OXFORD

Briefings in Bioinformatics, 20(2), 2019, 370–374

doi: 10.1093/bib/bbx040
Advance Access Publication Date: 18 April 2017
Paper

Bioinformatics in Germany: toward a national-level infrastructure

Andreas Tauch and Arwa Al-Dilaimi

Corresponding author: Andreas Tauch, German Network for Bioinformatics Infrastructure (de.NBI), Center for Biotechnology (CeBiTec), Bielefeld University, Universitätsstraße 27, Bielefeld 33615, Germany. Tel.: +49 521 106-8703; Fax: +49 521 106-89046; E-mail: tauch@cebitec.uni-bielefeld.de

Abstract

The German Network for Bioinformatics Infrastructure (de.NBI) is a national initiative funded by the German Federal Ministry of Education and Research (BMBF). The mission of de.NBI is (i) to provide high-quality bioinformatics services to users in basic and applied life sciences research from academia, industry and biomedicine; (ii) to offer bioinformatics training to users in Germany and Europe through a wide range of workshops and courses; and (iii) to foster the cooperation of the German bioinformatics community with international network structures such as the European life-sciences Infrastructure for biological Information (ELIXIR). The network was launched by the BMBF in March 2015 and now includes 40 service projects operated by 30 project partners that are organized in eight service centers. The de.NBI staff develops further and maintains almost 100 bioinformatics services for the human, plant and microbial research fields and provides comprehensive training courses to support users with different expertise levels in bioinformatics. In the future, de.NBI will expand its activities to the European level, as the de.NBI consortium was assigned by the BMBF to establish and run the German node of ELIXIR.

Key words: de.NBI; distributed network; bioinformatics infrastructure; big data analysis; bioinformatics service; bioinformatics training

Introduction

In recent years, life sciences research underwent a rapid development that was driven mainly by technical improvements in analytical areas in terms of miniaturization, parallelization and high-throughput analysis of biological samples and the generation of huge amounts of experimental data ('big data') [1]. Prominent examples of this development are the so-called 'omics' technologies to analyze the various levels of information storage and processing in living cells [2, 3]. The ever-growing application of these technologies and the exploitation of the resulting data have revolutionized many fields of research [4, 5] and are opening considerable opportunities for life sciences and society [6, 7]. The recent progress in life sciences and the generation of huge sets of experimentally derived data were accompanied by an ongoing development in bioinformatics in terms of data archiving, processing, analysis, visualization, management and integration [2]. The availability of bioinformatics services, appropriate hardware and highly curated databases is a fundamental prerequisite for the solid exploitation of large volumes of data in all areas of life sciences. As the volume and the complexity of data are continually growing, current bioinformatics tools and services need a constant further development in terms of expansion of technical infrastructure, i.e. computing power and storage capacities, and of engineering of new tools and software solutions for life sciences research.

In 2012, the German Bioeconomy Council, an independent advisory body to the German Federal Government, published the recommendations 'Requirements for a Bioinformatics

Andreas Tauch is a professor at the Center for Biotechnology (Bielefeld University), the head of the de.NBI Administration Office and a deputy coordinator of the German Network for Bioinformatics Infrastructure (de.NBI).
Arwa Al-Dilaimi is a postdoc at the Center for Biotechnology (Bielefeld University), a member of the de.NBI Administration Office and a service coordinator of the German Network for Bioinformatics Infrastructure (de.NBI).
Submitted: 13 February 2017; **Received (in revised form):** 15 March 2017

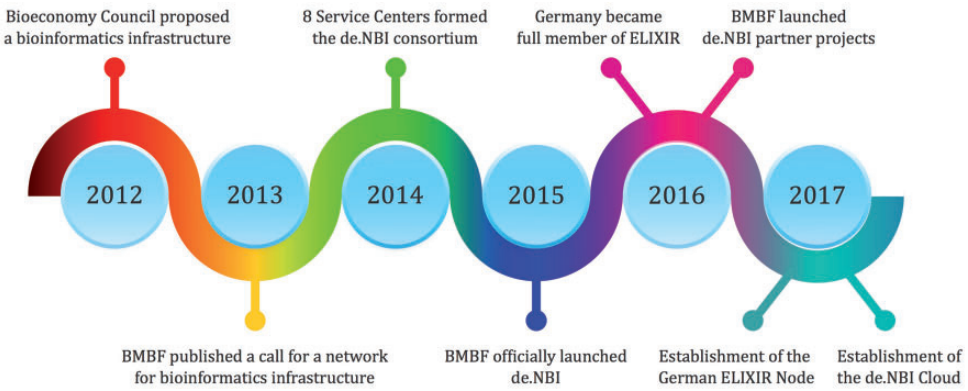

Figure 1. Time bar showing prominent steps in the development of the national infrastructure de.NBI that was initiated by the BMBF.

Infrastructure in Germany for future Research with bioeconomic Relevance', coming to the conclusion that a coordinated, networked bioinformatics infrastructure should be developed in Germany to ensure that biological data have the most efficient and sustainable use for research and industrial application [8]. The recommendations of the Bioeconomy Council were more precise regarding the bundling of technical know-how, as local expertise centers should ensure the development of bioinformatics approaches to solve specific scientific and big data problems. These expertise centers should already have an established reputation in their scientific field, should provide the necessary computing capacities and should have permanent structures for training users from life sciences. Supercomputing capacities should be set up by the infrastructure network for special applications and issues. In addition, a coordination board should act as an interface between the network's expertise centers, bioinformatics research institutes, users and other interest groups [8].

The recommendations of the Bioeconomy Council were translated by the German Federal Ministry of Education and Research (BMBF) into two announcements of funding guidelines to establish the German Network for Bioinformatics Infrastructure (de.NBI) (Figure 1) and to help in this way to achieve the Federal Government's objectives of the National Research Strategy BioEconomy 2030 [9]. The announcement of funding guidelines for a de.NBI (Deutsches Netzwerk für Bioinformatik-Infrastruktur—de.NBI) was published in May 2013. The aim of this announcement was to establish a national-level infrastructure, consisting of well-equipped and networked service centers with an appropriate coordination structure that will provide and continuously refine bioinformatics services [10]. The announcement of the funding guidelines for partner projects of the de.NBI was published in November 2015 to expand the expertise of the national infrastructure by linking new partners into the service centers [11]. In both cases, German public and private universities, nonuniversity research institutions and commercial enterprises were eligible to apply for funding. The current members of the service centers and the partner projects were selected from among the applicants after an international peer-review process. Funding for applicants not involved in commercial activities takes the form of nonrepayable grants awarded directly to each institution. The basis for calculating grants for commercial enterprises is formed by the eligible project-related costs, of which up to 50% are normally financed by the BMBF, depending on the relevance of the project.

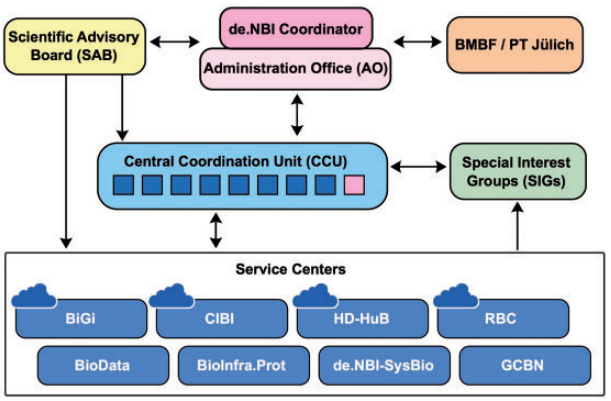

Figure 2. Organization chart of the de.NBI. The eight service centers of the network are shown by their acronyms: BiGi, CIBI, HD-HuB, RBC, BioData, BioInfra.Prot, de.NBI-SysBio, GCBN. Service centers contributing to the de.NBI cloud are specifically labeled. The de.NBI initiative is funded by the BMBF and managed by the Project Management Jülich (PT Jülich).

The de.NBI program was launched by the BMBF in March 2015, and the partner projects started their work in November 2016. The overall organization and coordination of the de.NBI were designed by the service centers in a conceptual phase that was funded by the BMBF from February to July 2014 (Figure 1). The resulting concept of the de.NBI underwent an additional international peer-review process.

The organization and the mission of the de.NBI are described in the following sections.

Organization of the de.NBI

The de.NBI is a distributed service and training network that was launched by the BMBF in March 2015. It now includes 40 service projects that are operated by 30 research institutes organized in eight service centers. The network is managed by the de.NBI coordinator and the team of the de.NBI Administration Office. The governance structure of the network is illustrated in the de.NBI organization chart (Figure 2).

The scientific core of de.NBI is the eight service centers that provide and develop further bioinformatics services according to their areas of expertise and their bioinformatics resources. The service centers complement each other in terms of thematic priorities and services offered to users in many areas of life sciences research (Figure 3). The services are complemented by comprehensive training activities of users and developers.

Each service center is composed of at least three scientific partners from different research organizations at various locations in Germany and is led by a service center coordinator.

The Central Coordination Unit is the decision-making body of de.NBI and responsible for the development of the network (Figure 2). The board of the Central Coordination Unit is composed of nine de.NBI members, including the de.NBI coordinator and the eight coordinators of the service centers. The Central Coordination Unit is on duty to control that the network activities comply with the mission to provide high-quality bioinformatics services and training to users in basic and applied life sciences research. The administrative and strategic work of the Central Coordination Unit is supported by Special Interest Groups. These subcommittees are small discussion groups of experts from all service centers and are established to focus on topics relevant for operational and strategic decisions by the Central Coordination Unit. Seven Special Interest Groups have been established so far to discuss the prominent topics: communication and outreach, service and service monitoring, training and education, infrastructure and data management, de.NBI development, de.NBI cloud (since 2016) and European life-sciences Infrastructure for biological Information (ELIXIR) cooperation (since 2017).

The International Scientific Advisory Board of de.NBI has been established by the BMBF to act annually as an independent consultative body and to provide recommendations to the BMBF, the Central Coordination Unit and the service centers on technical, organizational and strategic matters related to the de.NBI mission. The Scientific Advisory Board is currently constituted of six members who are internationally renowned scientists in bioinformatics.

Mission of the de.NBI

The core purpose and focus of the de.NBI is to build a sustainable national life sciences infrastructure

- to provide first-class bioinformatics services to users in basic and applied life sciences research from academia, industry and biomedicine;
- to offer bioinformatics training to users in Germany and Europe through a wide range of workshops and courses; and
- to foster the cooperation of the German bioinformatics community with international network structures such as the European life sciences infrastructure for biological information ELIXIR.

Mission 1: Bioinformatics services

de.NBI has been established by the BMBF as a distributed bioinformatics infrastructure, which builds on existing resources of the scientific partner organizations. According to the areas of expertise of the contributing project partners, de.NBI develops further and maintains almost 100 bioinformatics services for life scientists working in the human, microbial and plant research fields. Moreover, the de.NBI cloud will offer comprehensive compute resources to the research community in life sciences. An overview of prominent bioinformatics services provided by the eight service centers of de.NBI is illustrated in Figure 3.

The Bielefeld-Gießen Resource Center for Microbial Bioinformatics (BiGi) offers various services for the analysis of microbial genomic, metagenomic and metaproteomic data. The service center particularly addresses large-scale genomics and post-genomics data integration and exploitation in the microbial research field. The service center Bioinformatics for Proteomics (BioInfra.Prot) provides services regarding statistics and bioinformatics for proteomics and lipidomics to increase the performance of targeted experimental approaches. BioInfra.Prot also provides data standardization and conversion services for proteomics data sets to be submitted to public repositories. The Center for Integrative Bioinformatics (CIBI) brings together expertise in genome sequence informatics, computational mass spectrometry and workflow environments to provide life scientists with bioinformatics tools for proteomics, metabolomics and next-generation sequencing data analysis and with the workflow engine KNIME to integrate the respective tools into coherent solutions for reproducible multi-omics studies. The de.NBI Systems Biology Service Center (de.NBI-SysBio) offers standards-based management solutions for data and models in all kinds of systems biology projects. The German Crop BioGreenformatics Network (GCBN) focuses on bioinformatics for crop plant research and provides both tailored plant-specific data and infrastructure to the plant research and crop breeding community to improve gene and genome annotation and bridge genotypes with phenotypes. The Heidelberg Center for Human Bioinformatics (HD-HuB) provides numerous resources to efficiently evaluate highly complex sequencing data in human genetics and genomics to foster research in the area of metagenomics, such as human microbiome profiling and metagenome annotation, and to support applications in systematic phenotyping of human cells. The RNA Bioinformatics Center (RBC) supports all RNA-related research including (meta)transcriptome analysis, RNA structure analysis, prediction of noncoding RNA targets, definition and classification of RNA transcripts and the analysis of protein–RNA interactions. In addition, the Freiburg Galaxy Project is part of the service center RBC and a powerful central platform for RNA analysis.

The Center for Biological Data (BioData) is mentioned here as an exemplification in more detail. This service center consists of the four databases SILVA, PANGAEA, BacDive and BRENDA that provide highly curated reference data sets to the research community. SILVA provides access to and services for ribosomal RNA genes from all three domains of life [12]: PANGAEA is a library aimed at archiving, publishing and distributing georeferenced data from Earth System Research; BacDive merges detailed strain-linked information on different aspects of bacterial and archaeal biodiversity [13]; and BRENDA is a comprehensive enzyme and enzyme–ligand information system [14]. Further, the BioData services include a set of tools of the structure-based modeling support server ProteinsPlus that offers a rich functionality related to protein structures.

To ensure high-quality standards of the bioinformatics services provided by de.NBI partners and to collect and evaluate user feedback, de.NBI has established the Special Interest Group ‘Service and service monitoring’.

Mission 2: Bioinformatics training

de.NBI offers a portfolio of high-quality training courses to support life scientists with different expertise levels in bioinformatics and from various research fields. These training activities aim at enabling users to handle and interpret their biological (big) data more effectively by applying tools, standards and compute services provided by the de.NBI service centers. de.NBI also collects and hosts educational materials related to the bioinformatics services of the network. The training activities of de.NBI, ranging from basic skills to advanced data analysis, include 1–2 day training courses, webinars, online

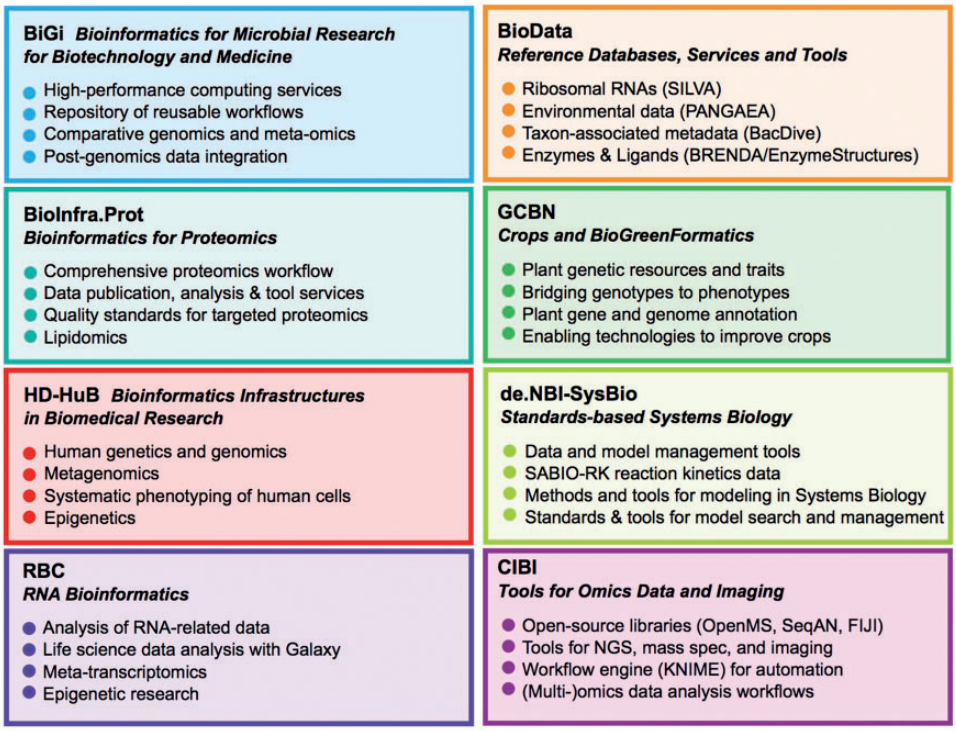

Figure 3. Prominent topics of bioinformatics services provided by the eight service centers of the de.NBI. A detailed list of services and associated training activities can be found at www.denbi.de.

training and themed 1 week summer schools. To effectively coordinate training courses and to plan new training activities, de.NBI has established the Special Interest Group 'Training and education' that is composed of training experts from each service center.

The training activities of de.NBI in the last quarter of 2015 reached already 329 participants of 16 training courses and of a late summer school in microbial bioinformatics. In 2016, the network partners organized 39 training courses with various bioinformatics topics and a themed summer school in bioinformatics for proteomics and metabolomics. In total, 882 participants attended these training activities of de.NBI and deepened their bioinformatics knowledge.

Mission 3: International collaboration

Since August 2016, Germany is full member of the European life sciences infrastructure for biological information ELIXIR (www.elixir-europe.org) [15], following the ratification of the ELIXIR consortium agreement by the BMBF. By joining the ELIXIR consortium, the BMBF will increase the visibility of the German bioinformatics infrastructure and will be able to take part and contribute to ELIXIR's governance and operations. The core mission of ELIXIR is to bring together European life science resources, including databases, software tools, training materials, cloud storage and supercomputers, and to coordinate these resources, so that they form a single infrastructure. Each ELIXIR member can establish a national node, which is usually a network of research organizations within a member country that run resources and services as a part of ELIXIR [15, 16]. Each ELIXIR node has a lead organization that coordinates and oversees the national ELIXIR activities.

The de.NBI has been designated by the BMBF to establish and run the German ELIXIR node and to provide bioinformatics services and training to the scientific user community in Europe. The lead organization of the German ELIXIR node will be Bielefeld University in Bielefeld (Germany). The next steps to further formalize the membership of Germany and the role of de.NBI in ELIXIR are the submission of the ELIXIR node application by the member state and the signing of the ELIXIR collaboration agreement by the lead organization.

The German life sciences community will particularly benefit from the international collaboration of de.NBI in terms of training via ELIXIR's training portal Training eSupport System (TeSS). TeSS is a central portal that provides researchers with comprehensive information about bioinformatics training courses and training materials aggregated from ELIXIR nodes. A parallel strategy to internationalize the training activities of de.NBI is to join the Global Organisation for Bioinformatics Learning, Education and Training [17]. The access to an international foundation experienced in developing standards and guidelines for successful courses and training materials will help to permanently improve the quality of bioinformatics training in Germany.

International collaboration of the de.NBI is not limited to ELIXIR and the ELIXIR nodes, as a strategic priority for the European Union is to be 'open to the world' and to foster international cooperation in research and innovation to tackle global challenges more effectively [18].

Key Points

- de.NBI was launched in March 2015 and includes 40 service projects run by 30 partners and organized in eight service centers that are located all over Germany.
- de.NBI develops further and maintains almost 100 bioinformatics services for the human, plant and microbial research fields.

- de.NBI provides bioinformatics training courses to support users with different expertise levels in bioinformatics and from various research fields in life sciences.
- de.NBI will establish and run the German Node of the intergovernmental organization ELIXIR that brings together life science resources from across Europe.

Funding

The German Federal Ministry of Education and Research (grant number 031A532B).

References

1. Frelinger JA. Big data, big opportunities, and big challenges. *J Investig Dermatol Symp Proc* 2015;**17**:33–5.
2. de Brevern AG, Meyniel JP, Fairhead C, *et al*. Trends in IT innovation to build a next generation bioinformatics solution to manage and analyse biological big data produced by NGS technologies. *Biomed Res Int* 2015;**2015**:904541.
3. Yu P, Lin W. Single-cell transcriptome study as big data. *Genomics Proteomics Bioinformatics* 2016;**14**:21–30.
4. Noor AM, Holmberg L, Gillett C, *et al*. Big data: the challenge for small research groups in the era of cancer genomics. *Br J Cancer* 2015;**113**:1405–12.
5. Popescu GV, Noutsos C, Popescu SC. Big data in plant science: resources and data mining tools for plant genomics and proteomics. *Methods Mol Biol* 2016;**1415**:533–47.
6. Vicini P, Fields O, Lai E, *et al*. Precision medicine in the age of big data: the present and future role of large-scale unbiased sequencing in drug discovery and development. *Clin Pharmacol Ther* 2016;**99**:198–207.
7. Alyass A, Turcotte M, Meyre D. From big data analysis to personalized medicine for all: challenges and opportunities. *BMC Med Genomics* 2015;**8**:33.
8. Bioeconomy Council. *Requirements for a Bioinformatics Infrastructure in Germany for Future Research with Bio-Economic Relevance*. Berlin: Bioeconomy Council, 2012.
9. Bundesministerium für Bildung und Forschung. *National Research Strategy BioEconomy 2030 - Our Route Towards a Biobased Economy*. Berlin: BMBF, 2011.
10. Federal Ministry of Education and Research. *Announcement of Funding Guidelines for a 'German Network for Bioinformatics Infrastructure'*. Berlin: BMBF, 2013.
11. Federal Ministry of Education and Research. *Announcement of the Funding Guidelines for Partner Projects of the German Network for Bioinformatics Infrastructure—de.NBI*. Berlin: BMBF, 2015.
12. Quast C, Pruesse E, Yilmaz P, *et al*. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013;**41**:D590–6.
13. Söhngen C, Bunk B, Podstawka A, *et al*. BacDive–the bacterial diversity metadatabase. *Nucleic Acids Res* 2014;**42**:D592–9.
14. Placzek S, Schomburg I, Chang A, *et al*. BRENDA in 2017: new perspectives and new tools in BRENDA. *Nucleic Acids Res* 2017;**45**:D380–8.
15. Crosswell LC, Thornton JM. ELIXIR: a distributed infrastructure for European biological data. *Trends Biotechnol* 2012;**30**:241–2.
16. Eijssen L, Evelo C, Kok R, *et al*. The Dutch Techcentre for Life Sciences: enabling data-intensive life science research in the Netherlands. *F1000Res* 2015;**4**:33.
17. Attwood TK, Bongcam-Rudloff E, Brazas ME, *et al*. GOBLET: the Global Organisation for Bioinformatics Learning, Education and Training. *PLoS Comput Biol* 2015;**11**:e1004143.
18. European Commission. *Open Innovation, Open Science, Open to the World—A Vision for Europe*. Brussels: European Commission, 2016.