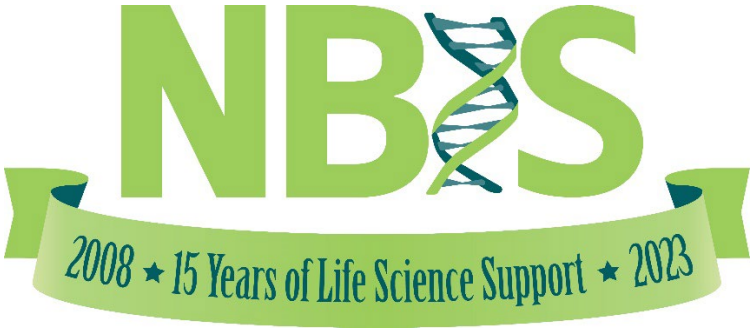


Annual Report 2023

Table of Contents

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 EU, ELIXIR, NordForsk (NeIC), and Chan Zuckerberg Initiative (CZI).

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, epigenomics, metagenomics, systems biology, single-cell/spatial biology, structural biology, biostatistics, machine learning 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 2023, NBIS provided support to **265 PIs** (111 female, 154 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 55 study design consultations. These numbers are at about the same levels as 2021 and 2022.

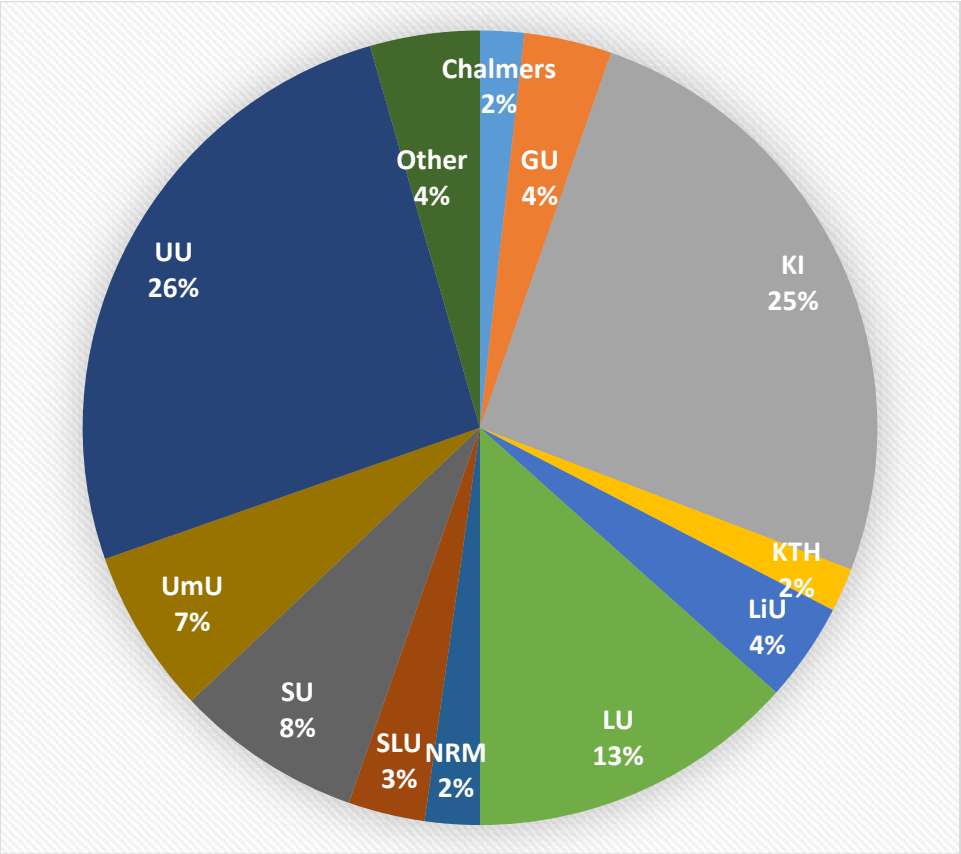


Figure 1. Distribution of PIs that have received support during 2023

In 2023, our user fee support track had 160 projects and our peer review support track had 88 projects. In addition, NBIS was engaged in 7 partner projects, and the BioImage Informatics Facility provided support to 21 projects. In addition, the AIDA Data Hub served 47 users related to data accessibility including ethical and legal consultations. 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.

During 2023, NBIS has expanded our support areas to also include Cryo-EM support. Structural biology is a constantly developing research field, e.g. with exciting opportunities in combining experimental data with rapidly advancing methods for protein structure prediction. Thanks to the DDLS programme, we are now able to broaden our services to include cryo-EM and cryo-ET data in a joint effort with the cryo-EM units and the Integrated Structural Biology platform. More information at:
https://www.scilifelab.se/news/scilifelab-enhances-support-for-advanced-data-analysis-services-in-cryo-em-applications/

Furthermore, our AIDA Data Hub unit offers advanced support in AI development, systems design, and sharing of sensitive data. During 2023, 60 AIDA-external data sharing events have

been supported, as well as 8 AI development projects, and a Barncancerfonden pre-study in system design of a multimodal data platform for childhood cancer.

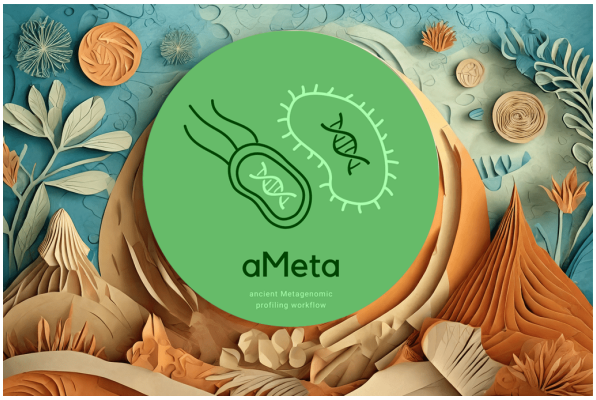
Examples of important achievements from NBIS support projects

NBIS has a total of **92 publications** in 2023 from our direct project support work (excluding Compute & Storage), listed in the Annex. A few achievements are shortly highlighted here.

Novel diagnostic test for metastatic prostate cancer in clinical practice. Together with the Lindberg group at KI and molecular pathologists Olauson and Haglund at Karolinska University Hospital, NBIS has developed novel diagnostics based on sequencing of circulating tumor DNA (ctDNA) enriched from plasma. As of May 2022, men with metastatic castration-resistant prostate cancer carrying a mutation in either the BRCA1 or BRCA2 genes could be treated with Olaparib. NBIS staff improved the sensitivity to detect loss of BRCA2, a key biomarker which has previously been particularly challenging. Initially part of the ProBio study (www.probiotrial.org), the new analysis is now offered at the Karolinska University Hospital.
https://www.scilifelab.se/news/precision-medicine-diagnostic-test-for-metastatic-prostate-cancer-transitions-from-research-to-application/

The Swedish childhood tumour biobank - to improve paediatric diagnostics and care. NBIS bioinformatics and data management staff have aided The Swedish Childhood Tumour Biobank (BTB), a national infrastructure for tissue samples and genomic data from paediatric patients with central nervous system and other solid tumours. A pilot study using blood-derived DNA and methylation profiling showed potential to enhance diagnostic accuracy, and identified alterations that may represent novel driver events. Data will be available at FEGA Sweden, and we expect BTB to impact the clinical care for children with cancer towards personalized treatments (https://doi.org/10.1186/s12967-023-04178-4).
https://www.scilifelab.se/news/barntumorbanken-reports-high-quality-genomic-research-data-generated-at-scilifelab-in-a-new-breakthrough-publication/

aMeta - A microbial lens to our past environment, lifestyle, and disease. Analysis of microbial data from archaeological samples is a rapidly growing field with great potential for understanding ancient environments, lifestyles, and diseases. However, high error rates have been a challenge in ancient metagenomics, and accurate and scalable methods have been lacking. With the Centre for Paleogenetics, NBIS has released aMeta, a metagenomics profiling workflow for ancient DNA with superior performance in microbial detection and authentication at low memory usage (https://doi.org/10.1186/s13059-023-03083-9).



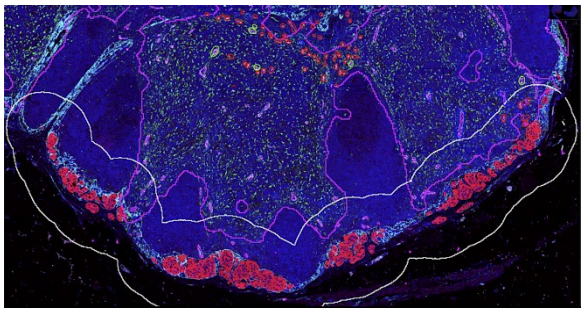
https://www.scilifelab.se/news/ameta-metagenomic-profiling-workflow-tool-for-adna/

BeyondFold – A new leap for integration of data and AI predictions in structural biology. Structural biology has been revolutionised by accurate protein structure prediction methods like AlphaFold2 and RoseTTAFold. BeyondFold has fused the strengths of SciLifeLab structure biology research groups, the ISB platform, AI competence at NBIS, and computational power at BerzeLiUs, opening new frontiers to tackle large protein complexes. The project has so far

contributed three novel methods, one workflow, and new training material, with a major manuscript under review in Nature Methods
(preprint: https://doi.org/10.1101/2023.09.20.558579).

Molecular and spatial landmarks of early mouse skin development. In a study with the Kasper group at KI using single-cell transcriptomics, spatial omics, and cell-lineage tracing, we deconstructed early embryonic mouse skin during the key transitions from seemingly uniform developmental precursor states to a multi-layered, multilineage epithelium, and complex dermal identity. The study uncovered the emergence of various cell types involved in hair production, protection, sensation, and temperature regulation, providing a comprehensive blueprint of early skin development, with a rich portal (https://kasperlab.org/embryonicskin) for data exploration and data re-use (https://doi.org/10.1016/j.devcel.2023.07.015).

New AI deep learning tool detects changes in blood vessels in cancer. In collaboration with the BIIF unit, the Ulvmar group from Uppsala University have created HEV-finder, a deep learning AI tool capable of automated detection of morphological changes in blood vessels caused by invasive breast cancer. Vascular remodelling often occurs in the tumour draining lymph nodes and potentially both contribute to tumour spread (metastasis) into the TDLNs and tumour-induced suppression of the anti-tumour immune response. To foster further development of the tool it is publicly available as an open-source resource and all raw data will be shared.



https://www.scilifelab.se/news/new-ai-deep-learning-tool-detects-changes-in-breast-cancer-blood-vessels/

Biodiversity

NBIS is heavily involved in biodiversity-focused projects, both on the Swedish and European level. High quality reference genomes are produced to enable research otherwise restricted to model organisms. Population genomics analyses support sustainable use of commercial fish species and also facilitate studies on threatened species, with results being fed directly into ongoing conservation efforts in Sweden. Samples from biodiversity hotspots in the Mediterranean area are also analysed by our support team in collaborative efforts with researchers in Spain and Slovenia in the Biodiversity Genomics Europe project.

ERGA - Addressing the biodiversity crisis in Europe. The availability of reference genomes is still limiting research and monitoring for many species. NBIS is collaborating in both national and international projects (European Reference Genome Atlas, ERGA) aiming to rectify this by delivering genomes of the highest possible quality. The genomes are used for example in conservation efforts like the European green toad (Grönfläckig padda) and for sustainable use of wild populations of commercial species like the European cisco (siklöja), the source of a highly praised caviar (löjrom).

The Tasmanian tiger – The first transcriptional study of an extinct animal. A study on the transcriptional profile of the extinct Tasmanian tiger by the Friedländer and Dalén groups received a lot of attention both in national and international press, as the first example of sequenced historical RNA recovered from an extinct animal species. NBIS staff contributed with

metatranscriptomic analyses designed to authenticate the RNA data from the Tasmanian tiger and explore sources of potential contamination from other organisms (https://doi.org/10.1101/gr.277663.123).

User Fees

NBIS has five support tracks:

- Study Design Consultation (≤ 3h)
- User Fee Support
- Partner Projects (PP)
- Peer Review Support
- BioImage Informatics (BIIF)

For User Fee Support, PP and BIIF, NBIS charges user fees, while Peer Review Support is provided for free according to the funding requirements by the Knut and Alice Wallenberg foundation. In 2023, the income from user fees was 8.9 MSEK, thus a little lower than 2022.

Data Management and Open Science Support

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. A large focus has been put on submission of genome sequence data to support the international evolution and biodiversity efforts that NBIS and NGI are engaged in. 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 continued to put a large effort into strengthening an operational user helpdesk for the Swedish node of the Federated EGA to enable a straightforward and efficient submission process for researchers. This work has been in the form of major contributions to the international efforts and projects on human data where NBIS is engaged, such as the European Genomic Data Infrastructure (GDI) project. The team has also been active in contributing to the European Joint Program on Rare Diseases (EJP-RD) on a demonstrator of federated analysis for genomic and phenotypic data.
On the training side, we have continued to provide the course on Introduction to Data Management practices, as well as running a tailored edition of this course for the SNP&SEQ unit at NGI. A large number of outreach events have also been performed, nationally and internationally, among them at the SciLifeLab days at the new sites in Umeå, Linköping, Göteborg, and Lund. Together with the SciLifeLab Data Centre we have continued to develop and improve the joint support portal and guideline knowledge-base for research data management for Swedish life science (https://data-guidelines.scilifelab.se/).

In the European landscape, NBIS has led the establishment of an ELIXIR Research Data Management Community, which functions as a pan-European competence network for research data management in the life sciences with over 200 RDM professionals in over 20 participating national nodes. This competence community has a focus on knowledge sharing, capacity building,

and training for data management. The data management team has also been active in other international efforts and organisations such as the EOSC and the Research Data Alliance (RDA). Notably, the team has chaired EOSC Association's task force on Semantic Interoperability, chaired the RDA's Life Science Data Infrastructures Interest Group, and organised a series of international workshops bridging European life science infrastructures to global and cross-disciplinary research data communities. Additionally, a member of the team joined RDA's Technical Advisory Board to contribute expertise and advice to the technical direction of the organisation.

Human Data

NBIS contributes to Swedish and major European efforts for handling sensitive human data, and has since long been engaged in building these infrastructures.

Late 2023, NBIS submitted the first synthetic datasets to the **Federated European Genome-phenome Archive (FEGA)** and expect to submit the first pilot case datasets early 2024, https://fega.nbis.se/. 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, through the NeIC Tryggve and NeIC Heilsa Tryggvedottir projects.

To support the **European 1+ Million Genomes Initiative (1+MG)**, EU funding has been obtained – 2020–2023 with the B1MG coordination and support action for writing recommendations on standards and methods and Nov 2022 – Jan 2027 with the subsequent implementation project **GDI (Genomic Data Infrastructure)**, having a total budget of 40 MEUR.

A major part in the GDI project is construction of the human genomic data infrastructure, led by NBIS as ELIXIR-SE together with ELIXIR-FI, building on experience from B1MG and FEGA. During 2023, NBIS has organised two onsite technical workshops for defining the 10 products needed for the infrastructure and later sharing expertise of the infrastructure and plan for the next improvements. The workshops engaged about 60 and 90 European project members, respectively.

Sweden is one of the "vanguard nodes" in GDI. In June 2023, NBIS and four other partners demonstrated the ability to run the infrastructure in its current state with synthetic data, from discovery of data to retrieval and analysis.

The AIDA Data Hub Sensitive Data Services (SDS) provides sharing and GPU processing of medical imaging data for the

AIDA and SCAPIS communities. Preparations have been made for a 15 MSEK v2.0 extension during 2024 using Bigpicture technologies, to serve further co-funding communities including CMIV, as well as EUCAIM and ASHA (cf below). SDS 2.0 introduces a service portfolio priced for sustainability while prioritizing scientific impact and incentivizing FAIR data sharing for Open Science.

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 based on the work on FEGA and GDI. The repository is in operation, based on ELIXIR technologies developed for genetic data. Service integration was demonstrated in Jan 2023, hands-on using live services, showcasing data submission, discovery, access request and approval, download, and on-platform data use. During 2023, the Bigpicture repository received its first non-clinical datasets, from large pharmacological companies in the European Federation of Pharmaceutical Industries and Associations (EFPIA)

EUCAIM is a 4-year 36 MEUR EU project started in 2023 in order to build a European cancer image federation similar to GDI. AIDA Data Hub contributes a collaboration platform on SDS 2.0, set up in collaboration with the SysDev unit. Furthermore our SysDev unit will also work on interoperability with GDI.

ASHA is a 4-year 73 MSEK VINNOVA systems demonstrator started in 2023 to construct interconnected spaces for primary and secondary use of standardized health data, engaging Swedish healthcare regions, EHR systems provider Cambio, and start-up company PredictMe. AIDA Data Hub provides a secondary use health data space on SDS 2.0, with a driver use case in PredictMe development of innovative epigenetic/multimodal AI tools for personalised medicine in long COVID and other systemic diseases. In a parallel 2 MSEK MedTech4Health funded focus area collaboration with East Sweden Medtech (ESMT), we develop business models for scaling out this type of use to further businesses and efforts.

AIDA Data Hub has implemented support for sharing datasets containing synthesised medical images, for example as supplementary data to support scientific publication of generative AI algorithms. The value of such data is not so much in the shared synthetic images themselves, but more in the capability of the AI algorithms to produce much more data of the same kind on demand, showcasing defining characteristics of images from a specific clinic or cohort, but where no synthetic image has a 1-to-1 correspondence to any specific real person. This would mean that the synthetic images would not constitute personal data in the GDPR sense. Such synthetic images could therefore be shared more easily than the original, very sensitive, real images. The vision is that if larger quantities of such synthetic images were made available, from more sources, then new ideas for research or AI development could be "tried out" or "roughed in" using only the more easily accessible synthetic data, and any prototype showing promise could then potentially be "fine-tuned" using real data.

Computational Infrastructure

A fundamental part of NBIS is the formation of a sustainable compute 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 NAISS 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 NAISS and SUNET.

Computational Resources

The AIDA Data Hub has prepared for a 15 MSEK v2.0 extension of our Sensitive Data Services (SDS) during 2024 based on Bigpicture / GDI technologies such as SDA, REMS, Beacon network, and Life Science Login. These technologies, originally developed for genetics and extended in Bigpicture to support also pathology, will thereby be further extended to also support use cases in radiology using for example Grand-Challenge.org software components, and open standards health data using OpenEHR (ASHA). Services are currently tailored to suit the needs of leading-edge national expert AI researchers. However, SDS 2.0 plans include feature additions to make the service more usable by further user competence profiles (such as clinicians) and other usage patterns (such as through a locked-down EHDS conformant remote desktop, or project private SaaS web applications like an open-source or Sectra PACS).

Computational Support

High throughput biomedical science depends on high-performance computers for bioinformatics analysis. The hardware is maintained by NAISS but NBIS provides expertise needed for domain scientists to efficiently access and use the computational and storage resources. Notably, over 1300 bioinformatics-related software packages are installed and maintained on NAISS compute clusters for life sciences at UPPMAX (Uppsala University), Rackham and, for sensitive data, Bianca. Support for Computational Resources also administers resource allocations and participates in help-desk support at the major national compute clusters, handling 2177 omics research projects with 825 unique PIs in 2023. These projects have used a total of close to 9 million core hours each month and over 12 PB of storage.

During 2023, NBIS continued the support that was previously provided by SNIC, as NAISS' support organisation was not yet in place. In the autumn, an activity was started in collaboration with UPPMAX and PDC (Paralleldatorcentrum at Royal Institute of Technology) to move the software environment and NAISS projects from Rackham at UPPMAX to Dardel at PDC. The aim of this activity is to relocate the user community to a new system with minimal interruption before Rackham reaches end-of-life at the end of this year.

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. Below we list a number of resources developed by NBIS.

PLUPP – A new pollen prognosis system. PLUPP is a collaboration between NBIS and the Swedish Museum of Natural History (NRM), to offer a single, comprehensive platform for pollen data. PLUPP advances the field of aerobiology and public health by a new open API (https://api.pollenrapporten.se/docs), representing a major upgrade by integrating advanced pollen forecasting functionalities into the NRM's existing PollenDB system. The new platform increases efficiency at NRM, but also enables developers to seamlessly incorporate real-time pollen forecasts and counts into a wide range of applications, benefiting many users

Chemical Space – scaling up the search space in early drug discovery. Jens Carlsson's group at UU in collaboration with NBIS and CBCS have developed the "Chemical Space Navigator", an online tool to speed up and improve drug discovery. The "Chemical Space Navigator" aids the discovery of chemical compounds by a radical scale-up of the search space for guiding hit identification and optimization. In June 2023, the first version was released, enabling CBCS' users to search for analogues of the CBCS primary screening set (37,000 compounds) among billions of commercially available compounds. A second version for optimization of hits from fragment screening at MAXIV will be publicly released soon.

Metabolic Atlas is a web platform developed by NBIS in collaboration with SysBio at Chalmers university. It integrates open-source genome scale metabolic models (GEMs) for easy browsing and analysis. In 2023, despite a reduced amount of development activity, we have consistently maintained an excellent standard of service, prioritising the community's needs by continually improving the user experience. During 2023, Metabolic Atlas was visited over 31,500 times (an increase by ~10% compared to the previous year) by users from around 130 countries. Moreover, Metabolic Atlas is an ELIXIR-SE service. https://metabolicatlas.org/

NBIS is also developing a range of tools and workflows for other data types. For example, our workflows for metabolomics and metagenomics are openly available and have been applied in many projects (see https://www.nbis.se/about/resources/tools).

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.

The AIDA Data Hub has supported a Region Skåne led AIDA project aiming at constructing a database facilitating AI research and innovation in mammography. This collaboration served as one of the cornerstones for the AIDA led and VINNOVA funded incubator for validation environments for clinical imaging diagnostic AI (VAI), where AI providers can install tools that national healthcare providers can then evaluate in private using real clinical data. The VAI-B mammography environment is in production, and working to transition to sustainable operations mode, potentially as a national quality register.

AIDA Data Hub has also supported a Region Skåne led AIDA project aiming to develop and evaluate an AI algorithm for automatic identification of gold fiducial markers in an MRI-only

prostate radiotherapy workflow. In an extension to this project, Region Skåne transitioned to an MRI-only workflow in its clinical practice for prostate radiotherapy, obviating the need for complementary CT-scans and reducing the time and complexity of the clinical workflow.

AIDA Data Hub has during 2023 provided secure storage and GPU processing to a LiU-led AIDA project aiming to develop AI algorithms for automatic time-resolved cardiovascular segmentation of 4d flow MRI. This project is ongoing, and is currently receiving AI development support from the AIDA Data Hub, where initial results using 2D Generative Adversarial Networks are promising and have been submitted to ISMRM 2024.

Advanced Training

Advanced training is one of the pillars of NBIS, and our training programme 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 programme as well as collaborative NBIS training, e.g. supporting the KAW-funded DDLS programme 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 infrastructures.

In 2023, most of the NBIS training events were delivered in an on-site format, while a few were online. Our training catalogue is tailored to graduate students, post-docs and researchers. This year we had two new courses: DDLS Population Genomics in Practice and DDLS Biomarker discovery. In total, **NBIS had 46 training events**, whereof 21 courses part of the NBIS training programme and 25 collaborative efforts, cf. table below. During 2023, **NBIS had 1103 participants** including the collaborative efforts. For the NBIS training programme we had 565 participants and the acceptance rate was 78%. The gender distribution is close to equal with 52% females accepted.
During 2023, NBIS co-organised an EMBO Practical training course in Advanced methods in bioimage analysis reaching beyond the European life science community as a more targeted international user community was approached.

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	Location	Participants
Courses organised by NBIS				
Single Cell RNAseq data analysis	2023-01-30	2023-02-03	Stockholm	26
Introduction to bioinformatics using NGS data	2023-02-06	2023-02-10	Uppsala, Lund, Umeå	45
Omics integration and Systems Biology	2023-02-08	2023-02-10	online	95
RNAseq data analysis	2023-03-13	2023-03-16	Uppsala	22
Neural Nets and Deep Learning	2023-03-20	2023-03-24	Uppsala	25

Workshop on Data Visualization in R	2023-04-17	2023-04-19	Lund	19
Introduction to Data Management Practices	2023-04-18	2023-04-20	Uppsala	17
Introduction to Biostatistics and Machine Learning	2023-04-24	2023-04-28	Uppsala	25
ELIXIR Tools for Reproducible Research	2023-04-24	2023-04-28	online	27
Programming formalisms	2023-05-29	2023-06-02	Uppsala	20
RaukR, Advanced R for Bioinformatics Summer School	2023-06-11	2023-06-22	Visby	39
Introduction to Data Management Practices for NGI	2023-09-07	2023-09-19	Uppsala	18
ELIXIR-GOBLET Train-the-Trainer	2023-09-11	2023-09-14	Uppsala	18
Epigenomics Data Analysis: from Bulk to Single Cell	2023-09-18	2023-09-21	online	25
Python programming with applications to bioinformatics	2023-10-09	2023-10-13	Uppsala, Umeå	28
R Foundations for Life Scientists	2023-10-23	2023-10-27	Uppsala	19
DDLS Population Genomics in practice	2023-11-05	2023-11-10	Uppsala	19
Introduction to bioinformatics using NGS data	2023-11-13	2023-11-17	Uppsala	26
Tools for Reproducible research	2023-11-20	2023-11-24	Stockholm	12
Programming Formalisms	2023-11-27	2023-12-01	online	20
DDLS Biomarker Discovery	2023-12-06	2023-12-08	Stockholm	20
Courses co-organised by NBIS				
UPPMAX Introductory Course	2023-01-09	2023-01-13	Uppsala	20
To awk or not	2023-01-16	2023-01-18	Uppsala	12
Workshop on QuPath	2023-02-08	2023-02-09	Uppsala	15
LCI microscopy course - From sample preparation to image analysis	2023-02-16	2023-02-16	online/Stockholm	25
Introduction to Bianca: Handling Sensitive Research Data	2023-03-14	2023-03-14	Uppsala	5
Clinical Research Clinical School in Molecular Medicine 2023	2023-03-20	2023-03-20	Stockholm	25
Cell segmentation, Classification and Spatial Analysis in Multiplexed Fluorescence Microscopy Image Data	2023-04-02	2023-04-02	Porto, Portugal	30
Digital Image Analysis for Scientific Applications, DIASA	2023-04-11	2023-06-29	Uppsala	25
AI tools Biology	2023-05-04	2023-05-04	Uppsala	15
CellProfiler on the cloud	2023-05-09	2023-05-09	Porto, Portugal	25
AI course for clinicians	2023-05-16	2023-05-18	Linköping	25
Cell segmentation, Classification and Spatial Analysis in Multiplexed Fluorescence Microscopy Image Data	2023-05-30	2023-05-31	Lund	25
Introduction to QuPath	2023-06-07	2023-06-07	Uppsala	7
Introduction to Bianca: Handling Sensitive Research Data	2023-06-14	2023-06-14	Uppsala	8
Ethical, Legal and Social implications in Research Infrastructures and Core Facilities	2023-06-12	2023-06-22	online	25
EMBL Practical Course: Imaging-based spatial-omics	2023-06-25	2023-06-30	Rome, Italy	25
UPPMAX Introductory Course	2023-08-15	2023-08-19	Uppsala	18
Image Analysis and Data Processing in Super-Resolution Microscopy 2023	2023-08-23	2023-08-23	online/Prague	30

To awk or not	2023-08-29	2023-08-30	Uppsala	12
Introductory course to Image Analysis in Life Science	2023-09-04	2023-09-04	Göteborg	30
Image Analysis, Bioimaging and Cell Analysis, 3MG036	2023-09-05	2023-09-07	Uppsala	25
EMBO Practical Course: Advanced methods in bioimage analysis	2023-09-10	2023-09-15	online/EMBL Heidelberg, Germany	26
1MB438 Introduktion till bioinformatik	2023-09-02	2023-10-25	Uppsala	35
Swedish Bioinformatics Workshop 2023	2023-11-07	2023-11-07	Stockholm	25
AI course for clinicians	2023-11-27	2023-11-29	Linköping	25
Internal courses				
NBIS internal workshop: Population Genomics in practice	2023-08-28	2023-08-30	online	9
NBIS internal workshop: Spatial transcriptomics course sessions	2023-09-24	2023-11-20	online	23

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, during 2023 internal courses was organised in Population Genomics and Spatial Transciptomics.

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.

SciLifeLab Training Hub was established in 2023, which is in essence a scale-up of the NBIS Training support, tools and resources to be offered to the wider SciLifeLab Infrastructure ecosystem and beyond. Much of what the Training Hub at SciLifeLab will offer and provide will be built collaboratively with the NBIS Training team. Also, Training Hub will make use of the open source and open educational resources established by ELIXIR, where NBIS plays a significant role.

For the ELIXIR Training Platform Work Programme 2022–23, NBIS Training members co-led several Tasks and activities involving training certification, establishing the training collections and learning paths framework and FAIR-training. Furthermore, NBIS has continued to be active and engaged contributors to the Train-the-Trainer instructor network.

NBIS has during 2023 continued to strengthen the connection to the Open Science scene in Sweden and with EOSC in particular. In 2023, NBIS was invited to two Swedish strategic events on EOSC to represent and showcase Open Science from a Research Infrastructure perspective (Open Science-from policy to practice and National EOSC event med SUHF and VR). NBIS also continued the involvement in EOSC Association Taskforces: "Upskilling of countries to engage in EOSC" (Jessica Lindvall), "Semantic Interoperability" (Wolmar Nyberg Åkerström), "AAI architecture" (Jonas Söderberg), "Technical Interoperability" (Lars Eklund). In addition, NBIS co-delivered a course module within the RITrainPLUS programme on Open Science. The target audience was Research Infrastructure staff.

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 2023, we arranged 39 national drop-ins via Zoom.

In 2023 the NBIS website was relaunched following a significant amount of work from our systems development group. The new site is clearer and more well organized, improving the experience significantly for visitors and making our organisation more accessible.

NBIS has a community coordinator to oversee and plan NBIS outreach activities and so far two local site coordinators. In 2023 NBIS celebrated 15-year anniversary with a symposium presenting many of our achievements, current capabilities and plans for the future.

The Swedish Bioinformatics Workshop (SBW) 2023 took place in Stockholm and was sponsored in part by NBIS. The appreciated recurring event mixes renowned keynote speakers with workshop sessions, two of which were arranged by NBIS, and poster sessions. NBIS was also in the Scientific Advisory Board of the event.

The NBIS Artificial Intelligence and Integrative Omics seminar series, AI&IO, now an open seminar series (internal until 2022), held several events with both NBIS and external speakers. Notably two DDLS fellows presented.

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 also contributed to the establishment of a DDLS Data Science Node at Chalmers.

In **Linköping** a new bioinformatics gathering, dubbed "Bioinformatics Bash" was initiated where NBIS was presented. The event is planned to be held twice annually. NBIS is also in the infrastructure group of Linköping SciLifeLab site which is organising a series of seminars focused on local platforms.

In **Umeå**, NBIS has been present and presented at several events, the KGC days, KBC infrastructure seminar series, Research Infrastructure day and UCMR day.

Finally, NBIS has internal Outreach group meetings which maintain a github repository of outreach materials and presentations as well as discuss 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.

During 2023, NBIS has provided support to Key2Brain's development of antibody fragments for brain transport of therapeutics. Our bioinformatics expertise was used to analyse the first set of results from the company's human antibody therapeutics unit.
https://www.scilifelab.se/news/industry-case-scilifelab-infrastructure-a-prerequisite-for-key2brains-development-of-antibody-fragments-for-brain-transport-of-therapeutics/

In 2023, our AIDA Data Hub unit has provided industry support to Spearpoint Analytics for development of algorithms to improve prostate cancer diagnostics. The start-up company Spearpoint Analytics has used AI compute and data sharing services available at AIDA Data Hub to develop an outcome-based analysis (PCAI – Prostate Cancer Aggressiveness Index) that predicts the severity of prostate cancer by analyzing microscopy images of tissue sections. The AIDA Data Hub DGX-2 is a resource provided as a service for AI and training on sensitive medical image data. True FAIR and open science spirit have permeated the project – principles that are crucial for transparency and moving the research field forward faster.

International

NBIS is the Swedish node in the European infrastructure ELIXIR. NBIS also aims at strengthening the Nordic collaborations, and since 2011, we have regular meetings between the Nordic ELIXIR Heads of Nodes. Since 2013, we have funding from NeIC (Nordic eInfrastructure Collaboration) for development and provision of infrastructure for sensitive data (current project NeIC Heilsa Tryggvedottir). On the bioimage analysis side, NBIS is involved in Euro-Bioimaging, NEUBIAS, and Chan-Zuckerberg image analysis initiatives.

ELIXIR is unique in Europe encompassing all European national bioinformatics infrastructures into a coordinated distributed research infrastructure with currently 23 partners and more expected to join.

During 2023, 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. Since 2022, ELIXIR-SE 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.

Furthermore, NBIS as ELIXIR-SE has been active in multiple EU projects (as described above), e.g. GDI (Genomic Data Infrastructure), Bigpicture for digital pathology, EUCAIM providing infrastructure for cancer image data, and PHENET (where NBIS provides training on the European level).

During 2023, two EU projects were finalised – ELIXIR-CONVERGE and B1MG. The B1MG activities continue in GDI and 1+MG, and our data management activities continue in the form of an ELIXIR Community for Research Data Management, that is co-lead by ELIXIR-SE.

Staff

The table shows the number of staff in FTE (full time equivalents) during 2023 for the different functions in NBIS. Gender balance is 69% male and 31% female. Number of FTEs has increased from 91.4 in 2022 to 94.3 in 2023.

NBIS	Staff (FTE)
1 Central functions, incl. management	7.1
2A Support (SMS)	22.6
2B Support (LTS)	15.1
3A Data management	3.6
3B Human data	11.3
4A Systems development	8.5
4B Pipelines & Tools	2,1
5 Support for Computational Resources	3.9
6 Training	11.5
7 ELIXIR	3.0
8A BioImage Informatics	3.3
8B AIDA Data Hub	2.3
TOTAL	**94.3**

Economy report for 2023

NBIS	2023	Result 2023	Budget 2023	Result 2022
Incomes	VR Infrastructure grant	20 000 000	20 000 000	20 000 000
	VR grant EGA-SE	2 093 946	1 404 000	1 401 270
	VR grant Biodiversity	2 368 913	1 000 000	
	SciLifeLab National	28 850 000	28 850 000	26 550 000
	Universities + SciLifeLab SFO	19 273 000	19 273 000	16 558 000
	KAW	26 498 748	26 500 000	21 996 109
	Vinnova (incl. Co-funding EU-proj.)	2 657 710	3 200 000	
	EU + ELIXIR	12 190 014	11 881 000	4 092 853
	NeIC NordForsk	2 011 395	1 550 000	1 892 035
	Chan-Zuckerberg	1 000 000	1 000 000	1 000 000
	User fees	8 923 588	11 100 000	9 500 000
	Other		4 867 000	2 040 601
	SUM Incomes	**125 867 313**	**130 625 000**	**105 030 868**
Expenses	Personnel	81 707 508	96 759 259	71 585 049
	Equipment	3 567 277		3 416 967
	Travel	3 358 247		1 904 437
	Other costs	3 569 889		2 486 853
	Office space	5 948 944	33 865 741	5 149 211
	Indirect costs	25 420 555		20 488 351
	SUM Expenses	**123 572 420**	**130 625 000**	**105 030 868**

Distribution of costs on the different NBIS activities

Activity	kSEK	Budget
1 - Central functions[1]	9 454	8 750
2A - Support SMS (incl. Partner Projects)	29 250	33 327[2]
2B - Support LTS	19 938	28 350[2]
2C - Support Sysbio	included in 2B	3 400[2]
3A - Data management	4 869[3]	8 660[2]
3B – Human data	14 207[4]	
4A - Systems development	10 717	17 620[2]
4B - Pipelines & tools	2 768	
5 - Compute & Storage	4 505	6 870[2]
6 - Training & Nat networking	14 431	[2]
7 – ELIXIR + EU projects	5 918	10 264
8A - BioImage Informatics	4 239	5 850
8B - AIDA Data Hub	2 758	7 534
9 - Other	519	
SUM	**123 572**	**130 625**
SUM (excluding new modules 8A 8B)	**116 575**	**117 241**

[1]Central functions also include project management for Support
[2]Training costs are included in budget of modules 2–5
[3]Several activities in data management are filed under 6 Training and 7 ELIXIR
[4]Human data activities have been budgeted in modules 3, 4, 7 and 8B

Annex – Key performance indicators

1. Number of projects

During 2023, we have worked on 276 support projects and provided 55 consultations.
The 276 support projects involved 224 unique PIs (94 female, 130 male). In addition, the AIDA Data Hub served 47 users related to data accessibility including ethical and legal consultations. Altogether, we have supported 286 PIs.

2. Number of PIs distributed on universities

Univ	# Unique PI excl. consultations
Chalmers	4
GU	8
KI	57
KTH	4
LiU	9
LU	30
NRM	5
SLU	7
SU	17
UmU	15
UU	58
Other	10

3. Number of projects distributed on SCB codes

SCB code and subject	
102 Computer and Information Sciences	2
105 Earth and Related Environmental Sciences	1
106 Biological Sciences	108
107 Other Natural Sciences	2
209 Industrial Biotechnology	3
211 Other Engineering and Technologies	1
301 Basic Medicine	68
302 Clinical Medicine	51
303 Health Sciences	14
304 Medical Biotechnology	13
305 Other Medical and Health Sciences	6
401 Agricultural, Forestry and Fisheries	4
403 Veterinary Science	2
404 Agricultural Biotechnology	1

4. Gender balance

A total of 265 PIs have received support and/or consultations, of which 111 female (42%) and 154 male (58%).

5. Publications

A total of 92 publications in 2023 have been published from our direct project support work (excluding Compute & Storage); DOIs provided in the table below.

10.1186/s12864-023-09889-y	10.3389/fimmu.2023.1240597	10.1084/jem.20220741
10.1007/s10162-023-00920-3	10.1038/s41408-023-00914-7	10.1016/j.soilbio.2023.108994
10.1038/s41597-023-02800-0	10.1093/rheumatology/kead152	10.1186/s40168-023-01502-4
10.1021/acs.jproteome.3c00636	10.1126/science.adg0995	10.3389/fimmu.2023.1151754
10.1002/prot.26598	10.1002/art.42683	10.1186/s12864-023-09250-3
10.1002/prot.26593	10.1186/s12896-023-00804-7	10.1016/j.devcel.2023.02.007
10.3389/fpubh.2023.1289945	10.1007/s13402-023-00863-0	10.3354/meps14268
10.1111/tpj.16535	10.1016/j.devcel.2023.07.015	10.1186/s13058-023-01631-6
10.1093/nar/gkad1010	10.1016/j.ecoenv.2023.115321	10.1002/hbm.26165
10.1084/jem.20230101	10.12688/f1000research.134798.1	10.1007/s00018-023-04714-x
10.1016/j.scitotenv.2023.168159	10.1038/s41592-023-01944-6	10.1073/pnas.2217868120
10.1371/journal.pcbi.1011498	10.1101/gr.277663.123	10.1186/s12915-022-01509-7
10.1186/s13059-023-03083-9	10.3324/haematol.2023.282965	10.1093/evolut/qpac049
10.1242/bio.060113	10.3390/nu15143273	10.1038/s42003-023-04526-6
10.1158/1078-0432.CCR-23-0178	10.1186/s40635-023-00528-0	10.1111/mec.16799
10.1038/s41467-023-41919-x	10.1186/s12944-023-01863-7	10.1016/j.tox.2023.153425
10.1158/1078-0432.CCR-23-0538	10.1038/s41467-023-39421-5	10.1016/j.envint.2023.107746
10.1016/j.immuni.2023.08.011	10.7554/eLife.84874	10.1002/jbio.202200227
10.1038/s41596-023-00881-0	10.1016/j.isci.2023.106906	10.1016/j.dib.2022.108769
10.1186/s12967-023-04534-4	10.1016/j.ccell.2023.04.010	10.1038/s41556-022-01064-x
10.3389/fpls.2023.1210850	10.1186/s12859-023-05320-1	10.1016/j.isci.2022.105857
10.1093/genetics/iyad179	10.1371/journal.pgen.1010801	10.1186/s40478-022-01500-x
10.1111/tpj.16269	10.1136/ard-2022-223633	10.3389/fimmu.2022.1087843
10.1016/j.dci.2023.104760	10.1186/s12967-023-04178-4	10.3390/healthcare11020184
10.1111/1755-0998.13838	10.1016/j.xcrm.2023.101038	10.1093/nar/gkac831
10.1093/g3journal/jkad192	10.1038/s41598-023-34107-w	10.1186/s12859-022-05120-z
10.1016/j.ebiom.2023.104813	10.1093/molbev/msad095	10.1093/gbe/evac177
10.3389/fmicb.2023.1253371	10.1016/j.heliyon.2023.e15306	10.1371/journal.pgen.1010599
10.1080/0284186X.2023.2257873	10.1038/s41593-023-01312-9	10.1002/jmri.28221
10.1080/0284186X.2023.2254480	10.1161/JAHA.122.029248	10.1017/s2633903x23000053
10.1038/s41598-023-42506-2	10.1038/s41467-023-37835-9	

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 2023, a total of 82 applicants of which 19 were granted LTS support (23%). Female: 38 applicants of which 9 were granted (24%). Male: 44 applicants of which 10 were granted (23%). Thus, there is no difference in success rate between female and male.