

## VENUE



Gästehäuser am Hauptbahnhof  
Lehrter Straße 68,  
10557 Berlin  
GERMANY



<http://www.jgh-hauptbahnhof.de/>

This workshop was initiated by SeqAhead, AllBio, and ISBE

We thank the COST Action SeqAhead and the CSA AllBio for financial support of the workshop

Managing Big Data  
Berlin Germany

The organising team:

Martijn J. Moné  
Susanne Hollmann  
Martin Golebiewski  
Babette Regierer

VU University Amsterdam (NL)  
University of Potsdam (D)  
HITS (D)  
LifeGlimmer GmbH (D)



Contact: (+49) 331-977-702811

[plantsys@uni-potsdam.de](mailto:plantsys@uni-potsdam.de)

# Managing Big Data

## SeqAhead Workshop

Setting the standards for analysing and integrating big data

## AllBio Tutorial

New methods in genome annotation

July 9-11 2014,  
Berlin - Germany



Modern life sciences are facing a rapidly increasing amount of data produced worldwide. The diversity and heterogeneity of the data calls for standardized formats for the data itself, its description and its context (the metadata), as well as for the processes during generation and processing of the data, the Standard Operating Procedures (SOPs). Standardization is important to enable researchers to exchange and integrate the data and resulting models and to relate corresponding data to each other.

The rapid development of modern technologies, especially the Next Generation Sequencing (NGS) technologies, allows the generation of biological data with increasing speed and precision. Standards and SOPs, therefore, become an increasingly critical issue specifically in large scale, distributed or applied approaches assuring that data can be accessed, integrated, shared, and compared. The development of quality control and standards are key drivers for efficient and successful large-scale research pipelines in the life sciences.

The workshop will bring together experimentalists and modellers, as well as representatives from national and international standardization bodies and data management experts. It will include discussions of the current state of standardization in the different fields of the life sciences. The aim is to identify overlaps, possible interoperability and interfacing options between the corresponding standards.

## SeqAhead



The next generation sequencing data analysis network SeqAhead is a COST Action established in 2011 to coordinate efforts in the area of bioinformatics dedicated to data produced by Next Generation Sequencing (NGS) technologies and develop concepts to address the problem of massively increasing amounts of data produced by the new technologies.

seqahead.eu

## ISBE



The mission of ISBE is to establish and enable access to an integrated, distributed infrastructure of state-of-the-art facilities and expertise for systems biology across Europe, with a view to transforming our understanding of the life sciences, human health and the environment. To this end ISBE will create a solid pan-European technological, intellectual and training basis and provide services allowing researchers in academia and industry to (i) collect, analyse and process data sets for integration into effective models, and (ii) build predictive models that give insight into the functioning of biological systems. The infrastructure will also serve as a platform from which to launch large-scale integrative projects of great scientific and societal impact.

project.isbe.eu

## AllBio



Handling and understanding vast bodies of biological data is only possible with the help of sophisticated mathematical analysis. A broad spectrum of bioinformatics tools has been developed across all areas of the life sciences. AllBio coordinates the efforts for the generalisation and the broadening of the applicability of the many human-centric bioinformatics facilities to other areas of the life sciences. Via "Test Cases" AllBio collects information about the existing needs and tries to find solutions together with researchers from the bioinformatics community. One central aim of the AllBio consortium is to transfer the knowledge of existing bioinformatics tools and web-services among the various life science areas and to identify still unsolved bioinformatics challenges.

allbioinformatics.eu

## Program

### Workshop SeqAhead

Setting the standards for analysing and integrating big data

#### Wednesday, July 9, 2014

**09:30-10:00** Welcome and Introduction  
**10:00-11:15** Session 1: Data Generation  
**11:15-11:45** Coffee Break  
**11:45-13:00** Session 2: Data Standards  
**13:00-14:00** Lunch  
**14:00-15:30** Session 3: Data Integration  
**15:30-16:00** Coffee Break  
**16:30-18:00** Session 4: Data Processing  
**18:00-18:30** Discussion and formation of groups for breakout sessions  
*Get together*

#### Thursday, July 10, 2014

**09:00-10:30** Breakout sessions  
**10:30-11:00** Coffee Break  
**11:00-13:30** Presentations of the results & Discussion of a roadmap  
Closing remarks

### Tutorial AllBio

New methods in genome annotation

Hands-on training in new computational methods in next generation sequencing and genome annotation

**15:00 - 18:00** Lectures

#### Friday, July 11, 2014

**09:00 - 17:00** Hands on training\*

- genome annotation automatic pipelines
- specific metabolic pathways gene discovery
- co-expression correlation network construction

\*please bring your own laptop with you