Development and organization

of the Bielefeld Institute for Bioinformatics Infrastructure (BIBI)

A. Pühler, A. Schönhuth, A. Sczyrba, J. Stoye, A. Tauch and R. Wittler

Bielefeld Institute for Bioinformatics Infrastructure, Bielefeld University, 33501 Bielefeld

Content

Mission Statement

The BIBI institute is responsible for the development and operation of a bioinformatics infrastructure in Germany. Both service and training aspects play a role within the framework of the German Network for Bioinformatics Infrastructure (de.NBI) and the German ELIXIR node (ELIXIR-DE). The institute is also involved in the establishment and operation of a bioinformatics cloud. In addition, research questions on microbial bioinformatics are dealt with at the BIBI institute. It also runs a graduate program for data scientists in the life sciences.

Summary

The Bielefeld Institute for Bioinformatics Infrastructure (BIBI) harbors the central parts of the German Network for Bioinformatics Infrastructure (de.NBI). The BIBI institute at Bielefeld University, together with ZB MED - Information Centre for Life Sciences in Cologne, strives to become a member of the Leibniz Association. The BIBI institute was founded on June 1, 2019 at the Faculty of Technology of Bielefeld University. It consists of six departments responsible for the management of the institute, for the coordination and administration of the de.NBI network and the ELIXIR-DE node, for the bioinformatics cloud, for the service and training in the field of microbial bioinformatics, and for educating young scientists in the DILS graduate school in the bioinformatics service area. Most of the staff positions in BIBI are financed either by the German Ministry of Education and Research (BMBF) (46%) via the de.NBI project or directly by Bielefeld University (45%). Only the staff positions in the department “Head of the Institute”/"Institute Management" (9%) are jointly funded by Bielefeld University and ZB MED - Information Center for Life Sciences.

1. Background of the Institute at the Faculty of Technology of Bielefeld University

The establishment of the Bielefeld Institute for Bioinformatics Infrastructure at the Faculty of Technology of Bielefeld University is closely linked to the German Network for Bioinformatics Infrastructure (de.NBI). The de.NBI network was launched as a project funded by the German Ministry of Education and Research (BMBF) in March 2015. It consists of 40 individual projects distributed throughout Germany, which are combined in eight service centers. The network is managed by a coordinator. The tasks of the network focus on the establishment of a bioinformatics infrastructure that is available to all life science researchers involved in the analysis of large amounts of data. To this end, the de.NBI network offers both service and training, which are in the hands of the individual de.NBI projects and service centers. For evaluating large amounts of data, a de.NBI-owned compute infrastructure was set up consisting of a federated cloud located at six university sites. For the connection of the de.NBI network to European infrastructure initiatives, Germany joined the ELIXIR organization and commissioned de.NBI members to take over service and training tasks of the national ELIXIR node, ELIXIR Germany. With the start of the de.NBI Industrial Forum in 2019, the range of tasks of the de.NBI network was further expanded. An overview of network

activities can be found in the "Handbook of the German Network for Bioinformatics Infrastructure" (https://www.denbi.de/images/Downloads/deNBI_Handbook_V2_2019.pdf). Further information on scientific highlights of the de.NBI network is provided in the brochure "From data analysis to understanding complex biological systems", which was published in February 2020 (https://www.denbi.de/images/Downloads/deNBI_highlight_brochure_German.pdf). In the meantime, the de.NBI network has completed its establishment, i.e. the bioinformatics infrastructure is active and is used monthly by hundreds of thousands of researchers from the life sciences to analyze experimental data sets. Following the establishment of the national bioinformatics infrastructure in Germany, it is now necessary to work for a continued existence secured by a stabilization step.

In order to initiate the sustainable continuation of the de.NBI network, the following activities have been developed in the past two years: Firstly, a meeting was held in February 2018 at the Federal Ministry of Education and Research (BMBF), which was dedicated to the question of the sustainability of de.NBI. The division of the network into a central and a decentral part was seen as an option. The central part should be integrated into an institute of the Leibniz Association and the decentral part should be supported by long-term BMBF funding. This option was considered feasible and the de.NBI coordinator was given the task of developing a concept for the sustainable continuation of the de.NBI network. The concept paper entitled "Verstetigungskonzept des de.NBI-Netzwerks durch Integration als ein Teilinstitut für Bioinformatik-Infrastruktur in das Leibniz-Zentrum ZB MED" was submitted to the BMBF in early May 2018. The concept paper mainly dealt with the consolidation of the central de.NBI part, which is intended to represent the core of an institute for bioinformatics infrastructure at Bielefeld University. The latter plans to be merged with ZB MED, which strives to be re-integrated into the Leibniz Association. The concept paper describes in detail the establishment of the institute located at the University - later called **Bielefeld Institute for Bioinformatics Infrastructure (BIBI)**. A total of six departments are planned in the institute. The institute should be headed by a **professor for "Service Science in the Life Sciences"**.

The approaches of the de.NBI network were subject of another BMBF meeting in May 2019. This meeting also reported on the establishment of an institute in Baden-Württemberg that will combine the de.NBI individual projects located at the universities in Freiburg, Heidelberg, Konstanz and Tübingen. This Institute for Bioinformatics Infrastructure (IBIS) thus includes projects of the de.NBI network that were previously assigned to the decentral part.

As a result of the BMBF discussion, three approaches were proposed to solve the de.NBI problem of a sustainable continuation:

1. **Small-sized solution**: BIBI should be integrated into the Leibniz Association together with the institute ZB MED.
2. **Medium-sized solution**: At a later stage, IBIS should be connected to the Leibniz Institute ZB MED/BIBI.
3. **Large-sized solution**: The entire de.NBI network including the de.NBI cloud is considered to be integrated into the Leibniz organization.

Currently, the small and the medium-sized solutions are considered as feasible. For the large-sized solution there is little chance of realization.

2. Foundation of the institute

The aim to establish the central part of the de.NBI network as an institute at Bielefeld University with the goal to be integrated together with ZB MED into the Leibniz Association, was pursued with vigour after the sustainability concept had been established. First, the sustainability concept was presented to the Scientific Advisory Board on May 30, 2018 and to the Board of Trustees of ZB MED on July 20, 2018. After both boards had approved the plans, Bielefeld University and ZB MED signed a cooperation agreement. This contract regulates in particular the joint call for a W3 professorship for Service Science in the Life Sciences at Bielefeld University, which was created to manage the BIBI institute. This W3 professorship was announced on 29 October 2018. Subsequently, at Bielefeld University the Administrative and User Regulations (VBO) for the BIBI institute were drawn up, which finally came into force on 1 June 2019. These VBO regulates all the legal details of the institute and was decisive in enabling the institute to start operating within a very short time. Thus, the founding board of directors met on 25th June 2019, followed by a general meeting and a foundation ceremony on 31st June 2019. Finally, on 9 October, four positions for doctoral students were filled in the graduate school "Digital Infrastructure for the Life Sciences (DILS)". A BIBI/ZB MED workshop was held on 24 October 2019 to prepare joint research projects of the participating institutions. In the meantime, the BIBI institute already has 40 positions that are financed by Bielefeld University, the BMBF and ZB MED. The chronological sequence of steps in the development of the BIBI institute can be seen in Fig. 1.

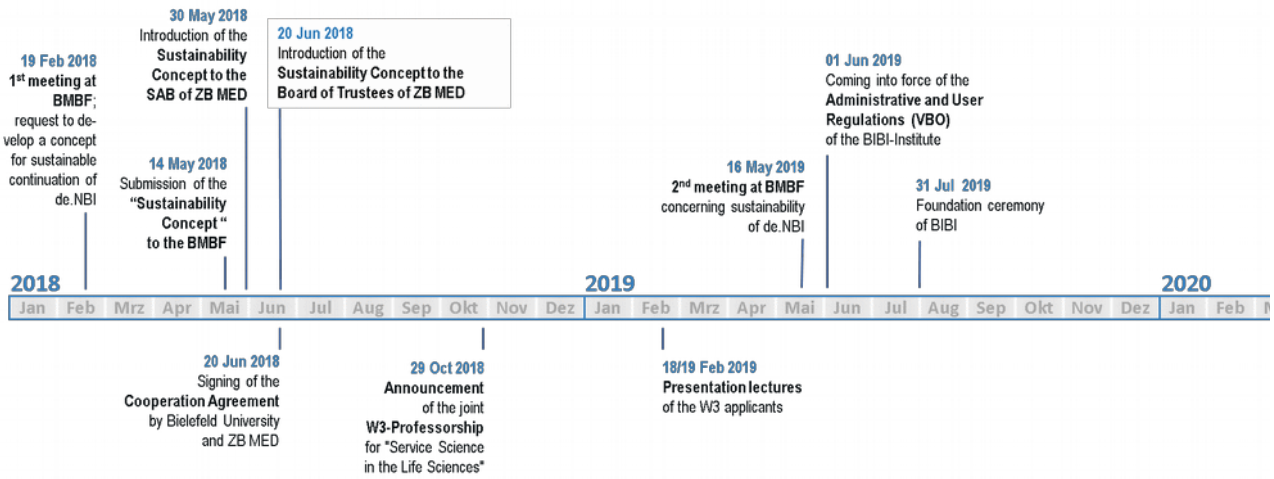

Fig. 1: Time sequence of steps in the development of the Bielefeld Institute for Bioinformatics Infrastructure (BIBI).

3. Structure of the Institute

The BIBI institute was founded to transfer central parts of the de.NBI network into a structure that can be integrated, together with ZB MED, into the Leibniz Association. As shown in Fig. 2, the institute consists of six departments focusing on the following tasks: **Institute Management** (Department 1), **de.NBI Coordination and Administration Office** (Department 2), **ELIXIR-DE Coordination and Administration Office** (Department 3), **Bioinformatics Cloud** (Department 4), **Microbial Bioinformatic**s (Department 5) and **Graduate School** (Department 6). The central components of the de.NBI network are the two coordination and administration offices for the de.NBI network and the German ELIXIR node. Further central components of the de.NBI network are the bioinformatics cloud at the Bielefeld University site and the graduate school. The two Departments of Microbial Bioinformatics and the Institute Management complement the BIBI institute's range of tasks in the field of bioinformatics services. The staff positions in the six departments of the institute are shown in Fig. 2.

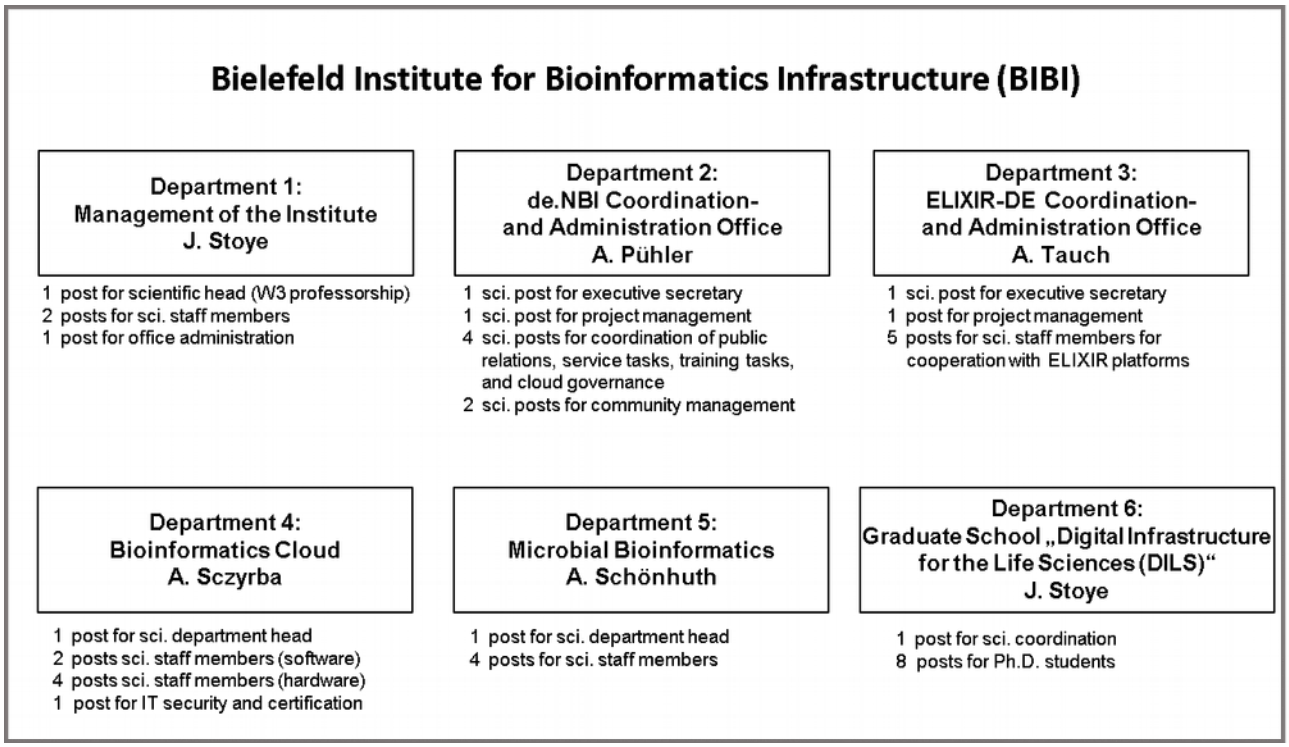

Fig. 2: Structure and personnel of the Bielefeld Institute for Bioinformatics Infrastructure.

4. Range of tasks of the BIBI departments

Figure 2 shows that the institute is divided into six departments with clearly separated areas of responsibility. In the following, the range of tasks of these departments will be presented in more detail.

4.1 Management of the BIBI Institute (Department 1)

The W3 professorship "Service Science in the Life Sciences", which is jointly funded by ZB MED and Bielefeld University through a cooperation agreement, is intended to head the institute in future. The joint W3 professorship was announced by Bielefeld University on 29.10.2018. Six applicants were invited for a presentation at Bielefeld University on 18-19 February 2019. The Appointments Committee drew up a list of candidates which the Rectorate of Bielefeld University will use to fill the advertised W3 professorship. The appointment procedure is not yet completed.

The W3 professorship, which is located in Department 1, is equipped with two positions for scientific staff and one position for an office administrator. The scientific staff members support the W3 professor in the field "Service Science in the Life Sciences".

4.2 The de.NBI Coordination and Administration Office (Department 2)

Department 2 of the institute houses the coordination and administration office of the de.NBI network, which was established in March 2015. This office was set up to plan and execute all business processes in the network from a central location. In particular, the meetings of the Central Coordination Unit (CCU) of the network have to be organized on a quarterly basis. The CCU is the sole decision-making body of the de.NBI network. The CCU body is made up of the heads of the eight service centers and the de.NBI coordinator. Furthermore, the de.NBI Coordination and Administration Office is responsible for the annual meeting of the Scientific Advisory Board and for the plenary assembly of the de.NBI network. In addition, the de.NBI Coordination and Administration Office reports to the BMBF on work processes and successes achieved in the de.NBI network and prepares the semi-annual major project controlling report (Großprojektkontrolling-Bericht) on the network.

The department is equipped with two scientific positions for the management and administration of de.NBI tasks. The scientific management of the de.NBI Coordination and Administration Office is responsible for the definition of work procedures in the department. It also works closely together with the coordinator, mainly discussing organizational issues of the de.NBI network. The scientific management position was established to realize the many different tasks and events of the de.NBI network.

In addition, four scientific positions are located in the department, which deal with the tasks of public relations, service, training and government of cloud use. The de.NBI public relations work is multi-layered and includes the maintenance of the website, the maintenance of the Twitter channel, the publication of the de.NBI quarterly newsletter and the production of information material. In the de.NBI service area, more than 100 software programs have to be tested for usability, quality and demand. In the de.NBI training area, more than 80 training courses per year have to be checked for content, quality and success. Both task areas of the de.NBI Coordination and Administration Office

act in coordination with the network partners and fulfil central monitoring tasks for service and training. The central administration of the de.NBI cloud activities includes the support of the user portal including authentication and authorization of users as well as the evaluation of submitted projects with regard to suitability and allocation of computing resources. Finally, two scientific posts have recently been made available for community management. One of these positions is used to supervise the members of the de.NBI Industrial Forum, the other position is used to determine - together with users of the de.NBI cloud - the deployment and demand planning of the cloud infrastructure.

4.3 The ELIXIR-DE Coordination and Administration Office (Department 3)

The establishment of the German ELIXIR node (ELIXIR Germany) was started in August 2016 after the ELIXIR Consortium Agreement was signed by the BMBF. After the formal accession of Germany to the ELIXIR organization, the ELIXIR Node Application of Germany that defines the basic structure of the German ELIXIR node, was submitted by the BMBF in February 2017. After an international peer review process, the ELIXIR Board accepted the application in April 2017. Subsequently, two further agreements were initiated to put Germany's full membership in the ELIXIR Consortium on a legal basis. The basic national cooperation agreement was finally signed by 21 de.NBI partners and came into force on 9 January 2020. Since January 2020, the BMBF funding of de.NBI also supports the separate ELIXIR-DE Coordination and Administration Office, which is to ensure the cooperation at the national level and with ELIXIR.

The funding of the ELIXIR-DE Coordination and Administration Office includes a position for the scientific management ("Node Coordinator") and a position for the project management of the ELIXIR node. The Node Coordinator supports the "Head of Node (HoN)" in his work task for the strategic development of the German node and is the central contact for the ELIXIR Hub to carry out international cooperation projects. The project management, on the other hand, is the central contact for the members of the German ELIXIR node in the context of the work tasks to establish the German ELIXIR node and to integrate the German ELIXIR members into the European consortium. In addition, five further positions for scientific staff are planned, which are thematically structured in analogy to the five ELIXIR platforms (Compute, Data, Interoperability, Tools, Training). With the new staff for platform management the scientific cooperation with ELIXIR shall be advanced. From the current ELIXIR program for the period 2019 to 2023 it is clear that ELIXIR activities in future will be based on the five platforms. By mirroring the platform topics, the manifold activities of ELIXIR Germany can be well structured thematically. The platform management staff is especially required to involve the de.NBI service centers and the de.NBI projects located there in the integration and implementation of ELIXIR activities. The platform management personnel thus forms a hinge function between ELIXIR Germany and the five ELIXIR platforms.

4.4 The Bioinformatics Cloud (Department 4)

When the de.NBI network was founded in March 2015, it was already clear that the provision of computing and storage resources was essential for a bioinformatics infrastructure. In July 2016, following the BMBF's funding commitment, it was therefore decided to establish the de.NBI cloud at the universities in Bielefeld, Freiburg, Gießen, Heidelberg and Tübingen. In the meantime, further installations have been put into operation at the DKFZ (Heidelberg) and Berlin Institute of Health (Charité Berlin). In addition to the investment funds for the establishment of the infrastructure, the BMBF has agreed to finance a total of six staff positions, which were increased by six more in 2019, with special tasks in the field of certification and IT security. These staff positions are distributed across the locations of the de.NBI cloud, with the universities and research organizations providing additional staff positions at the locations to operate the cloud.

The de.NBI cloud has been in the production phase since the beginning of 2018 and offers life scientists from German universities and research institutions free access to computing and storage resources, but also to reference data and relevant methods of the respective work area. The cloud supports projects from the entire field of life sciences in a trusted environment. The staff of the de.NBI Cloud offers the expertise for the operation and embedding of the respective projects in the Cloud, starting from small projects for the promotion of young researchers up to large international projects with German participation.

Department 4 of the BIBI institute harbors the Bielefeld site of the de.NBI Cloud. Table 1 shows its current expansion. It provides computing and storage resources with the Openstack virtualization solution. Besides "general purpose" hardware, the infrastructure also offers special hardware. For example, "high memory" machines are provided for special storage-intensive bioinformatics applications, e.g. (meta-) genome assemblies. Machine learning applications require graphics cards for their calculations. The cloud infrastructure allows bioinformatics analyses to be scaled both vertically and horizontally. In order to give analyses running in the cloud the fastest possible access to databases and reference data, 9.2 PB of storage capacity is available for the storage of databases and reference data.
The authentication and authorization of the cloud users takes place via ELIXIR AAI and is controlled via a portal that receives project applications from users and also advises them. The portal is located in the de.NBI Administration Office and is managed by the Cloud Governance.

Table 1: Expansion of the bioinformatics cloud at the Bielefeld site.

General Purpose	Mid Class	High Memory	GPUs	Storage
56 x (28 cores/384 GB) + 38 x (32 cores/512 GB)	12 x (56 cores/768 GB)	4 x (56 cores/3 TB)	10 x Nvidia (P100), 20 x Nvidia (V100), 4 x Nvidia (T4)	9.2 PB

Department 4 consists of a scientific department head and seven scientific staff members. The head of department is responsible for the coordination of the cloud activities, initially in consultation

with the other de.NBI cloud locations, but also with international ELIXIR and EU projects (e.g. EOSC-Life). To this end, the department management works closely with the cloud governance of the Administration Office and the Special Interest Group 6 - de.NBI-Cloud of the de.NBI network.

The cloud employees ensure the operation of the cloud and sufficient support of users. Here, four scientific staff members in the technical area look after the computing and storage resources as well as the network infrastructure for operating the bioinformatics cloud. This activity includes the maintenance of the hypervisor at operating system level as well as the Ceph storage infrastructure. Furthermore, they operate the openstack infrastructure and other necessary services (e.g. ELIXIR-AAI). The operation of the federated infrastructure requires coordination with the other de.NBI cloud sites via regular conference calls. Another two scientific staff members are responsible for the software area. They will (further) develop cloud-based scientific services for bioinformatics and offer special solutions based on OpenStack. This includes software developed for the simple handling of cloud resources (SimpleVM), as well as software for the simple deployment of a compute cluster in the cloud (BiBiGrid). In order to make it easier for both developers and users to get started in "cloud computing" environments, the employees regularly offer training courses or support other de.NBI partners in organizing courses in the de.NBI cloud area. In individual cases, cloud projects are supported by special consulting services. Support requests are also handled in daily operations.

In addition, the BMBF has created a further post for IT security and certification. The de.NBI cloud invests extraordinarily heavily in IT security and data protection measures in order to comply with the basic data protection regulation and to provide a secure environment for processing sensitive data. Here, the cloud sites are subjected to complex evaluation and testing processes. In the end, an internationally recognized certification is planned.

4.5 Microbial Bioinformatics (Department 5)

Department 5 "Microbial Bioinformatics" of the BIBI institute has two areas of responsibility. These are (1) the performance of service and training in the area as part of the de.NBI service center “BiGi”, and (2) the scientific expansion of the area "Microbial Bioinformatics".

To continue the tasks within the de.NBI project, two scientific staff members were hired. Their tasks include the continuous provision of software programs for the processing and analysis of data from high-throughput transcriptomics, proteomics, qualitative and quantitative metabolomics and metagenomics, as well as their maintenance and servicing. These tasks also include the updating of the software programs in response to technical innovations and improvements initiated by users. In addition, user inquiries are processed. At the same time, project-related individual consulting is provided, which starts with the virtual project setup and data transfer and can extend to the evaluation of post- and metagenome data.

To assess the quality of the bioinformatics services offered, de.NBI criteria and de.NBI measures were developed. The connection to the infrastructure (the tracking solution "Matomo", the survey platform "Surveymonkey" and the ticket system "Requesttracker") to record the individual criteria (KPIs, user satisfaction, user requests), as well as the maintenance of corresponding implementations, the aggregation of usage and user statistics and their provision also fall within this task area.

A central aspect of the de.NBI project is the training of researchers in the use of the offered software programs and workflows. The de.NBI network offers training courses in the key areas of genome analysis, polyomics data integration and metagenomics. Since the courses are mostly cloud-based, content on cloud computing for scalable high-throughput data processing is also provided, while at the same time enabling the provision of online course materials. The conception and implementation of the courses is part of the de.NBI tasks.

For the scientific expansion of the "Microbial Bioinformatics" department, two further positions for scientific staff are available. The basis for this is provided by the research topics located in Bielefeld, which originate from biotechnology, agriculture or the environment. In particular, microbial communities from biogas plants and agricultural soils are analyzed. It is planned to investigate the interaction of members of a microbial community and to describe novel non-culturable microorganisms. By integrating multi-omics data, i.e. metatranscriptome, metaproteome and metabolome data, complex questions regarding the interaction between members of microbial communities will be addressed. In the future, it is planned to analyze the human microbiome in cooperation with the newly founded medical faculty at Bielefeld University (Medical School OWL).

4.6 Graduate School DILS (Department 6)

The graduate school "Digital Infrastructure for the Life Sciences" (DILS) contributes significantly to the educational aspect, but also to the research profile of the institute. By training data scientists with a focus on the development of new bioinformatics methods and their sustainable provision as services, the graduate school supports the institute in establishing the new research profile "Service Science in the Life Sciences" on an international level.

The institute benefits not only from the trained data scientists, their research activities in the context of their doctorate and the newly developed methods. Interdisciplinary doctoral projects and the joint supervision of doctoral students by researchers from different disciplines within Bielefeld University (biology, biotechnology, bioinformatics, computer science) have a long and successful tradition. Furthermore, the DILS graduate school also offers a platform for initiating further bilateral research projects with ZB MED.

The faculty of the graduate school currently consists of eleven professors from Bielefeld University and three professors from ZB MED. In June 2019, the Strategy Fund of Bielefeld University granted three years of start-up funding for the graduate school in form of four PhD. In addition, four further

doctoral students are currently affiliated with the graduate school (see Table 2). Based on the size of the faculty, the long-term plan is to finance 12-16 PhD positions. In addition, an academic position is planned for the management of the graduate school.

Table 2: Overview of doctoral students at the DILS graduate school. Affiliated doctoral students are marked with [a]. All supervisors are members of the DILS faculty, except: [1]) Alexander Goesmann, Justus-Liebig University Giessen, Germany, [2]) Faraz Hach, Simon Fraser University, Burnaby/Vancouver, Canada.

Name	Ph.D Topic	Supervisor
Frey, Katharina	Computational pangenomics in plants	Weisshaar, Stoye
Jünemann, Sebastian[a]	Computational quality assessment of sequencing data	Stoye, Goesmann[1]
Osterholz, Benedikt[a]	Characterizing biogas microbiomes by meta analysis of metagenomes	Sczyrba, Schlüter
Rempel, Andreas	Software for computational pangenomics	Stoye, Förstner
Schulz, Tizian[a]	Applications of colored de Bruijn graphs	Stoye, Hach[2]
Sielemann, Janik	Machine learning approaches for the characterization of biological systems	Bräutigam, Hammer
Tubbesing, Tom	Bioinformatics solutions for microbiome meta-transcriptome analyses	Sczyrba, Schlüter
Wulf, Donat[a]	Machine learning based analysis of crop regulatory networks	Bräutigam, Sczyrba

The doctoral students not only benefit from the usual advantages of a structured doctorate (two supervisors, quality management, networking, etc.), but also from the general conditions offered by the institute. In addition to the embedding of the doctoral projects in an overarching scientific program with joint research focus, bioinformatics methods developed during the doctorate can also be promoted as de.NBI services. The institute provides the doctoral students with a forum for scientific exchange in the form of own workshops, seminars, training courses etc. The PhD program "Bioinformatics" of the Faculty of Technology represents the formal basis of the structured doctorate within the graduate school.

As a "Leibniz Graduate School", DILS complements the institute's plans for continued existance. The conditions for this are ideal, because the career guidelines formulated by the Leibniz Association are largely in line with practice at Bielefeld University: employment contracts as scientific staff members with appropriate contract terms, compatibility of work and family life, mentoring and further education opportunities for a career on and off the academic track, structured doctoral programs, etc. Furthermore, the "Bioinformatics" in Bielefeld looks back on a history of more than 20 years of doctoral training.

5. Financing of the Institute's staff posts for the year 2020

The personnel positions currently available in the various departments of the BIBI institute are listed in Fig. 2. A description of the tasks of the personnel positions listed in the figure has already

been made in Section 4 "Range of tasks in the departments". This section will now provide information on the origin of these posts and the costs incurred for the year 2020 (Table 3). The personnel positions are listed for each of the six individual departments of the BIBI institute.

Table 3: Staff positions in the six departments of the institute in 2020

Function	Financing of the job
Department 1:	
1 post for sci. head (W3 professorship)	Uni Bi/ZB MED
2 posts for sci. staff	Uni Bi/ZB MED
1 post for office administration	Uni Bi/ZB MED
Department 2:	
1 sci. post for executive secretary	BMBF (de.NBI)
1 post for management of the de.NBI network	BMBF (de.NBI)
4 sci. posts for coordination of • public relations • service tasks • training tasks • cloud governance	BMBF (de.NBI)
2 posts for community management	BMBF (de.NBI)
Department 3:	
1 sci. post for executive secretary	BMBF (de.NBI)
1 post for project management	BMBF (de.NBI)
5 posts for sci. staff for cooperation with ELIXIR platforms	BMBF (de.NBI)
Department 4:	
1 post for sci. department head	Uni Bi
2 posts for sci. staff (software)	Uni Bi / BMBF (de.NBI)
4 posts for sci. staff (hardware)	Uni Bi
1 post for cloud security and certification	BMBF (de.NBI)
Department 5:	
1 post for sci. department head	Uni Bi
4 posts for sci. staff	2x BMBF (de.NBI) / 2x Uni Bi
Department 6:	
1 post for sci. coordination	Uni Bi
8 posts for Ph.D. students	Uni Bi

The personnel costs calculated on the basis of Table 3 show that 45% of the personnel posts at the BIBI institute are financed by Bielefeld University. A further 46% of the posts are financed by the de.NBI/ELIXIR-DE project of the BMBF. 9 % of the posts are the result of mixed financing between Bielefeld University and ZB MED (Land NRW).