



# Data Management in Life Sciences

## Daten in den Lebenswissenschaften

Wolfgang Müller, HITS gGmbH, Heidelberg,  
Can Türker, Functional Genomics Center Zürich, Switzerland

When you hear about biological data, so-called high-throughput data come mostly to the fore. These are data obtained in experiments, which are executed in parallel through clever planning and automation. With this focus on high-throughput data and their specific data access and management problems (e.g., “How do I get my data in less than 10 hours from my machine to the Data Center”), it is easily overlooked that data processing in the life sciences tends to be very complex by nature. For example, the integration of life sciences data is a very big issue that is gaining interest. Biologists are trained to capture the diversity of life to perceive small differences, and respond to them. In contrast, computer scientists try to abstract away data differences as far as possible and to identify common structures.

This special issue focuses on problems and challenges of data management in the life sciences. The articles of this issue tackle some of the central problems and challenges from four different angles.

Prasser et al. illuminate the situation in translational medicine research, which concerns with the transfer (translation) of bio-medical research results into medical practice. The aim here is to establish a bidirectional information flow between the experimental setup and the patient's bedside. There are several obstacles to overcome. In addition to a variety of standards that are to be used skillfully, the privacy of patients' data has to be respected. Especially the latter brings a number of technical challenges.

While biological research is characterized in the previous article as “borderline chaotic”, managing the data of this boundary of chaos is the daily business of Bernd Rinn. His article distills the experiences he made with working on data management solutions for institution-crossing systems biology projects.

The article of Weise et al. presents Metacrop and discusses another important area of data management

described in the biomedical field: the curated data collections. These data collections provide an important interface between the user and the results of published scientific literature. Data curating in the context of data quality is coordinated and controlled manually.

Finally, the article of Kormeier et al. gives an insight into the area of biological data integration and data warehouses. It outlines data warehousing approaches that are widespread in the life sciences and bioinformatics, where mostly curated data sources are brought together.

Each article of this issue gives a glimpse of an interesting part of the data management area in the life sciences. Together they illustrate the diversity of research and engineering issues and show the differences between domain specific data management and the mainstream of data management.

PD Dr. Wolfgang Müller

Dr. Can Türker



**PD Dr. Wolfgang Müller** studied physics in Constance, Germany, Computer Science at the Fern-Universität Hagen and obtained a Diplom (master's degree) in both subjects. In 2001 he obtained a PhD about visual search from the Computer Vision and Multimedia Lab at the University of Geneva and habilitated 2008 about Information Retrieval at the University of Bamberg. Since 2009 he leads the Scientific Databases and Visualisation (SDBV) group at HITS, the privately funded, not-for-profit Heidelberg Institute for Theoretical Studies. SDBV is concerned with development, maintenance and operation of a curated data base of reaction kinetic data (SABIO-RK), as well as data management software in state-funded large systems-biology projects such as the transnational SysMO project and the German Virtual Liver Network.

Address: HITS gemeinnützige GmbH, Schloss-Wolfsbrunnenweg 35, D-69118 Heidelberg,  
e-mail: wolfgang.mueller@h-its.org



**Dr. Can Türker** heads the Data Integration Group of the Functional Genomics Center Zurich (FGCZ). Before joining FGCZ, he was a Post-Doc in the Database Group at ETH Zurich. Can Türker holds a diploma and Ph.D. degree in computer science from the Technical University of Darmstadt (Sept. 1994), and the University of Magdeburg (Oct. 1999), respectively. He is co-author of the several lecture books, among others "Object-Relational Databases" and "SQL:1999 & SQL:2003". His current research focuses on data and application integration in the field of systems biology. He is the chief architect of B-Fabric, an open source system infrastructure with applications on top, which is in daily use for managing all project and life sciences data generated at FGCZ.

Address: Functional Genomics Center Zürich, UNI/ETH Zürich, Winterthurerstraße 190, CH-8057 Zürich, Switzerland,  
e-mail: tuerker@fgcz.ethz.ch

