and Umeå
Introduction to Data Management Practices	11.10.2022	13.10.2022	Stockholm
Epigenomics Data Analysis: from Bulk to Single Cell	24.10.2022	28.10.2022	online
Programming formalisms for life scientists and bioinformaticians	24.10.2022	28.10.2022	online
AI course for clinicians	09.11.2022	11.11.2022	Linköping
AIDA Advanced AI training workshop with ENCCS	07.12.2022	20.12.2022	online
Courses co-organised by NBIS			
The Swedish Bioinformatics Workshop	27.10.2022	28.10.2022	Umeå
Training Data Stewards for Life Sciences	15.11.2022	17.11.2022	Oeiras, Portugal
EMBO practical course: Single-cell omics: deeper to genomics	13.11.2022	18.11.2022	Heidelberg, Germany
ERGA Nextflow training and hackathon	30.11.2022	30.11.2022	Online
Workshop on Genomics	22.05.2022	03.06.2022	Cesky Krumlov, Czech Republic
EMBL Course - Deep learning for image analysis	17.01.2022	21.01.2022	online

Introduction to Image Analysis Software	09.06.2022	09.06.2022	Uppsala
3MR103 Bioinformatics Course	06.05.2022	06.05.2022	Uppsala
Digital image analysis for scientific applications - focus MAX IV	18.01.2022	24.02.2022	Uppsala
QuPath in Experimental Oncology	01.02.2022	01.02.2022	Lund
Microscopy: from sample preparation to image analysis	09.02.2022	10.02.2022	Stockholm
Defragmentation: bringing BioImage Analysts to the cloud	07.10.2022	07.10.2022	online
ImageJ macro writing	21.09.2022	22.09.2022	Göteborg
ZIDAS 2022	10.07.2022	11.07.2022	Lausanne
Advanced Microscopy Course: Image Analysis	05.12.2022	05.12.2022	Umeå
BIIF - Spatial Omics Workshop	10.11.2022	10.11.2022	Stockholm
Spatial Omics Data Analysis	29.08.2022	02.09.2022	online
To awk or not	18.01.2022	18.01.2022	Uppsala
To awk or not	29.08.2022	30.08.2022	Uppsala
BK0001 Bioinformatics	29.08.2022	31.10.2022	Uppsala, SLU
1MB438 Introduktion till bioinformatik	02.09.2022	25.10.2022	Uppsala, UU
UPPMAX Introductory Course	14.02.2022	17.02.2022	Uppsala
UPPMAX Introductory Course	15.08.2022	19.08.2022	Uppsala
Programming formalisms	24.10.2022	28.10.2022	Uppsala
Internal NBIS Training			
Internal NBIS workshop: Git for reproducible research	10.05.2022	10.05.2022	online
Internal NBIS workshop: Nextflow for reproducible research	12.05.2022	12.05.2022	online
Internal NBIS workshop: Git GUI	07.11.2022	07.11.2022	online

The Training co-production is a model where NBIS reaches 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 2022, we arranged 38 national drop-ins via Zoom and were represented by speaking or poster at 77+ events.

NBIS has a community coordinator to oversee and plan NBIS outreach activities and so far two local site coordinators. Collaboration continued in 2022 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 the VR application season.

The Swedish Bioinformatics Workshop (SBW) was in Umeå 2022 sponsored in part by NBIS and with several NBIS staff in the organising committee. During the two days of the event, 14 speakers gave lectures ranging from 15 to 45 minutes each, nine of which were short talks given by PhD students and PostDocs, and the rest were longer talks given by established PIs and well-renowned keynote speakers. There was also a poster session with a total of 25 posters shown at the event and a career panel discussion where 3 professionals representing academia, NBIS and Clinical Genomics answered questions about their careers and gave advice for future graduates. 129 people registered for the event, of those around 108 people (~84%) showed up. Attendees were almost exclusively from Sweden, with the exception of one person from Finland. A bit more than half of attendees (54 %) were from Umeå.

NBIS is engaged in multiple local activities across the country. In Gothenburg, NBIS is involved in managing the local bioinformatics network GOTBIN. NBIS is also active in the SciLifeLab Gothenburg site and provides input in the advisory board for the Core Facilities at University of Gothenburg and Sahlgrenska Academy. NBIS staff has contributed to the establishment of a DDLS Data Science Node at Chalmers. In Stockholm, NBIS has connections with Clinicum providing common support structures for clinical and translational research within KI and the Stockholm Region; Data-driven research at KI to promote data-driven research across the whole range of medical science and education practised at KI; Precision Medicine Center Karolinska (PMCK) to create the conditions for taking precision medicine all the way into the everyday life of healthcare. In Umeå, NBIS organises bioinformatics mingle events to increase local contacts.

Finally, NBIS has internal Outreach group meetings which maintain 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 from 2022 provides the Metabolic Atlas as an ELIXIR service.
- 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.
- During 2022, several ELIXIR-related EU projects were funded:
    - o GDI (Genomic Data Infrastructure)
    - o EUCAIM providing infrastructure for cancer image data
    - o PHENET (where NBIS will provide training on the European level)
- 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.

Staff

The table shows the number of staff in FTE (full time equivalents) during 2022 for the different functions in NBIS. Gender balance is 72% male and 28% female. Number of FTEs has increased from 79.4 in 2021 to 91.4 in 2022.

NBIS	Staff (FTE)
1 Central functions, incl. management	9.4
2A Support (SMS)	19.4
2B Support (LTS)	16.2
2C Support (Sysbio)	0.1
3 Data management	4.1
4 Systems development	13.3
5 Compute & Storage	4.1
6 Training	11.7
7 ELIXIR	4.3
8A BioImage Informatics	4.4
8B AIDA Data Hub	4.0
Other	0.5
TOTAL	**91.4**

Economy report for 2022

NBIS	2022	Result	Budget
Incomes	VR	20 000 000	20 000 000
	VR grant EGA-SE	1 401 270	
	SciLifeLab National	26 550 000	22 000 000
	Universities + SciLifeLab SFO	16 558 000	16 218 000
	KAW	21 996 109	17 000 000
	EU + ELIXIR	4 092 853	3 000 000
	NeIC NordForsk	1 892 035	1 000 000
	User fees	9 500 000	14 000 000
	From surplus 2020 and 2021	3 791 415	
	SUM Incomes	**105 781 682**	**93 218 000**
Expenses	Personnel	72 108 240	
	Equipment	3 416 967	69 160 326
	Travel	1 904 661	
	Other costs	2 526 954	
	Office space	5 178 511	24 206 114
	Indirect costs	20 646 349	
	SUM Expenses	**105 781 682**	**93 366 440**

Distribution of costs on the different NBIS activities

Activity	kSEK	Budget[1]
1 - Central functions[2]	12 405	9 486
2A - Support SMS	23 713	22 690[3]
2B - Support LTS	18 891	20 400[3]
2C - Support Sysbio	162	5 202[3]
3 - Data management	4 864[4]	10 396[3]
4A - Systems development	13 217	12 442[3]
4B - Pipelines & tools	2 640	
5 - Compute & Storage	4 165	6 120[3]
6 - Training & Nat networking	12 160	[3]
7 - ELIXIR	5 856	6 732
8A - BioImage Informatics	2 767	
8B - AIDA data hub	3 951	
9 - Other	990	
SUM	**105 782**	
SUM (excluding new modules 8A 8B)	**99 064**	**93 366**

[1]Budget version 9 Nov 2019 submitted to VR
[2]Central functions also include project management for Support
[3]Training costs included in the modules 1–5
[4]Several activities in data management are filed under 6 Training and 7 ELIXIR

Annex – Key performance indicators

1. Number of projects

During 2022, we have worked on 276 support projects and provided 62 consultations.
The 276 support projects involved 236 unique PIs (95 female, 141 male).
Altogether, we have supported 283 PIs.

2. Number of PIs distributed on universities

Univ	# Unique PI support and/or consultation	# Unique PI excl. consultations
Chalmers	1	1
GU	8	4
KI	73	65
KTH	3	3
LiU	11	8
LU	44	38
NRM	4	4
SLU	14	10
SU	17	15
UmU	19	14
UU	71	61
Other	18	13

3. Number of projects distributed on SCB codes

SCB code and subject	
102 Computer and Information Sciences	4
105 Earth and Related Environmental Sciences	2
106 Biological Sciences	94
107 Other Natural Sciences	2
301 Basic Medicine	63
302 Clinical Medicine	53
303 Health Sciences	11
304 Medical Biotechnology	2
305 Other Medical and Health Sciences	38
401 Agricultural, Forestry and Fisheries	6
404 Agricultural Biotechnology	1

4. Gender balance

A total of 283 PIs have received support and/or consultations, of which 117 female (41%) and 166 male (59%).

5. Publications

A total of 79 publications in 2022 have been published from our direct project support work (excluding Compute & Storage); DOIs provided in the table below.

10.1186/s13073-022-01144-5	10.1109/JBHI.2022.3177602	10.1186/s12864-022-08432-9
10.1038/s42003-022-04279-8	10.1016/j.chemosphere.2022.1347	10.3390/insects13050403
10.1002/hipo.23482	10.1371/journal.pcbi.1010366	10.1038/s41467-022-29608-7
10.1016/j.nbt.2022.06.005	10.1002/1873-3468.14451	10.1038/s42003-022-03309-9
10.3389/fgene.2022.1042483	10.1038/s41598-022-16077-7	10.1016/j.jchromb.2022.123200
10.7554/eLife.75749	10.1186/s13395-022-00299-4	10.1016/j.reprotox.2022.02.004
10.1007/s10278-022-00731-7	10.1002/ijc.33972	10.1111/jsr.13472
10.12688/f1000research.126734.1	10.1007/s11270-022-05709-3	10.1111/1755-0998.13614
10.1093/molbev/msac242	10.1002/ece3.9050	10.1038/s43705-022-00104-9
10.26508/lsa.202201530	10.3897/mycokeys.90.84866	10.1038/s41385-022-00497-9
10.1016/j.cub.2022.08.042	10.1073/pnas.2118852119	10.3892/or.2022.8269
10.1093/nar/gkac831	10.1016/j.celrep.2022.110982	10.1002/ece3.8676
10.3390/cancers14194553	10.1016/j.xcrm.2022.100657	10.1038/s41593-022-01011-x
10.1002/dvdy.399	10.1038/s41389-022-00409-3	10.4148/1941-4765.2177
10.1371/journal.pgen.1010419	10.1101/2022.06.14.496119	10.1016/j.compbiomed.2022.105282
10.1111/acel.13684	10.1186/s12859-022-04757-0	10.1002/1878-0261.13158
10.1002/path.5981	10.3389/fimmu.2022.874490	10.1371/journal.pgen.1010067
10.1126/sciadv.abo1543	10.1016/j.isci.2022.104303	10.1111/mec.16291
10.1016/j.isci.2022.104826	10.1038/s41598-022-11826-0	10.3389/fphys.2022.832417
10.1016/j.isci.2022.104850	10.1038/s41598-022-11341-2	10.3390/ma15030708
10.48550/arXiv.2112.05760	10.1038/s41467-022-30208-8	10.1182/bloodadvances.2021004962
10.1038/s41467-022-32395-w	10.3390/ijms23095098	10.1093/molbev/msab349
10.1101/gr.276665.122	10.1091/mbc.E21-10-0500	10.1016/j.gim.2021.09.003
10.1098/rspb.2022.0971	10.1371/journal.ppat.1010494	10.1016/j.annonc.2021.09.007
10.1164/rccm.202203-0463LE	10.1016/j.soilbio.2022.108650	10.1007/978-3-030-76394-7_2
10.1098/rspb.2022.1115	10.1002/ueg2.12227	
10.1093/rheumatology/keab912	10.1111/mec.16484	

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 2022, a total of 56 applicants of which 17 were granted LTS support (30%). Female: 26 applicants of which 5 were granted (19%). Male: 30 applicants of which 12 were granted (40%). NBIS has observed that the gender distribution in 2021 and 2022 was skewed, and we have tried to investigate explanations in order to avoid any future unfair bias, but so far without any clear-cut conclusion on the reason for the skew.