

# BioDATEN - Bioinformatics DATa ENvironment

Jens Krüger

High Performance and Cloud Computing Group  
IT Center (ZDV), University of Tübingen  
Tübingen, Germany  
jens.krueger@uni-tuebingen.de

on behalf of the  
BioDaten Consortium

## ABSTRACT

The digitalization changes the working habits in almost all scientific disciplines with vast impact on teaching, research and interdisciplinary collaboration. The challenges faced in bioinformatics arising from vastly growing amounts of data are exemplary for a discipline establishing new ways of scientific work and insight.

Through the Bioinformatics DATa ENvironment (BioDATEN) [1] consortium, comprising of bioinformaticians, core facilities, compute centers and libraries, a science data center for life sciences for Baden-Württemberg is going to be established. The four-year project will start in summer 2019. We will report about the anticipated goals and the work plan intended to achieve them.

BioDATEN will use the technical infrastructure available through the project partners such as bwSFS, BinAC and de.NBI Cloud or the repositories provided by the university libraries in Konstanz and Tübingen. State-wide activities will be coordinated through the work group on research data management (<https://www.forschungsdaten.info/>). The consortium will establish the regulations for the preservation of and the access to well-annotated research data. The infrastructure and scientific methods for research data analysis will be further extended. Questions related to annotation and curation or the consistent collection of organizational and discipline-specific metadata will be addressed. A particular challenge will be the juridical and technical aspects arising for sensible data.

The data repositories hosting the well-annotated research data will form the basis for the research data infrastructure, using unified standards and procedures. As such, the access to data and infrastructure is democratized in a transparent way, improving the equality of opportunity for all researchers. It is anticipated to establish a science gateway for BioDATEN to offer a uniform point of access. This approach also facilitates the collaboration within national and international networks, necessary to be able to process the vast amounts of data with a bioinformatics context. The challenge to overcome lies not

only in the sheer size of bioinformatics data, but also in its diversity with respect to its characteristics and related research questions.

A central pillar of BioDATEN is the education of young academics, transferring knowledge about cutting-edge methods in data sciences complemented by the latest approaches in information technologies to form the basis for a solid research data management.

As BioDATEN follows a generic view on the underlying infrastructure other communities may use and adapt developments in particular the anticipated developments addressing the long-term access of research data.

To strengthen Baden-Württemberg as a key site for top-ranking research and excellence is a crucial goal for BioDATEN. A multi institutional network among the strongest bioinformatics research sites will form the backbone for the science data center.

The consortium comprises of the compute centers of the universities of Freiburg and Tübingen; the DKFZ, the EMBL and the university (BioQuant) in Heidelberg; the university libraries of Konstanz and Tübingen; the Quantitative Biology Center Tübingen; Bioinformatics and Plant Biotechnology Freiburg; CLARIN-D Center Tübingen; Livestock Microbial Ecology Hohenheim; Disease Ecology and Evolutionary Genetics Konstanz; Clinic of Neurology Ulm.

*Keywords—bioinformatics; science data center; whole data life cycle; research data infrastructure; reproducible science*

## REFERENCES

- [1] BioDATEN project page, accessed on 25 March 2019  
<[www.biodaten.info](http://www.biodaten.info)>